I T H E A®

Intelligent Data Processing in Global Monitoring for Environment and Security

Sofia – Kiev

2011

Krassimir Markov, Vitalii Velychko, Krassimira Ivanova, Iliya Mitov, Nataliya Bilous, Andrii Ostroushko, Andriy Bugriy, Benoit Depaire, Koen Vanhoof, Rumyana Dimitrova, Krassimira B. Ivanova, Anatolij Krissilov, Levon Aslanyan, Venko Bojilov, Pavel Burak, Natalia Ivanova, Olga Korobulina, Kristian Milenov, Pavel Milenov, Lyudmila Milenova, Radko Radkov, Hasmik Sahakyan, Anatolij Shutko, Oleksii Kravchenko, Nataliia Kussul, Andrii Shelestov, Sergii Skakun, Matthias Becker, Georgi Milev, Ivo Milev, Guenter Stangl, Keranka Vassileva, Igor Galelyuka, Volodymyr Hrusha, Nikola Kolev, Martin Nenov, Aleksandr Palagin, Volodymyr Romanov, Yevgeniya Sarakhan, Dmitri Rachkovskij, Elena Revunova, Michael Bondarenko, Victor Borisenko, Gleb Kobzar, Dmitry Makivsky, Roland Haarbrink, Vladimir Krapivin, Victor Krissilov, Eugenij Novichikhin, Igor Sidorov, Alexandr Kuzemin, Vyacheslav Lyashenko, Eugene Nickolov, Dimitrina Polimirova, Tatyana Stupina, Vladimir Kirilyuk, Vladimir Norkin, Alexandr Golodnikov, Pavel Knopov, Vladimir Pepelyaev, Elena Ilina, Victor Gladun. (Authors are ordered in accordance of appearing)

Intelligent Data Processing in Global Monitoring for Environment and Security

First Edition
ITHEA®, 2011, Sofia (Bulgaria) and Kiev (Ukraine).
ISBN: 978-954-16-0045-0 (printed)
ISBN: 978-954-16-0046-7 (CD/DVD)
ISBN: 978-954-16-0047-4 (online)
ITHEA® IBS ISC No.: 21
Includes 561 bibliographical references and indexes of 12 tables and 183 figures.

Editors: Krassimir Markov, Vitalii Velychko
Reviewers: Leonid Hulianytskyi, Volodymyr Opanasenko
The monograph is recommended for publication by the Scientific Councils of:

- *V.M.Glushkov Institute of Cybernetics* of National Academy of Sciences of Ukraine
- *International Institute of Information Theories and Applications FOI ITHEA*

This collective scientific monograph is aimed to present several important aspects of Intelligent Data Processing in Global Monitoring for Environment and Security, which are investigated by the authors. The implementing of the results in the corresponded program systems for intelligent data processing in GMES is outlined. It is represented that book chapters will be interesting for experts in the field of intelligent technologies for global observation as well as for practical users.

Printed in Bulgaria and Ukraine by:
ITHEA®, Sofia-1090, P.O.Box 775, Bulgaria. e-mail: info@foibg.com; URL: www.ithea.org
General Sponsor: Consortium FOI Bulgaria, URL: www.foibg.com

ISBN: 978-954-16-0045-0 (printed)
ISBN: 978-954-16-0046-7 (CD/DVD)
ISBN: 978-954-16-0047-4 (online)
C\o Jusautor, Sofia, 2011

Preface

This book contains a collective scientific monograph, which presents several important aspects of Intelligent Data Processing in Global Monitoring for Environment and Security. It is a result from two years collaborating of the authors in the frame of Special Interest Group on "Intelligent Data Processing" of the ITHEA International Scientific Society. Many of the results were approbated at the ITHEA International Conferences and were published in the ITHEA International Journals and Thematic Collections.

Global Monitoring for Environment and Security (GMES) is a joint initiative of the European Commission and European Space Agency, which aims at achieving an autonomous and operational Earth observation capacity. The key for operational GMES services is to have an appropriate governance and business model structure supporting these services. GMES is the European Union contribution to the Global Earth Observation System of Systems (GEOSS).

It is clear, the Global Observation is impossible without common scientific work between scientists from all over the world. This monograph is a good example of such collaboration. It unites sixty-two authors from nine countries: Armenia, Austria, Belgium, Bulgaria, Germany, Netherlands, Russia, Sweden, and Ukraine. Editors of the text are *Krassimir Markov* and *Vitalii Velychko*. Technical editors are *Krassimira Ivanova and Iliya Mitov*. A brief explanation of the content of the monograph is given below. The authors of the chapters are given in alphabetical order.

The book is structured in introduction, fifteen chapters collected in three parts, regarding ***practical***, ***theoretical,*** and ***technological aspects*** of Intelligent Data Processing in GMES, and conclusion. In addition, indexes of tables and figures, bibliography and authors information are given.

✓ Introduction

The introduction presents the United Nations initiative UN-SPIDER and European initiative GMES – observing our planet for a safer world, which are target area of implementation of results outlined in this monograph. Special attention is given to intelligent data processing in GMES and especially to data mining and, in particular, the association rule mining and class association rule classifiers. The introduction is written by *Benoit Depaire*, *Rumyana Dimitrova, Krassimira Ivanova, Krassimira B. Ivanova, Anatolij Krissilov, Krassimir Markov, Iliya Mitov, Koen Vanhoof,* and *Vitalii Velychko.*

✓ Part I. Practical Aspects

➢ Chapter 1. Managing Risk and Safety

Every year many collapses happen in the world, which can be divided into three groups: (1) Natural disasters, (2) Information disasters, and (3) Technogenic disasters.

Natural collapses are hurricanes, floods, earthquakes and other acts of nature, which destroy the whole cities and lead to mass deaths of people. Technogenic collapses are big accidents on industrial and transport objects which have been caused by failure in work of technical systems. Technogenic collapses are accompanied by victims among people and ecological disasters. Information collapses are collapses, which occur in information computer systems. Mainly they occur because of viruses and other harmful programs.

Several examples of these types of collapses and opinions of scientists concerning their reasons are outlined. Examples of different intelligent systems for prediction, risk analysis, and intrusion detection systems are given.

The authors of this chapter are *Levon Aslanyan*, *Venko Bojilov*, *Pavel Burak*, *Natalia Ivanova*, *Olga Korobulina*, *Anatolij Krissilov, Krassimir Markov, Kristian Milenov*, *Pavel Milenov*, *Lyudmila Milenova*, *Radko Radkov*, *Hasmik Sahakyan*, and *Anatolij Shutko*.

- **Chapter 2. High-performance Intelligent Computations for Environmental and Disaster Monitoring**

In this chapter, we present different approaches to multi-source data integration for the solution of complex applied problems, in particular flood mapping and vegetation state estimation using satellite, modeling, and in-situ data. Since these applications are data- and computation-intensive, we use Grid computing technologies. In such a case, computational and informational resources are geographically distributed and they may belong to different organizations. For this purpose, we also investigate benefits of different approaches to the integration of satellite-based monitoring systems.

This chapter is result of common work of *Oleksii Kravchenko*, *Nataliia Kussul*, *Andrii Shelestov*, and *Sergii Skakun*.

- **Chapter 3. Investigation of Geodynamics of Central and Eastern Europe, Balkan Peninsula and Bulgaria**

The investigations related to the global, regional and local geodynamics are particularly important taking into consideration the latest achievements of the science and engineering and as an element of the global monitoring of the Earth (for example Global Monitoring for Environment and Security – GMES, Europe). A number of projects and programs are realized or in process of realization within the framework of an international cooperation – CERGOP (Central European Regional Geodynamic Project), GGOS (Global geodetic Observing System), GMES, etc. The subject is the genesis, character, and development of the natural hazardous processes and their investigation and countermeasures or mitigation. The region of the Balkan Peninsula is one of the most active geodynamic regions not only within Europe. That is the reason for its intensive study. Within the international project CEGRN, the Balkan Peninsula and Central Europe (CE) are subjects of studies of several working groups. The results from the investigations accomplished by the CERGOP-2 project funded by the European Union are presented here. They are based on GNSS (Global Navigation Satellite Systems) campaigns, operation of GNSS permanent stations, data processing, interdisciplinary analysis and interpretation and respective generalizations presented as a special monograph for the Balkan Peninsula (BP). The existing geodynamic picture of BP is confirmed, supplemented, and expanded. Respective measures for reaction are proposed. Within these investigations, a considerable part is devoted to the territory of Bulgaria. Concrete results of them are given here. Additionally an overview is given concerning the movements of the permanent stations from different Balkan networks in the frame of CERGOP. The complex interdisciplinary significant investigations outline the contemporary approach of regional and national study of such a type.

This chapter is contribution of *Matthias Becker*, *Georgi Milev*, *Ivo Milev*, *Guenter Stangl*, and *Keranka Vassileva*.

- **Chapter 4. Intelligent Tools for Environment Monitoring: Features and Applications**

In this chapter, we present different approaches and tools for monitoring of environment. It is described Bulgarian and Ukrainian experience in development of measurement devices and an organization of express and continuous field electronic measurements for evaluation of main environment parameters.

The authors of this chapter are *Igor Galelyuka*, *Volodymyr Hrusha, Nikola Kolev*, *Martin Nenov*, *Aleksandr Palagin*, *Volodymyr Romanov, Yevgeniya Sarakhan*, and *Vitalii Velychko*.

- **Chapter 5. Intelligent Gamma-Ray Data Processing for Environmental Monitoring**

This chapter discusses machine-learning techniques for intelligent data processing in environmental and security-related monitoring employing gamma spectroscopy. We provide a concise survey of a diversity of methods and algorithms from the machine-learning arsenal applied to multichannel gamma spectroscopy and airborne gamma-ray spectrometry in order to improve the performance of the previous methods. Intelligent data processing techniques observed included sparse approximation, model selection, blind source separation, as well as the Tikhonov regularization and random projections. The chapter is enhanced with experimental results provided to illustrate the comparative performance of the discussed methods.

The contribution is given by *Dmitri Rachkovskij* and *Elena Revunova*.

- **Chapter 6. Acquisition, Processing and Analysis of Space Images at Risks Management of Natural and Technogenic Emergencies**

The chapter is devoted to one of possible approaches to automation in the field of the risk management, based on data processing of remote space monitoring of the spatially-distributed natural and technogenic objects for timely detection maintenance, diagnostics and the development predicate of the dangerous phenomena and emergencies. For acquisition of a digital image set for Earth's surface it is offered to use shooting from space satellites and unmanned aerial vehicles. We propose to eliminate a disadvantage inherent in standard unmanned aerial vehicles management schemes, which are associated with a limited range of management using digital imaging systems, included in the control loop. The visualization system synthesizes a three-dimensional image of cockpit-exterior space based on the unmanned aerial vehicles position and terrain. The set of mathematical methods and stage-by-stage procedures of computer processing of the space images are offered, allowing making a preliminary filtration, to estimate them information compatibility, to carry out qualitative recognition, fixing and tracing of artificial objects. The considered integrated automation means complex provides necessary reliability and quality of achieved results, and also differs high speed that allows to use it and for the analysis of situations in real time.

This chapter is result of common work of *Nataliya Bilous*, *Michael Bondarenko*, *Victor Borisenko*, *Andriy Bugriy*, *Gleb Kobzar*, *Dmitry Makivsky*, and *Andrii Ostroushko*.

✓ Part II. Theoretical Aspects

➢ Chapter 7. Elaboration of Geoinformation Regional Monitoring Environmental System ("GERMES-I") Enriched by Artificial Intelligence Instruments

In last decades and especially within the last few years, there was a steady growth of interest in joint use of both the new methods (and techniques) of advanced geophysical measurements and GIS-technologies, on one hand, and intellectual methods of information processing on the other hand in environment monitoring and territorial management. The chapter presents some leading ideas: use of effective methods and algorithms of estimation, comparison, generalization and decision making under uncertainty; use of contemporary methods of basic earth and water parameters measuring, specifically aerospace microwave radiometry; combining these effective means in the big Regional Monitoring System. Superposition of means mentioned above becomes an indispensable tool for the present when problems of Risk Collisions and Risk Assessment are to be solved.

The presented results reflect experience and common work of *Roland Haarbrink*, *Vladimir Krapivin*, *Anatolij Krissilov*, *Victor Krissilov*, *Eugenij Novichikhin*, *Anatolij Shutko*, and *Igor Sidorov*.

➢ Chapter 8. Microsituation Concept in GMES Decision Support Systems

A set of situation representation levels of subject domain naturally increases in emergent situations. This is connected with increased influence of subject domain factors on prototype system, independently of emergency conditions. Therefore, for emergent situations of possible snow avalanches microsituations could differ of small differences of temperature or changes in wind speed. Development of financial crisis as emergency economic situation is characterized by separate characteristics of microsituations of different facilities, involved in economical process.

Hence, situation includes a set of separate microsituations, each describing properties of prototype system in some characteristic category of its subject domain. Such categories for subject domain could represent internal and external processes, characteristics, advices from both prototype system's point of view and set of factors, which influence from subject domain to prototype system. Separate set of such characteristics without interconnections could hardly precisely describe a situation, because such characteristics are interconnected with each other, involved in different process of prototype system studying.

The presented results are structured and formed by *Alexandr Kuzemin* and *Vyacheslav Lyashenko*.

➢ Chapter 9. Methods and Means for Protection of Software Critical Infrastructures

In this chapter main methods of prevention, protection, and recovery of file objects, exposed to information attacks are discussed. In particular, analysis of different file objects, information attacks, and protection methods most frequently used in Software Critical Infrastructures is made. The analysis is made on the base of available information of National Laboratory of Computer Virology at the Bulgarian Academy of Sciences for accomplished attacks in Bulgaria, Balkan Peninsula, and Southeast Europe.

This overview is written by *Eugene Nickolov* and *Dimitrina Polimirova*.

- **Chapter 10. The MLRP-method for Analysis of Some Problems in Climate and Seismology**

An application of the heterogeneous variables system prediction method to solving the time series analysis problem with respect to the sample size is considered in this chapter. It is created a logical-and-probabilistic correlation from the logical decision function class. Two ways is considered. When the information about event is kept safe in the process, and when it is kept safe in depending process. The very actual problems from hydrological and seismology domains are presented here by MLRP-method.

The chapter represents main results of work and experience of *Tatyana Stupina*.

- **Chapter 11. Polyhedral Coherent Risk Measures and their Application to Investment Decisions Support under Catastrophic Flood Risks**

In this chapter we review some results of the PCRM theory, consider applications to decision making support in conditions of risk, and develop numerical methods for searching optimal decisions. An investment decisions making under catastrophic flood risks is considered as a particular application.

The results are presented by *Vladimir Kirilyuk* and *Vladimir Norkin*.

- **Chapter 12. Techniques for Robust Bayesian Estimation**

This chapter analyzes statistical techniques that are usually used for estimation of reliability parameters. It compares two main approaches: the classical sampling theory methods and the Bayesian approach. It is known that sampling theory methods are inappropriate for treating scarce data samples. In contrast to sampling theory methods, the Bayesian approach allows naturally to incorporate data from various sources in reliability parameters estimates by considering each source as a sample from the same population. However, the justification of a prior distribution frequently is a practical difficulty in the application of the Bayesian approach.

In cases when small datasets of past reliability data are available, it is desirable to estimate how far the calculated Bayesian estimate is from the true Bayesian estimate. Therefore, when only partial prior information is available it is necessary to search upper and lower bounds for Bayesian estimates which can be derived for any prior distribution satisfying the given partial prior information. The chapter considers models for searching such bounds.

The chapter is written by *Alexandr Golodnikov*, *Pavel Knopov*, and *Vladimir Pepelyaev*.

- **Chapter 13. Application of Information Theories to Safety of Nuclear Power Plants**

To this date, strategies aiming at a safe operation of nuclear power plants focused mainly on the prevention of technological breakdowns and, more recently, on the human attitudes and behaviors. New incidents and challenges to safety, however, motivated the nuclear community to look for a new safety approach. The solution became a strong focus on knowledge management and associated theories and sciences as information theories, artificial intelligence, informatics, etc. In all of these, the fundamental role is played by a category of information.

This chapter reviews a number of information interpretations and theories, among which of great relevance are those capturing the fundamental role information plays as a mean to exercise control on the state of a system, those analyzing information communication between agents involved in safety-related activities, and, finally, those, which explore the link between information and the limits of our knowledge. Quantitative measures of information content and value are introduced. Completeness, accuracy, and clarity are presented as attributes of information acquired by the receiver. To conclude, suggestions are offered on how to use interpretations and mathematical tools developed within the information theories to maintain and improve safety of nuclear power plants.

The text reflects main viewpoints of *Elena Ilina*.

✓ Part III. Technological Aspects

➢ Chapter 14. Growing Pyramidal Networks

The key enabler of increase of search operations efficiency is use of network structures for modeling environments in which problems solving. Orientation to real applied environments essentially raises a level of requirements to network models. The real environments, in which the problem-solving processes operate, have some typical features, such as multicoupling, heterogeneity, hierarchiness, and dynamism. In order to give proper representation of the examined processes, the network structures, used for representing the environment, must take into account these specifics. By their construction, Pyramidal Growing Networks, proposed by Prof. Victor Gladun, were created as answer of these requirements.

Methods for solution of regularities discovery tasks based on pyramidal networks, and methods of using of the retrieved regularities for decision-making are implemented in program complex CONFOR (Abbreviation of CONcept FORmation). In the case of decision-making in risk management, the described objects are assigned to specific disasters and/or emergency situations. This makes it possible to apply universal approach of growing pyramidal networks to analysis of attributive risk management and disaster emergencies.
This chapter is written by *Victor Gladun* and *Vitalii Velychko*.

➢ Chapter 15. Multi-dimensional Information Spaces as Memory Structures for Intelligent Data Processing in GMES

The advantages of multi-dimensional information spaces, used as a memory structures in the processes of data mining and knowledge discovery, is discussed.

Data mining analysis environment "PaGaNe" collects variety of algorithms for statistical analysis and association rule mining. The main focus in the realization is using the advantages of multi-dimensional information spaces, such as: the possibility to build growing space hierarchies of information elements; the great power for building interconnections between information elements stored in the information base; the possibility to change searching with direct addressing in well structured tasks.

The "class association rules" (CAR) algorithms have their important place in the family of classification algorithms. The advantages of associative classifiers can be highlighted in several very important directions, such as: very efficient training; possibility to deal with high dimensionality; no assumptions for the independence of attributes; very fast classification and the result is understandable by humans. The latest two advantages make CAR algorithms irreplaceable assistant in the processes of disaster risk management. Special attention is made on the description of MPGN classifier and its using in the field of disaster prediction.

The authors of this chapter are *Benoit Depaire*, *Krassimira Ivanova*, *Krassimir Markov*, *Iliya Mitov*, *Koen Vanhoof*, and *Vitalii Velychko*.

Acknowledgements

The idea to write this book was inspired from several initiatives of the Bulgarian Information Office for GMES, which role is to support this process through exchange of information, transfer of knowledge and good experience, promotion of initiatives, and establishing contacts between potential partners, services providers and users. Main events in this area are First and Second Workshops on GMES Operational Capacity – a joint initiative of the Bulgarian Government and the European Commission. The First Workshop was officially opened by the Prime Minister of Bulgaria – Mr. Boyko Borissov. The participants were also welcomed by a video address from European Commissioner for International Cooperation, Humanitarian Aid, and Crisis Response – Mrs. Kristalina Georgieva. More than 160 participants from about 15 EU Member-states and representatives from different EU organizations and institutions such as DG ENTR, DG ENV, DG ECHO, DG JRC, European Space Agency, European Environmental Agency, EUMETSAT, etc. took part in the event. The second GMES Operational Capacity Workshop has been planned to be in March 2011, again in Sofia, at the same high governmental level. This event is after time of publishing of this monograph. Because of this, we recommend readers to visit http://gmes-bg.org where one can find more information about the event as well as presentations, pictures, and movies.

The Bulgarian EO-GMES initiative aims to support the participation of Bulgaria and other member-states and candidate countries in the European Earth Observation Program – Global Monitoring for Environment and Security. The establishment of a single national point called Bulgarian Information Office for GMES (BIOG) concentrates information, advisory and support activities related to GMES in one institution, what should result in an increased efficiency and synergy.

The project focuses on the following priorities:

- support for participation of Bulgarian institutions, organizations in GMES thematic projects and introduction of institutions from other candidate countries into these projects;
- support in developing national and regional operational capacity in GMES services;
- support for growth of user segment of GMES-related projects;
- enhancing the active participation of the Community’s small and medium enterprises, especially those from the (new) member states still not so experienced in the space applications market;
- support to the realization of a special capacity building unit as part of the GMES program (including the space segment) in the (new) member states;
- clear distinction and different approach between environmental monitoring and risk and security management;
- support to the close cooperation between EU Program GMES and EU Directive INSPIRE and the use of reference information systems (data bases) for harmonization in the process of integration of the spatial data base;
- support for the establishment of a European Mediterranean Network for risk and security management

In addition, the activities of UN-SPIDER Regional Support Office (RSO) in Ukraine, which is established, based on the Space Research Institute of the National Academy of Sciences of Ukraine and the National Space Agency of Ukraine inspired the tide cooperation between authors of this monograph.

This work is partially supported by the General Sponsor Consortium FOI Bulgaria as well as by:

- Project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine.
- Project "Information Modeling" of the Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences;
- South-East Europe Transnational Cooperative programme under the project Monitor II: "Practical Use of MONITORing in Natural Disaster Management", 06.2009/05.2012;
- Joint project of the Science & Technology Center in Ukraine (STCU) and the National Academy of Sciences of Ukraine (NASU), "Grid Technologies for Multi-Source Data Integration" (No. 4928), and the Ministry of Education and Science of Ukraine, "Development of Integrated Remote Sensing Data Processing System using Grid Technologies" (No. M/72-2008);
- Bulgarian National Science Fund under the project D002-308/19.12.2008 "Automated Metadata Generating for e-Documents Specifications and Standards";
- Bulgarian National Science Fund under the joint Bulgarian-Ukrainian project D002-331/19.12.2008 "Developing of Distributed Virtual Laboratories Based on Advanced Access Methods for Smart Sensor System Design" as well as Ukrainian Ministry of Education under the joint Ukrainian-Bulgarian project No: 145/23.02.2009 with the same name;
- Ukrainian State Fund for Fundamental Researches under the joint Belarusian-Ukrainian project "Hybrid Methods and Technologies of Intelligent Data Analysis, Modeling and Forecasting in Corporate Information Systems of New Generation", 01.2011/12.2012;
- Ukrainian State Fund for Fundamental Researches under the joint Belarusian-Ukrainian project "Development of Theoretical Foundations of Pattern Recognition and Visual Information Processing Technology with Application in Decision Support Systems", 01.2011/12.2012.

We express our thanks to all authors, editors, and collaborators. Many of the presented in the book results were recognized by national or international awards, successful PhD projects, etc. For instance, let point the results presented in Chapter 4.3 awarded by the "Prize of the President of Ukraine for young researchers" for 2010 year.

We are grateful to ITHEA International Scientific Society; Association of Developers and Users of Intelligent Systems (ADUIS), Ukraine; ITHEA Institute of Information Theories and Applications and ITHEA Publishing House; V.M.Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine; Institute of Mathematics and Informatics at Bulgarian Academy of Sciences; Hasselt University, Belgium; for supporting the work on this monograph.

Special thanks to reviewers Dr.Sc., PhD, Leonid Hulianytskyi and Dr.Sc., PhD, Volodymyr Opanasenko; to the Scientific Council of the *International Institute of Information Theories and Applications FOI ITHEA;* and especially to the Scientific Council of the *V.M.Glushkov Institute of Cybernetics* of National Academy of Sciences of Ukraine (protocol No.:3/24.02.2011) for recommendation this monograph for publishing

It is represented that book chapters will be interesting for experts in the field of intelligent technologies for global observation as well as for practical users.

Krassimir Markov, Vitalii Velychko

Sofia-Kiev, February 2011

Table of Contents

Index of Tables

Index of Figures

Introduction

In modern age, the intelligent systems are finding wide applications in various fields like Micro-Electro-Mechanical Systems, Robotics, Manufacturing, Medical, Aerospace, Drives and Controls, Business Intelligence, etc. A very important area for applications of the intelligent systems is the World as well as European Earth observation programs. This monograph is closely connected to two of them:

- **UN-SPIDER** – The United Nations Platform for Space-based Information for Disaster Management and Emergency Response;
- **GMES** – European programme for Global Monitoring for Environment and Security, which provides data useful in a range of issues including climate change and citizen's security.

UN-SPIDER

In its resolution 61/110 of 14 December 2006, the United Nations General Assembly agreed to establish the "United Nations Platform for Space-based Information for Disaster Management and Emergency Response – UN-SPIDER" as a new United Nations programme, with the following mission statement:

"Ensure that all countries and international and regional organizations have access to and develop the capacity to use all types of space-based information to support the full disaster management cycle".

The United Nations Office for Outer Space Affairs (UNOOSA) implements the decisions of the General Assembly and of the Committee on the Peaceful Uses of Outer Space and its two Subcommittees, the Scientific and Technical Subcommittee and the Legal Subcommittee. The Office is responsible for promoting international cooperation in the peaceful uses of outer space, and assisting developing countries in using space science and technology.

UN-SPIDER is the first programme of its kind to focus on the need to ensure access to and use of space-based solutions during all phases of the disaster management cycle, including the risk reduction phase, which will significantly contribute to the reduction in the loss of lives and property.

Whereas, there have been a number of initiatives in recent years. They have contributed to making space technologies available for humanitarian and emergency response, UN-SPIDER is the first to focus on the need to ensure access to and use of such solutions during all phases of the disaster management cycle, including the risk reduction phase, which will significantly contribute to reducing the loss of lives and property.

The UN-SPIDER programme is achieving this by being a gateway to space information for disaster management support, by serving as a bridge to connect the disaster management and space communities and by being a facilitator of capacity-building and institutional strengthening, in particular for developing countries.

UN-SPIDER is being implemented as an open network of providers of space-based solutions to support disaster management activities. Besides Vienna (where UNOOSA is located), the programme also has an office in Bonn, Germany and will have an office in Beijing, China. Additionally, a network of Regional Support Offices multiplies the work of UN-SPIDER in the respective regions.

UN-SPIDER aims at providing universal access to all types of space-based information and services relevant to disaster management by being a gateway to space information for disaster management support; serving as a bridge to connect the disaster management and space communities; and being a facilitator of capacity-building and institutional strengthening.

UN-SPIDER ensures that all countries and international and regional organizations have access to and develop the capacity to use all types of space-based information to support the full disaster management cycle. UN-SPIDER is achieving this by being a gateway to space information for disaster management support; serving as a bridge to interlink the disaster management and space communities; and being a facilitator of capacity-building and institutional strengthening.

Capacity-building refers to the process of facilitating the strengthening of the competency of individuals, teams, and agencies to use space-based information to prevent, mitigate, and respond effectively to the challenges posed by natural hazards and related humanitarian crises.

The objective of the capacity-building efforts of UN-SPIDER is to ensure that countries recognize the value of all types of space-based information, and therefore access it to reduce the impacts of disasters and to respond more efficiently in case of such disasters through improved use of this type of information.

To increase capacities regarding the use of space-based information for disaster-risk management and emergency response at the national level, UN-SPIDER pursues three lines of action:

- Institutions: UN-SPIDER will promote the adoption of policies, which ensure that operating procedures within institutions incorporate the use of space-based information, and that they support activities targeting all phases of the disaster management cycle.
- Individuals: UN-SPIDER facilitates access to training programs conducted by Centers of Excellence and Specialized training centers to enhance the knowledge and skills of staff working in institutions, which conduct activities targeting all phases of the disaster management cycle.
- Infrastructure: UN-SPIDER facilitates access to infrastructure (hardware, software, related equipment) to ensure the capacity to access and make use of space based information.
- The respective activities are conducted in a systematic fashion to ensure that capacities within agencies are strengthened. This is coordinated with the network of Regional Support Offices and the National Focal Points.

Curricula are being elaborated targeting both disaster-risk management and emergency response in order to design training courses to be conducted through the Regional Centers for Space Science and Technology Education affiliated to the United Nations, Centers of Excellence, UN training centers linked to UN-SPIDER, and other national or regional training centers where remote sensing and earth observation applications are taught. Complementary efforts are being made to establish an e-learning environment as a means to build a repository of learning material, and information is being compiled to provide a database of training opportunities.

More information for UN-SPIDER is given at URL: [UN-SPIDER, 2010].

Regional Support Office of UN-SPIDER in Ukraine

UN-SPIDER Regional Support Office (RSO) in Ukraine was established based on the Space Research Institute of the National Academy of Sciences of Ukraine and the National Space Agency of Ukraine.

The Institute signed the cooperation agreement with the United Nations Office of Outer space Affairs (UNOOSA) during the forty-seventh session of the Scientific and Technical Subcommittee of the Committee on the Peaceful Uses of Outer Space (COPUOS) on 10 February 2010 in Vienna, Austria.

UN-SPIDER aims at providing universal access to all types of space-based information and services relevant to disaster management

UN-SPIDER RSO is a regional or national centre of expertise that is set up within an existing entity by a Member State or group of Member States that has put forward an offer to set up and fund the proposed regional support office.

Activities of UN-SPIDER RSO in Ukraine:

- horizontal cooperation;
- outreach and capacity building;
- technical advisory support.

The Space Research Institute of the National Academy of Sciences of Ukraine and the National Space Agency of Ukraine (SRI NASU-NSAU) and the United Nations Office of Outer space Affairs (UNOOSA) signed the cooperation agreement on the establishment of a UN-SPIDER Regional Support Office (RSO) on the 47th Scientific and Technical Subcommittee sessions on 12 February 2010.

The Space Research Institute was established in 1996 at the National Space Agency of Ukraine and National Academy of Sciences of Ukraine for organization of scientific space researches in the country, conducting, and co-ordination of scientific and engineering activities in the area of peaceful exploration and use of outer space.

The Institute's main activities are:

- Pure and applied research in outer space, astrophysics research of objects in the universe, including in ranges unavailable from the earth surface;
- Development of strategy and principles of universe exploration means use in solving scientific and applied issues for the needs of the economy;
- Development and testing, in the space environment, of scientific space exploration equipment and relevant technological processes;
- Development of new spacecraft navigation and control systems and earth and space monitoring systems; improvement of existing ones; creation of information space systems;
- Working out suggestions on the conception and strategy for space programs

The Institute's main projects include:

- use of earth remote sensing and geographic information systems (GIS) for informational support of environmental control;
- monitoring, estimation, and forecasting of underwater petrochemical pollution in Ukraine;
- interball project on exploration of solar-earth relations;
- variant project on measurement of electromagnetic field and electric current flows in ionosphere;
- warning project: satellite complex for exploration of ionosphere phenomena related to seismic activity;

– planning and controlling system for science and engineering experimentation aboard the Ukrainian explorer unit of the International Space Station.

For additional information, please visit URL: [SRI NASU-NSAU, 2010].

GMES – observing our planet for a safer world

Land, sea, and atmosphere – each Earth component is observed through GMES, helping to make our lives safer. The purpose of GMES is to deliver information, which corresponds to user needs. The processing and dissemination of this information is carried out within the "GMES service component" [GMES, 2010].

The thematic areas within the GMES service component comprise:

– land, marine and atmosphere information – ensuring systematic monitoring and forecasting the state of the Earth's subsystems at regional and global levels;
– climate change information – helping to monitor the effects of climate change, assessing mitigation measures and contributing to the knowledge base for adaptation policies and investments;
– emergency and security information – providing support in the event of emergency and humanitarian aid needs, in particular to civil protection authorities, also to produce accurate information on security related aspects (e.g. maritime surveillance, border control, global stability, etc.).

Managing natural resources and biodiversity, adapting to sea level rise, monitoring the chemical composition of our atmosphere: all depend on accurate information delivered in time to make a difference.

The GMES service component depends on Earth observation data, collected from space (satellites), air (airborne instruments, balloons to record stratosphere data, etc.), water (floats, shipboard instruments, etc.) or land (measuring stations, seismographs, etc.). These facilities are called the GMES infrastructure component; non-space based installations in the GMES infrastructure component are generally referred to as "in situ component".

By securing the sustainability of an information infrastructure necessary to produce output information in the form of maps, datasets, reports, targeted alerts, etc., GMES helps people and organizations to take action, make appropriate policy decisions and decide on necessary investments. GMES also represents a great potential for businesses in the services market, which will be able to make use of the data and information it provides according a full an open access principle.

Earth observation-based services already exist in Europe, but they are dispersed at national or regional level. Because of this, they cannot rely on a sustainable observation capacity. With the exception of meteorological services, long-term availability and reliability of information is not guaranteed. This is why, in order to contribute to improve its response to ever-growing challenges of global safety and climate change, Europe develops a sustained and reliable Earth observation system of its own [GMES, 2010]. Let remark that one of the priority topics to be discussed on the second GMES Operational Capacity Workshop is planned to be in March 2011 again in Sofia is the Bulgarian initiative for regional networks for integrated risk and security management with main targets:

– strengthening and regional cohesion of GMES operational capacity in Europe, taking into account local and regional specific characteristics and capacities;

- space and in-situ infrastructure policy, taking into account a bottom-up approach – attention on "user" needs for spatial data, collected under the "space" and "in-situ" GMES components as well as better data access and efficiency in their practical use, without losing high standards of data security;
- regional "User" driven approach in defining and developing services in support for better prevention, better forecast and transborder/transnational preliminary measures, reducing nature and anthropogenic crisis and disasters impact.
- risk and security prevention based on regional network of servers and service providers, thus providing flexibility in the use of infrastructure and human power depending on the area of impact, complexity and importance of the problem.

Intelligent data processing in GMES

Let consider a short example.

If one needs to know what is the time, he or she may ask someone and will know "the local" time. However, if one wants to know what time is pointed by all watches in the town or in the country, it is impossible to solve such problem because of the high number of devices – it is impossible in the same moment to scan and receive all information. During processing one part of watches, the other will show new time. So, one will never know "the total" time, i.e. what time is on all watches in the world. (Let make difference with "the global time" because there is a convention all watches to be synchronized with the standardized Greenwich Mean Time (GMT). It is the same all year around – there is no Summer Time, i.e. "Daylight Saving Time".)

In other words, along with modern technologies appears a lot of information. Often, when we want to collect and understand it, the traditional methods are not enough. They might be too general and they need additional processing and transformation. Therefore, we use advanced methods of data processing such as rough sets, genetic algorithms, neural nets, or fuzzy sets. Rough sets have application e.g. in problems of excessive data, problems with correct classification or problems with retrieval hidden relations between data. Placing these methods in environment of distributed applications, based on. Net platform and XML Web Services, introduces new possibilities in usage of artificial intelligence methods in data processing systems [Zielosko and Wakulicz-Deja, 2005].

Technological advancement using intelligent techniques has provided solutions to many applications in diverse engineering disciplines. In application areas such as web mining, image processing, medical, and robotics, just one intelligent data processing technique may be inadequate for handling a task, and a combination or hybrid of intelligent data processing techniques becomes necessary. The sharp increase in activities in the development of innovative intelligent data processing technologies also attracted the interest of many researchers in applying intelligent data processing techniques in other application domains [Wai-Wong et al, 2007].

A briefly classification of the existing methods of data processing by the complexity of results that we want to achieve is given in [Nguyen et al, 1997]:

- in some problems, all we want is one or several numerical values. It may be that we measure some characteristics, or it may be that we know the model, and we want, based on the experimental data, to estimate the parameters of this model. These problems are usually handled by statistical methods;

- in other problems, we want to know a function. We may want to reconstruct an image (brightness as function of coordinates), we may want to filter signal (intensity as a function of time), etc. These methods are usually handled by different regularization techniques;
- finally, there are many complicated problems where we want to reconstruct a model of an analyzed system. Methods that handle such problems are called **intelligent data processing methods**. Many of these methods are based on logic programming, a formalism that (successfully) describes complicated logical statements algorithmically, in a kind of programming language terms. Other methods are based on advanced information modeling technologies, which use rough sets, genetic algorithms, neural nets, fuzzy sets, etc. Very important case is the methods of data mining.

Data mining in GMES

Data mining technologies have important place in Geosciences. Here we will remember only the main areas of data mining. For further readings see, for instance, [Ramachandran et al, 2006].

Over the past few centuries, the quantity of accumulated information is constantly growing. Because of rapid development of all areas of human activity in modern society, production, economic and social processes became greatly complicate. Most companies, which use information technology resources, collect and store large amounts of data. The challenge that most companies face today is not how to collect and store adequate amounts of data but how to derive meaningful conclusions from this mass information. The answer is in technology of data mining and, in particular, the association rules. Association rules can be classified as data based set of rules that are similar to expert system. They show worthy conditions – attributes that often occur together in a data set. Association rules provide information of the "if-then" structure. Those rules are calculated from the data, and unlike in the last rules of logic, association rules are probabilistic in nature. The first part ("if") is called support of the rule. The second part ("then") is like confidence of this rule. In data mining, association rules are useful for analyzing and predicting the behavior of customer, system, natural processes, etc. Developers use association rules to build programs capable of machine learning. Machine learning is a form of artificial intelligence that seeks to build programs with the opportunity to become more efficient without being explicitly programmed.

Data Mining binds the terms "knowledge discovery in database" and "intellectual analysis of data". The emergence of all these terms is associated with the emergence of new trends in development of tools and methods for data processing. Data mining is a process of analysis of stored databases in the direction of mining new useful information by revealing deep and hidden relationships between seemingly unrelated and unknown to each other values. An important feature of his is that it enables processing multi-dimensional arrays and retrieval of multidimensional relationships while automatically reveals exceptional situations – the information, which is not included in the general laws. Data mining analysis automatically makes assumptions to detect relationships between different components and parameters. The work of analysts who deal with these systems is limited to verification and clarification of the resulting hypotheses. The emergence of data mining is associated with the need to improve techniques for recording and storing data that summarize the work of thousands of people in huge flows of information in various fields. By the time, it became clear that without productive data processing, redundant information is generated. The need of the development of modern technologies of processing such data can be summarized as follows: unlimited data volume, variety, and heterogeneity of data (quantitative, qualitative, and textual); need for concrete and understandable results; processing tools providing data easy use.

Data mining derives its name from the similarities between searching for valuable information in a large database — for example, finding linked products in gigabytes of store scanner data — and mining a mountain for a vein of valuable ore. Both processes require either sifting through an immense amount of material, or intelligently probing it to find exactly where the value resides. Given databases of sufficient size and quality, data mining technology can generate new opportunities by providing these capabilities [Davenport and Harris, 2007]:

1. ***Automated prediction of trends and behaviors***. Data mining automates the process of finding predictive information in large databases. Questions that traditionally required extensive hands-on analysis can now be answered directly from the data — quickly. Main goal is forecasting. It is a method to reveal the models in the relationship between data that are needed for forecasting and modeling. This type of algorithms prepares input and output data for the preparation of illustrative examples. A typical example of a predictive problem is targeted marketing. Data mining uses data on past promotional mailings to identify the targets most likely to maximize return on investment in future mailings. Other predictive problems include forecasting bankruptcy and other forms of default, and identifying segments of a population likely to respond similarly to given events. In GMES, the forecasting of disasters and collisions is one of the main tasks. The data mining methods may help to predict possible risk by analyzing large databases of satellite, in-situ, and other information.
2. ***Automated discovery of previously unknown patterns***. Data mining tools sweep through databases and identify previously hidden patterns in one-step. An example of pattern discovery is the analysis of retail sales data to identify seemingly unrelated products that are often purchased together. Other pattern discovery problems include detecting fraudulent credit card transactions and identifying anomalous data that could represent data entry keying errors. Very important is the possibility for segmentation – analysis of existing data form specific groups, based primarily on the parameters of the data, for instance, of the customer – demographics and purchasing power. Clustering algorithms make it possible to delineate homogeneous groups or types of customers for each group can determine the intrinsic and group clients. This enables us better to evaluate its customer base and planning of marketing activities. In GMES, clustering the prime causes of collapses and disasters is important to find regularities in the data and this way to recognize properly the risk situation.

Two technologies are in the base of data mining: machine learning and visualization.

The quality of visualization defines graphic representation of data in their colors, shapes and other elements representing the hidden data relationships.

The effectiveness of methods of machine learning is determined by the capabilities of data mining to explore the much larger amounts of data to reveal connections between them than a person.

The most commonly used techniques in data mining are [Davenport and Harris, 2007]:

- ***Artificial neural networks***: Non-linear predictive models that learn through training and resemble biological neural networks in structure. The knowledge is presented in the form of links, joining a set of conditions. The strength of the relationship is determined by the relationship between factors and data;
- ***Decision trees***: Tree-shaped structures that represent sets of decisions. It is designed to classify data using the gravity of the partition coefficients of elements of data in ever smaller and smaller groups. The decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID);

- ***Genetic algorithms***: Optimization techniques that use processes, such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution. They define natural "breakdown" of the data based on the target variables. Each branch of the tree is a separate part of the rules;
- ***Nearest neighbor method***: A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where $k \geq 1$). Sometimes it is called the k-nearest neighbor technique;
- ***Rule induction***: The extraction of useful if-then rules from data, based on statistical significance. Such rules can be generated using the process for requesting and checking various combinations of rules or extraction of any of the decision tree.

How exactly could data mining foresee important things, which are still not known or which may occur later? The answer to this question relates to the technique used for data mining, namely modeling. ***Modeling*** is the act of creating a model of a situation based on existing experience and knowledge and understanding of the response to this situation. This model is then applied to another situation in which the answers are not known [Davenport and Harris, 2007]. This act of building models is something that people made long ago, certainly before the advent of computers and technology to retrieve data. What happens with computers is not very different from the way people build models. In computing, a set range of information about different situations in which knowing the answer, the software for data mining should be extended to these data and derive the characteristics of those data should appear in the model. Once the model is built, it can be used in similar situations where looking for unknown response. While data mining represents a significant progress in analytical tools that are available now, there are limitations on its capabilities. One of the limitations is that although data mining can help to identify patterns and relationships, it reveals what is the value and importance of these models. The user alone must make these distinctions. The second limitation is that although data mining can identify connections between behavior and/or variables, it does not necessarily identify causality. To be successful, data mining requires technical and analytical specialists who can structure the analysis and interpret search results [Davenport and Harris, 2007]. The associative rules represent causal links and determine the likelihood or the coefficient of reliability, allowing drawing appropriate conclusions. Rules presented in the form “IF <condition> THEN <conclusion>” is used for forecasting and evaluating the unknown parameters and meanings.

Data mining stands at the crossroad of databases, artificial intelligence, and machine learning. Association rule mining is a popular and well-researched method for discovering interesting rules from large collections of data. The contemporary databases are very large, reaching giga- and terabytes, and the trend shows further increase. Therefore, for finding association rules one requires efficient scalable algorithms that solve the problem in a reasonable time. The efficiency of frequent item set mining algorithms is determined mainly by three factors: (1) the way candidates are generated; (2) the data structure that is used; and (3) the implementation details. Most investigations focus on the first factor, some describe the underlying data structures, and implementation details are almost always neglected [Bodon, 2003].

The main pillar of association rule mining algorithms is Apriori [Agrawal and Srikant, 1994]. It is the best-known algorithm to mine association rules, which uses a breadth-first search strategy to count the support of item sets and uses a candidate generation function, which exploits the downward closure property of support. Over the years, many improvements of Apriori, supported with different types of memory structures, are proposed.

Recent association rule mining algorithms, based on graph mining can be roughly classified into two categories. The first category of algorithms employs a breadth-first strategy. Representative algorithms in this category include AGM [Inokuchi et al, 2003] and FSG [Kuramochi and Karypis, 2001]. AGM finds all frequent induced sub-graphs with a vertex-growth strategy. FSG, on the other hand, finds all frequent connected sub-graphs based on an edge-growth strategy. Algorithms in the second category use a depth-first search for finding candidate frequent sub-graphs. A typical algorithm in this category is gSpan [Yan and Han, 2002], which was reported to outperform both AGM and FSG in terms of computation time.

A different approach for association rule searching is used in ECLAT (Equivalence Class Clustering and Bottom-up Lattice Traversal) [Zaki et al, 1997]. It is the first algorithm that uses a vertical data (inverted) layout. The frequent item sets are determined using sets of intersections in a depth-first graph.

In graph approaches, the bottleneck is the necessity of performing many graph isomorphism tests. To overcome this problem, alternative approaches use hash-based techniques for candidate generation. The representatives in this direction are DHP [Park et al, 1995] based on direct hashing and pruning, [Özel, and Güvenir, 2001] which proposed the use of perfect hashing, and IHP [Holt and Chung, 2002] that uses inverted hashing and pruning.

FP-Tree [Han and Pei, 2000], Frequent Pattern Mining is another milestone in the development of association rule mining, which breaks the main bottlenecks of the Apriori. FP-tree is an extended prefix-tree structure storing quantitative information about frequent patterns. The tree nodes are arranged in such a way that more frequently occurring nodes will have better chances of sharing nodes than less frequently occurring ones. The efficiency of FP-Tree algorithm has three reasons: (1) FP-Tree is a compressed representation of the original database; (2) it only scans the database twice; (3) it uses a divide and conquers method that considerably reduces the size of the subsequent conditional FP-Tree. The limitation of FP-Tree is its difficultness to be used in an interactive mining system, when a user wants to expand the dataset or change the threshold of support. Such changes lead to repetition of the completely mining process.

The Hmine algorithm [Pei et al, 2001] introduces the concept of hyperlinked data structure "Hyper-structure" and uses it to adjust dynamically links in the mining process. Hyper-structure is an array-based structure. Each node in a Hyper structure stores three pieces of information: an item, a pointer pointing to the next item in the same transaction and a pointer pointing to the same item in another transaction.

The innovation brought by TreeProjection [Agarwal et al, 2001] is the use of a lexicographical tree, which requires substantially less memory than a hash tree. The number of nodes in its lexicographic tree is exactly that of the frequent item sets. The support of the frequent item sets is counted by projecting the transactions onto the nodes of this tree. This improves the performance of counting the number of transactions that have frequent item sets. The lexicographical tree is traversed in a top-down fashion. The efficiency of TreeProjection can be explained by two main factors: (1) the transaction projection limits the support counting in a relatively small space; and (2) the lexicographical tree facilitates the management and counting of candidates and provides the flexibility of picking efficient strategy during the tree generation and transaction projection phrases.

Another data structure that is commonly used is a "trie" (or prefix-tree). Concerning speed, memory need, and sensitivity of parameters, tries were proven to outperform hash-trees [Bodon and Ronyai, 2003]. In a trie, every node stores the last item in the item set it represents, its support, and its branches. The branches of a node can be implemented using several data structures such as a hash table, a binary search tree, or a vector.

Another algorithm for efficiently generating large frequent candidate sets, which use different data structures, is proposed by [Yuan and Huang, 2005] and is called the Matrix Algorithm. The algorithm generates a matrix with entries 1 or 0 by passing over the cruel database only once, and then the frequent candidate sets are obtained from the resulting matrix. Finally, association rules are mined from the frequent candidate sets. Experiment results confirm that the proposed algorithm is more effective than the Apriori Algorithm is.

An interesting theoretical framework for association rule mining algorithm is brought in [Aslanyan and Sahakyan, 2009] that combines the known "unit cube chain decomposition structures" introduced in [Hansel, 1966] and [Tonoyan, 1976] into the frequent itemset generation algorithm. [Hansel, 1966] established the chain split theory. [Tonoyan, 1976] invented an excellent chain computation framework which brings chain split into the practical domain. Chain Split Data Mining integrates these technologies around the rule mining procedures. Effectiveness of this approach is related to the low complexity of rules mined, which is a usual case for many applications. Complexity of the procedure composed is complementary to the known Apriori family of algorithms.

In general, the "Affinity Analysis" (or association rules) technique discovers which events are likely to occur together. It is a data analysis and data mining technique that discovers co-occurrence relationships among natural events or, human activities performed by (or recorded about) specific individuals or groups. In general, this can be applied to any process where agents can be uniquely identified and information about their activities can be recorded [Shen et al, 2005].

The "class association rules" (CAR) algorithms have its important place in the family of classification algorithms. As it is mentioned in [Zaïane and Antonie, 2005], the advantages of associative classifiers can be highlighted in four major ones:

- the training is very efficient regardless of the size of the training set;
- training sets with high dimensionality can be handled with ease and no assumptions are made on dependence or independence of attributes;
- the classification is very fast;
- the classification model is a set of rules easily understandable by humans and can be edited.

The first associative classifier CBA was introduced by [Liu et al, 1998]. During the next decade, various other associative classifiers were introduced, such as CMAR [Li et al, 2001], ARC-AC and ARC-BC [Zaïane and Antonie, 2002], CPAR [Yin and Han, 2003], CorClass [Zimmermann and De Raedt, 2004], ACRI [Rak et al, 2005], Arubas [Depaire et al, 2008], etc.

The idea in CAR-algorithms is relatively simple. Given a training set with transactions where each transaction contains all features of an object in addition to the class label of the object, the association rules are constructed, which have as consequent a class label. Such association rules are named "class association rules" (CARs). Generally, the structure of CAR-algorithms consists of two major data mining steps:

(1) An association rule mining stage

(2) A classification stage, which uses the mined rules from the first stage directly.

During the first stage, several techniques for creating association rules are used, which mainly are based on:

- Apriori algorithm [Agrawal and Srikant, 1994] (CBA, ARC-AC, ARC-BC, ACRI, Arubas);
- FP-tree algorithm [Han and Pei, 2000] (CMAR);

– FOIL algorithm [Quinlan and Cameron-Jones, 1993] (CPAR);
– Morishita & Sese Framework [Morishita and Sese, 2000] (CorClass).

Generating association rules can be made from all training transactions together (such it is in ARC-AC, CMAR, CBA) or can be made for transactions grouped by class label (such it is in ARC-BC), which offers small classes a chance to have representative classification rules.

In order to reduce the produced association rules, pruning in parallel with (pre-pruning) or after (post-pruning) creating association rules is performed. Different heuristics for pruning during rule generation are used, mainly based on minimum support, minimum confidence, and different kinds of error pruning [Kuncheva, 2004]. In post-pruning phase, criteria such as data coverage (ACRI) or correlation between consequent and antecedent (CMAR) are also used.

In the classification stage, three different approaches can be discerned [Depaire et al, 2008]:

(1) Using a single rule

(2) Using a subset of rules

(3) Using all rules

An example, which uses a single rule, is CBA. It classifies an instance by using the single best rule covering the instance. CPAR uses a subset of rules. It first gathers all rules covering the new instance and selects the best n rules per class. Next, it calculates the average Laplace accuracy per class and predicts the class with the highest average accuracy. CMAR, ARC-AC and ARC-BC use all rules covering a class to calculate an average score per class. CMAR selects the rule with the highest χ^2 measure from the candidate set, whereas ARC-AC and ARC-BC use the sum of confidences as score statistics. Different approach is proposed in [Coenen and Leng, 2004], which suggests to consider the size of the antecedent and favor long rules before making an allowance for confidence and support.

During the pruning phase or in classification stage, different ranking criteria for ordering the rules are used. The most common ranking mechanisms are based on the support, confidence, and cardinality of the rules, but other techniques such as the cosine measure and coverage measure (ACRI) also exist.

The "class association rules" (CAR) algorithms have their important place in the family of classification algorithms. The advantages of associative classifiers can be highlighted in several very important directions, such as very efficient training; possibility to deal with high dimensionality; no assumptions for the independence of attributes; very fast classification and the result that is easily understandable by humans. The latter two advantages make CAR algorithms an irreplaceable assistant in the processes of disaster risk management, where fast reaction and reliability of the systems are crucial.

Discussion

The main idea of association rules mining is to discover regularities in the incoming data. Such approach is very important in Global monitoring for environment and security. The incoming data in GMES is as a rule in huge volume and finding regularities in it is critical to discover incoming disaster.

In data mining there are two key factors for success. The first is to determine precisely the problem to be solved. Bundled demand usually leads to best results. The second factor is the use of appropriate data. After choosing from the available data, or the likely purchase of external data, you may need them to be transformed or combined in some way. Retrieval does not give automatically solutions without initial direction [Two Crows Corp., 2005]. Moreover, although the best tools for data mining avoid intricate statistical techniques, it is necessary to understand how the selected tools and algorithms on which they rely.

Data mining is a tool, not a magic wand, which will not sit in the database and send e-mail, to draw attention to some interesting dependency model. These rules are structured in a certain way to retrieve data from the system, which eliminated the need to know the business to understand the data and analytical methods. Each method has its advantages and disadvantages. One of the main disadvantages of data mining association rules is that the support-confidence framework often generates too many rules. The advantage of associative rules lies in their understanding – they are similar to the natural language.

The association rule mining is very important technology for the GMES. It is clear, the floods and landslides are connected disasters, and finding regularities between them is very important. The same we may say for the drought and forest fires. Finding regularities make possible the recognition of the incoming disasters. This is true for anthropogenic collapses, where the regularities in the collected data from great number measurement devices may show the incoming collapse.

Arising from the field of market basket analysis for discovering interesting rules from large collections of data [Agrawal et al, 1993], the association rule mining easily finds its applicability to model relationships between class labels and features from a training set [Bayardo, 1997]. Since then, many associative classifiers were proposed, mainly differing in the strategies used to select rules for classification and in the heuristics used for pruning rules.

The short overview of available algorithms and used structures shows the variety of decisions in association rule mining. As we can see graph structures, hash tables, different kind of trees, bit matrices, arrays, etc., are used for storing and retrieving the information.

Each kind of data structure brings some benefits and bad features. Such questions are discussed for instance in [Liu et al, 2003] where the comparison between tree-structures and arrays is made. Tree-based structures are capable of reducing traversal cost because duplicated transactions can be merged and different transactions can share the storage of their prefixes. However, they incur high construction cost especially when the dataset is sparse and large. Array-based structures incur little construction cost but they need much more traversal cost because the traversal cost of different transactions cannot be shared.

We consider the data processing as intelligent if it is referred to any rational activity. Particularly, functions of comparison, estimation, generalization, systematization, aggregation and decision-making are realized in the intelligent data processing – in difficult conditions, when there is a lack of information, time etc. The application of intelligent processing in environmental problems/tasks is conditioned at least by three groups of factors:

- first, this information is complicated: it comes from various sources, in a considerable volume and irregularly, it has various nature, ranges, and units of measurements;
- second, data that is being collected relates to the significantly important areas – monitoring the quality of water and air, danger of fires and level of ground waters, weather and climatic anomalies, sharp dangerous situations and "creeping" accidents, etc.;
- third, processing of this information in general requires application of such operations, as estimation, comparison, analysis, classification etc.

All this area, i.e. the specified information, its collection and processing, its analysis and decision-making, management and taking actions, control and estimation of results – all this is in general poorly structured, has more likely qualitative rather than quantitative nature and therefore is insufficiently formalized, and is beyond usual methods of representation and computer processing.

Within the last decades, researches and application results of new methods of work with complicated information in difficult conditions were started in a variety of areas (management, economics, business, computer science, etc.). They have the name of intellectual methods, and it is not always a metaphor.

We may define the intelligence as "ability to think abstractly", as "ability to operate effectively in the present conditions", "ability to react correctly to certain problems", "ability to study", "ability to receive knowledge from experience", "skill to get the abilities that lead to desirable results", "ability to adaptation", etc.

In Webster dictionary, the dictionary of authority, there is the following interesting definition: intelligence is an ability to be taught (to be learned) or to achieve the comprehension due to experience. Another definition from the same source is to have a ready and quick apprehension.

In Oxford dictionary of current English (ed. A. Hornby) which is an excellent source, the following definitions (besides others) are presented: intellectual – having or showing good reasoning power; intelligence – the power of perceiving, learning, understanding, and knowing.

About the intelligence, Schlesinger and Hlavach wrote, "You and we altogether should not see such a great nonsense in that one can learn about something, which has never been observed. The entire intellectual activity of individuals, as well as that of large human communities, has for long been turned to those parameters which are inaccessible to safe observation. We will not be speaking about such grandiose parameters as good and evil. We will choose something much simpler at first glance, for example the temperature of a body, which is regarded as an average rate of motion of the body's molecules.

The path leading to knowledge about directly unobservable phenomena is nothing else than an analysis of parameters which can be observed, and a search for a mechanism (model) explaining the relations between the parameters. This means an effort of exploring the relations between the observed parameters and the impossibility to explain them in another way (or more simply) than as an existence of a certain unobservable factor that affects all the visible parameters and thus is the cause of their mutual dependence. Recall astronomers who have been predicting a still unobservable planet by encountering discrepancies in observations from assumed elliptical orbits of observable planets since Kepler laws have been known. Such an approach is a normal procedure for analyzing unknown phenomena. The capability of doing such exploring has since long ago been considered to be a measure of intelligence" [Schlesinger and Hlavach, 2002].

The reality cannot be given in one definition. There exist many definitions of the concept "***intelligence***". For instance, a definition given from practical point of view is given in [Fritz, 1997]: the "intelligence is the ability to reach ones objectives. A system is more intelligent if it reaches its objectives faster and easier. This includes the ability to **learn** to do this. The intelligence of a system is a property of its mind. The mind is the functioning of its brain. An ***intelligent system*** is a system that has its own main objective, as well as senses and actuators. To reach its objective it chooses an action based on its experiences. It can learn by generalizing the experiences it has stored in its memories. Examples of intelligent systems are persons, higher animals, robots, extraterrestrials, a business, a nation. An ***artificial intelligent system*** is a computer program. We can say that it is like the proverbial black box; it has inputs and **learns** which outputs get the most approval by human beings. It stores experiences in its memory, generalizes them, and thus can deal with new circumstances (new inputs)".

The philosophical and practical points of view are two sides of the same idea. Nevertheless, from these definitions it is not clear what the main characteristics of the intelligence are. We need a definition that is more detailed and in the same time to be universal to cover natural and artificial intelligence.

In general, the definitions of the intelligence are covered by next definition [Mitov et al, 2010], which follows from the General Information Theory [Markov et al, 2006] and especially from the Theory of Infos [Markov et al, 2009].

The intelligence is synergetic combination of:

- ***(primary) activity for external interaction***. This characteristic is basic for all open systems. Activity for external interaction means possibility to reflect the influences from environment and to realize impact on the environment. For instance, in Walter Fritz' definition [Fritz, 1997] these are "senses" and "actuators";
- ***information reflection and information memory***, i.e. possibility for collecting the information. It is clear; memory is basic characteristic of intelligence for "the ability to **learn**";
- ***information self-reflection***, i.e. possibility for generating "secondary information". The generalization (creating abstractions) is well known characteristic of intelligence. Sometimes, we concentrate our investigations only to this very important possibility, which is a base for learning and recognition. The same is pointed for the intelligent system: "To reach its objective it chooses an action based on its experiences. It can learn by generalizing the experiences it has stored in its memories";
- ***information expectation*** i.e. the (secondary) information activity for internal or external contact. This characteristic means that the prognostic knowledge needs to be generated in advance and during the interaction with the environment the received information is collected and compared with one generated in advance. This not exists in usual definitions but it is the foundation-stone for definition of the concept "intelligence";
- ***resolving the information expectation***. This correspond to that the "intelligence is the ability to reach ones objectives". The target is a model of a future state (of the system) which needs to be achieved and corresponding to it prognostic knowledge needs to be "resolved" by incoming information.

In summary, ***the intelligence is creating and resolving the information expectation*** [Mitov et al, 2010].

This definition of the concept "intelligence" is a common approach for investigating the natural and artificial intelligent agents. It is clear; the reality is more complex than one definition. Fortunately, there exits good theoretical ground, especially in the areas of pattern recognition and data mining as well as the decision-making. Presented understanding of intelligence is important for realizations of the intelligent computer systems. The core element of such systems needs to be possibility for creating the information expectation as well as the one for resolving it. The variety of real implementations causes corresponded diversity in the software but the common principles will exist in all systems. Summarizing, the artificial system is intelligent if it has: (1) Activity for external interaction; (2) Information reflection and information memory; (3) Possibility for generalization (creating abstractions); (4) Information expectation; (5) Resolving the information expectation.

PART I. Practical Aspects

1 Managing Risk and Safety

The human safety and the surrounding his environment as well as the work sustainability and stability of technogenic and information objects are one of basic and strategic human problems on a way to a world sustainable development. In the present time, the quantity of technogenic dangers and the emergent situations on the globe menacing to a society and environment has considerably increased.

According opinion of a number of scientists, such events as natural disasters, technogenic and information accidents are characterized by growth of their number on 57%, damage growth — on 5,1%, the quantity of victims growth — on 6,1% annually. As is mentioned in the State strategy of sustainable development of the Russian Federation, the same tendency will preserve and increase until 2030.

For the decision of this problem it is necessary to carry out studying and evaluation of the ecological situation, forecasting of the dangerous situation development, to discover and classify the kind of dangers as well as to estimate of risk' level and to define operating effect for prevention of negative consequences from realization of ecological danger.

Every year many collapses happen in the world, which can be divided into three groups: natural, technogenic, and information collapses.

Natural collapses are hurricanes, floods, earthquakes and other acts of nature, which destroy the whole cities and lead to mass deaths of people.

Technogenic collapses are big accidents on industrial and transport objects which have been caused by failure in work of technical systems. Technogenic collapses are accompanied by victims among people and ecological disasters.

Information collapses are collapses, which occur in information computer systems. Mainly they occur because of viruses and other harmful programs.

1.1 The natural collapses

1.1.1 The earthquake on Haiti, January 2010

On 13 January 2010 near the Haiti coast happened an earthquake. Earth tremors were fixed on depth of 10 km in 15 km from capital of Haiti Port-au-Prince. The first and the most powerful tremor was equal to 7 magnitudes according to the Richter scale. After this, within several hours followed eight more earth tremors with magnitude from 5 to 5, 9 [Haritonov, 2010]. After the earthquake in the city the electricity and telecommunication was disconnected and pillages began.

The earthquake destroyed many buildings in Port-au-Prince. It is known that because of disaster at least 72 thousand people were killed. Local authorities do not exclude that during the life-saving works the number of killed people can reach 200 thousand [Haiti, 2010a].

Experts in seismology for many years have been warning the society about the possibility of destructive earthquake on Haiti, and present events did not surprise them. In 2008 five scientists at 18th Caribbean geological conference called the southern half of island Espanola (where Port-au-Prince is located) a zone of "large seismic threat". At that time one of the authors of the report, the senior scientific employee of Institute of geophysics in Texas Paul Mann took notice of the dangerous closeness of Port-au-Prince with the million citizens to a joint of two tectonic plates – North American and Caribbean [Haiti, 2010b].

According to the scientists, earthquake could set free the seismic pressure, which had been saved for 250 years. "This super cataclysm was simply time business", Roger Musson from the British geological service has told [Alekseev, 2010]. Plates moved and huge pressure grew.

It is rather difficult to predict the exact date of similar earth tremors: plates can lie motionlessly for hundreds of years and then unexpectedly start moving.

1.1.2 The snowfall in Saint-Petersburg, December 2009

The snowfall, which occurred in December 2009 (the night from 24^{th} to 25^{th}) in Saint Petersburg, was the strongest of snowfalls that had happened in December for last 130 years. According to hydro meteorological center, the snowfall became the strongest in December since 1881. The thickness of a snow cover exceeded 35 mm [Saint Petersburg, 2009a].

As a result of the snowfall, the traffic in the city streets was very complicated. The average speed of cars practically in all the streets of Petersburg was less than 40 km/h. In many places, the carriageway was narrowed to one line, because roadsides have turned to one continuous snow-drift, and between road-lines of different directions, the dividing line was formed by snow. Many drivers could not park their cars and left them directly on the carriageway, others tried to clear roadsides themselves. It was also difficult to walk on pavements: in many places, they simply had not been cleaned, and people had to go by carriageway [Saint Petersburg, 2009b].

Nevertheless, many scholars of authority claimed, that the snowfall in Saint Petersburg was not an anomaly. The formed in the city streets snow blockages were due to the inability of the city authorities to react operatively to the situation.

Anatoly Grabovsky, the head of the Northwest inter-regional territorial administration of federal service of Russia on hydrometeorology and environment monitoring, has noticed that in "anomalous" days only 15% of the monthly norm dropped out. For comparison: in Vladivostok only in one day the two-month norm dropped out [Grabovsky, 2009].

1.1.3 The earthquake in Italy, April 2009

On April 6th, 2009 in 3:32 a.m. in Italy occurred an earthquake, which power was equal to 6,3 according to the Richter scale. According to National institute of geophysics and volcanology of Italy the earthquake hypocenter was on the depth of 8,8 km and in 5 km from the centre of L'Aquila, that is located in 95 km to the north-east of Rome. Earthquake was both notable in Rome and at the coast of Adriatic Sea in the east of the country.

This earthquake became the most destructive one in Italy for the last 30 years. As a result of it, 300 people were killed, more than 1000 got wounds of different severity level, 29 000 lost their houses [Starzev, 2009].

Besides, because of the earthquake architectural monuments were damaged. In L'Aquile at least 4 temples were partly destroyed: Romance temples (XI-XII centuries) and the Renaissance temples. The castle, constructed in XVI century, in which there is a branch of the Italian National museum, was also damaged. The bell tower of the basilica of Sacred Bernarda fell down; the temple's apse was seriously damaged. In Italian capital, the terms of Karakally suffered from the earthquake [Italy, 2009a].

Joakino Juliani, a famous scientist of seismology in Italy, had predicted a strong earthquake in the central mountain region Abrutso some months before the tragedy. However, nobody had paid attention to his warnings, and even had obliged not to spread the information, which could cause a panic among local population.

The anxiety about the elements started to grow in a small town L'aquila from the middle of January, when the first weak underground tremor had been fixed. Then these tremors continued, but with variable intervals. Then Juliani, using the information about the raised concentration of radon gas around seismically unstable areas, made a conclusion on the future powerful earthquake.

In March, cars with loudspeakers drove around the town and urged the local population to begin evacuation. In police began to arrive many complaints about seismologist's panic distribution, and they caused the anger of mayor. As a result, the scientist had to remove from the Internet all information concerning the forthcoming earthquake. Moreover, on the 31 of March the Service of civil protection and the Committee on prevention of scale risks held special session to calm down the inhabitants of mountain areas. In the statement of its participants, it was said that underground tremors were normal for seismically active areas, such as L'Aquila. In addition, it was told that there were no occasions for anxiety, and that the supervision in these areas would be continued "just in case" [Italy, 2009b].

1.1.4 Tornado Katrina, August 2005

The tornado Katrina is the most destructive tornado in the history of the USA. The tornado reached the southeast coast of the USA on August 29, 2005. In the zone of its action were the following areas: Louisiana, the south and the centre of Mississippi, the south of Alabama, the West of Georgia, the West and the south of Florida. The speed of the wind reached 280 km/h.

The city of New Orleans suffered the most in those days. It was destroyed for 80%. The tornado Katrina killed almost 2000 human lives and brought an economic damage at a rate of 125 billion dollars.

The tornado had a very bad influence on the oil business. On September 2, the representative of the American Institute of Oil, Tim Sampson, established that Katrina destroyed 58 oil platforms and boring installations in the Gulf of Mexico, 30 of them were lost. Since the first days of tornado, the oil extraction in the Gulf of Mexico was reduced for 95%, and its price began to grow [Forex, 2005].

Tornado is a rotating warm air stream. Te speed of rotation can reach 1000 meters per second. Rarefied rain clouds and a powerful vertical stream of air between a cloud and an earth surface are necessary for its formation. Experts of meteorology consider that the most destructive tornados, which have become very frequent for last decades, are connected with global warming of a climate.

As the temperature in atmosphere is constantly growing, we should expect very strong tornados in the near future.

Experts in tornados have been repeating for many years that consequences of a blow of a strong tornado on New Orleans can be catastrophic. Two experts warned about serious danger to a city 6 weeks before an arrival of Katrina, but nobody trusted in their predictions in July 2005.

The atmospheric system, which turned into the tornado Katrina, arose on August 23 in the form of area of the lowered pressure at the coast of the Bahamas. This natural phenomenon not necessarily leads to tornado formation – for this purpose is necessary a much bigger quantity of energy. Within 6 days in 1 500 km from the coast the tropical zone of the lowered pressure slowly, but steadily gained its strength. At last, on August 29 it turned to a mature tornado in width more than 800 km. The speed of the wind was 280 km/h [Katrina, 2005].

The killing force of tornado is influenced by 4 basic elements.

The first element is the speed of the wind. During the tornado Katrina the speed of the wind reached 225 km/h, therefore it was given the fourth category on a scale of Saffira-Simpsona. For comparison: tornados of the first category have speed 120 km/h, the fifth category – 320 km/h.

The next element is the size of tornado. The more it is, the more its destructive force is. The zone of the destructions, caused by Katrina, was stretched on 160 km along the coast of the Gulf of Mexico and on 160 km deep into the continent. It is a territory with the population of 10 million inhabitants.

The third element is torrential rains. Hurricanes accumulate billion liters of water over the sea and bring it down on the earth in the form of downpours.

Besides, the tornado Katrina, as well as the majority of other tornados, had one more terrible weapon – the sea level abruptly rose and fell down upon the earth, sweeping away everything on its way. This phenomenon is called "stormy water tide". During the tornado Katrina in a resort town Biloxi the absolute record of stormy waters was set: more than 9 meters [Katrina, 2005].

These 4 elements are peculiar to the majority of tornados. They become the reason of the basic destructions and deaths of people in the zone of natural calamity. However, the tornado Katrina was not the champion in each of these categories, the destructions, caused by it, were apocalyptical.

1.1.5 The volcano Eyjafjallajokull eruption, April 2010

Starting on April 14, 2010, the volcano Eyjafjallajokull (Iceland) eruption rose in the air huge clubs of ash and steam. The last time the volcano woke up in 1821, and its eruption lasted for two years.

According to the head of the Meteorological Office of Iceland, Robert Matthew, on April 14 there were two ejections of volcanic lava, the second one was particularly big (Figure 1). As the magma immediately met with the ice the dirty jets of boiling water with stones and ash were ejected out of the crater [Newsru, 2010].

The airspace in Western Europe was blocked; flights to dozen countries were cancelled, because volcanic ash can damage plane engines.

"Firstly it is highly abrasive and can scour and damage moving parts. Secondly, if it enters a jet engine the intense heat of the engine can fuse it to the interior of the engine with a caking of hot glass, which ultimately can cause the engine to cut out completely", explained Dr. Mike Branney, senior lecturer in volcanology, University of Leicester [BBC, 2010].

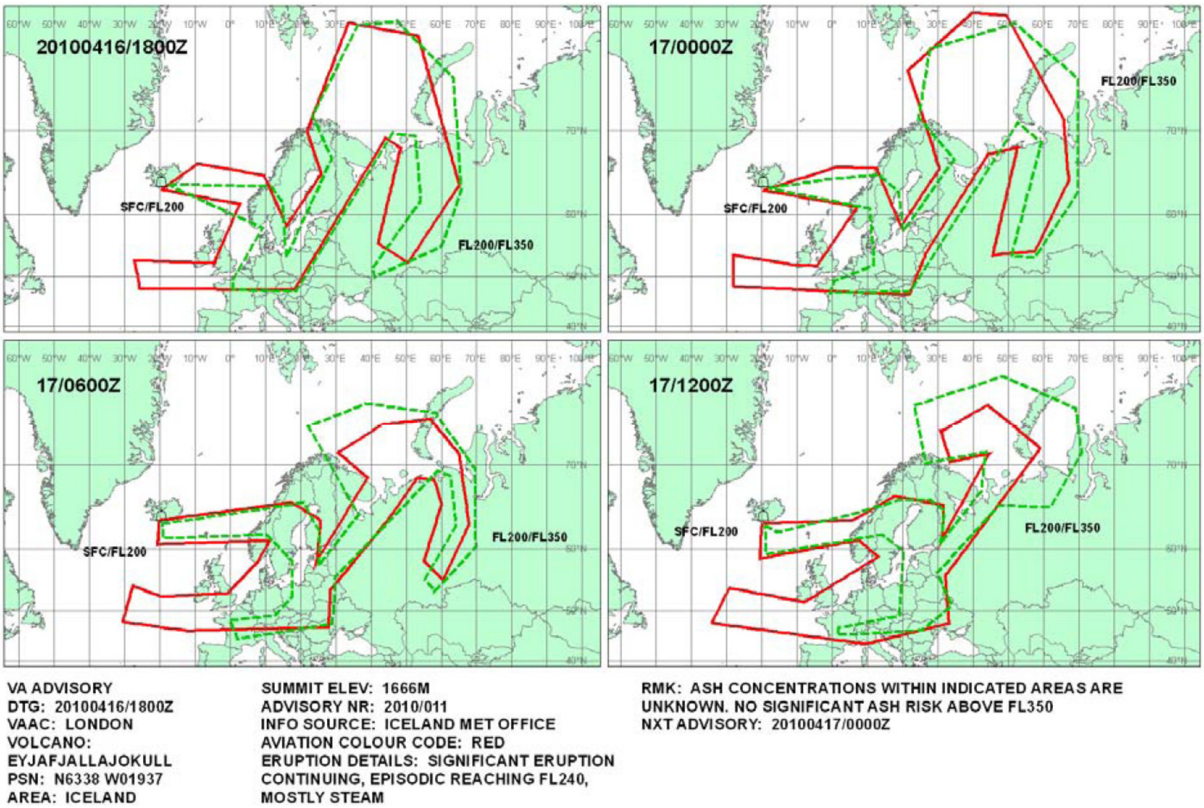


Figure 1. Ash concentration on April 17, 2010 [Metoffice, 2010]

The economic losses were extremely high. Airlines were losing about $200 million every day because of the shutdown of many European airports. In addition, trade goods transported by air were unable to reach their markets, so their prices grew, too [Kaufman, 2010].

According to the European flight control agency, about 500 flights in Europe were cancelled and about 20,000 delayed because of the ash cloud produced by the volcano in Island. The worst situation was in the airports in Great Britain, Spain, Portugal, and Austria [Ruvr, 2010].

Traffic jams once again raised the issue of high fragmentation of the European airspace. Because of this in Europe call for the adoption of the program "open sky", that will establish a common aviation area of the continent. The project should enter into force on January 1, 2012.

1.2 The technogenic collapses

1.2.1 Train collision in Belgium, February 2010

On February 15, 2010, two passenger trains collided near Brussels. Collision occurred in 8:30 a.m. local time.

The force of the stroke was so powerful, that the first carriages of trains reared and damaged the electric lines located at three-meter height over the railways. Some carriages were derailed, one of them turned over [Belgium, 2010a].

The incident picture looks as follows. The six-carriage train, going from Leuven in Braine-le-Comte, changed the track at station Buizingen and passed to the track number 96 of Belgian railways. Because of a strong snowfall, the engine driver did not notice the red signal of a semaphore. The track in that zone is equipped with the system of emergency braking, which does not allow a train to pass on a red light, but this train was not equipped properly, so it could not activate the system.

At the same time overcrowded with passengers 12-carriages express Kevren-Liege, going to Brussels, on full speed went on the same 96th track. The engine driver got the permission from the controller, who did not suspect that there was another train on this track.

Therefore, the human factor and technical insufficiency of security systems became the main version of the investigation at a preliminary stage.

Because of the accident, 18 people were killed, nearby 125 got wounds of various severity levels]. The railway traffic on the two main lines of the western Belgium was stopped: Brussels-Mons and Brussels-Turne, and the railway traffic of the international high-speed trains "Talis" and "Evrostar" was also stopped [Belgium, 2010b].

1.2.2 The accident at the Sayan-Shushenskaya hydro electrical station

Industrial technological disaster at the Sayan-Shushenskaya power plant occurred Aug. 17, 2009. 75 people died because of the accident. Equipment and premises of the station was severely damaged. Workstation electricity production has been suspended.

Consequences of the accident affected the ecological environment of water area, adjacent to the HPP. The accident has had a negative impact on the environment: oil bath lubrication of thrust bearings of hydraulic units from the destroyed management systems guide vanes and transformers fell into the river and formed a spot spread over 130 km. The total volume of oil leakages from the equipment of the station was 436.5 m ³, of which approximately 45 m ³ predominantly turbine oil has fell into a river.

Figure 2. Engine room Sayano-Shushenskaya power plant before (left) and after (right) the accident (www.lhp.rushydro.ru)

Because of the accident, the main hydropower number 2 and main hydraulic were destroyed. There are broken generators in hydro number 7 and number 9. Other hydraulic units were received substantial damage.

The walls and roof of the machine hall in the area of hydro-number 2, 3, 4 were destroyed too. Other equipment of the station – transformers, cranes, elevators, and electrical equipment was received varying degrees of damage. It was located in the engine room and near him. Total losses associated with damage to the equipment valued at 7 billion rubles.

This accident is the largest catastrophe in the history of hydropower projects in Russia and one of the most significant in the history of world hydropower. Nevertheless, the assessment of the consequences of the expert and political community is ambiguous. The accident caused a major impact on society, becoming one of the most talked about in the media events of 2009.

As a result of the investigation the direct cause of the accident was called fatigue failure studs cover attachment hydraulic turbines, which led to its disruption and flooding the engine room station (Figure 2).

It was found that it was not properly organized permanent control of technical condition of equipment operational and maintenance personnel (which should provide instruction on operation of Sayan-Shushenskaya HES approved by the chief engineer of HES 18.05.2009 g.). The main cause of the accident was the failure of taking prompt measures to stop the second hydraulic unit and determine the causes of vibration [GES, 2009].

1.2.3 The explosion at the chemical plant in Toulouse, France

At a chemical plant in Toulouse (France) on 21 September 2001 have occurred explosions, whose consequences are still considered like the largest industrial disaster in recent years.

The cause of the explosion was the violation of safety at the facility.

Thousands of homes and many institutions were destroyed, including 79 schools, 11 secondary schools, 26 colleges, 2 university, 184 kindergartens, 27 thousand apartments, 40 thousand people have become homeless, and 134 companies have stopped their work.

Government and insurance companies received 100 thousand claims for damages. The total damage amounted to 3 billion Euros. 30 people have died in the city, the total number of wounded has exceeded more than 3,5 thousand.

1.3 The information collapses

1.3.1 Sberbank, December 2009

On the 26th and 27th of December, 2009 the clients of Sberbank, who were paying for the goods and services with credit cards, were in an unpleasant situation: the money was removed from their accounts two times.

The representatives of Sberbank declared that the reason of the failure were technical defects at counterparts. The bank Press-service has confirmed "the doubling of some transactions, made on the 26th and 27th of December on purchases of clients of the bank in the networks, acquiring in which was carried out not by Sberbank". According to the bank, the failure influenced on less than 0,5% of the transactions served in the system of Sberbank every day [Gerashenko, 2009].

"Within several hours all doubled transactions will be searched, and the rests on the accounts of clients will be restored", was told in the Sberbank's communiqué.

But there is an opinion that these problems were not at the counterparts, but in the bank. For example, there could be a hardware failure. "The Sberbank has its own processing of cards, it means that the problems are somewhere in its structure", — tells a source in the sphere of bankcards service [Gerashenko, 2009].

1.3.2 Citibank, Summer 2009

On December 22, 2009 The Wall Street Journal, citing unnamed officials in the U.S. government announced the kidnapping of tens of millions of dollars from the accounts of depositors of the Citibank [Churumova, 2009]. The attack was noticed by the U.S. secret services in the summer of 2009, when it was discovered a suspicious traffic coming from Internet addresses previously used by the hacker group, Russian Business Network. It is quite possible that the kidnapping of money took place long before when he was tracked down by the suspect traffic. Moreover, as evidence of the involvement of Russian hackers for the crimes of U.S. investigators, allege that the hacking program Black Energy was used.

According to Secure Works expert Joe Stewart, Black Energy has written in Russia by hacker Crash.

This program infects the computer network and allows an attacker to perform actions using the resources of the infected computer. With its help, Georgian government portals and Estonian websites have blocked in 2007-2008 [RBS, 2009].

However, Black Energy sells online for $ 40, says Jose Nahariya, manager of Arbor Networks. According to director of Digital Stakeout Adam Mikruta, a software package YES Exploit System was selling for 700$ in April 2009 in the Internet. It allows you to steal the parameters of access to the bank accounts. It includes Black Energy also.

According to Joe Stewart, in summer Cr4sh has released a new version of the program, through which you can steal usernames and passwords for access to the bank servers [Churumova, 2009].

Citigroup denies any knowledge of the attack and the theft of a substantial amount of the bank. "We have not had any leaks in the system and no losses, both by consumers and by banks. Any suggestion that the FBI is working on a matter of Citigroup, associated with the theft of tens of millions of dollars, are not true", said Joe Petro, Managing Director of the Department of Security, Citigroup [Truhanov, 2009].

Russian experts also confirmed that Citibank was not hacking. "The incident is connected with the attack on the trade network 7-eleven, which occurred in the summer of 2008, Alexander commented to the guests, a global center of research from Kaspersky Lab. This trading network uses ATM Citibank.

Bank was providing their work from a technical point of view, although they were the property of 7-eleven. Hackers managed to access the payment system 7-eleven through vulnerability. They did not use malicious programs.

Then they worked with the so-called "loot" – people who received either ready-made counterfeit cards or instructions how to make them. After that drops, using these credit cards, withdraw money from ATMs" [Truhanov, 2009].

Robert Blanchard had trouble accessing account. It is an indirect proof that the hacking took place. He is co-owner of Bridge Metal Industries. At night, July 6, 2009, he tried over the Internet to gain access to the bank account of the company, but was unable to do so. Then he phoned to the bank, and there he was told that he would get a new password by e-mail. But even with the new password Blanchard was unable to gain access. By that time, hackers have been transferred from his account $ 1 million in banks in Latvia and Ukraine. As a result, Citibank returned most of the stolen Blanchard amount [Churumova, 2009].

It is worth mentioning that on December 22, the day with a message The Wall Street Journal, the White House decided to declare the name of "CyberKing" – the official, who will be responsible for cyber security departments. It became Howard Schmidt, who was cyber-advisor in the previous administration [Dubinskaya, 2009].

1.3.3 RBS WorldPay, November 2009

RBS WorldPay is organization that provides support payment transactions for the Royal Bank of Scotland (RBS) in the United States.

Hacking was detected in server of RBS In November 2009, located in Atlanta.

According to investigators, intruders broke into the bank's security system and have access to data of cards that receives the salary of bank customers. This information was passed by crackers accomplices, who cashed more than $ 9 million money with the help of charts around the world. Money was withdrawn in the U.S., Russia, Estonia, Ukraine, Italy, Japan [RBS, 2009].

The report of the Ministry of Justice of the United States called the names of criminals, these are Sergei Tsurikov from Tallinn, Victor Pleschuk from St. Petersburg, Oleg Kovelin from Chisinau, as well as people called in investigative materials "hacker 3". Living in Tallinn, Igor Grudizhev, Ronald Tsoi, Evelyn Choi, and Michael Yevgenov also accused of fraud with bankcards. 3 hackers could face up to 20 years in prison. In addition, they will have to pay compensation in the amount of $ 9.4 million. Other members of the group could receive up to 15 years in prison and a fine of $ 250000 [RH, 2009].

1.3.4 Spain, Summer 2009

In autumn 2009, the holders of cards of many banks have found their cards blocked. In most cases, operators of banks explained that the cards should be released again, as they have been compromised. Some customers have been informed that there were attempts of unauthorized transactions on their cards.

Most owners of blocked cards were in Spain in the summer or in September, where they used the cards [Spain, 2010].

Experts suggest that there was a hacking of servers, where information was stored on the cards. The Central Commission of credit cards in Germany called cards review like preventive measure, assuring cardholders that all potential victims will receive appropriate notice and in any case would not lose their money [Germany, 2009]. "Banks have been notified of the payment systems Visa and MasterCard (the latest reported first), the mass compromise of plastic cards (namely, their requisites: № cards, validity, etc.) used in Spain during the period January-September 2009 year" told the head of the interaction of "Alfa-Bank" Alexei Golenishchev [Jogova, 2010].

1.3.5 The service of state incomes of Latvia, February 2010

In 2000, the Service of state incomes of Latvia introduced the electronic declaring system (EDS) to simplify procedures of the report on the paid taxes by physical and juridical persons. Tax department counted on this IT-project, allocating for its realization the essential sums of money, and widely advertised it. It was especially underlined that all operations made in the system were safe: each document had its own password and SSL protocol was used during the transmission of the document [Kazakov, 2010].

But in February 2010 in the security system of the EDS was found a vulnerability through which for three months unknown persons have stolen 7,4 million documents (120 Gb) [Latvia, 2010a]. The information not only about finance of the state officials, but also about private persons and enterprises was stolen from system.

The group, called "National army of the 4th Atmody" (4ATA), has informed about the leak. As members of 4ATA said, last spring they found out about the existence of defects in the security safety with which help it was possible to get access to the EDS. This information one of the developers of the EDS told the members of 4ATA. He said that they had left the "hole" in the system, because a person of the high rank from the tax service had ordered them to do so. The "hole" was very simple and everyone could use it for all this time. Vulnerability allowed download documents by user without authorization, only the address and the number of document were needed [Latvia, 2010a].

It is necessary to underline that this vulnerability has existed for 10 years, as 4ATA claims.

From the point of view of potential harm, it is the greatest leak, which has ever happened in Latvia.

"Present or former leading persons of the Service of state incomes, guilty of the scale leak, will be punished", the minister of Finance Ejnars Repshe declared on February 15 [Latvia, 2010b].

1.4 Prediction of natural collapses

Earthquake early warning can provide a few seconds to tens of seconds warning prior to ground shaking during an earthquake. Earthquake early warning systems are currently operational in Mexico, Taiwan and Japan but not in the United States.

EalarmS, short for Earthquake Alarms Systems, is a methodology designed to provide warnings in California and other earthquake prone regions around the world. The warning messages provided by EalarmS can be used to reduce the damage, costs, and casualties in an earthquake [EalarmS, 2010].

EalarmS takes the form of a suite of algorithms designed to rapidly detect the initiation of an earthquake, determine the size (magnitude) and location of the event, predict the peak ground motion expected in the region around the event, and issue a warning to people in locations that may expect significant ground motion. The algorithms use data from regional broadband seismic networks (Figure 3).

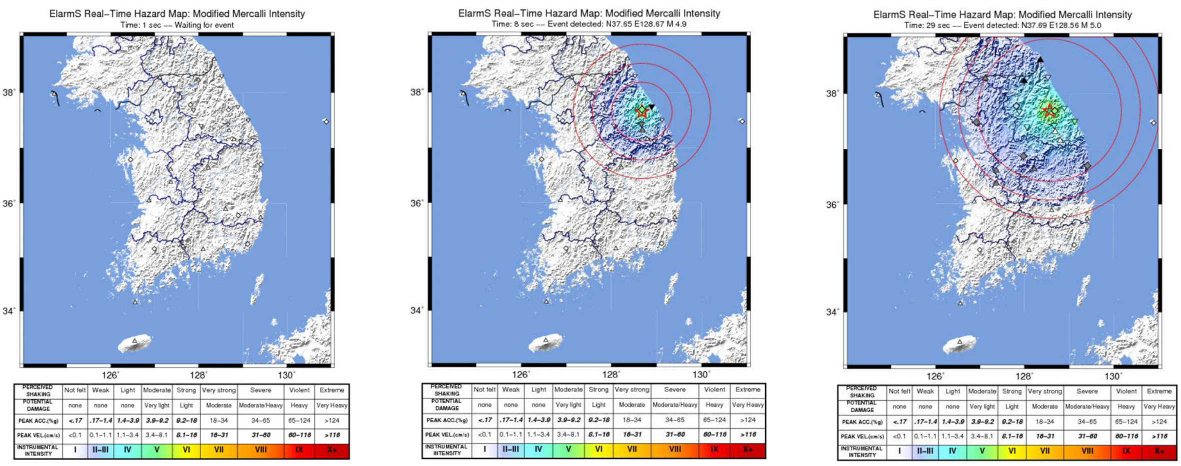
ElarmS is currently being tested by the CISN as part of the real time seismic system in California. This pre-prototype test will allow assessment of the likely timeliness and accuracy of warnings if early warning is implemented for general use in California. ElarmS has also been tested offline using data from Japan, Taiwan, Italy, Alaska, the Pacific Northwest, and California.

Earthquake early warning systems have been expanding rapidly around the world. There are currently warning systems in Japan, Taiwan, Mexico, Turkey and Romania

We need earthquake early warning because:

- current earthquake mitigation strategies focus on long-term ground shaking forecasts that can be used in building design and rapid post-event notification used for emergency response;
- early warning systems allow for short-term mitigation including slowing and stopping of transportation systems, switching industrial and utility systems to a safe mode, and taking personal protective measures;
- early warning would reduce the number of casualties and the cost of earthquakes in California.

Feasibility studies of the ElarmS methodology show that the amount of warning time would range from a few seconds to a few tens of seconds depending on your distance from the epicenter of the earthquake. This is enough time to slow and stop transportation such as trains, taxiing planes and cars entering bridges and tunnels; to move away from dangerous machines or chemicals at work, and take cover under a desk; to automatically shut down and isolate industrial systems. Taking these actions before shaking starts can reduce damage and casualties during an earthquake. It can also prevent cascading failures in the aftermath of an event. For example, isolating utilities before shaking can reduce the number of fire initiations.



(a) *(b)* *(c)*

Figure 3. Earthquake (a) not observed; (b) discovered; (c) 30th second after the first jolt (the red lines show the region of wave propagation)

Seismologists across California are currently planning real-time testing of earthquake early warning across the state. We want to hear from companies, institutions, government agencies, and individuals about how a few seconds to a few tens of seconds warning could be used to reduce the casualties and damage caused by an earthquake in the state.

AlertMaps are ElarmS 'prediction of the coming ground shaking. AlertMaps are updated every second as more information is incorporated into the ElarmS predictions. As time progresses and the ground, shaking spreads across a region the initially predictive AlertMaps evolve into a ShakeMap, which is a map of the observed peak ground shaking. You can view the AlertMaps as an animation or look at individual frames.

1.5 Rapid multi-sensor system for effective risk analyses

An international consortium of specialists from the Netherlands, Russia, Bulgaria, United States, and Australia has developed and demonstrated a new airborne multi-sensor system that is ready for implementation in two major areas of application, related to the Common Agriculture Policy (CAP): control of the Good Agriculture Environment Conditions (GAEC) and the update of the Land Parcel Identification System (LPIS) [Haarbrink et al, 2007]. The airborne multi-sensor system consists of several cameras including a soil moisture scanner for (underground) water and drought mapping, a thermal infrared camera for ground temperature data, a lidar scanner for elevation models, vegetation mapping, and deformation detection, and a digital photo camera for visual and near infrared imagery of the project area. All sensors are mounted on board one single light aircraft and are operated simultaneously to produce a wide range of critical information that is then processed into an Information Monitoring System (IMS). The international team has demonstrated the benefits of the system in a show case project in Bulgaria in the summer of 2007.

The project was successfully performed by MIRAMAP, a private company from the Netherlands that was founded at the European Space Incubator (ESI) initiative of the ESA Technology Transfer & Promotion (TTP) office. The TTP office is contributing to the capitalization of space-based technology and knowledge for the benefit of Europe's economy and science.

The MIRAMAP sensor suite simultaneously collects high accuracy soil moisture data, ground temperature data, digital elevation data, and ortho-photos. It consists of three microwave sensors in X-band, C-band, and L-band that are all GNSS integrated, a thermal infrared camera, a lidar scanner, and a digital photo camera. Table 1 lists the specifications of each different sensor and specific use.

Table 1. MIRAMAP Multi-Sensor Specifications

Sensor	Wavelength	Specs	Use
Passive Microwave Scanner	Microwave 2, 5, 21 cm	5-m GSD, 0.15 K precision	(Sub)surface detection of wet and dry areas
Thermal Camera	LW Infrared 7.5-13 micron	3-m GSD, 0.1° C precision	Surface Temperature
Lidar Scanner (Laserscanner)	SW Infrared 1064 nm	2-m GSD, 0.10 m precision	Elevation Model, Hydrological Model, Deformations
Digital Photo Camera	Visible 0.4-0.7 micron	10-cm GSD, Subpixel precision	Detailed Visible Interpretation, Cartography

The platform on which these instruments are flown is a reliable and safe twin Aero Commander (Figure 4). The aircraft is specially modified to carry simultaneously this range of instruments. The capability to measure such a comprehensive range of remotely sensed parameters from a single low-cost airborne platform is unique worldwide.

As for the soil moisture products, the X-band and C-band sensor makes a conical scan at constant incidence angle over a wide swath, while the L-band sensor for underground soil moisture mapping makes a twin-beam oscillating scan (Figure 5). The small instrument sizes and weights enable use of a low-cost light aircraft as the observing platform, providing decision makers with a new affordable tool.

Through both laboratory and field experiments it has been documented that the passive microwave radiometers with internal calibration, and the processing/retrieval algorithms from the Institute of Radio-engineering and Electronics (IRE) of the Russian Academy of Sciences (RAS) and Radio Corporation VEGA, are feasible to determine several soil, water and vegetation related environmental parameters and conditions. These are soil moisture content, depth to shallow water table, buried metal objects under shallow dry ground, contours of water seepage through hydro technical constructions (levees and dams), sliding zones, plant biomass above wet soil or water surfaces, temperature changes, oil slicks, salt and pollutant concentration in water areas, on-ground snow melting and ice on roads and runways.

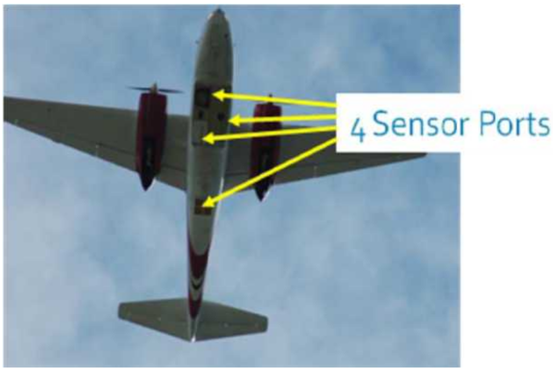


Figure 4. MIRAMAP multi-sensor aircraft

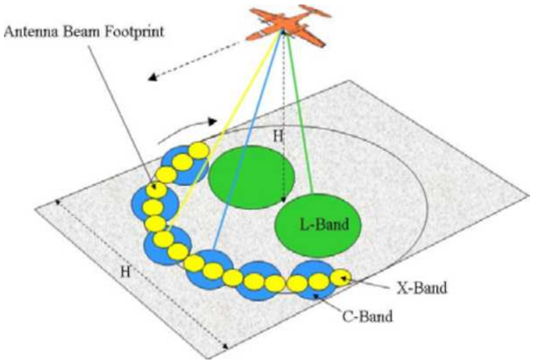


Figure 5. MIRAMAP soil moisture scanner

Simultaneously collected multi-sensor data from the air offer a high social-economic relevance. The approach brings government managers and decision makers the following benefits:

- *rapid*: Large areas can be monitored in a short time frame. The quick availability of information can also be of advantage to minimize the impact of disasters, or to prevent disasters. Adequate and rapid response in the first hours of an (pending) emergency situation is of great importance. Many data can also be collected at night;
- *effective*: Simultaneous acquisition of remote sensing data is a very effective tool to monitor water barriers and coastal zones. Detailed data and flexible deployment on a single platform enables water managers and decision makers to take adequate action, and therefore spend their financial resources effectively;
- *accurate*: A complete and detailed view of possible failure locations is given. The main failure indicators of water barriers and coastal zones are water seepage, dry areas, deformations, and vegetation quality. The results can be easily imported and combined with existing geo-information;
- *safe*: The measurements are collected from the air, so field inspection specialists are not faced with potentially dangerous situations, such as rising water.

Airborne multi-sensor data was collected over two selected areas in Bulgaria – river Rusenski Lom and dams at village Nikolovo – near town Ruse, and additional areas over river Danube, river Yantra, and a forest fire between July 30 and August 2, 2007. The project locations are shown in Figure 6. Unique information was received related to:

- land surface cover, including true ortho-photo at 10-cm ground sample distance, ground temperature data at 0.1 degrees precision, and relief features at 2-m grid and 15-cm vertical precision of terrain impacting moisture parameters;
- surface soil moisture, dry conditions 0.05 to 0.10 m^3/m^3 and saturation up to 0.35 m^3/m^3;
- underground soil moisture reached 0.35 m^3/m^3 in the condition of saturation;
- revealing over-moistened zones near water barriers.

Microwave surface moisture related data revealed over moistening in the areas near the Yantra River outflow zone, which is indicated with red arrows in Figure 7. Water seepage of Danube River through the dike or in land surface depression located in these spots could be reason of this phenomenon.

All different datasets (soil moisture maps, ground temperature data, digital elevation data, and ortho-photos) were processed and delivered to the product specs listed in Table 1 in UTM35 coordinates, so that all data can be quickly combined with existing datasets such as digital topographic maps and satellite imagery.

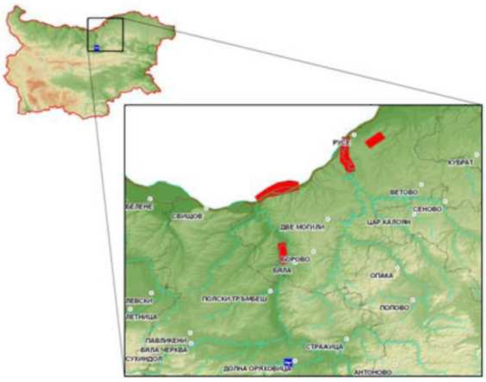


Figure 6. Project locations in Bulgaria

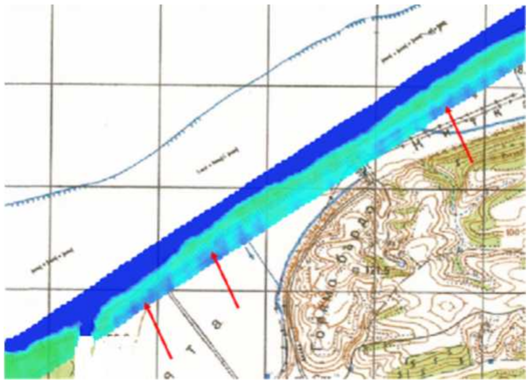


Figure 7. Water seepage at Danube river

During the project weather changed rapidly from dry to very wet condition, so that unique information was received related to soil moisture changes. Figure 8 illustrates surface moisture related data of July 31 under dry condition. Figure 9 illustrates surface moisture related data of August 2 under wet condition of the same area.

The colors in Figure 8 and Figure 9 correspond to the soil moisture gradations in the upper 0-5 cm thick layer. Dark green ranges between 0 and 0.05 m^3/m^3, light green ranges between 0.05 and 0.10 m^3/m^3, light blue ranges between 0.10 and 0.20 m^3/m^3, and dark blue ranges between 0.20 and 0.35 m^3/m^3, and represents open water.

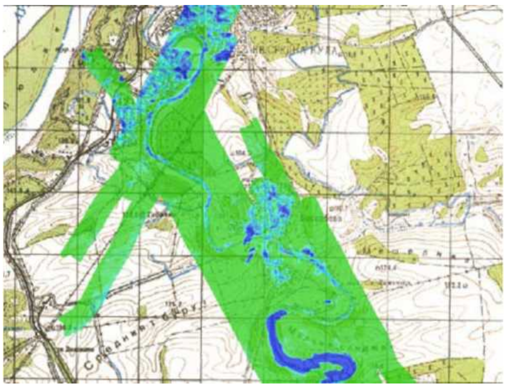


Figure 8. Soil moisture under dry condition

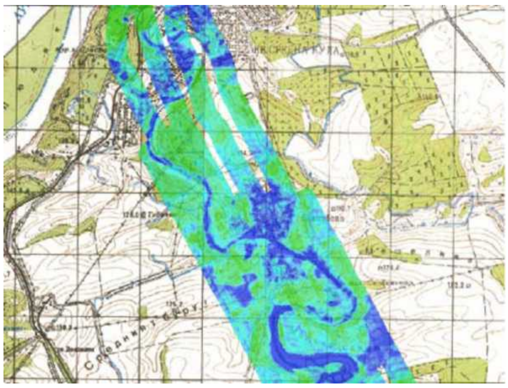


Figure 9. Soil moisture under wet condition

The land surface cover from 10-cm true ortho-photo (Figure 10) was combined with the lidar relief features (Figure 11) in a three-dimensional (3-D) fly-through by ReSAC. A snapshot of this product is shown in Figure 12.

The simultaneously collected multi-sensor data was finally combined in an Information Management System (IMS) for effective risk analyses. The block diagram of Figure 11 illustrates the various elements of the IMS: the airborne data that is geo-referenced using positioning and inertial data, a backup and retrieval database system that includes existing a-priori geodata, and an integral processing part using various mathematical procedures, such as normalization, weighting, and distribution.

Figure 10. 10-cm Orthophoto for LPIS

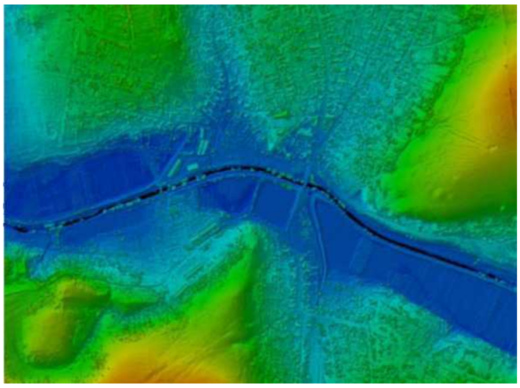


Figure 11. Lidar elevation model for GAEC

Figure 12. 3-D Fly-through of orthophoto and Lidar elevation

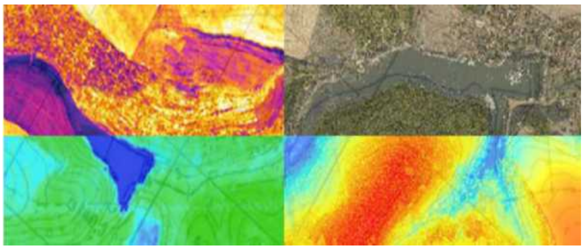


Figure 13. Multi-sensor composite

Figure 13 illustrates a composite of all datasets at Dams at Nikolovo. The surface temperatures (top left) range between 25°C (dark blue, open water) and 50°C (white, dry and open fields), and correlate with the soil moisture data (bottom left) and lidar elevation data (bottom right).

The main conclusions and recommendations of the show case project in Bulgaria are:

- the project has successfully proven the effectiveness of the used MIRAMAP multi-sensor technology and approach for preventive risk management of natural hazards, part of which is flood preventive measures and preventive control of flood protection facilities;
- the MIRAMAP system is ready for implementation in two major areas of application, related to the Common Agriculture Policy (CAP): control of the Good Agriculture Environment Conditions (GAEC) and the update of the Land Parcel Identification System (LPIS). As for the GAEC, the system has especially proven successful for standards; temporary channeling of surface water on sloping terrain, and defense of ground structure through maintenance of surface water drainage.

The conditions of over-moistening revealed on the August 2nd data were the important indicator of expected flooding.

1.6 Network protection – policy and technology reinforcements

Critical monitoring infrastructures that serve emergency management require strong network protection means. Security is not a finite state and is a permanent process. Basic technical means of network security include firewalls, secure protocols, PKI and electronic signature systems, virus and spam filtering. The final network protection level still depends on correct integration of mentioned components. It is sensitive the use of proper knowledge on known software bugs and intrusions. In general it is important to know about the existence of alternative means for security above the PKI [Aslanyan et al, 2002; 2003].

The issues mentioned above are treated in terms of several EC funded research project results. These projects were developing pattern recognition and data mining algorithms, estimating complexities of discrete optimization problems, they dealt with software agent structures, which are mostly implemented, in Java algorithmic language. Here we apply these results to the Network Protection schedule. First, in this section we discuss and conclude that the complications/combinations of basic means for security provide the same or comparable level of protection. Then, we show that the distributed software agent systems may provide additional means for security, being able to monitor wide network areas and analyze the monitored information – mining useful knowledge on intrusions, on malfunctioning, etc.

1.6.1 INTAS 04-77-7173 "Data flow systems: algorithms and complexity"

The project research objectives in global terms can be characterized as:

1. Comparative study and systematization of the achievements in combinatorial and computational algorithms and their complexities – intractability, approximation, lower bounds, and heuristics
2. Elaboration of new algorithmic approaches in cases, when the known algorithms are proven non-data flow optimal
3. Study of complex systems: asynchronous computational models (such as hybrid agent societies, WEB search engines), their computational and knowledge extraction power, the complexity counterpart of the overall behavior and functionality of cellular autonomous systems

One of the specific research objectives studies the encryption related algorithms. From one side – it is the well known the one way functions are valuable instrument in this area, from other side – it is directly unknown and important to know if a mixture and complication/combination of several one-way function will bring additional hardness in reversing these functions. It appears that combining combinatorial – computational means for security composes the mixture, which is the same level protection structure. Such postulation is one of the valuable inputs of INTAS 04-77-7173 project.

The formal framework is as follows [Grigoriev et al, 2006].

Let (G,E,D) be probabilistic worst-case polynomial time algorithms for key generation, encryption and decryption respectively of a PKI system S. The PKI scheme defined is δ(n)-correct iff for large n $Pr\{Dsk(Epk(m))=m\} \geq \delta(n)$ for $(pk,sk) \leftarrow G(1^n)$ /key pair generation by given security parameter/. Here sk, pk are secret and public keys and m is a plaintext message.

Probabilistic black-box A ε(n)-breaks PKI scheme if for infinitely many parameters n $Pr\{Apk(1^n,Epk(m))=m\} \geq \varepsilon(n)$ for $(pk,sk) \leftarrow G(1^n)$.

An encryption scheme (G_1,E_1,D_1) is reducible to an encryption scheme (G_2,E_2,D_2) if there exists a probabilistic polynomial time oracle machine R, such that for any probabilistic black-box A that breaks (G_2,E_2,D_2), R^A breaks (G_1,E_1,D_1).

We denote the class of all 2/3-correct public key encryption schemes by PKCS.

Theorem: There exists a complete PKCS /complete, as an analogy from the NP complexity theory means that all schemes are reducible to the considered one/.

Polynomial reduction of problems and complectness concepts appeared earlier in algorithmic intractability studies. Any NP complete problem is the hardest representative of class of nondeterministic polynomial recognition problems. The given result on encryption proves a similar property of being the hardest representative for encryption algorithms when an error is allowed. Our interest is not in details but is in PKI systems being reducible to each other. This means that they have the same protection power and then we conclude that alternative algorithmic forces are necessary for security out of combinatorial and computational counterparts, and then we try to strengthen the level of securities through the intelligent network monitoring and data analysis means.

A similar notion is placed in ESFORS (European security forum for WEB services, Software and Systems), FP6-027599 project documentation. Prof. Lioy (Politecnico di Torino) mentioned limitations and conflicts between business objectives and effectiveness of security services insisting to incorporate the ability of prediction of systems behavior. Prof. Malek from University of Berlin also mentioned the role of prediction with some specific details and suggestions. In the same line, the prediction through Data Mining and intelligent software agents is innovated by European Framework Program SPARTA project.

1.6.2 SPARTA "Security policy adaptation reinforced through agents" project

Security Policy consists of:

1. Detailed description of any information, which might be monitored operationally and which might be of some interest for data security reasons
2. Archiving of existing knowledge – systems, structures, technologies, viruses, hacking
3. The data analysis algorithms – to be designed and realized, by the above data descriptions according to the basic tasks and requirements
4. The set of laws, rules, and practice that regulate how an organization implements, manages, protects, and distributes its information and computing resources to achieve security objectives

SPARTA system started by the mentioned European Framework Program Project http://cordis.europa.eu and was further developed after the project. This framework considers the mobile agent server system, which monitors the implementation of security policies, identifies security problems, and performs intrusion detection tasks. Security checks are flexible, at run-time and without interrupting the system's activity. The two main use cases are Surveillance (of a given security policy) and Intrusion Detection (ID).

✓ System design, architecture, main components

SPARTA architecture supports both distributed and centralized use cases (Figure 14).

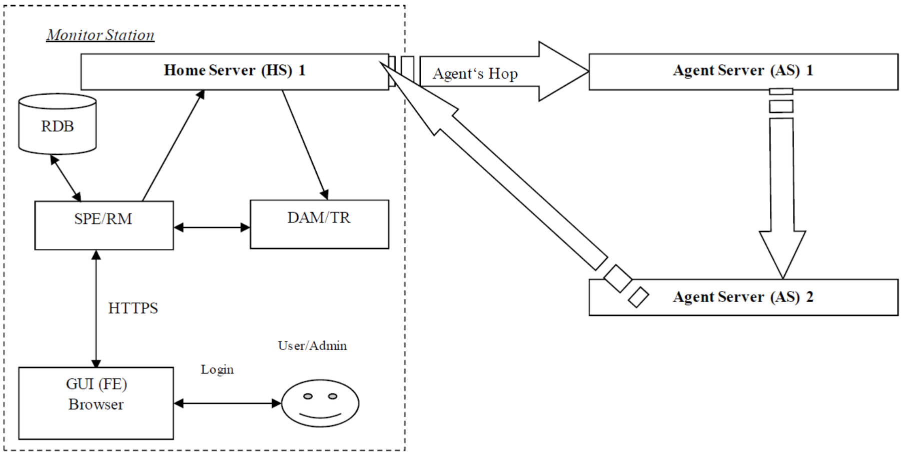


Figure 14. SPARTA architecture

Main components in SPARTA architecture are: Agents (A); Agent Server (AS); Home Server (HS); Secure Infrastructure with Secure Information Space (SIS); Data Analyzer Module (DAM); Security Policy Editor (SPE); and User Front End (FE).

✓ Architecture, Agents (A)

Agent is the centerpiece of SPARTA design. It bears mobile software code and serves automating application tasks. Main types of SPARTA agents are: One-hop, Multi-hop, and Embedded. Each agent consists of two parts:

1. Agent State: agent's operational data together with management information (e.g. user ID)
2. Agent Code: its source code as Java class file, which is separately downloaded from a Code Base Server

✓ Architecture, Agent Server (AS)

Each agent is initiated by an agent server (AS), running on an AS in a certain place. A place provides a run-time environment for an agent by allowing it to call certain functions. A communicator is an AS module, which is responsible for sending and receiving agents among the servers. An Agent Security Manager (ASM) prevents attacks from agents, which are directed against the AS or the underlying host.

✓ Architecture, Home Server (HS)

The home server (HS) has two main duties:

1. It allows agents that finished their work to return to a special place (User Place), where they are stored and wait for user login.
2. It provides an interface for components like FE, SPE, or DAM to access returned agents or to launch the new ones.

Home server supports detached computing, i.e. users might be disconnected from the network, while the agents are performing their work. User can optionally be notified by email or SMS when the agent has returned.

1.6.3 SPARTA implementation

In Figure 15 some aspects of SPARTA implementation are shown.

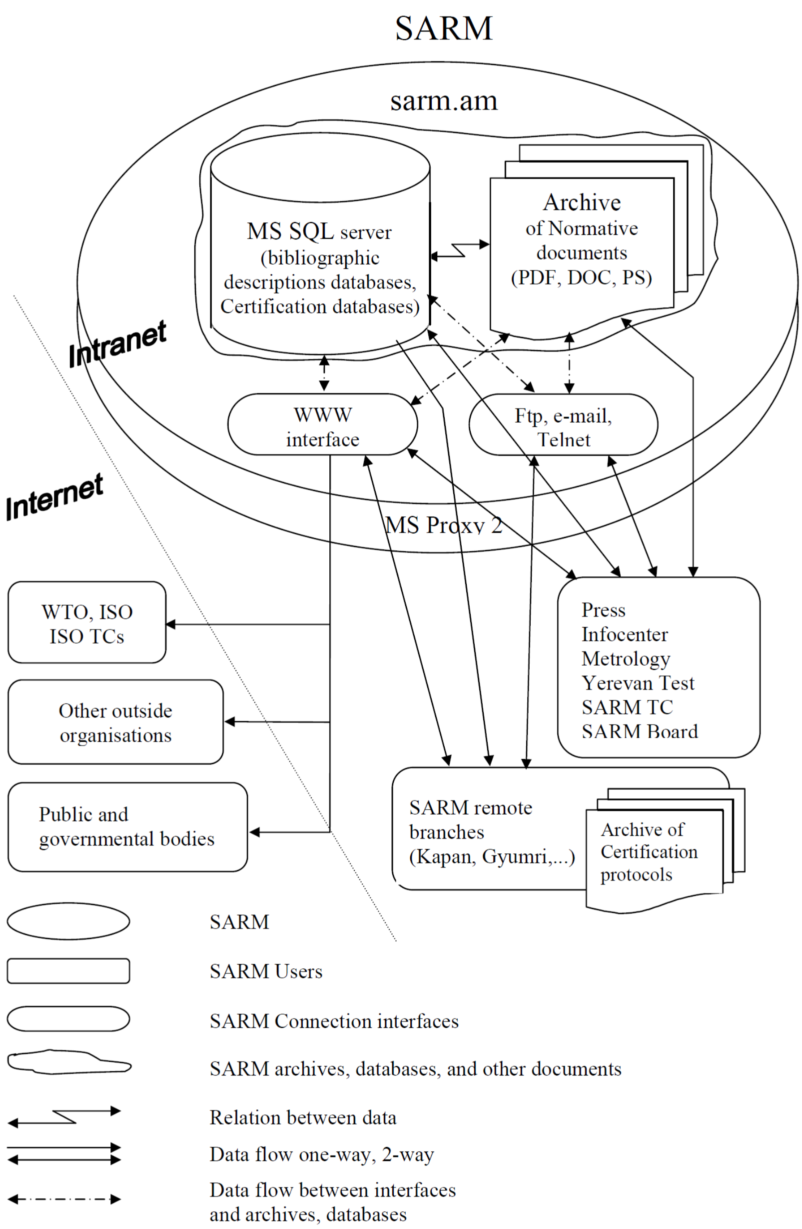


Figure 15. SPARTA implementation

- ✓ **Case studies**

SARM (State Department for Standardization, certification and Metrology), and its branch organization in Kapan (Armenia) (Figure 16):

SARM's Branch organization in Kapan manages provision of certification and accreditation process and documents, in relation to the product, services, quality system, etc. The certification body works with standardization documentation to make a conclusion regarding the kinds and services under the consideration:

– production certification (based on use of Technical Conditions);
– certification of delivered product (based on laboratory analysis);
– services certification;
– quality correspondence certification (ISO 9000).

After making conclusion, certification body sends the concluding certification document to the SARM server. Communication uses network of Infoservice Company.

Figure 17 is the scheme of the technical environment for Independent Verification and Validation tasks after the SPARTA Design and Implementation. The system specific part is validation, which is divided into subtasks:

– security policy validation, monitoring and enforcing;
– monitoring of systems integrity;
– vulnerability assessment;
– general purpose intrusion detection.

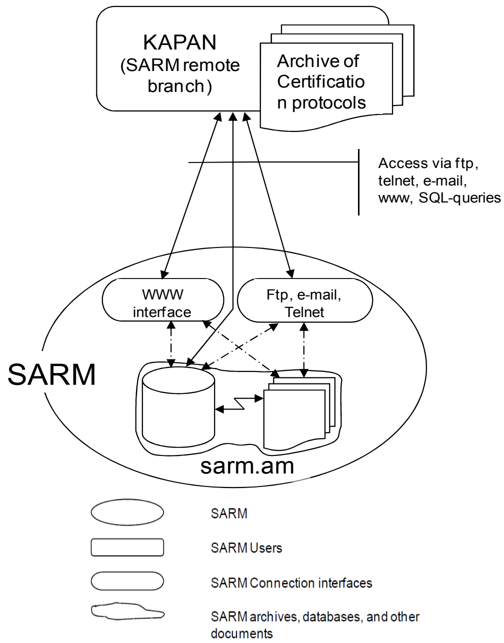


Figure 16. SARM's branch organization in Kapan

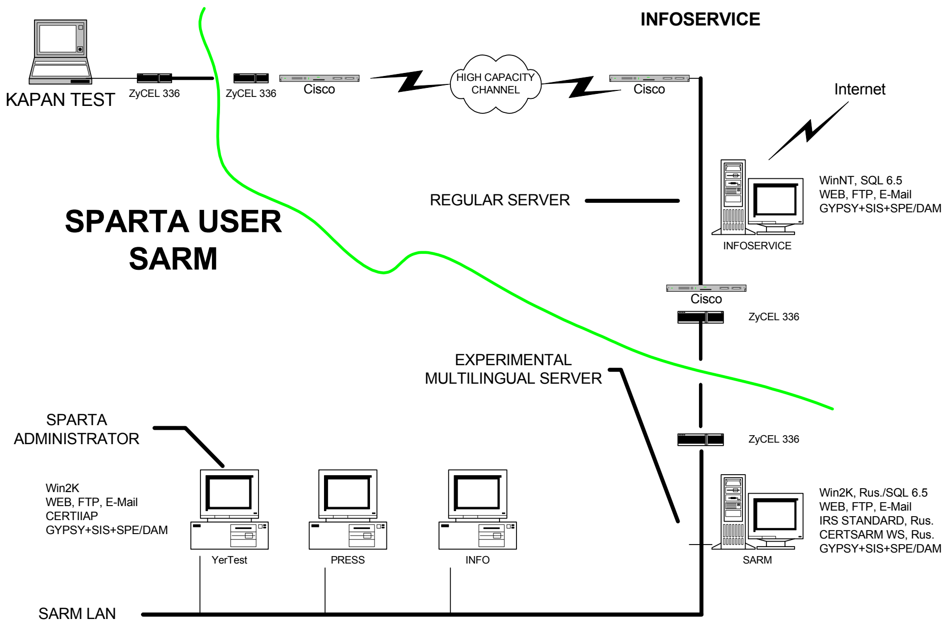


Figure 17. Scheme of the SPARTA technical environment for independent verification and validation tasks

✓ Data analysis module architecture

Next, the most intellectual component of system is the data analysis algorithmic tool. Here the main technology used is Data Mining where the practical algorithms were selected during the studies of Data Mining technologies through the INTAS research project described below (Figure 18).

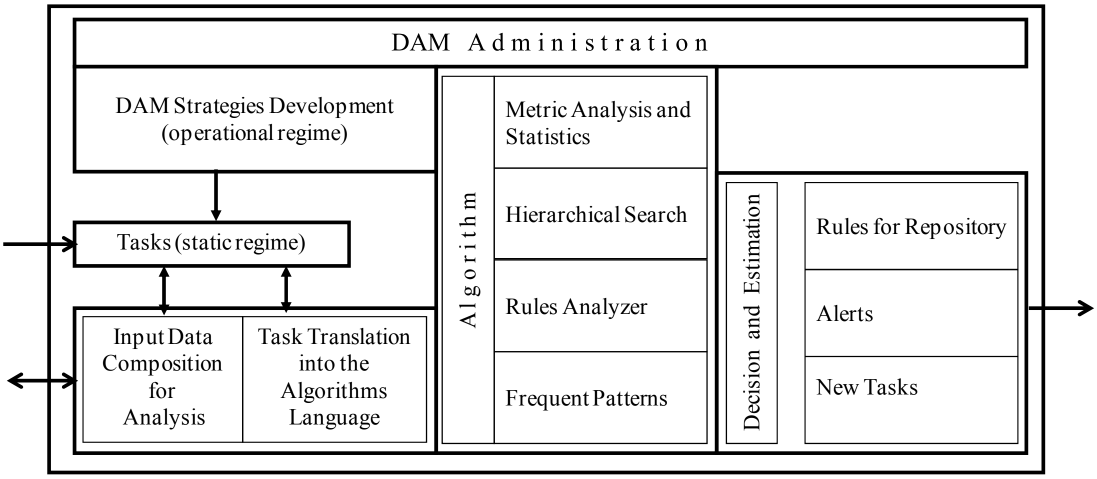


Figure 18. SPARTA data analysis module architecture

Each task of DAM has been formed by determining the input data, the data analysis method and – the output actions and forms (alerts, e-mail, SMS, special files, etc.). Three classes of task scenarios are assumed:

1. Fixed task scenario
2. Formal task scenarios by selection and composition (for experimentation by administrator)
3. Background (long term or permanent) search for regularities

✓ Fixed task scenario

In this scenario the input data types, sources, data analysis methods and output forms are determined and fixed during the design time on base of study of user's typical requirements. The individual tasks might be selected and activated by a user from the task list (it is the responsibility of system administrator to specify the data types for data collection, data analysis methods, by the ordered data and the type of interpretation of data analysis results and decision).

✓ Formal task scenarios by selection and composition

This class assumes the following cases:

- input data format is fixed/selected and user selects the data analysis methods for/by this format;
- data analysis method is fixed/selected and user selects data format for the input data;
- user selects any combination of input data formats and data analysis methods and the system checks the validity of the formed combination before the implementation.
- background search for regularities: in this case, the regular user similarly to the previous cases has formed new (experimental) DAM tasks. The functionality mode is automatic.

1.6.4 INTAS 00-652 "Data mining algorithm incubator"

- **Hybrid recognition schemes**

Monitoring systems collect huge information amounts, which require novel algorithmic approaches to be able to provide an online analysis. In recognition and classification where learning set is known as very limited in size, the first priority is the detailed information analysis. The shift of input from learning set to monitoring information requires restructuring of recognition algorithms. Hybrid recognition works in two stages. First is a quick tree based procedure. Then works a metric recognition procedure, but these work is with "error classes" after the first stage, which are much small in sizes [Aslanyan et al, 2007].

- **Frequent patterns mining**

Association rule mining is one of the basic data mining tools. The known realizations use "growing" of frequent subsets as the way of finding association rules. An alternative approach is developed, which uses the n-cube geometry elements. Monotone Boolean functions given by an oracle are recognized optimally through the special n-cube partitioning into the monotone growing chains. A modification of this structure is supported by chain computations, proven productive for frequent subsets finding [Aslanyan and Sahakyan, 2009].

- **Logic separation recognition**

Logic based Pattern Recognition extends the well-known similarity models, where the distance measure is the base instrument for recognition. Initial idea is under consideration since 70s and it reduces the logic based recognition models to the reduced disjunctive normal form of partially defined Boolean functions. An alternative pattern recognition approach combines the metric and logic hypotheses and features, and leads to studies of logic forms, hypotheses, hierarchies of hypotheses and effective algorithmic solutions. Current results provide probabilistic conclusions on effective recognition by logic means in a model environment of binary attributes and of data flows [Aslanyan and Castellanos, 2007], [Aslanyan and Ryazanov, 2008].

1.7 Intrusion detection systems

All information systems are vulnerable to computer attacks, in which they can completely fail. Computer viruses can also distort the functions of information system and add the undeclared capabilities. To prevent such situations it is required a robust information security systems. One element of such systems is intrusion detection systems. Initially, system administrators detect intrusions, sitting in front of the console and analyzing users' actions. Then began to be used logs user actions, the volume of these journals has increased with increasing number of users of information systems, so examine all records at once, administrators could not.

In the early 1990-s, there was developed intrusion detection system in operational mode, which looked through logs as soon as they are generated. Intrusion detection systems (Intrusion Detection System, IDS) is a software or hardware, designed to detect unauthorized access to the facts of a computer system or network or unauthorized management (mainly through the internet). Intrusion detection systems are used to detect malicious activity that may compromise the security of computer systems. Such activity includes network attacks against vulnerable services, an attack aimed at increasing the privileges of unauthorized access to sensitive files, as well as the actions of malicious software (Viruses, Trojans, and Worms).

An important characteristic of intrusion detection systems is the method of data analysis. There are two methods:

1. ***Detection of anomalies:*** In this case, the models of expected behavior of users and applications, the deviation from "normal" behavior of the system is interpreted as a violation of the protection [Denning, 1997]. The main advantage of such systems is that they can identify previously unknown attacks, but this is a high probability of false alarm due to significant variability in the information system.
2. ***Detection of abuse:*** Such systems contain descriptions of attacks (signatures) and looking for compliance with the descriptions in a controlled flow of data to detect manifestations of known attacks [Ilgum and Kemmerer, 1995]. The number of false positives in this case, very little, but these systems can only identify known attacks; therefore, the base of signatures must be constantly updated.

Typical architecture of an intrusion detection system is shown in Figure 19.

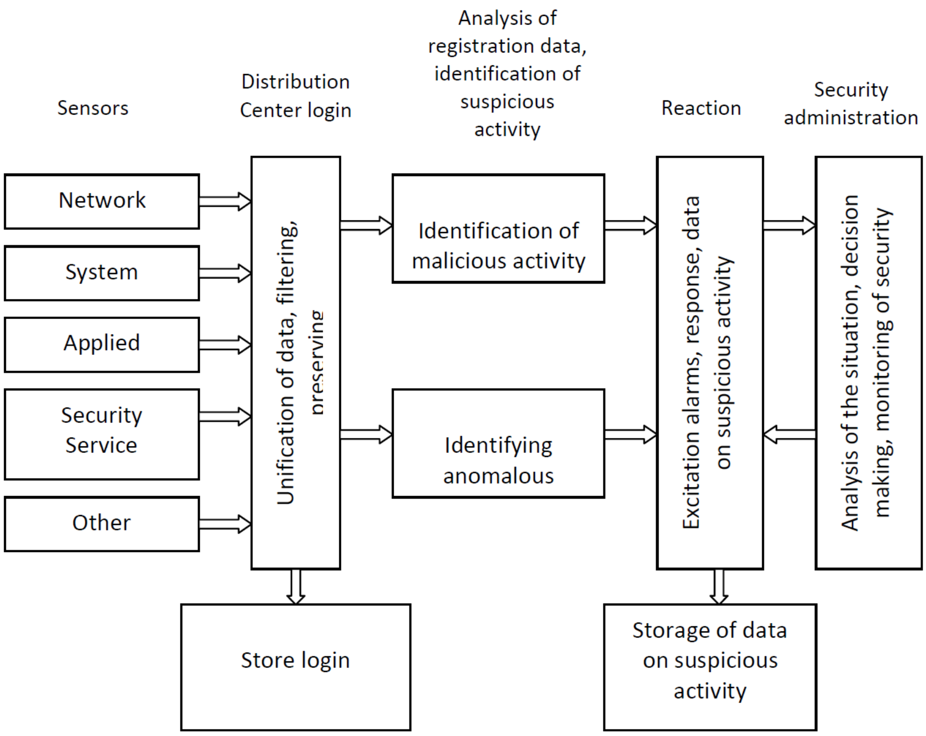


Figure 19. Architecture of an intrusion detection system [Galatenko, 1999]

Primary information gathering carried out by various sensors. To facilitate the work of all components of the system, the sensors can perform filtering of recorded data, but it is significantly complicate the structure of sensors. The sensors transmit the information received in the distribution center, which leads her to a single format, if necessary, was further filter and store in a database and send for the analysis of statistical and expert component. If the process of statistical or expert analysis identified suspicious activity, a message will send to the center response, which identifies whether the alarm to be justified, and chosen method of response.

1.7.1 System SNORT

Snort is a network intrusion detection system (IDS) with open source, which can perform real-time analysis of IP-packets transmitted on the monitored interfaces. Snort detects attack by combining two methods: the signature and protocol analysis [Yaremchuk, 2007].

All the information collected by the detector Snort could be saved to the log files that may have a different format. In addition, for ease of analysis, information can be entered in the database (Postgresql, MySQL, unixODBC).

The system, built on Snort, has the ability to collect and process information from multiple spaced sensors. All of the cases are in performance of computers that are used as sensors. In order to improve performance, sharing the quick work of IDS to capture packets and relatively slow in entering the information, you must use Barnyard, which is available on the upload page draft Snort. In this case, Snort creates a special binary output "unified" format, which working Barnyard.

Snort not directly able to work with 802.11 wireless networks, but connected to such a device will be able to interpret the information received. Today Snort in principle does not distinguish whether what type of network he has to do, no specific options to install and no need to configure.

There is a special branch of Snort – Snort-Wireless (snort-wireless.org), just designed for the detection of attacks aimed at the network 802.11. Snort-Wireless is backwards compatible with Snort 2.0, when it contains some specific rules for processing packets that are configured on the vulnerability and the typical attack wireless networks.

1.7.2 System OUTPOST

Intrusion Detection System "Outpost" version 1.3 is designed to identify and block network attacks in information systems (IS) [Outpost, 2010].

System of "Outpost" reveals an attack on user workstations, servers and communications equipment IP. It allows equally effectively identify and block attacks from both external and internal perpetrators.

The system of "Outpost" reveals an attack based on information collection and analysis of IP data packets on the network, transport, and application level stack TCP / IP.

"Outpost" uses signature-based method of detecting network attacks. It provides detection of attacks based on special patterns (signatures), each of which corresponds to a specific attack.

Getting network traffic, the system analyzes it to ensure compliance with these patterns of attacks available in the database system, which is constantly updated by the developer. In case of detection signatures in the original data, the system fixes the detection of network attack. The administrator has the ability to add to the system the new attack signature.

1.8 User-oriented policy for managing risk and safety

The growing frequency of natural disasters, their serious impacts, and huge damage have become to be considered one of the modern challenges of the countries in Europe and all over the World. The issue of having a preliminary assessment of the likely damage and ways to mitigate or even avoid it has attained an ever-bigger significance.

The experience gained in recent years from various national and international projects aimed at finding solutions to minimize damage from natural disasters and creating workable models of forecasting and assessing potential future damage reveals two components as being major factors for achieving good results: a good methodology and access to the necessary spatial data and information systems. Given that there have already been developed methodologies and simulation models of the respective disaster, what remains an open issue is ensuring the necessary data, in an appropriate format, with the relevant compatibility so that they can be combined with other data and processed as a package so as to obtain the required synergy effect and results as close as possible to the real ones. Thus good methods, simulation models, and software are reduced to zero when data sources are not available or they are not accurate, updated, and easy for use. The problem is related to the adoption of common rules for data exchange, data storage and formatting, etc., which are the subject matter of European Directive 2007/02/EC setting the legal framework for the establishment and operation of an Infrastructure for Spatial Information in the European Community (INSPIRE), but which also call for the introduction of rules vis-a-vis interoperability of the information systems.

Therefore the presented article is combining research & development topic – as modeling preliminary losses from natural disasters (earthquake in this case), and conceptual topic – as the development of a quality assurance reference data base and regional networking for data storage & processing and for risk and security management. One possible approach on the last item is creating an infrastructure to ensure such exchange of spatial data and their processing, and to set up a network of application servers for spatial information in different parts of Europe, which work in harmony and constitute parts of a unified system. Such an approach is considered in many respects be better, more effective, and yielding faster and better quality results than the creation of a central server, where primary data are fed by various municipalities, regions and countries.

1.8.1 General overview

In the course of several millennia, the human evolution turned the human being from someone who inhabited forests and various natural earth cavities into a geological factor who built first small, and then ever bigger and bigger artificial complexes – buildings grouped in even greater zones: population and industrial agglomerations. The matter moved to be input in these urban agglomerations is commensurate with the quantity expelled by major geological disasters – volcanic eruptions, or moved by tectonic processes – earthquakes.

Irrespective of his incessant effort, man cannot yet boast that he has tamed the natural force and the cyclic occurrence of hurricanes, earthquakes, fires that wreak havoc and claim many lives. No one has yet "won a battle" against a natural disaster, one can rather speak of avoiding the worst and mitigating the damage. This is why a "winning war" with the dark side of the nature would need a careful preparation of "the next battle" – an appropriate system for forecasting of possible disaster-induced damage and the strategic planning and preventive measures that go together with it.

1.8.2 Local and global approach

It is common knowledge that proper measures taken beforehand reduce many times over the damage from a possible natural disaster as well as the resources and money necessary to overcome the consequences. Given that a natural disaster may strike at a particular place but can affect a large geographical area sometimes covering several countries, the work should be aimed at achieving coordinated estimates and measures simultaneously, on the local, as well as on the inter-regional, community or even global, level.

This is why the respective methodologies, spatial databases, simulation models and outputs need to be compatible, aimed to achieve a synergy effect and create a platform of interoperability.
Five levels of management are identified:

- municipality;
- country;
- region;
- European Union;
- global level.

Halfway levels are possible too, such as unification of policies for risk management in a given area of impact, for example the catchments areas of big rivers. A good example of this would be the creation of an integrated system for prevention against flooding and water pollution along the Danube within the future European Danube Strategy.

In respect of earthquakes, an intermediate form of unification could be sought of policies for risk management in areas affected by known seismic foci or geological areas, e.g. significant faults in contact areas of tectonic plates.

Set out below is an opinion about the necessary exchange, and in certain cases, integration of spatial databases and information systems regarding seismic risk.

1.8.3 Methodological basis

The methodology for forecasting damage caused by natural disasters, particularly earthquakes, mainly targets the residential urban environment and the public utility and industrial infrastructure related to it, which make up the modern city. The methodology complies with the requirements to yield results that are useful for the end users (local and central powers, business entities, civil protection), as well as to tackle the existing problems with the available spatial data, unreliable information and absence of information.

Why is such a methodology an indispensable element of the effective policy for managing risk and safety? This is so for a number of reasons, namely:

- buildings differ in type, mode of construction and age, and thus have different vulnerability to catastrophic seismic impacts, so that the damage they will sustain will be different;
- the modern urban society has never had such surplus of energy, resources and temporal capacity so as to allow the construction of buildings able to withstand the possible maximum seismic impacts;
- rigorous construction supervision is missing (especially in the time passed); many violations are made on town-planning indicators and the requirements for free spaces and distances between buildings, laid down for safety reasons;
- the policy for sustainable development calls for judicious spending of resources for prevention against seismic impacts, allowing such expenditure only where the alternative to building a new earthquake-proof building or the reinforcing of an old one would yield an immediate positive effect for the public, either in terms of safety and quality of the living space, or as industrial

economic infrastructure, or also as protection of an intangible value of historical or emotional essence;

- recently we have witnessed (in Bulgaria) a dangerous scheme used by developers who opt for fulfilling their interests by contravening the law and even creating a threat to people's lives and health. In the scheme certain buildings which have public significance, constitute monuments of culture, or are banned from demolition by law, are left in disrepair or else their process of degradation is accelerated artificially, so that they reach the stage of self-destruction and can thus be freely removed and replaced by a new development project which usually exceeds considerably the permitted town-panning parameters. The scheme in question poses a double threat – on the one hand the disintegrating building poses a threat to the members of the public and to the buildings and infrastructure next to it (a fatal accident of this type occurred in Sofia only two years ago), and on the other, the new construction usually fails to comply with the safety criteria and standards and becomes a potential threat during a future earthquake.

Driven by various reasons the modern society (and particularly United Europe) has developed adequate methods for observing a given disaster and for managing economic flows during recovery; however, a single methodology is yet missing for forecasting disaster-induced damage, including in terms of individual risks and topics. This is especially true of places where the data necessary for making the estimates are rarely available in the necessary volume, accuracy, and credibility.

In such a case, the option is to look for a methodology that is "sustainable" and resilient to data incompleteness, a methodology that is nevertheless applicable, albeit with a reduced accuracy of results. The most important thing in this case is for the methodology to provide a good idea about the areas with increased risk and a reliable general assessment of the likely damage. The methodology needs to be able to be applied both at the local level – a neighborhood or a town, as well as at a greater one – a region or a country. At the next stage partnership is to be sought with experts and structures from neighboring countries to ensure interaction between such methodologies of two countries or more, as well as the exchange of data, models and results.

Such a methodology will prove very useful for risk assessments vis-a-vis urban centers with limited data about housing and public utility infrastructure.

The methodology takes into account the state of the spatial databases and information systems in Bulgaria and the actual difficulties in providing full, detailed and objective information. In this respect the situation in Bulgaria is similar to that in many other EU member states, meaning that the methodology can be successfully applied in the EU, as well as in other geographical regions.

The methodology, developed by a team of experts in different areas: geophysics, economy, town planning, informatics, has three successive layers:

1. risk assessment of housing
2. risk assessment of public utility and economic infrastructure
3. assessment of economic damage

Assessment of the so-called "intangible damage" has not been envisaged, as at this stage it proved impossible to validate in the respective software the connection between intangible or information values and current tangible values and assets of the modern city.

Several versions of the methodology have been set out, for example – for assessment of earthquake-induced damage of residential buildings; for an overall assessment of the seismic risk; for estimation of the economic losses from natural disasters.

1.8.4 Methodology for assessment of earthquake-induced damage of residential buildings at the level of municipality, district, country, region

The methodology has been used on test cases for the Bulgarian cities of Sofia, Ruse, Vratza, and for the whole country. First priority was given to results, which can be of use for the local and state administration for planning their economic, and resources actions as well as for urban planning, building rehabilitation, and civil protection.

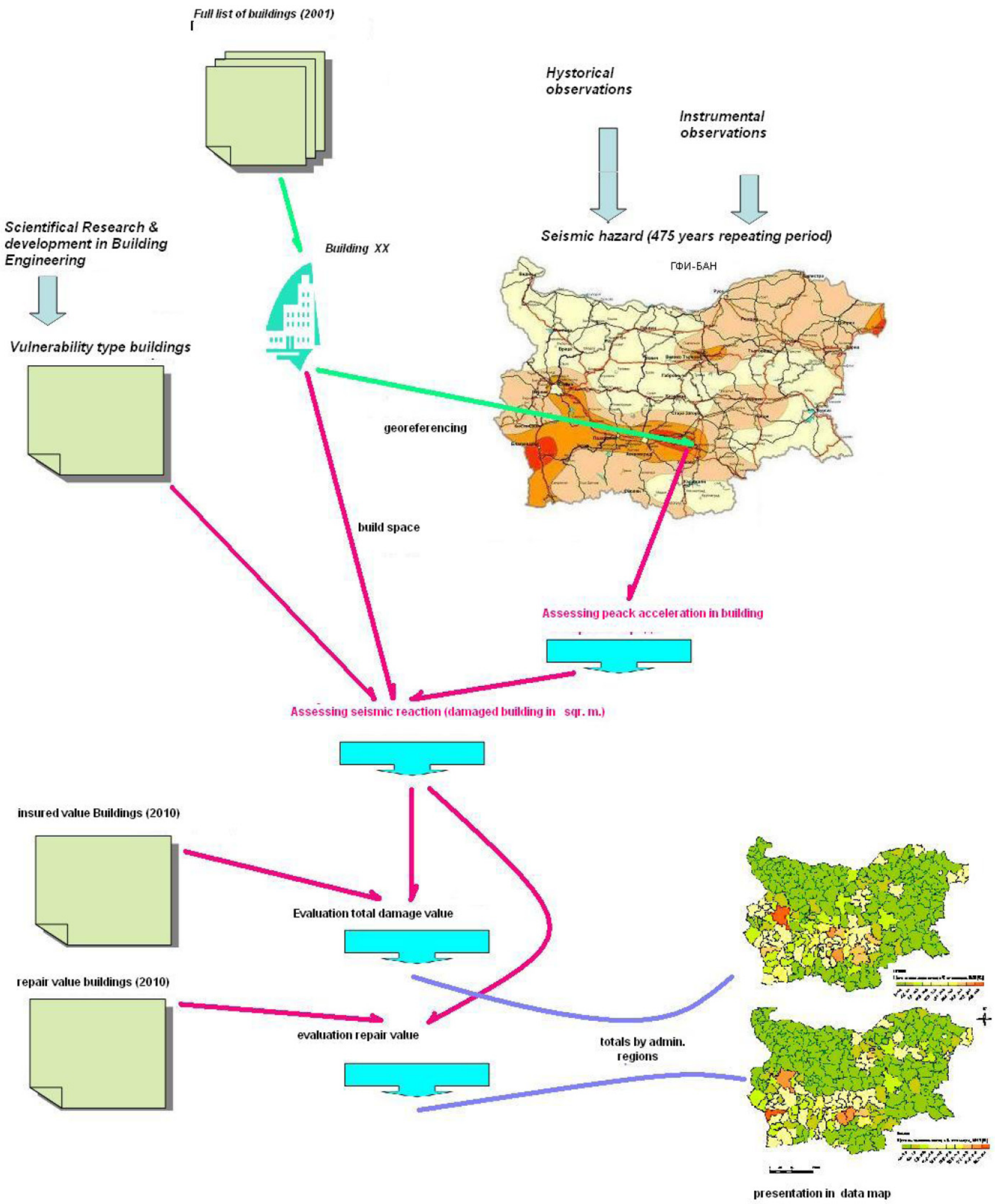


Figure 20. Simplified technology for seismic damage assessment of dwellings

The methodology is based on experience gained from work under the RISK-UE project in Europe and HAZUS in the United States in the late 1990s and the beginning of the new millennium. It starts from:

– availability of data about catastrophic impact (in the region);
– availability of data about housing (in the region);
– availability of a vulnerability function (vis-a-vis the respective housing related to catastrophic impact by type of buildings and type of impact).

To obtain an assessment of the expected damage, systematic modeling is made of the catastrophic impact for each registered housing object, depending on the planned maximum catastrophic impact.

The methodology concerns assessment of the likely damage on residential buildings caused by an expected earthquake with a given magnitude. It is applied on information about housing and the occupants thereof available in the respective National Statistical Service. The methodology is apt to be applied in assessing the value of small to medium damage (for example, up to 20-25% of the area), whereby the recovery value increases non-linearly when the percentage of damage is higher, as costs will be provided for clearance.

A brief scheme of the methodology is shown in Figure 20.

1.8.5 Methodology for overall assessment of earthquake damage

The methodology is in the process of being developed; the authors have set themselves the task of extending it (for housing infrastructure) to include the public infrastructure (the urban infrastructure in particular) in the actual situation of missing workable databases about buildings and other urban infrastructure. To this end, partial use is made of the methodology for structuring of the urban infrastructure in the European project RISK-UE (WP3) (Figure 21).

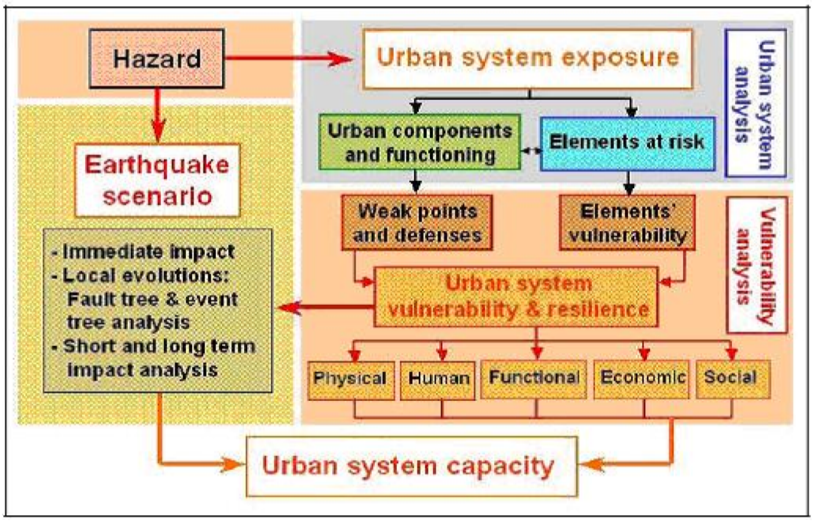


Figure 21. GEMITIS' seismic risk analysis

From the overall approach, the part covering the elements at risk complementing housing is being extracted.

The simplest option would be to apply the basic methodology on sets of infrastructure descriptions. Nevertheless, here lies the main obstacle, in this case – for Bulgaria – the absence of a detailed and comprehensive description of this infrastructure.

1.8.6 Extrapolation of the housing damages to damages in the public infrastructure

A more complex option would be to arrive at a system for extrapolation of the public housing data to a mechanism for assessment of infrastructure data as a main task in risk assessment.

Solving such a problem is not possible using simple linear extrapolation of housing damage (a purely physical assessment).

In the first place, the purely housing damage – provided that it does not exceed a minimum percentage (e.g., 5%), can be serviced in fully functioning infrastructure (i.e., without, or with only negligible losses in the public infrastructure). Then recovery is seen as part of the normal functioning of society, rather than an emergency activity. This is why losses in the public infrastructure should be considered from a certain threshold and above.

After the initial version, mechanisms can be applied as well for taking into account the impact of losses in the public infrastructure in the sub-scheme of their structuring according to the above-mentioned model of WP3-RISK-UE:

- vis-a-vis physical damage (collapsed buildings) of the public infrastructure;
- vis-a-vis functional damage of the public infrastructure (impossibility to function fully);
- vis-a-vis economic damage (from own collapsed facilities and from a diminishing workforce busy with recovery of primary needs);
- vis-a-vis functional and economic damage from newly emerging urgent social & material needs;
- vis-a-vis "social" (non-material historical, marketing and similar) damage as a result of the disaster which diminishes the economic value of the respective area.

Further, mechanisms are applied for transformation of housing damage to damage on the public and economic infrastructure. Some preliminary conditions are accepted as:

- going beyond a certain percentage of damage (e.g., over 10%) suggests additional costs for rescue, evacuation of population without shelter, setting up of medical aid, security and safety, provision of food for the population remaining in place;
- hurricane Katrina, and even more so, the earthquake disaster in Haiti, revealed the considerable shortage of capacity of the security services in the event of a large-scale disaster;
- in the case of more than 10% affected (destroyed) housing, emergency costs are to be provided for temporary housing to shelter the surviving population until recovery of at least 60% of the volume of housing from before the disaster;
- going beyond a certain percentage of damage (e.g., over 5-10%) suggests non-linear increasing of costs for site clearing. If more than 35-50% of the housing has been affected (has collapsed), from 10% to 40% costs (of the building value) should be provided for site clearing and environmental-friendly processing of waste;
- as a rule, the normal construction activity for a given region is determined as being designed to build/recover no more than 5% of the available housing (i.e., a term of restoration of housing of no less than 20 years). Above this percentage all activities for clearing and restoration constitute ADDITIONAL capacity for construction and repair works which is not available in situ;
- more than 35-50% affected (destroyed) housing suggests total destruction and a need to build anew the houses affected by the disaster;
- when the affected (destroyed) housing exceeds 10%, a relevant percentage of destruction of public utility infrastructure should be assumed – schools, hospitals, which could vary from 5% to

15-20% of the value of the housing (depending on the characteristics of the social environment at the place of the disaster). This should be taken into account, irrespective of the application of specially heightened seismic requirements vis-a-vis such public buildings in certain areas (e.g., in the United States). When more than 40-50% of the housing has been affected, total destruction of the subsurface public utility infrastructure servicing the given housing should be assumed. Respectively, rebuilding of the public utility infrastructure should be provided;

- in recovery actions in a region heavily hit by a seismic disaster, costs for the transport infrastructure should also be provided for, which need to be higher than the prices of such infrastructure during normal times of construction;
- when more than 15% of the public utility infrastructure has been affected (destroyed), urgent building of temporary replacement public utility infrastructure should be provided;
- in the event of a seismic disaster (affecting more than 5% of the housing), shortage of construction capacity should be assumed at the place of recovery and a subsequent increase (sometimes multiple) of the economic value of the restoration in relation to the own value of the destroyed infrastructure (both housing and public infrastructure).

1.8.7 Extrapolation of the housing damages to the economy

Together with the social and public utility infrastructure, a seismic disaster will also damage the economic infrastructure of the given region.

In the modern societies the economy constitutes up to 70% services (commerce, public utilities, energy supply and supply of staples, education, medicine, administrative services), which are proportionate to the available housing infrastructure and which are affected/destroyed in the same degree as the housing infrastructure.

For example, in the observation of the restoration works after the 2008 earthquake in China, destruction of a considerable part of the existing economic sector was observed on the one hand, and on the other hand the replacement thereof by a newly-emerged construction sector for restoration funded by the Chinese budget.

The industrial economic infrastructure will reduce its capacity by a given percentage of the overall damage to the housing but only within certain limits of the total disaster (for example, up to 20-30% of the housing). Individual sectors may witness less reduction when they are related to restoration of the infrastructure (construction, etc.). Above this percentage anticipatory increase of the damage of the industrial infrastructure can be expected, in parallel with the reduction of the workforce being used on the one hand (which has been redirected to salvage and restore housing), and on the other, on account of the destroyed transport infrastructure, and still other, on account of a change in the local consumption caused by the restoration boom.

Depending on the geographical location of the industrial facilities, they too sustain damage, proportionate to that of the housing infrastructure. In the majority of industrial productions, the damage caused by a seismic disaster can cause an additional disaster: fire, pollution with chemicals, bio-agents, radioactive substances, etc., gazing, flooding, etc., which (depending on the concrete industry) can increase the damage to more than 500% from the initial value of the industrial facility.

In large-scale disasters, there inevitably occur economic losses due to a need for restructuring the economic infrastructure.

An example for a non-linear, non-proportionate (as set against that of housing) increase of damage is shown in Figure 22.

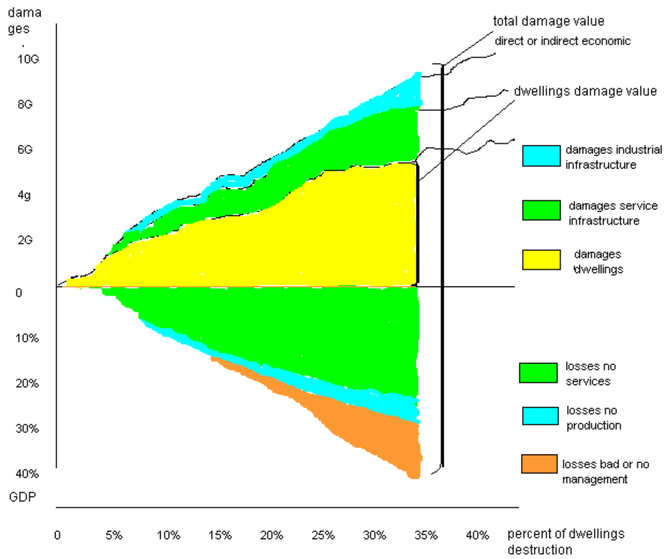


Figure 22. Non-linear, non-proportionate (as set against that of housing) increase of damage

1.8.8 Methodology for assessment of economic damage from natural disasters

This methodology constitutes an extension of the methodology for determining the total damage from a seismic disaster. In the presence of a baseline system for assessing housing damage caused by a catastrophic natural disaster (with taking into account of the "national" peculiarities), it is possible to make a multiplication scale which identifies parallel and subsequent damage on public utility and industrial infrastructure, and thus, an assessment of the relevant economic losses. A next step should analyze the cumulative effect from non-direct impacts, intangible assets, stress and psychological breakdown of big masses of the population.

A percentage of loss/destruction of public utility and industrial infrastructure commensurate to the housing infrastructure should be accepted as a basic criterion for economic loss, whereby the said infrastructures generate:

- public utility services (including transport, energy, communications, water and sewerage);
- material flows from industrial infrastructure;
- as a consequence of a reduced GDP – reduced tax revenue, restrictions/cuts in the government budget intended for restoration;
- additional reduction of the capacity for restoration and, respectively, the rates of GDP recovery.

Undoubtedly, a more precise assessment would be yielded by precise modeling of a "facility at risk" from a detailed map of the locations of public utility and industrial infrastructure compared to the map of the magnitude of the catastrophic impact via the vulnerability function, when such a map appears. Now information about the presence of such an integrated map is not available for any country in Europe.

As we said earlier, the public utility services are closely linked with the housing and the destruction of the former ties in very well with the damage on the latter. It is logical to recalculate the total volume of affected public utility services, via the square meters of housing area and the total lost/damaged square meters of area.

By way of an example, we offer below the preliminary assessments about Sofia, without claiming that they will not be adjusted when additional information and up-to-date data are provided.

Output data about Sofia, 2007:

- Gross Domestic Product 20,576,000,000 leva
- Gross Surplus Product 16,900,000,000 leva
- Redistributed via budget – 40% of the GDP, or 6,800 M
- Breakdown: Services 78.5%
- Industry 21.2%
- Forestry and farming 0.3%
- In services – 8.5% (increased, prior to that about 5%) construction, which will not decrease in the case of disaster.

With a total GDP per annum (2007) of 20,576 M for the territory of Sofia City, and projected maximum damage of 32% of the housing area, the total planned economic loss from services will be 5,530 million Bulgarian levas, or about the size of the direct damage of the housing lot.

The impact on the industrial sector (extracting and processing industry) of a seismic disaster as an assessment from a correlation with damage on housing is weaker than that of the service sector. Here we first have an impact threshold (above 5-10% affected homes from a separate industrial area). Secondly, the industrial activity is concentrated in separate industrial zones, mostly outside population centers, so that heavy damage of high-rise housing would not have a significant impact on the industrial capacity. In this context, it can be safely assumed that the percentage of damage of the industrial infrastructure will be no more than half of that of the housing infrastructure.

If we again take as an example Sofia City, the estimated projected economic losses would be to the tune of 411 million BG leva, or nearly a ten-time lower impact than a disaster in the service sector.

The occurrence of additional damage due to events caused by the primary disaster, such as fires, people/equipment buried under ruins, floods, pollution, etc., has relevance only for specific industrial infrastructure. In conducting a preventive analysis, a study should be made of available major industrial or energy facilities where a fire, pollution, flooding could occur in the case of disaster.

An additional damage from a disaster in the area of Sofia could be expected in the form of above-proportionate shrinkage of wholesale trade (up to 20-30%) and mostly, above-proportionate reduction of administrative and governance services since the administration is concentrated in the area of Sofia. Then the ensuing economic effect could reach up to 20-50% of the economic losses from services, or in the case of Sofia, this would mean another 2,000 million levas in damage.

The damage from necessary restructuring of the economy following a disaster is difficult to estimate precisely enough.

1.9 Discussion

Every year many collapses happen in the world, which can be divided into three groups:

- natural disasters
- technogenic disasters
- information disasters

Natural disasters are hurricanes, floods, earthquakes and other acts of nature which destroy the whole cities and lead to mass deaths of people.

Technogenic disasters are big accidents on industrial and transport objects which have been caused by failure in work of technical systems. Technogenic disasters are accompanied by victims among people and ecological disasters.

Information disasters are collapses, which occur in information computer systems. Mainly they occur because of viruses and other harmful programs.

Several examples of these types of collapses and opinions of scientists concerning their reasons were outlined. Examples of different intelligent systems for prediction, risk analysis, and intrusion detection systems were given.

It is clear for most experts that there is little likelihood for a comprehensive methodology yielding as truthful results as possible to be devised in the next couples of years in the European Union. Examples for such failures are found in risk prevention in the United States in cases such as Hurricane Katrina, in Europe in the recent floods in Central and Eastern Europe, the earthquake in China, etc.

Thus, above all we want to promote the pragmatic systematic approach in the creation and application of methodologies and models for risk management first on regional level, with a possibility for parallel verification in situ.

Developing of regional focal units, engaged in a large network of intelligent systems, covering the EU and later larger regions is a probably the best sustainable solution.

The absence of sufficient information in terms of diversity, up-to-date relevance, and correctness makes it necessary to look for effective use of the available data, whereby in most cases these constitute statistical information.

Two mainstays need to be used in order to achieve good results:

- European directives and regulations;
- strengthening the cooperation within the broad European area, which includes both EU member states and neighboring countries, based on the user-oriented approach which is better achieved if the regional and intelligent networking concept is accepted.

We need elaboration of national methodologies, simulation models, and analytical models for damage assessment, based on local data, which are subject to constant coordination and harmonization and in direct connection with the user – national governments, civil protection units, local authorities. In this sense a main partner of the regional and national teams could be some structures of the European Commission as the Directorate General Enterprise and Industry (more particularly, the department for Competitiveness, internal market for goods and sector policies), Directorate General Humanitarian Aid, Directorate General Joint Research Centre (more particularly, the Institute for the Protection and Security of the Citizen).

Meanwhile, a number of countries, as well as European structures, have set up relevant websites with up-to-date information for expected or past disasters. It would be a good thing if, on the basis of peer initiatives, and in addition to the information they upload, these sites adopt a unified intelligent system for presentation of information, including:

- automatically uploaded up-to-date information about actual events, including by regions and topics, e.g. fires, floods, etc.; the information should be stored, analyzed, and accessible by time periods, e.g. weekly, monthly, yearly;

- the system should be open for inclusion of other sources of information, as well as for making comparative analyses;
- the system should also include addresses of local, national, and European structures responsible for making of forecasts or for actions in the case of natural disasters.

In this presented above sense good opportunities are presented by:

- the European Directive 2007/02/EC setting the legal framework for the establishment and operation of an Infrastructure for Spatial Information in the European Community (INSPIRE), Directive 2003/98/EC on the re-use of public sector information and others;
- the creation of regional, and subsequently, European network of application servers for analyses and risk monitoring, which exchange data and information based on common standards and rules for exchange and processing and which can work as a super-computer when fed with particularly large amounts of data (Figure 23).

This can be developed jointly with several DG,s as DG ECHO, DG ENTR, DG JRC, under the Global Monitoring for Environment and Security (GMES) Program and the future realization of Global Earth Observation System of Systems (GEOSS).

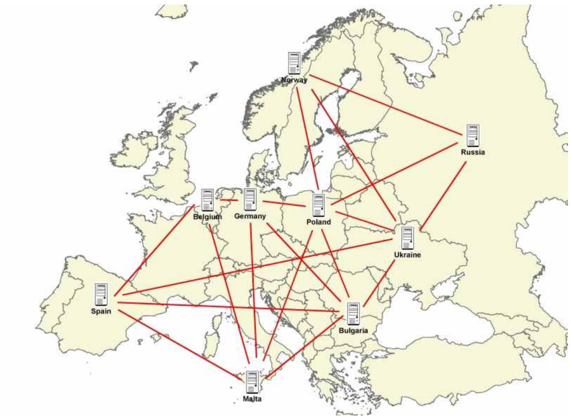


Figure 23. European network of application servers for analyses and risk monitoring

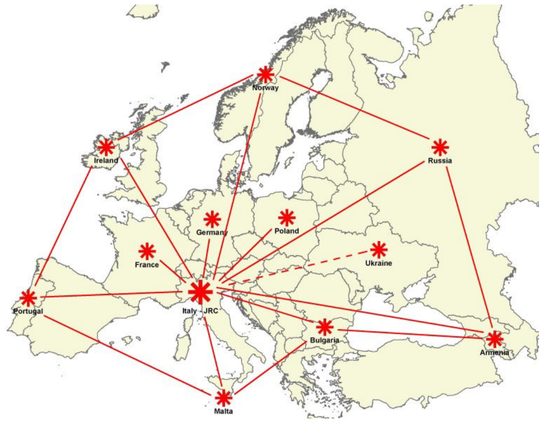


Figure 24. European network of regional units for risk and security management, land cover changes monitoring and data quality assurance

In close cooperation with the previous servers network, another regional network is needed to be developed – a network of regional cores (or units) for risk and security management as well as monitoring land cover changes and spatial data quality assurance, and subsequently, a European (or European-Mediterranean) network, also within the framework of the operational application of the GMES and the future GEOSS (Figure 24). The latest concept is very much open to the new tendencies and priorities in the EU 2020 strategy, as well as quality assurance (QA). It will take into account as well, the risk for the citizens, created by the ICT approaches and relevant technologies.

2

High-performance Intelligent Computations for Environmental and Disaster Monitoring

2.1 Specifics of Earth observation problems

At present, global climate changes on the Earth made a rational land use, environmental monitoring, prediction of natural and technological disasters, etc the tasks of great importance. The basis for the solution of these crucial problems lies in the integrated use of data of different nature: modeling data, in-situ measurements and observations, and indirect observations such as airborne and space borne remote sensing data [GEOSS, 2010].

In particular, models can be used to fill in the gaps in the data by extrapolating and estimating necessary parameters to the site of interest; to better understand and predict different processes occurring in the atmosphere, land, ocean and sea, etc; they can help to interpret measurements and to design new observing systems. In-situ measurements are often used for assimilation into models, calibration, and validation of both modeling and remote sensing data. Satellite observations have an advantage of acquiring data for large and hard-to-reach territories, as well as providing continuous and human-independent measurements. Many important applications such as monitoring and predictions of natural disasters, environmental monitoring, etc. heavily rely on the use of Earth observation (EO) data from space. For example, the satellite-derived flood extent is very important for calibration and validation of hydraulic models to reconstruct what happened during the flood and determine what caused the water to go where it did [Horritt, 2006]. Information on flood extent provided in the near real-time (NRT) can also be used for damage assessment and risk management, and can benefit to rescuers during flooding [Corbley, 1999]. Both space borne microwave and optical data can provide means to detect drought conditions, estimate drought extent and assess the damage caused by the drought events [Kogan et al, 2004], [Wagner et al, 2007]. To assess vegetation health/stress, which is extremely important for agriculture applications, optical remote sensing data can be used to derive biophysical and biochemical variables such as pigment concentration, leaf structure, water content at leaf level and leaf area index (LAI), fraction of photosynthetically active radiation absorbed by vegetation (FPAR) at canopy level etc. [Liang, 2004].

The EO domain is characterized by the large volumes of data that should be processed, catalogued, and archived [Fusco et al, 2003], [Shelestov et al, 2006]. For example, GOME instrument onboard Envisat satellite generates nearly 400 Tb data per year [Fusco et al, 2003]. The processing of satellite data is carried out not by the single application with a monolithic code, but by the distributed applications. This process can be viewed as a complex workflow [DEGREE, 2008] that is composed of many tasks: geometric and radiometric calibration, filtration, reprojection, composites construction, classification, products development, post-processing, visualization, etc. For example, calibration and mosaic composition of 80 images generated by ASAR instrument onboard Envisat satellite takes 3 days on 10 workstations of Earth Science GRID on Demand that is being developed in ESA and ESRIN [Fusco et al, 2003]. Dealing with EO data, we have to also consider the security issues regarding satellite data policy, the need for processing in NRT for fast response within international programs and initiatives, in particular the International Charter "Space and Major Disasters" and the International Federation of Red Cross.

It should be also noted that the same EO data sets and derived products could be used for a number of applications. For example, information on land use/change, soil properties, meteorological conditions etc. is both important for floods and droughts applications as well as for vegetation state assessment. That is, once we develop interfaces to discover and access the required data and products, they can be used in a uniform way for different purposes and applications. This represents one of the important tasks that are being solved within the development of the Global Earth Observation System of Systems [GEOSS, 2010] and European initiative Global Monitoring for Environment and Security [GMES, 2010]. Services and models that are common for different EO applications (e.g. flood monitoring and crop yield prediction) are shown in Figure 25.

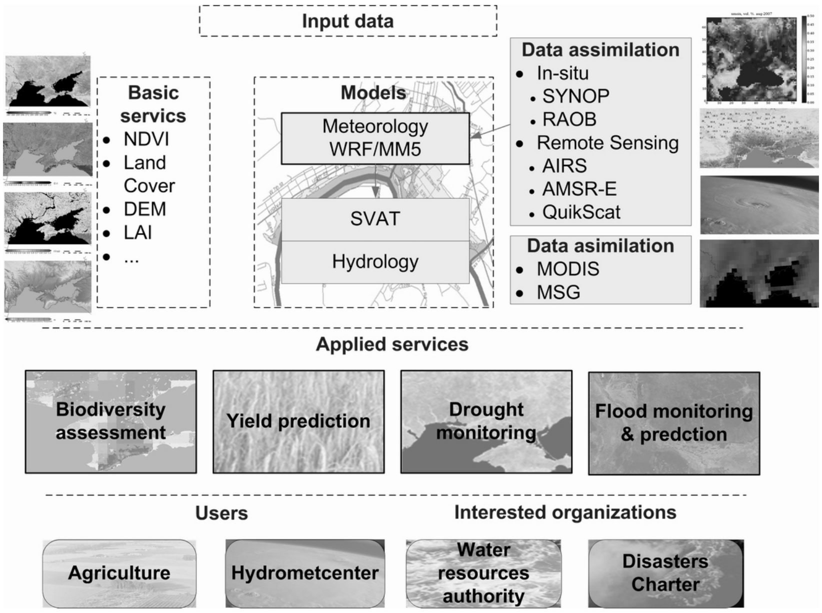


Figure 25. Common services and models for a variety of applications

A considerable need therefore exists for intelligent methods an appropriate infrastructure that will enable the integrated and operational use of multi-source data for different applications domain. From technological point of view, Grids can provide solutions to the above-mentioned problems [Foster and Kesselman, 2004], [Fusco et al, 2003], [Shelestov et al, 2006]. In this case, a Grid environment can be considered not only for providing high-performance computations, but, in fact, can facilitate interactions between different actors by providing a standard infrastructure and a collaborative framework to share data, algorithms, storage resources, and processing capabilities [Fusco et al, 2003].

In this part, we focus on the description of the Grid infrastructure that is under the development in the Space Research Institute NASU-NSAU (SRI). We will describe several real-world applications that are solved using the Grid infrastructure, namely numeral weather prediction (NWP), flood monitoring, and vegetation state assessment. We also review issues regarding the integration of the Sensor Web and Grid technologies for flood applications.

2.2 Existing tendencies and initiatives

2.2.1 Challenges

Increasing numbers of natural disasters have demonstrated to the humanity the paramount importance of the natural hazards topic for the protection of the environment and the citizens. Climate change is likely to increase the intensity of rainstorms, river floods, droughts and other extreme weather events. All these problems are among benefit areas of GEOSS (Global Earth Observation System of Systems) aiming to integrate efforts of different countries for exploiting the growing potential of Earth observations to support decision making in an increasingly complex and environmentally stressed world.

Floods. Floods are among the most devastating natural hazards in the world, affecting more people and causing more property damage than any other natural phenomena [CEOSDMSG, 2001]. In the period of time between 1900 and 2006 a total of 415 major flood events occurred in Europe alone, with an average death toll of 22 and 35159 affected people (Source: EM-DAT: The OFDA/CRED International Disaster Database, August 2006).

Ukraine is vulnerable to floods, in particular in the Carpathian region where it occurs almost every year. During the floods in 2001, 9 people were killed and 12000 citizens were evacuated, more than 1500 buildings were destroyed, and more than 30000 buildings were flooded.

In January-February 2008 heavy flooding were affecting a number of southern Africa countries including Mozambique, Zimbabwe, and Zambia. Media provisionally reported that almost 7,000 households have reached resettlement centers and more than 30,000 hectares of crops have been lost (Source: International Charter "Space and Major Disasters").

Drought: In spring-summer, 2007, southern regions of Ukraine were heavily affected by droughts. As a consequence, crops area of approximately 1,4 million ha were totally destroyed, and 8,5 million ha of crops were damaged. Due to this severe drought, financial losses for Ukraine were approximately 100 million of U.S. dollars. The event was considered as a disaster of national level (source: Ministry of Emergent situations of Ukraine).

EU countries are also constantly hit by severe drought. In the last thirty years, EU has been affected by major droughts, in particular in 1989, 1990, 1991, and 2003. The overall impact of droughts in the last thirty years is estimated to 100 000 Million Euro.

Vegetation state assessment: In the following 2 years, Ukraine is planning to launch its own remote sensing satellite Sich-2. One of the applied problems that will be solved using data acquired from Sich's instruments is vegetation health assessment in support of agriculture. Thus, it is needed to develop appropriate method for vegetation state assessment using space-borne remote sensing data.

Computational complexity: It should be stated that efficient monitoring of agricultural resources and natural disasters is almost impossible without the use of Earth Observation (EO) data from space. Satellite observations enable acquisition of data for large and hard-to-reach territories; can provide continuous measurements and human-independent information, etc.

In turn, the EO domain is characterized by large volumes of data that should be processed, catalogued, and archived. For example, GOME instrument onboard Envisat satellite generates nearly 400 Tb data per year. Space Research Institute NASU-NSAU beginning from 2006 have installed EUMETCast system for environmental data dissemination. EUMETCast that is part of global

GEONETCast system of GEOSS enables acquisition of more than 50 Tb of processed and unprocessed information per year. Moreover, the processing of satellite data is carried out not by the single application with monolithic code, but by distributed applications. This process can be viewed as complex workflow that is composed of many tasks: geometric and radiometric calibration, filtration, reprojection, composites construction, classification, products development, post-processing, visualization, etc. For example, calibration and mosaic composition of 80 images generated by ASAR instrument onboard Envisat satellite takes 3 days on 10 workstations of Earth Science GRID on Demand that is being developed in ESA and ESRIN.

2.2.2 GEOSS and GMES

The globalization and integration processes are dominant tendencies in the development of new solutions for complex problems solving. At present, international cooperation efforts are focused on the implementation of GEOSS. GEOSS is a distributed system of systems built on current international cooperation among existing Earth observing and data management systems – in situ and remote sensors and systems [GEOSS, 2010].

GMES is a European initiative for the implementation of information services dealing with environment and security; support for emergency management in the case of natural hazards; forecasting for marine zones, air quality or crop yields and so on (GMES 2004). The GMES capacity is based on four inter-related components: services, observations from space, in-situ, and data integration and information management capacity. The data integration and information management will enable user access and the sharing of information.

In both GEOSS and GMES, it is stated that the areas that are data and computationally intensive require high-performance networks and Grid-based computing for the essential data mining, sharing and analyzing and visualization of the results.

In the following subsection, we briefly describe several projects and initiatives that deal with the application of Grid technology for the EO domain.

2.2.3 Grid projects for EO applications

At present, Grid technologies are widely applied in different domains, in particular the EO domain.

European DataGrid Project (EDG) was the first large European Commission-funded grid project (www.eu-datagrid.org). Many of the results of EDG project have been included in the European project Enabling Grids for E-sciencE (EGEE). EGEE aims to develop a service grid infrastructure, which is available to scientists 24 hours-a-day.

Based on the gained experience, the European Space Agency (ESA) and the European Space Research Institute (ESRIN) have focused on the development of Earth Observation Grid Processing on-Demand infrastructure (G-POD) [Fusco et al, 2003]. Grid is considered as a comfortable "open platform" for handling computing resources, data, tools, etc., and not limited to only high performing computing. G-POD enables access to different data and products from Envisat satellite (http://envisat.esa.int), SEVIRI instrument onboard MSG (Meteosat Second Generation) satellite, etc. One of the most important applications is the analysis long-term data. For example, the analysis of 8 years of GOME on-board temperatures (overall 525 Gb of data) took less than 2 days on 40 computer elements of ESRIN "Grid-on-demand" structure (overall 38460 files were processed). At present, G-POD infrastructure consists of more than 150 working nodes with ability to store and handle of about 100 Tb of data.

DEGREE (Dissemination and Exploitation of GRids in Earth science) project is a European-funded project that aims to build a bridge linking the Earth Science and Grid communities throughout Europe [DEGREE, 2008]. Grid is considered to be the appropriate platform for integration of heterogeneous data resources, processing tools, models, algorithms, etc. The following applied problems are within the scope of DEGREE: earthquake analysis, floods modeling and forecasting influence of climate changes on agriculture, etc.

The Japan Aerospace eXploration Agency (JAXA) and the KEIO University started establishing the Digital Asia system aimed at semi-real time data processing and analyzing. They use Grid environment to accumulate knowledge and know-how to process the remote sensing data. The Digital Asia project is a part of the Sentinel Asia project that is targeting on building natural disasters monitoring system (http://dmss.tksc.jaxa.jp/sentinel).

The Wide Area Grid (WAG) project is initiated by the CEOS Working Group on Information Systems and Services (WGISS), and aims to develop the "horizontal" infrastructure in order to integrate computational, human, intellectual, and informational resources of the space agencies within a large distributed system. Implementation of geospatial-related services and Grid-enable EO data archives are among the priority tasks in this project [Kopp et al, 2007].

The Space Research Institute NASU-NSAU have created a basic computational Grid infrastructure, provided the proof of concept for the solution of complex problems arising in the space weather, hydro-meteorological modeling and flood monitoring [Kussul et al, 2008a]. The Grid infrastructure is developed within several international; projects, namely INTAS-CNES-NSAU project "Data Fusion Grid Infrastructure", STCU-NASU projects "Grid Technologies for Multi-Source Data Integration" and "Grid technologies for environmental monitoring using satellite data".

In this paper we present different approaches to multi-source data integration for the solution of complex applied problems, in particular flood mapping and vegetation state estimation using satellite, modeling and in-situ data. Since these applications are data- and computation-intensive, we use Grid computing technologies. In such a case computational and informational resources are geographically distributed and may belong to different organizations. For this purpose, we also investigate benefits and approaches to the integration of satellite-based monitoring systems.

2.2.4 Scientific approaches

There are some methods for solving of aforementioned problems. Below we examine them in details.

Flood monitoring and prediction: Hydrologic and hydrodynamic models play a major role in assessing and forecasting flood risk. Model's predictions of potential flood extent can help emergency managers to develop contingency plans well in advance of an actual event to help facilitate a more efficient and effective response. One of the main stages of flood prediction is runoff-rainfall simulation. Traditionally for this stage, lumped or semi-distributed hydrological models were used (for instance HSPF model [Singh, 1995]. In general, such models have several parameters that are subject for calibration using input-output time series and/or expert's knowledge. The modern way of hydrological prediction is application of distributed physically based models, e.g. TOPKAPI model [Liu and Todini, 2002]. Such models are better suited for representing heterogeneous hydrological features and for using gridded meteorological data available from modern regional Numerical Weather Prediction models.

These models require several types of data as input, such as rainfall amount/intensity, water extent, land use, soil type and moisture, Digital Elevation Models (DEM), etc. Complex terrain and land use in many regions result in a requirement for high spatial resolution data over very large areas, which can only be practically obtained by remote sensing systems.

Remote sensing data are widely used for flood extent extraction, since it is impractical to acquire the flood area through field observations. Flood extent can be used for hydraulic models to reconstruct what happened during the flood and determine what caused the water to go where it did, for damage assessment and risk management, and can benefit to rescuers during flooding. In order to extract flood extent from satellite imagery we can use data in both optical and microwave range of electromagnetic emission.

The flood extent maps using optical sensors can be extracted using information provided in visible and infrared channels. Different vegetation indices, such as NDVI (Normalized Difference Vegetation Index), could also be used for these purposes. However, the use of optical imagery is limited by severe weather conditions, in particular clouds.

In turn, SAR (synthetic aperture radar) image acquisition is independent of daytime and weather conditions. The use of SAR data for flood extent mapping is motivated by the fact that smooth water surface provides no return to antenna in microwave spectrum and appears black in SAR imagery. Existing methods for flood extent mapping are based on the use of multitemporal technique (http://earth.esa.int/ew/floods/), pixel-processing methods with threshold [Cunjian et al, 2001], [De Chiara et al, 2006]. The authors of the project have developed neural network method for flood extent extraction [Kussul et al, 2007] that is based on image segmentation with sliding window.

Therefore, there exist sophisticated methods for flood extent extraction from satellite imagery. However, in order to provide comprehensive system for flood monitoring and forecasting one need to integrate data different from different sources: modeling, satellite, and in-situ measurements.

Drought monitoring. Both radar and optical data can provide means to detect drought condition, estimate drought extent and assess damage caused by drought events. The temporal resolution of current low resolution optical data as well as wide swath low and medium resolution radar data is enough to monitor vegetation condition and is close to be enough to monitor moisture changes. For instance, MERIS and MODIS data are available once per day for middle latitudes and ASAR WS/GM data are available with temporal resolution up to 2 images per week.

Estimation of drought condition using radar data is possible due to radar's sensitivity to soil/vegetation moisture content. However, the complete decoupling soil and vegetation scattering effects is hard using current C-band single polarization wide swath data. Due to this drought monitoring using radar data can be provided using time series analysis of ASAR WS/GM backscatters [Wagner et al, 2007].

Drought monitor can be done using optical data, e.g. MERIS and MODIS VIS/NIR to create vegetation indices, MODIS TIR to monitor surface temperature. Methods for drought monitoring using NOAA AVHRR data were developed in [Kogan et al, 2004].

Drought monitoring will benefit from merging both optical and microwave remote-sensing data with comprehensive Land Surface Models driven by meteorological data from regional Numerical Weather Prediction models.

Vegetation state assessment: To assess vegetation health/stress the derivation of several biophysical and biochemical variables from optical remote-sensing data was considered by scientific community. Such variables include pigment concentration (e.g. chlorophyll a+b), leaf structure, dry matter content (e.g. lignin, cellulose, protein), water content at leaf level and leaf area index (LAI), leaf angle distribution (LAD), fraction of photosynthetically active radiation absorbed by vegetation (FPAR) at canopy level [Liang, 2004].

Roughly, two main approaches were investigated. The first is empirical or physically based derivation of biophysical parameters from so-called spectral indexes. For instance, relations between LAI and Normalized Difference Vegetation Index (NDVI) and between reflectance in NIR/SWIR domain with vegetation water content were established [Carlson and Ripley, 1998], [Gao, 1996] [Ceccato et al, 2002].

The second approach consists in inversion of physically based leaf, canopy, and atmosphere models. These models were used to estimate structural parameters as LAI and FPAR [Knyazikhin et al, 1998] and biophysical variables such as water content [Zarco-Tejada et al, 2003].

As a conclusion, existing methods for solution of aforementioned applied problems are quite fragmentary and designed to work with some particular sensor data. Significant progress in monitoring of floods, drought, and vegetation's state can be achieved through simultaneous use of data from different sensors, in-situ observations, and modeling approach. Moreover, the same data and core services (e.g. land cover/land use) could be used as inputs for different applications. For instance, regional Numerical Weather Prediction models are valuable for both drought and flood monitoring, as well as assessment of vegetation's state and drought conditions will benefit from using Land Surface Models. Grid technologies will provide the platform for development of such methods and deployment of core and applied services.

2.3 Data assimilation approach

As we can see from overview, there is an urgent need for operational services solving environment-monitoring problems using heterogeneous data. Such approach is implemented within GMES program. Within Ukrainian segment of GEOSS/GMES, we develop new methods for integration of data of different nature (in-situ measurements, modeling, and remote sensing) and Grid technologies for their implementation and data visualization. The particular tasks that are solved are as follows:

1. Development of new method for data integration, in particular remote sensing data from space, in-situ measurements, and modeling data
2. Development of Grid-technologies for heterogeneous data integration
3. Application of developed methods to agricultural and natural disaster monitoring

The overall flowchart of Ukrainian segment of GEOSS/GMES (models, methods, information flows) is depicted in the Figure 26. In this scheme, filled rectangles represent models, methods, and processes. Rounded dotted rectangles represent input data for models and methods or results of previous processing, while rounded rectangles with solid line show end-users.

We consider given blocks in details.

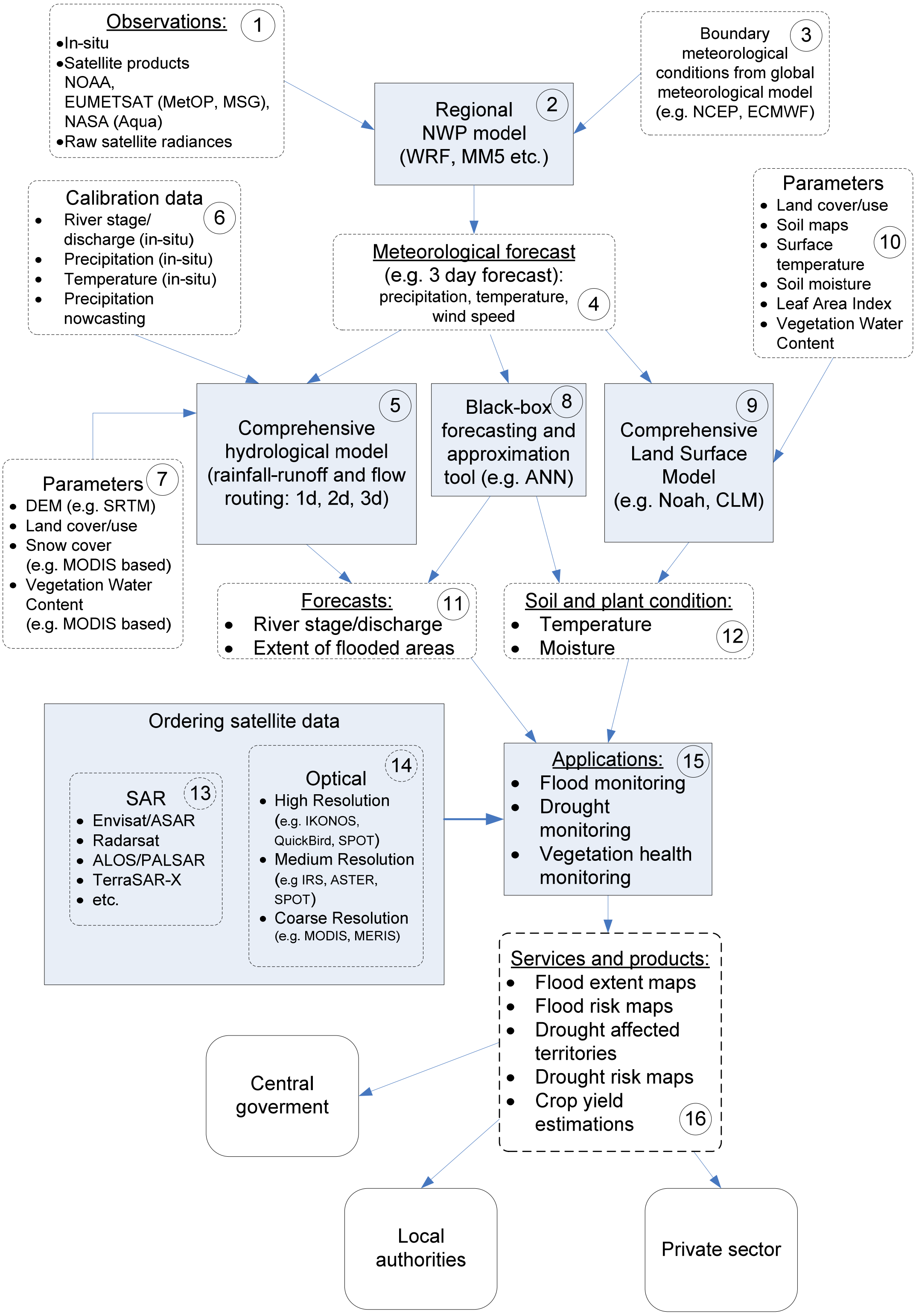


Figure 26. Overall data flowchart at Ukrainian segment of GEOSS/GMES

✓ Meteorological observations

Within Ukrainian segment of GEOSS/GMES, it is planned to create data assimilation system for regional Numerical Weather Prediction model (NWP). In particular, the following satellite data will be assimilated:

— NOAA (microwave & IR instruments: AMSU, MHS, HIRS etc.);
— EUMETSAT (MetOP instruments, MSG geostationary data);
— NASA (Aqua satellite: AMSR-E, AIRS data).

This will increase the accuracy of regional meteorological forecasts and as consequence will increase the quality of solving of several applied problems namely flood monitoring and forecasting, drought monitoring and vegetation health monitoring.

✓ **Regional NWP model**

Because of previous joint STCU-NASU project "GRID technologies for environmental monitoring using satellite data" (2005-2007), we have adapted regional NWP model for the territory of Ukraine. Now SRI runs Weather Research&Forecasting NWP model in operational mode for the territory of Ukraine (see http://dos.ikd.kiev.ua/index.php?option=com_wrf).

✓ **Boundary meteorological conditions from global meteorological model**

To create regional NWP forecasts it is necessary to obtain forecast frames from global meteorological models. This data is used to specify boundary conditions of regional model (vertical profiles of wind, temperature, humidity, pressure etc.). Currently global NCEP forecasts using GFS model are easily available in Internet. In this project, global forecasts will be obtained using NOAA NOMADS system (http://nomad5.ncep.noaa.gov/).

✓ **Meteorological forecasts**

Meteorological forecasts provided by NWP model will be used as inputs for hydrological models, Land Surface Models and additionally they will be used for initiation of retrieving additional datasets in case of possible natural disasters (blocks #13 and #14). For example, using 3 day forecast we could order satellite images at least in 3 days in advance before actual flood event occurs. Without using of forecasts to order satellite data, timely will be much more complex because satellite operators need some time to reprogram satellite and make changes in queue schedule and priorities. Depending on satellite, such procedure can take from few hours to several days. Additionally it should be noted that acceleration of retrieving of satellite data will significantly increases the price of the images.

✓ **Comprehensive hydrological model**

In Ukraine, a detailed hydrological model for river Tisza basins was deployed. Using this model in the framework of the project, we can obtain detailed forecasts for flood dynamics, water levels, and flooded areas. Other basins require adaptation of existing models using additional in-situ observation for calibration.

✓ **Calibration data**

These data are required to estimate parameters in both physically based and empirical black-box hydrological models. Project members have access to such data for Tisza river basin (Carpathian mountain region of Ukraine). For other regions, data from local authorities is required. Additionally we can use satellite based rainfall estimations, for instance, using geostationary data (e.g. MSG or GOES).

✓ **Parameters of hydrological models**

Some of these parameters can be estimated using remote-sensing data. For instance, we will use digital elevation model (DEM) data from SRTM project that have spatial resolution of 90m and available for free while buying more precise DEM as needed. Land cover/land use maps will be obtained for local watershed using medium/coarse resolution imagery. Additionally, some state variables of hydrological models can be estimated from optical/radar data (for instance, MODIS snow cover or LAI (Leaf Area Index) products).

✓ **Black-box forecasting and approximation tool**

Such tools that are based on Artificial Neural Networks (ANN) can be used as additional source of information complementary to comprehensive environmental models. For instance, such ANN-based models are useful for rainfall-runoff simulations. In addition, this tool will be used for fusion of model-based, in-situ and satellite-based retrievals of environmental parameters (temperature, moisture, etc.). For these purposes, modular neural networks can be applied. Modular NN combine different modules, which can be neural networks with various parameters in order to exploit advantages of modules and to improve the global performance.

✓ **Comprehensive Land Surface Model (LSM)**

Such model describes interactions between atmosphere, soil and vegetation including such processes as infiltration, evapotranspiration, soil moisture and heat transport. With appropriate meteorological forcing and soil/land cover data; these models are capable to predict soil temperature and moisture profiles, surface temperature, plant water content, snowpack etc. LSMs are commonly included into meteorological models to provide bottom boundary conditions (for instance, Noah LSM within WRF, MM5 or NCEP Global Forecast System models, Community Land Model (CLM) within Community Climate System Model) but can be run in so-called off-line mode (decoupled from meteorological model). In the latter case, LSM can be run with high spatial resolution (up to 1 km). SRI has experience to operate Noah LSM in coupled mode within WRF modeling system. In the framework of the project, we will develop method for assimilation of satellite data into such models. The results of assimilation will be used for creation of several products (block #12) that will become the basis for end-user services (block #16), in particular for drought indicators and vegetation stress estimations.

✓ **Parameters of LSM**

As in the case of hydrological models used for flood prediction, several LSM's parameters can be estimated using remote-sensing data. Within the proposed project, we will use land cover/land use maps, leaf area index, surface temperature, soil moisture. These data will be assimilated into LSM using intelligent techniques (in particular by ANN) and evolutionary computations (e.g. genetic algorithms).

✓ **Hydrological forecasts**

This block provides results of hydrological model. Ideally, we would like to have forecasts of flooded areas. Worse case if we have only river stage/discharge data. In the case of flood, we can issue alert message to the local authorities and order remote sensing data in advance to estimate flooded areas during the flood (blocks #13 and #14).

✓ **Soil and plant condition**

Soil and plant condition are obtained as a merge of remote-sensing retrievals, in-situ data and results of modeling of land surface. Such data will be used to produce dedicated products in the field of drought and plant condition monitoring (block #16).

✓ **SAR (synthetic aperture radar) data from space-borne instruments**

SAR imagery are most valuable satellite data for estimation of flooded areas due to all-weather SAR functioning. At first stage, we propose to use Envisat/ASAR (ESA) and ALOS/PALSAR (JAXA) data. These sensors are included into International Charter "Space and Major Disasters" (http://www.disasterscharter.org/main_e.html). In addition, these data can be obtained via ESA Cat-1 projects "Wide Area Grid Testbed for Flood Monitoring using Spaceborne SAR and Optical Data" (#4181) in which SRI takes part.

✓ **Optical data**

Taking into account possible cloud cover problems we can use optical imagery for flood extent estimation as well as to assess state of vegetation. Within this project, coarse resolution sensors such as MODIS from Terra satellite or MERIS from Envisat satellite will be used. Medium/fine resolution imagery can be ordered to produce local/regional products and services.

✓ **Applications**

Within Ukrainian segment of GEOSS/GMES, we focus on the following applications:

– flood monitoring;
– drought monitoring;
– vegetation state monitoring.

✓ **Services and products**

The following products and services are provided for the central government, local authorities and private sector:

– flood extent maps;
– flood risk maps;
– areas affected by drought;
– drought risk maps;
– crop yield estimation for agricultural regions.

2.4 Applications

In this section, we describe in details EO applications that were deployed in the Grid infrastructure. In particular, we focus on the weather modeling application, flood monitoring, and vegetation state estimation. The motivation for the selection of these applications comes from the following:

(i) numerical weather prediction belongs to computational intensive applications;

(ii) flood applications need the fast response to the emergencies, and thus require a reliable infrastructure for data management and processing;

(iii) vegetation state estimation belongs to data intensive application where different data and products are analyzed in order to produce the final product and requires intelligent data assimilation techniques.

Prediction of meteorological parameters represents one of the core services for a number of applications (e.g. floods, droughts, agriculture, etc). Currently, we run the Weather Research and Forecasting model (WRF) (Michalakes et al. 2004) in operational mode for the territory of Ukraine. The meteorological forecasts are generated every 6 hours with a spatial resolution of 10 km. Forecast range is 72 hours. The horizontal grid dimensions are 200x200 points with 31 vertical levels. We use NCEP GFS (Global Forecasting System) forecasts as boundary conditions. This data is available via Internet though the NOMADS system (National Operational Model Archive & Distribution System).

The workflow of the model run is composed of the following steps (Figure 27):

(i) data acquisition;

(ii) data pre-processing, computation of forecasts using WRF model and data post-processing;

(iii) visualization of the predicted parameters.

Data acquisition: To run WRF model, it is necessary to obtain boundary and initial conditions for territory of Ukraine. This data can be extracted from GFS model forecasts. To get the required data, the dedicated script was developed. This script downloads global forecasts every 6 hours. To decrease the data volume, our script uses special Web-service capable of selecting subsets of the GFS data for the territory of Ukraine. The acquired data is transferred to the storage subsystem and marked as unprocessed (i.e. it has to be processed by the WRF model). After the GFS data has been downloaded, the Karajan script initializes a workflow for data pre-processing, WRF run, and data post-processing.

Data pre-processing step is intended to transform the downloaded data into the format that is used to run the WRF model. GFS data is delivered in the GRIB format in the geographical projection. This data is transformed into the internal WRF format by the grib_prep.exe command, warped into the Lambert Conformal Conic projection (by executing hinterp.exe command) and vertically interpolated using the vinderp.exe command. (grib_prep.exe, hinterp.exe, and vinterp.exe commands are tools from WRF Standard Initialization (SI) package.) The results of these transformations are stored in the netCDF format. After that, the real.exe command is used to produce initial and boundary conditions for WRF model run. The inputs to real.exe command are GFS data in netCDF format and WRF configuration file (namelist.input).

Data processing step consists in performing WRF run using wrf.exe command. The output of the command is forecasts of the meteorological parameters. This is the most computationally intensive task.

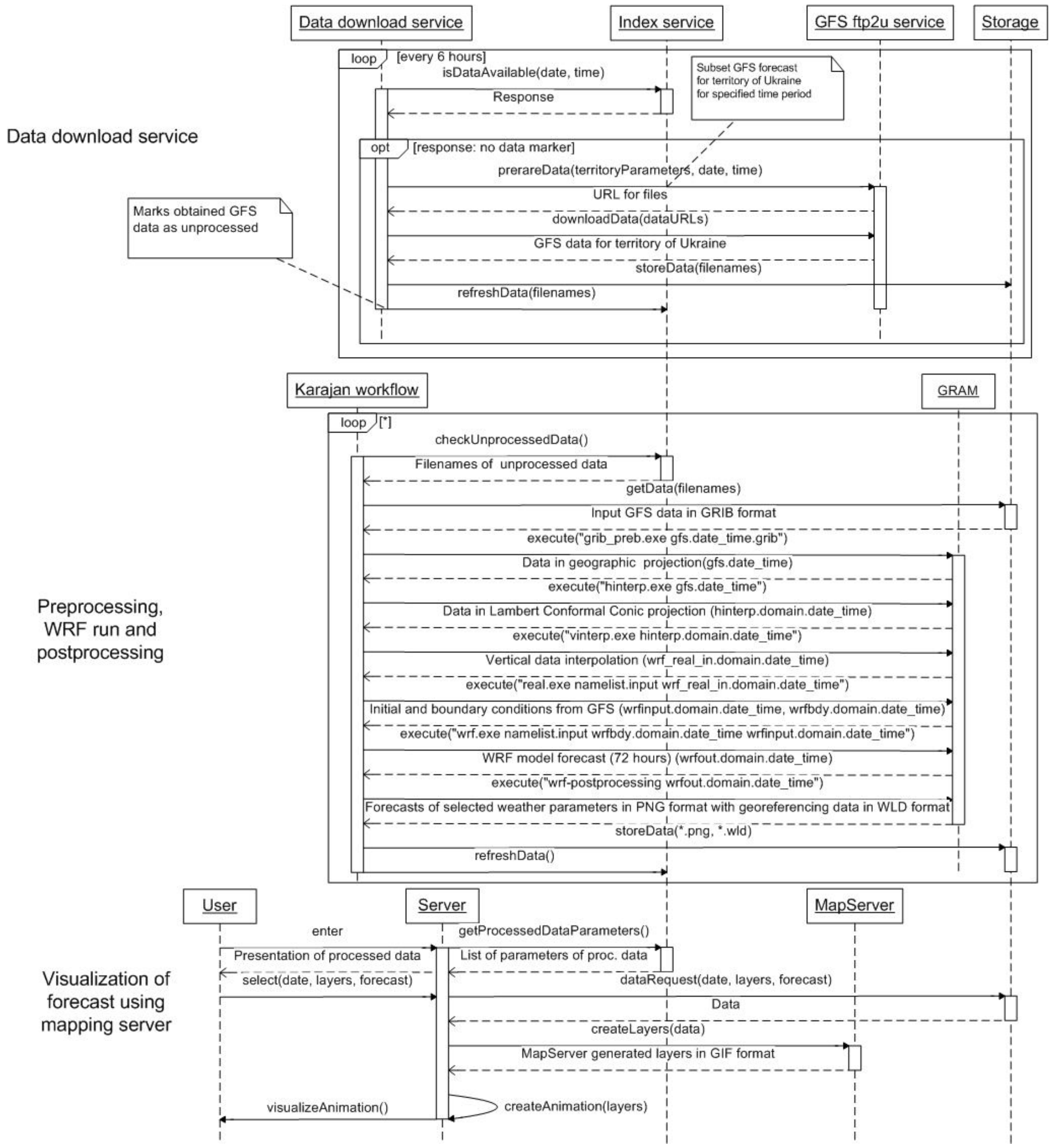


Figure 27. UML sequence diagram [Larman, 2004] for the NWP application

After WRF model run, **post-processing step** is carried out. For specified weather parameters and for each forecast frame (3 hours), a graphic representation (in PNG format) of spatial distribution is created. Additionally, special files containing georeferencing information are created (files with *.wld extension). The results of the post-processing phase are used to visualize the WRF forecasts via the mapping service. This service is available via http://dos.ikd.kiev.ua, and provides to the users animations of the weather forecasts (Figure 28).

The service provides tools to select a forecast time, forecast frames (up to 72 hours ahead), and weather parameters to display. Selected by the user information is packed into the request to the server. To process the request, all required data (in PNG and WLD formats) is retrieved from epy storage subsystem and passed to epy mapping server in order to create the maps. Maps are further processed by the script to generate weather animation in GIF format. Finally, this animation is presented at user side.

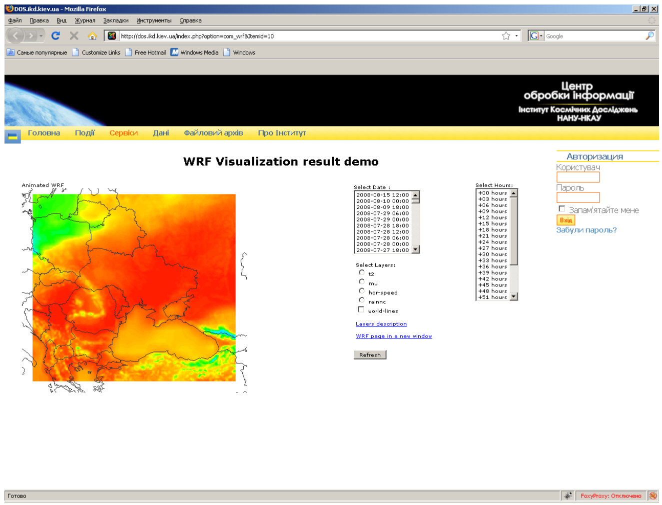


Figure 28. The example of land temperature forecasts using WRF model

We have also tested the performance of the WRF model in dependence of the number of computational nodes. For test purposes, we used the WRF model version 2.2 with a model domain identical to those used in operational NWP service (200x200x31 grid points with horizontal spatial resolution 10 km). We observed almost linear productivity growth within increasing number of computation nodes. For instance, 8 nodes of the SCIT-3 cluster of the Grid infrastructure gave the performance increase in 7.09 times (of 8.0 theoretically possible) when compared to the single node. The use of 64 nodes increases the performance in 43.6 times (see Figure 29).

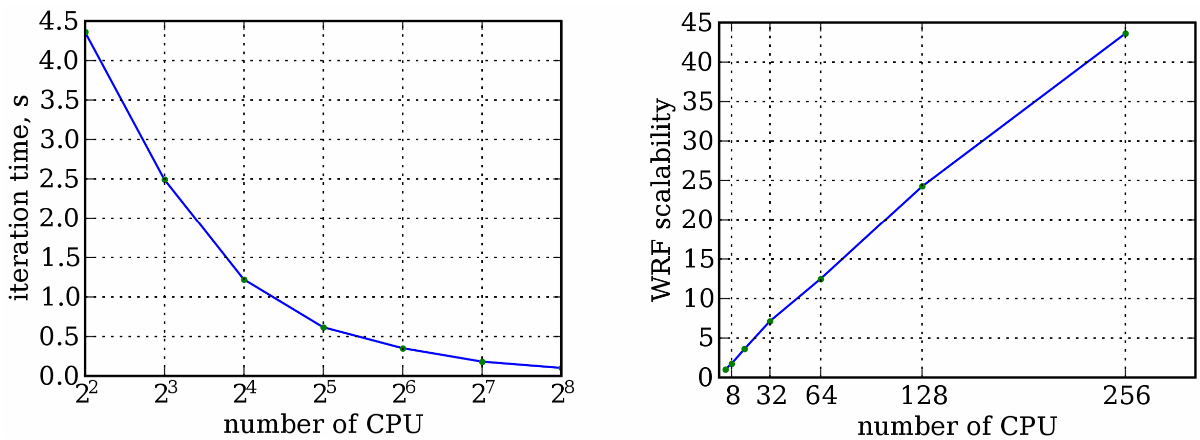


Figure 29. The results of WRF performance on the SCIT-3 cluster:
computation time for 1 iteration (left);
acceleration of the WRF model with respect to a number of nodes (right)

2.4.1 Flood prediction and mapping from satellite imagery

In recent decades, the number of hydrological natural disasters has considerably increased. According to [Scheuren et al, 2008], we have witnessed in recent years a strengthening of the upward trend, with an average annual growth rate of 8.4% in the 2000 to 2007 period. Hydrological disasters, such as floods, wet mass movements, represent 55% of the overall disasters reported in 2007, having a tremendously high human impact (177 million victims) and causing high economic damages (24.5 billion USD) [Scheuren et al, 2008].

EO data from space can provide valuable and timely information when one has to respond to and mitigate such emergencies as floods. From satellite imagery, we can determine flood areas, since it is impractical to provide such information through field observations. The use of optical imagery (in visible and infrared range) for flood mapping is limited by severe weather conditions, in particular by the presence of clouds. In turn, synthetic aperture radar (SAR) measurements from space are independent of daytime and weather conditions and can provide valuable information to monitoring of flood events. This is mainly due to the fact that smooth water surface provides no return to antenna in microwave spectrum and appears black in SAR imagery [Rees, 2001].

Flood mapping procedure from SAR imagery represents a complex workflow and consists of the following steps. The first step consists in re-constructing a satellite imagery taking into account the calibration, the terrain distortion using digital elevation model (DEM) and providing exact geographical coordinates. The second step is image segmentation, and the third step consists in the classification to determine the flood extent.

In this subsection we describe a neural network approach to flood mapping from satellite SAR imagery that is based on the application of self-organizing Kohonen's maps (SOMs) [Kohonen, 1995], [Haykin, 1999]. The advantage of using SOMs is that they provide effective software tool for the visualization of high-dimensional data, automatically discover of statistically salient features of pattern vectors in data set, and can find clusters in training data pattern space, which can be used to classify new patterns [Kohonen, 1995]. We applied our approach to the processing of data acquired from different satellite SAR instruments (ERS-2/SAR, ENVISAT/ASAR, RADARSAT-1 and RADARSAT-2) for different flood events: river Tisza, Ukraine and Hungary (2001); river Huaihe, China (2007); river Mekong, Thailand and Laos (2008); river Koshi, India and Nepal (2008); river Norman, Australia (2009); and river Zambezi, Mozambique (2008) and Zambia (2009).

To this end, different methods and approaches were proposed to flood mapping using satellite imagery:

- multi-temporal technique (http://earth.esa.int/ew/floods);
- threshold segmentation [Cunjian et al, 2001];
- statistical active contour model [Horritt, 1999];
- edge-detection techniques [Niedermeier et al, 2000];
- analysis of time-series of SAR images [Martinez and Le Toan, 2007].

The following shortcomings of the existing approaches can be identified: manual threshold selection and parameters identification; statistical models require a priori knowledge of image statistical properties; application of complex models for noise (speckle) reduction; no spatial neighborhood between pixel is considered. A more detailed description of the existing techniques is given in [Kussul et al, 2008a].

Data set description. We applied our approach to the processing of remote-sensing data acquired from different satellite SAR instruments for different flood events:

- ERS-2/SAR: flood on Tisza river (Ukraine), 2001;
- ENVISAT/ASAR Wide Swath Mode (WSM): river Huaihe, China, 2007; river Zambezi, Mozambique, 2008; river Mekong, Thailand and Laos, 2008; river Koshi, India and Nepal, 2008; Ha Noi City, Vietnam, 2008; river Zambezi, Zambia, 2009;
- RADARSAT-1: river Huaihe, China, 2007;
- RADARSAT-2: river Norman, Queensland, Australia, 2009 (see Figure 30).

Data from European satellites (ERS-2 and ENVISAT) were provided from the ESA Category-1 project "Wide Area Grid Testbed for Flood Monitoring using Spaceborne SAR and Optical Data" (№4181). Data from RADARSAT-1 satellite were provided from the Center of Earth Observation and Digital Earth (China). RADARSAT-2 data were provided by the Canadian Space Agency (CSA) within the GEOSS Architecture Implementation Pilot Phase 2. (AIP-2, www.ogcnetwork.net).

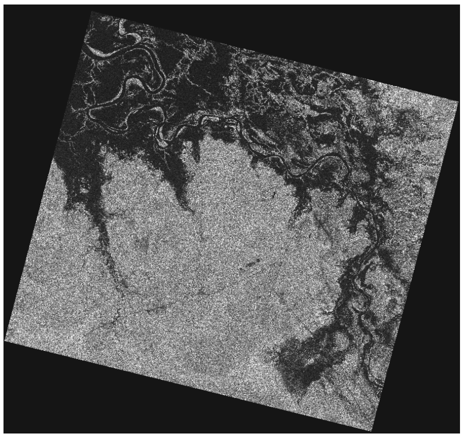


Figure 30. SAR image acquired from RADARSAT-2 satellite during the flood on the river Norman, Australia(14.02.2009)

A pixel size and ground resolution of ERS-2 imagery (in ENVISAT format, SLC – Single Look Complex) were 4 m and 8 m, respectively; for ENVISAT imagery – 75 m and 150 m; and for RADARSAT-1 imagery – 12.5 m and 25 m; for RADARSAT-2 imagery – 3 m both. We used auxiliary data to derive information on water bodies (Landsat-7/ETM+, European Corine Land Cover CLC 2000) and topography (SRTM DEM v.3).

Neural network is built for each SAR instrument separately. In order to train and test neural networks, we manually selected the ground-truth pixels with the use of auxiliary data sets that correspond to both territories with the presence of water (we denote them as belonging to a class "Water") and without water (class "No water"). For ENVISAT/ASAR instrument, data from Chinese flood event were used to construct and calibrate the neural network. This neural network, then, was used to produce flood maps for other flood events. Collected ground-truth data were randomly divided into the training set (which constituted 75% of total amount) and the testing set (25%). Data from the training set were used to train the neural networks, and data from the testing set were used to verify the generalization ability of the neural networks, i.e. the ability to operate on independent, previously unseen data sets [Haykin, 1999].

Methodology description: Our flood mapping workflow with input and output data is shown in Figure 31 [Kussul et al, 2008a].

SOM is a type of artificial neural network that is trained using unsupervised learning to produce a low dimensional (typically two-dimensional), discretized representation of the input space of the training samples, called a map [Kohonen, 1995], [Haykin, 1999]. The map seeks to preserve the topological properties of the input space. SOM is formed of the neurons located on a regular, usually 1- or 2-dimensional grid. Neurons compete with each other in order to pass to the excited state. The output of the map is a, so-called, neuron-winner or best-matching unit (BMU) whose weight vector has the greatest similarity with the input sample **x**.

The network is trained in the following way: weight vectors $\mathbf{w}_j$ from the topological neighborhood of BMU vector i are updated according to [Kohonen, 1995], [Haykin, 1999]

$$i(\mathbf{x}) = \arg\min_{j=\overline{1,L}} \left\| \mathbf{x} - \mathbf{w}_j \right\|,$$

$$\mathbf{w}_j(n+1) = \mathbf{w}_j(n) + \eta(n) h_{j,i(\mathbf{x})}(n)(\mathbf{x} - \mathbf{w}_j(n)),\ j = \overline{1,L} \qquad (1)$$

where η is learning rate (see Eq. 3), $h_{j,i(\mathbf{x})}(n)$ is a neighborhood kernel around the winner unit i, **x** is an input vector, $\|\bullet\|$ means Euclidean metric, L is a number of neurons in the output grid, n denotes a number of iteration in the learning phase.

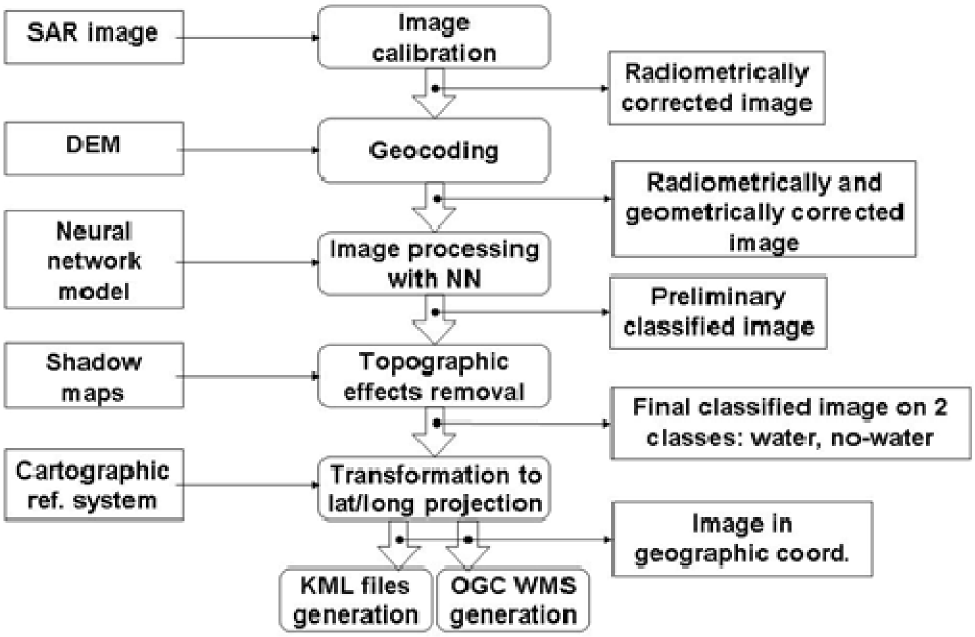


Figure 31. Flood mapping from SAR satellite imagery: workflow

The neighborhood kernel function $h_{j,i(\mathbf{x})}(n)$ is taken to be the Gaussian

$$h_{j,i(\mathbf{x})}(n) = \exp\left(-\frac{\|r_j - r_{i(\mathbf{x})}\|}{2\sigma^2(n)}\right) \tag{2}$$

where $r_j, r_{i(\mathbf{x})}$ are the vector locations in the display grid of the SOM, $\sigma(n)$ corresponds to the width of the neighborhood function, which is decreasing monotonically with the regression steps.

For learning rate, we used the following expression:

$$\eta(n) = \eta_0 \cdot e^{-\frac{n}{\tau}}, \eta_0 = 0.1 \tag{3}$$

where τ is a constant. The initial value of 0.1 for learning rate was found experimentally.

Kohonen's maps are widely applied to the image processing, in particular image segmentation and classification [Kohonen, 1995], [Haykin, 1999]. Prior neural network training, we need to select image features that will be give to the input of neural network. For this purpose, one can choose original pixel values, various filters, Fourier transformation etc. In our approach, we used a moving window with backscatter coefficient values for ERS-2 and ENVISAT images and digital numbers (DNs) for RADARSAT-1/2 image as inputs to neural network. The output of neural network, i.e. neuron-winner, corresponds to the central pixel of moving window. In order to choose appropriate size of the moving window for each satellite sensor, we ran experiments for the following windows size: 3-by-3, 5-by-5, 7-by-7, 9-by-9, and 11-by-11.

We, first, used SOM to segment each SAR image where each pixel of the output image was assigned a number of the neuron in the map. Then, we used pixels from the training set to assign each neuron one of two classes ("Water" or "No water") using the following rule. For each neuron, we calculated a number of pixels from the training set that activated this neuron. If maximum number of these pixels belonged to class "Water", then this neuron was assigned "Water" class. If maximum number of these pixels belonged to class "No water", then this neuron was assigned "No water" class. If neuron was activated by neither of the training pixels, then it was assigned "No data" class.

Results of image processing: In order to choose the best neural network architecture, we ran experiments for each image varying the following parameters: (i) size of the moving window for images that define the number of neurons in the input layer of the neural network; (ii) number of neurons in the output layer, i.e. the sizes of 2-dimensional output grid. Other parameters that were used during the image processing are as follows:

- neighborhood topology is hexagonal;
- neighborhood kernel around the winner unit is the Gaussian function (see Eq. 2);
- initial learning rate is set to 0.1;
- number of the training epochs is equal to 20.

The initial values for the weight vectors are selected as a regular array of vector values that lie on the subspace spanned by the eigenvectors corresponding to the two largest principal components of the input data [Kohonen, 1995].

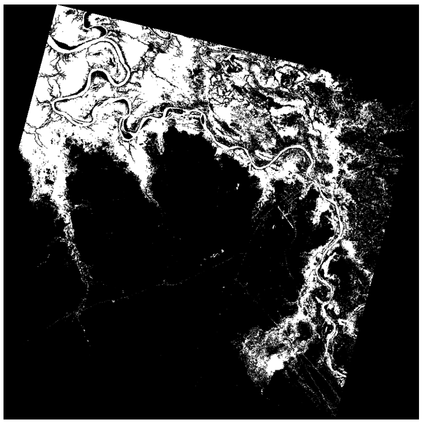


Figure 32. The resulting flood extent shown with white color for the river Norman, Australia (RADARSAT)

We applied our approach to determine flood areas from SAR images acquired by the following instruments: ERS-2/SAR, ENVISAT/ASAR, and RADARSAT-1. Classification rates for these sensors using independent testing data sets were 85.40%, 98.52% and 95.99%, respectively.

For the images with higher spatial resolution (i.e. ERS-2 and RADARSAT-1), the best results were achieved for larger moving window 7-by-7. In turn, for the ENVISAT/ASAR WSM image, we used the moving window of smaller size 3-by-3. The use of higher dimension of input window for the ENVISAT image led to the coarser resolution of the resulting flood extent image and reduced classification rate.

The example of resulting flood extent map derived from RADARSAT-2 data acquired for the river Norman, Australia (see Figure 30) is shown in Figure 32.

Implementation: We developed a parallel version of our method and deployed it at the Grid infrastructure. Parallelization of the image processing is performed in the following way: SAR image is split into the uniform parts that are processed on different nodes using the OpenMP Application Program Interface (www.openmp.org). The use of the Grids allowed us to reduce considerably the time required for image processing. In particular, it took approximately 30 min to process a single SAR image on a single workstation. The use of Grid computing resources allowed us to reduce the time to less than 1 min.

2.4.2 Vegetation State Estimation

Estimation of vegetation state from satellite data has proved to be very helpful for agriculture monitoring, climate modeling, natural disasters management [Liang, 2008]. Parameters that can be estimated using optical data include Leaf Area Index (LAI), Fraction of Photosynthetic Active Radiation (FPAR), leaf pigment concentration, water concentration. Here, we will focus on plant moisture estimation from satellite data. This is very important for drought monitoring that becomes one of the major disasters in agricultural countries like Ukraine. For example, drought in Ukraine in 2007 resulted in $100 millions losses.

Water shortage in plants and plant stress in general can be detected by optical satellite data. Vegetation moisture determination is possible mainly due to significant differences in reflectance in Shortwave Infrared band of electromagnetic spectrum (SWIR) of vegetation under water stress and under normal conditions. However, in solar optical domain vegetation reflectance is controlled not only by moisture but also by several other factors: leaf structure, pigment concentration, LAI, soil reflectance [Liang, 2004]. Due to this plant moisture, estimation is far from trivial.

This estimation task is a massive parallel problem since estimation has to be performed on the per pixel basis. In addition, even if the problem is not computationally complex for a single pixel, it has to be solved for each pixel of the satellite imagery. For current moderate resolution sensors such as MODIS 1 million pixels has to be processed per day, and new satellite systems such as RapidEye will deliver billions pixels per day. Nevertheless, this problem is highly parallelizable and, thus, is a good candidate to be executed in a Grid environment.

Earlier approaches to vegetation moisture estimation were based on so-called Vegetation Indexes [Ceccato et al, 2002], [Gao, 1996]. Index is a simple combination of reflectance in different bands of satellite image, which has increased sensitivity to target variable like moisture content and low sensitivity to other factors. For example, one of the popular indexes is a Normalized Difference Water Index (NDWI):

$$NDWI = \frac{\rho_{0,8} - \rho_{1,6}}{\rho_{0,8} + \rho_{1,6}} \tag{4}$$

where $\rho_{0,8}$ and $\rho_{1,6}$ are reflectance value in Near Infrared band (NIR) and SWIR band.

Vegetation Indexes uses only a limited number of spectral bands (2-3) while modern sensors like MODIS, MERIS have 7-15 bands. In addition, indexes remain only indirect measures of target variables, and additional regressions have to be used to estimate it. Usually, such regressions require additional calibration using local data, which further complicates utilization of Vegetation Indexes. That is why, at present, the modern way to estimate vegetation parameters is based on more sophisticated approach – physical modeling of satellite signal using canopy radiative transfer models [Liang, 2004].

Problem statement: Under modeling approach, the estimation problem is considered as inverse to the problem of simulation of satellite signal. For the latter task the wide range of models exists [Liang, 2004], among which several models (like PROSPECT [Feret et al, 2008] and SAIL [Verhoef et al, 2007]) are widely used in remote sensing. For our purpose we will formulate radiative transfer model as a mapping $h:\mathbf{R}^{n_x} \to \mathbf{R}^{n_d}$ that maps state of vegetation $x \in X \subset \mathbf{R}^{n_x}$ into reflectance in different bands $h(x) \in D \subset \mathbf{R}^{n_d}$:

$$d = h(x) + h(x)\varepsilon \tag{5}$$

where d is measurement vector and ε is noise vector. This problem is characterized by multiplicative noise [Bacour et al, 2006].

For instance, for PROSPECT leaf radiative transfer model the dimension of x is four $x = (N, C_{ab}, C_w, C_m)^T$, where N — leaf structure parameter, while C_{ab}, C_w, C_m — concentration of chlorophyll, water and dry matter. Dimension of model output vector $h(x)$ is 2100, however for remote sensing purposes model output has to be aggregated to be comparable with current multispectral sensors. So usually the dimension of observation vector d is much smaller, for instance for MODIS sensor it will be 7.

In this chapter, the Bayesian approach to inverse problems is considered [Tarantola, 2005]. Within this approach, uncertainty in a priory estimate of state vector x and in process of measurement of reflectance vector $h(x)$ has probabilistic nature. Let $\mathbf{x}$, $\mathbf{d}$, $\boldsymbol{\varepsilon}$ — random vectors of a priory estimate of model input, observations and noise in observations, $p(x)$, $p(d)$ and $p(\varepsilon)$ — densities of probability distributions of these vectors. It is assumed that random vectors $\mathbf{x}$ and $\boldsymbol{\varepsilon}$ are independent, while densities $p(x)$, $p(\varepsilon)$ and function h is such, that random vectors $\mathbf{x}$ and $\mathbf{d}$ have common density $p(x,d)$ and components of these vectors have variance.

The solution of inverse problem is conditional density of model input $\mathbf{x}$ with respect of known value of observations vector $\mathbf{d}$ [Tarantola, 2005]:

$$p(x \mid d) \propto p(d \mid x) p(x), \quad x \in \mathbf{R}^{n_x}, d \in \mathbf{R}^{n_d} \tag{6}$$

However, for practical purposes we have to estimate some properties of above conditional density, like mean, standard deviation, median, most probable value etc.

Neural network method to solve inverse problem: There are several methods to estimate properties of (6): Monte-Carlo [Qingyuan et al, 2005], variational [Bacour et al, 2002], lookup tables [Combal et al, 2002] and neural networks [Bacour et al, 2006]. However, in recent years neural networks gain a lot of attention due to their ability to approximate arbitrary continuous function and computational efficiency [Haykin, 1999].

To solve inverse problem (6) within traditional neural network approach the approximation $f: D \to X$ of inverse mapping to $h: X \to D$ is constructed using neural network, for instance Multilayer Perceptron (MLP). This is performed through minimization of quadratic functional:

$$J(w) = \frac{1}{2} \sum_i \left\| x_i - f(d_i, w) \right\|^2 \tag{7}$$

where function $f(\cdot, w)$ is defined by neural network with weight coefficients w, $\{(d_i, x_i), i = \overline{1,n}\}$ is learning sample set created via sampling from density $p(x,d)$.

It can be shown (see for instance [Bishop, 1996], [Kravchenko, 2009]) that given sufficient number of learning samples neural network with quadratic error criteria will approximate conditional mean $\mathrm{E}[\mathbf{x} \mid \mathbf{d} = d] = \int x p(x \mid d) dd$ of network output x given input d. Therefore, in the framework traditional neural network approach we can obtain only point estimate of parameters. To overcome this deficiency of traditional neural networks for inverse problem solving we propose to apply neural networks with non-quadratic error criteria, such as Mixture Density Networks (MDN) [Bishop, 1996]. Such networks allow modeling of conditional density $p(x \mid d)$ as a mixture of Gaussian densities.

$$p_{MDN}(x \mid d,w) = \sum_{l=1}^{L} \alpha_l(d,w) \cdot \phi\left(x; m_l(d,w), \sigma_l(d,w)\right) \tag{8}$$

where $\phi(x;m,\sigma) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^{n_x}} \exp\left(-\frac{\|x-m\|^2}{2\sigma^2}\right)$ — Gaussian density with mean m and diagonal covariance matrix $\sigma^2 \mathrm{I}$, α_l — mixture coefficients ($\sum_l \alpha_l = 1$), L — number of elements of mixture. Functions $\alpha_l(d,w)$, $m_l(d,w)$ and $\sigma_l(d,w)$ are constructed using MLP with modified output layer. MDN is learned through minimizing the following error criteria:

$$J(w) = \frac{1}{n} \sum_{i=1}^{n} -\ln p_{MDN}(x_i \mid d_i, w) \tag{9}$$

Unlike MLP, MDN with even one Gaussian component in mixture can approximate both conditional mean and variance of $p(x \mid d)$ [Kravchenko, 2009].

Numerical experiment with PROSPECT model: Here we will demonstrate use of MDN to solve inverse problem of leaf moisture estimation. To formulate forward problem we will use PROSPECT leaf radiative transfer model. In this case x vector consists of 4 parameters: $x = \left(N, C_{ab}, C_w, C_m\right)^T$, while observation vector d consists of seven leaf reflectances in MODIS-like spectral bands. To pose inverse problem we will assume uniform a priory density $p(x)$ and independent Gaussian noise model for $\boldsymbol{\varepsilon}$ (5% standard deviation). To estimate plant moisture we will use MDN with 7 neurons in input layer, 5 neurons in hidden layer and one-dimensional mixture containing one Gaussian component. This network is used to estimate mean and variance of conditional density $p(C_w \mid d)$. Increasing number of mixture's components or number of neurons in hidden layer does not improve the quality of solution in this problem.

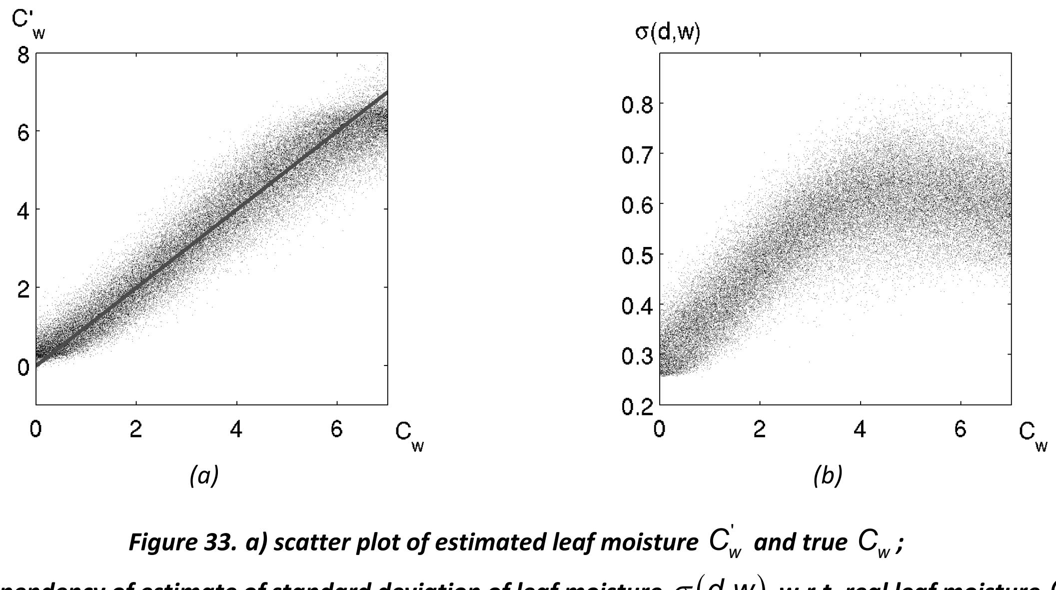


Figure 33. a) scatter plot of estimated leaf moisture $C_w^{'}$ and true C_w;
b) dependency of estimate of standard deviation of leaf moisture $\sigma(d,w)$ w.r.t. real leaf moisture C_w

Scatter plot of conditional mean of leaf moisture $C_w^{'} = m_1(d_i, w)$ estimated by MDN given observation d_i and true value C_w is shown on Figure 33a (identical dependency is shown by strait line), while dependency of estimate of standard deviation of leaf moisture $\sigma(d_i, w)$ given observation d_i with respect to true value C_w is shown in Figure 33b. Standard deviation is increased with increase of moisture C_w and stabilized for large C_w (4-7 cg/cm^2). This is in accordance with the fact that sensitivity of SWIR reflectance is decreased for large leaf moisture values.

Validation results: To validate our algorithm we used LOPEX leaf optical properties database (Leaf Optical Properties EXperiment). This database contains over 1250 plant reflectance spectra. For validation purpose, 330 fresh leaf spectra of 66 plant species at different moisture level were used. Spectra were aggregated using MODIS band relative spectral response functions. Figure 34a shows the scatter plot of estimated leaf moisture ($C_w^{'}$) and observed (C_w), while Figure 34b shows the histogram of moisture estimation error normalized by estimate of standard deviation $\delta = \left(C_w^{'} - C_w\right) / \sigma(d_i, w)$. Most of the departures (90%) are located in $[-2;2]$ interval (in $\pm 2\sigma$ interval) that confirms adequacy of standard deviation estimates using MDN.

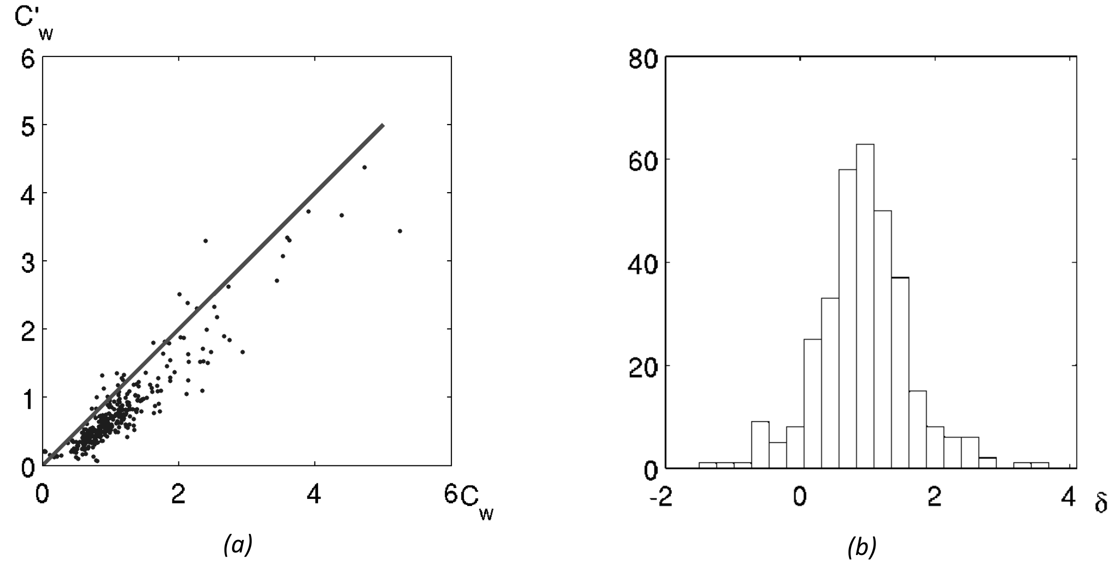


Figure 34. a) scatter plot of estimated leaf moisture $C_w^{'}$ and true C_w;
b) histogram of normalized errors δ

2.5 Levels of integration: main problems and possible solutions

Modern tendencies of globalization and development of the "system of systems" GEOSS lead to the need of integration of heterogeneous satellite-based monitoring systems. Integration can be done at different levels: (i) data exchange level, (ii) task management level. Data exchange is supposed to provide infrastructure for sharing data and products. This infrastructure enables data integration where different entities provide various kinds of data to support joint solution of complex problems (Figure 35). Task management level envisages running applications at distributed computational resources provided by different entities (Figure 36). Since many of the existing satellite monitoring system rely on Grid technologies appropriate approaches and technologies should be evaluated and developed to enable Grid system integration (so called InterGrid).

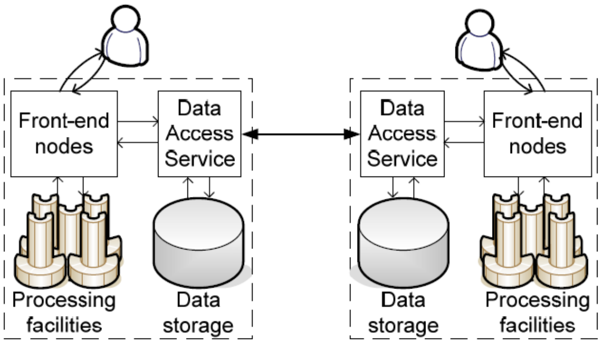


Figure 35. Data integration level

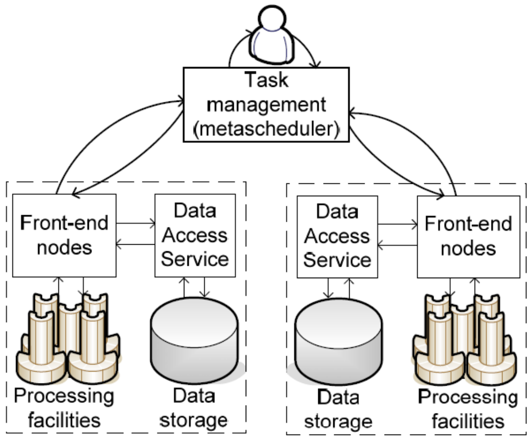


Figure 36. Task management level

This section highlights main challenges and possible solutions for satellite monitoring systems integration at both levels, and provides the case studies for both cases.

Integration at data exchange level could be done by using common standards for EO data exchange, common user interfaces, and common data and metadata catalogues. Considering the task management level, the following problems additionally should be tackled: the use of joint computational infrastructure; development of jobs submission and scheduling algorithms; load monitoring enabling; security policy enforcement.

✓ **Data exchange level**

At present the most appropriate standards for data integration is Open Geospatial Community (OGC) standards. Data visualization issues can be solved by using the following set of standards: WMS (Web Map Service), SLD (Style Layer Descriptors), and WMC (Web Map Context). OGC's WFS (Web Feature Service) and WCS (Web Coverage Service) standards provide uniform ways for data delivery. In order to provide interoperability at the level of catalogues CSW (Catalogue for Web) standard can be applied.

Since data are stored at geographically, distributed sites there can be issues regarding optimization of visualization schemes. In general, there are two possible ways for distributed data visualization: centralized visualization scheme and distributed visualization scheme. Advantages and faults of each scheme were described in [Shelestov et al, 2008].

This approach is implemented in the International vegetation state estimation system, developed jointly by Space Research Institute NASU-NSAU, Space Research System of Russian Academy of Science and Institute of Informatics of Slovak Academy of Science.

✓ **Task management level**

In this subsection, we present main issues and possible solutions for Grid-system integration. Main prerequisite of such kind of integration is certificates trust. It could be done, for example, through EGEE infrastructure that nowadays brings together the resources of more than 70 countries. Another problems concerned with different Grid systems integration are as follows: enabling data transfers and high-level access to geospatial data; development of common catalogues; enabling jobs submission and monitoring; enabling information exchange.

Data transfer: GridFTP is an appropriate and reliable solution for data transfer. The only limitation is the requirement of transparent LAN (local area network) infrastructure.

Access to geospatial data: High-level access to geospatial data can be organized in two possible ways: using pure WSRF services or using OGSA-DAI container. Each of this approach has its own advantages and weaknesses. Basic functionality for WSRF-based services can be easily implemented (with proper tools), packed, and deployed. Nevertheless, advanced functionality such as security delegation, third-party transfers, indexing should be implemented by hands. WSRF-based services can also pose some difficulties if we need to integrate them with other data-oriented software.

OGSA-DAI framework provides uniform interfaces to heterogeneous data. This framework makes possible to create high-level interfaces to data abstracting hiding details of data formats and representation schemas. Most of problems in OGSA-DAI are handled automatically, e.g. delegation, reliable transfer, data flow between different sources and sinks. OGSA-DAI containers are easily extendable and embeddable. Nevertheless, comparing to WSRF basic functionality implementation of OGSA-DAI extensions is more difficult. Moreover, OGSA-DAI require preliminary deployment of additional software components.

Task management: There are two possible approaches for task management. One of them is to use Grid portal (Figure 37) supporting different middleware platforms, such as GT4, gLite, etc. Grid portal is an integrated platform to end-users that enables access to Grid services and resources via standard Web browser. Grid portal solution is easy to deploy and maintain, but it does not provide application interface and scheduling capabilities.

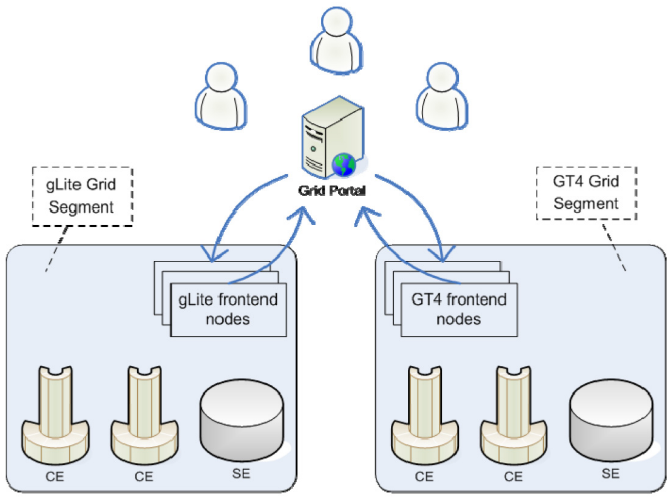


Figure 37. Portal approach to grid system integration

Another approach is to develop high-level Grid scheduler (Figure 38) that will support different middleware by providing some standard interfaces. Such metascheduler interacts with low-level schedulers (used in different Grid systems) enabling in such way system interoperability. Metascheduler approach is much more difficult to maintain comparing to portals; however, it provides API with advanced scheduling and load-balancing capabilities. At present, the most comprehensive implementation for the metascheduler is a GridWay system. The GridWay metascheduler is compatibility with both Globus and gLite middlewares. Starting from Globus Toolkit v4.0.5 GridWay become standard part of its distribution. GridWay system provides comprehensive documentation for both users and developers that is an important point for implementing new features.

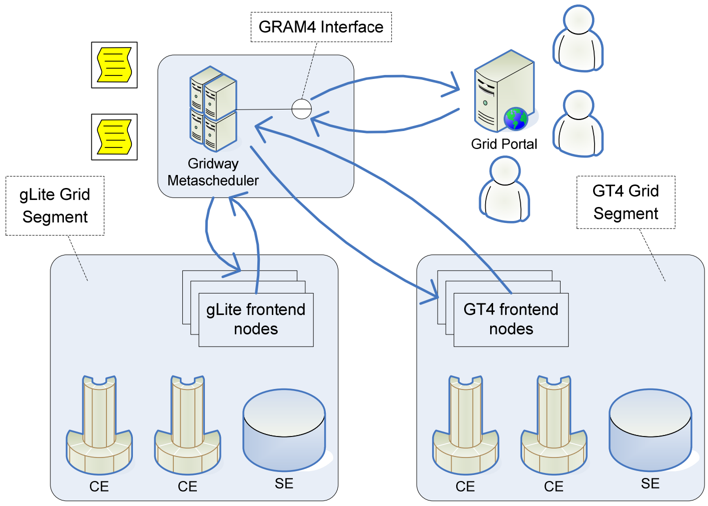


Figure 38. Metascheduler approach

In the next section, we show the examples of application of described approaches to integration of satellite monitoring systems and development of InterGrid environment.

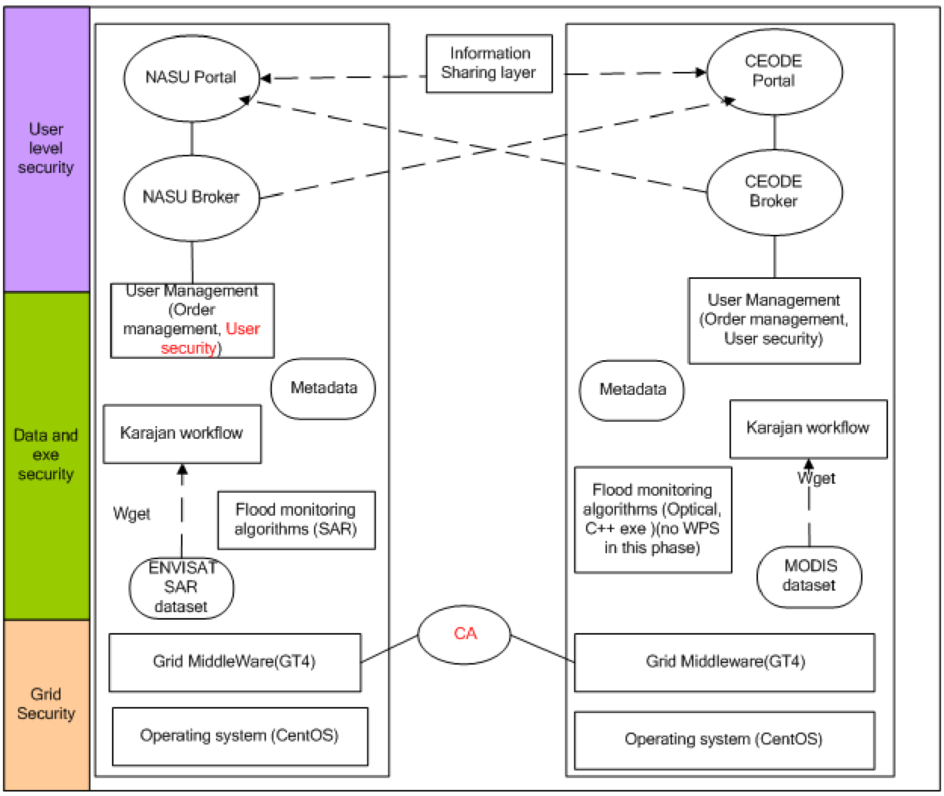


Figure 39. Three-level architecture of Wide Area Grid (WAG)

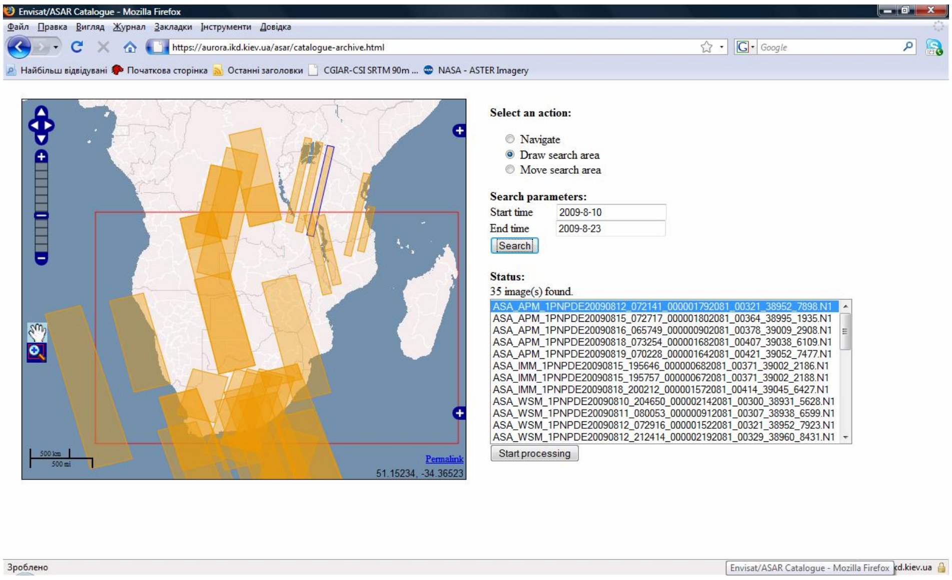


Figure 40. User interface of WAG geoinformation system

This approach is implemented within Wide Area Grid, developed within the project of CEOS by Space Research Institute NASU-NSAU, French Space Agency CNES, and Center of Earth Observation and Digital Earth. Structural scheme of its implementation is shown in Figure 39 and user interface – in Figure 40.

2.6 Implementation: lessons learned

✓ Integration of satellite monitoring systems

The first case study refers to the integration of satellite monitoring systems of NSAU (Ukraine) and IKI RAN (Russia). The overall architecture for integration of data provided by two organizations is depicted in Figure 41. The proposed approach is applied for the solution of problems for agriculture resources monitoring and crop yield prediction. Within integration NSAU provides WMS interfaces to NWP modeling data (using WRF model) [Kussul et al, 2008b], in-situ observations from meteorological ground stations in Ukraine, and land parameters (such as temperature, vegetation indices, soil moisture) derived from satellite observations from MODIS instrument onboard Terra satellite. IKI RAN provides WMS interfaces to operational land and disaster monitoring system. Both NSAU and IKI RAN provides user Web-interfaces to monitoring systems that support OGC WMS standards.

In order to provide user interface that will enable visualization of data from multiple sources we use open-source OpenLayers framework (http://www.openlayers.org). OpenLayers is "thick client" software based on JavaScript/AJAX and operational on client side. Main OpenLayers features also include: support for several WMS servers, support for different OGC standards (WMS, WFS), cache and tiling support to optimize visualization, support for of both raster and vector data. The provided data and products are accessible via Internet http://land.ikd.kiev.ua. The example of OpenLayers visualization of data from multiple sources is depicted in Figure 42.

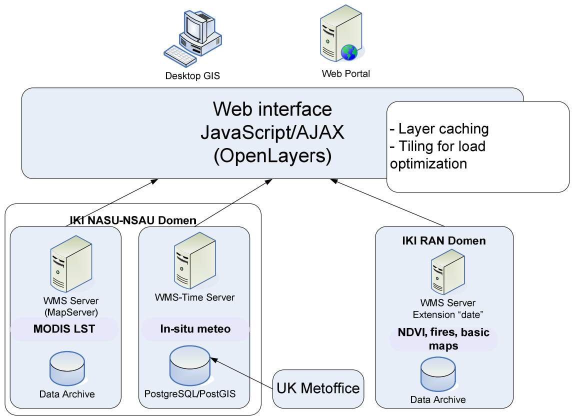


Figure 41. Architecture of satellite monitoring system integration

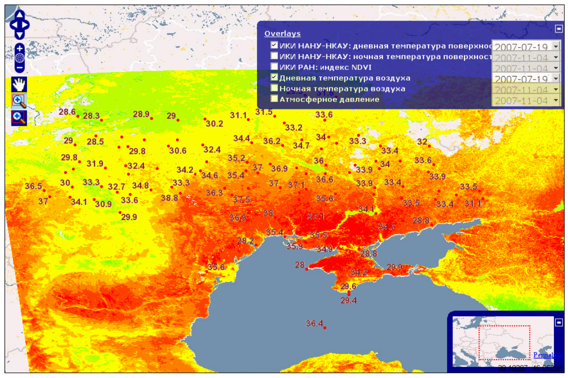


Figure 42. OpenLayers interface to multiple data

✓ InterGrid testbed development

The second case study refers to the development of InterGrid for environmental and natural disaster monitoring. InterGrid integrates Ukrainian Academician Grid (with Satellite data processing Grid segment) and CEODE Grid (Chinese Academy of Sciences) and is considered as a testbed for Wide Area Grid (WAG) implementation—a project initiated within CEOS Working Group on Information Systems and Services (WGISS).

The important application that is being solved within InterGrid environment is flood monitoring and prediction. This task requires adaptation and tuning of existing hydrological and hydraulic models for corresponding territories and the use of heterogeneous data stored at multiple sites. Flood monitoring and prediction requires the use of the following data sets: NWP modeling data (provided by Satellite data processing Grid segment), SAR imagery from Envisat/ASAR and ERS-2/SAR satellites (provided by ESA), products derived from optical and microwave satellite data such as soil moisture, precipitation, flood extent etc., in-situ observations from meteorological ground stations and digital elevation model (DEM).

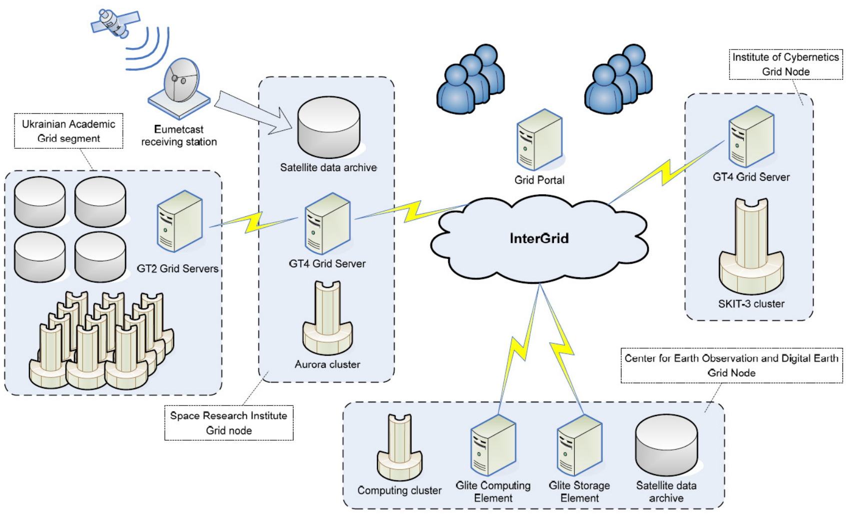


Figure 43. InterGrid architecture

The process of model adaptation can be viewed as a complex workflow and requires the solution of optimization problems (so called parametric study). Satellite data processing and products generation tasks also represent complex workflow and require intensive computations. All these factors lead to the need of using computational and informational resources of different organizations and their resources into joint InterGrid infrastructure. The architecture of proposed InterGrid is depicted in Figure 43.

GridFTP was chosen to provide data transfer between Grid systems. In order to enable interoperability between different middleware (for example, Satellite data processing Grid segment is using GT4; CEODE Grid is using gLite 3.x; Ukrainian Academician Grid is based on NorduGrid) we developed Grid portal that is based on GridSphere portal framework (http:// www.gridsphere.org). The developed Grid portal allows users to transfer data between different nodes and submit jobs on computational resources of the InterGrid environment. The portal also provides facilities to monitor statistics of the resources such as CPU load, memory usage, etc. The further works on providing interoperability between different middleware are directed to the development of metascheduler using GridWay system. In the nearest future, we are intended to provide integration with ESA's EO Grid-on-Demand infrastructure.

The system is used within the UN-SPIDER project for flood monitoring and prediction (Figure 44).

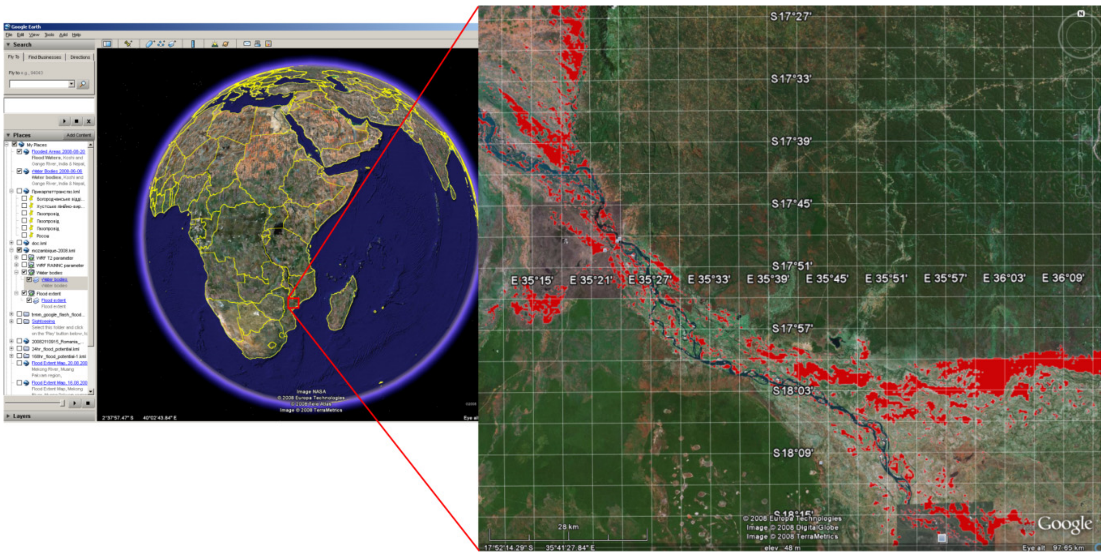


Figure 44. Global UN-SPIDER flood monitoring and risk assessment system

In conclusion, we need to point, that the Ukrainian segment is implemented under the standards of GEOSS. We use intelligent data processing technique for geographically distributed information, and this allows us to provide visual data mining and risk assessment for large-scale disasters. We studied different approaches to system integration allowing uniting different national risk assessment systems into common international infrastructure, for example, UN-SPIDER.

3

Investigation of Geodynamics of Central and Eastern Europe, Balkan Peninsula and Bulgaria

3.1 Introduction

Within the initiative of the European Community for monitoring of the Earth **–** Global Monitoring for Environment and Security (GMES) information about the physical, chemical and biological systems of the planet is collected. Among them are the elements of the geodynamics. However, its multilateral aspects of manifestation are subject of independent purposeful intensive investigations especially in recent time. On the one hand it is determined by the great scientific and practical importance of these investigations and on the other hand – by the direct and very often negative or adverse effect of the hazardous geodynamic processes and the necessity of their multilateral study and finding of methods and measures for reaction and prevention. In this aspect, the advanced scientific and practical achievements are very important. GNSS (Global Navigation Satellite Systems) and InSAR (Interferometric Synthetic Aperture Radar) are presently the main sensors of the Earth's surface deformation. This trend will be strengthened in the next decade because of the ongoing modernization of the Global Positioning System (GPS) and the deployment of GLONASS, GALILEO and other GNSS systems. Similarly, the launch of several new SAR sensors is now schedule for the next 5 years, such as SENTINEL-1. Presently GNSS can monitor the Earth's surface deformation in global Reference Frames like the ITRF 2005 and GPS velocity fields today can monitor site motions at a level of 0.5 mm/yr for the horizontal and 1 mm/yr in the vertical component. Regional studies are deriving the crustal motion and strain patterns for particular areas, like [Nocquet and Calais, 2003], [Caporali et al, 2008]. The strain rates can be resolved at the nanostrain or 1 ppb (10^{-9}) level, provided long enough time series are available. In all present studies, the vertical rates, essential for monitoring the present day status of the lithosphere, are deteriorated due to imperfect modeling. For the combination with other space and terrestrial techniques a uniform and homogeneous (re-)processing of all data according to the latest models is required. The derivation of the related products for position time series and the derivation of a numerical velocity gradient tensor field solution (i.e., spatial variations of horizontal strain rate tensor components and rotation rates) for the Central European platform are required, too. Dense data in the combination with other geodetic data on Earth-surface change, like InSAR and levelling will provide more insight to the present state of change of our environment.

Future projects in the frame of GMES therefore have to aim at an integrated approach of all techniques and multi-disciplinary projects. By this, the determination of strain rates and vertical rates can be extended to the monitoring of hazards and regional or local deformation regimes. Therefore, systematic effects, like periodic or episodic motion and the individual characterization of region may be separated. Main questions are the determination of recent tectonic displacement rates in the central part of the Central and southern parts of the European Platform, their separation from glacial-isostatic adjustments as well as environmental and anthropogenic effects. A major scientific question is how to combine multiple methods with different temporal and spatial resolutions to receive a single homogeneous product with information from the scale of tectonic plates to a fault or a landslide. InSAR and GNSS time series will allow the identification of surface and mass changes and in combination with spatial gravimetry data, a complete picture on the changing Earth can be obtained.

Many international projects are already realized, for example, the GNSS infrastructure implemented by the EU FP 6 project CERGOP-2 in the framework of the Consortium for Central European GPS Geodynamic Reference Network (CEGRN) and the others are in the process of realization. Solutions and results are presented here from the cooperation within the framework of EC related to the regional collaboration in the Central and Eastern Europe (CEE) and in the Balkan Peninsula (BP) as a whole and in particular countries and specially in Bulgaria. It is a good example for fruitful cooperation with a particular importance. The results reported in this study are therefore a step in the direction to the global monitoring of the system Earth and have to be continued in the future GMES framework.

3.2 General information about the investigations of the global geodynamics in the context of Central and Eastern Europe and the Balkan Peninsula

The transition zone between the African and Eurasian Plate at the Balkan Peninsula (BP) is the most active one in Europe concerning seismicity and tectonic movements (Figure 45).

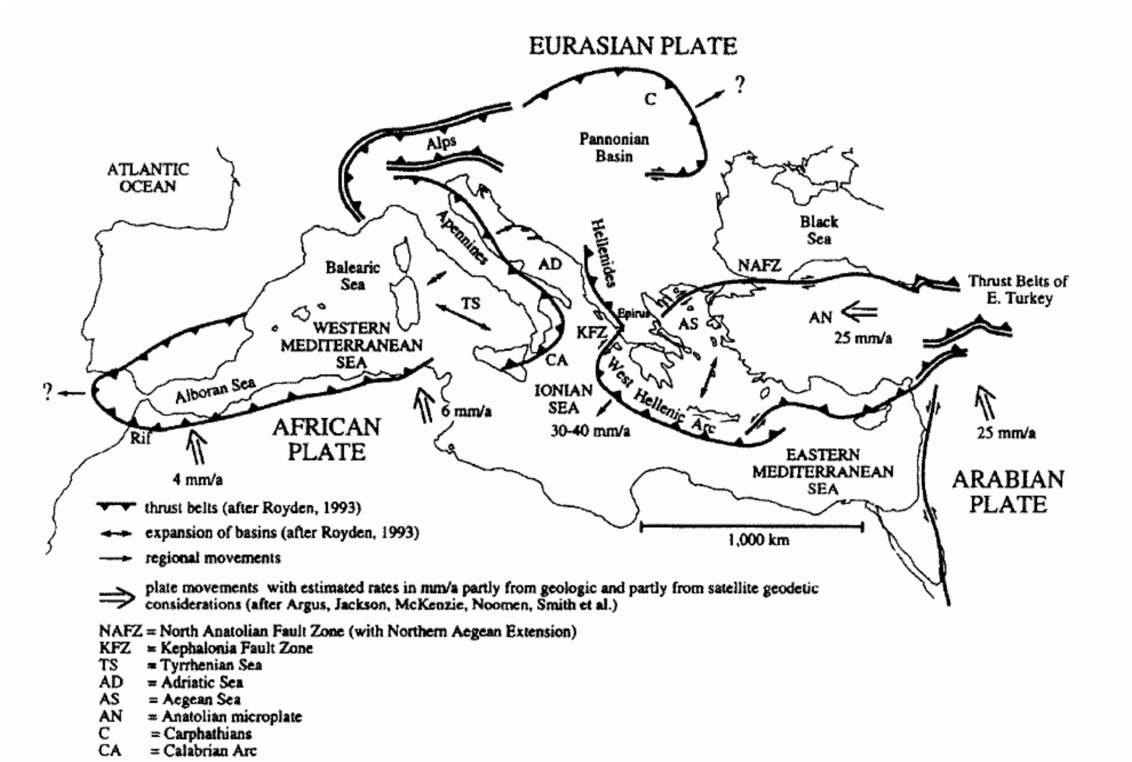


Figure 45. Simplified sketch of main tectonic plates and sub-plates in the Mediterranean area [Plag et al, 1998]

Therefore, the study of geodynamics and geokinematics of this region is very important for understanding natural hazards, like earthquakes, tsunamis, and volcanic activity. From geology it is known that the transition zone consists of several tectonic parts with different behavior, like the Adriatic Microplate, the Aegaean Plate, the Antaolian Plate and others. Apart from the seismic information, new methods of very accurate positioning by satellites like GPS allow to study crustal deformations at the surface at the millimeter level. Starting with the late nineteen eighties with laser and GNSS campaign measurements twenty years after a geokinematic model was computed for the region between the Hellenic Arc and Northern Greece and Anatolia [Reilinger et al, 2006]. The tectonic units can be described by abrupt changes in the velocity fields generated from the movements of the different epoch stations. Using a block model approach the associated kinematic energy can be compared to the energy released during earthquakes, which is somehow related. With the more frequent usage of GNSS permanent stations, more detailed studies can be done revealing more features [Stangl and Bruyninx, 2006]. For example, time series of coordinates of permanent stations show significant offsets after an earthquake in the neighborhood [Hollenstein, 2007]. With yearly velocities between 10 and 30 mm, accurate results can be achieved in the Hellenic region in short time. Farther north in the Balkan Peninsula the movements relative to the Eurasian Plate become much smaller in the range of 1-5 mm/year which are more complicate to detect and to interpret. These intra-plate movements at the Balkan Peninsula are the main object of research of the long-term project CERGOP (Central European Regional Geodynamic Project), covering the region between the Baltic and the Mediterranean Sea. Starting in 1994 CEGRN campaigns at well-determined sites try to deliver a picture of the movements in Central and Eastern Europe and the Balkan Peninsula. A network of permanent stations in that region is added to monitor the movements permanently.

3.3 Geodynamic investigations of CEE

The CERGOP networks have been adjusted to form consistent time series of the stations and to derive station velocities [Hefty et al, 2009], Figure 46 and Figure 47).

The general difficulties for deriving a velocity field for geodynamic investigation are the limited coverage of observations in time and space and the separation of local deviations from the general picture. As can be clearly seen in the figures there are empty spaces because of missing participation of some countries. Campaigns as well as permanent sites need a sufficient sequence of observations to derive accurate velocities. A minimum of three occupations for campaign sites is requested which means an occupation time of at least 2-3 years, for CEGRN sites presently 4-6 years [Stangl et al, 2008a,b]. Permanent sites require at least two years of observations to avoid misinterpretation by seasonal effects and potential local disturbances. The Istrian sites in Figure 46 show the problem of short occupation leading to high velocities, which are not reliable. The very high velocities of the sites CLUJ and DEVA in Romania (Figure 46 and Figure 47) are well determined, but were found to experience heavy local movements associated with the mining activities there. When these two stations are neglected, both figures show two velocity provinces in Southeast Europe which differ significantly from the general horizontal rotation of the Eurasian Plate. Around the Adriatic Sea the stations move generally to the North by 3-5 mm/year which is caused by the Adriatic Microplate. South of the Carpathians a transition zones seem to be pulled south with a small rotation focused at Bulgaria by 2-3 mm/year [Aichhorn et al, 2008]. Lack of observations in the region of the Dinarides inhibits to draw accurate boundaries inside the Balkan Peninsula. Combination of velocity fields from various sources leads to a slightly better picture [Caporali et al, 2007; 2008; 2009], but the interior of the Balkan Peninsula is still largely uncovered by observations.

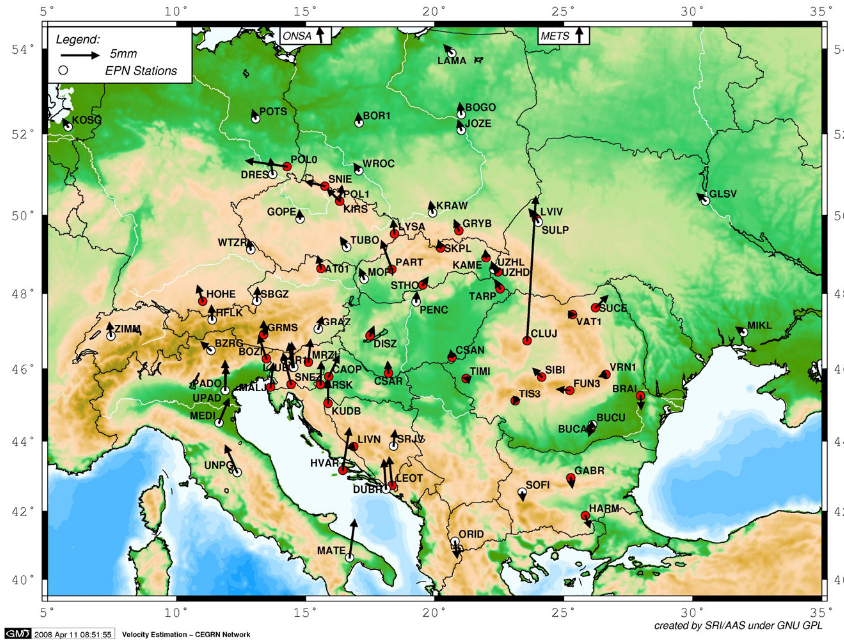


Figure 46. Horizontal velocity estimations from reprocessed time series of CEGRN campaigns 1994-2007

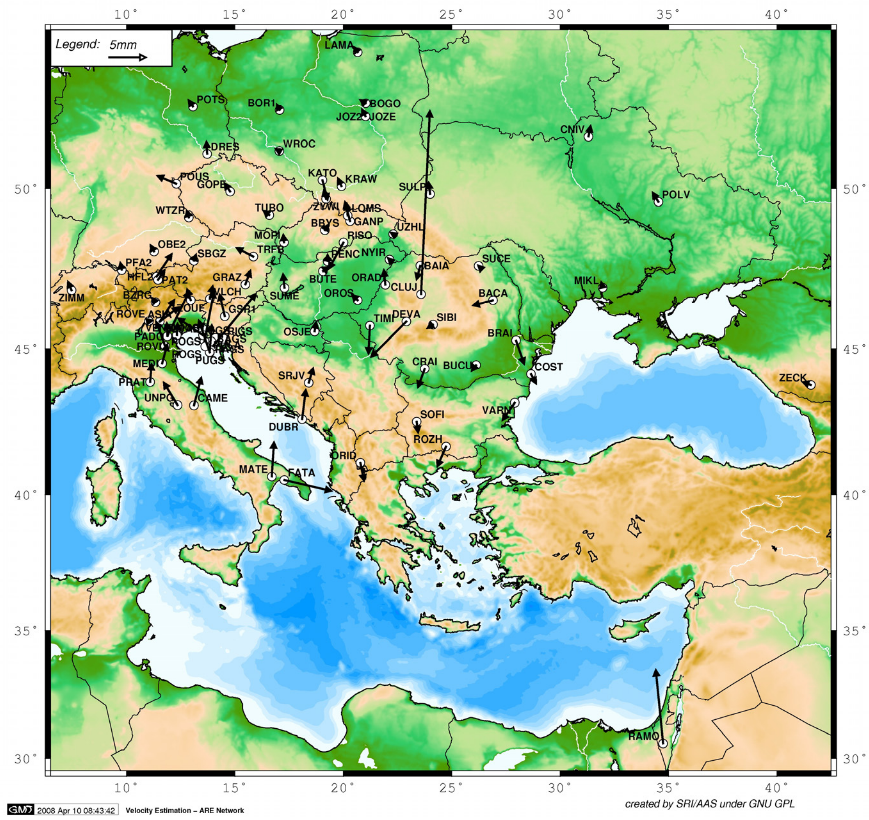


Figure 47. Horizontal velocity estimations from cumulated weekly time series of CERGOP permanent sites 1999-2008

3.4 Geodynamic investigations of BP

The territory of Balkan Peninsula is characterized by active geodynamics. It is the most active region in Central and Eastern Europe in geodynamical respect. A number of hazardous geodynamic processes of endogenic (earthquakes, contemporary movements of the Earth's crust, mud volcanoes) and exogenic origin (natural and technogenic), including landslides, abrasion, erosion, subsidence, collapse, rockfalls, mud-stone flows, deformations caused by mine workings, karst, etc., are observed in the area.

The tectonics on the Balkan Peninsula is determined by the fold structures of the Alpine-Himalayan orogen. The territory of the peninsula had been subjected to older orogeneses too. The Alpine-Himalayan orogen itself, which builds the three big South-European peninsulas and the greater part of Middle Europe, is bilateral. Its northern branch is represented by the Alps, the Carpathians, and the Balkanides and is called the Carpathian one. The southern branch, called the Dinarian, comprises the real Dinarides and the Hellanides. Both branches are equally well expressed and are in close contact on the Balkan Peninsula.

At present the space of the Balkan Peninsula falls within the zone of collision between three large plates – Euroasian, African and Arabian, which are themselves divided in smaller ones (Figure 45).

The geostructural situation and the geological evolution represent a substantial element of the general situation of relatively high seismicity in the region. The present-day geodynamics of the Balkan region is controlled by the active tectonic processes in the Eastern Mediterranean (Figure 48).

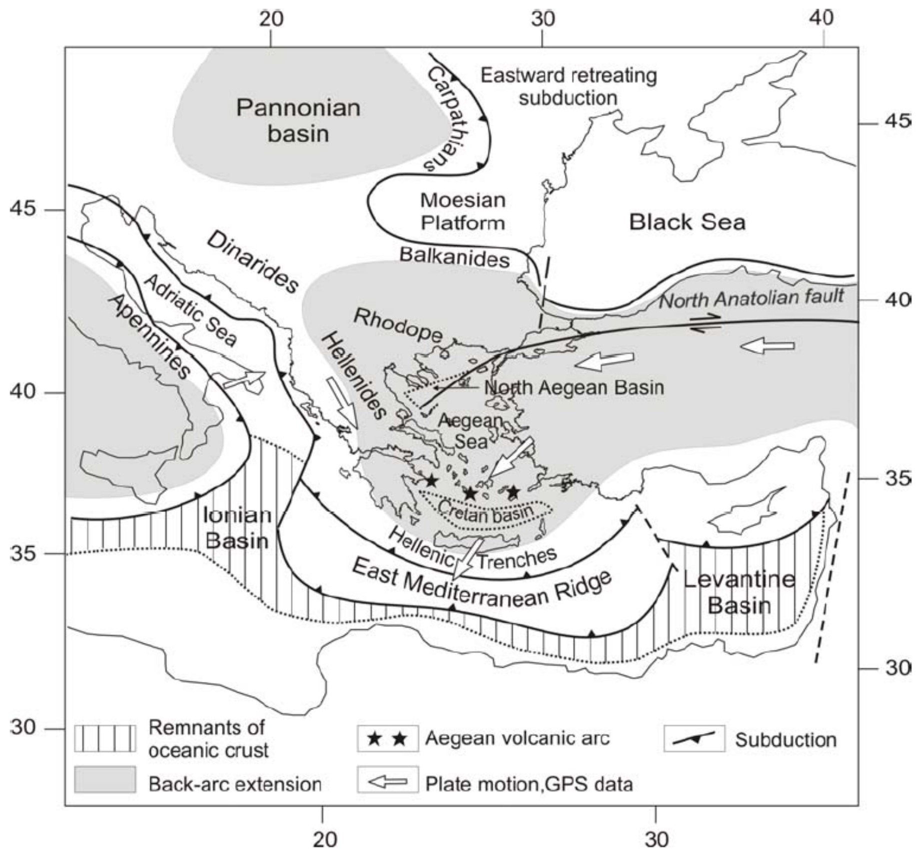


Figure 48. Active processes in the Eastern Mediterranean (based on [Carminati et al, 2004])

There exists a subduction of the Adriatic (Apulian) microplate beneath the Dinarides, a subduction of oceanic Ionian and Levantine lithosphere under the Hellenic arc-and-trench system and a collision between Eurasia and Arabia with related westward escape of Anatolia along the North Anatolian dextral strike-slip fault.

The Balkan Peninsula was a subject of study in a special working group of the international project – Central European Regional Geodynamic Project (CERGOP 2). The investigations performed are based on the GNSS campaign observations, operation of GNSS permanent and epoch stations, data processing, interdisciplinary geodynamic analysis and interpretation.

Except geophysical methods, recently GNSS have been widely applied for precise determination of earth crust movements in the region. The permanent GNSS stations are extensively increased.

3.4.1 Investigations of the particular Balkan countries

A concise review of the tectonic setting for each BP country is outlined. Detail information can be found in the Monograph "Geodynamics of the Balkan Peninsula" [Milev and Dabovski, 2006].

From the geological standpoint, **Albania** belongs to the Dinarides, the southern branch of the Alpine folded belt. The Dinarides are separated into two: the Dinarides and the Hellenides. The Dinarides pass into the Hellenides in Albania, most of the country being encompassed by them.

The Aegean (Hellenic) Arc, visible in the southern part of the Hellenides, is situated between the Arabic-Africa/Eurasia collision to the east and the Adriatic collision to the west.

The neotectonic zonation of Albania is based on tectonic regime and types of deformations, which occurred. Four large neotectonic units have been recognized, each of them based on the sense, intensity and chronology of vertical movements. The external margin of the fold and thrust belt in Albania and its surroundings was thrust on Adria microplate, partly over Apulian platform and partly over Albanian Basin.

The area of **Bosnia and Herzegovina** is included in the middle parts of the Dinaridic Mountain System and it is positioned between Apulia (Adriatic Microplate) in the south and the Panonian and South Tisia, respectively. It is located NE from active compressional geotectonic contact between the Adriatic mass and the Dinarides. The Adriatic mass, as part of Africa, is impressed between Apennines and Dinarides along strike-slip active faults.

According to the activities during the last 100 years, the Bosnia and Herzegovina was divided into 5 seismogenic zones and 57 potential seismoactive structures. Their lengths are between 6 km to 40 km. Based on analysis of tectonic data it is concluded that the stronger seismic activity occurs at the boundaries of geotectonic units (directions NW-SE), then along the longitudinal dislocation (directions NW-SE), at the intersections of transversal faults (directions NE-SW and N-S). In the year 2000 the National Offices for Surveying and Mapping in Bosnia and Herzegovina set up a GPS network comprised of about 30 stations.

The major role in geodynamics of Croatia belongs to the Adriatic microplate Dinnarides. Seismically the most active part is the coastal part of Croatia, especially its southernmost part. Two permanent GPS EUREF stations were installed on the territory of Croatia – Osijek and Dubrovnik. Several GPS campaigns were carried out. For monitoring the tectonic movements around Zagreb a local GPS networks is established. It is studied that the zone of Dinnarides delineated by the Alps in the north, the Adriatic in the southwest and the Pannonian basin in the northeast is seismically and tectonically very active area. GPS measurements yield very accurate displacements on regional or local level.

The area of western **Greece**, a seismically active region, exhibits substantial tectonic complexity mainly dominated by the counter clockwise motion of the Apulia microplate in the north and the active subduction of the front part of the African plate beneath Eurasia, in the south.

Monitoring of the displacements of the stations of the geodetic control network in the seismic zone of Volvi has been a task carried out for approximately 25 years. Starting 1994, the extended Volvi network was measured by GPS methods in five epochs (1994, 1995, 1996, 1997, and 2003). The analysis of the geodetic data shows that significant deformation occurred in the area until 1994. The highest rate of deformation observed after the 1978 earthquake (about 6 mm/year) is strongly correlated with it. Afterwards, GPS data show that there is a relaxation of the deforming body in general, with the exception of a few points.

The long-term seismicity in the **Macedonia-Bulgaria** border region has been critically considered. The revised picture suggests that the earthquake catastrophes in the region occur at intervals of 200 – 400 years on the average. Two peculiarities are marked as common to these series: tightening of the preceding seismic activity towards the location of forthcoming main shock and lack of strong aftershocks after coupling of the catastrophic events. The first GPS surveys in Macedonia were conducted in 1996. At that time, GPS observation on seven points was carried out in Macedonia and those points were merged with the global ITRF Network. During the conveyance of the GPS campaign in 1996, a permanent station in Ohrid was placed, as a point being part of the European GPS Network. At the beginning of 2004, large-scale GPS surveys were undertaken by which a larger number of points on the territory of R. Macedonia were encompassed.

Plate boundaries on the **Romanian** territory and their dynamics have been recently reviewed with special emphasize on the role played by the Black Sea opening. Short-distance large variations in the lithosphere thickness advocate for the presence of at least three plates/sub-plates in the area. East European plate (EEP), with a lithosphere thickness of more than 150 km, Intra-alpine micro-plate (IaP) with its 80-90 km lithosphere, and Moesian micro-plate (MoP), with 120-150 km lithosphere thickness, meets each other in the Vrancea unstable triple junction, where the asthenosphere was revealed at more than 250 km in depth.

The territory of **Serbia and Montenegro** was subject of complex geological investigations for different research and applied purposes. Reason for so significant interest is in very complex geological relationships and presence of geological formations at the area. At the territory of Serbia and Montenegro, the highest seismic activity is characteristic of Dinaridic seismogenous block (Montenegro and SW Serbia), with over 70% events.

The strong earthquakes caused by intensive tectonic processes, predominantly occurring in the coastal part of the territory, produce destructive effects in the form of landslides, avalanches, and soil liquefaction.

Although **Slovenia** is a small country, it lies in a tectonically very interesting area. Its territory is the meeting-point of three different geological units: Alpine, Dinaric and Pannonian. The larger part of Slovene territory, i.e. the northern and northwestern parts of the country, belongs to the Alpine region. The smaller, southern part of Slovenia with the Karst belongs to the Dinaric region, and the Pannonian massif is limited to the eastern part of the Prekmurje region.

Turkey is the country affected by several different faults such as Black Sea plate, Eurasian plate, Aegean plate, African Plate, Arab Plate, and Anatolian Plate. Due to this structure, almost 92% of the country is under the risk of earthquake. Most of the micro geodetic networks have been established for monitoring geodynamic activities on North Anatolian Fault (NAF). Turkish National Fundamental GPS Network (TUTGA) has been established in between 1997 and 1999 and some of the stations have been re-surveyed due to the earthquakes happened in 1999. The total number stations are 596, for each station 3D coordinates, and their associated velocities have been computed. The Turkish national permanent GPS network (TUSAGA) is still in establishment phase with its 16 operational stations of which data can be used.

3.4.2 Generalized investigations and results

✓ GPS data processing

An important indicator in geodynamical point of view is an eventual station movement of GNSS network stations. That concerns the permanent and epoch GNSS stations on the territory of the Balkan Peninsula. Determination of the size and direction of the vectors of movement, their analysis, and assessment is of significant importance for their further interpretation with a view to present the geodynamical picture of the region.

Geodynamical behavior of GNSS stations (Figure 49) of the CEGRN sub-network covering Balkan Peninsula has been studied and analyzed in number of papers. Data from different GPS CEGRN measurement campaigns concerning the Balkan Peninsula stations (permanent and epoch) have been involved.

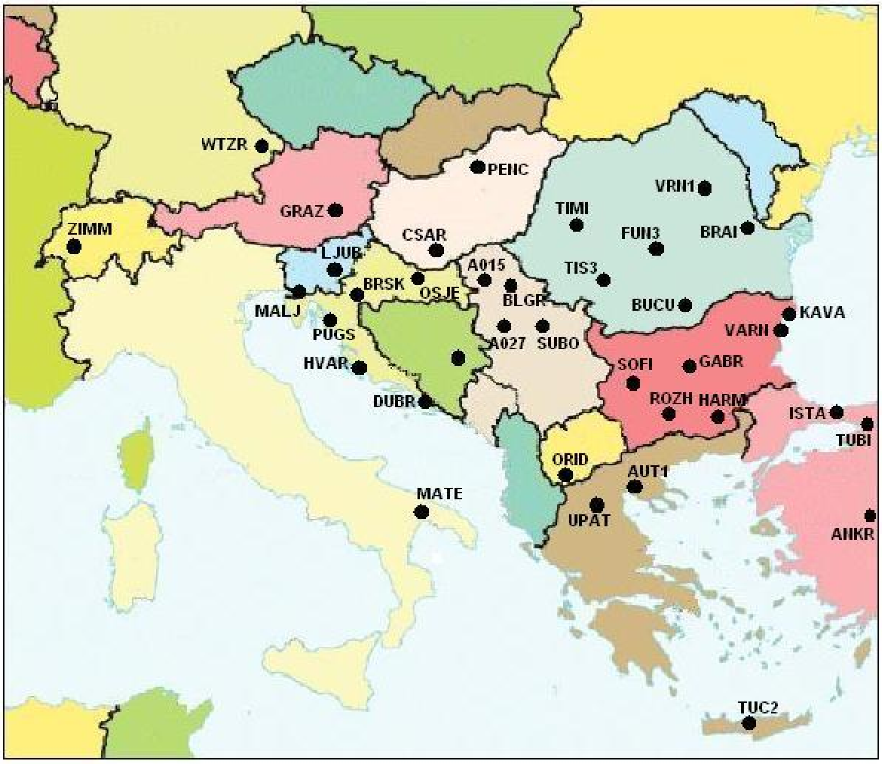


Figure 49. Balkan Peninsula CEGRN sub-network stations

Data processing has been accomplished with the scientific software Bernese Software, version 5.0. Before comparing and analyzing the most important from geodynamical point of view results, namely, velocity vectors the quality and reliability of station coordinate estimations obtained have been studied and analyzed.

Combined solutions of different BP’CEGRN have been done and published in [Milev et al, 2005]. The estimated velocities from different combined solutions have been compared and analyzed. The problematic stations have been localized and some explanation for the problems within the respective CEGRN campaigns concerning Balkan Peninsula sub-network have been done.

Figure 50. Balkan Peninsula permanent stations

The latest investigations covered the period 2007-2009 [Vassileva, 2009a]. For more precise and detailed monitoring of the tectonic movements by GNSS, the number of the permanent GNSS stations in the region of Balkans is permanently increased. Within this period, they increased from ten in 2007, up to twenty-two in 2009.

All free available stations (Figure 50) on the territory of the Balkan Peninsula are used for study of their movement within three years – 2007, 2008, 2009.

GPS data of one week in each year and their combinations are processed. The obtained station velocity estimations have been analyzed and compared with other results.

✓ Analysis of velocity estimations

All possible combinations of the obtained weekly solutions in 2007, 2008 and 2009 have been processed and station velocity estimations have been obtained in system ITF2005 [Vassileva, 2009b]. Minimum constrained adjustments have been applied for all combinations using the same reference stations. All possible combinations of two years give different results for most of the stations to some extent. Comparison with the annual station velocity estimations of the IGS, EPN, and CEGRN results shows deviations, which are considerable for most of the stations. It confirms again the conclusion drawn in [Vassileva, 2009a] that two years solutions do not generate reliable results especially using two consecutive years.

Velocity estimations of IGS/EPN/CERGOP stations obtained are compared with those ones obtained from the respective IGS/EPN/CERGOP solutions and the differences are within 1-3 mm (Figure 51). For some of the stations greater differences between the velocity components have been obtained and an eventual reason like antenna changes has been assumed.

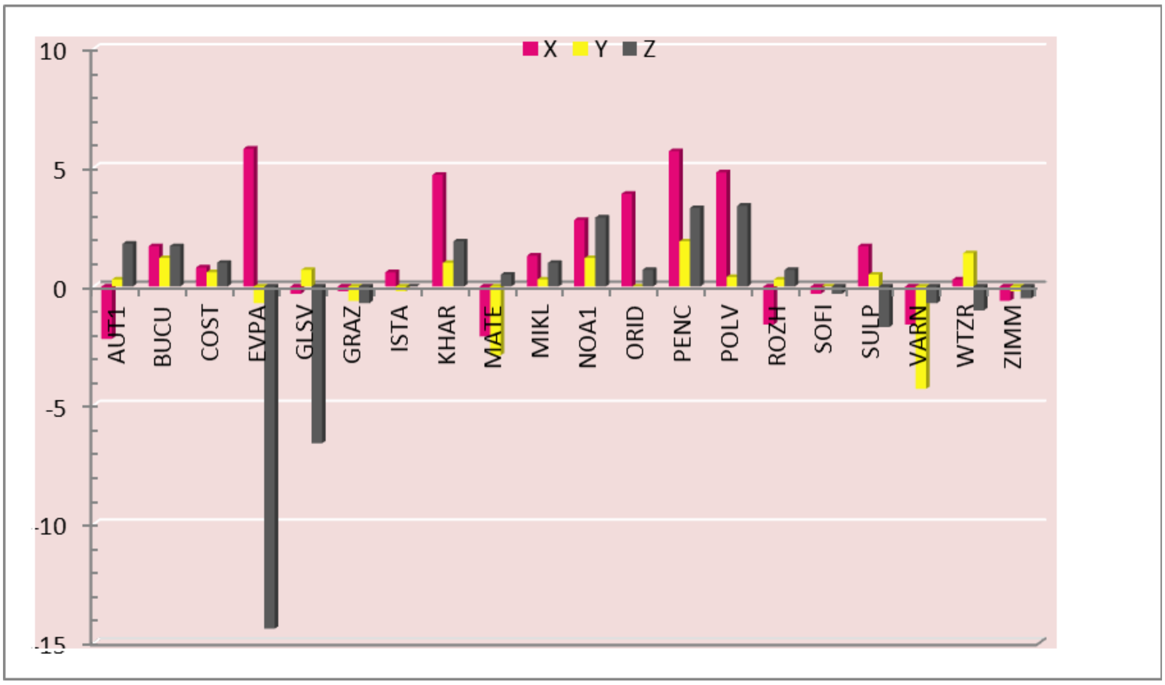


Figure 51. Station velocity differences between results of this study and estimations from IGS/EPN/CEGRN

Three years weekly data processing and their combined solution have been accomplished and the comparison of obtained results for the velocity estimations of most of the stations shows a good consistence. For some of them differences in the different components of the velocity vectors are higher and they are not acceptable. Conclusions, which can be drawn here, confirmed these ones drawn in [Vassileva, 2009a] and they are:

– velocity estimations obtained from two years data independently of the time interval between the years are not reliable and they should not be used;

- for obtaining reliable estimations of the velocity vectors of permanent stations GPS data from more than three years should be used;
- velocity estimations obtained from weekly data are not very reliable. For obtaining representative velocity vectors, it is advisable to be used annual data solutions.

These conclusions undoubtedly confirm the suggestion given in [Vassileva, 2009a] about determination of velocity vectors of national network reference points and geodynamic network points. The suggestion concerns involving of long term data of several years in obtaining of point velocity estimations for such a type of points. Velocity vectors obtained are crucial for further analysis and interpretation and they should be reliable estimated.

3.4.3 The monograph "Geodynamics of the Balkan Peninsula"

The monograph (Figure 52) is a result of the investigations carried out by the Work Group on Geodynamics of the Balkan Peninsula within the framework of the "Central European Regional Geodynamic Project – CERGOP-2" of the 5th Framework Program of the European Commission [Milev and Dabovski, 2006]).

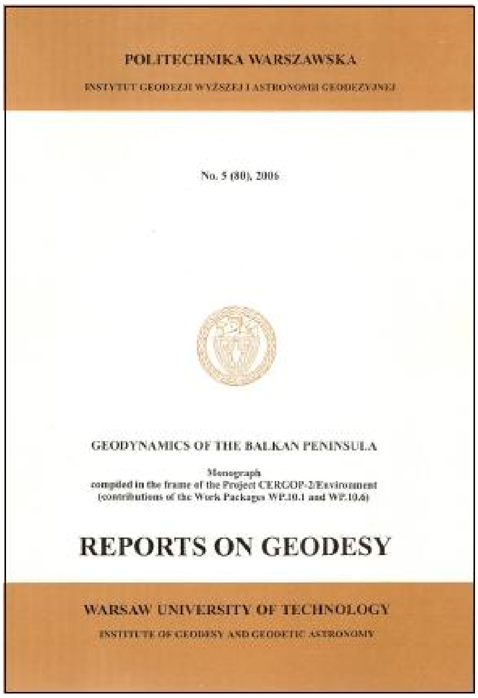


Figure 52. The monograph "Geodynamics of the Balkan Peninsula"

The monograph structure is:

1. Preface
2. Introduction
3. General data and prerequisites for the part of global geodynamics comprising BP
4. Geodynamic investigations in the single Balkan countries
   4.1. Seismological investigations
   4.2. Seismotectonic investigations
   4.3. Geological and geomorphologic investigations; engineering geodynamics
   4.4. Geodetic investigations
   4.5. Complex analysis and interpretation
   4.6. Views on BP geodynamics in the single countries
5. Generalized investigations
   5.1. Analysis and generalization of the results obtained from the GNSS – geodetic investigations
   5.2. Complex interpretation of the velocity and deformation vector field
   5.3. Generalized geodynamics of BP
6. Conclusion

A special attention is paid to the fight against the hazard geodynamic processes and to the registration and documentation of the effects of their impact, including creation of special geographic information systems.

The Monograph provides an accurate picture of the position and particular features of BP in the overall geodynamic picture of the Earth and the region. It comprises 650 pages, contributions of 96 authors. It is addressed to a broad range of specialists involved in the different Earth sciences and protection against natural disasters.

3.5 Geodynamics of Bulgaria

3.5.1 Tectonic setting

Bulgaria is situated in the eastern part of the Balkan Peninsula. The eastern part of the Balkan region comprises variable in type structural units differing in age, origin, tectonic style, and contemporary geodynamics.

The earthquake activity in Bulgaria is the most apparent manifestation of contemporary geodynamics on its territory [Christoskov et al, 2006]. Over the past centuries, Bulgaria has experienced strong earthquakes. Some of the Europe’s strongest earthquakes 20-th century occurred in Bulgaria. The hazard map for a 1000-year return period is presented in Figure 53. The adopted Bulgarian building construction code is based on this hazard map.

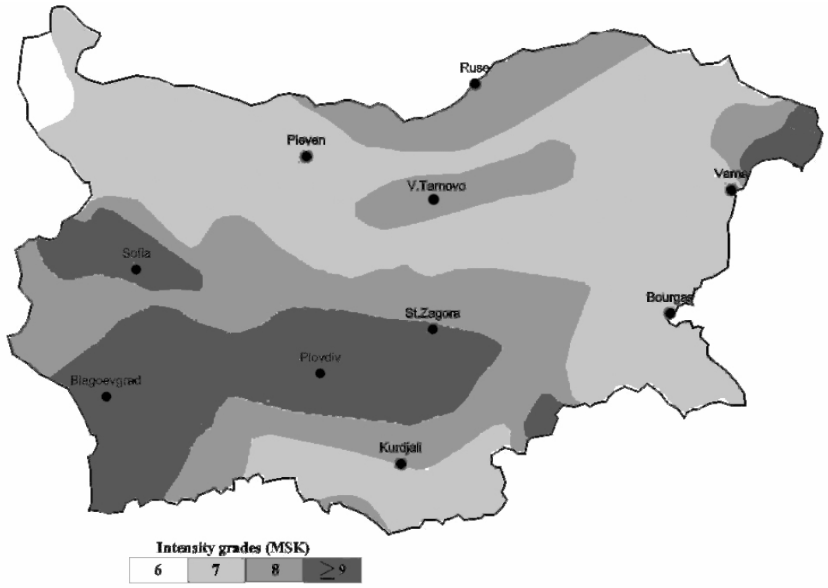


Figure 53. Hazard map for a 1000-year return period

The historical seismicity in Bulgaria has not been satisfactory covered by the existing information sources [Botev et al, 2006]. The supporting dataset quality has been improving in the course of time, although the data points’ number has not been significantly changed.

After establishing (in 1891) Bulgarian Service for regular observing and documenting earthquakes felt in the country, the seismicity over its territory as well as seismic impacts from seismic zones around are known quite better. The nowadays monitoring of the weak earthquakes in Bulgaria is based on a modern National seismological network, which was established in 1980. More than 15 000 events recorded during the operation of this National network are located on the territory of Bulgaria and its close vicinity.

3.5.2 Geodetic measurements

First geodetic measurements for the territory of Bulgaria on the CERGOP project were carried out in 1996. Only two Bulgarian stations (SOFI, HARM) of the four planned stations (SOFI, HARM, GABR, KAVA) were involved in the CEGRN’96 and in the CEGRN’97 GPS campaigns (Figure 54).

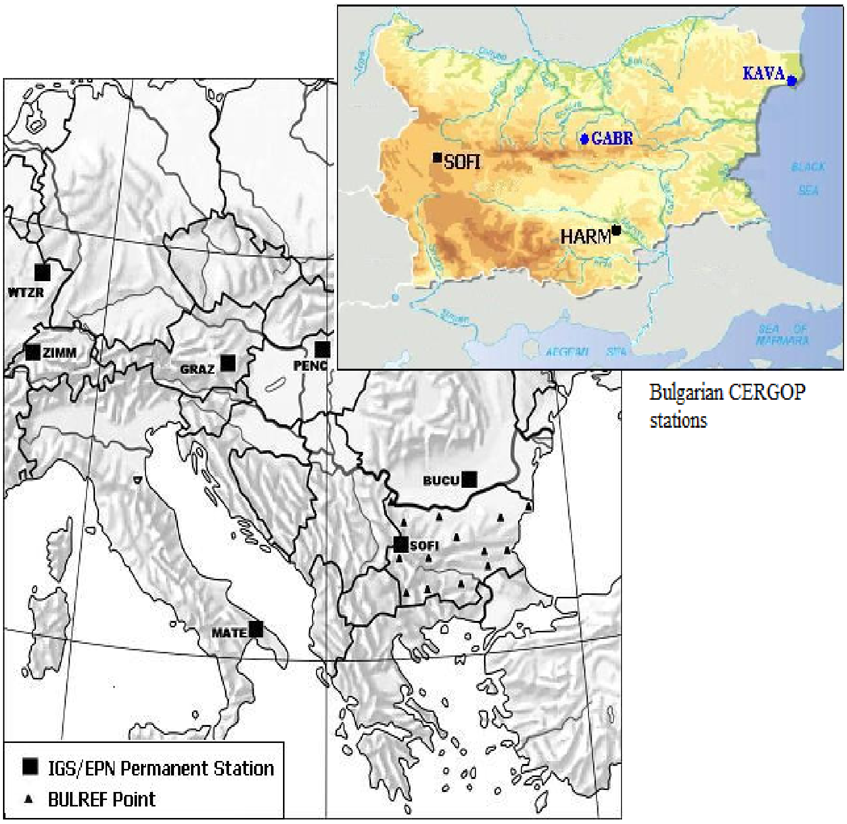


Figure 54. Bulgarian stations participated in CEGRN03 campaign

In the second stage of the project – CERGOP-2 during the CEGRN'03 GPS campaign Bulgaria participated with all 15 BULREF points as three of them were the CERGOP epoch stations – GABR, HARM, KAVA and one IGS/EPN permanent station – SOFI (Figure 54) [Milev and Vassileva, 2004]. The BULREF points were observed earlier in 1993 during the EUREF – BULREF GPS campaign and as a part of the WEGENER-MEDLAS geodynamic project. In the CEGRN'05 GPS campaign, the country was involved with one EPN/IGS permanent station (SOFI), two newly established CERGOP permanent stations (VARN, ROZH) and the three epoch stations (Figure 55).

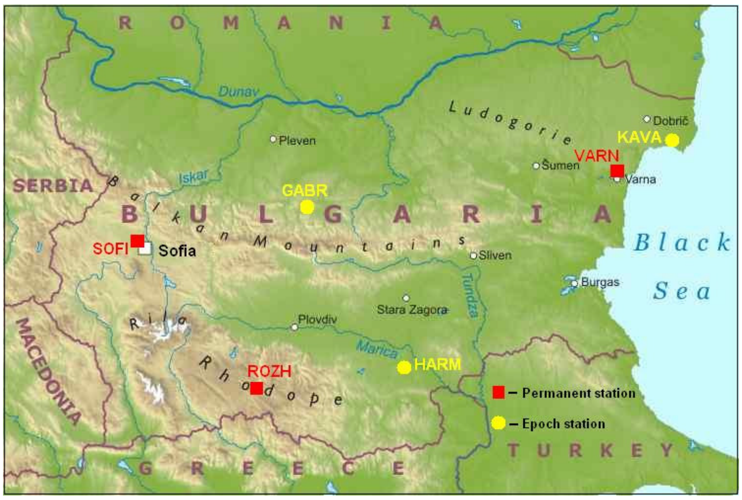


Figure 55. Bulgarian stations participated in CEGRN05 campaign

3.5.3 Geodetic investigations and analysis

For investigation of station movements GPS data of all 15 BULREF stations available in two campaigns – BULREF93 and CEGRN03/BULREF03 were processed with the Bernese GPS Software, version 4.2 [Milev and Vassileva, 2003; 2004], [Vassileva, 2004] regarding the accepted processing CERGOP principles. A combined processing of BULREF'93 and BULREF'03 was accomplished and estimations of the ITRF2000 coordinates for the mean epoch and station velocities were computed. The estimated GPS horizontal velocity vectors (blue), calculated NNR-NUVEL1A velocity vectors (green) and the obtained relative velocity vectors (red) are shown in the Figure 56.

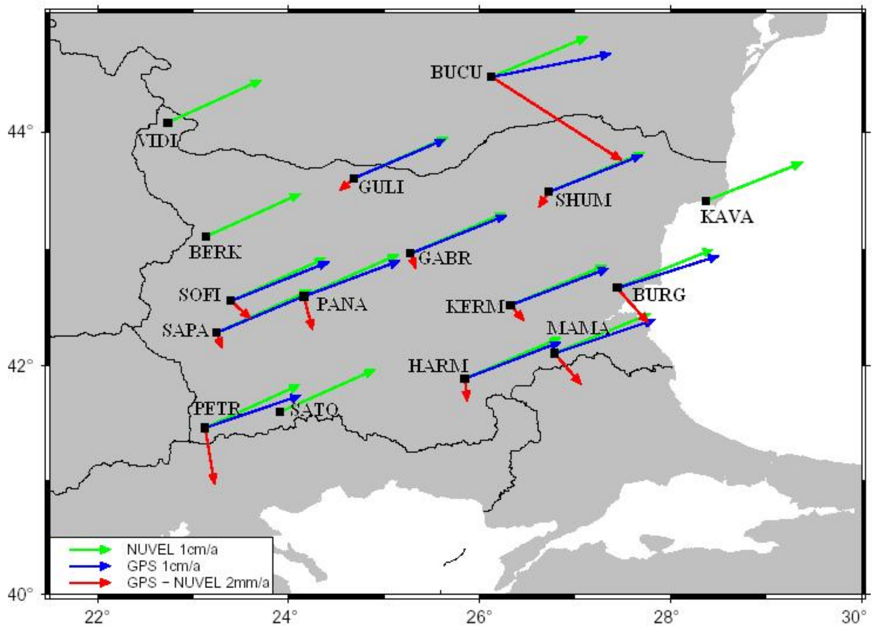


Figure 56. GPS and NNR-NUVEL1A velocity vectors of BULREF stations

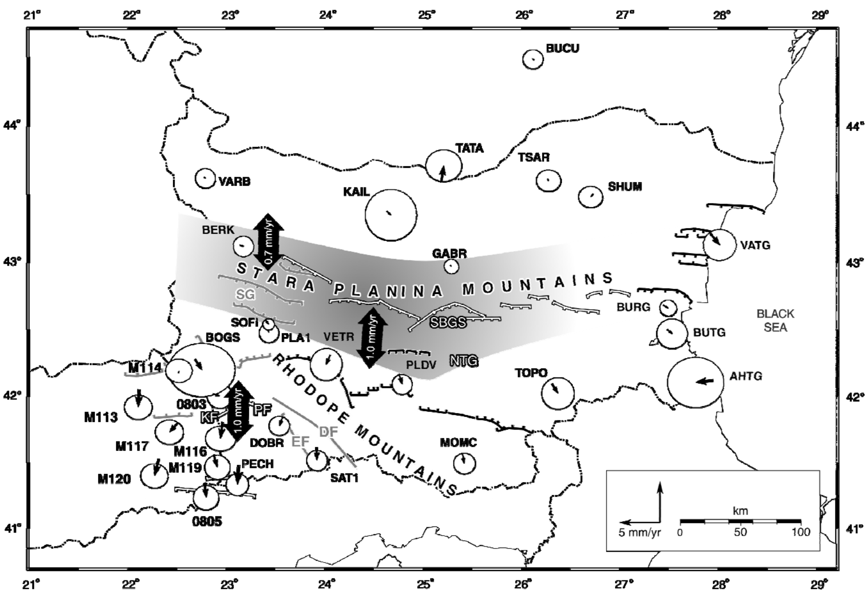


Figure 57. GPS station velocities in south-west Bulgaria

The velocity estimations obtained from the study show undisturbed behavior for all BULREF stations for the period of 10 years, period between the two campaigns. The values of station velocities are very similar and differences vary with an amount of maximum up to 2 mm in north component and in east component. In addition, the deviations to the NUVEL model velocities are in a reasonable size. This indicates that no unexpected jumps or outliers occurred in the behavior of stations during that time.

For study of the active tectonics in south-west Bulgaria, a number of local GPS measurements and data processing were carried out [Nakov et al, 2006]. The obtained stations velocities are shown in the Figure 57 are part from a velocity solution for about 80 GPS stations in the southern Balkan region [Burchfiel et al, 2006], [Kotzev et al, 2006].

3.5.4 Permanent DGNSS networks

At present for the purposes of many applications in the life on the territory of the country, the following DGNSS networks and systems are established or in process of establishment or operating in some extent: BULiPOS, NAVITEQ, Geonet of Zenith-GEO, Leica. The BULiPOS is the only system, which operates over the whole country within the EUPOS (EUropean POsition Determination System) as its Bulgarian segment.

BULiPOS (BULgarian Inteligent POSition determination System) is a DGNSS and a system of permanently operating reference stations for position determination and for space orientation, established and operating within the framework of EUPOS [Milev and Vassileva, 2009].

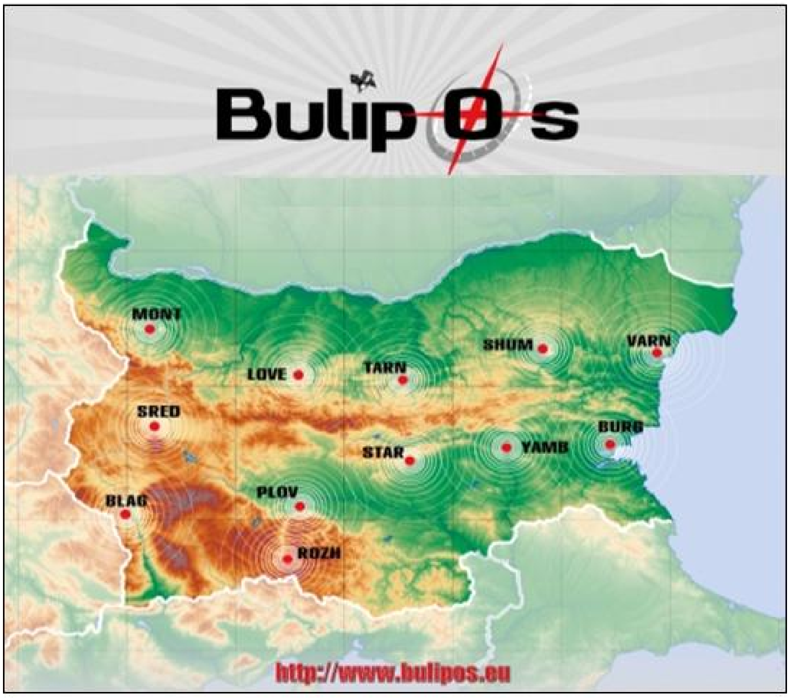


Figure 58. BULiPOS DGNSS stations

The establishment of the DGNSS network BULiPOS was carried out in 2007 [Milev et al, 2009]. The number of the established active reference stations is 12, shown in Figure 58. The stations have been spread over the country evenly, each of them covers an area of 70-100 km in radius, and thus the required DGNSS corrections can be provide with the necessary accuracy and reliability. DGNSS reference stations are multifunctional and they are established according to the respective international standards of EUPOS. The used software for processing and transmitting the differential corrections is Geo++® GNSMART professional.

The BULiPOS is unified homogeneous reference network and system for determination of position and space orientation. It is a high efficient system and meets the technical requirements for solving the problems of all fields of geodesy and cadastre and a variety of other applications in management, legal regulation, and economy of the country.

BULiPOS provides DGNSS corrections for position determination with high precision in real time as well as data for post processing within the territory of the country by the established network of active reference stations.

The coordinates of the stations are provided in reference system ETRS89 (European Terrestrial Reference System 89) more precisely in its realization ETRF2005.

BULiPOS provides, through various kinds of connections, additional information to correct data in the use of the Global satellite systems to determine the position – GPS and GLONASS. The variety of communication media (radio UKW / RDS), 2 m-band, GSM-mobile phone, Internet, Systems for data transfer) allows the transfer of data and accurate determination of location for diverse fields of application with different level of accuracy, from several meters to centimeters and millimeters. BULiPOS offers four types of services, depending on the required level of accuracy shown in Table 2.

Table 2. BULiPOS services

NAME	DESCRIPTION	ATTAINABLE ACCURACY	DATA FORMAT	TRANSFER	SATTELITE SYSTEM
BULiPOS PP	Post Processing	5 mm	RINEX	Internet	GPS+ GLONASS
BULiPOS VP	Virtual Station	5 mm	RINEX	Internet	GPS+ GLONASS
BULiPOS RT Precise	Precise Real Time	2 cm	RTCM2.3	CSD GPRS Internet	GPS+ GLONASS
BULiPOS RT	Real Time Kinematics	0.5 – 3.0 m	RTCM2.3(3.0)	CSD GPRS Internet	GPS+ GLONASS

The multifunctional BULiPOS network for position determination does not only replace a great part of the classical geodetic control network but it allows effective and economic favorable solutions in geodesy, photogrammetry, and other fields. Moreover, it proposes a possibility of position determination of moving objects to all who need to know their position in a fixed moment.

The BULiPOS stations BURG, LOVE, MONT, SHUM, and TARN (Figure 58) have been included in the data processing of the CEGRN network and since 2009.

Furthermore, BULiPOS, respectively DGNSS are actively used for different aspects of imaging and monitoring of the earth surface (aerial imaginary, laser scanning), investigation of different geodynamic processes, designing, etc.

3.6 Discussion

Decisions and results from the cooperation in the frame of EC related to the regional cooperation within the CEE and Balkan Peninsula as whole and particular countries of it is a good example of a fruitful cooperation with significant importance and it should further be developed and used.

The outlined in a large-scale measurements, data processing, investigations, generalizations, and interpretation of the geodynamic processes on the territory of the Balkan Peninsula, their manifestation and impact, the consequences, measures for their exploration and protection are grounded on the complex interdisciplinary basis within the Earth sciences and related fields. They are based on the use of modern systems, devices, instruments and methods, technologies, algorithms as well as the modern knowledge, algorithms and software to determine the present outlook and state of the different Earth sciences and they are reliable enough. They confirm, supplement, and enlarge the existing active geodynamic picture of the Balkan Peninsula and in particular this one for Bulgaria within the framework of the Central and Eastern Europe and the Mediterranean. They outlined the contemporary approach for regional and national investigations of this type and they can be used as an effective base for future measurements and studies in the region of the Balkan Peninsula, and in the particular countries of it.

The usage of GNSS observations for determining crustal intraplate movements in the Balkan Peninsula was successfully demonstrated by the CERGOP project. The missing gaps need to be filled by using time series of permanent stations, which are known to exist in the countries without present data. A data provision would help all countries to a better understanding of the geodynamics of the Balkan Peninsula.

Establishing and using the DGNSS (as BULiPOS a Bulgarian part of EUPOS) unified systems, based on the common standards will solve a great number of problems and applications including those ones required very high precision. Their multilateral possibilities and realized aspect of application in the countries are of great important as well. There is no doubt these are the systems of the present and the future useful for geodynamic the investigations and the GMES monitoring.

4
Intelligent Tools for Environment Monitoring: Features and Applications

4.1 Introduction

Remote sensing for agricultural purposes has intensive development in Bulgaria and Ukraine during the last twenty years. The researches in this area are supported by Governments of both countries. Scientists and specialists from the Bulgarian Academy of Sciences, the Agricultural Academy of Bulgaria, National Academy of Sciences of Ukraine, and National Academy of Agricultural Sciences of Ukraine participate in this investigations.

The data of remote sensing can be used in the management only when they are combined with data from synchronous or quasi-synchronous land surface agrophysical and microclimatic measurements and purposeful soil investigations. Bulgarian and Ukrainian teams have accumulated rich experience in this collaborative research on the last years.

Development of agricultural investigations requires quantitative evaluation of the state of main energy budget elements of the soil-plant-surface air layer system on the field for successful management of soil fertility and soil pollution protection. Many of the processes are essential to soil and to nutrition movement and plant growth, such as exchange of heat, moisture and movement of contaminants depended significantly on the soil properties and canopy foliage [Nielsen et al, 1983], [Perry et al, 1988].

Application of electronic devices in field measurements has to be organized with the extreme care, because of their high cost and unique character. Emphasis is placed on the author's task, namely to carry out a characteristic evaluation of the "soil – plant – surface air layer" system's parameters are capable of being subject to electronic measurements, as well as, to propose structures for measurement systems. The measurements have to carry out in the representative places on the field.

The team of the N. Poushkarov Institute organized mobile laboratories for field measurements, included portable devices for soilmoisture and temperature, pH and soil salinity, main elements of the microclimate on the field and microcomputer agrometeorological stations. The purposeful electronic measurements enable to create a database of the heat, water, and mechanical properties of the main Bulgarian soils and to well-ground the main agroecological regions of Bulgaria.

Scientists from V.M. Glushkov Institute of Cybernetics of NAS of Ukraine developed family of portable devices "Floratest" and proper methods for express-diagnostics of plant state. The devices serve as tool for surface validating remote satellite data about vegetation.

There is a common methodology in the European countries for stratification of agricultural lands and by choice of test segments for land surface investigations complementing one another remote sensing estimation and forecasting of the crops growing.

The purpose of the chapter is to show Bulgarian and Ukrainian experience in development of measurement devices and an organization of express and continuous field electronic measurements for evaluation of main water and heat balance elements and plant state on the field.

4.2 Measurement methods and structures

Only properties of the water and heat balance elements of the soil-plant-surface-air layer system that can be transformed into corresponding electric signals are included in electronic measurements. They are underground water level; soil moisture, temperature and salinity; water potential of soils and plants; air humidity and temperature; wind velocity and direction; rainfall; solar net radiation; transpiration and evapotranspiration from the field etc.

The authors consider their main task to propose of applicable methods of approach to optimum organization and control of the field electronic measurements, both express and continuous, by using special-purpose sensors, algorithms, and structures of measuring devices and systems. There is an example of combined multi-channel measurement device for soil and leaf temperature and humidity (Figure 59).

The method of sorption sensors in the form of electrical resistance gypsum blocks is used widely based on the relationship between the electric resistance and the equilibrium moisture or matric water potential of the gypsum blocks at the particular soil suction. As a result of studies carried out in the course of several years sorption sensors for moisture of heavy Bulgarian soils are constructed and produced by Agrophysics Department of the Institute of Soil Science 'N. Poushkarov'. The soil moisture meter covers a range of 50 to 100% field capacity with 1.5% accuracy.

It is known that tensiometers have a comparatively simple construction and consist of a porous ceramic cup, connected with a manometer by means of a tube. The team established the production of ceramic cups with the highest possible water conductivity and a pore diameter of maximum about 2 micrometers. The tensiometers can measure a suction of 0 to 800 mbar. The range reflects the energy state of about 50% of available water in most Bulgarian arable lands [Kolev et al, 1986].

Our experience in soil moisture measurements includes laboratory and field evaluation of soil moisture and electrical impedance based on application of TDR meter. The principle of operation of the TDR technique, as applied to dielectric constant measurements is based on the relationship between the velocity of electromagnetic wave propagation and dielectric constant of the medium the wave propagate in [Malicki and Skierucha, 1989]. For continuous measurements on the field we use type of device with 1,5 m plastic tubes.

The instrumental evaluation of soil profile water content allows rapid determination of the irrigation rate of parcels of the agricultural field. More than 20 years working team is developing methods and devices for irrigation management based on soil moisture and temperature evaluation.

The soil properties have different values in different zones of the agricultural field. No effective method for evaluating one or other property distribution is established, but there is an opportunity to use a pattern for an ordered set of closed isolines. In other words, it is possible to consider the unhomogeneous field as numerical homogeneous zones displaying these with tree-root structure, a concept first suggested by [Uchitomi and Mine, 1988].

Figure 59. Multichannel device for field measurements

Figure 60. An IR thermometer for field measurements

A sensor consisting of two-metal electrode system joined with temperature sensible microdiode, all in a ceramic body, has been designed to monitor the soil profile salinity and temperature in the field. An alternating current digital bridge was used with sensors of soil moisture, temperature, salinity, and frost, every type with own calibration (Figure 59). Temperature measurements of the soil surface and canopy were obtained by infrared thermometer (Figure 60). A portable conductometer and pH-meter was used for evaluation of electrical conductivity and pH of the soil solution in the field conditions.

There are examples of canopy and soil surface temperature and moisture maps shown in Figure 61 and Figure 62.

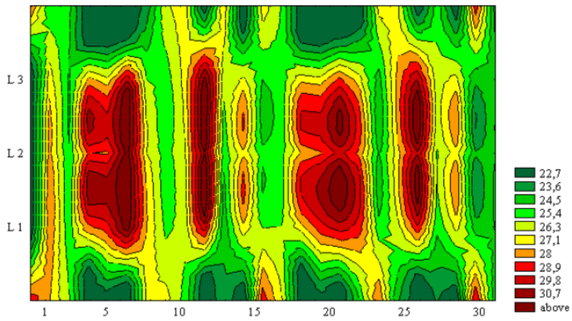


Figure 61. Cotton cannopy temperature map on 12:00, based on infrared measurements on the field

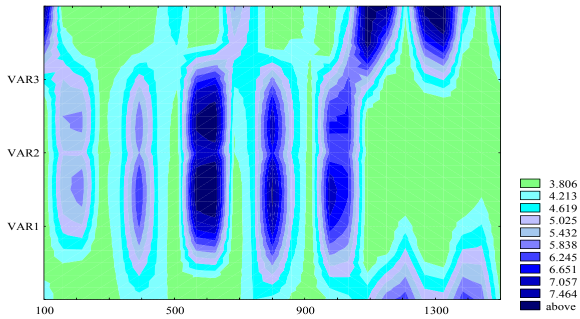


Figure 62. Soil moisture map of the soil surface of the cotton field

A portable conductometer and pH-meter was used for evaluation of electrical conductivity and pH of the soil solution in the field conditions.

Agrometeorological parameters, including heat balance elements, have been estimated using several American ARAX systems (Automatic Reporting Agricultural Weather System) [Sampey, 1986]. These types of systems use satellite measurement stations (up to 16) mounted in representative places of crop growing areas. Figure 63 shows an ARAX satellite installation.

The soil moisture and temperature sensors are placed at two depths in the crop root zone. The air temperature and humidity are measured at two levels over the canopy. In addition, a rain gauge, wind velocity, wind direction, and solar radiation sensors are situated on the satellite top. The satellite can be set up to collect the data at time intervals of 3 min to 4 hours. Frequent measurements need more power and reasonable compromise should be achieved especially in winter. ARAX satellite units transmit the data collected to the base computer station in the High VHF Band (150-174 MHz). In perfectly flat terrain, the maximum range for reception is not more 22 km.

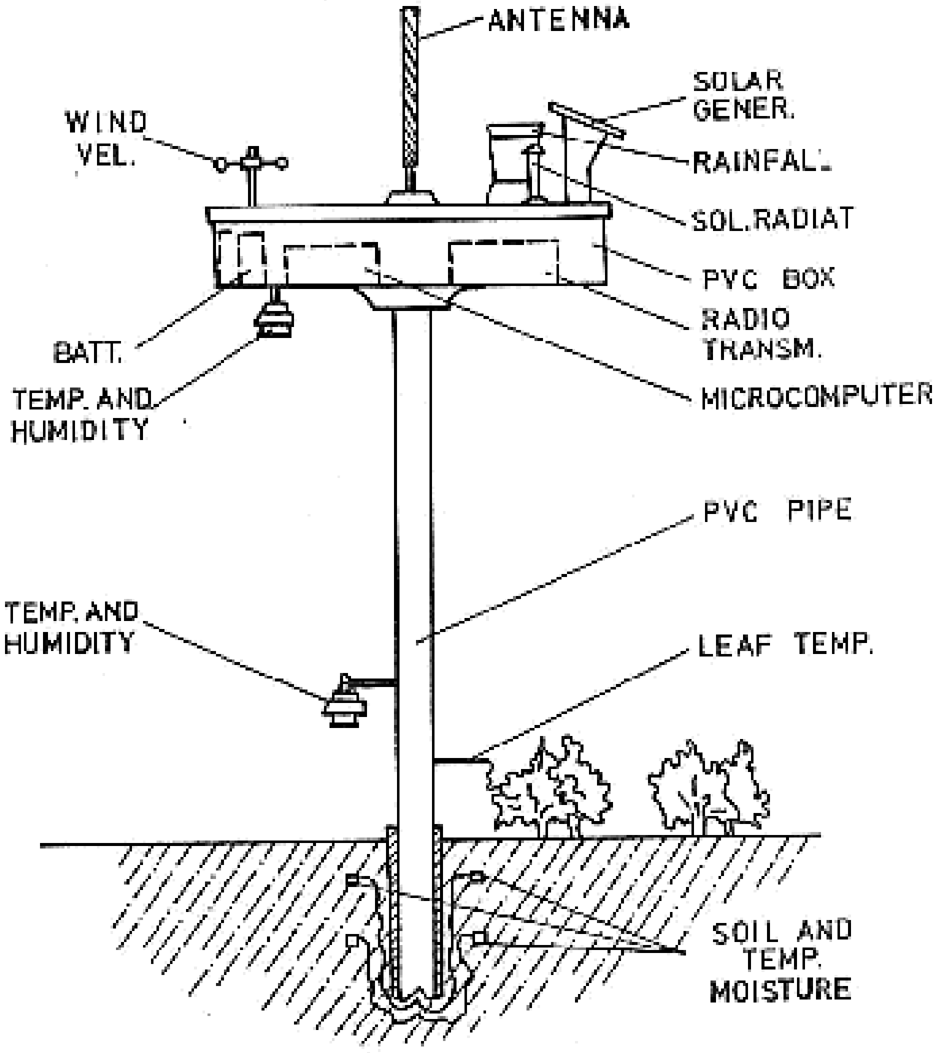


Figure 63. Structure of a satellite station of ARAX system

Reports received by the base computer are filed in memory as current weather reports. Regardless of the report period on each hour, the current weather reports from every one of the satellites are averaged or totaled as hour reports. On the same way, 24 hourly and monthly reports are filed. The basic structure of a space distributed in different agrofarms or co-operatives ARAX system is shown in Figure 64.

Registration of main heat balance elements on the field are shown in Figure 65.

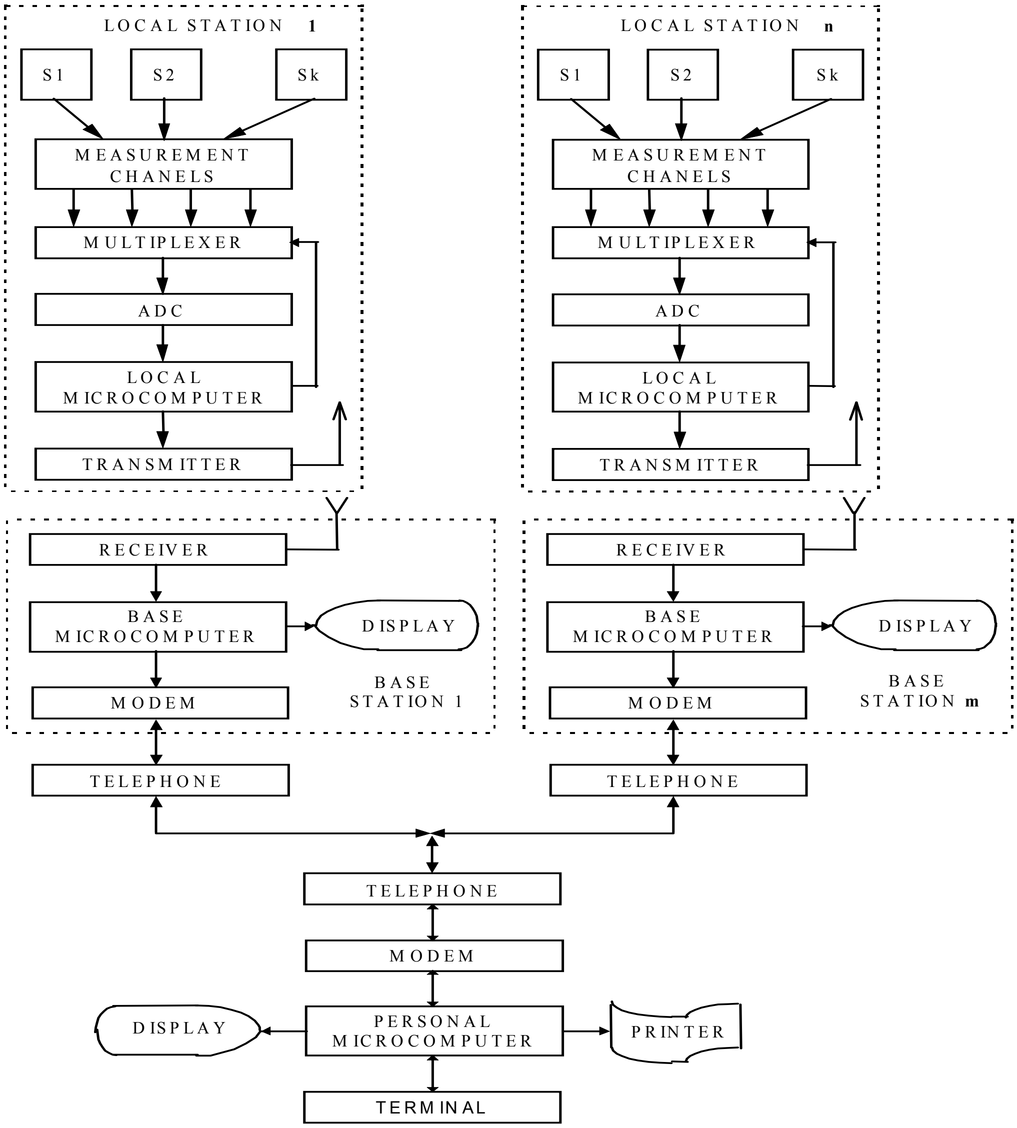


Figure 64. Block diagram of the microcomputer measurement system structure

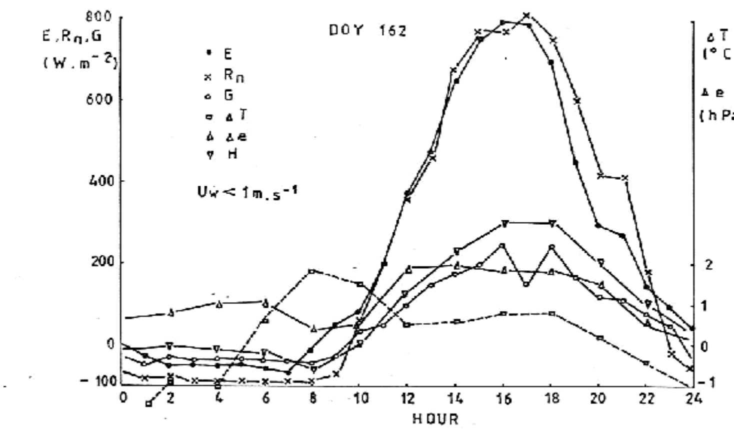


Figure 65. Diurnal curves of the evapotranspiration and other heat balance elements registered by one local station of the system in the Razgrad region

Having the data and information collected by microcomputer meteorological systems agronomists can monitor the dynamics of the processes in the field and their influence on crop growth, plant diseases, insect activity, crop water stress, etc. Direct measurements of relative humidity and temperature differences in the field allow calculation of heat balance elements and evapotranspiration. The agrometeorological systems described are very sophisticated devices designed to operate automatically all year round.

The instrumental evaluation of soil profile water content allows rapid determination of the irrigation rate of parcels of the agricultural field. More than 20 years working team is developing methods and devices for irrigation management based on soil moisture and temperature evaluation.

The soil properties have different values in different zones of the agricultural field. No effective method for evaluating one or other property distribution is established, but there is an opportunity to use a pattern for an ordered set of closed isolines. In other words, it is possible to consider the unhomogeneous field as numerical homogeneous zones displaying these with tree-root structure, a concept first suggested by [Uchitomi and Mine, 1988].

Canopy structure is another factor setting restrictions on the homogeneity of the soil properties. The soil temperature and the soil moisture, for example, differ under the plants and between the rows of row-crop canopies, and depend on the fraction of soil surface exposed to direct sunlight. The area of this fraction can be calculated based on the relationship between the sunlight transpiration, leaf area index and leaf angle distribution. The concept of this relationship can be approximated by Beer's Low.

The soil surface temperature in the regions H_i is T_S^1 and in the regions L_i is T_S^2. Then:

$$\left(T_S^{\,1}, T_S^{\,2}\right) = \varphi(f) \tag{1}$$

$$f = \exp(-k.LAI) \tag{2}$$

where f is the fraction of soil surface in direct sunlight, k is an absorption coefficient and LAI is the leaf area index. The coefficient k is a function of the canopy leaf angle distribution, random of the leaves and the sun zenith angle θ. For a given sun angle, k represents the fraction of leaf area projected onto the soil surface. The latitude of Bulgaria is 42°N, and the solar zenith angle, θ, changes from 30^{o} to 70^{o} during the active agricultural season.

For canopies where the foliage is horizontally homogeneous, the absorption coefficient is:

$$k = \int_0^{\pi/2} \frac{g(\alpha).\cos\delta}{\cos\theta}.d\alpha \tag{3}$$

where α is the angle between the leaf normal and the vertical, δ is the angle between the normal to the leaf surface and the incoming ray of sunlight, and $g(\alpha)$ is the density function of leaf normals with the vertical.

Combining eq.2 and eq.3, the fraction of soil surface in direct sunlight becomes:

$$f = \exp(-LAI\int_0^{\pi/2} \frac{g(\alpha).\cos\delta}{\cos\theta}.d\alpha) \tag{4}$$

Equation (4) defines the integral relationship between canopy structure and area of soil surface in direct sunlight. This equation is based on the assumptions of an uniform distribution of leaf azimuth angles and homogeneous positioning of the foliage in the horizontal.

To solve equations (1) and (4) for the unknowns $g(\alpha)$ and LAI, the fraction f, can be measured at various sun angles θ, and a system of integral equations be defined. The leaf angle distribution and leaf area index LAI can be estimated using methodology of direct measurement of canopy structure, described by [Perry et al, 1988].

4.3 Portable device "Floratest" for express-diagnostics of photosynthesis

In the context of the program of Presidium of National Academy of Sciences of Ukraine (NASU) "Development in the field of sensor systems and technology" in Glushkov's Institute of Cybernetics of NASU there was developed portable smart device "Floratest" for express-diagnostics of influence of stress factors on the plant's state [Romanov et al, 2007]. The portable device measures chlorophyll fluorescence induction (CFI) without plant destruction. Using the curve of CFI (alike the cardiogram) allows diagnosing influence of one or other influential factor on the plant's state.

4.3.1 Principles of device operation

As a result of external influence, different objects, including biological ones, can generate plenty of radiation that is independent of these objects temperature.

All the types of radiation that were caused by some external sources of energy are called luminescence. Duration of luminescence after external influence stopping exceeds period of light fluctuations. Luminescence is conditioned by fluctuations of relatively small number of atoms or molecules of substance that become excited under energy source activity. Radiation is a result of transformation of atoms' or molecules' states into fundamental (unexcited) or less excited (they have less energy) states.

This is well adjusted with quantum theory, according to what every stationary orbit conforms to definite value of atom's energy (Bore's postulate). Being placed on stationary orbits an electron does not radiate and does not absorb electromagnetic waves. According to the second Bore's postulate radiation and absorption can happen only when atom changes its state from one stationary state to another:

$$h\varpi_{mn} = h\nu_{mn} = E_n - E_m \quad (5)$$

where ϖ_{mn} or ν_{mn} – photon's frequency, E_m, E_n – energy values of the states m and n, h – Planck's constant, m and n – the numbers of energy states. At the same time electron switches from one stationary orbit to another.

Luminescence is defined by the structure of substance energy spectrum, the average time of staying in excited states and rules of selection, which allow absorption or radiation of light of defined frequency. Short-timed luminescence is also called fluorescence. Luminescence, which appears during lighting of substance (phosphor) with visible or ultraviolet light, is called photoluminescence. Usually process of luminescence satisfies Stocks' rule that claims that wavelength λ' of radiated light is greater than wave λ of excited light. According to the quantum theory, this means that photon's energy $h\varpi(h\nu)$ is used partially for non-optical processes:

$$h\varpi = h\varpi' + E, \varpi > \varpi' \quad (6)$$

where ϖ' – luminescence's frequency, E – energy waste on another process.

Luminescence is characterized by energy output, which equals to ratio of luminescence energy to energy that was absorbed by substance under stationary conditions.

Energy efficiency of photoluminescence increases proportionally to wave length λ of absorbed light up to the definite maximum value at $\lambda = \lambda_{\max}$ and then rapidly decreases to zero at $\lambda > \lambda_{\max}$ (Vavilov's rule). A sharp decrease of energy at $\lambda > \lambda_{\max}$ is explained by the fact that at these wavelengths λ the energy of absorbed photons is not enough for the process of phosphor atoms and molecules transfer to the excited states.

Ratio of luminescence photons number to absorbed photons with fixed energy is called quantum yield of photoluminescence. According to Vavilov's rule, which is under Stocks' rule, quantum yield of photoluminescence does not depend on wavelength of excited light and rapidly decreases for anti-Stocks radiation.

Intensity of luminescence I depends on behavior of elementary processes that causes this radiation. In case of spontaneous luminescence, when radiation starts after light absorption during which atoms or molecules are transmitted to the excited level that is placed higher than the level at which radiation takes place and then these atoms (molecules) are transmitted to the luminescence level, intensity is subordinate to exponential rule

$$I = I_0 \exp(-t/\tau) \tag{7}$$

where I – lighting intensity at the moment t, I_0 – lighting intensity in a moment of excited radiation stopping, $\tau \approx 10^{-9} - 10^{-8}\,\mathrm{sec}$ – an average duration of excited state of phosphor atoms or molecules. Luminescence of compound molecules and phosphorescence (after lighting) of organic substance are subordinate to the low (7).

Under influence of light there can be happened photochemical transformation of substance (including photosynthesis), which is called photochemical reactions. In a process of such reactions, light absorption takes place. Energy is spent on compound molecules and polyatomic ions decomposition to component parts and creation of compound molecules of primary ones. An example of photochemical reactions is decomposition carbon dioxide under influence of light

$$2CO_2 + 2h\varpi \rightarrow 2CO + O_2 \tag{8}$$

Carbon dioxide decomposition takes place in green parts of plants under sun light influence, as photochemical process, which is a part of photosynthesis.

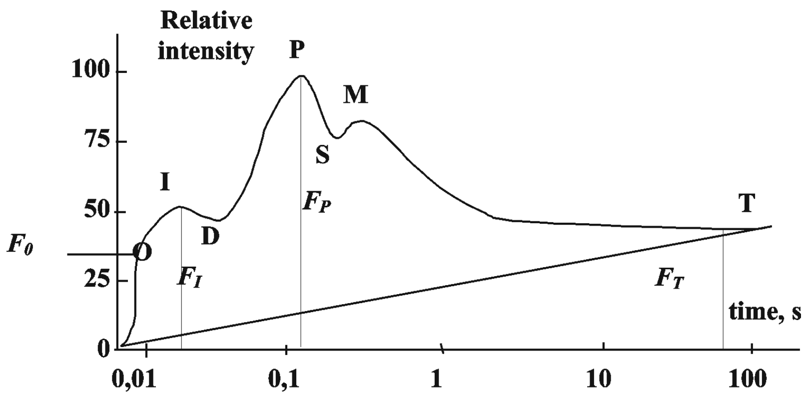


Figure 66. Chlorophyll fluorescence induction curve

One of the most important properties of the molecule of chlorophyll, which is the basic pigment of plant cell, is ability to fluoresce. For the first time this phenomenon was researched by Kautsky [Kautsky and Hirsch, 1931; 1934]. Dependence of chlorophyll fluorescence induction on time passed after start of lightning of plant's leaves is known as an induction curve or a chlorophyll fluorescence induction curve (Figure 66). The form of this curve is rather sensible to changes in the photosynthetic apparatus of plants during adaptation to different environmental conditions. This fact is a basic for extensive usage of Kautsky effect in photosynthesis research. The advantages of the method of CFI are the following: high self-descriptiveness, expressiveness, noninvasiveness, and high sensibility.

Abilities to estimate plants states using changing of Kautsky curve form are experimentally verified. Therefore, there are examples of changing form of this curve under influence external factors.

Increase of environment temperature relative to optimal for definite plant type causes decreasing of difference $F_V = F_P – F_0$. The reason is decreasing of activity of electron-transport chain or lighting activity of photosynthesis. During increasing of temperature to destruction level (45–50 °C), the level of intensity F_0 increases noticeable. It is possible to choose plants sort that are stable to high temperature influence using these parameters.

Decreasing of environments' temperature relative to optimal for definite plant type causes also decreasing of difference $F_V = F_P – F_0$, because of oppression of photochemical activity photosystem PSII. Ratio $(F_1 – F_0) / F_V$ is increasing. These features allow selecting cold-resistant plants.

Salinity of ground results in decreasing of level of F_P and F_0. Ratio $(F_P – F_0) / F_P$ is decreasing. Reason of that is oppressing of photo system PSII activity. Using these features, it is possible to choose plants that are stable ground salinity.

Water deficit results in decreasing of subtraction $F_P – F_0$ in direct proportion with decreasing of water potential of leaf. Most probable reason is slowing down of photo system PSII recovery of primary acceptor because of oppression of excretion oxygen and intersystem transport of electrons.

4.3.2 Device features

Device and relevant diagnostic methods refer to the area of biological object researches by detecting their biophysical properties, particularly native chlorophyll fluorescent induction. Device is defined as smart biosensor with fragment plant as sensing element.

Express diagnostic of plant state is carried out by functional features and is based on using of features of separate specific sections of IFC curve, which refer to separate areas of photosynthesis chains as diagnostic features. By IFC curve form, it is easily to detect influence of one or another factor on the plant state.

Figure 67. Appearance of portable device "Floratest"

Appearance of portable device "Floratest" is shown on the Figure 67.

Application areas of portable device for express-diagnostics of plant state:

- express-estimating of plant vital activity after drought, frosts, sorts coupling, pesticide introduction;
- express-detection of optimal doses of chemical fertilizers and biological additives, what lets to optimize amount of fertilizers and additives and reduce nitrates content in vegetables and fruits;

- express-detection of level of pollution of water, soil and air by pesticides, heavy metals and superpoison;
- economy of energetic and water resources during man-made watering;
- developing of precision agriculture technology for increasing the quality of agricultural products;
- using of the device in the insurance agriculture to get predicted results of future yield;
- automation of researches in the plant physiology field.

Functional diagram of the device is shown on the Figure 68. Data processing unit and displaying unit are built on the base of micro converter ADuC842 and graphical display with resolution capability of 128*64 pixels. Micro converter is system-on-chip for data acquisition and processing, which includes analog-digital and digital-analog converters, reference supplies, temperature sensor, timers, power supply monitor, embedded industry standard 8052 microcontroller, external and internal data memory, program memory etc.

Remote optical sensor is built as "reflection diagram" on the base of four light-emitting diodes and one photodetector. "Reflection diagram" means that light-emitting diodes and photodetector are situated from the same side of researched leaf. To research chlorophyll fluorescence in the red spectral region the filter is placed on the input of photodetector. Emission intensity of light-emitting diodes and photodetector sensitivity can be changed during measuring process. Integrated algorithm of device work is shown on the Figure 69.

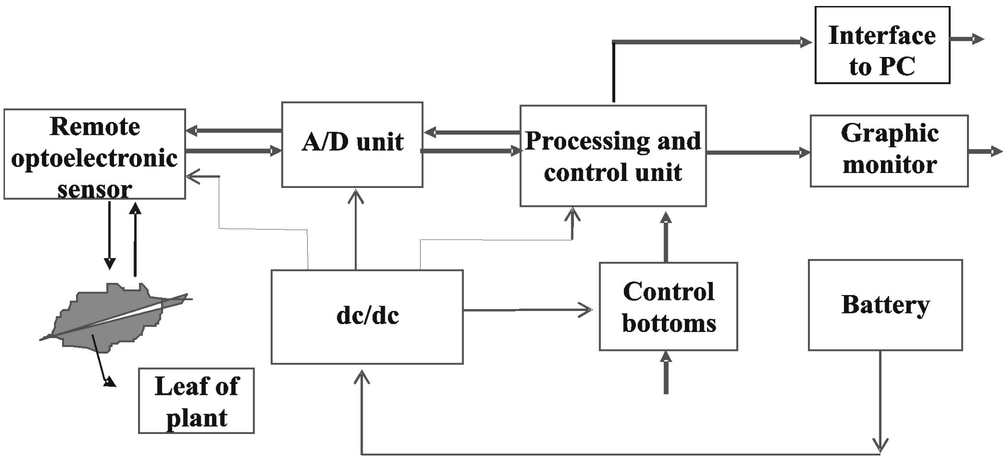


Figure 68. Functional diagram of portable device "Floratest"

Today it is not enough to acquire and save measurement result in the portable device memory. It is urgent to transmit measurement results from places of measurement to laboratories or centers of operative estimation of condition and necessary decision-making. For data transmitting from measuring channel to receiving point, it is proposed to use mobile communication by means of midget GSM-unit with GPS-subsystem, which is embedded in the portable device, and GSM-modem, which is connected to computer or workstation. During such measurements, the transmissions of a small amount of data are required, so it is reasonable to use GPRS standard. Data acquiring, processing, and transmitting system on the base of portable device with radio channel is shown on the Figure 70.

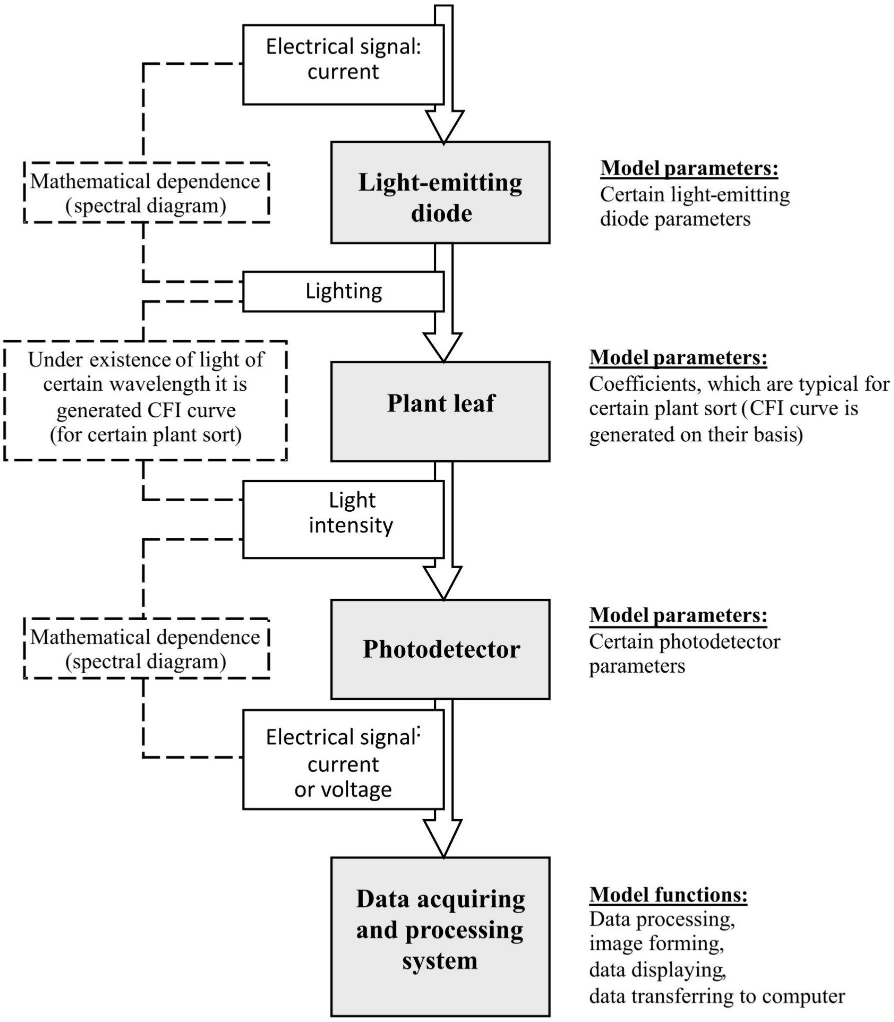


Figure 69. Integrated algorithm of device work and proper models

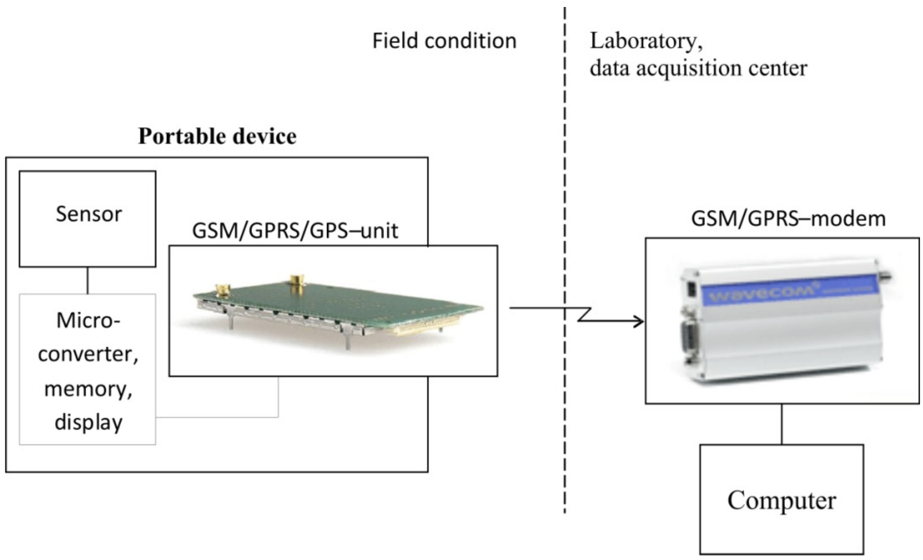


Figure 70. Data acquiring, processing and transmitting system on the base of portable device with radio channel

4.3.3 Device application

The experimental researches of the "Floratest" were conducted in National Scientific Center "V.E.Tairov's Institute of viticulture and winemaking" of Academy of Agrarian Sciences of Ukraine [Romanov et al, 2007].

Examples of the practical usage of fluorometer "Floratest" in the National Scientific Center "V.E.Tairov's Institute of viticulture and winemaking" are shown on Figure 71 and the graph of CFI on the device display are shown on Figure 72.

Figure 71. The sensor of the "Floratest" on the vine leaf

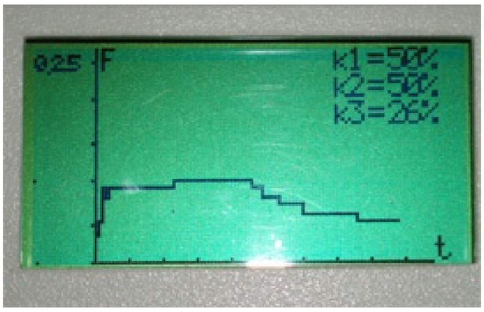


Figure 72. The image of CFI on the device's display

The conditions and results of the experimental researches are listed below.

Mature leaves of vine were used in the researches. Under changes of soil watering conditions there were observed sharp changes in behavior of induction transitions of chlorophyll fluorescence which were accompanied by quite essential changes of leaf tissue spectral characteristics.

Determination of fluorescence spectral characteristics was done by placing the device's sensor on the leaf's surface without integrity disturbance directly in a pot or in a field. It allowed to research on plastid and vacuolar pigments in their natural state and in that way approaching to understanding of the biophysical and physiology-biochemical processes which take place in the live leaf, and determination of important sides of photosynthetic activity.

Fluorescence intensity of the sample was determined in relative units.

It is significant that under natural conditions in the middle latitudes the drought is accompanied simultaneously by high temperatures of air, and that intensifies bad influence of ground water lack on agricultural plants.

Even in the first variant of experiment (drought) there appeared considerable changes of the behavior of fluorescence induction comparing to the control samples. Changes show in weakening of penetrability of the chloroplasts' membrane structures. That results in substantial increase of time characteristics of fluorescence induction slow decrease. At the same time noticeable variety differences become apparent. Sharp decrease of its value is typical for profound functional injuries of photosynthetic structures and cells of particular variety entirely.

Accordingly in this stage of drought influence, significant variety differences in exsiccate factor resistance of both photosynthetic structures and lamina's parenchymal cells entirely became apparent.

More deep changes of destructive nature may be observed in case of high temperatures (+40 °C), which influence on leaves complementary to drought. In this case, for all the varieties being studied significant and almost irreversible functional changes of plastid structures are noted. These functional changes show in sharp decrease of CFI intensity.

Disastrous changes of life activity of vine leaf cells, which take place during these processes show in oppression of biosynthetic processes, intensive decomposition of cytoplasmic structures and intensification of oxide catabolism of plant cell's content. The consequence of these processes is decrease of CFI intensity because of its oxidizing transformation.

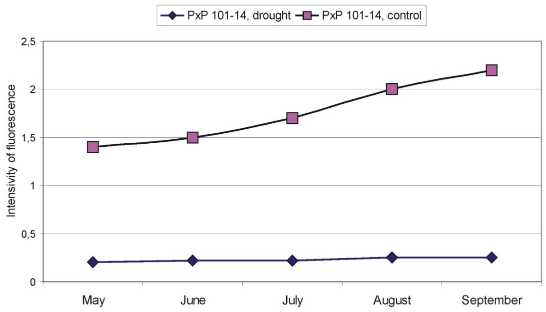


Figure 73. CFI intensity of vine plant (sort PxP 101-14) under drought and normal conditions

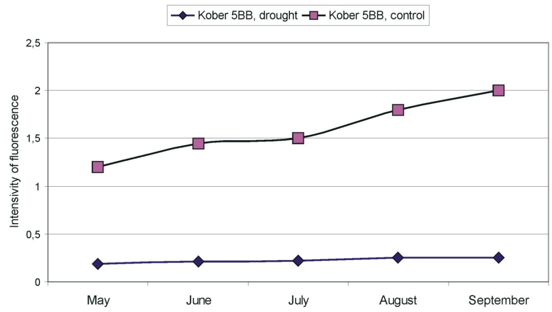


Figure 74. CFI intensity of vine plant (Kober 5BB) under drought and normal conditions

On the Figure 73 and Figure 74 there are shown the diagrams of measuring of chlorophyll fluorescence intensity for two sorts of vine plants (PxP 101-14 and Kober 5BB) during 5 months. The vine plants were under drought influence and normal conditions.

Thus, a water deficit shows up on the Kautsky curve as difference of fluorescence ($F_p – F_0$) decrease. The most credible reason of this is oppression of oxygen emission, which is related with slowing down of electrons transfer. Assuming that F_0 almost does not change for the test and control plants, in a maximum point the chlorophyll fluorescence intensity value can define the level of water deficit.

In 2008 together with Megalopolis Ecomonitoring and Biodiversity Research Center of National Academy of Sciences of Ukraine experimental researches of portable device "Floratest" were carried out to detect influence of heavy metals and harmful substances (e.g. lead, sodium, chlorine etc.) in leaf and soil on the plants state in Kiev. Today long-term phytomonitoring methods are used. They consist of visual observations and chemical analysis of soils and plant fragments, and needs complicated equipments and lasts more than one week.

Long duration and complexity of existing methods of heavy metals and harmful substances detecting in live plants and necessity to involve skilled specialists to perform these researches set necessary conditions to develop special diagnostic methods and tools for this aim. Preliminary researches indicate that portable device "Floratest" can be used for detecting of heavy metals and harmful substances influence on state of plants by measuring of CFI curve. The form of CFI curve changes versus level of harmful substances influence.

Joint researches were carried out in Kiev green regions by means of common phytomonitoring methods and portable device "Floratest". After processing of research result there were built dependences for searching correlations between chlorine content in trees' leaf, which are got by common phytomonitoring methods (Figure 75), and readouts of portable device "Floratest" (Figure 76). Even one can see some dependence between chlorine content in trees' leaf and readouts of portable device "Floratest" (stationary region of IFC curve). Calculations, made by mathematical methods, show certain correlations between these values. Such researches were made for other harmful substances, such as sodium, magnesium.

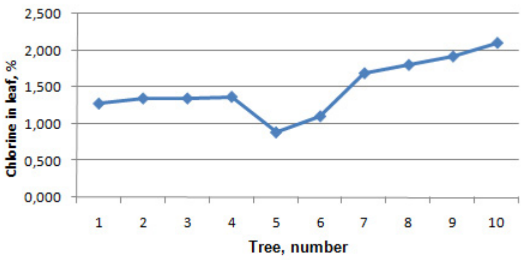


Figure 75. Chlorine content in researched trees' leaf

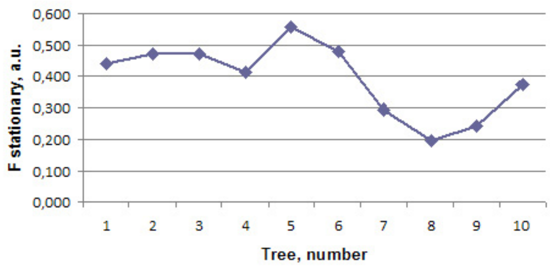
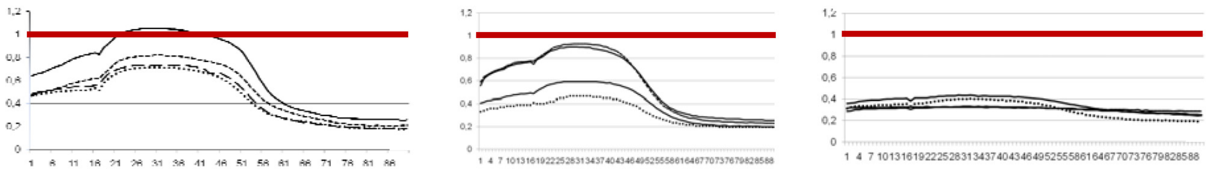


Figure 76. Readouts for researched trees

It is easily to concede that IFC curve form expresses not only contents of separate harmful substance, but general state of tree versus influence level of harmful substances.

Researches of developing smart biosensor device "Floratest" for detecting of water deficit of plants were executed in the National scientific centre "Institute of viticulture and wine-making named after V.Ye.Tairov" of National academy of agrarian sciences of Ukraine (see Figure 77).



(a) before predrying (b) after 10 hours of predrying (c) after 24 hours of predrying

Figure 77. Measuring chlorophyll fluorescence induction curve

4.4 Discussion

A way to manage electronic measurements for evaluation of soil properties at heterogeneous fields based on tree-root method and a single representative area has been presented. Gypsum blocks for soil moisture and temperature, and salinity sensors, interfaced to alternating current bridge made at "Nikola Poushkarov" Institute, proved to be useful in soil profile temperature and moisture measurements and in determination of the soil heat, water, and solution fluxes following rain or irrigation and fertilization. These measurements in single representative areas, combined with space distributed soil surface measurements done by an infrared thermometer can be widely used to evaluate the main components of the heat and water budget of the heterogeneous agricultural field.

The structure of the microcomputer measurement systems submitted above suggests a direct and correct estimation of the evapotranspiration from the system "soil-grass and trees-atmosphere", based on space distributed continuous measurements. The determination of the evapotranspiration by the energy balance method using the Bowen ratio allows for measuring only some of the heat and water balance elements. The ever-changing fluctuations in environmental conditions need a relatively large instrumental effort. Measurements carried out by microcomputer systems and analysis of the data obtained showed that such systems could be useful both for the research and for practice. The possibility of creating a large microclimatic database of the forests for long periods contributes to a careful study of these territories.

In Glushkov’s Institute of Cybernetics of NAS of Ukraine there was developed portable smart device "Floratest" for express-diagnostics of influence of stress factors on the plant's state. The portable device measures chlorophyll fluorescence induction without plant destruction. Using the curve of chlorophyll fluorescence induction (alike the cardiogram) allows diagnosing influence of one or other influential factor on the plant's state. On basis of preliminary researches there were shown that using of portable device "Floratest" let to detect in express mode the worsening of photosynthetic apparatus of plant by measuring fluorescence of native chlorophyll on the early stages. During experimental researches, there were developed methodical tools, which allow evaluating the state of vine plants under drought conditions and conditions of insufficient water capacity in express-mode.

5 Intelligent Gamma-Ray Data Processing for Environmental Monitoring

5.1 Introduction: gamma-ray spectroscopy for environmental monitoring and security

An important part of monitoring for environment and security is gamma-ray radioactivity monitoring. It extensively uses gamma-ray spectroscopy techniques. Gamma-ray spectroscopy is the science (or art) of identification and/or quantification of radionuclides by analysis of the gamma-ray energy spectrum produced in a gamma-ray spectrometer [Ortec, 2010]. It is a widely used technique with example applications in security and nuclear materials safeguards, geology and mineralogy, materials testing and reactor corrosion monitoring, industrial process monitoring, nuclear medicine and radiopharmaceuticals, health physics personnel monitoring, forensics, etc. Here we are especially interested in environmental and security-related radioactivity monitoring.

This areas use both remote and close-range gamma-ray spectroscopy techniques. Close-range gamma-ray spectroscopy analyses the spectra of individual objects or samples. Remote gamma-ray spectroscopy is used when there is no possibility of approaching or close contact with the monitored object, or when monitoring of vast territories or multiple sites is required.

Gamma-ray spectrometry is done ex-situ, i.e., by bringing the samples to laboratory, and in-situ, i.e. from aircraft, field vehicles, on foot, in boreholes, on the sea bottom, etc. Classical methods of gamma-ray spectroscopy involve taking samples into a counting room for analysis in a controlled environment and geometry. However, with laboratory analyses, there is much labor involved, and a long turnaround time for the analysis results. When the sample is spread over a wide area or is in a fixed location such as a pipeline, it is much easier and less costly to measure samples in-situ [Canberra]. These measurements, while simpler to carry out and available immediately, are prone to difficulty, usually related to calibration. Ex-situ and in-situ gamma-ray spectroscopy often complement each other.

An important part of environmental gamma-ray monitoring is gamma-ray surveys [IAEA, 2003]. Ground and airborne gamma-ray measurements may cover large areas of the earth’s surface. They often result in maps of terrestrial radiation and radioelement concentrations. Radiometric surveys and maps of natural background levels are used for solving geological and environmental problems in a number of fields, including mineral exploration or prospecting, geochemical mapping, soil mapping for agriculture, environmental assessment, etc. In addition to geoscience, it can be also applied in emergent situations to monitor environmental contamination in the event of an accidental release or major nuclear event, including radioactive spill from nuclear reactors or during transportation and storage of radioactive wastes, accidental loss of industrial and medical radiation sources, etc. as well as nuclear weapons incidents or tests.

Gamma-ray surveys provide a baseline against which man-made contamination can be estimated. This makes the radiological assessment of the environment possible, showing general regional trends in radionuclide distribution, etc. They may be used to estimate and assess the terrestrial radiation dose to the human population and to identify areas of potential natural radiation hazard, to determine the effectiveness of cleanup efforts after a spill or an industrial accident, to trace and to clean up old contamination, and to verify that new contamination is not occurring. For example, surveys are regularly carried out in the radiation risk zones around facilities where radioactive materials are used or produced, such as nuclear power plants and other facilities dealing with nuclear fuel cycle, radioactive waste and material repositories and their transportation routes, science, industrial, and medicine sites, etc. Similar surveys are used to assess the contamination in areas of former mining. Besides territories, of interest is gamma-ray monitoring of objects that use radioactive materials or fuel, contaminated buildings, containerized waste, industrial processes, laboratories, etc., as well as moving objects (vehicles, movable facilities, cargo, people, etc.).

Gamma-ray monitoring is also closely related to security, which in this context is mainly associated with the dangers of nuclear proliferation and, more recently, radiological and nuclear terrorism. Nuclear terrorism denoting the use, or threat of the use, of nuclear or radiological weapons ("dirty bomb") in acts of terrorism is nowadays becoming an immediate challenge for the entire world [Ruff, 2006], [Markman et al, 2003]. The challenges of coping with the consequences of possible acts of nuclear terrorism or of the detection of radiological or nuclear materials in a planted device are closely related to the issues of environmental gamma-ray monitoring and spectroscopy mentioned above.

This also concerns the challenges of pre-empting nuclear and radiological terrorism that requires detection, interdiction, and identification of smuggled nuclear materials, as well as preventing or intercepting the illicit trafficking, theft and loss of nuclear and other radioactive materials and weapons. Monitoring is required at the borders and ports of entry, including land, rail, marine, or air border crossings, as well as the facilities where radioactive materials and weapons are produced or stored. To detect illicit transport of radioactive material in vehicles, people, luggage, mail, cargo, containers, etc., radiation portal monitors are commonly used.

All those mentioned pressing challenges and applications require ever-enhancing accuracy, precision, stability, and reliability of gamma-ray monitoring. In its turn, this requires novel gamma-ray spectroscopic techniques to employ state-of-the-art approaches and methods of intelligent data processing and machine learning. In this chapter, we discuss gamma-ray spectrometry techniques for both close-range and remote monitoring. First, we consider gamma-ray spectroscopy with multichannel spectrometer both for reproducible and unknown geometry of measurements. Then, techniques for airborne gamma-ray spectrometry are discussed. Results of experimental comparison are provided to illustrate the performance of the employed machine learning techniques.

5.2 A machine learning approach to multi-channel gamma-ray spectroscopy

Gamma-rays are the most penetrating radiation from natural and manmade sources. Radionuclides emit gamma-rays of specific energies characteristic for elements and isotopes. Gamma-ray measurements are conducted in two modes. Total count measurements register gamma-rays of all energies and are used to monitor the gross level of the gamma radiation field and to detect the presence of anomalous sources. Gamma-ray spectrometers, on the other hand, measure both the intensity and energy of radiation, thus enabling the source of the radiation to be diagnosed [IAEA, 2003].

In order to interpret spectrometer measurements in real-world conditions, a number of intelligent data processing techniques have been developed or adapted. In this section, we review available techniques, as well as machine learning approaches such as model selection, sparse approximation, and blind source separation that we develop and/or apply to gamma-ray spectroscopy. Performance of those techniques is illustrated using artificial and real gamma-spectrum data.

5.2.1 Gamma-ray spectroscopy for fixed and non-fixed geometry of measurements

Gamma-ray spectroscopy is the quantitative study of the energy spectra of gamma-ray sources. The result of the gamma-ray spectrum measurement by a multi-channel spectrometer (Figure 78) may be considered as L numbers z_i (i=1,...,L) each representing the count $f(z_i)$ of gamma-rays of particular energy z_i. This output signal of the spectrometer can be decomposed as

$$f(z_i) = \sum_{j=1,\dots,N} w_j \psi_j(z_i) + \varepsilon(z_i), \quad (1)$$

where z_i is the gamma-ray energy in the i-th channel, ψ_j is the detector response function, w_j is its weight, N is the number of detector response functions in the decomposition, $\varepsilon(z_i)$ is the noise realization.

Figure 78. Multi-channel spectrometer "Vector"

The detector response function $\boldsymbol{\psi}(z)$ (Figure 79) is the spectrometer output when the input is gamma-rays having the energy z. Since a multi-channel spectrometer should be able to deal with the gamma-ray sources of any energy from its working range, one needs to have detector response functions for all z_i (i=1,...,L). However, real monochromatic gamma-ray sources do not exist for much energy. Therefore, the detector response functions are usually generated using Monte-Carlo simulations of gamma-ray interactions with the detector crystal [Zerby, 1963], [Breismeister, 2000], [Hendriks et al, 2002], [Sempau et al, 2003].

The task is, using the results of spectrum measurement $f(z_i)$ and knowledge of the detector response function set corresponding to possible radionuclides, to estimate non-zero w_j (weights or activities) and thereby to identify radionuclides with their activities that produced the measured spectrum (Figure 80).

Gamma-ray spectrometry is considered in fixed and non-fixed geometries of measurements. For measurements in a **fixed geometry**, position of the detector relative to the radiation source and geometry of the source are known; if the radiation source is shaded by a substance that absorbs radiation, the geometry and nature of this screen are known as well. So, detector response functions can be generated for this known geometry. Since the maximum number N of detector response functions is equal to the number of channels L, and each function has its value in each of the L channels, the total set of detector response functions forms the L×L matrix.

Gamma-ray spectrum measurements in a fixed geometry are used to estimate the radionuclide content in the samples of food, soil, construction materials, etc., often using the Marinelli beaker.

The task is complicated by fluctuations in the natural radiation background, low levels of radionuclide activity, a complex spectral composition of radiation.

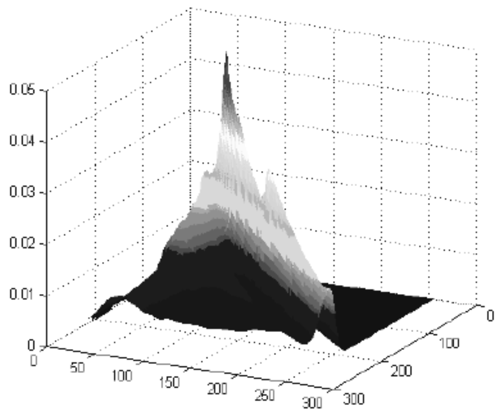


Figure 79. ***Detector response functions (200 functions of 256 channels each***

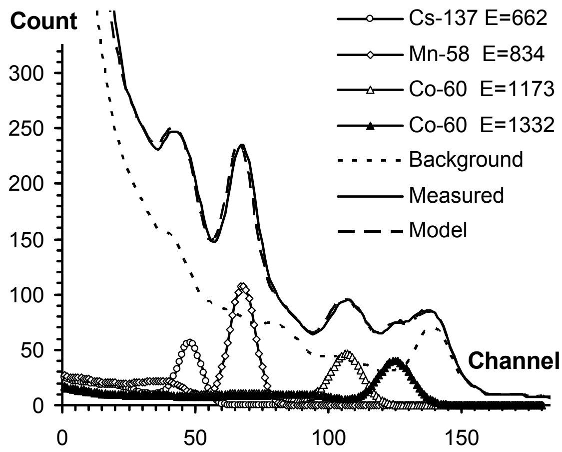


Figure 80. Spectrum decomposition using detector response functions

Measurements in a **non-fixed geometry** (unknown and complex) have the following peculiarities: position of the detector relative to the radiation source and the geometry of the source are unknown; the radiation source is shaded by a substance that absorbs radiation, geometry and nature of this screen are unknown.

So, the detector response functions for a non-fixed geometry of measurements cannot be generated in advance, and the decomposition of Eq(1) cannot be immediately applied to this case. To cope with this problem, we proposed [Zabulonov et al, 2004a] to split each detector response function into three separate functions $\boldsymbol{\psi}^A(z)$, $\boldsymbol{\psi}^C(z)$, $\boldsymbol{\psi}^R(z)$ corresponding to three characteristic spectrum regions: the total absorption peak, the Compton part and the reverse flight peak. So, each detector response function is represented as

$$\boldsymbol{\psi}(z) = w^A\, \boldsymbol{\psi}^A(z) + w^C\, \boldsymbol{\psi}^C(z) + w^R\, \boldsymbol{\psi}^R(z). \qquad (2)$$

Therefore, for this case the maximum set of detector response functions becomes ψ_{js}, j = 1,…,L, s={A,C,R} and those functions form the L×3L matrix.

Gamma-spectroscopy in a non-fixed geometry of measurements often occurs in practice, since placing the assessed environmental objects into the fixed geometry conditions is often difficult or impossible. It concerns radiation assessment of cargo, vehicles, facilities, field sources, as well as gamma-ray surveys, etc.

5.2.2 Previous gamma-ray spectroscopic methods

Some traditional approaches to gamma-ray spectroscopy are shown in Figure 81.

The traditional Regions-of-interest (ROI) method and its modifications [Desbarats and Killeen, 1990] make spectrum analysis in certain regions of interest. Usually the number N of those regions is a small fraction of L. At the calibration stage, the "contribution matrix" is calculated. Its columns correspond to the expected radionuclides, rows correspond to ROIs, and elements are the gamma-ray counts registered in the appropriate ROI from the expected radionuclides.

The measured spectrum is processed as follows. All ROIs are checked, starting from the one having the maximum energy. If some ROI has count value above the threshold, the presence of the corresponding nuclide is reported and its contribution is subtracted from the total spectrum, weighted by the relevant elements of the contribution matrix from all ROIs. This is repeated for all ROIs, providing the radionuclides with their activities. The stripping method [Desbarats and Killeen, 1990] uses the detector response functions of expected radionuclides instead of the gamma-ray counts in the ROIs.

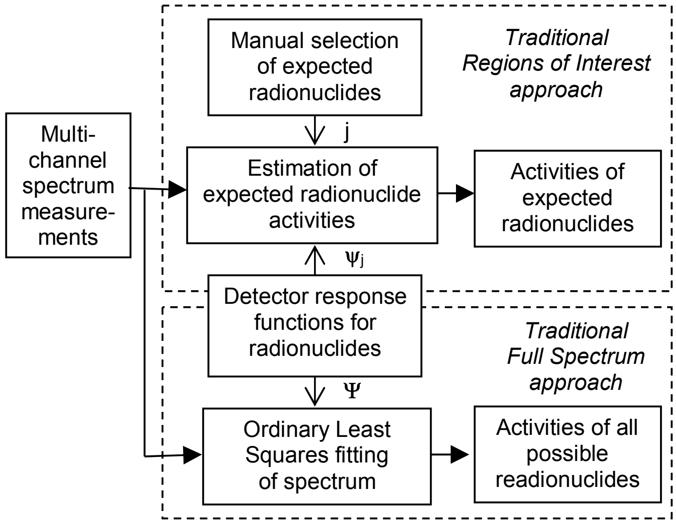


Figure 81. Regions-of-Interest and Full Spectrum approaches to gamma-ray spectroscopy

Thus, traditional methods require a-priori knowledge of radionuclide composition in order to estimate radionuclide activities and are not intended to assess samples of arbitrary composition. In order to overcome this drawback, full spectrum processing techniques were proposed [Minty, 1992], [Hendriks et al, 2001], [Guoa et al, 2004], [De Meijer, 2007], [Newman et al, 2008]. Here, the total matrix of detector response functions Ψ is used, and the activity vector **w** is obtained by the Ordinary Least Squares (OLS) solution:

$$\mathbf{w} = \Psi^{+} \mathbf{f}, \qquad (3)$$

where Ψ^{+} is the pseudoinverse matrix.

However, due to the noise, this method assigns non-zero weights to nuclides actually not present in the spectrum. To increase noise resistance, maximum noise fraction (MNF) [Dickson and Taylor, 1998; 2000] and noise-adjusted singular value decomposition (NASVD) [Minty and McFadden, 1998], [Minty and Hovgaard, 2002], [Mauring and Smethurts, 2005] are sometimes used. Regretfully, those techniques require information about noise.

For a non-fixed geometry of measurements, performance of the traditional methods is even poorer than for a fixed geometry. If one uses Eq.(1), detector response functions should be known for all possible geometries and their proper set should be chosen in-situ, that cannot be implemented in practice. If one considers Eq.(2), the drawbacks of the traditional methods described above become more severe since the number of weights (activities) to be estimated grows.

Another approach to the non-fixed geometry spectroscopy consists in the following. Some hypotheses about the composition and activity of nuclides in the measured spectrum are set up. Then the Monte Carlo simulation of photon transport is done for those hypotheses, taking into account the particular geometry of measurements. If the measured spectrum is exactly reconstructed by the simulation, the desired composition and activity are found [Gal et al, 2001], [Toubon et al, 2006] This approach has a large computational complexity and so cannot provide real-time results, since a lot of hypotheses should be verified in case of poor a-priori information.

5.2.3 Intelligent data processing methods for spectrometry

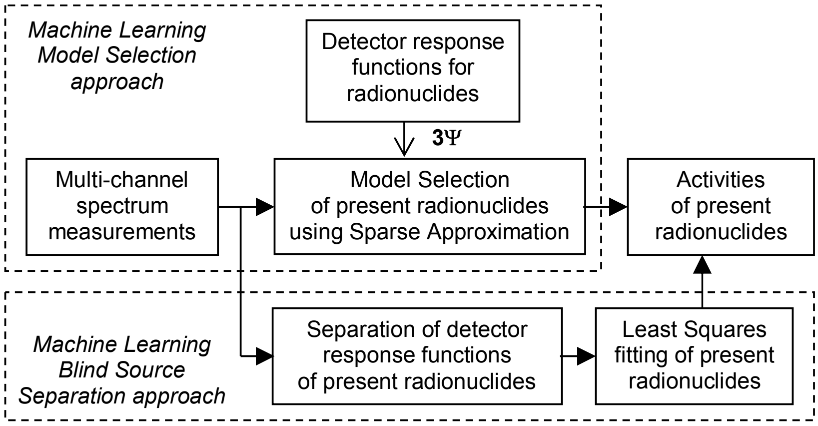


Figure 82. Gamma-ray spectroscopy by machine learning

To improve the accuracy and stability of radionuclide activity estimation, we develop an approach to full spectrum processing using machine learning methods of model selection, sparse approximation, and blind source separation (Figure 82).

✓ Model selection and sparse approximation

Model selection is a problem arising in the areas of machine learning and data mining. Given data usually consisting of input-output pairs, a model is built to relate the input and the output. A number of approaches have been proposed to build the model, including linear models, neural networks, classification and regression trees, kernel methods, etc. The task of model selection is to choose a close-to-optimal model in terms of the particular model usage and application.

A number of model selection methods use various model selection criteria [Akaike, 1974], [Mallows, 1973], [Hansen and Yu, 2001]. Model selection criteria are formulated in such a way that they balance the model complexity vs. the approximation error, and so automatically reduce the model complexity with increasing noise levels. At the optimal models, criteria reach minimum. Model selection criteria have been developed based on various assumptions, i.e. using predictive training error [Mallows, 1973]; generalization error [Sugiyama and Ogawa, 2001], [Niyogi and Girosi, 1996] and related criteria, such as based on information statistics [Akaike, 1974], description length [Rissanen, 1978; 2002], etc.

Let the data set D_L be represented by L pairs $D_L=\{(\mathbf{x}_i,y_i)\}_{i=1,\dots,L}$, $\mathbf{X}\in\Re^{L\times N}$, $y_i=y_{0i}+\varepsilon_i$, ε is Gaussian additive noise. The predictive training error is adopted as the error measure [Sugiyama and Ogawa, 2001] between estimated and true values at sample points contained in the training set:

$$I_{PTE} = E_\varepsilon||\mathbf{By}-\mathbf{y}_0||^2 = ||\mathbf{By}_0-\mathbf{y}_0||^2 + \mathrm{tr}(\mathbf{BQB}^T), \quad (4)$$

where $\mathbf{B}=\mathbf{X}_s(\mathbf{X}_s^T\mathbf{X}_s)^{-1}\mathbf{X}_s^T$ is the mapping from $\mathbf{y}$ to its estimation by the model, $\mathbf{Q}$ is the noise covariance matrix, $\mathbf{X}_s$ is the data for the model corresponding to $s\leq N$ non-zero components of $\mathbf{w}$ corresponding to the model, E_ε denotes the ensemble average over the noise, $||.||$ denotes the norm, tr(·) denotes the trace of an operator.

The Mallows criterion C_P [Mallows, 1973] gives an unbiased estimate of the predictive training error:

$$C_P = RSS + 2\sigma^2 s. \quad (5)$$

where RSS is the residual sum of squares, σ^2 is the noise variance.

Generalization error is the difference between the expected risk of the obtained solution $f_{H,L}$ and the expected risk of the "optimal" (true) solution f_0 [Niyogi and Girosi, 1996]:

$$I_G=I(f_{H,L})-I(f_0),\ f_{H,L}=\text{argmin}_{f\in H}\, I_{emp}(f),\ I_{emp}(f)=1/L \sum_{i=1,L}V(y_i, f(x_i)), \quad (6)$$

where $f_{H,L}$ is the solution obtained using the limited data set of length L and the model H with limited parametrization, I_{emp} is the empirical risk, $V(\cdot)$ is the loss function.

Subspace Information Criterion SIC gives an unbiased estimate of the generalization error [Sugiyama and Ogawa, 2001]. For least mean squares learning,

$$SIC = [\langle \mathbf{Ty},\mathbf{y}\rangle - \sigma^2 \text{tr}(\mathbf{T})]_+ + \sigma^2 \text{tr}(\mathbf{T}_S^+),\ \mathbf{T} = \mathbf{T}_S^+ - \mathbf{T}_F^+\mathbf{T}_S\mathbf{T}_S^+ - \mathbf{T}_S^+\mathbf{T}_S\mathbf{T}_F^+ + \mathbf{T}_F^+, \quad (7)$$

where $(\cdot)^+$ denotes the matrix pseudoinverse, $\mathbf{T}_S$ and $\mathbf{T}_F$ are the L×L matrices with the elements obtained by calculation of kernel function values K(x,x') for the selected $\mathbf{X}_s$ and for the whole $\mathbf{X}$ respectively, $[\cdot]_+$ is defined as $[t]_+ = \max(0,t)$.

Akaike's information criterion [Akaike, 1974] evaluates the generalization error from the information statistics point of view. It is calculated as

$$AIC = 2s + L \ln RSS, \quad (8)$$

where L is the number of observations, s is the number of parameters.

Minimum description length criteria for regression can be written as

$$I_{MDL} = DL(\mathbf{y}|\mathbf{X}_s) + DL(s), \quad (9)$$

where $DL(\mathbf{y}|\mathbf{X}_s)$ is the description length of $\mathbf{y}$ by the model of complexity s, DL(s) is the description length of the model.

In [Hansen and Yu, 2001] expressions for LyMDL and gMDL criteria were proposed, whose minima correspond to the minimum description length criterion:

$$\begin{gathered} LyMDL = 0.5(\mathbf{y}^T\mathbf{y} - FSS)/\sigma^2 + 0.5\, s_1\, [1+ \log(FSS/ (\sigma^2 s))]+ 0.5\log(L) \text{ for } FSS>s\sigma^2, \\ L(\mathbf{y}) = 0.5(\mathbf{y}^T\mathbf{y})\, \sigma^2 \text{ otherwise,} \end{gathered} \quad (10)$$

$$\begin{gathered} gMDL = 0.5L \log(RSS/(L-s_1)) + 0.5s \log(F) + \log(L) \text{ for } R^2 \ge s/L, \\ gMDL = 0.5L \log (\mathbf{y}^T\mathbf{y}/L) + 0.5 \log(L) \text{ otherwise,} \end{gathered} \quad (11)$$

where $FSS = \mathbf{y}^T\mathbf{X}_S(\mathbf{X}_S^T\mathbf{X}_S)^{-1}\, \mathbf{X}_S^T\mathbf{y}$, $F= (L-s)(\mathbf{y}^T\mathbf{y}-RSS)/ (s\ RSS)$, R is the usual multiple correlation coefficient.

For the spectroscopy task, the model selection approach is assumed to avoid the situation when the model includes elements actually not present. In order to use the approach, we must generate the models and choose the one with the minimal model selection criterion value. Since exhaustive search is impossible for practical settings, we use greedy techniques based on the ideas of sparse approximation [DeVore, 1998], [Donoho et al, 2004].

The field of sparse approximation is concerned with the problem of representing a target vector using a short linear combination of vectors drawn from a large, fixed collection called a dictionary This problem arises in applications as diverse as machine learning, signal and image processing, imaging sciences, communications, numerical analysis and statistics. It is known that sparse approximation is computationally hard in the general case. Nevertheless, it has recently been established that there are tractable algorithms for solving these problems in a variety of interesting cases. It was also shown that, under certain conditions, sparse approximation and Support Vector Machine technique proposed by V. Vapnik give the same solution [Vapnik and Chervonenkis, 1970].

Sparse approximation methods are used to solve $\Psi\mathbf{w} = \mathbf{y}_0$ for the complete basis Ψ (i.e. the basis that exactly represents $\mathbf{y}_0$) with the maximum sparseness in terms of the l_0 norm ($\mathbf{w}^*=\min_\theta||\mathbf{w}||_0$). They include step-wise regression [Seber, 1977], k-term approximation [Temlyakov, 2003], matching pursuit [Mallat and Zhang, 1993], [Tropp and Gilbert, 2007].

Matching pursuit at the (k+1)-th step calculates the approximation of $\mathbf{y}$, $\mathbf{f}_{k+1}=\mathbf{f}_k+w_{k+1}\psi_{k+1}$ choosing ψ_{k+1} ($\psi_{k+1}\subset\Psi$ without Ψ_k) and w_{k+1} that minimize the residual sum squared $||\mathbf{R}_k-w\psi||^2$:

$$(w_{k+1}, \psi_{k+1}) = \arg\min_{w,\psi} ||\mathbf{R}_k-w\psi||^2,\ \mathbf{R}_k=\mathbf{y}-\mathbf{f}_k. \quad (12)$$

The drawback of matching pursuit is that its application to noisy $\mathbf{y}$ requires stopping criteria.

We proposed to use model selection criteria as stopping criteria [Revunova, 2007b]. It works as follows.

Initialize $\mathbf{f}_0=0$, $\mathbf{R}_k=\mathbf{y}$.

Find the basis function (column ψ) of $\mathbf{X}$ that maximizes the scalar product with the current residual $\mathbf{R}_k$:

$$\gamma_k = \arg\max_{i=1,\ldots,N} |\langle\mathbf{X}(\cdot,i),\mathbf{R}_k\rangle|. \quad (13)$$

In order to get a more accurate solution proposed in the framework of orthogonal matching pursuit [Tropp and Gilbert, 2007], recalculate the solution for the already chosen basis functions Ψ_k by performing the OLS fitting:

$$\mathbf{w}_k=\Psi_k^+\,\mathbf{y}. \quad (14)$$

Calculate a new residual:

$$\mathbf{R}_{k+1} = \mathbf{y} - \mathbf{w}_k\,\Psi_k. \quad (15)$$

Calculate the value of the used model selection criterion CR(k). If CR(k)≥CR(k–1), select the model with s=k–1 as optimal. Otherwise, go to Eq.(13) and repeat the cycle.

Let us compare traditional methods and our model selection approach for spectroscopy in a fixed and non-fixed geometry of measurements [Revunova, 2008], [Zabulonov et al, 2004b], [Zabulonov et al, 2009a]. In our approach, Ψ contains detector response functions, and $\mathbf{y}$ is the result of spectrum measurement.

For a fixed geometry of measurements, we compared the relative error of the traditional region of interest (ROI) and stripping (STR) methods vs. orthogonal matching pursuit with the stopping by the model selection criterion (MPCR). Fixed geometry of measurements was ensured by placing the test samples in a standard 0.5l Marinelli beaker and using a detector with a collimator. Traditional ROI and STR methods were set up for processing of ^{137}Cs (Cesium) and ^{40}K (Potassium). Without changing this setup, an additional nuclide ^{230}Th (Thorium) or two additional nuclides ^{230}Th, ^{226}Ra (Radium) were then added to the sample.

The relative error was estimated for the radionuclide with the lowest energy ^{137}Cs, since its spectrum is "hidden" by the other radionuclides. Model selection criteria of Mallows, Akaike, Bin Yu were used with MPCR. Slightly better results were obtained for the Mallows criterion Cp. The experimental results are given in Table 3.

Table 3. Relative error of ^{137}Cs activity in the multi-nuclide sample obtained by ROI, STR, MPCR

Nuclides	Cs, K	Cs, K, Th	Cs, K, Th, Ra
ROI	0.1%	11%	19%
STR	0.5%	5%	12%
MPCR	0.35%	0.94%	3%

When processing the spectrum of known composition (^{137}Cs, ^{40}K), the ROI method provides the best relative error, as it was set up precisely for those nuclides and is influenced by noise only inside its regions of interest. MPCR provides a smaller error compared to STR. When processing the spectrum with additional radionuclides (Th or Th, Ra) unknown to the ROI and STR methods, MPCR provides the smallest relative error measuring of the ^{137}Cs activity.

For the first experiment in a non-fixed geometry of measurements, the samples including ^{60}Co(Cobalt) (2 spectrum lines), ^{137}Cs, ^{134}Cs, ^{58}Mn (Manganese) have been simulated by their 3-part detector response functions Eq.(2) (Figure 83). The weight of the reverse flight peak was 0 and the weight of the total absorption peak was 1 for all radionuclides. The weight of the Compton part was 2 for ^{137}Cs and ^{134}Cs and 1 for the remaining radionuclides. This simulates the situation when ^{137}Cs and ^{134}Cs which fell on the soil surface gradually seep into its lower levels and thus the gamma-rays from the total absorption peak redistribute to the Compton part, so that their coefficients in Eq.(2) become different. The reverse flight peak in this situation is practically not observed and so its coefficient is set to zero.

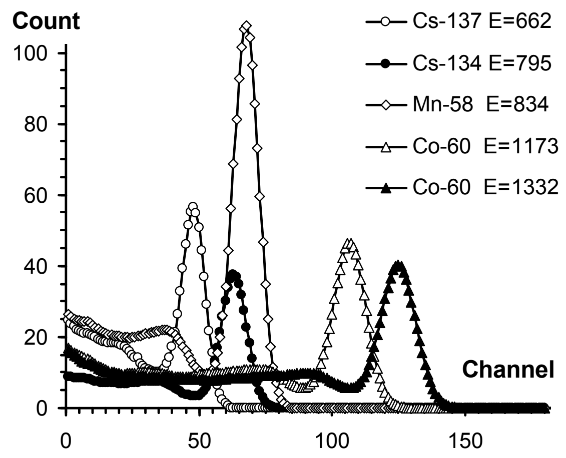


Figure 83. Detector response functions of the true model

So the number of non-zero detector response functions was 5×2=10. Then the full spectrum was simulated by the linear model Eq.(1), where the activities of radionuclides were set to w(^{137}Cs)=3, w(^{134}Cs)=4.5, w(^{58}Mn)=4, w(^{60}Co)=7. At last, a realization of an additive noise with the Gaussian distribution was added to the spectrum. To simulate unexpected radionuclides, 15×2=30 additional detector response functions ψ were added to Ψ, so the total number of basic functions in Ψ was N=40. The number of spectrometer channels L was 512. The tested methods used to estimate the radionuclide activities were: OLS with all basis functions (FULL), OLS with the detector response functions actually in the model (TRUE), MPCR using model selection criteria of Mallows C_P, Akaike AIC, BinYu L_Y (MDL). We compared the errors calculated as the l_2-norm of the difference between the true activity vector and the estimated activity vector at different noise levels. The results in Figure 84 show that MPCR for all model selection criteria and all noise levels performs better than FULL, but worse than TRUE. The error for MPCR with MDL is somewhat less than that for C_P and AIC.

The next experiment was conducted using the really measured spectra of a point gamma radiation source on the background of a distributed source with a higher energy. ROI, STR, and MPCR with the MDL model selection criterion were compared. In the experiment, the point source was ^{137}Cs and the distributed source was ^{40}K; the ^{40}K activity was 51% of the ^{137}Cs activity. ROI and STR were set up for ^{137}Cs and ^{40}K. The relative error of the ^{137}Cs activity was 92% for ROI, 85% for STR, and 12% for MPCR.

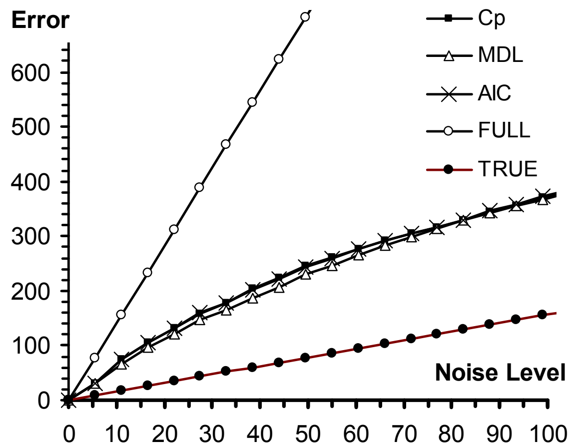


Figure 84. Radionuclide activity error vs noise level for various models and model selection criteria

✓ Model optimality tests

In order to determine why different model selection criteria vary in accuracy, we compared [Revunova, 2008], [Zabulonov et al, 2004a], [Zabulonov et al, 2009a] the dependence of the model dimensionality and the model error on the noise level. The comparison showed that the best accuracy is provided by the criterion that keeps the true dimension of the model at the higher noise levels than the other model selection criteria.

In [Gribonval et al, 2006] a test was proposed for checking if the particular set of basic functions in the model is true. It is based on the notion of l_0-optimal solution that provides both the minimum approximation error and the maximum model sparsity. Experiments showed [Revunova, 2008] that with the increasing noise levels the error using such a test is smaller than that of the traditional model selection criteria [Revunova, 2005a], [Revunova and Rachkovskij, 2005], [Zabulonov et al, 2005].

The drawback of this model optimality test is that it can only be used for some bases with valid "basis connectivity" condition [Gribonval et al, 2006]. So, research on how to solve this problem for certain types of detector response basis functions could lead to further improvement in the accuracy of radionuclide activity estimation under noisy conditions.

✓ Blind source separation

Blind source separation (BSS) is the separation of a set of signals from a set of mixed signals, without the aid of information (or with very little information) about the source signals or the mixing process. In machine learning, BSS is considered as one of the most powerful unsupervised learning techniques. Unsupervised learning is distinguished from supervised learning and reinforcement learning in that the learner is given only unlabeled examples. Theoretical foundations of the BSS methods are well developed [Chichocki and Amari, 2002], [Comon, 1994], [Hyvarinen, 1999].

The most straightforward case is linear BSS with the number of observed mixtures equal to the number of unknown sources. Here the data D_L are represented by the set of L mixture vectors $\mathbf{y}_i \in \Re^N$: $D_L=\{\mathbf{y}_i\}_{i=1,..,L}$. The mixture vector **y** depends on the source vector **x** as a linear model **y**=**Ax**, where the full rank mixture matrix $\mathbf{A}\in\Re^{N\times N}$ is unknown. The task is to obtain L unobserved vectors $\mathbf{x}$ (matrix $\mathbf{X}\in\Re^{L\times N}$) using L known vectors $\mathbf{y}$ (matrix $\mathbf{Y}\in\Re^{L\times N}$) by estimating the un-mixture matrix $\mathbf{B}=(\mathbf{b}_1,\ldots,\mathbf{b}_N)^T\in\Re^{N\times N}$ and then calculating $\mathbf{X}=\mathbf{YB}$.

Well-known approaches to solution of the BSS problem use a priori information about the sources. The sources are supposed to be statistically independent and having nongaussian distribution (independent component analysis ICA [Amari and Cichocki, 1998], [Amari et al, 2002]), or uncorrelated and having the Gaussian distribution (principal component analysis PCA [Chichocki and Amari, 2002]), or sparse (sparse component analysis SCA [Li et al, 2004]), etc. Based on those assumptions, a cost function is constructed (e.g., mutual information of sources) and then minimized by the optimization algorithms, providing the solution.

Usually, BSS-related methods such as MNF [Dickson and Taylor, 1998; 2000] are used in spectroscopy for noise suppression. We propose using BSS in spectrometry to reconstruct detector response function as follows [Zabulonov et al, 2009b]. A system of several detectors located at various distances from the radiation sources simultaneously measures the gamma radiation spectra. Those multiple measured spectra are processed using the BSS method and at the output we get the separated detector response functions of individual radionuclides as reconstructed mixture components. This BSS approach to spectroscopy does not require measurement of the background radiation, knowledge of detector response functions, and the number of reconstructed hidden sources only depends on the number of detectors.

We carried out experiments [Zabulonov et al, 2009b] investigating BSS performance in spectrometric tasks. In one of the experiments, the radiation sources Cs^{137} and K^{40} were measured simultaneously by three detectors located at various distances from the sources.

The measured spectra (Figure 85) were input to the Independent Component Analysis algorithm FastICA [Hyvarinen, 1999] belonging to the BSS family. The separated hidden sources corresponding to ^{137}Cs, ^{40}K and background radiation are presented in Figure 86.

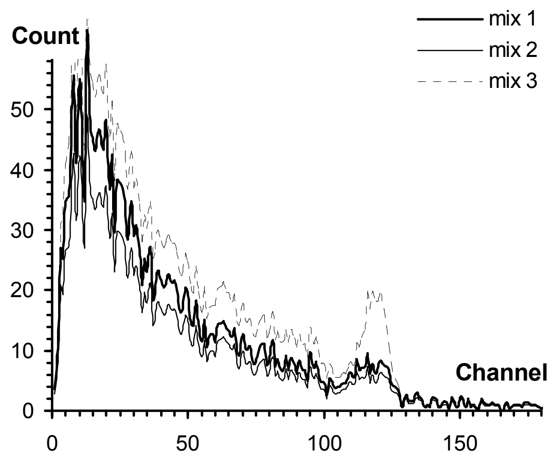


Figure 85. The results of gamma-ray spectra measurements by 3 detectors. ^{137}Cs, ^{40}K, and are mixed in the spectra

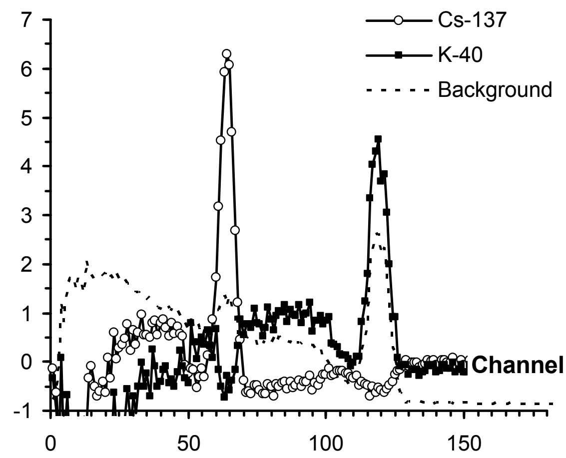


Figure 86. Independent components separated from the spectrum mixtures: ^{137}Cs, K^{40}, and background

The BSS methods are sensitive to noise and require large data sets, so that their usage in non-stationary applications is rather problematic [Bach and Jordan, 2002]. To improve BSS performance in this situation, we have developed a cost function based on the algorithmic complexity approach, and algorithmic mutual information in particular [Grunvald and Vitanyi, 2003]:

$$I(x_1{:}x_2) = K(x_1) - K(x_1 | x_2), \tag{16}$$

where $K(x_1)$ is the algorithmic complexity of x_1, $K(x_1|x_2)$ is the complexity of x_1 conditioned upon x_2.

The algorithmic information theory provides much of the foundation of machine learning and statistical and inductive inference. However, it is non-trivial to calculate it. We propose [Revunova, 2005b; 2005c] to calculate K(·) using minimum description length [Rissanen, 1978; 2002] based on a universal model nMDL proposed by Bin Yu [Hansen and Yu, 2001]:

$$\begin{gathered} nMDL = (L/2)\log(RSS/(L-k)) + 0.5\,k\log(F) + \log(L-k) - 3/2\log(k), \\ F = (\mathbf{x}^T\mathbf{x} - RSS)/(k\,RSS/(L-k)),\ RSS = \|\,\mathbf{x} - \Psi\mathbf{w}\,\|^2. \end{gathered} \tag{17}$$

Here $\mathbf{x}$ is described by a linear model M as $\mathbf{x}=\sum_{i=1,k} w_i \psi_i$, where $\mathbf{w}\in\Re^k$ are model parameters, $\psi\in\Re^L$ are basis functions selected in process of constructing M and used to approximate $\mathbf{x}\in\Re^L$, L is the number of data samples used to construct the model, k is the number of parameters, ψ_{ij} compose matrix $\Psi\in\Re^{L\times k}$.

The developed cost function demonstrated robust performance in the BSS tasks in the presence of noise as well as a higher signal to noise ratio compared to FastICA at small level of noise and to PCA in the whole range of tested noise values [Revunova, 2007a, 2007b]. We believe that application of the BSS methods will increase the quality of nuclide identification.

Research of combined blind source separation approach and sparse approximation with model selection approach to gamma-ray spectroscopy looks promising (Figure 82).

5.3 Machine learning with regularization for airborne gamma-ray surveying

Airborne gamma-ray spectrometry is a remote sensing technique that reconstructs the surface density of radioactive sources. It is an important part of remote environmental radioactivity monitoring and is used for the direct detection of ore bodies and as a lithological mapping tool, assessing health risks associated with radon in houses, ground water discharge and salinity studies, soil mapping, the mapping of fallout from nuclear accidents, etc. [IAEA, 2003].

Airborne gamma-ray spectrometry measures gamma-ray emissions from radioactive isotopes within the earth's near surface (<0.5 m) or from the surface itself. Generally, 512-channel spectra are collected and analyzed. Each radiometric measurement represents a complex average of a relatively large region, with areas closer to the aircraft contributing more than those do farther away. Hence, individual features on the ground appear blurred in the aircraft image. This blurring of spatial detail can be removed partially by intelligent data processing, the quality of the result depending on the noise level within the data, the sampling density, and the mathematical model describing the smoothing process [Billings and Hovgaard, 1999], [Billings et al, 2003], [Dickson, 2004].

In this section, we consider two reconstruction techniques for. First, we observe best available convolution-based signal processing approach and its solution using the Fourier transform. Then, we consider an alternative machine learning method that we develop based on direct regularized solution to the ill-posed discrete problem in space domain. We get a stable solution based on random projections, pseudoinverse, and show that its accuracy is comparable to the Tikhonov regularization, but is less expensive computationally.

5.3.1 Airborne survey technique

Airborne gamma-ray surveying is done as follows. The aircraft, a plane ("fixed wing") or helicopter that carries a gamma detector (Figure 87) moves over the target area with a given topographic relief (topography). Airborne geophysical surveys are normally flown on a regular grid along parallel lines ("flight lines"). The height and line spacing of a radiometric survey determine the spatial and spectral resolution that can be achieved. The line spacing (generally 100–400 m) controls the sampling density in the across-line direction, which can be significantly less than the sampling density along the line (generally 50–70 m). The footprint of a single measurement increases with flying height, while the count level decreases, thus reducing both spatial and spectral resolution. Ideally, to maximize both spatial and spectral resolution, a survey should be flown at the lowest possible safe altitude. Since gamma radiation is completely absorbed by the air layer of 200-300 m, surveys are typically flown at a constant height above the ground of between 40 m and 100 m, with helicopters able to fly considerably lower than most fixed-wing aircraft. There is obviously a trade-off in data acquisition between the observed count rates, and hence the accuracy of the measurements, and sampling time, aircraft speed, and spatial resolution of the data. Survey grids and traverse spacing should reflect the expected strength, size and distribution of sources [Billings & Hovgaard, 1999], [IAEA, 2003].

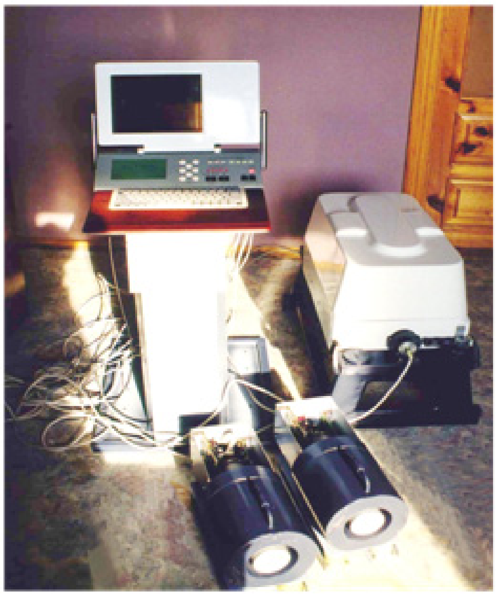


Figure 87. Onboard airborne gamma-ray system

Airborne gamma-ray spectrometric data can be severely distorted in rough terrain as follows: the radiation signal decays at an approximately exponential rate with increasing distance to the source; the rate of decay of radiation signal with distance depends on the source geometry; the resolution of discrete sources decreases with aircraft height; terrain clearance varies from line to line; the source geometry varies depending on the aircraft height and the severity of the topography; the geometry of the detector may have a directional bias.

Gamma-ray data are collected at height, and the detector cannot be focused like a camera. This results in a blurring of spatial detail, which gets worse as the height increases.

The height of the (x, y) point relative to some zero level is h (x, y). The time interval during which the gamma-ray count is accumulated before its recording is called the exposure time. Gamma-ray spectrometric data are usually acquired over a sample interval of one second. Thus, each flight line results in the array $F(t_i)$, $i=1,\ldots,N^d$ of detector readings. The carrier coordinates $(x^d(t),y^d(t))$, topography h(x,y), and height $h^d = h^d(x^d(t),y^d(t))$ are considered known.

Survey results are usually reported in units of gamma-ray dose rate (total-count surveys) or concentrations of the radioelements (spectrometer surveys). Measured count rates are dependent not only on the ground radioelement concentrations, but also on the equipment used, and on the nominal height of the survey. This is undesirable, as measuring units should have a direct significance and be independent of the instrumentation and survey parameters. Count rates should therefore be converted to ground concentrations of the radioelements.

5.3.2 Problem formulation and intelligent data processing techniques

The task is to restore the radionuclide surface density function g(x,y) by the detector readings $F(x^d,y^d)$. Due to the large angular aperture of the collimator, the measurement results (detector readings, count rates) are smoothed and do not reflect in detail the surface density pattern (Figure 88).

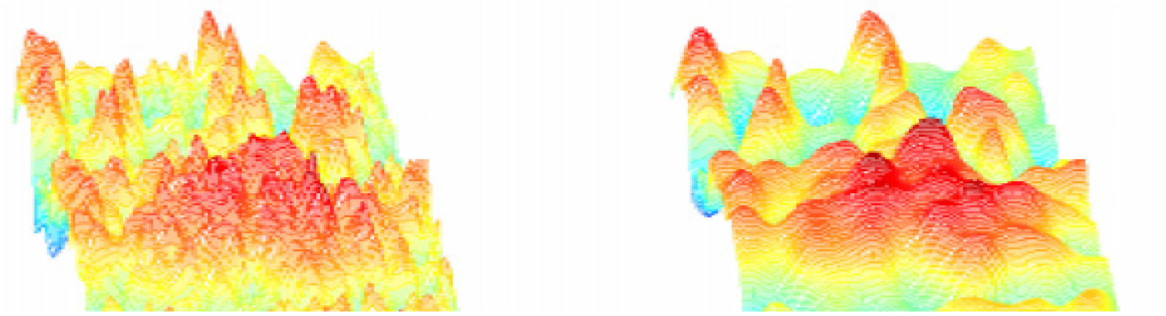


Figure 88. The surface density function (left) and its measurements (right)

Let us consider intelligent information processing techniques for reconstruction of the surface density. The detector count rates are related to the unknown surface density by a Fredholm integral equation of the first kind. For the case of stationary detector placed above the point (x^d,y^d) at the height h^d this equation is

$$\iint dx\, dy\, g(x,y)\, K(x^d,y^d,h^d,x,y,h) = f(x^d,y^d,h^d), \qquad (18)$$

where $f(x^d,y^d,h^d)$ is the number of gamma-rays counted per time unit, g(x,y) is the surface density expressed in the number of decays per square meter per second, h(x,y) is the topography altitude at the point (x,y), $K(x^d,y^d,h^d,x,y,h)$ is the kernel function (also known as "point spread function" [Billings et al, 2003]) describing the smoothing process. The aircraft movement can be easily taken into account given solution to Eq.(18) [Billings et al, 2003], [Zabulonov et al, 2006].

Recent progress in the processing of airborne gamma-ray data was connected with multichannel processing techniques. For example, NASVD [Hovgaard and Grasty, 1997], [Minty and McFadden, 1998], [Minty and Hovgaard, 2002], [Mauring and Smethurts, 2005] and MNF [Green et al, 1988], [Dickson and Taylor, 1998; 2000], see also [Dickson, 2004], [Ramos et al, 2007], use information contained in the whole 512-channel spectrum to reduce the noise and improvement the quality of measured data f and thereby in the sought-for g.

Assuming small variations in survey altitude and topography (h=const) the kernel (point spread function) of the integral equation Eq.(18) is presented as

$$K(d,c,h) = K(d-c,h), \qquad (19)$$

where $d=(x^d,y^d)$, c=(x,y), so that Eq.(18) becomes the convolution equation.

In [Billings and Hovgaard, 1999] the kernel is expressed as

$$K(d-c) = D(d-c,h) \exp(-\mu r)\, C/r^3, \qquad (20)$$

where C is a constant that can be determined by standard calibration procedures, $r = ((x^d-x)^2 + (y^d-y)+h^2_{const})^{1/2}$ is source-detector distance; μ is the linear attenuation coefficient of the air, and D incorporates the detector response.

The exponential term accounts for attenuation of gamma-rays in the air, a factor $1/r^2$ accounts for geometrical dispersion, and 1/r arises from attenuation within the earth (with its own attenuation coefficient). The detector model is derived from geometrical arguments. [Gunn, 1978], [Craig, 1993], [Craig et al, 1999] present a point spread function for deconvolving airborne radiometric data that neglects aircraft movement and assumes that the detector response is non-directional. The point spread function, which take into account aircraft movement and the detector characteristics, is considered in [Billings and Hovgaard 1999], [Billings et al, 2003].

In deconvolution approach, solution of Eq.(18-20) is obtained in the frequency domain using the Fourier transform (Figure 89). In the Fourier domain, Eq (18) has a very simple form

$$W(\mathbf{u}) = K(\mathbf{u})\, G(\mathbf{u}), \qquad (21)$$

where and $\mathbf{u} = (u_x, u_y)$ is the spatial frequency, and upper case means Fourier transformation.

A naive way to estimate $G(\mathbf{u})$ is by spectral division

$$G(\mathbf{u}) = W(\mathbf{u}) / K(\mathbf{u}). \qquad (22)$$

However, $W(\mathbf{u})$ is not known exactly and at high frequencies it is dominated by noise. Since $K(\mathbf{u})$ decreases rapidly with frequency the spectral division amplifies noise. So, the reconstruction usually uses the Wiener filter $V(\mathbf{u})$:

$$G(\mathbf{u}) = W(\mathbf{u})\, V(\mathbf{u}). \qquad (23)$$

The quality of the resulting deblurring of spatial detail depends on the noise level within the data, the sampling density, and the kernel describing the smoothing process. The improvement achieved by the deconvolution is limited by the noise levels in the data and the degree of smoothing imposed by the detector height.

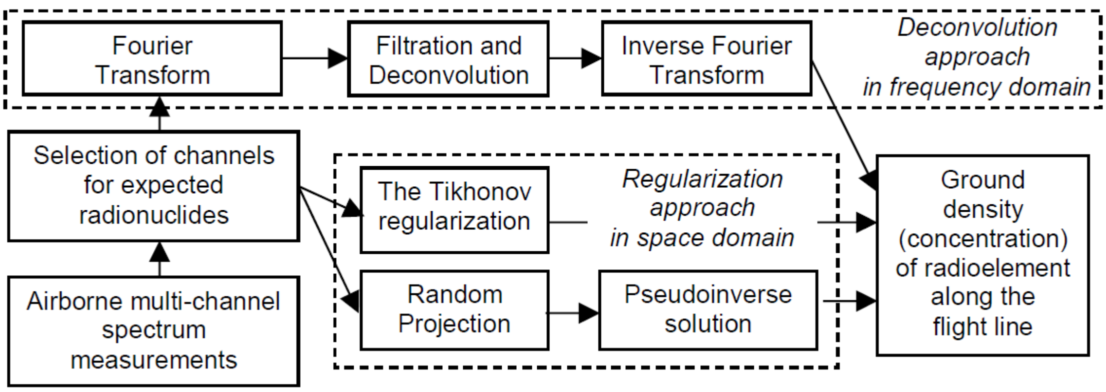


Figure 89. Deconvolution and regularization machine learning approaches to airborne gamma-ray surveying

The deconvolution approach has the following drawbacks. Derivation of the Wiener filter requires knowledge of spatial statistics of both the actual distribution density g(x,y) and the measurement noise $\varepsilon(x,y)$. Since both are unknown, some assumptions should be made, that may be not always valid. For example, determining the noise levels after multichannel processing such as NASVD, and particularly after application of the adaptive filter, is difficult. So, in [Billings et al, 2003], after getting a ballpark estimate of the noise, manual adjustments are used to visually tune the deconvolution.

Also, the blurring in the data is assumed to be the same everywhere. This means that no accommodation can be made for changes in the detector height or for 3D terrain effects.

The noise is assumed additive and the same everywhere. Due to the Poisson nature of radioactive decay and gamma-ray detection, this is clearly not the case. To mitigate this effect, sometimes an adaptive 2D Lee filter [Ristau and Moon, 2001] is used to remove random noise from an image while maintaining image edges.

In situations where the terrain clearance (variations in survey altitude and topography) and noise vary significantly over the survey area, a partial solution would be to break the survey up into smaller segments and apply deconvolution separately to each sub-area.

An alternative approach [Zabulonov et al, 2006] is to use a full space domain formulation. For a particular example, let us consider a detector whose field of view is defined by a vertically oriented collimator of the horn type with a rectangular cross section and angular aperture (θ_x,θ_y), e.g. $0.5\theta_x = 0.5\theta_y \leq 60°$. Therefore, the detector's field of view looks at a square piece of the surface topography, called "window" hereafter. The window size is determined by the angular aperture of the collimator and the height of detector. The window movement during the exposure time makes a rectangular exposure strip.

This setup is described by Eq.(18), where integration is over the area of the browsing window on the ground surface S^d, and

$$K(x,y, x^d,y^d) = s\, q(x^d,y^d,x,y)\, \exp(-\mu\rho)/4\pi\rho^2, \qquad (24)$$

where $q(x^d,y^d,x,y) = (h^d-h(x,y))/\rho$, $\rho = ((h^d-h(x,y))^2 + (x^d-x)^2 + (y^d-y)^2)^{1/2}$ is the distance from the detector to the surface point, $q(x^d,y^d,x,y)$ is the detector sensitivity function. s is the effective area of the detector inlet, including the calibration constant, and taking into account, for particular radionuclide, the quantum yield, the detection efficiency, and the inlet area. Here, only surface radiation sources resulted from fallout are taken into account.

The major drawback of this full space domain formulation of the airborne gamma-ray survey problem is that it requires solution of discrete ill-posed inverse problem. Below, we consider an approach to a stable solution of this problem.

5.3.3 Regularized solution of the ill-posed inverse problem for a full spacedomain formulation

After discretization of the Fredholm integral equation Eq.(18), we get

$$\mathbf{A}\mathbf{x}=\mathbf{b}, \qquad (25)$$

where the matrix $\mathbf{A}\in\Re^{N\times N}$ and the vector $\mathbf{b}\in\Re^N$ are obtained from discretization of the kernel K and $f(x^d)$, respectively; $\mathbf{x}$ represents the unknown g(x,y=const) along the flight line to be reconstructed. The vector $\mathbf{b}\in\Re^N$ is distorted by the additive noise $\varepsilon\in\Re^N$: $\mathbf{b}=\mathbf{b}_0+\varepsilon$.

In case when the singular values σ_i of $\mathbf{A}$ decay gradually to zero, the ratio between the largest and the smallest nonzero singular values is large, the problem is known as discrete ill-posed problem [Hansen, 1998]. Actually, these properties are present in the spectrometry task considered here.

Approximate solutions of discrete ill-posed problems as the least squares problem

$$\mathbf{x}' = \operatorname{argmin}_{\mathbf{x}} ||\mathbf{b}-\mathbf{Ax}||_2, \tag{26}$$

using standard methods of numeric linear algebra such as LU, Cholesky, QR factorizations are unstable. It means that small perturbations in the input data lead to large perturbations in the solution.

To overcome the instability and to improve the accuracy of solutions to discrete ill-posed problems, regularization methods have been proposed [Hansen, 1998], [Tikhonov and Arsenin, 1977], [Morozov, 1984], [Engl et al, 2000]. In the fields of machine learning and inverse problems, regularization involves introducing additional information in order to solve an ill-posed problem or to prevent overfitting. This information is usually of the form of a penalty for complexity, such as restrictions for smoothness or bounds on the vector space norm. Regularization imposes some constraints on the desired solution that stabilizes the problem and leads to meaningful and stable solution. For example, Tikhonov regularization [Tikhonov and Arsenin, 1977], [Hansen, 1998] penalizes solutions with large l_2-norms.

The standard form of Tikhonov regularization problem is formulated as follows:

$$\mathbf{x}_{reg} = \operatorname{argmin}_{\mathbf{x}} (||\mathbf{Ax}-\mathbf{b}||_2+\lambda||\mathbf{x}||_2), \tag{27}$$

where λ is the regularization parameter.

Regularization has a rich history, which dates back to the theory of inverse ill-posed and ill-conditioned problems inspiring many advances in machine learning, support vector machines and kernel based modeling techniques. Determination of the regularization parameter in the Tikhonov scheme is considered an important problem.

Solution of Eq.(27) can be obtained by the method of filtered SVD [Hansen, 1998]:

$$\mathbf{x}_{reg} = \mathbf{V} \operatorname{diag} (f_i / \sigma_i) \mathbf{U}^T \mathbf{y}, \tag{28}$$

where $f_i = \sigma^2_i /(\sigma^2_i + \lambda^2)$ are filter factors.

One problem here is using the SVD of **A**, since it is a computationally expensive decomposition method. Another problem, that may appear even more important, is selecting the proper regularization parameter λ.

A number of methods for selecting the regularization parameter have been proposed [Engl et al, 2000]. The L-curve method [Hansen and O'Leary, 1993] makes a plot of $||\mathbf{x}_{reg}||_2$ vs $||\mathbf{Ax}_{reg}-\mathbf{b}||_2$ for all valid regularization parameters. For discrete ill-posed problems the L-curve, when plotted in log-log scale, often has a characteristic L-shape appearance, hence its name. A distinct corner separates the vertical and the horizontal parts of the curve. The regularization parameter not far from the corner is selected as optimal.

The discrepancy principle [Morozov, 1984] chooses the regularization parameters such that the residual norm for the regularized solution satisfies $||\mathbf{Ax}_{reg} - \mathbf{b}||_2 = ||\mathbf{e}||_2$, where **e** in the norm of perturbation of the right-hand side, and therefore requires an estimation of $||\mathbf{e}||_2$, i.e. noise.

The generalized cross-validation [Wahba, 1990] method is based on the idea that an arbitrary element b_i of the right-hand side **b** can be predicted by the corresponding regularized solution, and the choice of regularization parameter should be independent of an orthogonal transformation of **b**. This leads to choosing the regularization parameter that minimizes $||\mathbf{Ax}_{reg} - \mathbf{b}||^2 / D^2$, where D is a squared effective number of degrees of freedom (which is not necessarily an integer) that can be calculated as $D = m-\Sigma_i f_i$. Here the errors of **b** are considered as uncorrelated zero-mean random variables with a common variance, i.e., white noise.

✓ Solutions using random projections

The drawbacks inherent in the methods of solving discrete ill-posed problems based on Tikhonov regularization include their high computational complexity and the difficulty of selecting the proper regularization parameter (penalty weight) which influences the solution stability. At the wrong values of the regularization parameter, the error of solution may be substantial. Therefore, alternative approaches are required for solving discrete ill-posed problems that would have the accuracy comparable to Tikhonov regularization at lower computational costs.

We develop such an approach using the ideas of random projections [Johnson and Lindenstrauss, 1984], [Papadimitriou et al, 2000], [Achlioptas, 2003] and of our previous work on distributed representations (e.g., [Misuno et al, 2005]) inspired by the idea of information representation in neural networks of the brain. Random projections have recently appeared as a tool for dimensionality reduction and have been used to produce a number of interesting results, both theoretical and applied, including those in context of inductive supervised learning using machine learning methods [Halko et al, 2009], [Revunova and Rachkovskij, 2009s].

The technique plays a key role in several breakthrough developments in the field of algorithms. In other cases, it provides elegant alternative proofs. Recently, researchers working in the area of numeric linear algebra applied similar ideas to get fast randomized algorithms for the least squares problem, matrix factorization, principal component analysis, etc. [Sarlos, 2006], [Drineas et al, 2007], [Rokhlin and Tygert, 2008], [Tygert, 2009], [Halko et al, 2009]. It is therefore of interest to study those techniques and apply them to discrete ill-posed inverse problems.

Let us use the randomized algorithms not only to accelerate, but also to stabilize the solution **x'** of the ill-posed problem, as follows [Revunova and Rachkovskij, 2009s]. Multiply both sides of Eq.(25) by the matrix $\Omega \in \Re^{K \times N}$, K≤N, whose elements are realizations of a normal random variable with zero mean and unit variance. The number of columns N of matrix Ω is determined by the dimension of the matrix **A**, the number of rows K is a priori unknown since the numerical rank of **A** is ill-determined and the required numerical rank of approximation is unknown. We obtain

$$\mathbf{\Omega A\, x} = \mathbf{\Omega b}, \text{ where } \mathbf{\Omega A} \in \Re^{K \times N},\ \mathbf{\Omega b} \in \Re^{K}. \tag{29}$$

Then the least-squares problem is

$$\mathbf{x}_{Pr} = \text{argmin}_{\mathbf{x}}\ ||\mathbf{\Omega A\, x} - \mathbf{\Omega b}||_2. \tag{30}$$

Signal reconstruction based on pseudo-inverse using random projection Ω is obtained as

$$\mathbf{x}_{pinPr} = (\mathbf{\Omega A})^{+}\, \mathbf{\Omega b}. \tag{31}$$

Signal reconstruction based on the pseudo-inverse using the projection matrix **Q** obtained by the QR factorization of **ΩA** is done as

$$\mathbf{x}_{pinQ} = (\mathbf{Q}^T\mathbf{A})^{+}\, \mathbf{Q}^T\mathbf{b}. \tag{32}$$

The pseudo-inverse $\mathbf{P}^+$ of a matrix **P** is actually computed based on SVD as:

$$\begin{gathered}\mathbf{P}^{+} = \mathbf{V}\ \text{diag}\,(\phi_i / \sigma_i)\ \mathbf{U}^T, \text{ iff } \sigma_i > \text{tresh } \phi_i = 1, \text{ otherwise } \phi_i = 0.\\ \text{tresh} = \max(K,N)\ \text{eps}(\max(\sigma_i)),\end{gathered} \tag{33}$$

where **U**, **V, S** are obtained by the SVD of $\mathbf{P} = \mathbf{USV}^T$; σ_i = diag **S** are singular values, the elements of a diagonal matrix **S**; floating-point relative accuracy eps(z) is the positive distance from abs(z) to the next larger in magnitude floating point number of the same precision as z.

In order to compare the quality of reconstruction of the surface density of radioactive sources, we conducted an experimental study of techniques for solving discrete ill-posed problems using the simulated data of radioactivity monitoring at h=100m.

The matrix Φ (i.e., **A** in Eq.(25)) obtained by discretization of the kernel Eq.(24) has dimensionality of 200×200 (Figure 90), large condition number ($\sigma_{max}/\sigma_{min}$>>1), and singular values σ_i gradually decaying to zero. The right-hand side **y** (i.e., **b** in Eq.(25)) is distorted by an additive noise with the Gaussian distribution and various amplitudes. Figure 91 shows the measurement **y** produced by the doublet signal **x** to be restored.

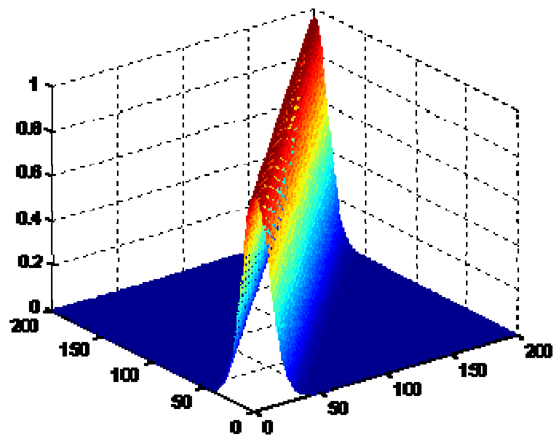


Figure 90. The kernel matrix Φ for gamma-ray surveying

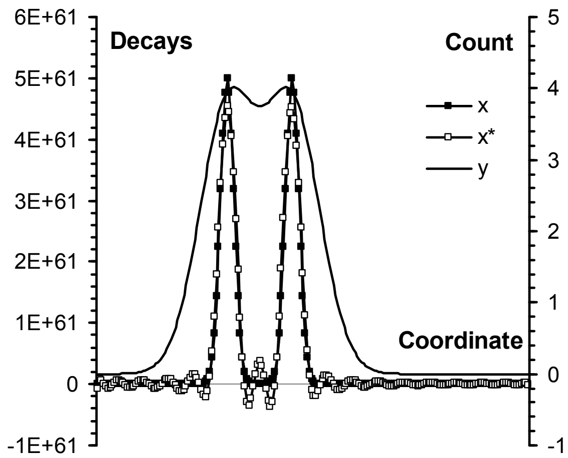


Figure 91. The modeled 1D surface density x, the measurement y, and the restored density x* for gamma-ray surveying

First, we have compared performance of three kinds of techniques: the pseudo-inverse solution Eq.(26), the Tikhonov regularization Eq.(28) with the selection of the regularization parameter λ using the L-curve, the generalized cross-validation, and the discrepancy principle [Hansen, 1998], [Wahba, 1990], [Morozov, 1984], and the pseudo-inverse with projection using the random matrix $\boldsymbol{\Omega}$ Eq.(31) and with the ortogonalized matrix **Q** Eq.(32). The random projection matrix $\boldsymbol{\Omega}\in\Re^{K\times N}$, N=200, K≤N is the Gaussian random matrix with the entries of zero mean and unit variance. Three noise levels are used $\{10^{-2},10^{-6},10^{-10}\}$. The results in terms of the signal reconstruction error vector norm are given in Table 4.

Without projection, the standard pseudo-inverse provides an acceptable error e_{Pin} only at the lowest noise level and cannot be used at all at the larger noise levels due to very large error. The errors for the Tikhonov regularization with the selection of λ by the L-curve e_{RegT} Lcur, the generalized cross-validation e_{RegT} GCV, and the discrepancy principle e_{RegT} Dsc are generally comparable for all three noise levels. However, we observe an outlier for e_{RegT} GCV at the noise level 10^{-6}, giving the instance of unstable regularization parameter selection. The lowest signal reconstruction error is provided by the Tikhonov regularization with the L-curve.

Table 4. The signal reconstruction error for the discrete ill-posed problem of gamma-surveying obtained by the regularization methods without and with random projecting. K is the dimensionality of projection matrix.

Noise level	Error, solutions without projection				Min error, with projection		Error, with projection		Error, with projection	
	e_{Pin}	e_{RegT} Lcur	e_{RegT} GCV	e_{RegT} Dsc	e_{PinR} (K)	e_{PinQR} (K)	e_{PinR} gMDL(K)	e_{PinQR} gMDL(K)	e_{PinR} Lcur(K)	e_{PinQR} Lcur(K)
10^{-2}	$7.3\ 10^{8}$	48.9	47.3	64	53.8(26)	50.3(29)	53.8(26)	51.3(27)	53.8(26)	51.9(30)
10^{-6}	$7.3\ 10^{4}$	21.1	$1.6\ 10^{3}$	22.8	21.9(35)	21.7(39)	21.9(35)	21.8(35)	22.0(36)	21.7(39)
10^{-10}	11.1	9.4	11.1	15.5	9.01(77)	8.3(47)	21.9(35)	12.5(45)	N/A	N/A

With the random projection, the pseudo-inverse becomes stable and the reconstruction error values at the minimum (with the proper choice of K) are comparable and small for the projector matrix **Ω** (Eq.(31)) e_{PinR} and **Q** (Eq.(32)) e_{PinQR}.

In order to see how to choose the optimal K for the projective methods, let us consider the dependence of the signal reconstruction error e_{PinQR} on the dimensionality K of the projector matrix **Q** (Figure 92, left Y-axis). The error minimum can be observed for all noise levels. With increasing noise level, position of the error minimum shifts to the lower values of K and the error value at the minimum increases. To choose the projection matrix dimensionality K at which the solution error is close to the minimum for real situations, i.e. where the exact solution is unknown, we propose using various types of model selection criteria:

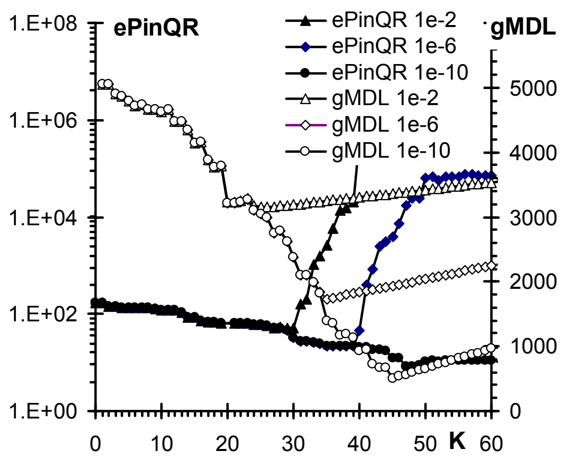


Figure 92. Dependence of the signal reconstruction error e_{PinQR} (left Y-axis) and gMDL values (right Y-axis) on the dimensionality K of the projection matrix at the noise levels 10^{-2}, 10^{-6}, 10^{-10}. Minima are at close K

(1) Criteria for model selection discussed above in context of sparse approximation and matching pursuit [Mallows, 1973], [Akaike, 1974], [Hansen and Yu, 2001];

(2) Criteria proposed to select the regularization parameters of Tikhonov-like regularization using the L-curve, the generalized cross-validation, and the discrepancy principle [Hansen, 1998], [Wahba, 1990], [Morozov, 1984].

Figure 92 (right axis Y) shows the values of Bin Yu model selection criterion gMDL Eq. (11) vs K. We observe that the minimum value of gMDL is near the true solution minimum. The results of K selection using gMDL, as well as those obtained with the L-curve, are shown in Table 4 and demonstrate the reconstruction error near or equal to the optimal one. An example of the reconstructed result **x*** is shown in Figure 91. So, the model selection criteria can be used for getting the required K providing the near-optimal solution.

Figure 93 shows the computation times of different techniques vs. K. As expected, the solution based on pseudo-inverse without random projection has a constant and rather large computation time compared to the solutions using random projection at the smaller values of K. For example, at K=30, the computation time with random projection (~0.003s for both considered projection methods) is 20 times less than that of ordinary pseudoinverse without random projection (0.06s). This time reduction is because the singular value decomposition is performed on the resulting (N×K) matrix after the projection, where K is a small fraction of N of the original (N×N) matrix Φ. Particularly large gains are achieved at the higher noise levels because their optimal K is small.

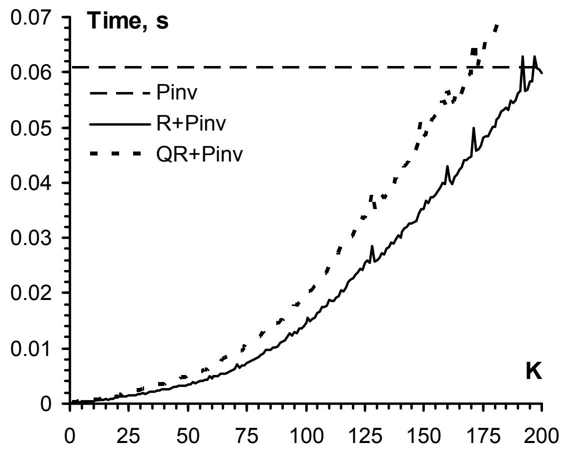


Figure 93. The computation times of pseudo-inverse solutions vs K: Pinv – ordinary, R+Pinv – with random projection, QR+Pinv – random projection orthogonalized

Thus, the study and application of intelligent regularization-based techniques for solving discrete ill-posed problems based on pseudo-inverse with a random projection is a promising direction due to the lower computational costs, and also because of their stability, manifested in the smooth change of the signal reconstruction error with the increasing noise and K.

Directions of further work in this area include:

- solution methods that take into account nonnegativity of solution, such as non-negative matrix factorization, non-negative least squares, etc.;
- techniques for a computationally efficient choice of the optimal K based on the knowledge about solution, such as its smoothness, nonnegativity, etc.;
- high-performance hardware implementation for real-time applications using efficient systolic architectures [Brent, 1988], [Brent and Luk, 1985], [Zabulonov and Revunova, 2006].

5.4 Discussion

In this chapter, we discussed machine-learning techniques for intelligent processing of gamma-ray spectroscopy data in environmental and security-related monitoring.

In particular, we developed and applied to intelligent processing of multichannel gamma-ray spectrometer data the supervised learning techniques of sparse approximation with model selection that resulted in a substantial increasing of data processing quality. Development of a new model selection test is suggested for further improving the quality of gamma-ray spectrometry. For this task, we also considered blind source separation approach of unsupervised learning, including a novel objective function based on algorithmic complexity. It is expected to become an important step in the development of the new generation of intelligent information technologies for multichannel and multidetector gamma-ray spectroscopy under a complex and unknown geometry of measurements.

We also described an advanced approach to processing of airborne gamma-ray spectrometry data in order to reconstruct the surface density of radionuclides. This approach requires solution of the discrete ill-posed inverse problem resulting from direct space domain problem formulation. In order to improve the speed and stability of the Tikhonov regularization traditionally used to solve such problems, we developed and employed a novel method based on random projection, which is conceptually close to distributed representations in neural networks. This resulted in substantial computational savings for noisy measurements, which is always the case. Again, model selection criteria were employed to choose the random projection parameter.

We would like to note that the developed and deployed intelligent data processing techniques are not limited to the case study of particular gamma-ray processing task considered. They can be applied to other kinds of environmental gamma-ray monitoring, as well as to other kinds of environmental monitoring in many areas beyond those considered in this chapter – those that require machine learning techniques for solving inverse problems that are ill-conditioned or ill-posed and call for model selection and regularization.

In addition to the machine learning methods considered in this chapter, another way to increase the level of intelligence in monitoring for environment and security is construction of and integration with knowledge bases and expert systems, particularly those employing analogical reasoning and knowledge mining [Gladun 1994], [Gladun and Vashchenko, 2000], [Gladun et al, 2008], [Markman et al, 2003], [Rachkovskij, 2001; 2004], [Rachkovskij and Kussul, 2001], [Slipchenko and Rachkovskij, 2009s]. This requires formation of data and knowledge bases and other intelligent memory structures accumulating the measurement results together with the results of their processing for varied monitoring situations, as well as with their expert estimations. Those bases could be used both for machine learning and for example-based reasoning in order to further raise the intelligent component of gamma-ray monitoring and of other kinds of environmental and security-related monitoring. An advanced access to available and emerging worldwide, national, and local databases, and their computation-intensive processing would benefit from the high-performance intelligent computations and multi-source data integration of geographically distributed Grid computing technologies.

6

Acquisition, Processing and Analysis of Space Images at Risks Management of Natural and Technogenic Emergencies

6.1 Base principles of automation for risks emergency management

6.1.1 On the basis of remote space monitoring data

Maintenance of all-round safety of the person, society, the state and the world community became a priority of the next decades, has turned to one of overall objectives for strategy of a civilization existence in modern and predicted conditions.

Scientific-technical and social-economic progress of last centuries has considerably changed the world, essentially having improved working conditions, quality of a life, education and culture. At the same time, technological progress has shown the negative side connected with an exhaustion of resource possibilities of the Earth and extensive character of their operation, a number of the crisis phenomena in social, economic, political spheres, and with the advent of new technogenic threats. Interdependence of natural and technogenic spheres has essentially amplified.

Natural and technogenic dangers are an integral part of a life for the population of modern Europe.

The probability of occurrence of natural and technogenic accidents considerably increases possibility of that in a zone of emergency there will be the territories sated with engineering constructions and utility-power systems. In the largest cities natural disasters are capable to cause a series of technogenic failures (fires, explosions, emissions of chemical substances etc.), having catastrophic consequences.

Natural cataclysms, such as flooding, typhoons, a tsunami, volcanic eruptions, droughts and forest fires, lead to different consequences and depend how the society prepares for them, predicting their occurrence, and what measures accepts after their approach. For example, building constructions, which are not answering to standards of seism stability in area where the probability of earthquakes is high, obviously increases number of possible victims, and thereby human scales of misfortune.

For decrease in losses, increases of stability for functioning of productive-economic and social systems at all levels (interstate, the countries, territories, the enterprises, and the organizations), it is necessary to carry out measures of protection from emergency. However, realization of protection measures is connected with considerable expenses that demands to correlate with expense from a prevented damage.

For potentially dangerous events speech should go about prevented risk based on monitoring operative data for natural and technogenic objects. At risk management of emergencies it is necessary to estimate efficiency of actions for their decrease based on all saved up data sharing (the integrated data ware). These actions should provide realization of management strategy by the risks, reflecting activity mainstreams on achievement of a comprehensible risk level in all its aspects. Thus, each strategy demands reception and processing of the certain kind and type information. Correct interpretation of the information about a condition and functioning of dangerous objects, about their interrelations, and also about geophysical both other phenomena and processes is provided with their adequate and correct description with application of a corresponding mathematical apparatus. Set of these descriptions makes subject domain model (SDM). In actions three basic strategy of risk – management by emergencies should be presented:

- strategy of prevention of the reasons of natural and technogenic failures and accidents' occurrence;
- strategy of natural and technogenic accidents' localization and prevention of dangerous conditions formation when the reason of accident' occurrence on technical, economic, social or other reasons to eliminate it is impossible;
- strategy of the maximum possible easing of natural and technogenic factors' influences on people and environment, liquidations of accident and its consequences in the shortest terms.

For maintenance of effective computer information-program support for realization of the given strategy it is offered to develop the integrated automated information-analytical emergency system (IAES) of remote space monitoring of the spatially-distributed natural and technogenic objects that allows timely to detect, diagnose and forecast of the dangerous phenomena and emergency' development. Creation of IAES is the modern, perspective, scientifically proved strategic, system approach to complex automation of emergency risk-management.

One of the most informative and technological sources of data for IAES are pictures of observable territory [IEOS, 2005]. They can be received as by means of space vehicles (satellites) [UNCOSA, 2003], so be means of modern unmanned aerial vehicles [Voronov, 2009].

The modern level of facilities development and Earth' research methods from space, program complexes for space data processing and a wide dissemination of geographical information systems (GIS), allow to receive qualitatively up-to-date information on a condition of territories, land objects, processes and dynamics of their condition change. New information quality defines new methodological approaches, and perspective technologies in acquisition and target application of results of the earth remote sensing (ERS) for complex researches, the analysis and efficient control development of regions. The basic directions of these works concern the most urgent questions and the problems facing to the state, regional and municipal controls. Such problems concern:

- prediction, search and development of new mineral deposits on insufficiently explored and remote territories;
- rational use and periodic inventory of natural resources;
- an operative dataware of the state, regional and municipal controls;
- the account of the earth’s and the organization of rational land tenure;
- monitoring of emergencies, ecological disasters, natural and technogenic accidents;
- space diagnostics of an infrastructure, including technical communications.

The urgency of these problems defines necessity of creation of the organizational-technical structure providing their effective decision. The space monitoring centers and the geospatial analysis, which can be created at different levels of regional, state, and interstate management based on the corresponding scientific, industrial, and educational organizations, can become similar structures [SMC, 2009].

Main objective of The space monitoring center (SMC) creation: maintenance of reception, complex processing of archival and operative data of ERS, and also the all-round analysis of the territory researches results with a view of granting of the fullest, timely and objective information about nature-resource potential, an economic, ecological condition and dynamics of nature-technogenic complexes development. The goal attainment allows accepting of administrative decisions of various scales.

SMC represents a high-efficiency hardware-software complex, which allows receiving and processing operatively ERS' data. SMC is scaled system and can be easy complemented for the decision of new problems.

SMC activity is focused on decisions of following primary ERS' goals:

- operative pictures acquisition of the same territory with the various spatial permission;
- shooting conducting in various spectral ranges practically at any time;
- repeated pictures acquisition of any territories, including remote and insufficiently explored sections
- space pictures formation into regular time numbers.

As a rule, ERS methods are much more economic than traditional ways of spatial data acquisition. However, for reception of an objective picture for dynamics of processes resulting in investigated territory the space picture is necessary for recognizing and decoding correctly, i.e. to pass from brightness surface characteristics to physical properties of a surface or objects and to spend specialized complex processing for reception of necessary information quality. So the technology of researches on the basis of space monitoring data (Figure 94 [Serebryakov, 2008]) provides the decision of following technological problems:

- operative ERS data acquisition provides monitoring those or other kinds of territorial natural resources, branches of a national economy, environmental problems, emergencies;
- a pre-processing of space pictures, their preparation to further automated and expert decoding and also to visual representation for regional authorities;
- the deep automated analysis and thematic data ERS processing for preparation of a wide spectrum of analytical maps on various subjects, definitions of various statistical parameters, developments of competent administrative decisions and definition of technologic realization;
- preparation of analytical reports, notes, presentation materials on the basis of space shooting materials of region territory, formation of offers and recommendations about the decision of problems, attraction of investments, redistribution of forces and the means put in development of a national economy.
- however, use ERS based on the data received from modern satellite equipment, along with doubtless advantages, possesses also a number of lacks. Basic lacks ERS are huge volumes of the satellite photos databases, essential timetable for preliminary processing of primary satellite photos, decrease in quality of pictures under adverse weather conditions, necessity of licenses purchase since there is no status of a public condition of the photos received from modern commercial satellites, problems of the state secret preservation.

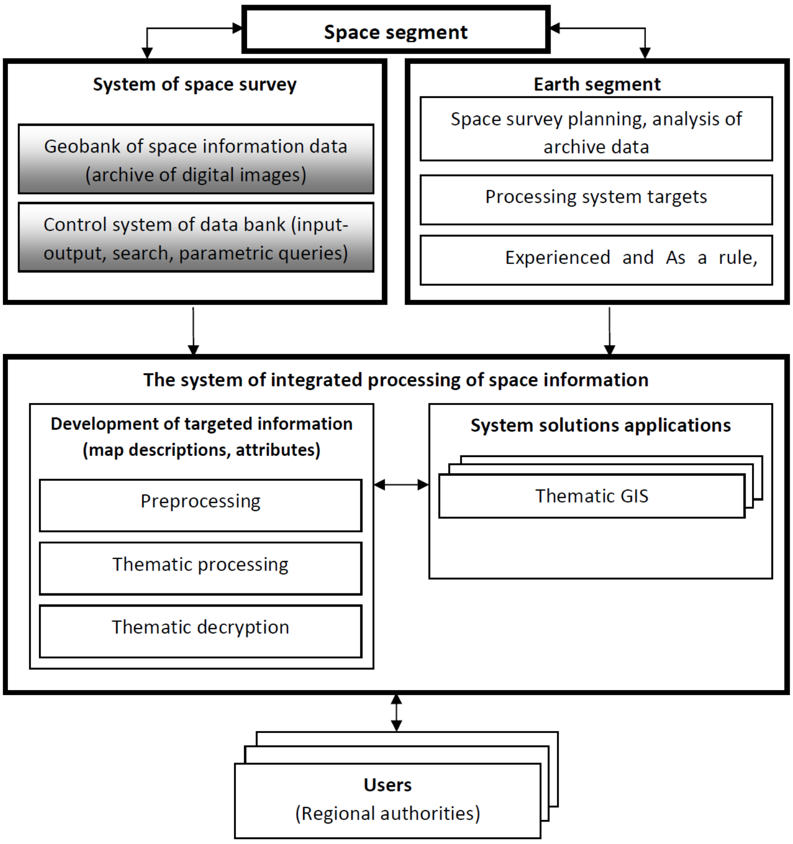


Figure 94. General technological scheme of remote space monitoring

Use of modern flying machines (aircrafts, helicopters, unmanned aerial vehicles (UAV), equipped with special equipment is one from the perspective approaches expanding and supplementing ERS possibilities.

Thus, especially it is necessary to allocate UAV use, capable to transfer not only the visual information, but also a number of other data, which acquisition it is complicated, and so it is impossible when other kinds of techniques are used.

UAV management method that is offered in the second subsection with use of digital imaging systems (DIS) allows expanding considerably a circle of solved problems – to raise quality and reliability of the achieved information, to lower risk of the device loss.

In the third subsection of the given work the method, allowing to estimate information capability of the achieved pictures is offered, and to reveal the areas containing the images of clouds or a smoke from fires. The stage-by-stage algorithm allowing with high degree of accuracy to allocate a zone of a possible arrangement of specified objects for the further processing is developed.

One of the important problems at the Earth's surface monitoring is detection and identification of objects. The mathematical method of restoration of the third coordinate (such as height) offered in the fourth subsection, is based on calculation of intensity curvature function for the image lines. Application of scale space representation (SSR) of images for the decrease of matching points amount is offered in this section, allows raising speed of image processing. It allows not only to conduct renewal of three-dimensional data of the Earth's surface in real time, but to track movements of transportable objects which were captured by satellite's camera. Efficiency of object tracking depends exclusively on frequency of data acquisition from the satellite which conducts certain Earth's surface region observation.

In the fifth subsection application questions of remote space monitoring territorially extended gas-transport networks (GTN) of a high pressure, management technological objects (MTO) having a significant amount the raised danger are considered.

6.2 Satellite imagery for Earth's surface monitoring

6.2.1 Earth's surface image acquisition from satellites

✓ General information

Space imagery of Earth's surface consists of photographs of Earth made by artificial satellites or other equipment, located outside the Earth atmosphere. Recently results of satellite imagery achieved wide distribution due to popularity and simplicity of processing. All satellite images, which were made and published by NASA, are distributed free and are considered as a public property. There is also a list of private companies, which execute commercial satellite imagery projects for different applications.

Satellite images have many applications in agriculture, geology, forestry, biodiversity conservation, regional planning, education, intelligence, and warfare.

Images can be made in visible colors and in other spectra. There are also elevation maps, usually made by radar imaging.

Satellite image acquisition could have different application, which depends on properties of the trajectory of satellite motion – orbit inclination, elevation, and synchronization of orbit. Depending on the chosen orbit, it is possible to take pictures of a certain territory of Earth's surface with different frequency, under a different corner and at different luminosity, resolution, and color detail levels.

Resolution of satellite images depends not only on the elevation of orbit but also from the type of photographic equipment.

On the first step of satellite image acquisition, transformation of raw sensors data to digital image representations is applied. Type of transformation depends on the frequency range of photographic equipment. On the second step, digital satellite images are put to different preprocessing operations: filtration of noise, merging of the color map of low-resolution and monochromatic high-resolution map to produce color images of the Earth's surface, edge detection, merging, etc.

Forming and processing of color-brightness constituent of satellite imagery in a greater measure depends on the used range of electro-magnetic waves. In addition, for the analysis of large territories at the high level of detail it is necessary to merge overlapping square satellite images to form rows and to merge rows of satellite images to form high-resolution maps of a certain territory.

Decoding and analysis of satellite images presently is mostly executed by automated or semiautomatic complexes of software.

✓ Photographing the earth's surface from satellites at different orbits

Depending on the chosen orbit, it is possible to achieve Earth's surface images at different resolution and color detail levels. Orbits differ by elevation over the level of sea – altitude. There are low Earth orbit, medium Earth orbit and geostationary orbit. Low Earth orbits (LEO) are ranging in altitude from 0–2,000 km (0–1,240 miles). Medium Earth orbit (MEO) is ranging in altitude from 2,000 km (1,240 miles) to just below geosynchronous orbit at 35,786 km (22,240 miles). High Earth orbits are located above the altitude of geosynchronous orbit 35,786 km (22,240 miles). Obviously, Low Earth orbits provide maximum resolution images.

Orbit altitude depends on the inclination and satellite motion synchronization. Satellite imagery for earth utilizes satellites flying at geostationary and polar inclined orbits for earth observation and geosychronization and geliosynchronization correspondently.

A geostationary or geosynchronous orbit is one in which the satellite is always in the same position with respect to the rotating Earth. The satellite orbits at an elevation of approximately 35,790 km because that produces an orbital period equal to the period of rotation of the Earth. By orbiting at the same rate, in the same direction as Earth, the satellite appears stationary – synchronous with respect to the rotation of the Earth).

Geostationary satellites provide global view, enabling coverage of weather events, which is especially useful for monitoring severe local storms and tropical cyclones. However, the use of geostationary satellites is limited to global change analyses only because of low resolution. Because a geostationary orbit must be in the same plane as the Earth's rotation, that is the equatorial plane; it provides distorted images of the polar regions with poor spatial resolution. This fact brings more limitation applied to the application of geosynchronous orbit satellites for satellite imagery.

Polar-orbiting satellites provide more global visual information, circling at near-polar inclination (the angle between the equatorial plane and the satellite orbital plane -- a true polar orbit has an inclination of 90 degrees). Orbiting at an altitude of 700 to 800 km, these satellites cover at high level of detail the parts of the world most difficult to cover on site.

Polar satellites operate in a sun-synchronous orbit. The satellite passes the equator and each latitude at the same local solar time each day, meaning the satellite passes overhead at essentially the same solar time throughout all seasons of the year. This enables regular data collection at consistent times as well as long-term comparisons and allows eliminating influence of shadow casting and illumination variation.

Inclined orbits are ate the range from polar to geostationary. They have an inclination between 0 degrees and 90 degrees. Inclined orbits are located at altitudes of few hundred kilometers and thus pass a certain region on Earth each several hours. Therefore, satellites at inclined orbits are able to provide extremely high-resolution images of a certain territory of the Earth's surface several times each 24 hours in a difference from polar satellites. These satellites are not sun-synchronous, however, so they will view a place on Earth at varying times, which sometimes is a problem for the Earth's surface analysis.

Thus, satellites circling at different orbits may be used for different real applications of satellite imagery. Global changes of certain Earth regions, such as weather changes, large fires and floods, could be detected at a high frequency with the use of geostationary satellites. High-resolution images of the Earth's surface could be photographed from polar satellites with a consistent illumination once a 24-hour cycle. Satellites circling at inclined orbits are able to observe a certain territory of the Earth's surface at extremely high level of detail several times each 24 hours.

✓ Technical properties of space imagery

As it has been already noted, the resolution of satellite images varies depending on the instrument used and the altitude of the satellite's orbit. For example, the Landsat archive offers repeated imagery at 30-meter resolution for the planet. For many smaller areas, images with resolution as high as 10 cm can be available. However, most of it has not been processed from the raw data yet.

GeoEye-1 satellite provided the world's highest resolution commercial satellite imagery [GeoEye-1, 2009]. The 0.41 meters resolution panchromatic images allow the satellite to distinguish between objects on the ground that are even smaller than half a meter apart.

However, satellite imagery is not able to provide satisfactory details and may be subject to atmospheric noise and obstacles. That is why satellite imagery is sometimes supplemented with aerial photography, which has higher resolution, but is more expensive per square meter. Satellite imagery can be combined with vector or raster data in a GIS if the imagery has been spatially rectified so that it will properly align with other data sets.

✓ Disadvantages of satellite imagery for the Earth's surface monitoring

Because the total area of the land on Earth is so large and because resolution is relatively high, satellite databases are huge and image processing (creating useful images from the raw data) is time-consuming. Depending on the sensor used, weather conditions can affect image quality: for example, it is difficult to obtain images for areas of frequent cloud cover such as mountaintops.

Commercial satellite companies do not place their imagery into the public domain and do not sell their imagery; instead, one must be licensed to use their imagery. Thus, the ability to legally make derivative products from commercial satellite imagery is minimized.

Privacy concerns have been brought up by some who wish not to have their property shown above. Google Maps responds to such concerns in their FAQ with the following statement: "We understand your privacy concerns... The images that Google Maps displays are no different from what can be seen by anyone who flies over or drives by a specific geographic location."

6.2.2 Application of UAVs for the tasks of environmental and technological objects monitoring

UAVs play one of the key roles in solving problems of emergency monitoring. They allow solving a wide range of tasks for the Ministry of Emergent situations (MES), Ministry of Defense, as well as the national economy. In particular, a UAV can be used as a robotic tool that can perform manufacturing operations in areas dangerous for humans. It is capable of conducting engineering, radiation, chemical and biological exploration; delivering extra special cargo; broadcasting and rebroadcasting the information in emergent situations dangerous for humans, and solving a wide range of other tasks. UAVs may well be engaged in environmental monitoring, protection of gas and oil pipelines and other strategic facilities including nuclear power plants (NPP), they can monitor the situation in the forests and peat bogs, traffic safety on the roads of the country, conduct ice reconnaissance.

In recent years the range of UAVs has expanded considerably that makes it possible to choose the most optimal variant for solving a particular problem. UAVs flight performance lies in a fairly wide range: the take-off weight is from several kilograms to several tons, the payload is from several to hundreds of kilograms, the flight duration is up to tens of hours, the range is up to several thousand kilometers, the speed is up to hundreds of kilometers per hour.

The advantages of such aircraft include:

- mobility and compactness;
- high degree of readiness for departure;
- variegated tasks performed;
- high flight and technical characteristics of the UAV;
- ease of maintenance and reliability;
- economic benefits in comparison with manned aircraft taking into account regular increase in prices of aviation fuel and services for execution of aviation operations and maintenance.

Besides, sometimes-using UAV is the only way of solving problems in case of natural disasters and accidents.

✓ UAV control by using DIS when solving emergency monitoring problems

The most flexible solution in emergencies is to use UAV remote control. The UAV control system should contain the on-board set (OS) which provides the following tasks:

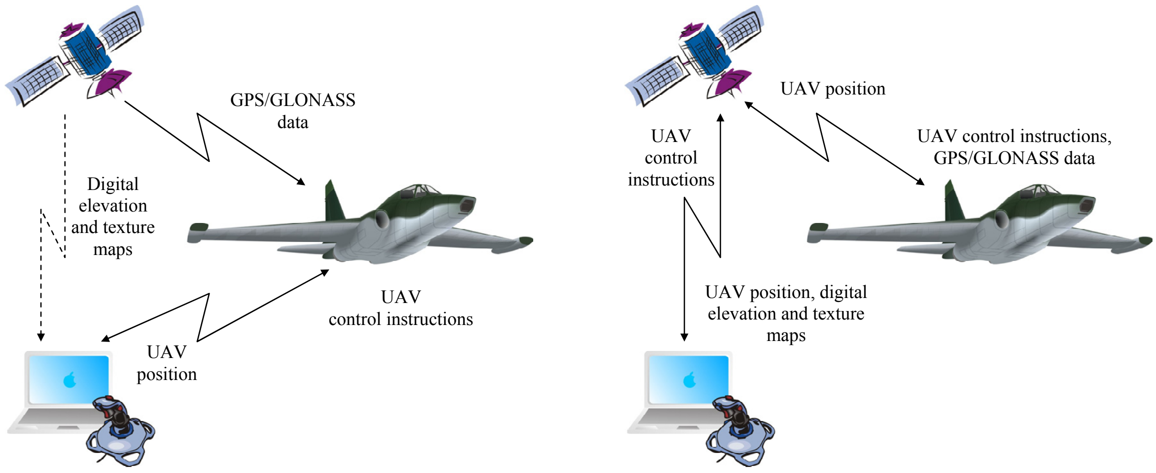
- definition of navigation parameters, angles of orientation and motion parameters of the UAV (angular velocity and acceleration);
- UAV navigation and control when flying along the target trajectory;
- stabilization of the UAV orientation angles in flight;
- delivery of telemetric information about navigation parameters, angles of orientation of the UAV to the transmitting channel;
- payload control.

The operator controls the machine either by using a video camera, mounted on the UAV, or by using visual observation of the aircraft. The main disadvantage of this method is the limited range of control, which varies from a few kilometers by visual observation up to several tens of kilometers using the video channel. In addition, control becomes impossible in the dark, in conditions of limited visibility or its total lack, in case of strong interference in the video channel.

Due to the absence of visual contact with the ground even when radio equipment is available, the operator-pilot undergoes a serious psycho-physiological stress that prevents the adoption of proper decisions. It may lead to the failure in accomplishing the task or to crash.

To date a large number of devices that enable to improve the accuracy and reliability of UAV remote control in poor visibility conditions has been developed. However they all have limitations that do not allow controlling the aircraft in rapidly changing external conditions typical for emergent situations (ES).

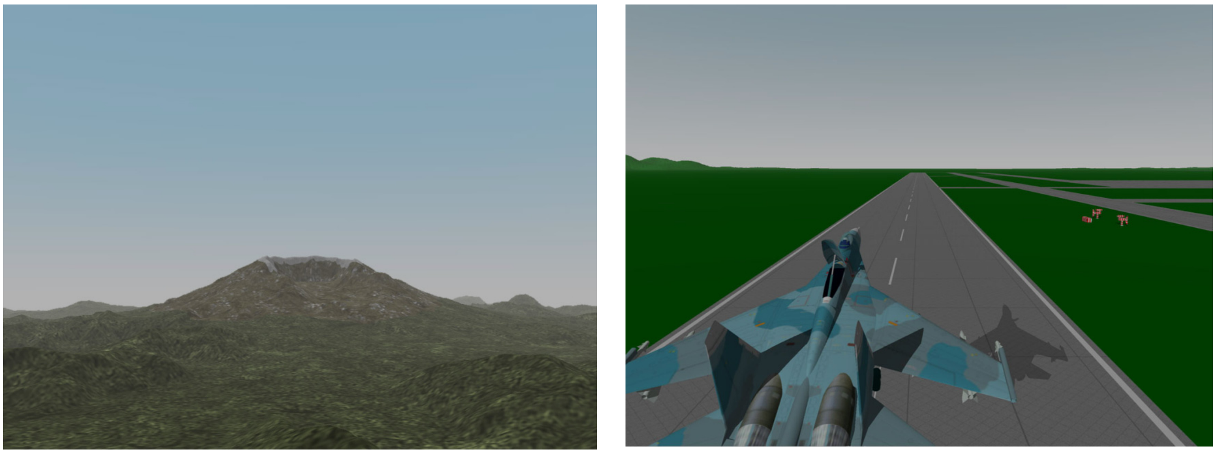
To eliminate these shortcomings we propose using DIS for UAV control [Gusyatin, 2002]. The principle of UAV control by the proposed method is shown in Figure 95a. DIS ensures the production of the image of the UAV surroundings on the basis of telemetry data (linear coordinates, determined by means of GPS or GLONASS receiver and the angular ones defined with the help of sensors installed on board the UAV), coming from aircraft OS, height map and, if possible, area images obtained by a satellite or from a database. As a result, the operator can see the qualitative synthesized image of terrain over which the flight is realized regardless of the actual weather conditions and control the UAV.



a) direct data link between the UAV and DIS; *b) data link between the UAV and DIS via satellite*

Figure 95. The principle of UAVcontrol with the use of DIS

Another advantage of this method is the reduction of the amount of information being transmitted, since there is no need in the video signal to control the UAV that will increase reliability and reduce power consumption.



a) image of terrain; *b) image of airport environs*

Figure 96. Images synthesized by DIS

Using a channel of communication via geostationary satellites will extend the UAV range because there is no necessity to install repeaters on the flight path and reduce the risk of accident due to miss of communication signal. The principle of UAV control by the proposed method is shown in Figure 95b. In this case, the data about the UAV position calculated by OS is transmitted via satellite to DIS, and control commands from the operator via satellite are received by the UAV. Figure 96 shows image frames obtained because of DIS work. Figure 96a shows image of terrain synthesized using digital elevation and texture maps. Figure 96b shows image of airport environs in the landing of aerial vehicle.

✓ DIS structure

DIS structure is shown in Figure 97. DIS contains the following blocks: a universal computer, a specially designed graphic processor unit (GPU) and a graphics display device (DD).

Data from the UAV OS represents three linear (x, y, z) and three angular coordinates (φ, θ, γ) determining position and orientation of the aircraft in space. Terrain data represents a map of heights H (x, y) and the area image in the form of a texture map T (x, y). In addition, if necessary, data about the position of artificial objects O containing information on their type, position, size, etc. can be transmitted to the mainframe.

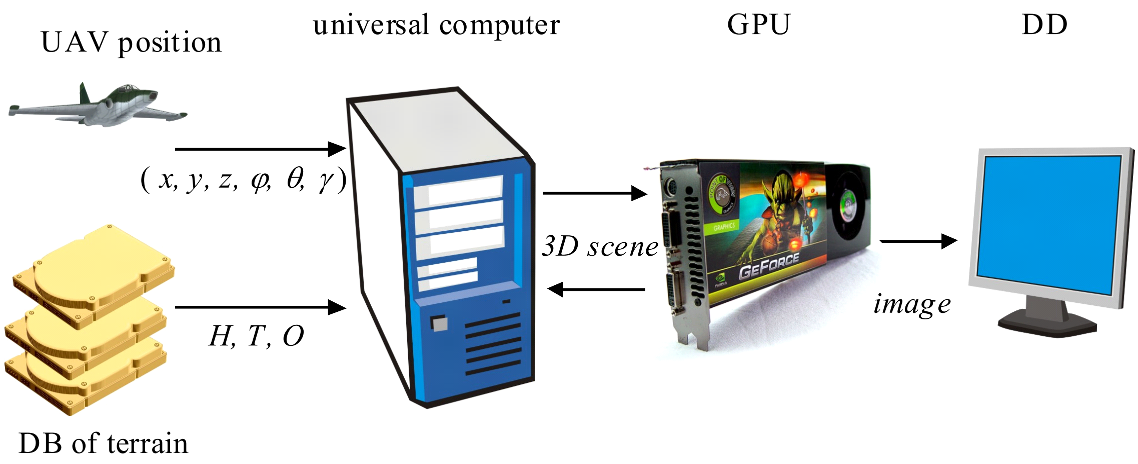


Figure 97. DIS structure

The mainframe performs calculations that do not require a great performance – "slow" calculations:

- processing of information on the position of objects in the scene;
- processing of information about changes in illumination (position of light sources);
- scanning the scene space;
- loading the data to GPU storage devices;
- implementation of the visualization system operator interface.

Partially processed data are then fed to the GPU where calculations are made requiring high performance ("quick" calculations). The frame image formed in the GPU is presented on the display device.

✓ Requirements to modern visualization systems

To obtain a realistic dynamically changing picture the visualization system must meet several requirements, which can be formulated when analyzing the human vision mechanism and the work of its nervous system. The perception of a light pulse entering the eye is realized with a delay. When the pulse action ceases a visual sensation retains for approximately 0.14 sec. [Vavilov, 1981].

In order to achieve a smooth motion during the scene synthesis it is necessary to form at least 30 frames per second [Babenko, 1984], [Krasovsky, 1995]. The output of generated images is displayed mainly on the dot-matrix output device. The image appears as a set of discreet elements called pixels (from 'picture element' in English). The size of pixels as well as their number is fixed. The richer the content of the scene is the greater is the amount of information contained in it, i.e. the higher the realism of the displayed scene is the greater the number of elements in the image must be for its full transmission.

The ability of a photorecording device to integrate a signal in space and time improves the signal to noise ratio [Hall, 1979]. When watching a television image with the fixed number of elements in the frame the eye can realize space integration at the stage of observing the image obtained. And since the vision inertia time is longer than the demonstration time of a single frame a time integration is implemented. Human vision is characterized by space integration in the angle ~ 0,5 º [Hall, 1979], [Lloyd, 1975] and time integration within the vision inertia time is about 0,1 – 0,2 sec. [Rose, 1973], [Leonov, 1977].

Thus on the basis of the analysis of the visual and nervous systems of man we can form the basic requirements to the visualization system operating in real time. The system must provide the image frame in no more than 0.14 sec from the moment of receiving the input signals. To obtain a realistic image the angular resolution of the system must be not less than 0.02°. In addition, the number of synthesized frames per second should be not less than 30.

✓ Methods for obtaining dynamically changing realistic images in real time

DI systems display information about processes or objects in the form of a synthesized image on the display screen. Unlike other systems the sources of input information, here are not physical processes or objects themselves but their mathematical models. In general, the models appear as sets of data, numerical characteristics, parameters, mathematical and logical relationships that reflect the structure, properties, interconnections, and relationships between the object elements as well as between the object and its surroundings.

The method most widely used when synthesizing the image is the one of ray tracing according to which we simulate a geometric path of each light ray being involved in image making. This process is based on the laws of geometrical optics such as the laws of refraction, reflection, linear propagation, etc.

The basic idea of the ray tracing method is reduced to doubling on a computer of all the geometric transformations which each light beam would make on the way the source – the object – the receiver. In this case, we only consider the rays falling to the center of receptors or coming from a limited number of points on the image surface. Based on the law of light rays reversibility one can model a beam path both on the way the object – the image and the reverse one. Therefore, two ways are distinguished: direct and reverse tracing. In direct tracing, we take a point calculated on the image surface as the primary position and simulate the path of a ray from it both to the light source and the receiver of the image. In reverse tracing, a receptor center at the image receiver is taken as the primary position and the path of a ray from it to the object and further from the object on the source of light is simulated.

Advantages of the direct tracing method: a smaller amount of computation needed for synthesis of low and medium complexity scenes; uniform representation of all the objects in the scene. Disadvantages: the need for objects approximation (the need for preliminary triangulation), the complexity of applying textures and calculating the brightness of points on the surface, the difficulty of parallelizing the computation process.

The reverse ray-tracing method is more preferable when processing very complex scenes because it is very easy to make a classification of the scene objects and identify those having appeared in the visible region. Moreover, it becomes possible to handle analytically the figures described without their preliminary triangulation thus avoiding extra transformations and significantly improving the quality of synthesized images.

Performance of today's universal computers has reached rather high level but it is still not enough to perform the tasks of image processing in real time. Therefore, at present DIS development is carried out in two directions: the use of mainframes with increased productivity and the use of specialized image processing systems.

Making realistic images in DIS requires transfer and processing of large volumes of data [Babenko, 1984], [Kovalev, 1988]. The number of natural (mountains, rivers, rugged terrain, forests, cloud layer, etc.) and artificial (houses, roads, factories, airfields, and various buildings) objects in the scene must be sufficiently large in order to achieve the necessary level of realism. This in turn also requires the transmission and processing of large volumes of information that greatly increases the system performance requirements.

Getting realistic images of a scene is impossible without processing the atmospheric and meteorological conditions in accordance with the existing standards. It is also necessary to be able to form the scene image for different seasons in the daytime, in the twilight and at night [Babenko, 1984], [Kovalev, 1991], [Mazurok et al, 1994].

Currently in the vast majority of systems, the isotropic model of the atmospheric layer is used [Mazurok et al, 1994], [Gusyatin, 2000]. This approach greatly simplifies the calculation of the atmosphere transparency and gives good results if there are minor changes in the position head of the observer and the objects in the scene. When these conditions are not met (for example, for the aircraft trajectory), the realism of synthesized images is greatly decreased.

Real time for DIS implies that the image frames change each other every 33 msec or more frequently. Since most of the calculations must be performed during a single frame image, the DIS performance should be about several trillion operations per second. Such operations must be complex like multiplication, division, exponentiation, the computation of trigonometric functions, etc.

To date the cost of systems with universal computers having this capacity, as it is evident from the above review, is quite considerable. Therefore, most of the world's largest DIS producers develop and use specialized computer systems.

Specialized microchips, which are the foundation of all modern three-dimensional graphics cards, have performance high enough for real-time (ATI Radeon R800, Nvidia GT200). These processors contain large quantities of ALU, which can be programmed using CUDA technology from Nvidia or Sream SDK from ATI to solve both the issues of visualization and mathematical problems.

✓ General structure of DIS graphic processor unit

Block diagram of special graphic processor unit is illustrated in Figure 98. GPU contains n separate channels. Each channel can handle objects of different types of scenes such as synthetic mobile and stationary objects, relief, cloud formations, light sources, etc.

Objects of each of these types are stored in its database (DB) and processed separately. With such a division, parallelization of scene processing and optimization of processing each type of object becomes possible.

The sources of data for imaging the earth's surface are digital height maps, which, depending on the situation can be transmitted to the GPU by a satellite in real time, or taken from the database formed in advance. To improve the quality of synthesized pictures the images of the area obtained in the aerospace photography can be drawn on the Earth's surface formed.

Artificial mobile and stationary objects included in the synthesized scene are taken from the database on the ground of the analysis of aerospace images of the terrain we are interested in and the subsequent identification of objects. The description the analysis is given in the following sections.

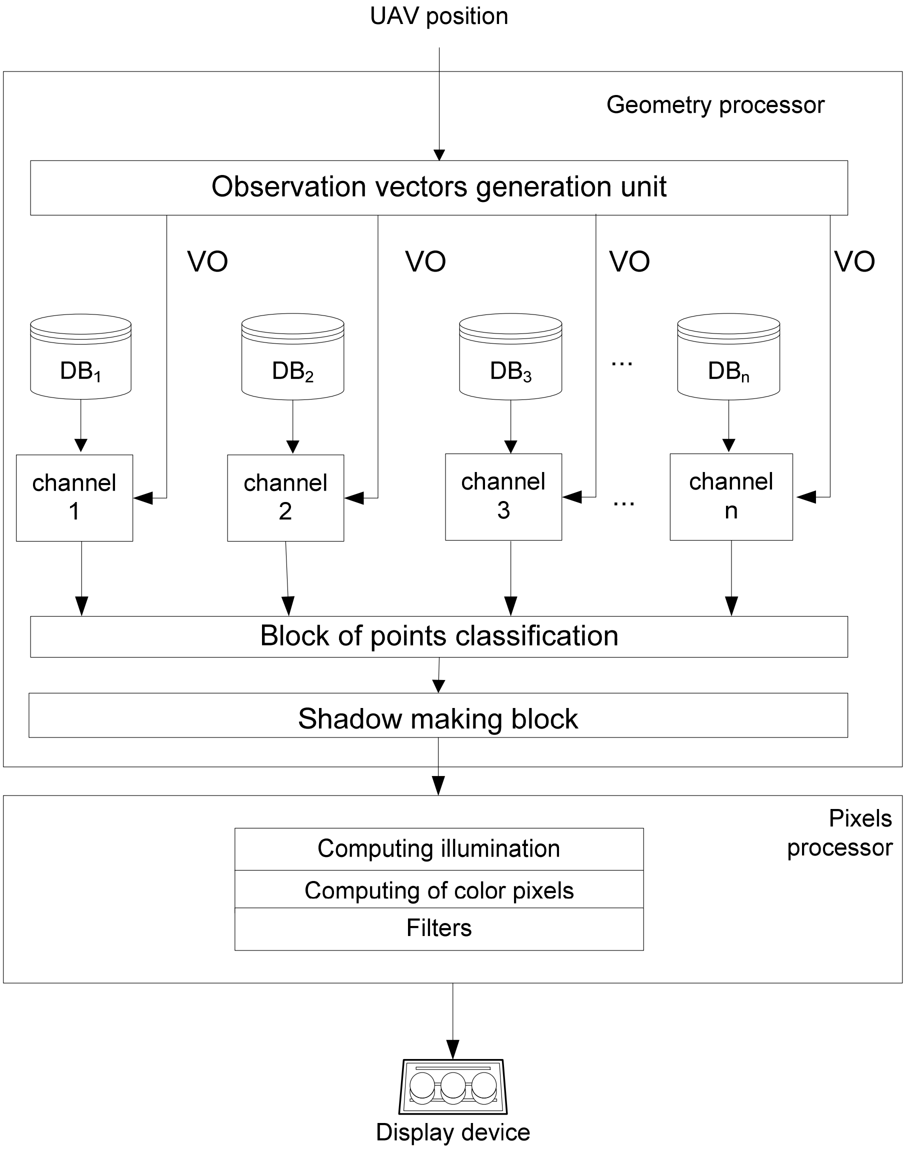


Figure 98. Graphic processing unit structure chart

Special graphic processor unit is a frame pipeline, so data about the scene changes enter DIS for each frame image: a new position of the observer from the channel of communication with the UAV, the location of objects in the scene and light sources because of area image analysis. At certain moments of time the position of the display system corresponding to the next frame image is fixed that completely defines the position of all the vectors of observation (VO) relative to the base coordinate system. Observation vectors generation unit produces all the vectors of observation VO corresponding to the current situation of the monitoring system. VO comes to specialized calculators of all channels and each of them produces rays tracing for a certain type of objects in the scene. Each type of the objects listed above has its own characteristics permitting to optimize each of the specialized calculators for the fastest possible visualization of objects of this type with the lowest hardware cost.

Each of the calculators generates the following output: position (coordinate information), surface color, normal (vector-gradient) to the surface at the given point of intersection, surface properties, transparency, etc. This data comes to the block of points' classification, which sorts them by distance and eliminates the invisible points (not requiring processing). The remaining points come to block the formation of terrain and artificial objects shadows where the information is supplemented by data as for the presence or absence of shading. After the shadow making block the information comes to the pixel processor, where the color of the pixel is determined based on information received [Bugriy, 2004], [Ostroushko, 2004]. In the pixel processor, the image filtering is realized to remove aliasing.

6.3 Analysis of information capability of aerospace Earth monitoring data

6.3.1 Analysis of detail level and overlapping of aerospace Earth images for information capability

Modern satellite systems have reached high precision, which allows for detailed monitoring of the Earth's surface, monitoring of atmospheric processes and even talk about the possibility of building security and timely response systems in emergent situations, based on aerospace survey systems. However, the use of these systems to monitor moving ground targets is highly expensive in terms of performance computing, because it requires the work of the satellite equipment at its maximum power to produce an image at maximum capacity. To effectively detect moving objects (ground transportation) and the danger on the original cosmic survey's images, a prerequisite is to obtain the source satellite imagery with a spatial resolution not less than 500 mm.

Peculiarity of space survey is the presence of atmospheric fronts, which may interfere geo-spatial monitoring. To detect clouds and atmospheric fronts it is necessary to develop a cloud detection algorithm, which will allow proper pruning of these areas of satellite image. Also, this system can be used for automatic monitoring of atmospheric fronts for the weather forecast and floods.

Along with the problem of partial or complete image's overlap by clouds, there is a problem of the observed images object deformation on the reason of the shadows. This problem, to date, no universal solution, so to solve this problem, uses different methods depending on the tasks and the observed objects.

An important characteristic of aerospace monitoring is the frequency of shooting. To meet the challenges of weather forecasting adopted the optimal time of shooting – 6 hours. Let us try to analyze the required frequency of shooting in order to achieve sufficient information capability of aerospace surveillance, while ensuring the safety of industrial facilities. GeoEye-1 satellite, owned by GeoEye Inc. (www.satimagingcorp.com), has the most modern parameters, to date, as satellite that used for civilian purposes Earth monitoring. For example, let us take parameters of GeoEye-1 satellite to calculate the frequency of shooting necessary to monitor moving ground targets, while ensuring safety at sites such industrial facility, the pipeline. According to the organization GeoEye Inc., satellite, GeoEye-1 for one image can cover a square with a side of 15.2 km. As a cross-country ground, transportation will take a main battle tank T-72. Speed of the T-72 cross country up to 45 km per hour. If we assume that the protected area is located in the center of the image, then we can conclude that it is necessary to capture the motion of the object before it reaches the center of the territory. Thus, it is necessary to photograph an object at least once while he covered a distance of 7.5 km (minimum for perpendicular directions). At a speed of 45 km/h, transport, moving on cross country, covers this distance in 10 minutes. It is therefore necessary to shoot the protected object and the contiguous territory frequently not less than once every 10 minutes. Given the characteristics of satellites, alleged by GeoEye, low-latency capture the same territory is 2 hours. Thus, we can conclude that the frequency of shooting from a satellite – not enough, so images from the satellite can be used only for monitoring the construction of buildings on the protected area.

Analysis of methods and technologies for Aerospace Surveillance shows that monitoring of large industrial facilities could more efficiently be conducted with the help of UAVs. However, the best monitoring results can be achieved by combining the two systems of monitoring: global monitoring of protected objects from satellites, and clarify the situation in the emergency areas, as well as those at heightened risk, with the help of the UAV.

6.3.2 Detection of clouds and fire based on the model of active contours

Images obtained by the means of aerospace vehicles are fuzzy. Weak contrast, usually caused by a wide range of reproducible brightness, often combined with non-linear characteristics of transmission levels. The nature of the variation of pixel brightness from minimum to maximum value also affects the image quality. Therefore, signs of objects can be distorted and not well identifiable. Correction of brightness palette significantly improves image quality. Contrast the original image will improve the efficiency of algorithms for segmentation and clipping of clouds in the image, so you first need to increase the contrast.

The clouds have a fractal structure because of their image on the photograph will have an arbitrary shape with partially or completely blurred. The most effective method to extract the boundary of arbitrary shape objects is a method of active contour. Problems of this kind have been solved by the authors in the measurement of intraocular pressure [Bilous et al, 2008].

Applying active contour method, it is assumed that the desired boundary in the image is a smooth curve on a two-dimensional image. The initial data is given an initial approximation to the boundary in the form of a closed curve, which does not necessarily correspond to the actual situation of the border, but close to it. This curve relies flexible and extensible, and under the influence of external forces, it is deformed and shifted to best meet the boundary.

Usually, when solving the problem of detection limits, on the original image are determined by the point of maximum gradient, which then bind to the border. The method of active contours using fit initially continuous closed line to the position of maximum gradient. The final state corresponds to the achievement of minimum potential energy of the model line. External force, under the action of which the line is deformed and displaced, depends on the source image. Inner strength, impeding change, is determined by the properties of the model. This inner strength, preventing too sharp bends strings, performs the same role that image smoothing before computing the gradient: reducing the influence of small-scale noise. This fits the usual force of elasticity [Soifer, 2003].

The main disadvantage of the active contour method is the need of the initial boundary approximation and big computational cost, which value depends on the accuracy of the initial approximation provided. In addition, a problem of forcing the premature stop of local minimal approach when the initial approximation of the set is not very accurate [Terzopoulos, 1987]. Solving the problem of cloud detection borders of averaged brightness regions, that respect to clouds and thus have maximum brightness value within an image, can be used as initial borders. Thus, the problem reduces to the allocation of closed regions with high brightness and subsequent refinement of the cloud on the given image.

Promising solution to this issue is the method of "Magic wand". Method "Magic wand" it was the first interactive segmentation methods, however, it can be successfully used to work in automatic mode, if properly set the initial point of the object [Davidov, 2007]. To use this method indicates one or more starting points of the object, and the algorithm selects the neighboring pixels with similar color and adds selection to the object. To assess the "similarity" is defined by thresholds distance between colors. At the same time, in the region "similar" colors are allocated only connected pixels.

The method works effectively for the allocation of sufficient monotone in color objects. With severe color variations, accurate distinguishing of objects against the background by the use of this algorithm is impossible. If the threshold is too low, sensitivity cannot devote a large portion of the object. Increasing the threshold leads to the fact that the selection "takes place" outside the facility. In the case of spotted an object or blurred boundaries between the background and the object of the algorithm is practically helpless [Soifer, 2003].

For use "Magic wand" require an initial reference point, and install them is problematic because of the possibility of covering the territory of the rented snow. The most effective is the segmentation of images on objects using the method of search boundaries based on the gradient and the subsequent use of the above methods.

Allocation limits based on the gradient is one of the easiest ways to select boundaries. Data is based on a spatial differentiation of brightness functions [Soifer, 1996]. For the two-dimensional intensity function A (x, y) differences in the directions x and y are recorded partial ∂ A (x, y) / ∂ x and ∂ A (x, y) / ∂ y, which are proportional to the velocity change of brightness in the respective directions.

The disadvantage of the algorithm is that it passes borders with small differences of brightness and includes a number of border parts of the image containing large changes of brightness. Noise source image will contaminate the filtered image, since it considers only the boundary points [Davidov, 2007].

Original image of aerospace survey carries a large number of small objects, which are not informative because of cutting off the clouds. Using of Gaussian smoothing filter will provide an opportunity to reduce the influence of the above objects in the original image. To restore the cloud, fuzzy Gaussian filtering, as well as to improve the clarity of the boundaries proposed to subsequent increase of contrast methods of linear contrast. Thus, for solving the problem of cloud pruning on the image the following algorithm of aerospace survey will be effective.

At the first stage is filtering by a Gaussian to smooth the boundaries of clouds.

At the second step of use contrast modification of images, this technique will provide a more accurate light cloud and smoke.

The third stage performs grading images for the separation of images into separate zones. For maximum accuracy and garbage collection, grading algorithm Sobel is used. The algorithm uses Sobel eight counts of brightness in the vicinity of the central point, (homepages.inf.ed.ac.uk).

$$G(x,y) = \sqrt{Gx_{x,y}^2 + Gy_{x,y}^2} \ ,$$

$$G(x,y) \cong \left| Gx_{x,y} + Gy_{x,y} \right| ,$$

$$Gx_{x,y} = \left[A_{x-1,y-1} + 2A_{x-1,y} + A_{x-1,y+1} \right] - \left[A_{x+1,y-1} + 2A_{x+1,y} + A_{x+1,y+1} \right],$$

$$Gy_{x,y} = \left[A_{x-1,y-1} + 2A_{x,y-1} + A_{x+1,y-1} \right] - \left[A_{x-1,y+1} + 2A_{x,y+1} + A_{x+1,y+1} \right].$$

G (x, y) – brightness values point at the coordinates x, y after the operation of the gradient;

$Gx_{x,y}$ – intermediate value of the gradient along the x parallels;

$Gy_{x,y}$ – intermediate value of the gradient along the y parallels;

A(x,y) – brightness values in the point of coordinates x, y;

The fourth step is the union of similar images in the zone of brightness using the algorithm "magic wand". Last, the final stage, divides the image into zones of high brightness (clouds), and the zone of low brightness (of the Earth's surface). Because of this algorithm, we obtain an image, divided into zones of high brightness and low brightness areas. Zones of high brightness may represent clouds or smoke. To determine whether clouds of smoke have been detected it is necessary to analyze the shape of clouds.

6.4 Object detection at areas conducive to accident

6.4.1 Aerospace image analysis for object appearance and disappearance detection

Object detection and tracking is one of the most important parts of the external surveillance of areas conducive to accidents. Tasks range, which requires prompt object detection that appear to be exposed to danger or present danger for the area of appearance, allows avoiding the emergent situations of local and global levels, is very wide. It includes detection of heavy transport and moving equipment above an underground gas pipelines, spontaneous building of dwellings and production structures near-by ground or above the underground chemical or explosive substance storages, illegal garbage dumping and many other tasks.

Large amount of different methods of aerospace object detection and tracking has been offered up to present. The methods of object detection based on certain characteristic features were developed by [Mayer, 1999] and [Gerke and Heipke, 2001]. Basic problem, attended with the use of the methods of this group is absence of temporal data utilization, simplifying detection of new object, and presence and position translation of shadows aside object data, that complicates introduction of stable features.

The methods of finding out appearance of transportable objects are offered based on spatiotemporal aerial and space image photigrammetric change analysis. Objects detection is carried out by forming background object and earth surface map, i.e. steady background. Appearance of objects is revealed by comparison of new images against the background map [Lei et al, 2008], [Hinz, 2004], [Grabner, 2008]. The basic lack of this approach is also influence of shadow cast and earthly surface illumination changes. In the case of satellite image analysis, this limits the application of the plain spatiotemporal approach out to global change detection that could be vegetation of the fields, finding out submergences, large fire spots as well as large buildings and others.

However development of methodology of shadow casting modeling, detection and removal applied to aerospace images makes possible the detection of small transport objects after this kind of preprocessing as shown by [Zhao and Nevatia, 2003] and [Hinz, 2004]. Methods of searching for transportable objects allow determining the size and the speed of objects and to track vehicles but require high frequency update of images and high resolution, which is not always possibly in the area of monitoring of large territories and is impossible for the satellite monitoring.

The other effective approach of object detection, no matter new or old they are, is application of photogrammetry for calculation of three-dimensional model of the Earth's surface [Paparoditis, 1988], [Nakagawa et al, 2002]. This approach is widely used in the area of in GIS map reconstruction. Reconstruction of 3D information is possible because of the directional flight of a satellite or unmanned aerial vehicles above the inspected territory and overlapping of neighbor images, which allows utilization of the pair/triples of neighbor images for point-to-point matching and reconstruction of 3D point coordinates of the territory surface (Figure 99). The separation of shadow is considerably simplified thus. Also new important information appears for the analysis – not only plane sizes but also height data of objects is detected. The problem of approach is low speed of 3D data recovery that eliminates the use of approach in the conditions of time limitations while monitoring emergency objects.

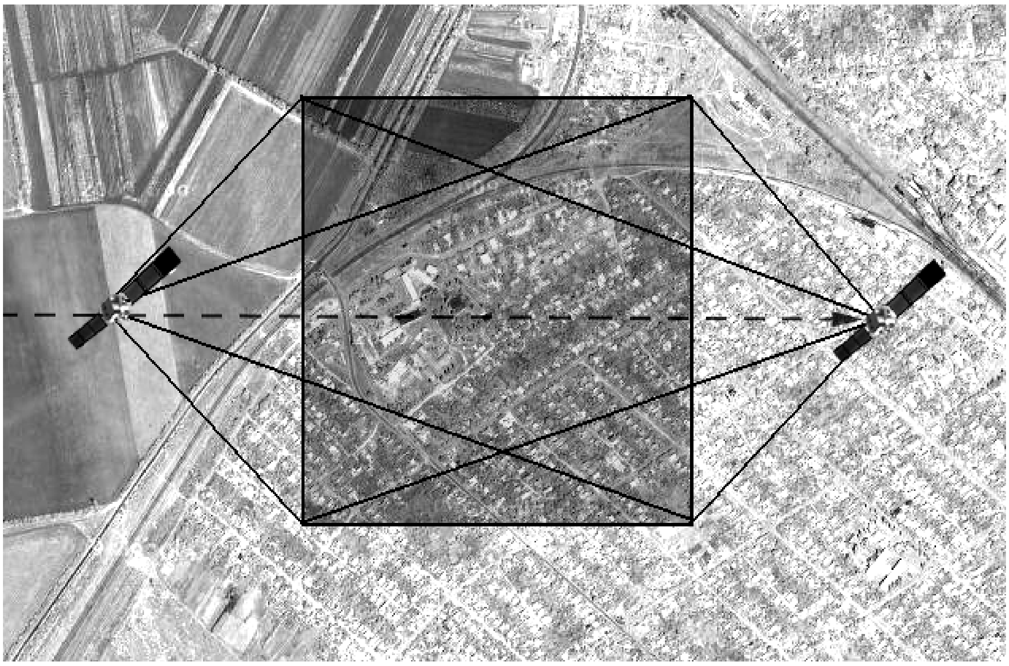


Figure 99. The overlapping of a terrain images made from the moving satellite

One of the most effective approaches to decrease temporal expenses at 3D data recovery is to decrease the amount of the compared points by the means of building of a representative set of characteristic points [Mikolajczyk and Schmid, 2002], [Lazebnik, 2005]. This approach is utilized mainly for stereometry surface recovery in the case when pair or sets of images are presented at different scales or are subject to different project transformations. However the application of 3D data recovery on the horizontally translated pair of images is obviously also possibly and will provide reducing of calculation expenses on point matching and 3D data calculation.

Application of the combined approach, utilizing spatiotemporal presentation of the recovered Earth's 3D surface data is offered in this work. Utilizing scale space representations of images is thus offered for renewal of three-dimensional information for the considerable decrease of point match operation amount by the means of selection of the most informative image points and regions. Application of scale space presentations has shown good results at different problem areas, where the selection of informative/characteristic points is providing high decision-making accuracy at considerably increased performance [Bilous and Kobzar, 2008], [Kobzar, 2008], [Bilous et al, 2008].

A task of 3D data reconstruction by the pair of aerospace images is less complex by calculation expenses as compared to similar tasks from other areas; where it is necessary to take into account not only horizontal translation of points at the change of viewpoint but also vertical translations, conditioned perspective distortions and rotation. I.e. the task terrain 3D surface map reconstruction consist in detection of matching points at both images, achieved from different view points, and the calculation of the third coordinate having the relative displacement of such points (Figure 100).

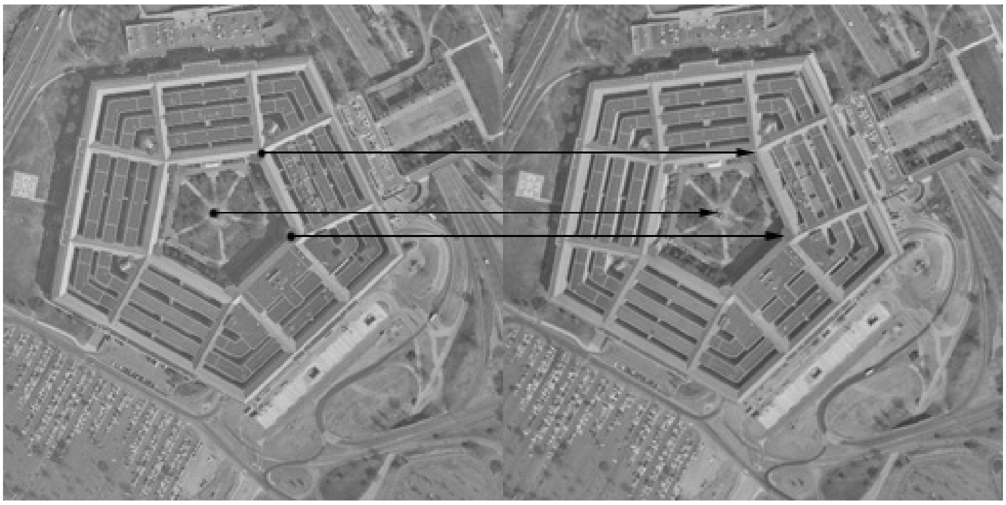


Figure 100. Point matching of the stereo pair of satellite images

The offered application of scale space representation of images for the decrease of matching point amount supposes the successive calculation of curvature function at different levels of scale in image rows while they are scanned in any order. Thus, by submitting intensity as the function of the point translation on a horizontal line of the surface, we get a continuous curve defined by equation in an obvious form y = y(x), for which curvature is calculated by a formula:

$$k = \left| \ddot{y} / (1+\dot{y}^2)^{3/2} \right|.$$

Real digital images acquired from satellites could be used to produce discrete curves only. Such discrete curves in general case can contain hollows and peaks. As differential description, curvature is very unsteady to discretization and noise and cannot be utilized for finding of characteristic points as it applies to an initial curve. For this reason, representation of curvature is offered at different levels of scale by the means of curvature function convolution with Gaussian kernels of the successively increased width [Linderberg, 1994]:

$$L(x;\sigma) = \int_{\xi=-\infty}^{+\infty} g(\xi;\sigma) f(x-\xi) d\xi \ ,$$

$$g(\xi;\sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\xi^2/2\sigma}$$

where $g : R \times R_+ \setminus \{0\} \to R$ it is the Gaussian kernel.

As a result of tracking curvature zero-crossing points, all the points that do not present considerable intensity changes of initial image rows are eliminated. Introduction of a certain scale threshold level provides selection of the set of characteristic point which influence on the image intensity gradient is considerably even at the high levels of scale. Such points are assuredly represented at two images of the same territory and, in most cases, represent edges of certain objects, even if the point of view changes. I.e. acquisition of characteristic point set makes it possible to recover basic surface height changes and detect objects, located on a certain Earth's surface territory.

Comparing surfaces of certain territory, recovered in different moments of time, it is possible on divergence of intensity and geometrical (height above the locally averaged level of the Earth's surface) difference with high accuracy to expose objects appearance or disappearance of size that would depend only on image resolution.

6.4.2 New approach to m-weight based calculation of curvature function estimation

As a differential characteristic, curvature is very unsteady to discretization and noising of images. Therefore, for providing of acceptable noise immunity at construction scale-space representations the function of curvature must be estimated indirectly.

Next formulation of m-weight of curve by which it is possible to formulate the estimation of curvature and satisfying to the requirement $\lim_{m \to +0} k_m = k$ (requirement for all estimations of curvature) is offered. For the calculation of such m-weight of curve, the points of curve are needed only. The initial geometric curvature estimation method was proposed in [Karkischenko, 1998].

It is proposed to use m-estimation of curvature to build curvature base scale-space representations to increase the rapidness of such representation construction. Proposed m-estimation is given in 2-dimentional parametric form of curve representation because it is easily converted to 1D form of ECG but may have a lot of other application in 2D [Bilous and Kobzar, 2009].

Lemma 1. Let m is neighborhood of point $\gamma(u) = (x(u), y(u))$, into which m-weight of curve given in parametrical view $\Gamma = \{(x(u), y(u) \mid u \in [0, L]\}$ is estimated. Then the following asymptotic formula $v_m = v_m^o + O(m), m \to 0$ takes place, thus v_m^o calculated on the following formula:

$$v_m^o = \left| S_m^o - \bar{S}_m^o \right| / \max(S_m^o, \bar{S}_m^o),$$

where $S_m^o, \bar{S}_m^o$ – are areas of sectors of circumference with a radius Δu :

$$S_m^o = \pi m^2 \angle(\overrightarrow{v^-}(u), \overrightarrow{v^+}(u)) / 360°$$
$$\overline{S_m^o} = \pi m^2 \angle(\overrightarrow{v^+}(u), \overrightarrow{v^-}(u)) / 360°$$

where $\overrightarrow{v^-}(u)$ is a vector connecting the point of curve $\gamma(u) = (x(u), y(u))$ and $\gamma^-(u) = (x(u-m), y(u-m))$ and $\overrightarrow{v^+}(u)$ is a vector connecting the point of curve $\gamma(u) = (x(u), y(u))$ and $\gamma^+(u) = (x(u+m), y(u+m))$ (Figure 101).

Lemma 2. Let $m = \Delta u$ is Neighborhood of point into which curvature of curve Γ given in parametrical view $\Gamma = \{(x(u), y(u) \mid u \in [0, L]\}$ is estimated. Then the following asymptotic formula takes place:

$$k = 3\pi v_m^o / (4m) + O_m^o(m),\ m \to 0.$$

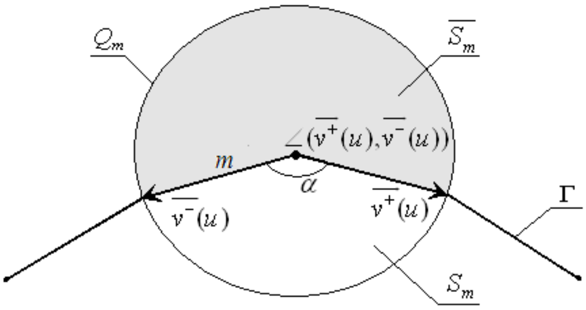


Figure 101. Neighborhood for m-rate of function of curvature of contour curve

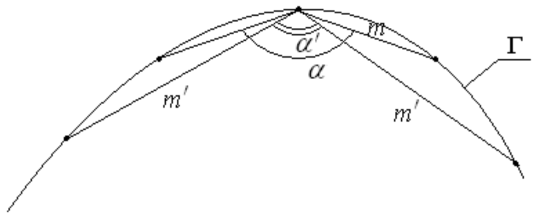


Figure 102. Neighborhood of sectoral m-rate of function of curvature of contour curve

Lemma 3. It is possible to assert that it is possible to obtain finding of estimation k_m^o with required accuracy at estimating of curvature to sectoral method by introduction of the following limitation:

$$v_m^o < \xi_m^o,$$

where ξ_m^o is some cut-off of m-weight.

Lemma 3 has a double value for the construction of SSR of curvature and proper handle. At first, changing of neighborhood m it is possible to check accuracy of m-rate of curvature in points of contour curve.

Secondly, as at $k \to 0$ error of sectoral rate $O_m^o \to 0$, it is possible to assert that for finding of transit points of curvature through a zero it is possible initially to take maximum size neighborhood m. In addition, for points with small curvature it is possible to take small neighborhoods m' (Figure 102) with guaranties here finding of estimation of curvature with given accuracy and fast-acting.

6.4.3 Adaptive discretization technique for fast curvature scale space reconstruction

To achieve rapid estimation of curvature with the given accuracy it is possible to connect m-neighborhood and step Δu of curve $\Gamma = \{(x(u), y(u) \mid u \in [0, L]\}$ discretization.

A curve is represented by the set of points $\Lambda = \{(x_i, y_i) \mid i \in [0, M_u]\}, M_u = L_u / \Delta u$, distance between any two points λ_i and λ_{i+1} of set Λ is equal thus to

$$\|\lambda_i - \lambda_{i+1}\| = \Delta u = 1,$$

where $\|\cdot\|$ is the Euclidean metrics.

For the construction of discrete scale-space representation with invariant to scaling, ration the amount M_u of points of curve Λ with losing or adding new points to some general value M, I.e. lead $|\Lambda| = M$. Thus, accept $L = M$ as new length of contour.

In addition, choose the step of discretization on a scale $\Delta\sigma$. Then representation of points of curve Λ on a scale σ, such that $\sigma(\mathrm{mod})\Delta\sigma \equiv 0$ (or $\sigma = \Delta\sigma \cdot j, j \in N$), can be easy found on the following formula of the discrete approaching of Gaussian convolution:

$$x_i^\sigma = \sum_{j=-[M/2]}^{j<[M/2]} x_{i-j} g(i-j,\sigma), \quad y_i^\sigma = \sum_{j=-[M/2]}^{j<[M/2]} y_{i-j} g(i-j,\sigma), \quad g(x;\sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-x^2/2\sigma} \qquad (4.1)$$

Thus, $x_{i-j}^0 = x_{i-j}$, $y_{i-j}^0 = y_{i-j}$.

For the construction of SSR of curvature, it is necessary to find the values of function of curvature in all M points of discrete curve for each of N_σ levels of scale. The calculation of position of some λ_i point on the levels of scale $\sigma > 0$ according (4.1) requires $O(M^2)$ operations, while due to rule 3σ only points in neighborhood 3σ bring in a meaningful contribution to the sums (4.1). It is possible to transform formulas (4.1) to the following kind for considerable decreasing of amount of operations that is ordinary practice at the use of Gaussian convolution in discrete spaces:

$$x_i^\sigma = \sum_{j=-[3\sigma]}^{j<[3\sigma]} x_{i-j} g(i-j,\sigma), \quad y_i^\sigma = \sum_{j=-[3\sigma]}^{j<[3\sigma]} y_{i-j} g(i-j,\sigma) \qquad (4.2)$$

Thus, at the maximal amount of levels on which the calculation of curve points is needed (and this amount in practice depends on how quickly on a curve will be not a single zero transit point of curvature) $\sigma = M/3$ for some $\Delta\sigma$ middle complication of calculations of every level will be order $O(M^2/2)$.

As for the exposure of maximums of SSR of curvature all levels of scale are needed, it is possible to apply another property of Gaussian convolution:

$$(f(x) * g(x,\sigma_1)) * g(x,\sigma_2) = f(x) * g(x, \sqrt{\sigma_1^2 + \sigma_2^2}).$$

Let σ_i is a level of scale for which the discrete curve $\Lambda^{\sigma(i)}$ of which is already calculated. For the calculation of curve $\Lambda^{\sigma(i+1)}$ of the following level of scale in accordance with chosen $\Delta\sigma$ it is necessary to execute Gaussian convolution of curve $\Lambda^{\sigma(i)}$ with a kernel σ^+ size of which can be calculated on the following simple formula:

$$\sigma^+ = \sqrt{\sigma_{i+1}^2 - \sigma_i^2}.$$

Thus for curve evolving at a certain range of scales required for discrete scale space reconstruction the following iterative calculation scheme may be used

$$\begin{gathered} \Lambda^0(u) = \Lambda(u) \\ \Lambda^{\Delta\sigma}(u) = \Lambda^0(u) * g(u,\Delta\sigma) \\ \dots \\ \Lambda^\sigma(u) = \Lambda^{\sigma-\Delta\sigma}(u) * g(u, \sqrt{\sigma^2 + (\sigma-\Delta\sigma)^2}) \end{gathered},$$

It allows to make average complication of calculations of every level to the order of $O(M \cdot \sqrt{M})$ at the maximal amount of levels (till to $\sigma = L/3$) on which the calculation of curve points is needed.

After the calculation of curve of every scale, directly the estimation of curvature can be expected by proposed sectoral m-weight of curve. Corner α_i^σ for the point $\lambda_i \in \Lambda^\sigma$ of curve at some level of scale $\sigma = \Delta\sigma \cdot j, j \in N$ it is possible to calculate as:

$$\alpha_i^\sigma = \arccos\left((x_i^\sigma - x_{i-1}^\sigma)(x_{i+1}^\sigma - x_i^\sigma) + (y_i^\sigma - y_{i-1}^\sigma)(y_{i+1}^\sigma - y_i^\sigma)\right) / |\overrightarrow{v_+^\sigma}||\overrightarrow{v_-^\sigma}|,$$

where $\overrightarrow{v_+^\sigma} = \lambda_{i+1}^\sigma - \lambda_i^\sigma$; $\overrightarrow{v_-^\sigma} = \lambda_i^\sigma - \lambda_{i-1}^\sigma$.

After this, it is simple to receive sectoral m- weight v_m^o and estimation of curvature k_m^o of point λ_i^σ due to formulas (4.1) and (4.2).

Considering such curvature scale space reconstruction approach a problem of optimal m level selection, and discretization step at $\Delta\sigma$ scale utilization for achievement of satisfactory curvature scale space estimation. Obviously, there is no right ahead solution of this task as it is hard to estimate how exact the CSS estimation should be for different application. However, proposed scale space construction approach based on the sectoral m-estimation of curvature function allows finding a tradeoff between estimation accuracy and robustness by changing the m -neighborhood and the power of M set.

The next method of construction of images of SSR of curvature based on adaptive binary simplification of contour curve and calculation of sectoral m- weight neighborhood m of which can be related to the level of simplification of curve is offered.

For finding of estimations of curvature of contour function at all levels of scale on the first step it is required to ration the contour curve so that length of discrete curve and amount of points were multiple 2: $L = M = 2^N, \Delta u = 1$.

Definition. Binary simplification of discrete curve is an operation of exception of every second point from set Λ^σ of points of curve at some level of σ scale. In other words set Λ_η^σ of points of simplified curve can be found as:

$$\Lambda_\eta^\sigma = \{\lambda_i \in \Lambda_{\eta-1}^\sigma \mid i = [0, M_{\eta-1}], i(\mathrm{mod})2 \equiv 0\}$$
$$\Lambda_0^\sigma = \Lambda^\sigma \qquad (4.3)$$

Definition. The level of simplification η of discrete curve Λ_η^σ on a scale σ is equal to the amount of simplifications in accordance with (4.3) created for receiving of curve Λ_η^σ from a curve Λ at all levels of scale $\sigma' < \sigma$. Obviously $\eta \le \log_2(M)$.

It is suggested to calculate the curves of the followings levels depending on the level of simplification, using for development of curve only those points, which was saved after binary simplifications on previous levels:

$$x_i^\sigma = \sum_{i \in \{\lambda_j \in \Lambda_\eta^{\sigma-\Delta\sigma} \mid \|x_i^{\sigma-\Delta\sigma} - x_j\| < 3\sigma^+\}} x \cdot g(\| x - x_i^{\sigma-\Delta\sigma} \|, \sigma^+), \qquad (4.4)$$

where x_i^σ is the first coordinate of located point $\lambda_i \in \Lambda_\eta^\sigma$ of curve on a scale σ at the level of simplification η ;

σ^+ – kernel of convolution which necessary for receiving of curve Λ_η^σ from curve $\Lambda_\eta^{\sigma-\Delta\sigma}$ calculated due to formula (4.3).

By analogy y_i^σ can be calculated.

We will enter cut-off ξ_m^o for m- weight of curves. We will simplify a curve each time during iterative development in accordance with a formula (4.4), when maximal weight of points of curve appears below than some cut-off:

$$\max_{\lambda_i \in \Lambda_\eta^\sigma} v_m^o(\lambda_i) < \xi_m^o + \Delta\xi \ ,$$

where $v_m^o(\lambda_i)$ is a value of m- weight of curve Λ_η^σ in a point λ_i;

$\Delta\xi$ – additional element which guarantees implementation of condition $v_m^o(\lambda_i') < \xi_m^o, \lambda_i' \in \Lambda_{\eta+\Delta\eta}^{\sigma+\Delta\sigma}$ on the curve of next scale and level of simplification.

Due to Lemma 3 it guarantees the calculation of estimation of curvature by sectoral m- weight of curve Λ_η^σ with given accuracy level. Actually, simplification of curve takes place when a curve is smoothed out so that the maximal estimation of curvature $v_m^o(\lambda), \lambda \in \Lambda_\eta^\sigma$ of points of discrete curve decreases to some a priory given level $C \cdot \xi_m^o$ and simplification of curve does not bring to the increase of maximal estimation of curve curvature $v_m^o(\lambda), \lambda \in \Lambda_{\eta+\Delta\eta}^{\sigma+\Delta\sigma}$ higher than some other level $C \cdot (\xi_m^o + \Delta\xi)$, $C = 4m/3\pi$ on next scale.

Application of method of adaptive discretization allows to considerably accelerate the process of construction of scale-space representations, that will influence on increasing of velocity of as process of recognition of ECG elements and determination of their boundaries so and the process of forming of average signal which was offered in this work.

Application of developed adaptive discretization technique provides considerable robustness improvement of scale space representation calculation, which by the proposed object detection approach leads to the increase of 3D surface data reconstruction performance, and characteristic point set acquisition process speed caused by fast point matching. It also provides stable set of feature points for accurate object detection.

Thus, it is proposed to analyze changes of 3D Earth's surface characteristic point model achieved by the means of scale space representation of image rows for object appearance and disappearance detection. Whereupon the utilization of 3D data provides the increase of detection accuracy eliminating shadow and illumination distortions and considerably increase the performance of 3D data reconstruction and analysis processes by achievement of characteristic point set by the means of developed curve adaptive discreatization technique for fast calculation of curvature scale space.

6.5 The main gas-transport system as object of complex remote space monitoring

The gas-transport industry is one of leading branches of economy for many countries of the world. The main pipelines of a high pressure, which are a part of gas-transport systems (GTS) of the modern states, represent the difficult territorially distributed network. GTS includes hundreds objects of the raised danger. Transported natural gas is dangerous explosive, large failures on a linear part of pipelines of a high pressure and other technological objects can have catastrophic consequences, especially in densely populated areas [Borisenko et al, 2004].

In spite of the fact, that in comparison with other types of transport pipeline transport is one of the most reliable. So there are till 25-30 failures, or to 0,2 failures on 1000 km a year on gas pipelines, for example, Russia annually. Throughout last 5-7 years, this indicator remains rather stable. However, there are a number of the preconditions testifying to presence of growth susceptibility on pipelines of a high pressure.

First, a gas-transport network ageing continues. As last 10-15 years, for example, in the states of the post-Soviet territory rates of new building, major repairs, and reconstruction of gas pipelines essentially lagged behind rates of their ageing. Now to 80% of gas pipelines operation term exceeds 20-25 years that essentially increases risk of occurrence of extreme situations.

Secondly, development of oil and gas deposits on a shelf of the seas, building of sea pipelines, and also their lining in mountain district where earth flows, avalanches, landslips, high waters, flooding are possible. All it causes occurrence of new dangers and threats, demands introduction of the no completely fulfilled and tested technologies.

Thirdly, last 5-7 years on pipelines emergencies, connected with subversive and terrorist actions have increased.

Space image are claimed in many industrial and social-economic branches, but especially steadfast attention to them the enterprises of a fuel-energy complex show, and first of all, the companies which are engaged in pipeline transport of hydrocarbons. For such enterprises are characteristic lengthy lines of pipelines, when only remote methods can prove, as the most operative and effective. For example, the gas-transport system of Ukraine has the general length more than 37 thousand kilometers. Hundreds technological objects are located in different areas of the country on distance from tens to thousand kilometers from each other.

Space image can be used as directly, and in a complex with other data for the automated decision of applied problems.

It is possible to allocate following base directions of complex remote space monitoring GTS objects application:

1). Use as visual replacement of a topographical basis of corridors and platforms of transport and gas storage objects (Figure 103), including territories outside of security zones, to scales 1:5000. Undoubtedly, all enterprises involved in GTS, have executive land shooting of the objects, the vulgar to scales 1:500, linear to scale 1: 2000. However, shooting, as a rule, is limited by a security zone of objects (for the main gas pipelines the security zone is established 100 meters from extreme threads of pipelines in a corridor of communications), that essentially limits information-spatial security at level of the concrete enterprise. Aerospace images of the high permission allow observing the territories located in 2 – 3 kilometer buffer zone round objects of transport and gas storage.

2). Updating of executive objects shooting of gas transport and storage to scale 1:2000. Annually tens new objects are put into operation in the enterprises of gas transport. Not always probably to make executive shooting operatively. Aerospace images of the high permission will allow updating and adding executive shooting of linear GTS objects, in time to spend inventory, certification, and their statement on the cadastral account.

3). Construction of digital district models (DDM) and digital relief models (DRM) of corridors and industrial platforms for GTS objects, for example, gas-compressor stations and adjoining territories (Figure 104) to 2 meters on height. DRM will allow to estimate deformation loadings on a gas pipeline in places of a relief excess, to reveal local falls on district (potentially explosive sites), to calculate zones of electromagnetic visibility of radio relay aerials.

4). The predesigned analysis of the territories intended for commissioning of new GTS objects, to scales 1:5000. Available topographic maps have hopelessly become outdated and do not give the complete information about the territories planned to development. Aerospace images possess a time urgency and spatial accuracy necessary for acceptance of the strategic decision on building of objects. Besides, on their basis carrying out of ecological examination within the limits of design activity for again put in operation of objects is possible.

5). Monitoring of corridors and industrial platforms of GTS objects, and also adjoining territories to scale 1:2000, that will allow to estimate operatively change of a situation in 2-3 kilometer zones from objects of gas transport (put into operation of new GTS objects, building in territory of "accessory manufacturers", changes of natural and landscape conditions). Work can be carried out in semi-automatic, and in some cases in an automatic mode in such program complexes of data processing of remote sensing, as ENVI.

Figure 103. An image from QuickBird satellite - a corridor of the gas main pipeline

Figure 104. An image from Quick Bird satellite - gas-compression station

Besides, following applied problems can be solved with use of remote monitoring means for an estimation of integrity and safety of GTS objects and environment surrounding them:

1). Revealing and mapping of zones of probable occurrence for technogenic failures and accidents, risks estimation:

- in crossing places of existing or projected objects for gas transport and storage with railway and highways, other kinds of pipelines, high-voltage transmission lines and other linear technogenic objects;
- in rapprochement places of existing or projected objects for gas transport and storage with all above listed objects;
- in places of a close arrangement of the large localized technogenic objects: factories, territories of mining operations, places of various kinds household and industrial wastes recycling;
- in places of a close cultural-historical monuments arrangement;

2). Revealing and mapping of zones of probable occurrence of natural accidents, an estimation of risks:

- in crossing places existing or projected objects of transport and storage of gas of natural objects: woods, bogs, the rivers, lakes, reserves, wildlife preserve;

- in crossing places existing or projected objects of gas transport and storage tectonic active geological structures, dynamically intense zones, deep breaks, seismically active areas;
- in crossing places existing or projected gas pipelines of zones for development modern ravine, river, wind, glacial, frost erosion, active development of gravitational processes (landslips, taluses, collapses, avalanches), water logging territories, places of a karsts development, soil slumping.

3). Detection and monitoring of the ecological negative processes proceeding in a security zone of gas pipelines, connected with extraneous sources of pollution and technogenic influence.

4). Revealing and mapping of changes in the geodynamic active zones connected with territories of underground gasholders, based on radar shooting and interferometry. Regular shootings by radar satellites of territories of underground storehouses (depth of occurrence to 1 km, capacity of ten billion cubic meters, pressure 120 atmospheres) will allow estimating motions of a surface to within the first millimeters, both in zones of depressions, and in possible zones of an earth surface rising. Early detection of shift superficial deformations in zones of a pipelines arrangement.

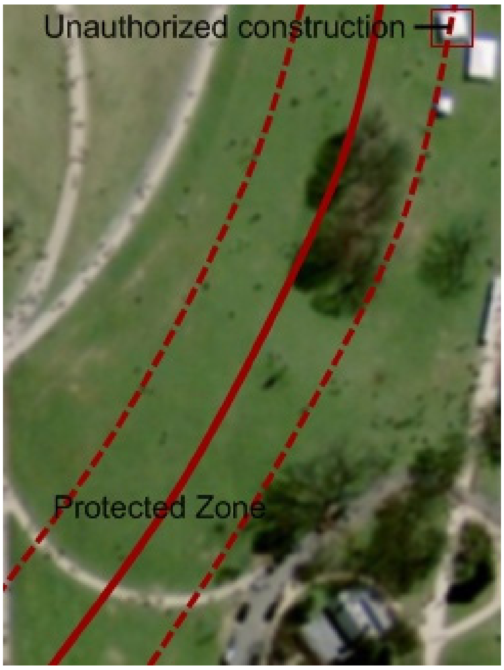


Figure 105. Automatic revealing of not authorised building in a security zone of the main gas pipeline with use of program complex ENVI

5). Monitoring of territories in security zones of the gas-transport enterprise about presence of self-capture zones and not authorized sites building inhabited and industrial targets. Such monitoring will allow stopping in time attempts of illegal use of the earth and if necessary to receive incontestable proofs of certificates of the earth's self-willed capture. Aerospace images occurring at different times (Figure 105) can serve as forcible argument by consideration of claim affairs to violators of security zones of transport gas objects.

As an example, it is possible to result developed and introduced in commercial operation in the Russian Federation geoinformation system of monitoring and forecasting of emergencies on objects of uniform system of gas pipelines of public joint-stock company "Gazprom". This system is supported in an actual condition and serves for the remote watching control over a condition of Russia GTS objects, and for working out of the preventive measures, capable to prevent growth of potential failures and to provide essential reduction of risks and level of emergencies damage at the enterprises of gas-transport branch.

In works [Borisenko et al, 2004], [Borisenko et al, 2008b] the base concept of construction, methods and effective implementers of the complex automated control system (CACS) are offered by gas-transport system of Ukraine. The subsystem of remote space monitoring is in it a case of one of the basic specialized component CACS.

Use of the digital spatial information of complex space monitoring together with a set is standard-help, the passport and factual data collected by systems of land basing (microprocessor systems of modular automatics, industrial control on the basis of modern SCADA, the automation system of operative-dispatch control on the basis of MES [Ponomarev, 2006]) allow to organize highly effective, reliable and qualitative support of decision-making on complex multilevel management of the distributed main gas-transport networks of a high pressure [Borisenko et al, 2008a].

6.6 Discussion

Studies carried out separately by Russian and U.S. experts clearly indicate that we are soon expected to undergo global climate changes. An ever-increasing number of natural disasters and technological accidents clearly confirm this. Therefore, tasks for improving the quality of forecasting, reconnaissance and monitoring the situation with the emergencies are relevant.

Using digital area images allows you to automate the process of monitoring and improving the quality of forecasting and detection of abnormal situations. Digital images obtained by satellites are not always informative because of time and weather factors. The use of UAV allows you to minimize the influence of these factors and improve the information content of images. The fact that UAVs can be used in areas hazardous to human is particularly valuable. It enables you to perform aerial surveillance of highways, pipelines, power lines, to fly in emergent situations, man-made and natural disasters, floods, large-scale fires at industrial enterprises, military depots; aerial photography, environmental and radiation monitoring with the possibility of mapping the extent of radiation (or other) contamination.

Using DIS for solving UAV control problem can serve as emergency monitoring in conditions of limited visibility. A possibility gained by the operator to control the UAV by means of synthesized image eliminates not only the visibility limitation but also the influence of meteorological conditions on aircraft control. This helps to improve considerably the quality of management, reduce the risk of accidents when monitoring natural disasters or man-made accidents.

The proposed methods of image processing enables to estimate the information content of the images obtained and to carry out restoration of the height of objects represented on the photo, with the aim of their further recognition and registration of movement within the monitoring zone. The proposed method uses the representation of the brightness function curvature at different levels of scale to increase the original data resistance to noise and sampling. The approach proposed for calculating the curvature of the curve is distinguished by improved performance and availability of monitoring the accuracy of curvature estimation at the points of the contour. Together with the method of adaptive sampling for fast curvature scale space representation it allows to solve the task of restoring the height of objects in limited time mode with high accuracy.

As area of effective practical use of the offered synergy of technologies, mathematical methods and algorithms of digital images processing the main gas pipelines of a high pressure which are objects with a high risk level of emergencies occurrence are considered.

PART II. Theoretical Aspects

7 Elaboration of Geoinformation Regional Monitoring Environmental System ("GERMES-I") Enriched by Artificial Intelligence Instruments

7.1 Introduction

Environmental problems as well as environmental and anthropogenic collisions in basins of Mediterranean Sea and Black Sea are closely linked with industrial, transport and agricultural activity and require considerable attention. We believe that one of the first protection and prevention measures to be taken today, while solving environmental tasks, is to combine the artificial intelligence (AI) means of information processing and decision making in uncertain situations – with new technologies in GIS, with new monitoring methods, aerospace remote sensing (RS) and some other techniques [Krissilov, 1999], [Krissilov et al, 2001], [Shutko et al, 1998, 2007], [Stepanov et al, 2001, 2003], etc. We call such intelligent system – "GERMES-I" (GEoinformation Regional Monitoring Environmental System-Intellectual).

It must be noted that talking risks assessments we mean at least 5-level interpretation or definition:

- 1st level: theory of different kinds of risks: risk structure, genesis of risk, the steps of danger development, theory of risk measurement, mutual influence and feedback links between different kinds of factors so on; it is new methodological and theoretical area now;
- 2nd level: concrete experimental studies of risks in condition of natural and man-made disasters;
- 3rd level: remote sensing application for risks determination by means of various sources (microwave radiometry, in particular, allows to obtain by practically direct way the risks assessments);
- 4th level: risks assessments in regional and/or national scale by means of advanced Geoinformation Environmental Monitoring System (examples – for Black Sea Region, for Aral Sea, GIMS for Bulgaria, others);
- 5th level: creation of state-of-the art technological monitor complexes for strategic determination of risks of hunger, another dangerous factors – throughout the World (e. g. by developing aerospace complex "RiceSAT" (rice satellite) with partners in Asia, America, etc.)

The problem of recognition and/or interpretation of monitored data occur in a lot of environmental tasks. All of them practically have similar characteristics: the objects and processes are complicated; features are both quantitative and qualitative, have various natures; one must take into account a lot of data, relations between them, etc. In addition, at the same time, we must use all this information for diagnostics, decision-making, and prediction; we have no other [Krissilov, 1999]. It is understood that without using certain arsenal of effective AI methods and algorithms we cannot make valid decision and solve agrotechnical, administrative, navigational, and environmental problems

[Stepanov et al, 2001]. Such tasks as measuring, recognition, estimation, considerable analysis and description of complicated objects and processes, etc. – may be solved by means of certain group of classification and evaluation methods, intellectual methods.

Therefore, planning and managing bodies must have full-scale, deep, and complex multilateral ecological monitoring.

At the same time any country and the Black Sea Region as a whole does not have general (common) big monitoring system. We have no interdisciplinary Expert-and-Analytical Center for collection and performance of OLAP (On-Line Analytical Processing) based on all information in order to save previous data, define and forecast the accidents and timely inform various users and agents in countries of the Region. There is no united and systemized database, as well as knowledge about links between parameters and probable results of its interaction. Almost all that we have now needs to be updated, is expensive, and cannot bring fast reaction.

Various forms of this issue were discussed in the framework of some events, some conferences of Balkan Environmental Association (BEnA), KDS events ("Knowledge – Dialog – Solution" Conferences and Schools), Danube and Black Sea round tables, GMES events, other scientific and public meetings, and included in certain Conclusions, etc. So far, administrative bodies in the basin and scientific organizations have unfortunately failed to get a grip on such Center creation. It is understood, that this work does not need a pot of gold, and new "Black Sea" Euro Region territorial unit can become both right way and right instrument for that.

We believe, that regional environmentally oriented monitoring and intellectual system GERMES–I can become the first step along this road (creation of such Center). Its general outlines and important special components are described in this chapter. As it can be seen, that in its algorithmic terms – a good deal of attention is paid to intellectual instruments for aggregation, evaluation, comparison, decision rules, – using original effective means [Krissilov, 1962], [Krissilov, 1984], [Krissilov et al, 2000], [Krissilov et al, 2001], [Krissilov A. and V. Krissilov, 2005], [Popov, 1971], [Krissilov V. et al, 1998], [Krissilov V. and A. Krissilov, 2000]. In measurement terms, major results of the long-term study and practical experience are presented [Shutko, 1982, 1987], [Shutko et al, 1995, 1998, 2007], [Stepanov et al, 2003], [Krapivin et al, 1996] and others.

7.2 Brief system description

The Black Sea and Mediterranean Sea basins and the coastal area of the region are a zone of extensive economic development with high capacity sea- and river ports, large industrial enterprises, heavy and growing traffic, owner-leased resource extraction on the shelf, and expanding recreational and agricultural uses of the shores.

These permanently acting factors, existing in a very complicated and vulnerable environment, require coordinated efforts and effective integrated management both of maritime and land economic activities. So the integrated environmental monitoring system can be seen as a tool of such management and risks prevention.

The following characteristics shall be monitored:

– environmental: direct survey and assessment of temporal and spatial data on oil pollution (where, what, how much, the dynamics); monitoring of all coastal areas as regards the outflow of waste waters and mapping of affected zones; detection of bilge water discharges; gas concentration surveys carried out above urban settlements and industrial zones, registering their wind drift and the nature of pollution; assessment of water turbidity, i.e. of practically all mineral suspended matter contained in water; assessment of chlorophyll concentrations (in other words – a self-cleaning capacity) in specific water reservoirs and water areas over the whole monitored basin, and a number of others;

- hydrological and meteorological: water temperature, presence and thickness of ice, wave characteristics, situation in straits, assessment of shore abrasion;
- land: assessment of water seepage in canals and through dams, the degree of soil pollution and salinity, the amount of swamping, assessment of crops conditions and that of the vegetation in general.

The system shall also deal with:

- detection and registering of various oil films: from a few micrometers to several centimeters thick;
- monitoring of general dynamics of the shore line (abrasion, landslides, building up and washout of sand bars, etc.) with high accuracy – of up to 2 to 5 m.
- agriculture activity; forestry, etc.

One can see that due to the range of characteristics having been monitored our System' objectives must be as following:

- collection and storage of current information on stationary and mobile pollution sources;
- detection and evaluation of risks presented by various critical and accidental situations in the technological, ecological, navigational and hydro meteorological spheres, taking into account the separate and cumulative effects of the considered factors;
- analyzing the background and current data in order to produce short- and long-range predictions for the observed objects and situations and collisions;
- presenting the current and predicted data in the form of maps describing temporal, spatial and material details;
- working out and justification of optimum technological and organizational solutions related to the prevention and combat against various dangers, and mitigation of the anthropogenic factor in the entire region.

So, instruments of "GERMES–I" system have provide to the administrative, economic, environmental bodies with reliable and fresh quantitative and qualitative data on the areas and objects, and compare them with the predicted data, etc.

In conformity with the set objectives and depending on specific conditions the monitoring shall be carried out:

a). regularly: a continuous observation and monitoring of assigned parameters over the whole land and water area and in specific sites;

b). as additional measurement and observation assignments required clarifying the situation (especially – the danger one) as regards the time, location, resolution, specified components, etc.

The work should be making on three levels: using satellite, air, and land means, with employment of a remote sensing system comprising optic, infrared and MCW instruments, as well as other appropriate equipment and software.

Basic parts of System are (Figure 106):

- block that realizes remote sensing (RS) of earth and water surface, by means of high technologies – Microwave Radiometry from planes or/and another platforms – for recognition and evaluation the green agricultural and forested (wooded) territories, overmoisted zones along the rivers, to contour it, to estimate leakage of the waters from the ponds, canals and pipelines, to define oil pollution, soil mineralization, iridizations, etc.;
- intellectual block, oriented to analyze data monitored, to assess various characteristics measured, to recognize and classify situations, objects and processes under the survey and control.

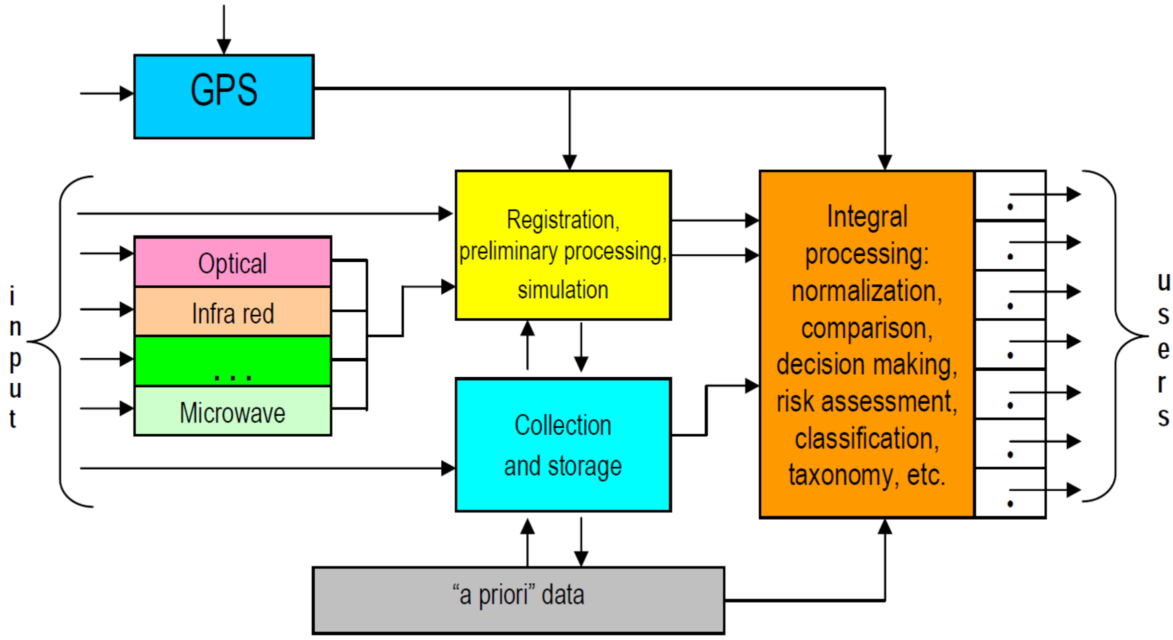


Figure 106. Preliminary version of GERMES-I structure

In our case a contour of GERMES–I includes some functional units [Krissilov, 1999], [Krissilov et al, 2001], [Haarbrink et al, 2007]:

- optical, infrared and microwave radiometers for data remote sensing;
- device for coordinates measuring (GPS – Global Positioning System);
- tools and instruments for various environmental, meteorological (etc.) parameters' measuring, registration and primarily processing;
- "a priori" data unit;
- unit for collection and storage of current data from various sources of pollution and accidents;
- unit for special intellectual processing of data measured.

The main attention in description below will be paid to first and last points, especially to remote sensing possibilities (MCW radiometry, see Chapter I), and to aggregated (integral) assessment methods, and to original effective decision rules. In addition, it must be underlined that last unit in our scheme is not simply mathematical one; it fulfils certain "mental" intellectual functions.

7.3 Some examples of practical MCW monitoring and risk assessment

It is needed to say several words about idea of "risk", referring particularly to Preface. As a rule, the risk related publications are discussing theoretical aspects of "risk assessment", "risk management", "risk development", "risk of artificial intellect utilization", "risk analysis", "risk theory and security", etc. One of the conclusions is: for obtaining the most representative "risk concept" definition, in the environmental situations especially, it is necessary, first of all, to reveal the most essential parameters of environmental objects or their models and patterns and, secondary, to determine the boundary characteristics for parameters under examination. It must be noted, that these two steps are completely adequate to the procedure of microwave radiometry data verification, validation and calibration. The team of the Institute of Radio engineering and Electronics, Russian Academy of Sciences, is working in this scientific and practical field more than 30 years (last decades – with certain collaborators, esp. "Miramap", Netherlands) and has accumulated a huge data bank on "what is microwave radiometry and what is an information content of measurements conducted by the microwave radiometers in different bands, from different heights, under different spatial resolution". Few examples of this methods and techniques application in various countries are placed below in this paragraph.

In addition, it must be clear, why just MCW methods and techniques can play the first fiddle among other sensors and methods.

What happened physically and historically, is that the objects to be remotely sensed in optical and infrared bands, namely, soils, vegetated areas, snow and ice, are characterized with big time constant, which varies within weeks–months and decades-centuries.

At the same time, water as it is (in stormy condition and in condition of chemical pollution), water on land surface (in condition of flood) and water inside the land surface (caused by rains and artificial watering) manifests as the most dynamic parameter of environment with the time constant of minutes-hours and days-months.

Fortunately, there exists a mighty remote sensing method of practically direct measurements of radio-physical water and moisten soil parameters under small dependence of parameter accuracy on surface roughness and a soil type, that is method of microwave radiometry (Figure 107). Main information can be found in [Shutko et al, 1998, 2007], [Krissilov, 1999], etc.

Devices for microwave radiometry are special (very high frequencies) radio-receivers perceiving by means of antenna the self-radiation of surface (objects) below in certain range of wavelength. These devices are able to determine the following soil, water and vegetation related environmental parameters and conditions:

- surface soil moisture,
- underground moistening,
- depth to a shallow water table (down to 2 meters in humid areas and down to 3-5 meters in arid/dry areas),
- located on the surface and shallowly buried metal objects of a reasonable size under the conditions of dry ground,
- contours of water seepage through hydrotechnical constructions (levees, dams, destroyed drainage systems, different kinds of leaks),
- biomass of vegetation above a water surface or wet ground,
- increase in temperature in land, forested and volcano areas,

- changes in salinity/mineralization and temperature of a water surface,
- water surface pollution, oil slicks on a water surface,
- on-ground snow melting,
- ice on a water surface and on the roads, runways.

Some figures below illustrate various kinds of application MCW-technique in different tasks during last decades.

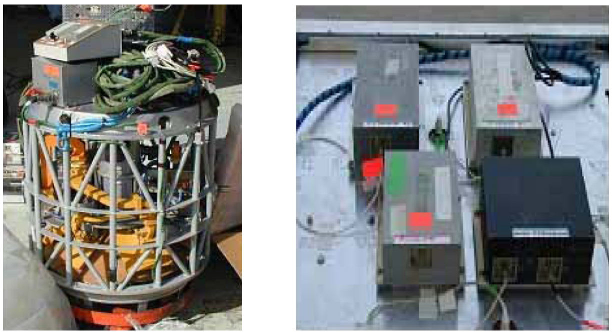


Figure 107. Scanning and non-scanning MCW-radiometers ready to work

Figure 108 shows the 5 micrometers film on the water in Odessa port area; left arrow indicates polluted piece from port Juznyj, the right one – from port Illichevsk. There was nice weather, no waves nor wind, we see background picture.

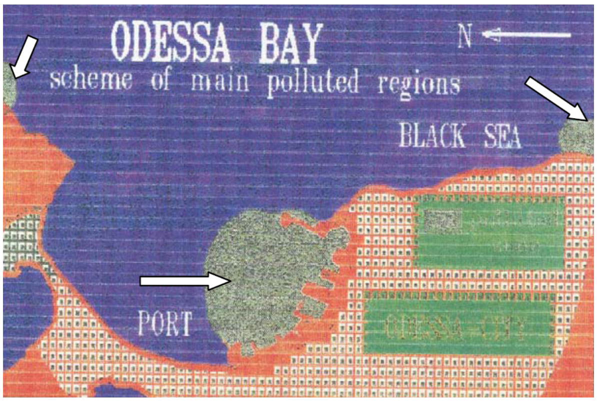


Figure 108. Thin oil-film recognition in Odessa harbor (end of 80s)

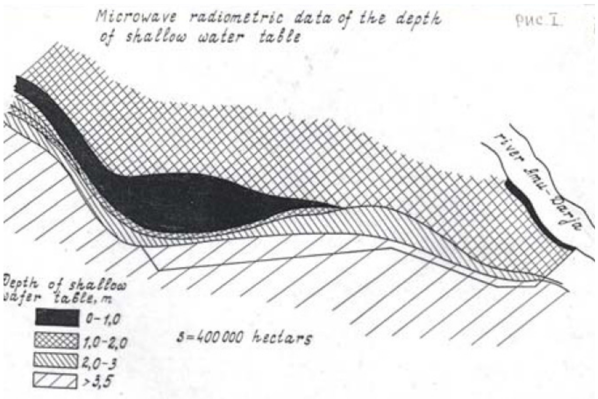
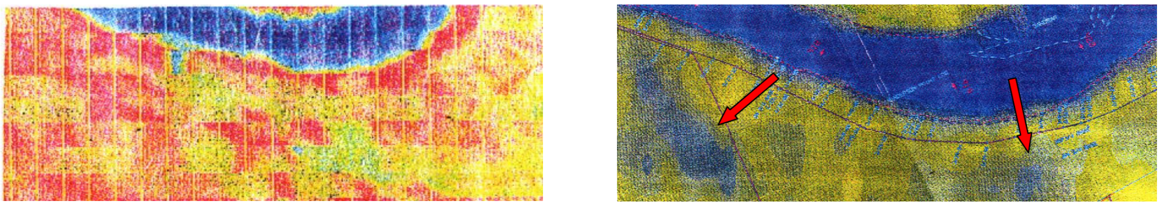


Figure 109. Map of underground water, Middle Asia (early 80s)

On the Figure 109 one can see that the boundary of moistened territory is situated in hundred kilometers from channel. This experiment and many other similar experiments indicating water seepage from existing canals pushed Soviet leaders to stop the Project devoted to turning some water from Siberian Rivers to Middle Asia and finally the project was declined.

Very interesting and demonstrative result of MCW measurement is presented on the Figure 110. Both pictures were taken at the same place on the bank of river. Left picture shows the surface of bank as dry territory (soil moisture levels are marked in colors – red and yellow indicate dry soil – water is absent). On the other hand, on the right picture, we can see two underground lakes just at the same place – the blue color indicates them. For obtaining such results, the measurements had to be made in different parts of MCW range.



(a) *(b)*

Figure 110. Illustration of surface state (a) and underground water table (b) measured in different bands (red and yellow colors indicate dry surface)

The accuracy of measurement one can see on the Figure 111: In its lower part, the vertical segments indicate results of hand-made in situ contact measurement, and dotted line – remotely sensed MCW data, from plane platform.

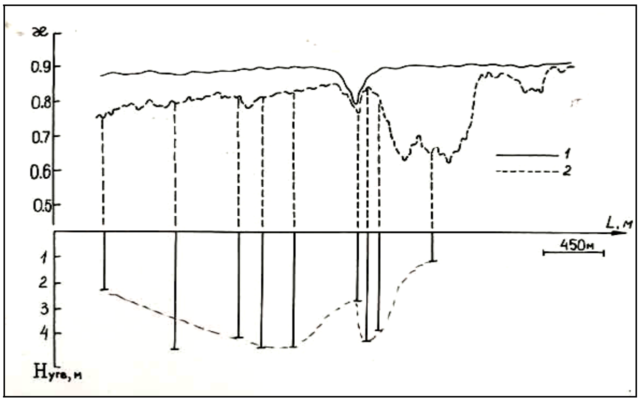


Figure 111. Experimental comparison of contact and remotely sensed data.
Upper Part: changes in emissivity at the wavelengths of 2 cm(1) and 30 cm (2)
Lower Part: "in situ" data of the depth to water table

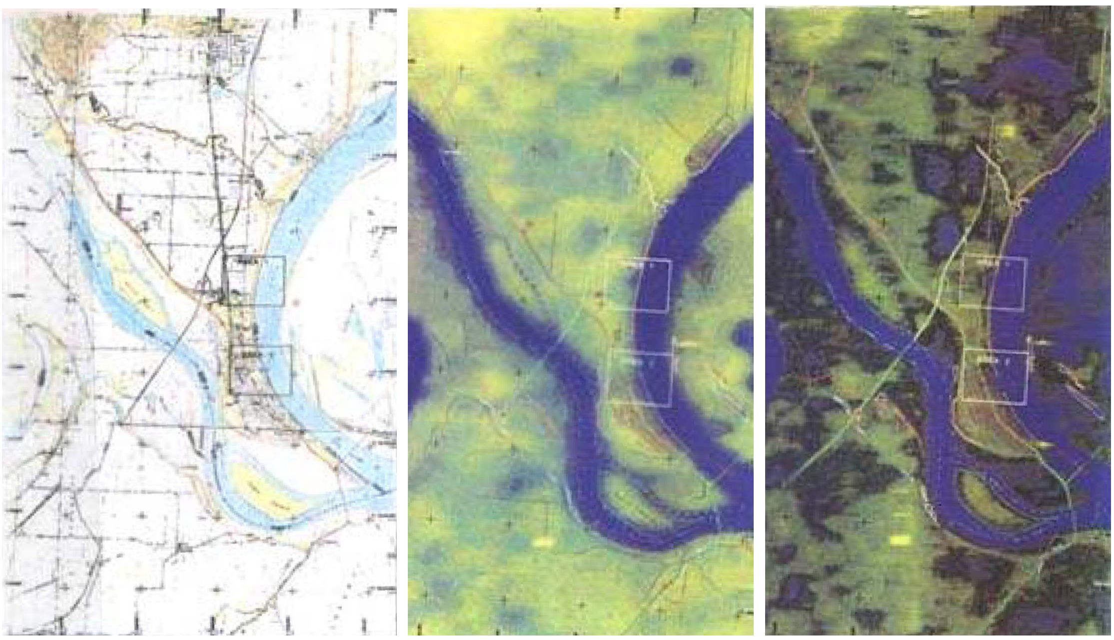
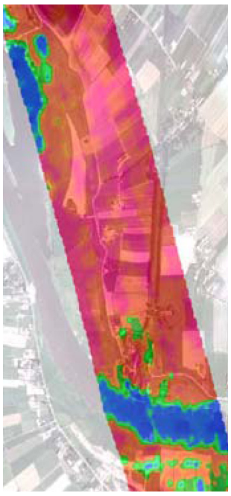


Figure 112. MCW monitoring in the region of Mississippi and Ohio Rivers, USA, end of 90s

Figure 112 demonstrates dual-wavelength procedure of water seepage/leakage determination through levees/dikes. The left picture is the map of the area, center – data at 5-cm wavelength, right – 21-cm data.

Results of superposition of different images (concerning the same object) are shown in the pictures below.



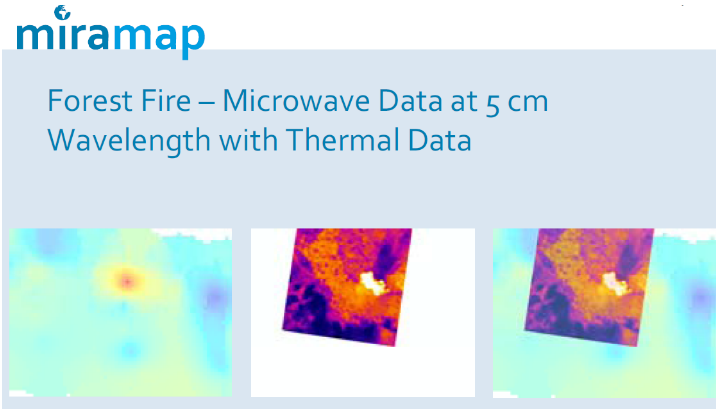


Figure 113. The complex (synthesized) pictures: left – area of water seepage detection in the Netherlands, 2005; right – forest fire in Bulgaria, 2007

Right part of Figure 113 contains three images: image of forest fire as seen by 5-cm wavelength microwave radiometer (left) and thermal infrared camera (center); the overlaid images are shown on the right.

Significant result can be seen in the figure below. Figure 114 is satellite multispectral picture with details of Mexican Gulf coastline. Figure 115 presents thematic maps obtained by MCW remotely sensed (from plane!!) data: surface soil moisture, depth to water table and vegetation index – different colors indicate different parameters. Both of these images are not "operational ": first one has no thematic information, the second one doesn't contain any geographic detail.

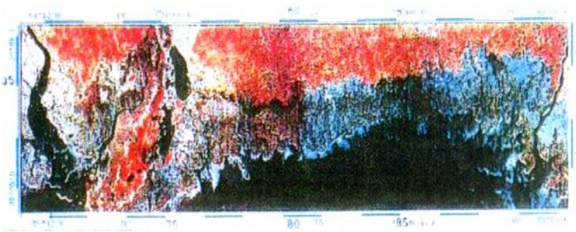


Figure 114. Satellite multispectral picture of Mexican Gulf coastline

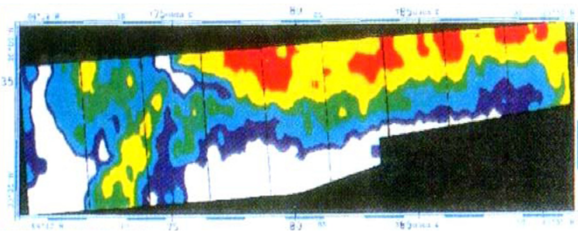


Figure 115. Thematic maps obtained by MCW remotely sensed data

On the initiative of the IRE RAS the National Wetlands Center in Lafayette synthesized all information available for this area and presented it in Figure 116. Almost all needed information together with surface geographic material is presented on it.

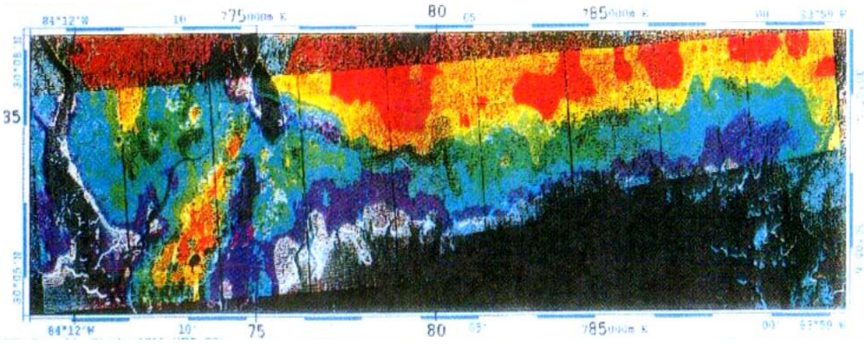


Figure 116. Mexican Gulf coastline synthesized picture, end of 90s

Some practical examples (from Bulgaria – 2007 expedition) devoted to the risk assessment of hydrotechnical construction leaking and of readiness to forest fire appearance and development will be presented below.

At the end of this brief review of some technical instruments and practical risk measurement and assessment, we want to say following. At first, the possibility for emergency response mapping, integrated with GPS and GIS technologies and special modeling and original algorithms, facilitates the monitoring and control of areas of water seepage through irrigation constructions, levees and dykes as well as revealing areas with dangerously high groundwater level. The passive microwave radiometry, used in GERMES-I system, is based on spectral measurements (own surface radiation receiving) in the millimeter to decimeter range of wavelengths. Moreover, comparing with other RS tools, such as color or infrared photography, thermal images and lidar, MCW radiometry is the only one taking measurements under the earth's surface and therefore is very well suited for any hydrological parameters monitoring in a fast and reliable way.

And, second, it can be seen, that all these kinds of activity, all monitoring tasks and problems are needed the aggregative assessment, quality evaluation, effective and right decision making, – whole advanced arsenal of intellectual data processing.

7.4 Artificial intelligence approach proposed and main principles of estimating/classifying model creation

There are some reasons in environmental studies to elaborate the Decision Support System for Complex Objects Evaluation now (it is understood, an estimated "object" is a very general concept, which can refer to a situation, an alternative, an action, a product and so on).

These reasons are as following:

- the fast increasing of diverse and combine information, which can be analyzed only with great difficulties if preliminary aggregation, classification, generalization etc. are absent;
- necessity to make analysis of complex objects with different nature in various fields of human activity (science, management, industry, and so on).

There are many advanced approaches to solve different AI tasks, such as Bayesian estimation, Fuzzy Sets Theory, Neural Network Theory and other. However, some tasks, such as decision-making, choice, classification, judgment, and pattern recognition, have similar nature and can be solved by means of certain common tools, in definite unified way.

This statement has following grounds:

- each task mentioned above is proceeded from the system of local parameters with various nature (quantitative and qualitative);
- solving process includes certain considering and transforming the primary features;
- all tasks are knowledge based;
- all of them consist such operation as calculating, selection, quantitative grounding.
- their principal dissimilarity has different types of result (numeric and symbolic).

So, we propose in our system to unify different methods and to consider AI tasks mentioned above from one point of view [Krissilov A. and V. Krissilov, 2005], [Krissilov V. et al, 1998], [Krissilov V. and A. Krissilov, 2000]. It becomes possible if one takes into account the aim of concrete environmental (protection, prevention, administrative, managing) task. Only formal aim's description permits to get high-quality results. The main point of approach presented is the aim-oriented vector model of evaluation task.

Evaluation of object is based on the aggregated characteristic calculating. The aim-oriented vector model permits to describe different AI tasks and to solve them in terms of complex object' evaluation (Figure 117).

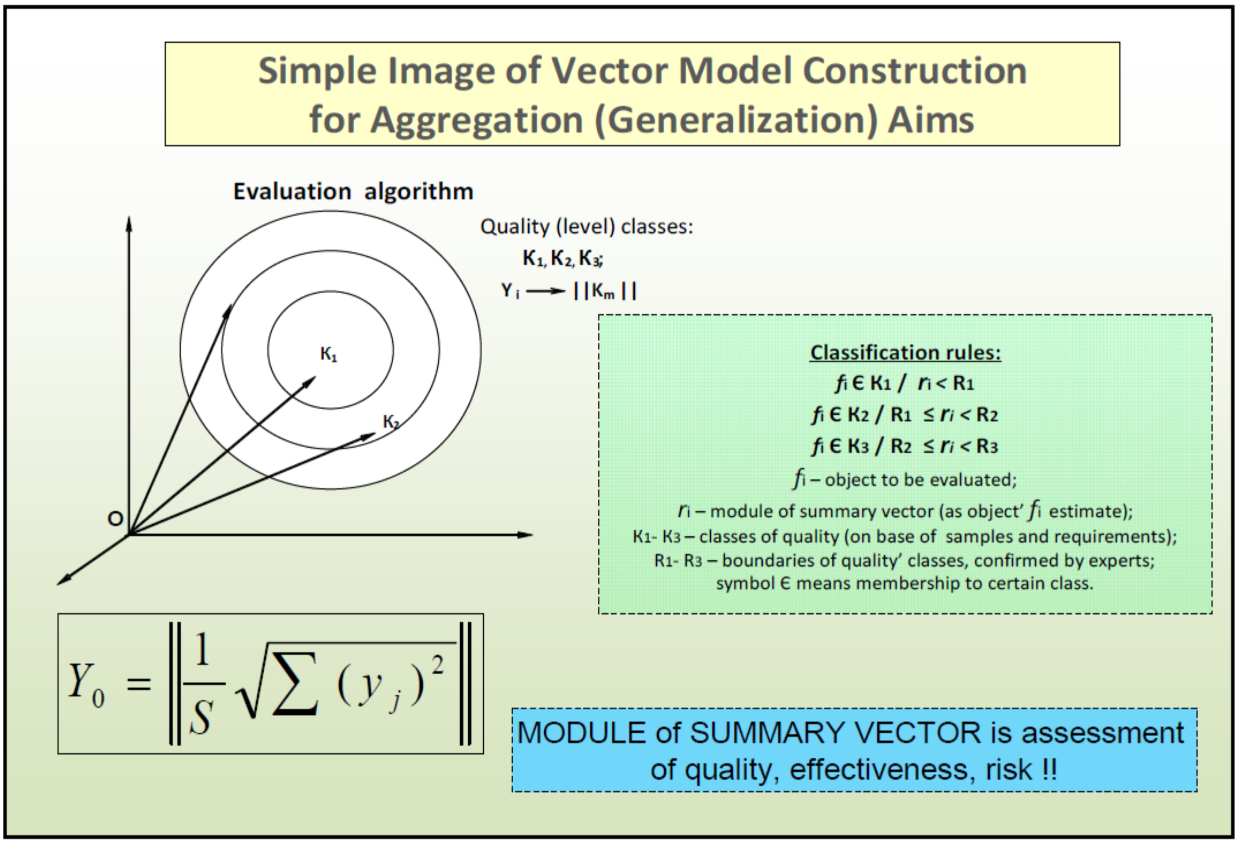


Figure 117. Presentation of vector assessment model

It may be shown that in essence all of AI tasks are based on evaluation with different goals. The resemblance of classification and evaluation can be demonstrated as example for proving it. Classification is the assignment of objects to groups within a system of categories distinguished by structure, origin, etc.

Usually the operations have been making by classification, by pattern recognition are following: comparison of concrete input object (in its feature' terms) with each sample of classes recognizing, the evaluation (calculation) of similarity level, and then the most coincident one' selection. So the classification can be produced substantially via evaluation.

Evaluation is a forming of an opinion or the approximate calculation of the qualities of an object. Thus, from that point of view, classification and evaluation are counterparts. The main difference is type of conclusion. Usually it is symbolic in classification and numerical in evaluation.

Analogously it can be shown that the pattern recognition, some kinds of decision making, choice, and judgment can be expressed in terms of evaluation.

Let us consider in detail the process of the evaluation, its participants, and component parts [Krissilov A. and V. Krissilov, 2005].

The correct evaluation process must be done with taking into account not only features of object, but also the requirements of subject, because the same object has to get different benchmark in cases of different requirements. In fact, the evaluation is formed as measure of correspondence (coincidence) between subject's requirements and object's features.

The object's properties can be both quantitative and qualitative, as it was told. Quantitative properties are characterized by possible maximum and minimum values. Qualitative properties are characterized by the set of their own values or description, – for instance, by means of so-called linguistic variable.

The concept of the requirement allows formalizing the aim of the concrete evaluation task. The requirements exactly should be attributed a certain weight factors in order to take into account the different importance of different factors (features). Thereby, forming the requirements means defining of ensemble of corresponding object's properties, and determination of the importance of the given requirement in respect to other requirements. Besides, presence of requirements determines necessity of forming of the set of dependent properties values, which the mostly satisfy to the given requirement. In general, it is possible to put these values as standards for given properties in case of concrete task.

7.4.1 Description of aim-oriented vector model

So, all this reasoning may be formulated as the main initial principles of vector model forming:

- there are two participants in every estimation process: object of evaluation and subject of evaluation;
- there are both quantitative and qualitative features;
- there is necessity of standard formation for every object's feature.

The vector model is intended for formalization of complex object evaluation with taking into account the purpose of concrete task.

The aim-oriented vector model composition

Let **S** is **n**-dimensional space of features. Each estimated object can be represented by a point $\mathbf{s'=\{s_1,\ldots s_i,\ldots,s_n\}}$ in this space.

Let **V** is **m**-dimensional space of requirements to objects. In general, **n** is greater or equal to **m**. The aim of every task is connected with point (or a zone) $\mathbf{v_O}$ in space **V**.

Example:	Demand:	Features:
	Good specialist	1. Education
		2. Experience
		3. Last place of job

Then there is point (or a zone) **E** in space **S**, which connected with $\mathbf{v_O}$ by the correspondence:

$$F:(V \Rightarrow S)$$

All objects from this zone satisfy all requirements on a 100%, -- so, it is a standard for this task. It is understood, that in case of classification task we have to form $\mathbf{v_O}$ and E for every classes.

The different requirements have different importance for concrete task (and/or for concrete experts). Moreover, each feature has particular influence to their demand. The vector of weight factors **P** is formed for taking into account the different importance of requirements and features.

In general case, the objects have quantitative and qualitative features. That is why all of features are presented as contribution function for identical treatment:

$$Z_i = f_i(X_i),\ i=l,...,n,$$

where $\mathbf{Z_i}$ is the contribution of the feature $\mathbf{S_i}$ to object evaluation producing and $\mathbf{X_i}$ is value (meaning) of feature $\mathbf{S_i}$.

The contribution function shows the influence of feature value (meaning) to final bench-mark for object. The contribution function must be defined for all feature $\mathbf{S_i}$ values (meanings) and it has maximum, which is the standard **E**. It is possible to say that every feature is exposed under the fuzzy transformation.

So, cortege **M = (S, V, F, P, Z)** is general context for evaluation task, and we propose name it aim-oriented vector model (AVM). Its view is shown on Figure 118.

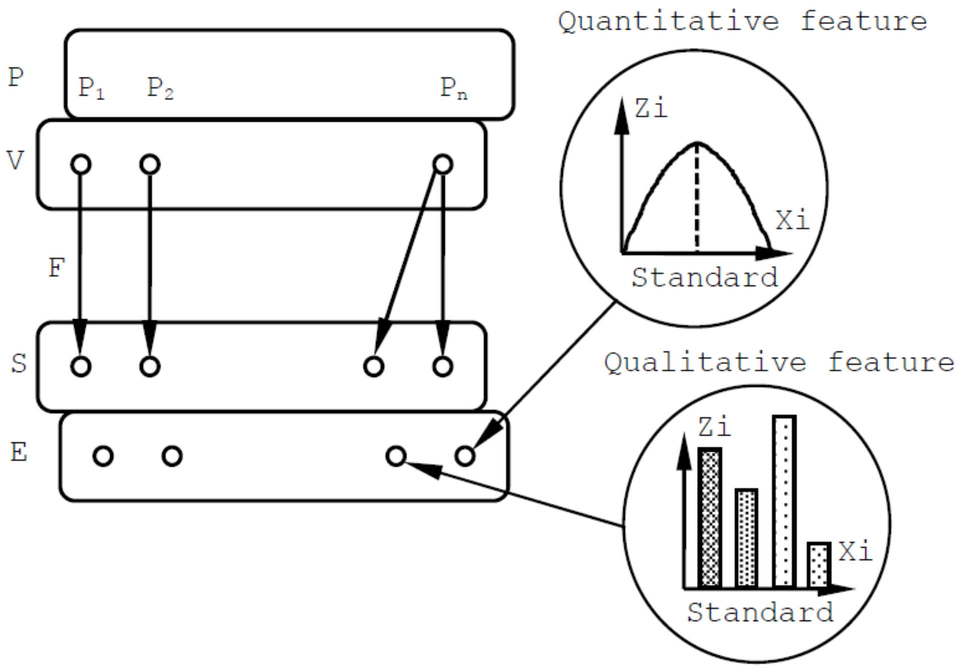


Figure 118. Relation between requirements (Vs) and features (S)

7.4.2 Solving of evaluation and classification task by means of AVM forming.

The aim-oriented vector model **M** allows calculating aggregated characteristic of the estimated object. The aggregated characteristic is:

$$K = ||PxF(Z'), Pxv_O||,$$

where x – sign of direct vector multiplication;

|| || – Euclidean metric in **V** space;

Z' – vector of z_i for point **s'**, which represents estimated object;

P – vector of weight factors.

The value of aggregated characteristic is defined as the distance between point **s'** (evaluated object) and point v_O (standard object, aim of evaluation) in transformed feature's space.

Analysis of the aggregated characteristics makes possible both estimation of objects and/or classification. It is possible to find out contribution of each feature and make retrospective analysis and determine which features raise the measuring level, which ones make it lower and which are indifferent ones (from point of view of the concrete requirement's list).

Now consider some especial aspects of classification task solving by means of the vector model proposed. In the first place, the task's context has to contain the set of classes must be recognized. Let us name it set **C**. Therefore, we have to form standards **E** and weight factors **P** for every class, and aggregated characteristic of classifying object is calculated for every class.

In the second place, the pattern recognition process is provided by means of some decision rule. One of them, namely the decision rule maximizing a right classification probability [Dubois et al, 1997] can be proposed for our problem solving. This decision rule is as follows:

$$s' \in c_j / R_1 = \max_j (K_j(s')),$$

where **j = 0, 1, …, L** is number of classes, and $K_j(s')$ is aggregated characteristic of object **s'** for class c_j.

This rule gives best result only in a case when the features are independent. There are many tasks where the objects are described by depending features. In this case, we propose to make special learning stage. Showing known objects from class c_j can measure for each **s'** the $K_j(s')$ and fix the probability $P(K_j(s')/c_j)$. Then the following decision rule may be proposed:

$$s' \in c_j / R_2 = \max_j (P(K_j(s') / c_j)),$$

The proposed approach allows getting correct solutions of classification tasks in case of depended features. This area of task will be described in details in fifth paragraph.

7.4.3 The stages of vector model forming

The process of aim-oriented vector model forming includes:

I. The stage of abstraction.

1) The feature set (space **S**) forming. There are possibilities to use both quantitative and qualitative features, but the measurable features are preferred.

2) The requirements set (space **V**) forming. It is understood; the requirements can be more general then features.

3) The correspondence **F:(V⇒S)** forming. Every requirement must be connected with one or more features.

II. The stage of scaling.

4) The vector **P** of weight factors forming. The weight factor takes into account the requirement importance and importance of feature for its requirement.

III. The stage of fuzzification.

5) The contribution function **Z** forming. The contribution function extremes are the standards for the feature.

In case of classification task, solving the steps 4 and 5 must be made for each class.

There are two fundamental ways to form the vector model: a) on a base of the statistical analysis; b) on a base of the expert's knowledge and opinion. Both of them were applied in the presented model.

The evaluating system above was realized with using certain effective decision rules and with calculating the aggregated estimations of objects/processes monitored. Moreover, in system described some original decision functions are used that allows operating with dependent features [Krissilov, 1984], [Krissilov et al, 2000], [Krissilov A. and V. Krissilov, 2005]. This property permits to solve complicated tasks with high accuracy.

Let us consider these interesting and original decision functions.

7.5 Decision making rules for situations both with independent and dependent features

7.5.1 Introduction notes

A lot of tasks, due to automatization and algorithmization different kinds of activity, deep computing and so on, – need now the effective decision rules application. There are dozens of works devoted to this problem, among them, particularly, [Banerji, 1978], [Schlaifer, 1979], [Zagoruiko, 1976], [Gladun and Vashchenko, 2000], many others. The neuro-computing methods and technologies also are used in the area of artificial intelligence (AI).

There are certain works describing data processing by means special spaces [Krissilov V. et al, 1998], [Krissilov A. and V. Krissilov, 2005]. It may be seen, that using geometrical explanation and imagination is very fruitful approach in various AI tasks and in processing of complicated data upon the whole.

A great many of decision functions and rules, that are applied in tasks of pattern recognition and support of the administrative decisions, are constructed in the assumption of independence of attributes (features) being used for description of analyzed and/or recognized objects. Desire to take into account dependence between parameters leads to the very large expenses of time and memory, or, at the best, comes to final fixing, for example, only of pair dependences, in nearest neighborhood. Besides, it happens seldom when users or developers have the information on real values of dependences features’ presence or absence from each other in various processes, classes of objects, etc.

At the same time, it is difficult to imagine a task, in which the description of analyzed objects is made by means of parameters independent in aggregate. Especially it looks clear by putting medical diagnosis [Popov, 1971], in solving the environmental tasks [Krapivin et al, 1996], in realization of geoinformation monitoring, etc. In these last cases, e.g., analyzing a natural situation or estimating anthropogeneus influence, the expert operates with such characteristics, as moisture of ground, water table level etc., are dependent from each other certainly.

Then if we have the problem to build full-scale monitoring system (and while another tasks solving) alongside with other means of processing of the information, it is necessary to use decision functions, which are able to take into consideration relations and dependencies between the features without using a big system resources.

7.5.2 Decision functions (1st learning stage)

Let some *k*'s object, having evaluated or recognized, is described by *n*-dimensional vector [Krissilov, 1962, 1984]

$$f_k = \{\text{v}_1, ..., \text{v}_i, ..., \text{v}_n\}$$

where $i = \overline{1,n}$ is number of feature and v_i is the measured value of ***i'***s feature in the ***k'***s object.

Dealing with classification problem, we can obtain (as result of learning stage) the matrix

$$C_n = \|p_{ji}\|,$$

where *j* is number of class and p_{ji} is probability of *i*'s feature in *j*'s class.

Therefore, we see, that objects recognizing have statistical nature. First step in learning is shown on Figure 119.

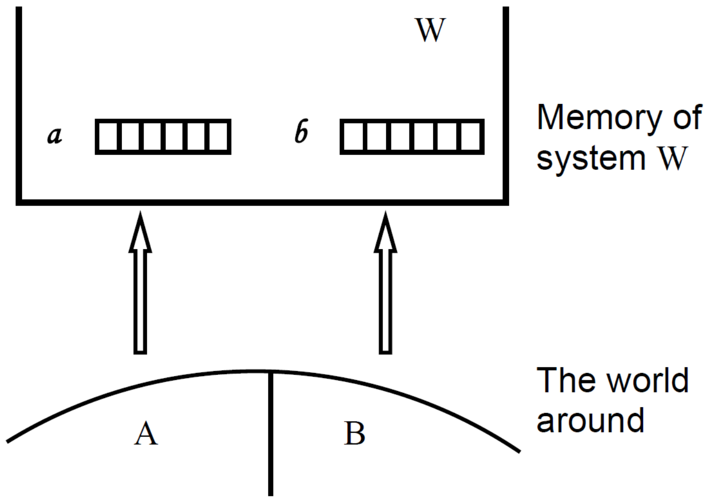


Figure 119. Creation of the memory (1st stage)

The pattern recognition stage is provided by means of some decision rules. One of them, namely the rule maximizing a right classification probability [Krissilov, 1984], [Krissilov et al, 2000] can be proposed for our problem solving. This decision function is as follows:

$$f_k \in M_j / R_f = \max_j P(M_j / f_k) = \max_j \prod_i^n p_{ji}^{(\text{v}_i)}, \quad (5.1)$$

where f_k is *k*'s unknown object, which must be classified;

– M_j is class with number *j*;

– $P(M_j / f_k)$ is probability of M_j when this concrete set of features f_k is presented;

– $p_{ji}^{(v_i)}$ equals to p_{ji} if v_t=1, and to $1-p_{ji}$ if $v_t=0$. And it is probability of measured value v_t, if the feature is not binary.

In expression (5.1) the value R_f, being found as the maximum of the value $\prod p_{ji}^{(v_i)}$ for various classes (j changes from 0 (unknown class!) up to S), – shows the class that can generate the present vector f_k with the most probability, – in a case the features are independent each from another.

However, there are many tasks, as it was told, where depending features describe the objects. So we have to construct corresponding decision rule. For these purposes, the second learning stage must be introduced.

7.5.3 Decision function strengthening (2nd learning stage)

The curves of probability distribution of concrete characteristics of examined objects are applied to this task solution. These curves are obtained as a result of the additional training stage by means of the representative choosing of objects/classes should be recognized (Figure 120).

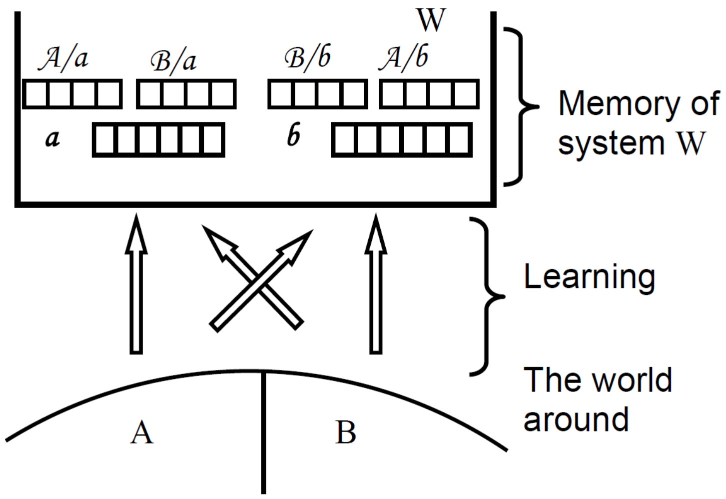


Figure 120. Second learning stage

Showing known objects from class M_j to recognizing system, we can find for each f_k the value $x_{kj} = \prod_{i}^{n} p_{ji}^{(v_i)}$ and fix the probabilities $P(x_{kj}/M_j)$ and organize its memory consequently (see 0).

Then the following decision rule may be proposed:

$$f_k \in M_j / R_\Sigma = \max_j P(x_{kj} / M_j), \tag{5.2}$$

Accordingly, this function of decision is made on the ground of a maximum probability of belonging the present input gamma of features (given f_k) to definite class, and comparing these probabilities for all classes. It needs to be noted, that we measure in this case the values of probability obtained by using only one line from matrix C_n, – for each class, i. e. using the sample of one class only.

The next our step can be generalizing this approach by means of extending the number of samples, of lines our matrix, being included in process, using for these purposes similar samples of classes and, lastly, all lines from matrix C_n entirely.

Then final version of decision rule looks as follows:

$$f_k \in M_j / R = \max_j \prod_{r=1}^{s} \left[P(x_{kr} / M_j) \right], \tag{5.3}$$

where $r = \overline{1, S}$.

Thus, we obtained the multidimensional decision function (MDDF) that allows using deeper data in comparison with previous rules of decision.

7.5.4 MDDF interpretation

Multi-Dimensional Decision Function' using and applying corresponding procedures form/build the memory of recognizing system in manner of holographic one, distributed and developed. Indeed, in cases when decision rule in expression (5.1) is realized, then each class' sample in memory unit contains information regarding that class. This information is represented by features' probabilities for objects belonging to given class.

When decision function expressed by (5.2) is realized, then the information, which forms each sample, becomes more complicated. It includes now the curve of probability' distribution of appearance certain values x_{kj} (results of evaluation) being measured/ weighed on definite line of matrix (with number *j*) by showing objects from corresponding class.

These curves of probability distribution (for all our classes), as it was told, are obtained as results of the additional training stage showing the representative sets of objects/classes should be recognized. In addition, we have for each class one curve only. Decision by (5.2) is made by means the finding maximum of evaluation found among classes compared, it is maximum probability of belonging just that combination of features (in this concrete recognizing object f_k) to given class.

Moreover, finally, when rule (5.3) is used, the sample of each class includes (in form of certain curves of probability distribution) the indices of estimates/assessments x_{kj}, which shows the results of weighing various objects from *given* class *on samples all other* classes. Just this result we obtain in 2nd stage of training. It looks as if just one object or situation is observed by means of several perception inputs, from different points of view, or is passed in parallel through various filters or prisms (see Figure 121 and Figure 122). It was found (some years ago, in certain neurobionic studies) that information from *each* mouse' vibrissa on its nose reflects (is sent) not only *in one* corresponding pull of neurons in mouse' brain, but it is in *all of them*; very close to the way being realized in holographic system.

We may say that increasing of grounding and quality of decision-making is got in this case in our decision or assessment system as result of more entire extraction of information about classes should be recognized, about interrelations between classes and latent links between its features. Besides it must be pointed out that such results are obtained due to some fuller and better disposing the knowledge mentioned above when Multidimensional Decision Functions in our AI-tasks and in various managerial and control problems are used, due to better use the "context information".

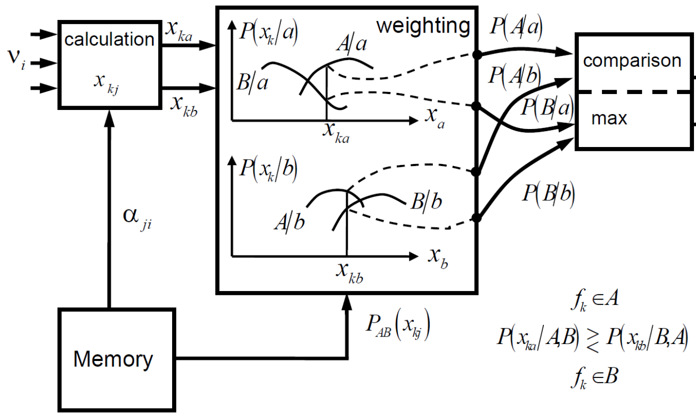


Figure 121. Using the probability distribution curves for advanced decision

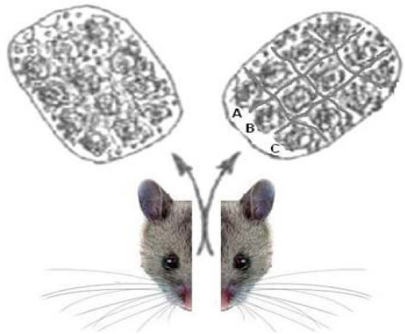


Figure 122. Structure of mouse memory belonging to the vibrissa's information

7.5.5 Some concluding remarks

Decision functions mentioned above was successfully used by solving various tasks:

- environmental tasks and radiometry data interpretation;
- risk assessment and estimation of various protective devices' effectiveness;
- recognition of printed and handwritten letters, either some voice commands;
- data interpretation in geophysical exploration (known task "Oil – Water Recognition");
- evaluation of the socio-economic level for some administrative territories (life quality);
- comparison of development level for south regions of Ukraine;
- medical diagnostics, staff management, and so on.

Interesting and hopeful results were obtained in all tasks listed above ([Krissilov et al, 2007], [Shutko et al, 1998, 2007] and others).

The system could recognize the different situations and objects, to make grounded decisions, is able to simulate monitored situations. Some important properties of outworked system were displayed:

- ability of work with qualitative and quantitative features;
- filling forecast information those of spatial-time intervals, where the measured data are absent;
- obtaining of aggregated, generalized evaluations for monitored processes, objects or situations in a case of action of various local or depending factors;
- ability of classification of the objects and situations, forecast variants evaluation, to provide support and quantitative decision-making.

Strictly speaking, we operate here with instruments that simulate the functions of intellectual processing of information.

7.6 Examples of intellectual data processing application in Bulgarian Expedition (2007)

7.6.1 Brief introductory notes

Probably, it is remarkable fact that Bulgaria became one of the first European countries for the last 10-15 years in which certain bodies have decided to spend multi-purpose monitoring environmental work with use of modern hi-tech means. This was done not at the level of private company, not for certain region or for narrow problem solving, but namely in the wide context and with the long-term purposes to obtain.

It is necessary to note that it was preceded by a long-term preparatory work at various levels: cooperation with the Bulgarian Academy of sciences, mass media publications, work with the profile parliament commissions, joint seminars, and conferences, etc.

In spring of 2007, the Bulgarian side has signed the contract with MIRAMAP Company, which has prepared for this purpose the international team with participation of specialists from Russia, Netherlands, Bulgaria, Ukraine, Australia, and United Kingdom. The Requirements Specification (Technical project) was developed together by both Bulgarian side and International team of scientists and engineers. Regions, problems, and terms were chosen as typical in order to show and to estimate the set of possibilities of applied methods and equipment and, on the other hand, to obtain an estimation of hydrological danger at the current environmental conditions. Owing to hot weather it has become actual to make the estimation of forest fires danger, and this problem also have been included into the program during fulfilling the work and was successfully solved.

It is necessary to tell some preliminary words about structure of the concepts connected with an estimation of risk, and about the role of information, of data – in this structure. Concept of risk is one of the widest and widely applied ones for the system representation of critical situations. As it was told, it belongs to number of badly formalizable concepts. However, some algorithms and methods, which appreciably overcome this difficulty, are entered into our monitoring system. They, anyway, make concept of risk operational, i.e. give the chance to operate with risk, in our case — to lower it that is essentially to be engaged in risks management.

The hierarchy of relations in system of the critical situations representations is following. At top level among other categories, there is a concept of risk. Further, one of the factors defining risk, namely degree of readiness for the prevention of a dangerous collision is situated (or readiness for minimization of damages at its occurrence from the uncontrollable reasons — a lightning stroke, flooding, at earthquake, etc.). Finally, degree of readiness for prevention of dangerous situations is appreciably defined by the knowledge factor. In a case the corresponding services have a few of data, dispose of the smaller information, then the more low degree of their readiness will be, the more poorly a measure on their prevention activity should be realized. It, in turn, leads to not optimum distribution of resources, to increase in risks and a possible damage.

The structural analysis of concepts risk and information (knowledge) and relations between them shows that they are connected by the quasi-hyperbolic relation (similar to curves presented on Figure 126).

Some considerations specify that these curves differ from one another by exponent of the hyperbola and distance from the axes. Presence of different curves is defined also by the fact that risks value depends on a number of factors: organizational ones, finance, personnel, other. However, appreciably exponent value (in the characteristic equation connecting risks with information!) and, consequently, the curve steepness, depends on parameter which we may to mark as quality of information. It is its completeness, system character, accuracy, etc.

It is obvious that for the best organization of readiness for the risks prevention (it conducts to reduction of risks and damages) — it is preferable to work with such curves for which in a transitive zone the insignificant increment of information conducts to considerable decrease in risk.

Last explaining introductory note is as follows. In order to describe quantitatively our situation we use concepts "risk classes» and "information classes». We shall consider risks low if they don't exceed 0,15 – 0,2, it will be I class of risks. Medium risks lie between estimations 0,15 – 0,2 and 0,45-0,55; it depends on variety of external and internal factors. Within these limits, there is a second class of risk. At last, all estimations above value 0,55-0,6 characterize III class, high risks. Approximately the same distribution is used by experts of insurance business. Classes of "volume" of information, with other borders of splitting, were analogously entered.

In Chapter 1 of the present monograph the short description of Bulgarian expedition is given. We would like only to emphasize the following examples of concrete use of "intellectual" algorithms of environmental data processing: for an estimation of protective properties of hydraulic engineering constructions and for an estimation of danger of forest fires. We want underline once again that our algorithms and methods are applied here neither for geophysical parameters interpretation, nor for increase of accuracy of measurements, but are used directly for modeling of abstract functions of a brain, for generalized concepts reception – regarding efficiency, risks, and similar to them.

Mentioned above examples of risk assessment of hydrotechnical construction leaking in the area of Rusenski Lom and of readiness to forest fire appearance and development in the area not far from Danube will be described below.

7.6.2 First example: assessment of dam hydro protective characteristics

On Figure 123 the dam near by the Nikolovo village (southwest from Ruse) is presented. The dam was constructed in the late fifties – the beginning of 60th, and has no clay inside basis. There is by-pass channel on right orthographical side, but its efficiency is insignificant. Because residential and administrative buildings are situated behind the dam, and due to some other reasons, the great attention was paid to the question: how high are hydroprotective properties of dam, does it hold water well or no, how much it is steady. This problem was especially important because of many residential buildings behind dam (on distance of 250 m and more) had had water in basements and cellars.

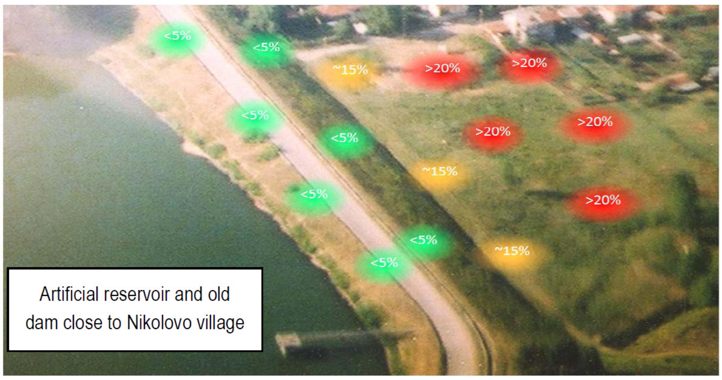


Figure 123. The dam nearby village of Nikolovo (from plane)

The figures put on photos, characterize humidity of a surface of the soil, measured by the radiometric equipment in the first day after long rains – about 5% on front side of a dam and more than 20% in territory behind a dam. This situation of the first day is presented on Figure 123.

The Dam first of all and surrounding area were covered by detailed RS measurement in the various conditions.

MCW maps with data measurements are presented on Figure 124 for dry and wet condition correspondingly. These maps contain the following parameters: both surface and underground soil moisture for dry and wet condition. Besides that: infrared assessments of surface temperature data in same area where the above microwave soil moisture assessments were collected; moreover, the lidar data of elevation were obtained in same area where the above microwave soil moisture assessments were collected.

In Figure 124a following colors are chosen for the following gradations of soil moisture in upper 0-5 cm thick layer in dry condition:

0.00 – 0.05 g/cc or 0 – 5% of volume (and forested areas) – dark green;

0.05 – 0.10 g/cc or 5 – 10% of volume – light green;

0.10 – 0.20 g/cc or 10 – 20% of volume – light blue;

0.20 – 0.30 g/cc or 20 – 30% of volume (and open water) – dark blue.

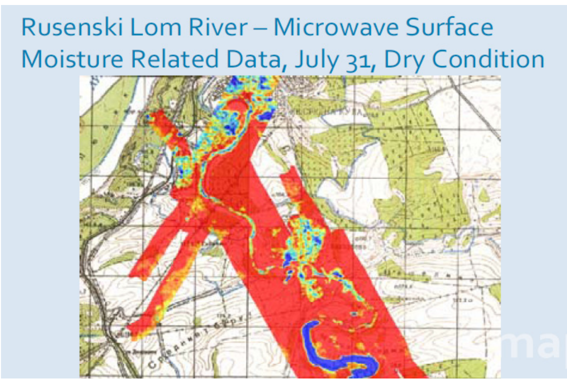
In Figure 124b following colors are chosen for following gradations of soil moisture in upper 0-5 cm thick layer in wet condition:

0.00 – 0.05 g/cc or 0 – 5% of volume (and forested areas) – red;

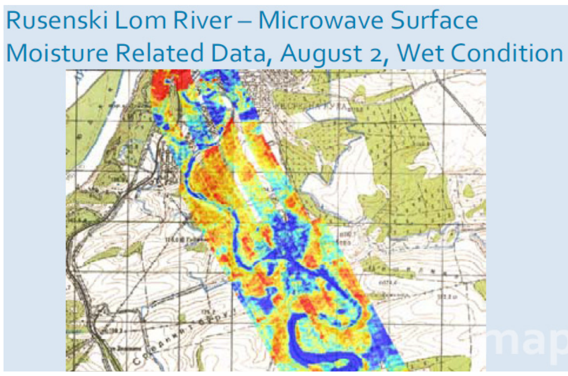
0.05 – 0.10 g/cc or 5 – 10% of volume – yellow;

0.10 – 0.20 g/cc or 10 – 20% of volume – light blue;

0.20 – 0.30 g/cc or 20 – 30% of volume (and open water) – dark blue.



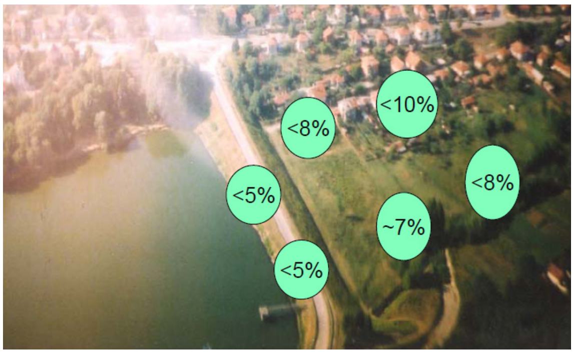
a) dry condition



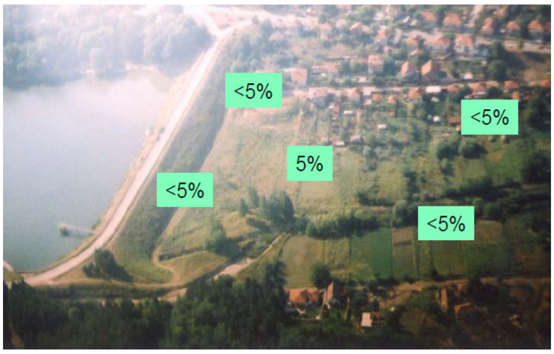
b) wet condition

Figure 124. MCW maps with data measurements at the Nikolovo area for various conditions

The remote gouging spent through 4 day and some calculations have shown that integrated humidity in three – five times has decreased (Figure 125). These total characteristics have been received as generalization (by means of model) the data on controllable area (some hundreds of sq. meters), and also taking into account surface and underground water – to one and a half – two meters.



a) Third day after a rainfall



b) Forth day after a rainfall

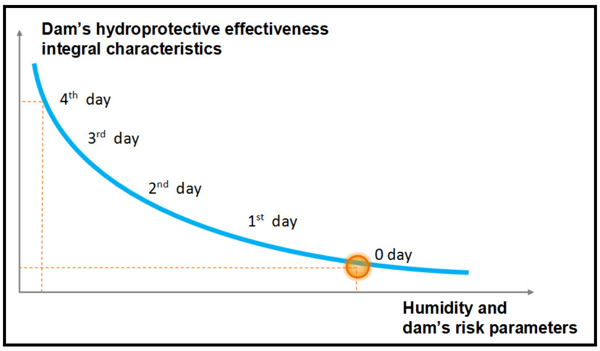
Figure 125. Situation with Nikolovo dam and area humidity after a rainfall

At performance of operation of generalization by means of the model described in given Chapter the certain weight factors have been used. In particular, at aggregation of the data of spatial distribution of humidity, estimation of this parameter for a back side of a dam were are taken with factors, on 30% by big, than for the front side and a dam surface; for slopes of underlying valley humidity of a surface of soil was considered with weight factor on 15% smaller, than for a ground part of this valley etc.

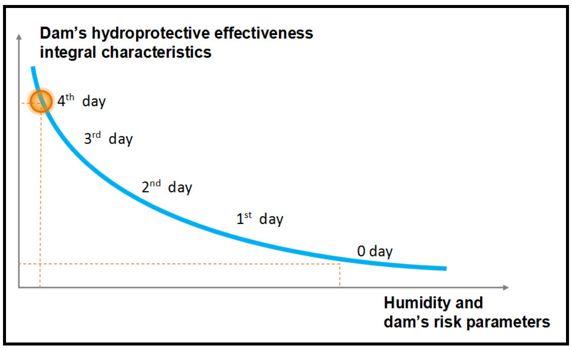
However, the account of the data mentioned above does not suffice for an estimation of efficiency of a dam as about its hydroprotective characteristics. It was necessary to consider: quality of road covering, age of the construction, first of all tixothropic properties of a ground (tixothropy is the set of its constructional characteristics and properties depending on a number of factors), and some parameters more.

The vector model described in paragraph 4 of the present chapter and solving rule (5.3) from paragraph 5 (in the modified kind) has been applied to construction of such generalized estimation of object quality (see Figure 126).

The conclusion has been made, as a result drawn, that hydro-protective properties of this dam are for today quite satisfactory; on Figure 126b the red circle indicates high level of hydro-protective effectiveness of the object assessed.



a) first view on effectiveness



b) final assessment of dam effectiveness

Figure 126. Illustration of use both aim-oriented vector model and advanced decision rules (assessment of dam effectiveness)

7.6.3 Second example – assessment of forest fire danger

Different situation has been observed in the forest area close to Danube as related to risk of forest fire development. Figure 127 presents infrared photograph of the woodland with two control areas – A and B tested in order to determine the probability of fire risk occurrence. As it has been revealed from dynamic observations of thermal infrared and microwave radiometric data, locations A and B could not be corresponded to the certain zone of risks, though preliminary version was that they belong to the middle class: about 30% of risk for A, and about 35% – for B locations.

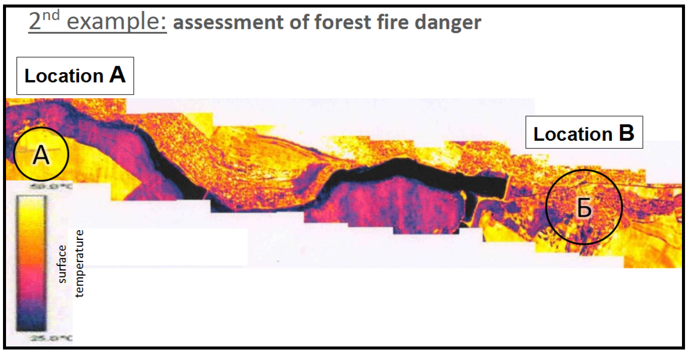


Figure 127. Infrared photograph of the woodland with two control areas – A and B

These preliminary calculation presented on the Figure 128 was based on the certain information. The expedition data bank contained the following maps:

– in situ data of soil moisture, temperature and precipitations (from Meteo) in selected areas,
– map of soil types from the text-books,
– map of surface relief from text books and from the lidar data
– map of brightness temperatures at 5 cm (most wet condition)
– map of brightness temperatures at 5 cm (most dry condition)
– map of brightness temperatures at 21 cm (most wet condition)
– map of brightness temperatures at 21 cm (most dry condition)

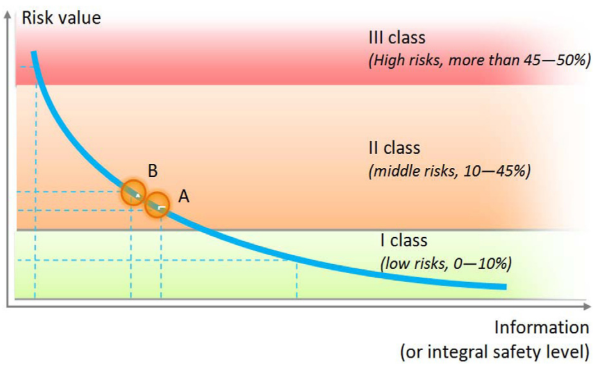


Figure 128. Preliminary assessment of forest fire danger for locations A (high risk) and B (low risk)

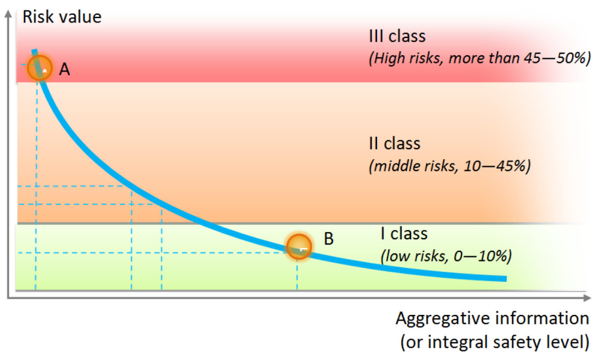


Figure 129. Final assessment of forest fire danger for locations A (high risk) and B (low risk)

None of that could give full picture in the forest. It must be known that deciduous or coniferous forest substrate (forest floor, pass-years leaves) plays a very important role.

Only having added data on a condition of a wood substrate, presence of dry trees and some points more; it has appeared possible to estimate danger of ignition really.

Then, according to the aggregate of parameters as mentioned above, total assessment of the fire risk occurrence for both areas was obtained. Upon the record of the temperature and other parameters of the forest floor it appeared that one of the areas is relatively safe (0,15 – I class), and as for the other one – the probability of ignition is rather high (0,65 – III class of risk) – which is presented at Figure 129.

7.7 Carrying platforms and some new technical points

In a problem of risk assessment as a whole, applied carriers of the metrological equipment play large practical role, from the point of view of efficiency, accuracy, and convenience of measurements carrying out.

According to various monitoring tasks and to different scales of objects under survey, a large variety of carrier platform for remote sensing are used. It looks fruitful when system "GERMES–I" being planned, to project system covering all levels of supervision: land (including water areas), from air – using planes and helicopters, and from space, using the data received from various satellites.

In Chapter 2 of the present monograph, the effective methods of various important information processing are considered in detail. These data are collected from a terrestrial surface, basically, by means of satellites and in situ regime. In the given Chapter, some examples of planes use as the basic carrier means have been mentioned.

However, throughout last 20 years, enough wide experience has been collected in use of various carrier platforms of the radiometric equipment developed by IRE RAS and radio corporation VEGA (RF), and at their joint activity in different monitoring projects and with various collaborating organizations worldwide [Shutko, 1987], [Shutko et al, 1995, 1998, 2007], [Haarbrink et al, 2007], [Krissilov et al, 2000, 2001], [Krapivin and Phillips, 2001], [Borodin et al, 1996], [Kondratyev et al, 2002], etc. Some examples illustrating this work are described below.

a) manned mobile platform

b) unmanned helicopter

Figure 130. Illustration of various mobile earth and air platform

Figure 130a presents photograph of the manned "Rover" type mobile platform equipped with three portable microwave radiometers, operating at the wavelengths of 6 cm, 18 cm and 21 cm. It is equipped with a folding panel of 1.5 m x 1.5 m for instrumentation installation along with GPS receiver, data acquisition system and power supply battery and is oriented for research activities related to soil moisture determination. These experiments have been successfully spent together with NASA Center for Hydrology, Soil Climatology and Remote Sensing, Agromechanical University in Alabama, and NASA Goddard Space Flight Center.

Figure 130b illustrates unmanned aerial vehicle (UAV) helicopter "Microwave Autonomous Copter System" (MACS) equipped with a 6 cm radiometer (is shown by red color), data acquisition system, GPS receiver and power supply battery (Miramap Co.).

Figure 131 contains the assessment of accuracy MCW measurement made by rover platform: comparison of microwave estimates of soil moisture at 21 cm wavelength unit with *in situ* soil moisture in 0-1 cm, 0-3 cm and 0-6 cm layer.

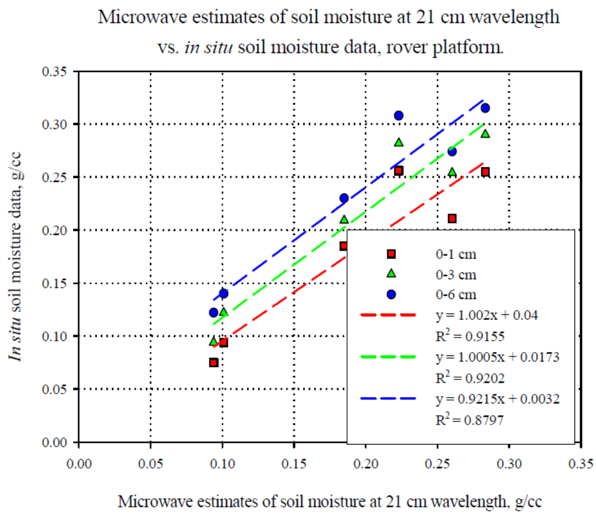
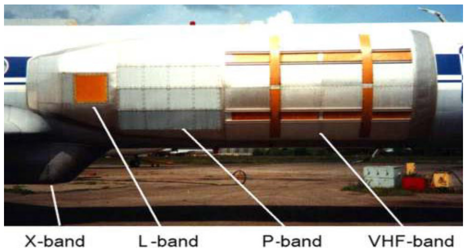


Figure 131. Comparison of remote sensed soil moisture with in situ data in 0-1 cm, 0-3 cm and 0-6 cm layer

Example of Airborne SAR system – IMARC, Vega Corporation, ("Flying Laboratory", SAR – Synthesized Aperture Radar) and allocation of certain kinds of antennas for different frequency bands one can see on Figure 132.

a)



b)

Figure 132. "Flying MCW Laboratory (a) and allocation of special antennas on the airplane (b)

In 2008, especially on Miramap request, Vega Radio Corporation designed and purchased the newest version of portable and light single-beam radiometers Ranet 6 (6-cm radiometer) and Ranet 21 (21-cm radiometer) (see Figure 133).

This is the newest generation of portable, autonomous microwave radiometers, which contain inside themselves radiometer, as it is, antenna, power supply unit, data acquisition system, and the GPS. The radiometer is designed for application in the laboratory conditions and inside a small cart for precise analysis of subsurface overmoistening or drought situation.

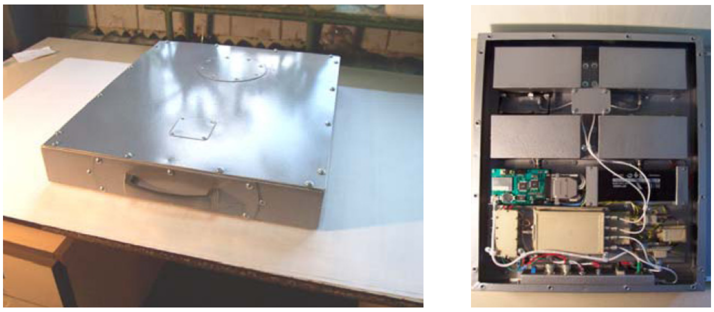


Figure 133. Ranet 21, the newest 21-cm radiometer design by Vega Radio Corporation

Last two examples from VEGA and IRE studies concern fifth level of works, according to the list in the first paragraph of the given Chapter: strategic international projects with use of high technologies for environmental monitoring, struggle against hunger, spontaneous and technogenic disasters. It can be power monitoring complex for strategic revealing and determination of various dangerous factors throughout the World, for instance creation and developing "RiceSAT" complex with partners in Asia and America, participation in GMES, so on.

On Figure 134 example of dual-frequency spaceborne radar with synthetic aperture is shown. This device has special antenna 12 x 3,3 m, two wavelength bands P (70 cm) with resolution 30 meters and L (23 cm) with resolution 3-30 meters. Originally, this complex is designed for:

– radar survey of Earth: land and water surface for Earth resources exploration, oceanography, geology, etc.
– monitoring of vegetation cover, particularly, index of vegetation;
– monitoring of emergent situations, etc.

The second example is shown on Figure 135. In this complex high resolution, radar is used – about one meter. Its wavelength is of 9,6 cm), antenna 12 x 3,3 m^2, exploitation period – 5 years.

Figure 134. Example of dual-frequency spaceborne radar with synthetic aperture

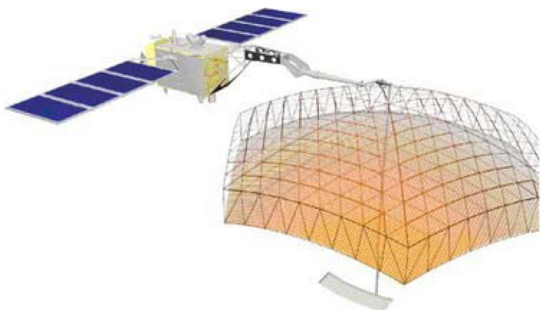


Figure 135. Example of high resolution spaceborne radar with synthetic aperture

Originally, this complex is designed for:

– Earth surface monitoring,
– geological and topographical mapping,
– ocean monitoring and ice patrol,
– environmental monitoring of sea and land surface,
– land use needs,

- emergency operative monitoring,
- navigation control.

Having such properties, this equipment can be the cream of the crop in risk revealing / assessment system.

7.8 Discussion

The description presented shows how joint use, definite combination of GIS-technologies, MCW-radiometry and AI methods should have looked like in multifunctional intellectual monitoring system – GERMES–I, and how they were successfully introduced in multiple projects and situations.

Important elements of this System are mostly implemented and its capabilities were examined in various areas in the World in a framework of projects similar to the System described.

Some of these works were performed for:

- monitoring and control of dangerous hydrological situations in dozens of countries;
- assessment of protection properties of some hydrotechnical constructions;
- risk assessment of forest fire appearance;
- determining the areas of underground moistening with shallow water table;
- detecting zones with high or low water level;
- detecting areas of water seepage from canals and through levees;
- evaluation of the environmental conditions in Bulgaria, Russia, Vietnam and Ukrainian southern regions;
- the geoecological and epidemiological mapping of the Earth, some others.

In sum, we can present the following considerations.

Risk assessment can be represented as process of study and analysis of environmental, technogenic, sanitary, epidemical and other dangerous effects – in order to quantify the probability of harmful environmental factors adverse effects on economic activity, public health, and socio-economic situation in general.

The internationally recognized methodology for risk assessment includes four stages.

1. Identification of risk (hazards): what factors in what levels and exposure pathways, from which media can have adverse effects on economic activity and on human health, how plausible and confirmed the links between factor and consequences are.

2. Evaluation of the exposition – description of pollution sources, routes of contaminants movement from source to man, to accident place, the path and impact point, the levels of exposure, etc. Currently, these objectives are widely used in various geoinformation systems (GIS) and are described in this monograph.

3. Establishing the "dose-reaction-response" correlation. For medical cases, it can be the connection recognition between health (for example, the proportion of individuals, who have developed a specific disease) and levels of exposure.

4. Risk profile and evaluation; that means: generalization and analysis of all data obtained in the descriptive, analytical and experimental studies of the risk estimates for similar kinds of situation, for population and its sub-groups; a comparison of risks of acceptable (suitable) levels, a comparative evaluation and ranking of the various risks in terms of their statistical, environmental, medico-

biological and social significance. The purpose of this phase is – the establishment of institutional and socio-environmental priorities and the risks that must be prevented or reduced to an acceptable level for a given society.

Note that considered in given Chapter algorithms and technical methods successfully work for each of these steps.

The work described and its objectives contribute to the scientific, technical, wider societal and policy objectives in the number of European programs, frameworks and EAS strategies, for example, 6th Framework Programme; Thematic Priority: 1.4 Aeronautics and Space; 2. SPACE; 2.3.2 Area: GMES; Application field: Risk management; etc.

Several years ago the ESA Book: "Down to Earth: Everyday Uses for European Space Technology" was published consisting of description of MCW measurements and application. Concerning this issue we could read short but significant and substantial opinion: "For me the message of the book is doubly valuable, as it shows us that the return on the investment that we make in Europe on space research is being significantly increased by the beneficial improvements in life it brings to us here on Earth" (Lord Sainsbury, Minister for Science and Technology in the UK Government). The time shows how right and correct these words were.

We hope that joint elaboration and international use of the system described will be fruitful and profitable for our countries, will promote and strengthen maritime and inland pollution control and prevention, will decrease existing risks.

It seems obvious that natural and anthropogenic risks are overlapping now more and more, and we are not on the horns of dilemma – we have no alternative: **the Flight to the Planet of Earth must continue; and Flight of the Planet of Earth must continue too.**

8
Microsituation Concept in GMES Decision Support Systems

8.1 Situational modeling for decision making in emergent situations

Situational modeling for GMES represents branch of system-analytical activity where modeling objects are defined such in relation to which the strict, exhaustive description is impossible. Thus, the situational simulations can be compared with the general theory of artificial intelligence in some way.

Nevertheless, in a general view the situation is understood as a combination of characteristics of problem area and prototype system where such characteristics represent interference model of such object with a subject domain. Thus various characteristics of a prototype system and a subject domain of studying of such object can be considered as separate levels of representation of research problem area that allows to speak about separate situations in such modeling (Figure 136).

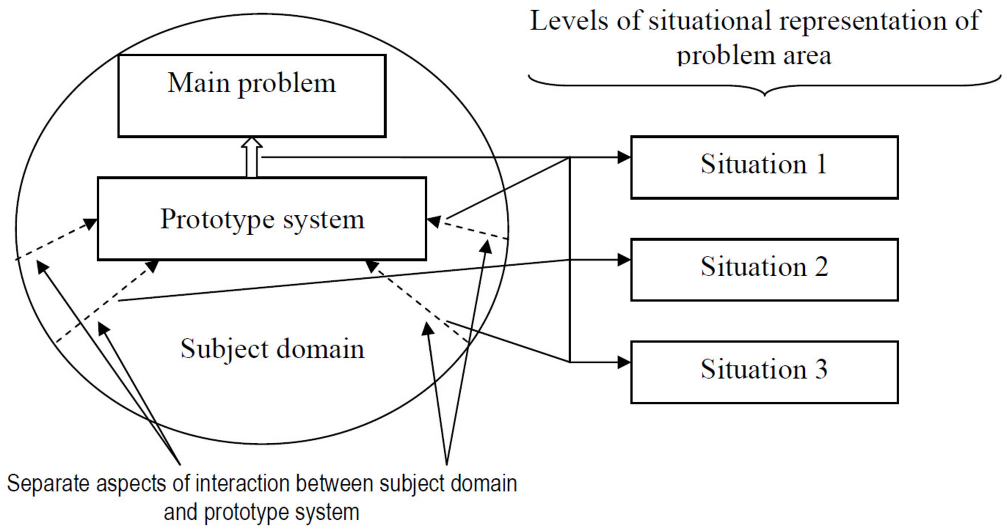


Figure 136. Generalization of separate situations of prototype system

Due to inability to describe whole model, it appears many levels of situational representations of subject domain that involves usage of methods of artificial intelligence for its better understanding. This leads us to speak, that each situation could be represented as a set of microsituations of prototype system. Such assumption is based on idea, that in process of studying object happens superposition of several interaction aspects between subject domain and prototype system due to set of objective reasons. Therefore, for example, if we study reasons of snow avalanches, it is hard to separate influence of environment temperature from influence of snow cover and hill's declivity. In the same time, for example, studying economical dynamics processes of banks we should take into account a great set of factors, which form such dynamics.

A set of situation representation levels of subject domain naturally increases in emergent situations. This is connected with increased influence of subject domain factors on prototype system, independently of emergency conditions. Therefore, for emergent situations of possible snow avalanches microsituations could differ of small differences of temperature or changes in wind speed [Kuzemin and Lyashenko, 2007b], [Kuzemin et al, 2007a; b]. Development of financial crisis as emergency economic situation is characterized by separate characteristics of microsituations of different facilities, involved in economical process [Kuzemin and Lyashenko, 2007a].

Hence, situation includes a set of separate microsituations, each describing properties of prototype system in some characteristic category of its subject domain. Such categories for subject domain could represent internal and external processes, characteristics, advices both from prototype system's point of view and from set of factors, which influence from subject domain to prototype system. Separate set of such characteristics without interconnections could hardly precisely describe a situation, because such characteristics are interconnected with each other, involved in different process of prototype system studying.

In some kind situation could be presented as hierarchy of separate microsituations using oriented graph, built on group descriptor. This allows represent situation as a set of n microsituations c_i:

$$C = \{c_i\},\ i = 1,...,n;\ c_i = \langle e_i R_e \rangle \tag{1}$$

where part of situation, represented by pair $c_i = \langle e_i R_e \rangle$, will be called microsituation, connected with some characteristic of prototype system.

Such representation of situation allows operating with situational model in terms of separate microsituations, which increases effectiveness of decision taking.

A general idea of calculating distance between microsituations is separating common part of two objects being compared and calculating how much each of them differs from common part. For example if we compare two microsituations c_1 and c_2, first of all their common part $c_1 \cap c_2$ is calculated. If microsituation c_1 is connected with some solution D, we should calculate how microsituations c_1 and c_2 match, in order to make decision if solution D is applicable in microsituation c_2. Measure of similarity S is calculated as a percent of common part $c_1 \cap c_2$ for each of microsituations being compared. But it should be taken into account that solution D is connected with microsituation c_1 and in case $c_1 \cap c_2 \subset c_1$ is also based on missing part – $c_1 \setminus (c_1 \cap c_2)$, which is missing from c_2. This fact decreases applicability level of solution D for searching general solution. $c_1 \cap c_2 \subset c_2$ describes for what part of microsituation c_2 solution D is applicable. In other words, similarity level of different microsituations is measure of their unity. In such way, we can perform more precise accounting of changing characteristics of prototype system during decision taking in emergency situation, which leads us to idea of using situation model for prototype system representation. That ultimately allows us to speak about the relationship of situational simulations with the general theory of artificial intelligence.

In next sections, we published a more detailed review on questions of microsituations usage for risks management in emergent situations of avalanche snowfalls and development of economical crisis an emergency economical situation

8.2 Microsituation as a base part of risks management in risks management of possible snow avalanches

✓ Methods of forecasting avalanche-dangerous situations and flaws of common approaches

Snow avalanches have a special place of natural emergent situations. They cause moving snow masses perform huge destructions in all 6% of earth surface, where avalanche-dangerous regions are located.

As a whole, one of the directions, i.e. the risk of emergent situations management and, in particular, situations initiated by the snow-slip, should be considered constant monitoring and building of interpretation models for prediction of such situations initiation. Hereinafter, such models form the basis for the system of decision–making support; this is favorable to development recommendations on modern performance of maintenance measures directed to natural calamities prevention.

Among the most essential and important problems in the given aspect one should note substantiation of the utility to use the corresponding mathematical apparatus intended both for investigation of the avalanche dangerous situations development dynamics and for development of methods for estimation of the potential avalanche cells, prediction of avalanches volumes and descent frequency. This concerns the fact that every avalanche can be regarded as a unique phenomenon of nature with its specific peculiarities. At the same time despite its uniqueness, it is possible to single out the climatic conditions variations characteristic ranges, which are prerequisites to prediction of the feasible avalanche descent. Eventually, the totality of these two factors defines the presence those approaches to prediction and warning of avalanches descent, at present these approaches are used in geoinformation systems (GIS) which make it possible to accumulate continuously meteorological information, carry out various calculations, reveal regularities and realize spatial tie of the obtained results [Durand,1993] and [Kuzemin and Toroev, 2006].

Considering methods and models of avalanches descent prediction the method images of similarity and regression analysis are singled out the most often [Buser et al, 1987]. At the same time, there is no doubt that the foundation of avalanche dangerous situations initiation prediction consists in the procedure of the preliminary analysis of such events. In this case, as a rule, the solution of the formulated problem is based on the statistical analysis methods. In particular, the approaches of such analysis make it possible to substantiate the most significant system of the facts, which is expedient to use in the avalanche-dangerous situations prediction procedures. Such approaches found their development in the predictions of snow avalanches descent based on application of the nearest neighbors' method or through the application of the regression equations [Buser et al, 1987]. But results of the prediction obtained with such methods are not always applicable and demonstrate a number of shortages: they require significant computational resources; they don't embrace existing variety of causes resulting in avalanches formation. The shortage is also the impossibility to define the degree of the avalanche hazard, number and dimensions of avalanches [Fuhn, 1998].

The data of nomograms, which in a general case extend the interconnection of such indices as temperature, value of snow cover, and precipitations are also used for estimation of the avalanches descent probability. Nevertheless, in spite of this the remaining non-predictive nature of the avalanche dangerous situation does not always allow to prevent negative consequences of emergencies caused by their descent. This is associated with that the available procedures of the avalanche dangerous situation initiation prediction are not sufficiently precise. At the same time, the severity of the problem and variety of the ways to solve it motivate the necessity to search alternative methods, which can give more argued answers.

✓ Representation of avalanche-dangerous and avalanche-safe situations

Analysis of different characteristics of the avalanche climate initiation medium makes generally the foundation of the avalanches descent prediction. Among such characteristics the most abundant ones are: the air temperature, humidity, atmospheric precipitations volume, wind velocity, angle of the slope of surface (descent angle) where the avalanche descent is possible. In general, the variation dynamics of both individual of the above characteristics the avalanche climate initiation and their totality can, with some probability, characterize initiation either avalanche-dangerous or avalanche non-dangerous event. As this takes place, a feasible range of the studied avalanche climate initiation characteristics variations describes a definite region of avalanche-dangerous and avalanche-non-dangerous situation.

Consider we have an avalanche-dangerous situation Ω_L and avalanche-safe situation Ω_N. That situations could be described as set of most statistically important characteristics X_1, X_2, X_3, X_4, X_5, where, for example, X_1 — air temperature; X_2 — air humidity; X_3 — wind speed; X_4 — amount of precipitation; X_5 — angle of hill slope.

By-turn situation Ω could be represented as a set of micrositutions $\Omega = \{\omega_i\}, i = 1,...,n$ each corresponding to a definite group of reviewed types of environmental data corresponding avalanche climate and reflecting, from one side, probability of avalanche-dangerous situation and from another – probability of avalanche-safe situation. In other words avalanche-dangerous situation Ω_L could be presented as homogenous unity of different micrositutions $\Omega_L = \bigcup_i \Omega_{L_i}$, and avalanche-safe situations represents unity $\Omega_N = \bigcup_i \Omega_{N_i}$. Each such micrositution represents probability of avalanche-dangerous or avalanche-safe situation in general.

In that way, estimation results could be interpreted as a base for forming according systems of micrositutions forming data for knowledge base of informational-analytic system of crisis situations management.

In the conceptual plan, the essence of the probabilistic aspect of the avalanche climate initiation analysis can be reduced to the definition of the probability to assign some point as the considered medium current characteristics either to the region of the avalanche initiation or to the initiation of avalanche non-dangerous situation. Otherwise, the given approach can be treated as a correspondence of the current characteristics of the avalanche climate initiation medium; parameters of these characteristics define some region using probabilistic distribution of avalanche dangerous or avalanche non-dangerous situations preceding this. Consequently, it is possible to speak about so-called probable conformity of the researched characteristics of the avalanche-dangerous climate environment to probabilistic distribution of the avalanche-dangerous or avalanche non-dangerous situations.

In particular, procedure of such analysis can be considered proceeding from the pair wise analysis of various characteristics of the avalanche climate initiation. The advisability of such a transition is related to the fact that at the stage of the preliminary analysis it is possible to omit less significant factors of impact on the avalanche dangerous situation initiation. Thus, the base element of the analysis procedure being considered is estimation of the probability of the avalanche climate initiation current parameters to fall within the regions typical and atypical for the avalanche climate initiation. The given regions can be presented in the plane in the form of the rectangle; its metric values correspond to definite parameters of variation of the avalanche- dangerous and avalanche non-dangerous situations initiation medium characteristics (Figure 137) [Kuzemin et al, 2007b].

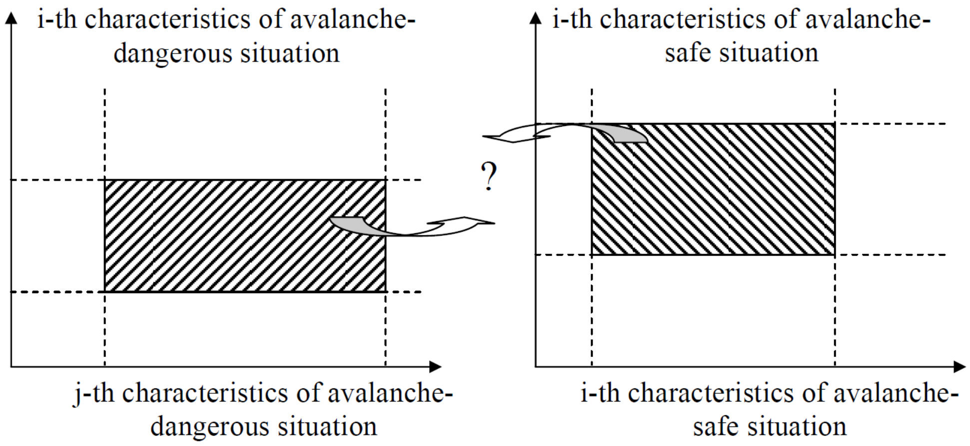


Figure 137. On the explanation of the probabilistic aspect of the avalanche climate initiation medium analysis

In addition, we should note, that reviewed characteristics of environment of avalanche climate, in general follows normal law of distribution. This allows using that low for preliminary calculations. In Table 5 the corresponding probabilities are presented taking into account avalanche-dangerous and avalanche-safe situations.

Table 5. Probabilities of correspondence of the current parameters of avalanche climate initiation medium to the avalanche dangerous and avalanche non-dangerous situations

Characteristics of the avalanche climate initiation medium analysis	Feasibility of correspondence
under condition of considering the avalanche dangerous situation and avalanche dangerous current parameters	
air temperature – wind velocity	0,854
air temperature – wind velocity	0,823
wind velocity – precipitations quantity	0,707
precipitations quantity – descent angle	0,809
under condition of considering the avalanche-dangerous situation and avalanche- non-dangerous current parameters	
air temperature – wind velocity	0,488

air temperature – wind velocity	0,582
wind velocity – precipitations quantity	0,317
precipitations quantity – descent angle	0,341
under condition of considering the avalanche-non-dangerous situation and avalanche- non-dangerous current parameters	
air temperature – wind velocity	0,798
air temperature – wind velocity	0,878
wind velocity – precipitations quantity	0,866
precipitations quantity – descent angle	0,939
under condition of considering the avalanche-non-dangerous situation and avalanche- dangerous current parameters	
air temperature – wind velocity	0,555
air temperature – wind velocity	0,482
wind velocity – precipitations quantity	0,403
precipitations quantity – descent angle	0,591

As can be seen from the data in Table 5 the assumptions made above are reasonably justified i.e. the probability of correspondence of the like situations and parameters is essentially significant. This allows generalizing even for estimation of probable initiation of the avalanche dangerous as a whole. To do this one should consider:

- either generalization of the obtained probabilities reasoning from the significance of different groups of characteristics of the avalanche climate initiation characteristics analysis in the assumption that the probabilities of correspondence can be considered as conditional probabilities of the concrete situations analysis;
- or a separate group of characteristics of the avalanche climate initiation medium analysis based on the greatest/least values of the correspondence probabilities.

✓ Forming of avalanche-dangerous and avalanche-safe micrositauations

For demonstrating proposed approach and proving its effectiveness let's analyze probability aspects of several characteristics of avalanche-dangerous climate using real snow avalanches data in Itagar-Chichkan region of Kirgizstan republic during 2001–2006 years. A main idea of such analysis is determination of wrong estimation – amount of avalanche-dangerous situation, which will be classified as avalanche-safe, and vice-versa.

For preliminary data analysis, we perform analysis of environmental statistical characteristics for forming set of data, which describes the probability of avalanches in general:

$$Z = \{X_1, ..., X_n\} \tag{2}$$

where Z – generalized characteristics of avalanche appearance.

$X_1, ..., X_n$ – set of most significant statistical characteristics connected with avalanche fall.

Using set 2 we determine interconnections of characteristics leading to avalanche and avalanche-safe situations and form groups of such characteristics in general (separate microsituations):

$$\{X_1,...,X_n\} \Rightarrow \begin{cases} \beta_1^L,...,\beta_n^L \\ \beta_1^N,...,\beta_n^N \end{cases} \quad (3)$$

where $\beta_1^L,...\beta_n^L$ – set of most valuable interconnections between factors and characteristics of avalanche appearance taking into account probability of avalanche appearance (avalanche-dangerous situations)

$\beta_1^N,...\beta_n^N$ – set of most valuable interconnections between factors and characteristics of avalanche appearance leading to no avalanche happening (avalanche-safe microsituations)

Each microsituation is described by a probability range $((\delta_i^d \div v_i^d)$, where d – separate microsituation, i – separate group of characteristics of avalanche appearance in general. If it is needed it is possible to take into account an overall dynamics of changing characteristics of possible avalanche appearance:

$$\left.\begin{matrix} \beta_1^L,...,\beta_n^L \\ \beta_1^N,...,\beta_n^N \end{matrix}\right\} \Rightarrow \begin{cases} \lambda_1^L,...,\lambda_n^L \\ \lambda_1^N,...,\lambda_n^N \end{cases} \quad (4)$$

where $\lambda_1,...,\lambda_n$ – set of adjustments of most valuable factors with different microsituations characteristics leading to avalanche appearance.

In such manner, we perform a generalization of data groups (situations classes or simply microsituations). Corresponding results for forecasting avalanche-dangerous and avalanche-safe situations are presented in Table 6 [Dyachenko et al, 2007].

Table 6. Results of forecasting of avalanche-dangerous and avalanche-safe situations

Groups of microsituations	Probability of classifying as correct avalanche-dangerous class	Probability of classifying as correct avalanche-safe class
X_1, X_2, X_3, X_4	0,54–0,82	0,51–0,87
X_1, X_2, X_4, X_5	0,50–0,87	0,52–0,77
X_1, X_2, X_3	0,52–0,77	0,51–0,70
X_2, X_3, X_5	0,51–0,69	0,50–0,73
X_2, X_3, X_4, X_5	0,51–0,96	0,50–0,80

Presented in Table 6 range of probabilities of correct classifying to any class of situation could be interpreted as some integral characteristic of corresponding microsituation. In other words each data group of avalanche-dangerous situation (Ω_L) and avalanche-safe situations (Ω_N) are the set of microsituations:

$$\left.\begin{matrix} \Omega_L \\ \Omega_N \end{matrix}\right\} \rightarrow < X_k, Z, R \rightarrow \begin{cases} \delta_i^L \div v_i^L \\ \delta_i^N \div v_i^N \end{cases} \quad (5)$$

where X_k – separate characteristics of avalanche appearance,

Z – generalized characteristic of avalanche appearance,

R – most valuable interconnections of avalanche appearance.

Thus, using this assumption and data presented in Table 6, avalanche-dangerous situation could be presented as a set of separate integral characteristics of such microsituations. As a confirmation of such generalization, we could serve some procedure of microsituations comparison. But, since not all microsituations are following normal distribution law, in order to check comparison hypothesis non-parametric tests are more suitable.

One of the possible tests, which could be used, is Wilcoxon test for connected sets, which answers a question: if any event in analyzed data happens, which leads to significant change in microsituation hierarchy [Prosvetov, 2005]. Wilcoxon test could be presented as such equation:

$$w_{g,j} = \sum_{m} \left| \lambda_g^m - \lambda_j^m \right|; g, j \in \{i\}; m = 1,...,n; \lambda_g^m, \lambda_j^m \in \{\Omega_L, \Omega_N\} \quad (6)$$

In other words, distinguishability between different microsituations is a subject of this research. The value of such test could be interpreted as a difference measure of reviewed microsituations.

The more value of Wilcoxon test, the more distinguish reviewed microsituations are.

Results of Wilcoxon test are presented in Table 7 (values 1-5 correspond to avalanche-dangerous situations, 6-10 – avalanche-safe one)

Table 7. Wilcoxon test value

	1	2	3	4	5	6	7	8	9	10
1	–	0,920	2,041	2,705	0,242	1,534	2,463	3,441	2,995	1,413
2	0,920	–	1,689	2,671	0,634	1,778	2,082	2,064	2,225	1,528
3	2,041	1,689	–	0,217	1,968	0,568	0,169	1,618	2,089	0,942
4	2,705	2,671	0,217	–	2,263	1,355	1,087	1,005	1,355	2,136
5	0,242	0,634	1,968	2,263	–	1,044	2,596	2,624	3,087	0,886
6	1,534	1,778	0,568	1,355	1,044	–	0,157	1,868	2,371	1,174
7	2,463	2,082	0,169	1,087	2,596	0,157	–	2,149	2,101	0,266
8	3,441	2,064	1,618	1,005	2,624	1,868	2,149	–	0,198	2,354
9	2,995	2,225	2,089	1,355	3,087	2,371	2,101	0,198	–	3,556
10	1,413	1,528	0,942	2,136	0,886	1,174	0,266	2,354	3,556	–

Data analysis of Table 7 results in a conclusion where obtained values of Wilcoxon test are significant for microsituations, which belong to different classes. This allows generalizing comparison of different microsituations. Let us review an integrated Wilcoxon criteria value (as sum of partial test values) from each microsituation to class of avalanche-dangerous and avalanche-safe situations. Corresponding values are presented in Table 8.

Analysis of Table 8 shows, that for avalanche-dangerous microsituations integral test value with avalanche-dangerous situations is lower than integral test value with avalanche-safe situations and vice-versa. That tells us, that in general avalanche-dangerous and avalanche-safe situatios are presented with homogenous microsituations.

Table 8. Integral Wilcoxon test values

Microsituations	Integral test value with avalanche-dangerous situations	Integral test value with avalanche-safe situations
1	5,908	11,846
2	5,914	9,677
3	5,915	5,386
4	7,856	6,938
5	5,107	10,237
6	6,279	5,570
7	8,397	4,673
8	10,752	6,569
9	11,751	8,226
10	6,905	7,350

Graphical interpretation of Table 8 (Figure 138) allows to establish a definite procedure of microsituations separation and to perform analysis of new microsituations using approach, which would be presented in next section.

✓ Avalanche-dangerous microsituation classes

Therefore, representing a set of factors and current nature environment state as a microsituation allows increasing stability of avalanche-fall forecasts. Every such a microsituation corresponds to a definite combination of factors of the avalanche initiation environment.

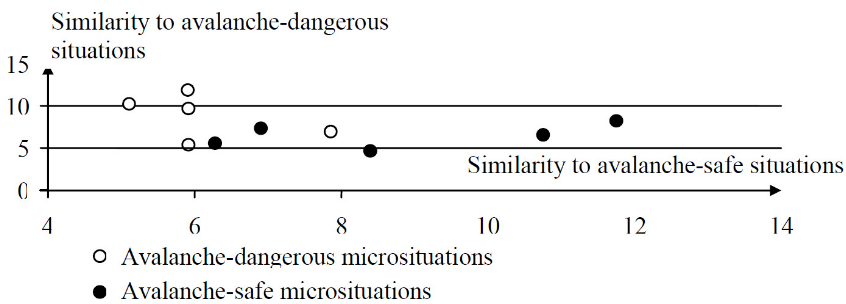


Figure 138. Position of microsituations in situations space

At the same time, such representation makes it possible to break up the whole set of causes affecting the avalanches initiation into two subclasses. One of subclasses characterizes a set of microsituations reflecting the avalanche initiation and the other subclass is typical for non-avalanche situation as a whole. Then the emergency avalanche situations risks management can be presented as a generalized description of the system with the help of a totality of different microsituations. Based on such interpretation the logical rules of the analyzed data set generalization for further their subdivision into classes of avalanche dangerous and non-avalanche dangerous situations:

$$\text{Avalanche dangerous}= \begin{array}{l}(\{F_L^L(X)\}/\{F_N^L(X)\})\cup(\{F_L^L(X)\}/\{F_N^N(X)\})\cup \\ (\{F_N^N(X)\}/\{F_L^L(X)\})\cup(\{F_L^N(X)\}/\{F_N^L(X)\})\cup \\ (\{F_N^L(X)\}/\{F_L^N(X)\})\cup(\{F_L^N(X)\}/\{F_N^N(X)\})\end{array} \quad (7)$$

$$\text{Non avalanche dangerous}= \begin{array}{l}(\{F_N^L(X)\}/\{F_L^L(X)\})\cup(\{F_L^L(X)\}\cap\{F_N^N(X)\})\cup \\ (\{F_L^N(X)\}\cap\{F_N^L(X)\})\cup(\{F_N^N(X)\}/\{F_L^L(X)\})\end{array}$$

where $F_L^L(X)$ ($F_L^N(X)$), $F_N^L(X)$ ($F_N^N(X)$) – probability function of referring avalanche dangerous (non-avalanche dangerous) micrositation to the avalanche dangerous (non-avalanche dangerous) class, respectively, on the set of factors of the avalanche danger initiation X [Kuzemin et al, 2007a].

Model 7 describes following possible choices of warning on possible avalanches:

Choice 1. Avalanche-dangerous set is formed as a subtraction of sets, limited by probability polynomials $\{F_L^L(X)\}/\{F_N^L(X)\}$. Avalanche-safe set — $\{F_N^L(X)\}/\{F_L^L(X)\}$ (Figure 139).

Choice 2. Border of avalanche-dangerous set is corresponding to symmetrical subtraction $(\{F_L^L(X)\}/\{F_N^N(X)\})\cup(\{F_N^N(X)\}/\{F_L^L(X)\})$. Avalanche-safe set is an intersection $(\{F_L^L(X)\}\cap\{F_N^N(X)\})$ (Figure 140).

Choice 3. Border of avalanche-dangerous set is $(\{F_L^N(X)\}/\{F_N^L(X)\})\cup(\{F_N^L(X)\}/\{F_L^N(X)\})$. Avalanche-safe set is an intersection of two sets: $(\{F_N^N(X)\}\cap\{F_L^L(X)\})$ (Figure 141).

Choice 4. Border of avalanche-dangerous set is expressed as subtraction $(\{F_L^N(X)\}/\{F_N^N(X)\})$, avalanche-safe set — $(\{F_N^N(X)\}/\{F_L^L(X)\})$ (Figure 142).

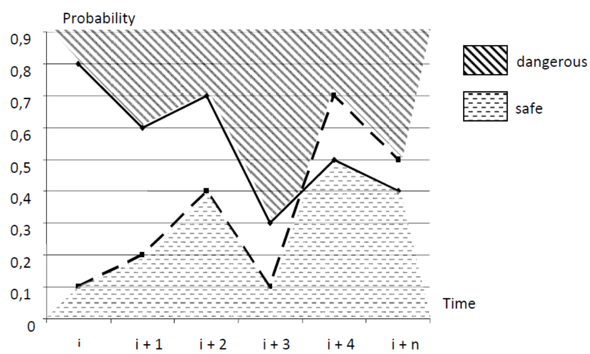


Figure 139. Probability distribution for 1-st version

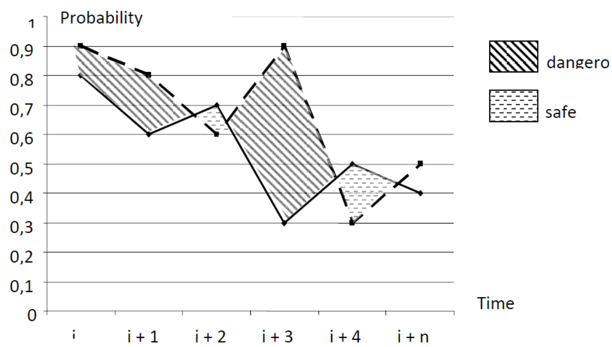


Figure 140. Probability distribution for 2-nd choice

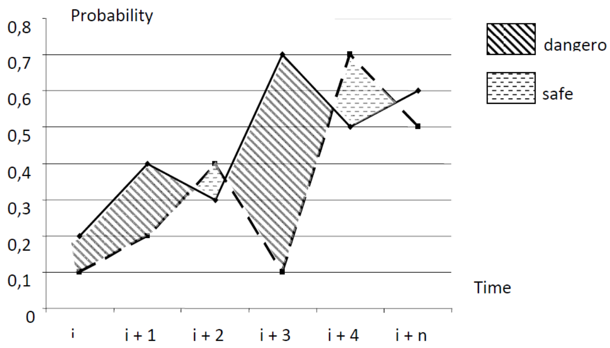


Figure 141. Probability distribution for 3-rd choice

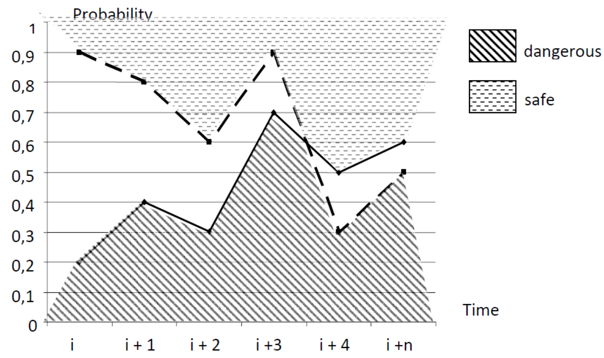


Figure 142. Probability distribution for 4-th choice

So, received probabilities allows us to answer quickly in which class belongs current situation (micrositeuation) and provides a base to review and analyze sets of avalanche-dangerous situations. Analysis of such sets allows to formalize sets of avalanche-dangerous and avalanche-safe situations and to build procedure of corresponding forecasting actions. At the same time presentation of the avalanche danger factors in the form of the microsituations classes allowed to get an objective correspondence between the probabilistic estimates of the avalanches descent and the avalanche danger scale degrees; eventually this makes it possible to correct time of the prediction system response to the possible avalanche descent. Essence of such estimate consists in construction and analysis based on the theory of fuzzy sets, corresponding functions of prediction time correction $\mu(X)$ (Figure 143) [Kuzemin et al, 2007b].

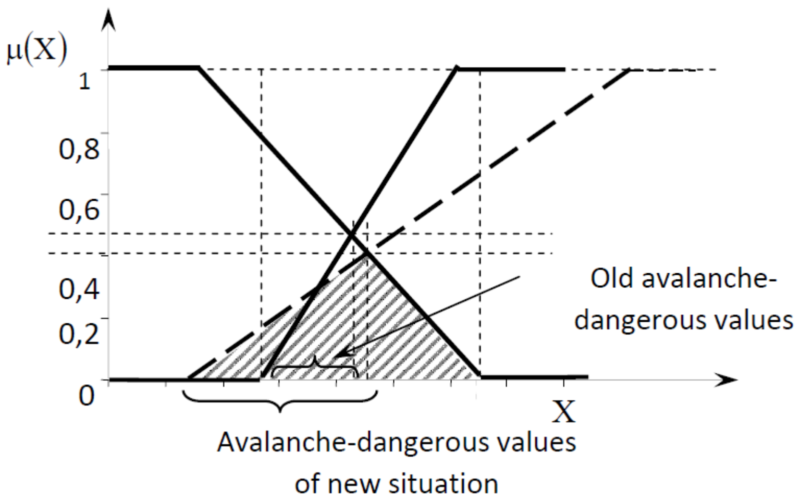


Figure 143. Methods of the fuzzy sets theory as the basis for correction of the prediction system time response to the possible avalanche descent

As a whole, the essence of such description reduces to construction of the fuzzy model for estimating temporal characteristics of the avalanche dangerous situations. Thus, for example, it is possible to suppose that in the case of the avalanche non-dangerous situation analysis the time until the assumed avalanche descent will be the more the greater is the probability of referring current parameters of analysis of the avalanche climate initiation environment to such situation. Respectively, in the case of insufficient probabilities of referring of the avalanche climate initiation environment analysis current parameters to such situation can testify to insignificant reserve of time until the avalanche descent moment. When considering the avalanche dangerous situations the corresponding characteristics are the opposite. This allows introducing into definition the functions of the avalanche descent time fuzzy set; their generalization makes it possible to predict the avalanche descent time.

It is possible to pass to the distribution functions in the estimate of the avalanche descent time based on the corresponding probabilities analysis of the fuzzy set available data separation into avalanche dangerous and avalanche non-dangerous situations.

Thus, the model construction generalized scheme and construction of the procedure for prediction of the avalanche-dangerous situations initiation is reduced to:

- sequential obtaining of the probabilistic characteristics of the avalanche climate initiation medium;
- construction of the corresponding sets of subdivision into avalanche-dangerous and avalanche-nondangerous situations;
- analysis of the avalanche descent initiation time using fuzzy models of its interpretation.

8.3 Microsituation as a base for comparative analysis of financial flows

Variety of emergency economic situations through prism of financial flows generalization. Introduced term "microsituation" could be extended also for performing analysis of financial flows of different economic users. In other words, objective laws of economic development consider constant interference of economic agent with different sides of external environment. The changes in environment results in rebuilding of internal structure of economic agent, cause this process leads to stability of economic agent and overriding economic crisis factors. In the same time effectiveness of financial flows could be increased introducing analysis subsystem. Direction and speed of financial flows movement demonstrates transiency of economic processes and speed of changes in external environment of reviewed economic agent. Thus, for example:

- transiency of economic processes closely connected with unpredictability and errors in forecasts about future economic development;
- globalization influence on overall economic processes all over the world. This could be easily confirmed by visual identity of index EMBI+ (The Emerging Markets Bond Index Plus, Figure 144) of such countries as Poland and Bulgaria, which allows assuming a possibility of same tendencies in financial flows. Nevertheless, for more significant conclusions a quantitative estimations is required, which leads to need of reviewing different levels of affecting of researched processes, which, by-turn brings us to review of separate microsituations;
- possibility of review of unexpected economic crisis as a set of specific situational aspects of management, which influence on effectiveness of taken decision;
- effectiveness of financial flows analysis greatly depends on existing system of their estimation and system of taking decision on those estimations, which directly connected with the process of formalization and analysis of many financial flows characteristics and, by-turn, leads again to microsituation statement in financial sphere. Necessity of such review is connected with limited applicability of traditional models for decision taking in social-economic systems [Voloshin, 2005].

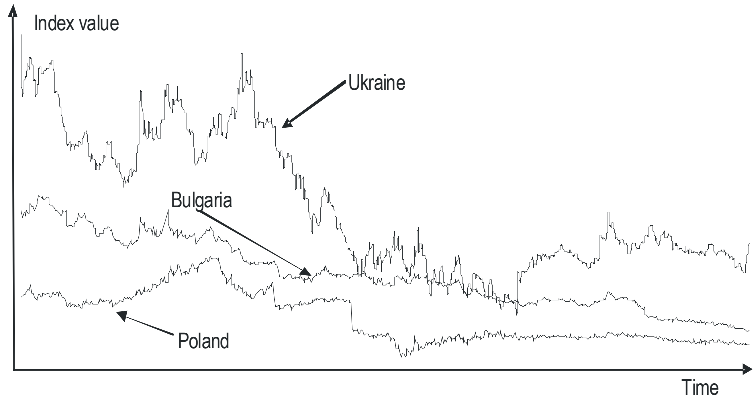
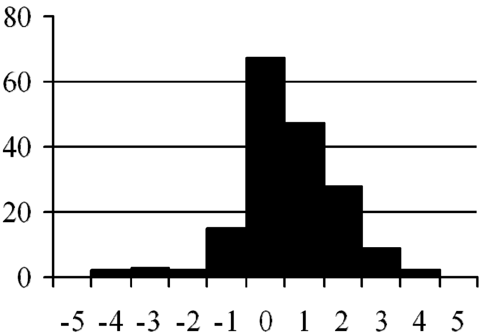
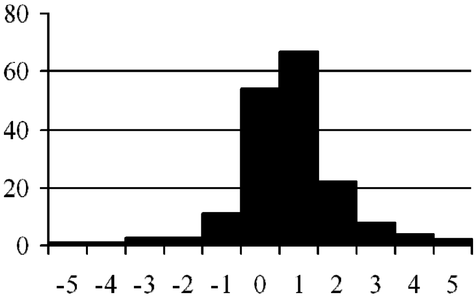


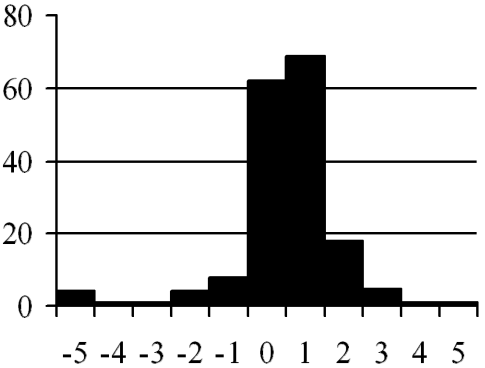
Figure 144. Dynamics of EMBI+ index change for several countries



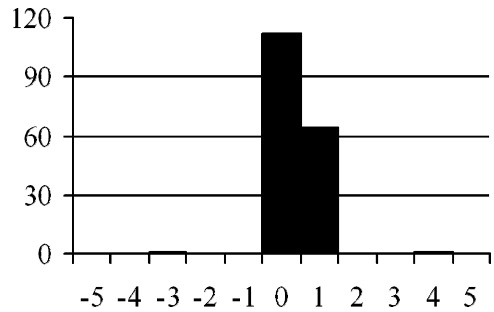
a) KAC-20



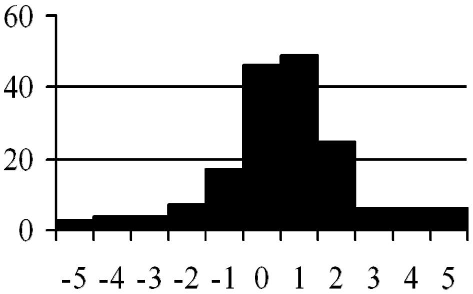
b) SOKRAT



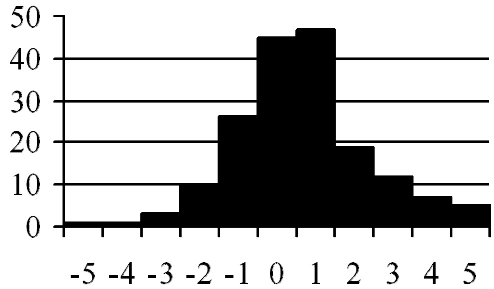
c) Sofia Bondar Priadka Ukraine



d) TEKT-KKOC Price Bonds



e) PFTS



f) KINDEX

Figure 145. Density of profitability distributions of some Ukrainian indexes

In the same time, complexity of solving economic tasks increases with the use of different estimations for the same processes. As an example of such estimation, we could name a set of tasks, connected with tendency estimation in funding market development, which from one side could be hardly forecasted for countries with transition economy, and for another side, is vitally important for taking decision in sphere of economic safety, cause it describes processes of reassignment of free money and financial resources.

Nevertheless, analysis and correct conclusion about processes is complicated because of different statistic characteristics of separate economic factors, describing the same processes (Figure 145) [Kuzemin and Lyashenko, 2006]. In addition, it is vitally important to estimate an interconnection influence of financial flows from different economic agents, which happening both on micro- and macro-levels [Kuzemin et al, 2005].

Therefore, in process of financial flow analysis we need to create a database, which in fact needs to be converted to real knowledge base. If we revive a structure of such process (Figure 146) it can be characterized as a procedure of receiving new informational images, which, in next step are grouped and analyzed in some kind, depending on microsituations in financial sphere.

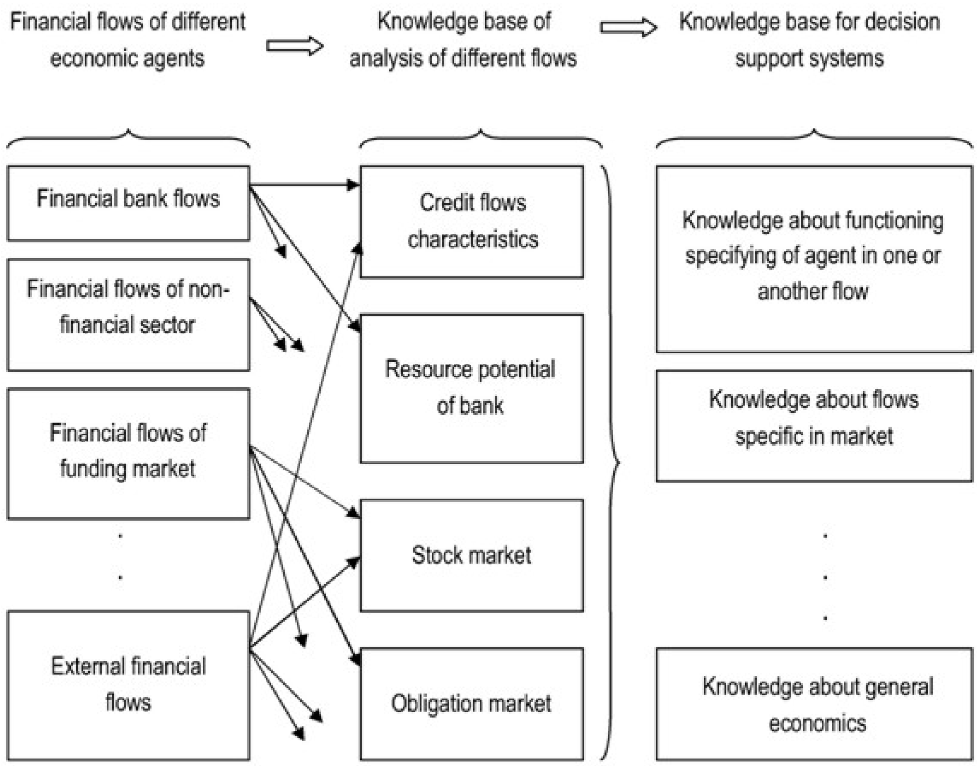


Figure 146. Generalized structure of process forming knowledge base in decision support system based on analysis of different financial flows, characterizing separate microsituations

8.3.1 Classical interpretation of effective bank management task and microsituation as base for performing comparison of bank functioning

One among possible approaches for comparative estimation of bank system or separate banks is comparison of their liquidity and profit-generation levels. Base idea of such comparison is connection between liquidity and profit-generation potential, which means that more risky bank operations could lead to extra income. Thus, when considering the probabilistic interpretation of banking activity management starting from a definite liquidity level one should take into account the fact that the bank tends to support the liquid assets volume at the level sufficient to ensure meeting previously taken commitments. At the same time, the bank defines the probability of the need for loan resources to meet its commitments. Then, for example, the interpretation of the banking system development based on the liquidity analysis can be considered as a probability for a random two-dimensional value to penetrate into some specified field where acceptable and admissible liquidity and profitability levels parameters manifest themselves as boundaries of such a field.

Thus, when considering the probabilistic interpretation of banking activity management starting from a definite liquidity level one should take into account the fact that the bank tends to support the liquid assets volume at the level sufficient to ensure meeting previously taken commitments. At the same time, the bank defines the probability of the need for loan resources to meet its commitments. Then, for example, the interpretation of the banking system development based on the liquidity analysis can be considered as a probability for a random two-dimensional value to penetrate into some specified field where acceptable and admissible liquidity and profitability levels parameters manifest themselves as boundaries of such a field.

The classical interpretation of the bank management efficiency can be considered in terms of the fuzzy sets theory. The given approach becomes possible through introduction into consideration of the ownership function of some set of the bank liquidity and profitability indices corresponding to a subset of efficient managing actions of the given indices.

Then, for example, the fuzzy interpretation of the bank management efficiency in the specified phase space is limited to building and estimation of the corresponding ownership functions characterizing the degree of reaching the bank efficient management in the specified variation intervals of the banking activity being analyzed. In this case, it is expedient to choose a fuzzy interpretation of the intended parameters variations in the limits of the admissible values of liquidity and profitability indices presented in the probabilistic model by the corresponding probabilistic curve as a formal description of such functions. The advisability of such a transition is motivated by that the fuzzy formalization of the corresponding probabilistic curve is possible based on the concept of the fuzzy number of L-R type, which in the given case can be regarded as a trapezoidal fuzzy number. Such an interpretation of the ownership function makes it possible not only to describe the processes under investigation formally but also to take into account existing economical aspects in their development.

As be marked before, for conducting of comparative analysis of functioning of banks an important instrument is the use of the finance flows, which makes it possible to give the most complete description of the banking on the basis of multiple presentation of the initial data (separate indices of activity) x_t^γ of their sets of γ at a certain temporal interval t in terms of the finance flows – $\{x_t^\gamma\}$.

This is associated with that the basis of the flow approach comprises the possibility to realize the structuring of data for complex dynamic systems; it is precisely the structuring that opens different directions for carrying out the necessary analysis [Kuzemin and Lyashenko, 2007a].

At the same time, the flow processes involve all spheres of the market economy, this is rather important as far as the banks is concerned as the centers of redistribution of monetary and reallocation of capital. This also allows taking into account the degree of various environment factors action, governing thereby the information saturation of the indices being considered.

It should be noted in this case that the flow approach can serve not only as the set of instruments for the banks functioning and development, but also act as the combining center of various approaches applications for carrying such analysis.

In the given aspect, to perform the comparative analysis of the banks functioning and development based on the flow approach, by the microsituation, variety of the banking description with the help of the corresponding parameters and indices should be meant. In this case the concrete microsituation S^L can be described in the form of a separate finance flow or some set of them being defined with a set of data $\gamma,(\gamma=1,...,m)$, characterizing the banking of some bank $L,(L=1,...,n)$:

$$S^L = (\{x_t^{k1}\})^L, k1 \in \gamma \tag{8}$$

$$S^L = (\{x_t^{k1}\},\{x_t^{k2}\},\{x_t^{k3}\})^L, k1,k2,k3 \in \gamma \tag{9}$$

Thus, comparing banks between themselves we, first, compare the microsituations, which in the given case describe the state of the banks functioning and development in terms of some parameter or their totality:

$$S^1 = \left(\left\{x_t^{k1}\right\},\left\{x_t^{k2}\right\},\left\{x_t^{k3}\right\}\right)^1 \approx S^2 = \left(\left\{x_t^{k1}\right\},\left\{x_t^{k2}\right\},\left\{x_t^{k3}\right\}\right)^2, 1,2 \in L, \tag{10}$$

where S^1 – is the microsituation describing the first of the banks being analyzed,

and S^2 – is the microsituation describing the second of the banks being analyzed.

At the same time it is possible to carry out comparison of the banks development and functioning as a whole fixing parameter t. Then, in the given case, variation of some of the banking parameters being analyzed x_{tp}^{γ} for a fixed date tp in terms of the whole variety of banks – $\left\{x_{tp}^{\gamma}\right\}^L, L,(L=\overline{1,n})$ is considered as a finance flow.

In this case, the concrete microsituation can be presented in the following form:

$$S_{tp}^L = \left\{x_{tp}^{\gamma}\right\}^L, tp \in t, \gamma, \left(\gamma = \overline{1,m}\right) \tag{11}$$

or

$$S_{tp}^L = \left(\left\{x_{tp}^{\gamma}\right\},\left\{x_{tp}^{\gamma}\right\},\left\{x_{tp}^{\gamma}\right\}\right)^L, tp \in t, \gamma, \left(\gamma = \overline{1,m}\right). \tag{12}$$

Then the comparison consists in performance of the analysis between the microsituations describing the state of the banking system functioning as a whole at some fixed dates of time:

$$S_{tp1}^L = \left(\left\{x_{tp1}^{\gamma}\right\},\left\{x_{tp1}^{\gamma}\right\},\left\{x_{tp1}^{\gamma}\right\}\right)^L \approx S_{tp2}^L = \left(\left\{x_{tp2}^{\gamma}\right\},\left\{x_{tp2}^{\gamma}\right\},\left\{x_{tp2}^{\gamma}\right\}\right)^L, tp1, tp2 \in t \tag{13}$$

Ultimately, we receive some set of microsituations $\Omega = \left\{S^L, S_{tp}^{\gamma}\right\}$, completely describing functioning and development of the banking system. Since, as mentioned above, not all microsituations can have the normal distribution, then we shall consider nonparametric tests to verify the hypothesis for coincidence of the microsituations being investigated. In the given case, it is expedient to use the test Wilcoxon for bound samplings [Kuzemin et al, 2005], which answers the question: whether some event essentially changing the microstructure hierarchy took place in the analyzed data, which characterize different samplings.

In other words, when carrying out the comparative analysis of banks functioning and development the analyzed microsituations distinguishability is studied. Then the value of the Wilcoxon test can be used as the measure of distinction (agreement) of the microsituations being considered. The greater is the value of the test being considered, the more distinguishable as a whole are the microsituations being considered and vice versa, the less is the value of the test being considered the closer are the microsituations being considered.

8.3.2 The initial data and results of the comparative analysis of banks functioning in Ukraine

The foregoing approach is being considered as an example of the banking in Ukraine in terms of such index as a share of the granted credits in the overall totality of bank assets.

The paramount importance of consideration of such banking values is associated with that just the credits

- on the one hand, constitute a considerable part of bank operations and, respectively, operating profits in total gains of a bank from such operations,
- on the other hand, the granted credits growth results in credit risks and, consequently, in the banks development destabilization.

Thus, the problem associated with the succession of the development dynamics of relation between the granted credits and total volume of banks assets both for the banking system as a whole, and in terms of separate banks functioning is rather significant. The more so the generalized dynamics of the relation between the granted credits and total volume of banks assets as a whole is indicative of the rise in the banks preferred weight with the increased part of the granted credits in their assets volume (Figure 147, generalized using the site www.finance.ua.). Hence the essence of the first question as to carrying out the comparative analysis of the banking activity consists in estimation of the succession in variation of the granted credits preferred weight in their assets volume during each year of the period being investigated. To analyze such a succession is possible because of investigation of the microsituations each of them describes the state of the banking system functioning as a whole for the fixed date of time tp in terms of the banking activity index x_{tp}^{γ} – the credits preferred weight in the banks' assets (see Eq. 8). The results of such investigation obtained within the periods of 2004, 2005, 2006 and 2007 years in section of each month represent a separate microsituation shown in Figure 148 (generalized based on the above approach and data of the site www.finance.ua). In this case, the black circles mark the microsituations the most consistent between themselves, the microsituations less consistent are not shown at all.

The dimension of each circle represents the degree of correspondence (consistency) of microsituations being investigated in section of every month of the years covered. The smaller is the circle, the greater is the consistency between the microsituations.

In other hand, the analysis performed according to the above methods of consistency in development of separate banks is no less interesting in the considered aspect. To perform such an analysis let us consider a group of 12 banks representing those representing and operating in the same region that makes it possible to consider indirectly the action of various factors on their functioning and development. For the microsituations, their comparison will represent consistency of separate banks development, generalization of their finance flows appears, this represents the specific weight of credits in the structure of such banks assets. Further comparison is carried out based on Wilcoxon criterion accordion to Eq.5. Figure 149 shows results of such consistency.

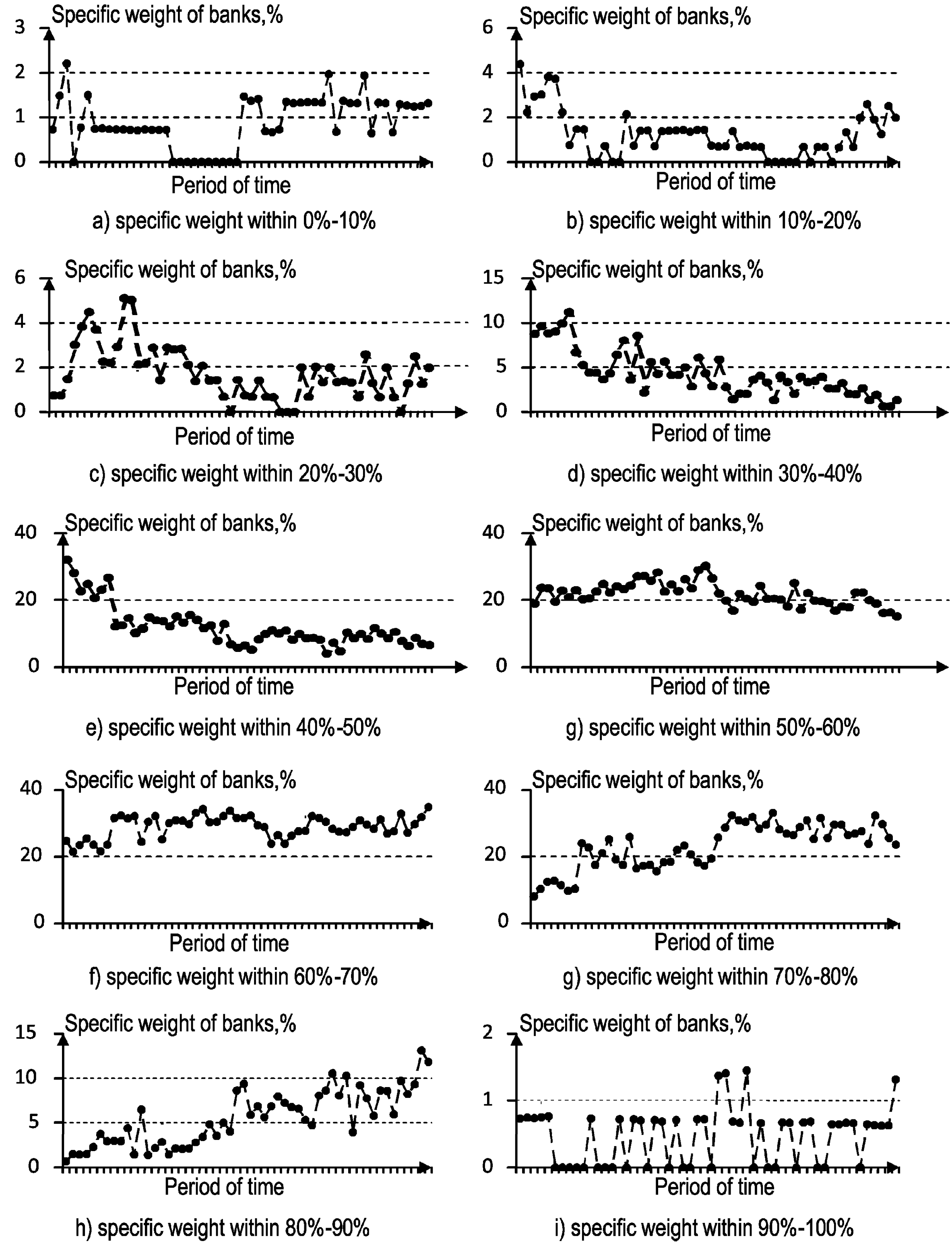


Figure 147. The specific weight dynamics of the granted credits to the total assets volume in the banking system as a whole during the period from 01.01.2004 till 01.05.2008

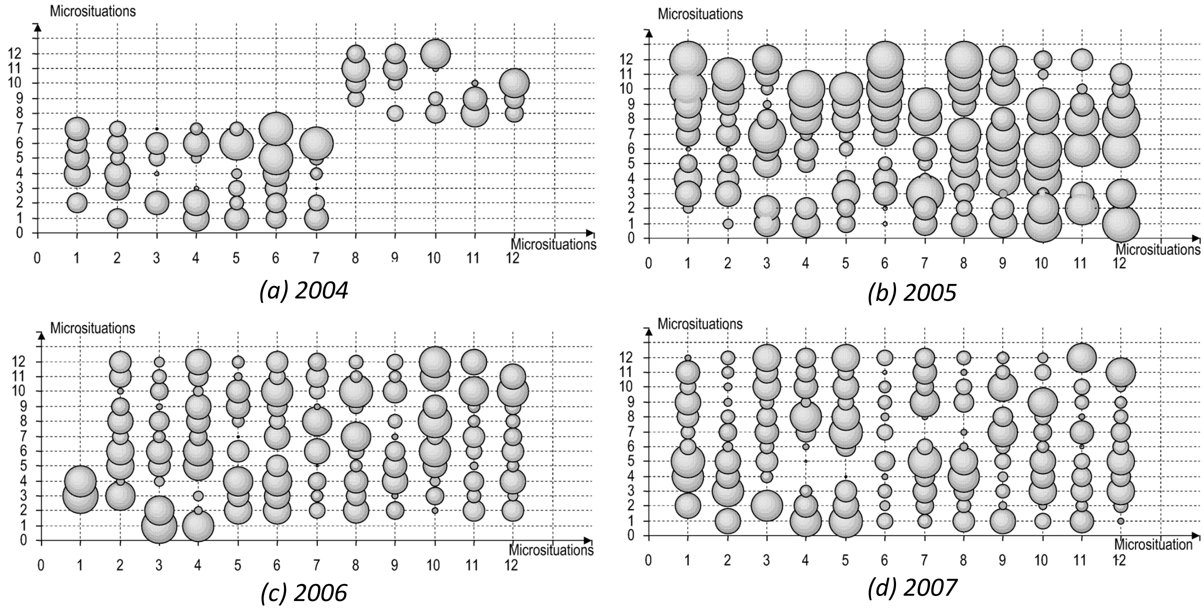


Figure 148. Consistency of microsituations representing variation of the credits specific weight in the banks assets volume according to the results of the banking system work

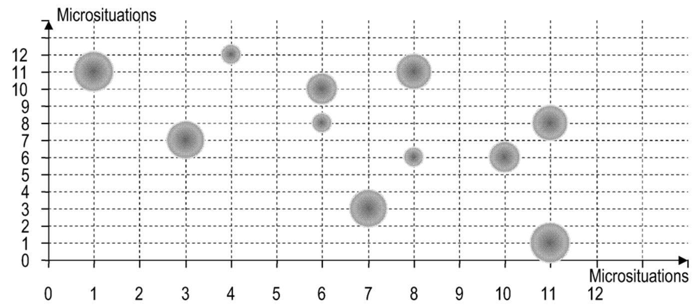


Figure 149. Consistency of microsituations representing variation of the credits specific weight in the banks assets volume according to separate banks of the group being studied by the results of their work within the period from 01.01.2004 till 01.05.2008

As evident from the data in Figure 149 the microsituations consistency in the considered aspect is not observed for the group of banks under study. Thus it may be concluded that each of the banks chooses its own strategy of increase of the credits being granted. Nevertheless, according to the data in Figure 148 such a strategy as a whole is aimed at increasing the credits specific weight in the bank assets structure. Consequently, the problem of the credit risk rise remains an urgent one.

If we speak directly on visualization of results of comparative analysis, we can use representation of such data in some kind of curve, which was described in previous parts of this paper. One suitable approach is taking into account distance between different microsituations according to Figure 148, where such distance is taken using different metric with coordination measure (in those examples we have a size of each circle describing degree of coordination of microsituations). In formalized equation, this could be described as follows:

$$R = M\left(S_n^m, S_{n+1}^{m+1}\right), \qquad (14)$$

Where R – is coordination distance between microsituations using some metrics M;

S_n^m – size of the circle, describing microsituation **nm** in previous analysis of microsituations set;

S_{n+1}^{m+1} – size of the circle, describing microsituation **nm** in next analysis of microsituations set.

Resulted curve using equation 9 will describe dynamics of coordination change of current bank and could serve as its dynamics profile. In the same time for such dynamics review, it is possible to use methods of nonlinear dynamics, which receive wide appliance for performing comparative analysis in economical researches.

8.3.3 Methods of nonlinear dynamics are in the estimation of development of banks

Methods of nonlinear dynamics are widely used in analysis and forecasting of parameters showing the development of stock exchange market, insurance market, and dynamics of investment handling. Simultaneously analyses of bank segments of finance market based on methods of nonlinear dynamics are not sufficient explored in scientific publications. One of boundaries of such approach to such type of markets is the necessary amount of sample data collected, which may characterize the development of bank sector. Even for such markets, the investigation of discontinuities of economic processes is quite important for taking into account existing dynamics and the possibility of weakening regarding to further development of banks.

Phase portrait of statistical data series is the key term of nonlinear dynamics, characterizing main parameters of bank's processing and their time-induced changes. Such series are e.g.

KI – data series, defining dynamics of bank's loan-investment portfolio;

KR – data series defining dynamics of loans handed over;

MK – data series, defining bank's activity on the markets of interbank loans;

ZP – data series, characterizing dynamics of amount of bank's securities;

D – data series, defining the general amount of resources, acquired as deposits;

DF – data series, generalizing amount of resources acquired as deposits of physical persons;

DY – time series generalizing amount of resources acquired as deposits of legal persons;

In this way, bank's activity may be described as an amount of data series mentioned above, which can be generalized as follows:

Data series, defining dynamics of bank's loan-investment portfolio as:

$$\mathrm{KI}(x_1, x_2, \ldots, x_t) = \mathrm{KR}(y_1, y_2, \ldots, y_t) + \mathrm{MK}(z_1, z_2, \ldots, z_t) + \mathrm{ZP}(d_1, d_2, \ldots, d_t), \quad (15)$$

and data series, defining the overall amount of resources, acquired as deposits:

$$\mathrm{D}(e_1, e_2, \ldots, e_t) = \mathrm{DF}(ef_1, ef_2, \ldots, ef_t) + \mathrm{DY}(ey_1, ey_2, \ldots, ey_t) \quad (16)$$

where $x_t, y_t, z_t, d_t, e_t, ef_t, ey_t$ – values of according series at a fixed time moment t.

Then, in a phase space of dimension 2 using Cartesian coordinates the phase portrait of statistical data series may be defined as a set of points:

$$\Phi(\mathrm{CHR}) = \{r_i, r_{i+1}\}, i = \overline{1, t-1}, \quad (17)$$

where CHR – one of series shown above according to equ. 15 and 16.

r_i, r_{i+1} – are the values of series shown, in defined time intervals.

According to the fundamentals of rating of bank’s development with methods of nonlinear dynamics in Figure 150 are shown phase portraits of data series, reflecting dynamics of interbank loans, taking into account the specifics of activities of different Ukrainian banks (values are taken from www.finance.ua).

As seen from Figure 150 generally for banks are characteristic different phase portraits of investigated data series. Simultaneously you may see that the dynamics of phase portraits of "Bazis" and "Grant" banks are most correlated compared with the dynamics of phase portraits of investigated series for "Big Energy" and "Nadra". This fact may be first explained by existing bank’s strategy to act on market of interbank loans.

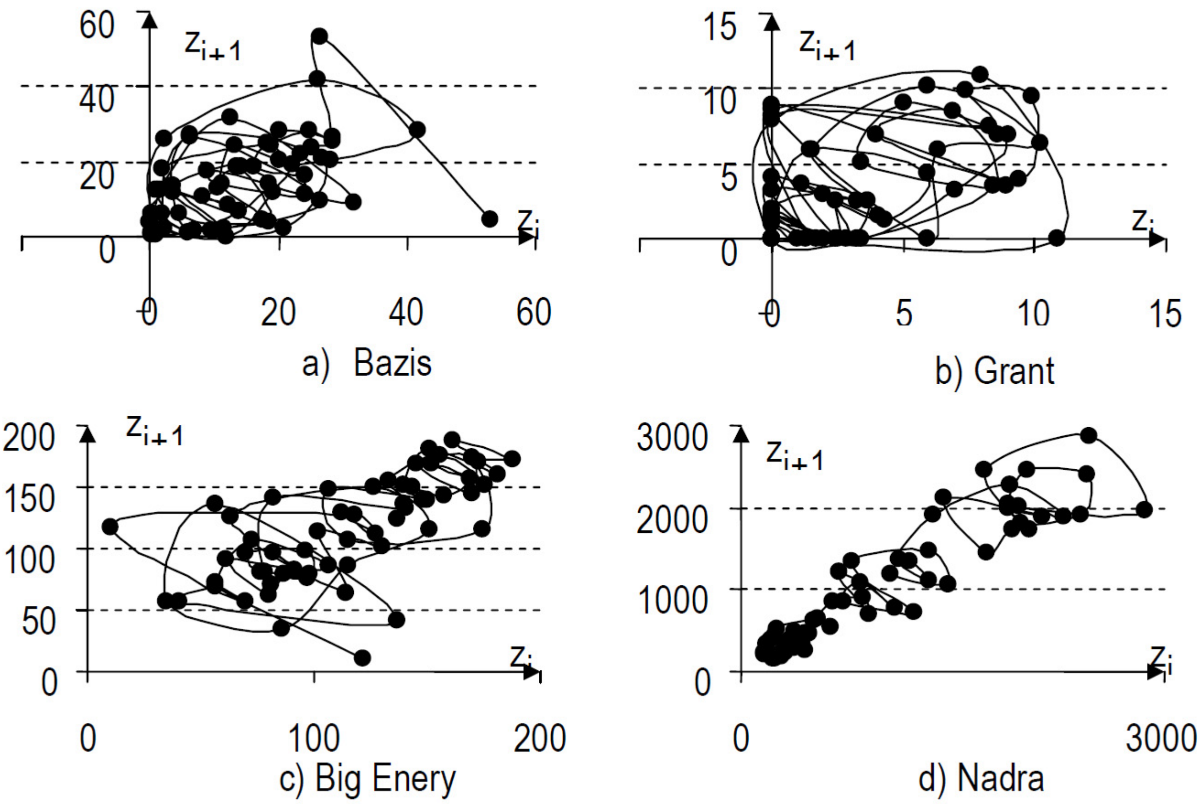


Figure 150. Phase portraits of data series defining bank's activities on the market of interbank loans for 2004-2008 (monthly)

Such strategy, however, will be defined based on existing conditions and factors, influencing bank’s activities. This fact is reflected in the phase portraits of investigated data series shown above. So the "Bazis" bank and the "Grant" bank are related to the same group of banks, which are additional administrated by intermediate management. "Big energy" and "Nadra" banks also are administrated by intermediate management (www.bank.gov.ua). Therefore, one may state, that phase portraits of data series, shown above, reflect existing conditions of functioning, belonging to different banks. With other words, methods of nonlinear dynamics may be used on equal rights for investigation and analysis of development processes of complex economic systems, banks belonging to.

In this chapter, a set of most perspective methods, developed during last 4 year by authors for comparative analysis of banks functioning were reviewed. As example of such methods should be mentioned:

- formalization of banks' management effectiveness using connection between liquidity and profit-generation potential based on fuzzy sets theory;
- integral structured representation of reviewed processes spatiotemporal dynamics;
- methods of performing comparative analysis of functioning and development both whole banking industry and one bank based on generalized conception of microsituations set, each describing such activity as a result of determined financial flows, which, in one's turn reflects that o another factors of banks activity. For microsituations comparison non-parametric tests based on Wilcoxon criteria are used;
- performing comparative analysis of banks function using non-linear dynamics, including phase portrait. It is shown a set of cyclic changes in temporal sets of data, which allows to describe a certain characteristics of banks development.

Adequacy and effectiveness of proposed approaches is tested on real data in different aspects of banks activities, which renders it usable for performing extended comparative analysis of separate banks and bank systems.

9
Methods and Means for Protection of Software Critical Infrastructures

9.1 Definitions

9.1.1 Infrastructures

According to "Etymology online" [Dictionary, 2010] a definition for the global infrastructure exists since 1927 and means: "The installations that form the basis for any operation or system". Originally, the term was used by the military.

Usually the term "infrastructure" refers to the technical structures that serve citizens. These are roads, water supply, sewage system, electrical network management systems in a flood, communications (internet, telephone lines, radio, and television), etc. In the past, these systems were usually owned and managed by local and central government authorities. These elements can be summarized as civil infrastructure, municipal infrastructure, or simply referred to as public facilities, although they can operate and develop in both the private sector and in state enterprises.

9.1.2 Critical Infrastructure

Infrastructure may refer to information technologies, formal and informal channels of communication, tools for creating software, political and social networks, or beliefs of certain groups of population.

The term "Critical Infrastructure" is used for those elements of infrastructure that if damaged severely or destroyed, would cause serious disruption of the depended system or organization. Storm, flood, or earthquake leads to the loss of key transport routes in the cities (e.g. bridges). This can prevent people to evacuate or rescue teams to do their work; these roads can be referred to Critical Infrastructure. Similarly, online registration systems can be Critical Infrastructure for airlines.

The notion of infrastructure lies not only on the public sphere and its facilities, but also on the working methods, management practices and policy developments in the direction of interaction between them all on the one hand, on the other – taking into account with the social needs and assuagement of public transport communications (for people and goods), providing drinking water and industrial water, safe waste disposal, energy supply and the need for dissemination of information among the population [21st Century, 1987].

IT infrastructure is an integral part of Critical Infrastructures and to achieve an effective protection of these Critical Infrastructures should primarily ensure the protection of Critical Information Infrastructures from malicious acts caused intentionally or unintentionally.

9.1.3 Software Critical Infrastructure

Once explained what we mean by Critical Information Infrastructure for the purposes of this part of the book, we can define the following working definition of Software Critical Infrastructure. As Software Critical Infrastructure will note a set of specific software solutions designed to operate and manage the relevant Critical Information Infrastructures.

In most cases, when talking about infrastructure security is understood security of Critical Infrastructures such as airports, railway highways, hospitals, bridges, network communications, media, power grid, dams, nuclear reactors, seaports, oil refineries and water systems.

Below when talking about the protection of Software Critical Infrastructure will mind the following points:

- protection during the design of the software solution;
- protection during the build of the software solution;
- protection during the test of the software solution;
- protection during the execution of the software solution;

To protect critical software infrastructure must be taken into account following:

- to analyze and identify threats;
- to reduce vulnerabilities in Software Critical Infrastructure;
- to reduce to a minimum the amount of possible injuries;
- to minimize to a minimum the time for reaction while trying to attack, and the time of response reaction and recovery of the caused damages;
- to analyze the causes of damages and the attack source (human action, incorrectly written source code, etc.);

9.2 The current situation

Nowadays information society requires the use of various types of information flows. These information flows are usually in the form of file objects. On various types of file objects, the following main operations are applied: creating, processing, transferring / accepting and saving / storage [Dale and Lewis, 2009], [Cohen and Kalbaugh, 2008] (Figure 151).

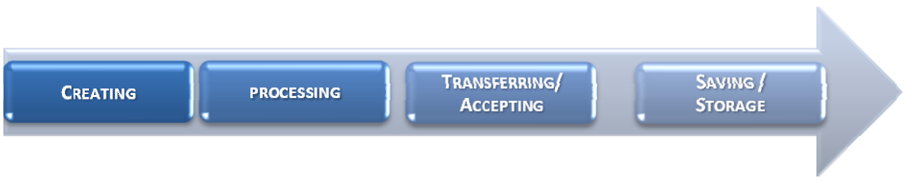


Figure 151. Main operations applied on file objects

In the process of creating of the object are generated its main parameters: filename, file format, header information, etc. This process is necessary to enable the opportunity of the object to be processed later, which includes some of the following operations related to the information included in the file: read, edit, copy, move, delete. Operation transferring / accepting of the file object is related to the possibility of the object to be sent to another device that is most often done by placing it on appropriate media or through the Internet, Extranet, or Intranet. Upon accepting of the object, it can be stored on local hard drive then its further processing could be possible.

During the investigation of the information security of objects, we should take into consideration they could be exposed to different information attacks. The information attacks can be provisionally divided into malware and malattacks (Figure 152).

In case of **malware** the direct participation of a user at the moment of the attack is missing, while in case of **malattack** the user's presence is required [Shaw, 2006], [Radhamani and Rao, 2007].

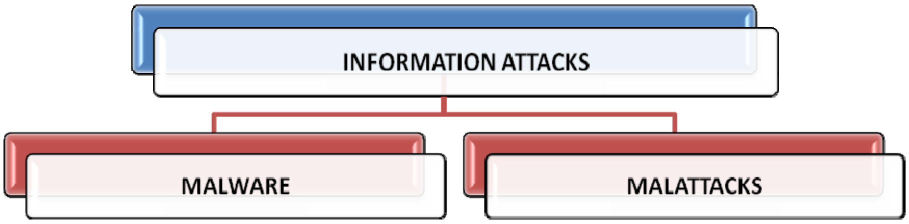


Figure 152. Main categories information attacks

The objects used in Software Critical Infrastructures exposed to attacks, may be possible in following states:

1) in cases where they are in an environment with method of protection (Figure 153);

2) in cases where they are in an environment without method of protection (Figure 154).

Figure 153. An object in environment without method of protection

Figure 154. An object in environment with method of protection

When methods of protection of object from information attacks are investigated, they can be divided into three main categories: methods of protection depending on the environment, methods of protection depending on the architecture and method of protection depending on CPU characteristics (Figure 155).

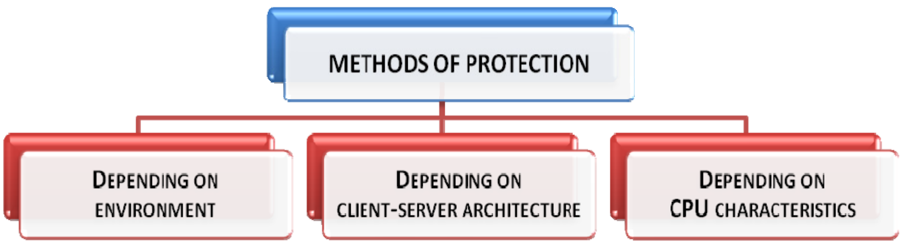


Figure 155. Main categories methods of protection

9.2.1 Information attacks groups

A description of the groups of attacks, referred to the categories "malware" and "malattack" is given below.

In category **malware**:

(I). Browser related attacks [Dingledine and Golle, 2009], [Camenisch, 2009], [Dubrawsky, 2009]. These attacks are characterized by the fact that they use the resources and capabilities of different browsers in order to fulfill its malicious content. The consequences of such an attack can be insignificant (such as crashing of the browser), but they can lead to theft of identity or other confidential information.

(II). Metadata [Clarke, 2009]. The attacks from this group use the possibility to add metadata to file object to achieve their malware goal.

(III). Cracking [Sarknas, 2006]. The attacks from this group succeed to neutralize procedures of authentication, check sums, registration, etc.

(IV). Spying [Ciampa, 2008], [Ciampa, 2009]. The attacks from this group use means for illegitimate gathering information. Absolutely everything in the attacked system is subject to espionage.

(V). DoS, DDoS [Docter et al, 2009]. The goal of the attacks from this group is to make computer resources unavailable to its intended users.

(VI). Exploits [Rabe, 2009]. They use a security hole. It is used as the base to create programs that search vulnerable systems and those who enter the already found vulnerable systems;

(VII). Scanners [Erjavec, 2009]. The attacks from this group use different tools and techniques to scan the system. In most cases, they precede other malicious components.

(VIII). Keyboard modifiers [Newman, 2009], [Dubrawsky, 2009].The attacks from this group use techniques to spy and manage the keyboard, mouse and screen activity.

(IX). Computer Trojan Horses [Kartalopoulos, 2009]. It transports program components that are later used for some malignant goal. The term is often used to describe the malignant software that is transported.

(X). Computer Backdoors [Newman, 2009]. They create security hole as supporting a communication port in opened state.

(XI). Computer Worms [Ao and Gelman, 2009]. An independent program that spreads data from one computer to the next using the network connections and frequently takes advantage of the weakness of the main Internet protocols in order to cause problems in computer systems and networks. It is usually combined with spy and advertisement components.

(XII). Computer Viruses [Esl, 2009]. A parasite program that without the user's knowledge or permission attaches to the infected object procreates and continues to spread. It usually has a malignant component aimed at harming the functionality of single programs or the whole computer system. This term includes currently 50+ functional types.

In category **malattack**:

(XIII). Using accessible information [Blyth, 2006]. The attack possibilities are based on a previous illegitimate acquisition of logs, registers, documents, etc., that reveal the functioning and security of a given system. The specific information needed for the next attack can be found by studying the volume and the internal structure of the incoming and outgoing traffic in a computer system or network.

(XIV). Overflow. Hackers cause an overflow of certain buffers in order to have a system or network dysfunction or to receive or execute malignant code.

(XV). Vulnerabilities [FTC, 2010]. Usually the actions of the attackers are based on a serious security breach in the system. The administration of a certain application or group of applications is taken over.

(XVI). Content [Osborne, 2006]. A large variety of techniques based on destructive office macros, executable trojans and ActiveX, and Java applications.

(XVII). Data Encapsulation [Wrembel and Koncilia, 2007]. Data is hidden or capsulated in such a way that it overcomes the firewall and is then free to unite or fulfill its malignant mission.

(XVIII). Denial of Service [Sisalem et al, 2009]. The main goal of these attacks is to receive denial of service due to planned and coordinated service requests. The important in this case is the large number of hosts that make the requests, which causes exponential increase in the number of requests.

(XIX). Spoofing [Haldar and Aravind, 2009]. To attacks by this group can be summarized that their aim is to successfully masquerade by falsifying data and thereby gaining an illegitimate advantage.

(XX). CrackPasswd [Miller and Gregory, 2009], [Chandra et al, 2009]. This attack is relatively rare due to the current password-based site protection system. Some serious preparation is needed to access the passwords. Should that happen the attack could last relatively long because people are used to being safe.

(XXI). Zombie Computers [Wang, 2009]. Completely controlled by a hacker's computer. In this case, the attacker can create for example two super-zombies that on their own control, 8 other zombies and so on. In this manner, a large number of computers simultaneously attack a computer until they achieve denial of service.

Note: Roman numbers in brackets are used later as identifiers of the names of the attacks groups.

9.2.2 Objects groups

The groups of the objects can be referred to the categories "directly executable" and "indirectly executable". Directly executable objects can be directly usable, while indirectly executable objects require secondary processing to become directly usable.

Table 9 shows the groups of directly executable and indirectly executable objects that are most often subject to information attacks in Software Critical Infrastructure.

Table 9. Groups of objects that are most often subject to information attacks

Object group	Total [%]	Malware [%]	Malattack [%]
DIRECTLY EXECUTABLE			
Archived and Compressed	11,11	9,54	1,57
Scientific	10,86	9,22	1,64
Data	17,49	14,58	2,91
Internet related	16,00	13,50	2,50
Binary	17,53	14,81	2,72
Virtual machine	9,34	7,70	1,64
Other	5,94	4,88	1,06
TOTAL DIRECLTY EXECUTABLE:	**88,27**	**74,23**	**14,04**
INDIRECTLY EXECUTABLE			
Graphic	6,17	5,15	1,02
Audio and music	2,31	1,94	0,37
Video	3,25	2,72	0,53
TOTAL INDIRECTLY EXECUTABLE:	**11,73**	**9,81**	**1,92**
ALL OBJECTS	**100,00**	**84,04**	**15,96**

9.2.3 Protection groups

Figure 156 shows the protection's groups used to protect file objects from information attacks in Software Critical Infrastructure.

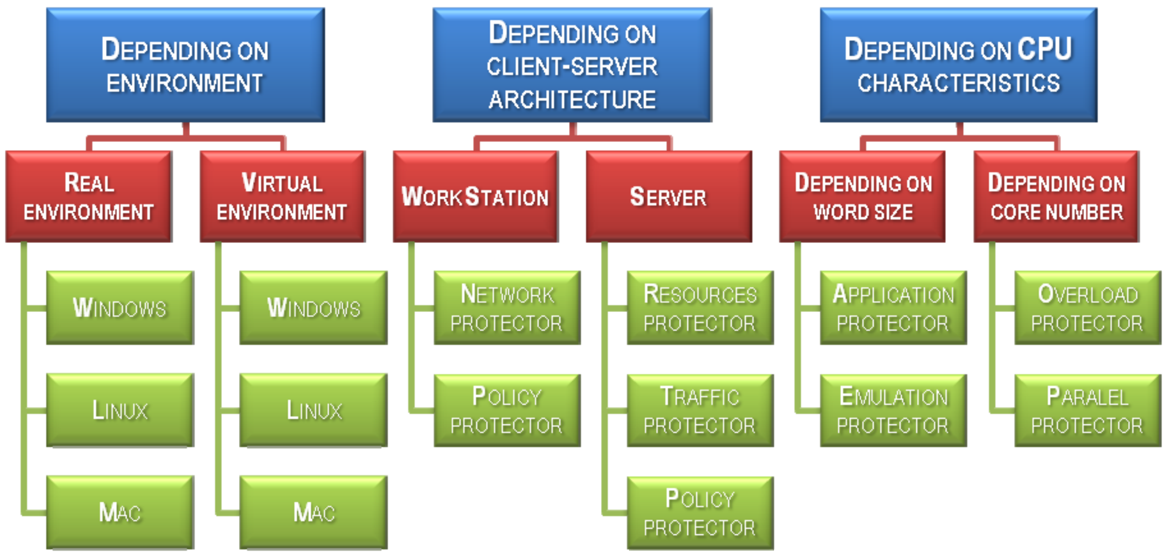


Figure 156. Main protection's groups

I. Protection types depending on environment:

1) in case of real environment. It is protected by:

- real-time memory protectors;
- real-time registry protectors;
- real-time media (disk, CD/DVD, flash, etc.) protectors;

The most commonly used operating systems in a real environment and their protection types are:

1.1) Windows real environment. The protectors include:

- real-time Windows memory protectors;
- real-time Windows registry protectors;
- real-time Windows media (disk, CD/DVD, flash, etc.) protectors;

1.2) Linux real environment. The protectors include:

- real-time Linux memory protectors;
- real-time Linux registry protectors;
- real-time Linux media (disk, CD/DVD, flash, etc.) protectors;

1.3) Mac real environment. The protectors include:

- real-time Mac memory protectors;
- real-time Mac registry protectors;
- real-time Mac media (disk, CD/DVD, flash, etc.) protectors;

2) in case of virtual environment. It is protected by:

- virtual memory protectors;
- virtual registry protectors;
- virtual media (disk, CD/DVD, flash, etc.) protectors;

The most commonly used operating systems in a virtual environment and their protection types are:

2.1) Windows virtual environment. The protectors include:

- virtual Windows memory protectors;
- virtual Windows registry protectors;
- virtual Windows media (disk, CD/DVD, flash, etc.) protectors;

2.2) Linux virtual environment. The protectors include:

- virtual Linux memory protectors;
- virtual Linux registry protectors;
- virtual Linux media (disk, CD/DVD, flash, etc.) protectors;

2.3) Mac virtual environment. The protectors include:

- virtual Mac memory protectors;
- virtual Mac registry protectors;
- virtual Mac media (disk, CD/DVD, flash, etc.) protectors;

II. Protection types depending on client—server architecture:

1) on behalf of Workstation the following protections can be referred:

1.1) Network protector. These type protections analyze the TCP/IP packets and recognize malware content;

1.2) Policy protector. A user profile, threat assessment and security policy choice is made;

2) on behalf of Server the following protections can be referred:

2.1) Resources protector. They protect single shared resources (memory, disk, periphery);

2.2) Traffic protector. Protection aims to prevent unusual, abnormal traffic;

2.3) Policy protector. A Work Station's profile is created, a malware content is recognized and correct security policy is chosen for each Work Station.

III. Protection types depending on CPU characteristics:

1) depending on word size the following protectors can be observed:

1.1) Application protector. They must be conformable to the application's word size as taking into account the relationship between the operating system and the word size;

1.2) Emulator protectors. These protections include defense against unwanted or incorrectly selected emulations of processes, applications, and resources.

2) depending on core's number the following protectors can be observed:

2.1) Overload protector. They include protection against overloading of dataflow with respect to one or more cores;

2.2) Parallel protectors. They include protection against malicious or accidental breach of parallel processing of the information by the single cores.

9.3 Available information for accomplished attacks

An analysis will be presented here, based on the current information base of National Laboratory of Computer Virology of Bulgarian Academic of Sciences. It collects information for the information attacks, which were carried out to a separate personal and/or corporate computers, and/or networks, and/or systems for 2009. This is a generalization of attacks, implemented in Bulgaria, Balkan Peninsula, and southeast Europe. Figure 157 shows the percentage distribution of accomplished attacks' groups. Summary, 84% from them belong to the **malware** category, and 16% belong to the **malattack** category.

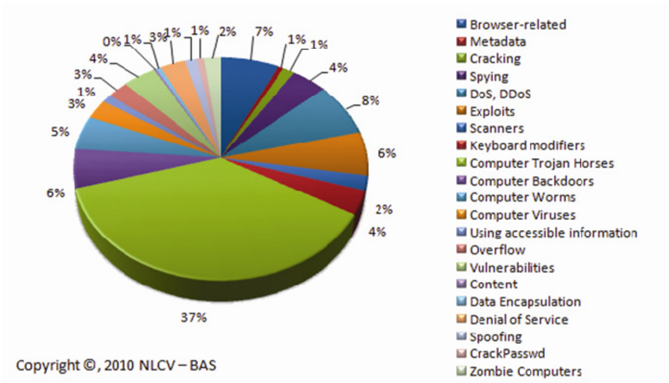


Figure 157. Percentage distribution of accomplished attacks' groups

Compared with the other structures (government, corporate and users), the number of accomplished attacks to Critical Infrastructures as a hole is about 23% (Figure 158).

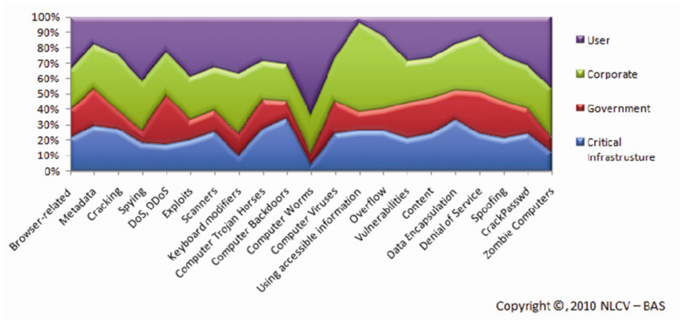


Figure 158. Percentage distribution of attacks' groups, accomplished to computers, systems and networks with respect to the separate structures

Figure 159 shows a comparison of the percentage distribution of single category's attacks with respect to the single structures with overall percentage distribution of accomplished attacks to separate structures. It is evident from the figure that the Critical Infrastructures are at the place before the last with respect to the accomplished attacks. In them, however, occurred almost equally effect of the two attacks' categories.

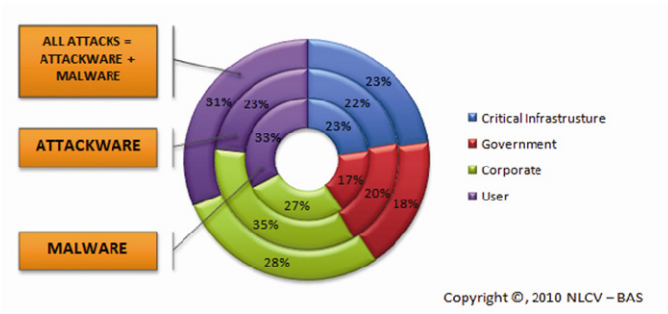


Figure 159. Comparison of the percentage distribution of single category's attacks with respect to the single structures with overall percentage distribution of accomplished attacks to separate structures

Table 10 shows the distribution of accomplished attacks by structures for each attack's group.

Table 10. Percentage distribution of accomplished attacks by structures for each attack's group.

Attack group	Total [%]	Government [%]	Corporate [%]	User [%]	Critical Infrastructure [%]
MALWARE					
Browser-related	7,06	1,48	1,27	1,84	2,47
Metadata	0,70	0,20	0,17	0,20	0,13
Cracking	1,41	0,38	0,17	0,51	0,35

Attack group	Total [%]	Government [%]	Corporate [%]	User [%]	Critical Infrastructure [%]
Spying	4,24	0,76	0,34	1,36	1,78
DoS, DDoS	7,77	1,32	2,49	2,17	1,79
Exploits	6,36	1,27	0,83	1,78	2,48
Scanners	2,12	0,53	0,30	0,59	0,70
Keyboard modifiers	3,53	0,35	0,49	1,38	1,31
Computer Trojan Horses	37,43	10,11	7,11	9,36	10,85
Computer Backdoors	5,65	1,92	0,62	1,36	1,75
Computer Worms	4,94	0,20	0,35	1,24	3,15
Computer Viruses	2,82	0,68	0,59	0,79	0,76
TOTAL MALWARE:	**84,03**	**19,21**	**14,72**	**22,57**	**27,53**
MALATTACKS					
Using accessible information	1,13	0,28	0,14	0,66	0,05
Overflow	2,57	0,67	0,36	1,21	0,33
Vulnerabilities	3,99	0,84	0,92	1,07	1,16
Content	0,29	0,07	0,07	0,07	0,08
Data Encapsulation	0,57	0,19	0,11	0,17	0,10
Denial of Service	3,14	0,75	0,85	1,13	0,41
Spoofing	1,42	0,30	0,34	0,41	0,37
CrackPasswd	0,86	0,21	0,14	0,24	0,27
Zombie Computers	2,00	0,24	0,18	0,64	0,94
TOTAL MALATTACKS:	**15,97**	**3,56**	**3,09**	**5,61**	**3,71**
ALL ATTACKS	**100,00**	**23,00**	**18,00**	**28,00**	**31,00**

Examined Critical Infrastructures include: Energy and Utilities; Communication and IT; Banking and Finance; Health care; Food and Agriculture; Water; Transportation; Safety; Government; Manufacturing.

With respect to individual objects' groups used in Software Critical Infrastructures, we can say that most attacks are carried out to Data, Internet related and Binary file objects (Figure 160, Figure 161).

Figure 162 shows the percentage distribution of accomplished attacks' groups which were carried out to all Critical Infrastructures.

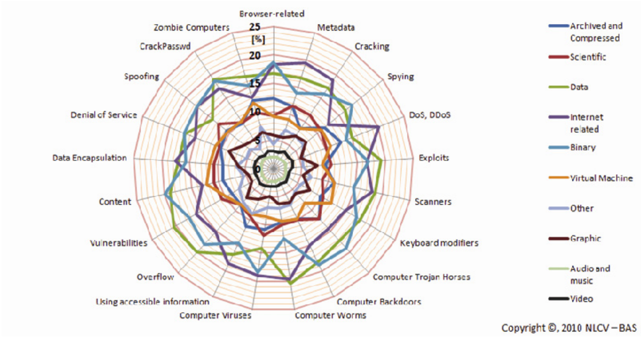


Figure 160. Percentage distribution of accomplished attacks groups to the Software Critical Infrastructures with respect to the objects' groups

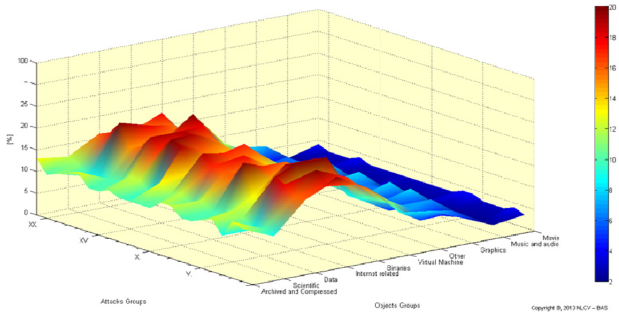


Figure 161. Percentage distribution of accomplished attacks groups to the Critical Infrastructures with respect to the objects' groups

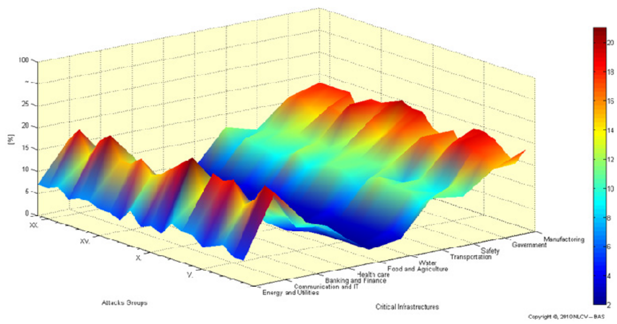


Figure 162. Percentage distribution of accomplished attacks groups to the Critical Infrastructures

9.4 Methods of prevention, protection and recovery

9.4.1 Methods of prevention

- ✓ **Human factor-related:**

– against disloyal employees or consultants;

- for loyal employees or consultants;
- against disloyal users;
- for loyal users;
- against disloyal system administrators;
- for loyal system administrators.

✓ **Media-related:**

- against theft;
- against copy/pasting;
- against physical harm/purposeful destruction;
- against copy/pasting (service maintenance).

✓ **Malware-related:**

- against viruses and the like;
- against future unauthorized access.

✓ **Related to electromagnetic emissions:**

- against screen information hijacking;
- against cable information hijacking.

✓ **Notebook related:**

- against unauthorized use;
- against theft.

✓ **Related to network topology:**

- against Internet effects;
- against network component manipulation.

✓ **Related to DoS attacks:**

Prevention of this type of attack is extremely difficult because it targets closed ports, can take different shapes, aims at many services and devices, and could be caused by legitimate packages if they cause a recursive effect such as opening and closing multiple simultaneous connections. A combination of the following is recommended:

- during network projection;
- for network perimeter protection.

✓ **Password related:**

- Use of strong password;
- Increase the length of the password (at least 8 symbols);
- Use of capital letters and numbers;

- Do not use personal information, names, whole words, easy to guess strings or hacker terminology in passwords;
- Do not write down passwords;
- Strict regulations regarding the number of stored used passwords (at least 10);
- Establishment of the minimum and maximum validity period of each password (0 days at least, 30 days at most);
- Use of password filtering programs such as passfilt.dll from Windows 2000 Resource Kit;
- Administrator testing with special password cracking tools;
- Complete ban on password sharing;
- Ban on password remembering of the program in question.

✓ **Virus related:**

- all of the software used should be bought from an authorized dealer;
- use of illegal software is unacceptable;
- all removable media (esp. those of unknown origin) should be scanned with an updated antivirus program;
- use of free publicly available software is to be avoided;
- floppy disks should be write-protected and removed from the computer when possible;
- if a non-virus protected Internet e-mail is used it should be done on an offline computer with updated antivirus.

✓ **Spam related:**

Spam blocking can be done on the mail server or the enduser's computer based on two models:
- black lists;
- contents filters.

✓ **Security policy related:**

According to SANS the security policy is viewed as a modular method for the development of specific targets or rules that are used to control the different aspects of a computer system such as password creation or real time application use. In brief, this document puts in writing how a company plans to defend its physical and informational assets.

9.4.2 Methods of protection

✓ **Basic functions**

The problem with absolute malignant software protection is practically impossible to solve but in order to protect a computer system we have to guarantee:

1) Intrusion protection
2) Resource inaccessibility
3) Inaccessibility of information in the working terminals

4) Cryptographic protection

- ✓ **Basic tasks**

The security protection of a company should have the following abilities.

1) Identify the attack.
2) Act on the attack.
3) Discover the attacks.
4) Reflect the attack.
5) Transparent work.

- ✓ **Basic elements**

The methods for computer system protection include the following basic elements:

1) Threat analysis.
2) Access control.
3) Authentication.
4) Confidentiality.
5) Data integrity.
6) Denial inability.
7) System reliability.

- ✓ **Basic principles**

The main principles for protection planning are:

1) Smallest privileges.
2) Deep protection.
3) Protection diversity.
4) Default denial.
5) Security by ambiguity.

- ✓ **Basic methods**

1) Data and file encoding
2) Methods for server antimalware protection
3) Methods for workstation antimalware protection
4) Methods for hoaxes protection
5) Methods for DDoS attack protection
6) Methods for protection against e-mail and spam bombings and IP spoofing
7) Protection against Sniffing
8) Protection against kernel level Rootkits
9) Protection against Smurf or Fraggle attacks

10) Protection against SYN Flood attacks
11) Protection against DNS attacks
12) Methods for wireless network protection
13) Methods for VPN protection
14) Methods for browser protection
15) Methods for mobile device protection (laptops, table computers, PDA etc.)

9.4.3 Methods of recovery

- ✓ **Common preliminary measures**

1) Strategy and recovery plan

The strategy and the plans should have clear and precise procedures for decision making regarding the people that should be informed about the incident the selection criteria for various recuperation procedures, supervisors, and responsibilities in order to prevent lengthy interruptions of regular activities.

The plan should include specific procedures for the different kinds of computer and network hardware, operating systems and application software.

The criteria and conditions on informing the authorities should also be included.

The final plan should be review by competent IT workers that did not participate in its development.

2) Concrete actions

- Creation of floppies and CDs for initial setup.
- A mail server for web-based back up should be installed outside the premises of the company for communication in case of a computer incident.
- Creation of backup e-mail accounts with global POP3 mail service sites.
- Installation of a backup Internet connection away from the organization's premises.
- Creation of an alternate site as an internal site (supported in a different location that has no connection to the mainframe), or as a Internet stored backup site.
- Frequent and complete backups.
- Backup storage in a secure place outside the computer system room as well as with Internet information storage sites.
- Backup storage on FireWire devices.
- Agreements to use office space at a different location in case of need.
- Gearing all IT employees and key workers with backup laptops in order to synchronize the main computer system restoration and not to interrupt the most important business activities.
- Backup system components ready for use in case of need.
- Standardization of the used hardware, software, and peripherals.
- The critical hardware components should be retail to avoid the delivery delay that comes with custom made equipment.

– Full documentation of system configurations and suppliers contact information for a quick switch in case of need.

✓ **Preliminary measures for big computer systems**

1) Constant updating off all critical databases.
2) Use of all fault-tolerant computer and network systems.
3) Use of extra components for critical systems.

✓ **Preliminary measures for computer networks**

1) Establishment of backup communication connections.
2) Use of several service providers.
3) Doubling of the devices creating the network connections.
4) Establishment of segmented networks.
5) Use of standard network technology.
6) Use of effective network security systems.
7) Use of all appropriate devices of network protections and prevention.

✓ **Computer incident reaction methods**

The various computer incident reaction methods include several different procedures. They all go through the following main stages.

1) Identification.
2) Qualification.
3) Limitation.
4) Source localization.
5) Target localization.
6) Deactivation.

✓ **Computer system recovery methods**

1) Coordination.
2) Review.
3) Evaluation.
4) Notification.
5) Description.
6) Shut down.
7) Removal.
8) Condition.
9) Recovery.
10) Return.

11) Check.

12) Guarantees.

✓ Recapitulation after recovery from the attack

Includes the answers to the following questions:

– How and why was there an attack?
– How were the consequences removed?
– Lessons that should be learnt for future reference.
– Having strictly determined goals that have quantified targets, deadlines and supervisors.
– Full documentation of the attack and the recuperation.
– Incorporating the necessary changes in the security policies and recuperation procedures to prevent future problems.

✓ Recovery after DoS attacks

DoS attacks require instant filtration measures depending on the nature of the attack. The traditional DoS attack usually does not cause host corruption, which makes recuperation extremely easy. If the client's IP address or the scheme of the attack can be quickly identified the traffic filtration in the router is easy to organize. The sophisticated DoS attacks as well as those committed by the Trojans Code Red and NIMDA, changed that perspective by damaging webhost platforms and generating different scanning and exploit schemes. Usually filtration is a temporary solution and it should be followed by the following to restore network services.

1) Access Control List (to limit malicious traffic).

2) Shut down of all unnecessary services.

3) Software update.

4) Fine tuning of Internet applications.

5) Eavesdropping services host access limitation through ACL.

In many cases the reaction and recuperation after a DoS attack requires intensive use of resources and bandwidth as well as close cooperation with the IPS to create a router filtering mechanism.

✓ Post-virus attack recuperation

The successful elimination of a certain virus from a computer system includes temporary restoration on removal of all infected or deleted files. The most effective way to restore damaged objects is to replace them with original copies. Thus, the frequent preparation of complete backup copies is very important and facilitates restoration.

After cleaning the virus from the computer system, it is necessary to scan all diskettes and other removable storage media to guarantee that they are virus-free. Should those measures not be performed a second infection of the computer system and the further distribution of the virus become a possibility.

9.5 Assessments

1. Sharp increase in the importance of security for Software Critical Infrastructures.

When establishing, Software Critical Infrastructure is required prior to making a roadmap for the movement of information flows between different components of the infrastructure. This can be described as follows:

1) Clarifying the number of the modules.
2) Clarifying the size of the modules.
3) Clarifying the number of the input, output, and input-output points for the information flows.
4) Reducing the number of the different points to one input-output point.
5) establishing a verifying mechanism that ensures the content of information flow during the input and output operation.
6) The verifying mechanism must be different for the input and output point.
7) Describing the duration as time for moving of information flow from point A to point B.
8) Adding additional time to perform supporting operations by certifying mechanism. A strict control during the movement of each information flow from point A to point B within the infrastructure is introduced by the roadmap.
9) Ensuring a relatively constant rate of movement of information flows through the overall management of the operating system.

2. Sharp increase in the price in establishing the Software Critical Infrastructures.

The increased requirements for how the infrastructure is created cause a significant increase in the price of its planning, the price of its creating and its using. This can be described as follows:

1) When planning, simulation experiments with different number of modules of the infrastructure, with different sizes of the modules and with different time and speed characteristics of the information flows, must be carried out.
2) Based on these experiments options in which to achieve the lowest price, average price and highest price, must be given.
3) Depending on the wishes of the client in the next phase of creating, an option is selected with respect to the three prices.
4) When creating the infrastructure is sought at an acceptable price in selected criterion for minimum, average, maximum price, while accomplishing an implementation of the prepared planning.
5) When using the infrastructure, an acceptable price is sought for the functioning of the infrastructure for certain periods such as microsecond, second, hour, day, month and year by managing the load of the infrastructure. The objective is to avoid peak values in energy consumption and repetitive processing of information flows.

3. Sharp increase in the complexity of establishing a Software Critical Infrastructures.

Increasing the number of the modules of the infrastructure, the number of the information flows and the long-term accumulation of recording information (logs) creates a condition for increasing the complexity of the infrastructure. This can be described as follows:

1) The number of the modules, their size, and the number of input-output points is directly related to the variants of running of the infrastructure whereby above a certain value of all

possible combinations mentioned above is reaching a level of complexity, which sharply aggravates the security of the Infrastructure.

2) The necessity of storing of recording information (logs) for the purpose of accomplishing of preventive and/or routine maintenance of the infrastructure, creates a critical threshold, which when is reached sharply aggravates the security of the infrastructure, because the ratio of useful and unnecessary information is changing catastrophically.

4. Sharp increase in vulnerabilities in the creation of Software Critical Infrastructures.

Security issues in the creation of the infrastructure have one side, but very important effect, namely, is ever increasing the vulnerabilities in the infrastructure. This can be described as follows:

1. The lack of sufficient time for completely full examination of possible combinations for connecting the modules, their sizes, the number of input-output points, the number of information flows, their speed and the number of the repetitions in the processing of the content of information flow leads to omissions in pre-planned functioning of the infrastructure. These omissions are the basis for the beginning of so-called infrastructure vulnerabilities.
2. The existence of free computational resources with a large number of processors and much more core numbers in these processors allows authors of attacking tools to receive serious advantage in the detection of vulnerabilities and their use for accomplishing attacks to the infrastructure.

9.6 Discussion

The problems of the Software Critical Infrastructure can be summarized as follows:

1. A change in the funding rules for establishing the infrastructure towards increasing the percentage of financing of test procedures and policies.
2. A change in the rules for scheduling for creating the Infrastructure in favor of increasing the percentage of the time for test procedures and policies in case of fully loading of the Infrastructure.
3. A change in the rules for individual modules, groups of modules and groups of software tools, composing the infrastructure in favor of adding a component to an existing configuration, only in case of fully one hundred percent completed testing procedure and policy.
4. A change in the rules for searching, detecting and correcting of vulnerabilities in favor of generous funding for each discovered vulnerability within the developer team, or forming a special team for discovering vulnerabilities with special funding for newly discovered single vulnerability.

In this connection in necessary to increase the funding, to create new methods for reducing the complexity of the Infrastructure and new methods for reducing the vulnerabilities to ensure the adequate security in creating of Software Critical Infrastructure.

10
The MLRP-method for Analysis of Some Problems in Climate and Seismology

10.1 Introduction

The problem of detection correlations and regularities by data, which is presented by time series or data table, is used in different scientific domains. We have so problem in sociology, medicine, geology and seismology, economics and climate, etc. [Box and Dgenkins, 1974]. The problem of detection correlations and regularities by data is the inverse and ill-defined problem; therefore, we have to consider every possible information about studied object or process in order to raise solution stability. This information first of all is system of characteristic of studied object: multidimensional and heterogeneity characteristic's space; "noisy" and "duplicate" characteristics; insufficient or excess and missing sample data.

We have the most difficult problem, when the priory information about process or object is absent. In addition to that several attendant problems are appeared. Therefore developing new method, we have to consider all above. Firstly, it is necessary to define a class of decision functions (models). Secondary, we must pick out a method of plotting optimal decision function, in other words to define optimality criterion by sample. In the third place, we must to test our model on adequacy and effectiveness (capacity for general conclusion or statistical stability).

It is very important to lean against practical and scientific experience of researchers when solving out practical problems. Experts in the certain area own the big share of the aprioristic information and have wide experience in interpretation of results of decisions of practical problems. It is very important in working out and testing of methods. Here it is important to underline an urgency of different countries scientists association at the decision of the important common to all humanity problems, for example, the risk estimation of occurrence the extreme situations due to man-caused and natural accidents. Unfortunately, availability of data occasionally is the exclusive property of the country or separate business. Therefore, association of scientists and experts of the different countries in work on the general projects is simply common to all humankind necessity.

Method presented here depends on statistical information about studied object and can be used in many domains of data mining also. That approach is relied to intelligent analysis to a greater extent, since it uses expert and apriory information generally.

At present time, there is joint initiative of the European Commission and European Space Agency, which aims at achieving an autonomous and operational Earth observation capacity. It is named Global Monitoring for Environment and Security (GMES) (http://www.gmes-bg.org/). The main objective of GMES is to monitor and better understand our environment and to contribute to the security of every citizen. Land, sea, and atmosphere – each Earth component is observed through

GMES, helping to make our lives safer. We think that our investigations are included to GMES-project properly.

There are many well-known scientific schools, what make researches to that line of investigation [Lukashin, 2003], [Bezruchko and Smirnov, 2003], [Lbov and Starceva, 1999]. However, universal method is not exists. Several suppositions and apriory information are used by every method. It says that problems are actually. Than method is preferred if it uses lame suppositions to respect with decision function class and has capability for change-over model of correlation during learning on sample data. Current methods use neural-network technological, pyramidal-network, wavelet analysis, logical structures and others approaches. Such methods we can name so as adaptive methods. A conception of adaptive has a more comprehensive sense [Lukashin, 2003], [Lbov and Starceva, 1999].

We will interpret concept of adaptive so as consecutive selection of model's structure in during process of learning on sample data in order to take effective prediction by time series. At the same time, it is appeared additional problem – detection time moment of changing model's structure (criterion of adaptive).

We suggest two ways to joint analysis of several univariate time series by using MLRP-method. When the information about event is kept safe in the process, and when it is kept safe in depending process. That method was applied to prediction of multivariable heterogeneous time series [Stupina and Lbov, 2006]. The solving of practice actual problems from hydrological and seismology domains are presented here by MLRP-method.

In a hydrological problem, the most informative signs influencing volume of a reservoir of water in the rivers is investigated. Probable catastrophic consequences from occurrence of high waters or a drought are obvious; therefore, the problem is actual for many regions of a considerable quantity of the countries on the Earth. The second problem is not less actual in global measurement. It consists in research of probable regional migration of active seismic zones. The problem of a short-term prediction of earthquakes remains important on present time. Therefore, the majority of problems of modern seismology are focused on studying of process of earthquake in order to come little nearer to the decision of problems of the forecast.

Model from the logical decision function class we will name as a logical-and-probabilistic correlation [Lbov and Starceva, 1999], [Stupina, 2005].

Let us note below priority properties of the logical decision function class at the solving of inverse problems:

- easily interpreted decision is constructed. It is especially important at the decision of difficultly formalizable problems, at finding-out of cause-and-effect interrelations in object, for the organization of a dialogue mode of forecasting etc.;
- the given class allows to process heterogeneous experimental data (without reduction all signs to one scale) and is invariant to admissible transformations of signs scales;
- it has a small measure of complexity. At the same time, this class is "rich enough" for the effective decision of the applied problems characterized by high level of aprioristic uncertainty about a functional kind of distribution. Mathematical properties of the given class are proved and brought in this chapter;
- gives the chance to realize simple optimizing procedures of searching the selective decision function;

– allows to work in the presence of admissions in empirical tables and to receive adequate enough decision at small volumes of sample.

10.2 MLRP-method of creating logical-and-probabilistic model

10.2.1 Basic notations

From the beginning, we consider a commonly probabilistic statement problem. Let the value (x,y) is a realization of a multidimensional random variable (X,Y) on a probability space $<\Omega, \mathrm{B}, \mathrm{P}>$, where $\Omega = D_X \times D_Y$ is μ-measurable set (by Lebesgue), B is the borel σ-algebra of subsets of Ω, P is the probability measure (we will define such as c, the strategy of nature) on B, D_X is heterogeneous domain of under review variable, $\dim D_x = n$, D_y is heterogeneous domain of objective variable, $\dim D_y = m$. The given variables can be of arbitrary types (quantitative, ordinal, nominal). For the pattern recognition problem, for example, the variable Y is nominal. Let us put $\Phi_\circ$ is a given class of decision functions. Class $\Phi_\circ$ is μ-measurable functions that puts some subset of the objective variable $E_y \subseteq D_Y$ to each value of the under review variable $x \in D_X$, i.e. $\Phi_\circ = \{f : D_x \to 2^{D_y}\}$. For example, the domain E_y can contain the several patterns $\{\omega_1, \ldots, \omega_k\}$ for pattern recognition problem.

The quality $F(c,f)$ of a decision function $f \in \Phi_\circ$ under a fixed strategy of nature c is determined as $F(c,f) = \int_{D_x} (P(E_y(x)/x) - \mu(E_y(x)))dP(x)$, where $E_y(x) = f(x)$ is a value of decision functions in x, $P(y \in E_y(x)/x)$ is a conditional probability of event $\{y \in E_y\}$ under a fixed x, $\mu(E_y(x))$ is measurable of subset E_y. Note that if $\mu(E_y(x))$ is probability measure, than criterion $F(c,f)$ is distance between distributions. If the specified probability coincides with equal distribution than such prediction does not give no information on predicted variable (entropy is maximum). On the nominal-real space $\Omega = D_н \times D_в$ a measure μ is defined so as any $E \in \mathrm{B}$, $E = \bigcup_{j=1}^{|E_н|} E_в^j \times \{z^j\}$, $\mu(E) = \sum_{j=1}^{|E_н|} \frac{\mu(E_в^j)}{|D_н|\mu(D_в)}$, were $E_н$ is projection of set E on nominal space $D_н$, z^j - item of $E_н$, $E_в^j$ - set in $D_в$ corresponding to z^j, $\mu(E_{\hat{a}}^j)$ – Lebesgue measure of set $E_в^j$. For any subset of domains D_X or D_Y the measure μ is assigned similarly. Clearly, the prediction quality is higher for those E_y whose measure is smaller (accuracy is higher) and the conditional probability $P(y \in E_y(x)/x)$ (certainty) is larger. For a fixed strategy of nature c, we define an optimal decision function $f_\circ(x)$ as such as $F(c,f_\circ) = \sup_{f \in \Phi_\circ} F(c,f)$, where $\Phi_\circ$ is represented above class of decision functions.

In commonly when we solve this problem in practice the size of sample is small and type of variables is may be different type. In this case is used class of logical decision function Φ_M complexity M [Lbov and Starceva, 1999]. For the prediction problem of the heterogeneous system variables class Φ_M is defined as $\Phi_M = \{f \in \Phi_\circ \mid f \sim <\alpha, r(a)>, \alpha \in \Psi_M, r(\alpha) \in R_M\}$ (the mark '~' denotes the correspondence of pair $<\alpha, r(a)>$ to symbol f), were Ψ_M is set of all possible partitioning $\alpha = \{E_X^1, \ldots, E_X^M \mid E_X^t = \prod_{j=1}^n E_{X_j}^t,\ E_{X_j}^t \subseteq D_{X_j}, t = \overline{1,M}, \bigcup E_X^t = D_X\}$ of domain D_X on M noncrossing subsets, R_M is set all possible decisions $r(\alpha)=\{E_y^1,...,E_y^M \mid E_y^t \in \Im_{D_Y}, t=\overline{1,M}\}$, $\Im_{D_Y}$ – set of all possible m-measuring intervals. For that class the measure $\mu(E_y(x)) = \frac{\mu(E_y)}{\mu(D_Y)} = \prod_{j=1}^{m} \frac{\mu(E_{y_j})}{\mu(D_{y_j})}$ is the normalized measure of subset E_y and it is introduced with taking into account the type of the variable. The measure $\mu(E_y(x))$ is measure of interval, if we have a variable with ordered set of values and it is quantum of set, if we have a nominal variable (it is variable with finite non-ordering set of values and we have the pattern recognition problem). A complexity of Φ_M class is assigned as M if we have invariante prediction (decision is presented by form: if $x \in E_X^t$, than $y \in E_y^t$), $M_\Phi = M$, and it is assembly $(k_1,...,k_M)$ if we have multivariate, i.e. $E_y^t = \bigcup_{i=1}^{k_t} E_y^i$, $t=1,...,M$ and $E_y^i \cap E_y^j = \varnothing$ for $i \neq j$ (decision is presented by form: if $x \in E_X^t$, than $y \in E_y^1 \vee E_y^2 \vee ... \vee E_y^{k_t}$). The class of logical decision function has universal property.

10.2.2 Properties of the method

Statement. For any function $f \in \Phi^\circ$ and $\varepsilon > 0$ there is M and several logical decision function $f_M \in \Phi_M$ so that $|F(c,f) - F(c,f_M)| \leq \varepsilon$.

Others good properties of the logical decision function class are presented in work [Lbov and Stupina, 2002], [Stupina, 2006] for prediction system heterogeneous variables problem.

Definition 1. The strategy of nature is c={p(x,y)=p(x)p(y/x))}, where a conditional probability p(y/x) is specified for any elements on B.

Theorem. For a fixed type of the predicate, the class Φ_M of logic decision functions is a universal class in the problem of prediction multivariate heterogeneous value by criterion F(c,f), i.e. for any strategy of nature c and any $\varepsilon > 0$ there exists a number M (M=1,2,3,…) and for some logical decision function f $\in \Phi_M$ (it is represented in the form of decision tree on M vertices) such that $\left|F(c,f) - F(c,f_\circ)\right| \leq \varepsilon$, where $f_\circ$ is optimal function in class $\Phi_\circ$.

The proof of this theorem readily follows from the property of μ-measurability and P-measurability of space D and its projections on the space D_X, D_Y correspondingly. This theorem for pattern recognition of two patterns was proofed in work [Lbov and Starceva, 1994].

We can introduce a complexity of distribution (strategy of nature c) using the class logical decision function. It is necessary for solving statistical stability problem of decision function [Lbov and Stupina, 2003].

Statement 1. For any nature strategy c the quality criterion $F(c,f)$ (risk function) of logical decision function f belonging to Φ_M is presented by following expression:

$$F(c,f) = \int_{D_X}\int_{D_Y} (1 - L(y,f(x)))p(x,y)dxdy = \sum_{t=1}^{M} p_x^t (p_{y/x}^t - \mu^t),$$

where the loss function $L(y,f)$ such as $L(y,f) = \begin{cases} p_o, & y \in \beta \\ 1 + p_o, & y \notin \beta \end{cases}$, $p_o = \mu(E_Y^t)$, $\beta = f(\alpha), \alpha \in \Psi_M$.

Definition 2. To each subclass Φ_M we put in correspondence the subset $L_\varepsilon(M)=\{c:\exists f\in\Phi_M,|F(c,f)-F(c,f_\circ)|\le\varepsilon\}$ of nature strategies; ε is an arbitrarily small number determining an admissible error level of this subset of strategies, where $f_\circ$ is optimal function in class $\Phi_\circ$.

The complexity measure of each subset $L_\varepsilon(M)$ is defined as the complexity measure of the corresponding subclass of decision functions: $\nu(L_\varepsilon(M))=\nu(\Phi_M)=M$. Accordingly, the nature strategies c belonging to $L_\varepsilon(M)$ has complexity measure M. The important statement follows from this theorem and definition.

Statement 2. The set of all possible strategies can be ordered according to complexity, i.e. $L_\varepsilon(1)\subset L_\varepsilon(2)\subset\ldots\subset L_\varepsilon(M)\subset\ldots\subset L_\circ$, and $\varepsilon^{M+1}\le\varepsilon^M$, where $\nu(L_\varepsilon(M))$=M is the complexity and ε^M is the admissible error level of the strategy class $\nu(L_\varepsilon(M))$.

We can suppose that the true (optimal) decision function belongs to Φ_M it is followed from this statement 1.

Definition 3. Define a nature strategy c_M (generated by logical decision function $f\in\Phi_M$) such as set of parameters satisfying the following conditions:

1) $\sum_{t=1}^{M} p_x^t = 1$,

2) $P(E_Y^t/E_X^t)=p_{y/x}^t$ (conditional distribution is same for any $x\in E_X^t$ and $y\in E_Y^t$),

3) $P(\bar{E}_Y^t/E_X^t)=1-p_{y/x}^t$,

where $E_X^t\in\alpha$, $E_Y^t\in\beta$, $<\alpha,\beta>\sim f\in\Phi_M$. The complexity of this strategy is M, i.e. $\nu(c_M)=M$. Note that c_M generated by logical decision function belongs to class $L_\varepsilon(M)$. Clearly, the decision function that generated this strategy is optimal function in class Φ_M.

Statement 3. For a fixed nature strategy $c_M\in L_\varepsilon(M)$ of complexity M, the quality criterion $F(c_M,\tilde{f})$ (risk function) of logical decision function $\tilde{f}\in\Phi_{M'}$ of complexity M' is presented in following form:

$$F(c_M,\tilde{f})=F(\tilde{\alpha})=\sum_{t'=1}^{M'}\tilde{p}_x^{t'}\rho^{t'}=\sum_{t'=1}^{M'}\tilde{p}_x^{t'}(\tilde{p}_{y/x}^{t'}-\mu_y^{t'}),$$

where $\tilde{p}_x^{t'}=P(x\in\tilde{E}_X^{t'})=\sum_{t=1}^{M}p_x^t\frac{\mu(\tilde{E}_X^{t'}\cap E_X^t)}{\mu(E_X^t)}$,

$$\tilde{p}_{y/x}^{t'}=\frac{1}{\tilde{p}_x^{t'}}\sum_{t=1}^{M}p_x^t\frac{\mu(\tilde{E}_X^{t'}\cap E_X^t)}{\mu(E_X^t)}\left(p_{y/x}^t\frac{\mu(\tilde{E}_Y^{t'}\cap E_Y^t)}{\mu(E_Y^t)}+(1-p_{y/x}^t)\frac{\mu(\tilde{E}_Y^{t'})-\mu(\tilde{E}_Y^{t'}\cap E_Y^t)}{1-\mu(E_Y^t)}\right).$$

Remark. If the nature strategy c_M such that some subset E_Y^t coincides with the space D_Y, than

$$\tilde{p}_{y/x}^{t'}=\frac{1}{\tilde{p}_x^{t'}}\sum_{t=1}^{M}p_x^t\frac{\mu(\tilde{E}_X^{t'}\cap E_X^t)}{\mu(E_X^t)}p_{y/x}^t\frac{\mu(\tilde{E}_Y^{t'}\cap E_Y^t)}{\mu(E_Y^t)}.$$

It is followed from that $p_{y/x}^t=P(D_Y/E_X^t)=1$, $\mu(D_Y)=1$.

Consequence 1. If the decision function $\widetilde{f}$ belonging to Φ_M coincides with the function f belonging to Φ_M, than $F(c,\widetilde{f}) = F(c,f)$.

Consequence 2. For the decision function $\widetilde{f}$ belonging to $\Phi_{M'}$ we have the expression $P(\overline{\widetilde{E}}_Y^{t'} / \widetilde{E}_X^{t'}) = 1 - \widetilde{p}_{y/x}^{t'}$.

Really, it is follows from the statement 3, where $\dfrac{\mu(\overline{\widetilde{E}}_Y^{t'} E_Y^t)}{\mu(E_Y^t)} = \dfrac{\mu(E_Y^t) - \mu(E_Y^t \widetilde{E}_Y^{t'})}{\mu(E_Y^t)}$,

$\dfrac{\mu(\overline{\widetilde{E}}_Y^{t'} \overline{E}_Y^t)}{\mu(\overline{E}_Y^t)} = \dfrac{1 - \mu(E_Y^t) - \mu(\widetilde{E}_Y^{t'}) + \mu(E_Y^t \widetilde{E}_Y^{t'})}{1 - \mu(E_Y^t)}$.

Consequence 3. If we have M=1 and the optimal function f generating c_1 such that $E_Y^1 = D_Y$, than $F(c_1,f) = 0$.

Really, for the express of criterion we have $F(c,f) = \sum_{t=1}^{M} \left(P(E_X^t E_Y^t) - P_o(E_Y^t) \right) = P_o(D_X D_Y) - P_o(D_Y) = 0$. It means that we have the event distribution in D for the nature strategy of the complexity M=1. It is case when the entropy is maximum.

Consequence 4. If we have M=1 and the optimal function f generating c_1 such that $E_Y^1 = D_Y$, than for any decision function $\widetilde{f} \in \Phi_{M'}$ the criterion $F(c_1,\widetilde{f}) = 0$.

Really, $\widetilde{p}_{y/x}^{t'} = \dfrac{\mu(\widetilde{E}_Y^{t'} D_Y)}{\mu(D_Y)} P_o(D_Y / D_X) = \mu(\widetilde{E}_Y^{t'})$, $\widetilde{p}_x^{t'} = \dfrac{\mu(\widetilde{E}_Y^{t'} D_X)}{\mu(D_X)} P_o(D_X) = \mu(\widetilde{E}_X^{t'})$,

$F(c_1,\widetilde{f}) = \sum_{t'=1}^{M'} \mu(\widetilde{E}_X^{t'})(\mu(\widetilde{E}_Y^{t'}) - \mu(\widetilde{E}_Y^{t'})) = 0.$

Consequence 5. If the decision function $\widetilde{f}$ belongs to Φ_1 and $\widetilde{E}_Y^1 = D_Y$, than we have $F(c_M,\widetilde{f}) = 0$ for any complexity $M \geq 1$.

Really, we have $\widetilde{p}_x = \sum_{t=1}^{M} p_x^t \dfrac{\mu(D_X E_X^t)}{\mu(E_X^t)} = 1$, $\widetilde{p}_{y/x} = \sum_{t=1}^{M} p_x^t \left(p_{y/x}^t \dfrac{\mu(D_Y E_Y^t)}{\mu(E_Y^t)} + (1 - p_{y/x}^t) \dfrac{1 - \mu(D_Y E_Y^t)}{1 - \mu(E_Y^t)} \right) = 1$.

If the strategy of nature is unknown the sampling criterion $F(\bar{f})$ is used by method $Q(v_N)$ of constructing sample decision function $\bar{f}$, $\bar{F}(\bar{f}) = \sum_{t=1}^{M'} \bar{p}_x^t (\bar{p}_{y/x}^t - \bar{\mu}_y^t)$, were $\bar{p}_x^t = \frac{N(\tilde{E}_X^t)}{N(D_X)} = \frac{N^t}{N}$, $\bar{p}_{y/x} = \frac{N(\hat{E}_y^t)}{N(\tilde{E}_X^t)} = \frac{\hat{N}^t}{N^t}$, $\bar{\mu}_y = \mu(\hat{E}_y)$, $N('*')$ is number of sample points, generating the set "*", $\bar{f} \sim < \alpha, r(\alpha) >$, $\alpha = \{\tilde{E}_X^1,...,\tilde{E}_X^{M'}\} \in \Psi_{M'}$, $r(\alpha) = \{\hat{E}_y^1,...,\hat{E}_y^{M'}\} \in R_{M'}$. The optimal sample decision function is $\bar{f}^* = \arg\max_{\alpha \in \Psi_{M'}} \max_{r(\alpha) \in R_{M'}} \bar{F}(\bar{f})$. In order to solve this extreme problem we apply the algorithm MLRP of step-by-step increase attachments of decision trees. It do the branching of top point on that value criterion $\bar{F}(\bar{f})$ is maximum and the top point is divisible or $\bar{F}(\bar{f}) \geq F^*$. The top point is indivisible if 1) number of final top point is $M' = M^*$ or 2) $\hat{N}^t \leq N^*$. That criterion and parameters F^*, M^*, N^* assign method of constructing sample decision function.

In order to estimate the MLRP – method quality we did statistical modeling. The average of the criterion of sample decision function on samples of fixed size $m_F(c) = \mathrm{E}_{V_N} F(c,\bar{f})$ is estimated for fixed nature strategy. Moreover we researched the averaging-out empirical functional quality $\varepsilon_N(c) = \mathrm{E}_{V_N} F(c,\bar{f}) - \mathrm{E}_{V_N} \bar{F}(\bar{f})$ for given strategy of nature with the purpose of estimating decision quality, and maximal removal of empirical functional quality average of distribution $\varepsilon_N^*(c) = \sup_{c:\bar{F}(\bar{f})=F_0} \varepsilon_N(c)$ for a given empirical quality value $F_\circ$. It was taken for some parametric nature strategy class, for given nature strategy complexity M, decision function complexity M'. The decision function is builder by MLRP-method on sample of size N. Parameters n, m (dimensions of domains D_X and D_Y) and a quantity of fixed type variables was considered in problem statement overall. It is defined the complexity of nature strategy and decision function in addition. The GenMLRP-algorithm was developed for modeling nature strategy parameters. Generation nature strategies were realized in accordance with definition, were parameters is established by random in the given interval. The properties of functional quality are presented in work [Stupina, 2006] for uniform distribution on set D_Y. Such approach to research of statistical stability was used in earlier works, but for simple and one-dimensionality cases [Raudis, 1976], [Startseva, 1995], [Berikov, 2002], [Lbov and Stupina, 1999]. The MLRP-method was applied for prediction multivariate time series. Three random processes were simultaneously considered instead of one. The feature systems (under review and predicted) was established. Procedure of building data table is offered in work [Stupina and Lbov, 2006]. The example of solving practical problem is presented in next paragraph.

10.2.3 Algorithm description

The MLRP- algorithm of construction of logical-and-probabilistic model algorithm carries out consecutive construction of fragmentation space D_X on two subspaces and representation of the decision in the form of a tree. Each fragmentation space let's compare with own node b. To initial fragmentation $\alpha = \{D_X\}$ there corresponds node b°, a set of corresponding decisions $r(\alpha) = \{D_Y\}$ and sample v of volume N.

Step 1. At the given stage on initial sample v of volume N for fixed fragmentation α search of the best set of decisions $r(\alpha) = \{E_y^1\}$ on empirical criterion of quality is carried out.

Procedure of consecutive truncation of space D_Y to cover $\hat{E}_y^1$ makes consecutive removal of "extreme" points of sample until value of criterion of quality will not start to decrease. Thus, for fragmentation α the maximum value of empirical criterion $\bar{F}(\bar{f}) = \bar{F}(\alpha)$ equal to $\bar{F}^1$ is received. Corresponding to the given value of criterion decision function is $\bar{f} \sim <\alpha, r(\alpha)>$, $\alpha \in \Psi_1(\nu)$, $r(\alpha) \in R_1(\nu)$, where $\Psi_1(\nu)$ – a class of fragmentations of complexity one, constructed on sample v, $R_1(\nu)$ – corresponding set of decisions.

Step 2. On the given step construction of fragmentation of initial area on two subspaces is carried out, i.e. branching of initial node (step 1) of which initial fragmentation consists, on two final nodes. If the node is indivisible, the algorithm finishes work and result is the fragmentation constructed on the previous step. If node is divisible then search on all variables $X_1,\ldots,X_j,\ldots,X_n$ is consistently made.

For each variable X_j various fragmentations of set D_{X_j} into two subsets E_{X_j} and $\bar{E}_{X_j}$ such, that $E_{X_j} \neq \varnothing, E_{X_j} \neq D_{X_j}$, where E_{X_j} – any subset for nominal variables, $E_{X_j} = \left\{ x_j \mid x_j \leq \gamma \right\}$ – for serial and real variables, $\gamma \in \{\gamma_1,...,\gamma_N\}$, $\gamma_i \in R$ are considered. Let's designate set of every possible fragmentations α of space D_X on variable X_j on two subspaces through $\Psi_2^j(v)$, i.e. $\alpha \in \Psi_2^j(v)$, $\alpha = \left\{ E_X^1, E_X^2 \right\}$, $E_X^1 = D_{X_1} \times \cdots \times E_{X_j}^1 \times \cdots \times D_{X_n}$, $E_X^2 = D_{X_1} \times \cdots \times \bar{E}_{X_j}^1 \times \cdots \times D_{X_n}$. Fragmentation $\alpha \in \Psi_2^j(v)$ is compared with final nodes b^1, b^2 and decisions $r(\alpha) = \left\{ E_Y^1, E_Y^2 \right\}$, $r(\alpha) \in R_2(v)$ corresponded to them. Thus, object a_i belongs to node b^t, $t = 1,2$, if point $x^i \in E_X^t$ corresponds to its description $X(a_i)$, where $E_X^1 = \left\{ x \mid x_j \in E_{X_j} \right\}$, $E_X^2 = \left\{ x \mid x_j \in \bar{E}_{X_j} \right\}$. Further fragmentation α is fixed and search of the best decision $r(\alpha) = \{\hat{E}_y^1, \hat{E}_y^2\}$ such that criterion $\bar{F}(\bar{f})$ accepted the maximum value is made. After full search on all variables $j = 1,...,n$ (i.e. on all fragmentations $\alpha \in \Psi_2^j(v)$) and on probable decisions $r(\alpha) \in R_2(v)$ it is found the best fragmentation $\alpha^* \in \Psi_2^{j_1}(v)$ of space D_X and the best decision $r^*(\alpha^*) \in R_2(v)$ corresponding to it.

Let for fragmentation α^* on variable X_{j_1} the empirical criterion of quality has maximum value $\bar{F}^2$ then the decision on branching of initial node b^o is made. If $\bar{F}^2 > \bar{F}^1$ we fix branching and the initial table (sample) v breaks on two tables v^1 and v^2. Table v^1 includes realizations $x^i \in E_X^1$, table v^2 – realizations $x^i \in E_X^2$. These tables are put in conformity to nodes b^1 and b^2, as a result we receive a tree consisting of two nodes. To nodes b^1 and b^2 we give the status of initial nodes. If condition $\bar{F}^2 > \bar{F}^1$ is not carried out, initial node b^o is left former, we appropriate to it the status of final node and exit from a step 2.

Step 3. The step 2 repeats for each initial node. The stop-condition of algorithm is indivisibility of final node or achievement on the given step the specified number of final nodes M^* (entrance parameter of algorithm).

Node b^t is considered final and not subject to division, if number of selective points N^* in the given node is less than some entrance parameter N^* defining the minimum admissible number of objects in node.

The choice of the best values of parameters N^*, M^* is connected with the question of estimation of quality of a method of construction of logical-and-probabilistic model on the limited sample.

Algorithm MLRP is software implemented in programming language Microsoft Visual C ++. In an interface window, it is necessary to specify a full path to file with entrance data and the file name where found solving rule in the form of a set of laws will be written. In a data file, it is necessary to specify:

- dimension of space of entrance variables;
- dimension of space of target variables;
- volume of sample of training;
- describe types of all sets of variables and a range of their change (binary, nominal, continuous, discrete, serial);

– enter the table of data according to order of the description of variables.

In an interface window parameters N^* – the minimum number of points in tree node, M^* – the maximum number of dangling nodes of a tree of decisions, Contin Grid Step – a step of sampling of an interval for a continuous variable, Infimum of Pyt – is minimum admissible estimation of number of the points which belongs to given space at fixed fragmentation are assigned.

Algorithm MLRP carries out the directed search of the best decision, therefore is locally optimum.

10.3 Application MLRP-method to prediction multivariate time series problem

10.3.1 The MLRP-method to prediction multivariate time series problem

Let us consider terminal time series $\{x(t), t \in T\}$, that it is realization of sometime-dependent random process η(t). One is supposed that simultaneous distribution $p(\eta_1)$, $p(\eta_1, \eta_2)$, $p(\eta_1, \eta_2, \eta_3)$,…,$p(\eta_1,\ldots, \eta_T)$ is exist. The value set $D_{\eta(t)}$ of variables may be quantitative, nominal, and ordinal type in a more case. Let the values of random process η(t) are measured at consequent moments of the time with the gap $\Delta t = t_k - t_{k-1}$. Denote this set of moments as $T = \{t_1,\ldots,t_k,\ldots,t_N\}$, N << ∞.

Classical problem of prediction time series is consist in that we must to take predict at time moment $t = t_R$ on time period $t_{R+\tau}$ by analyzing prehistory $\{x(t_k)\}$, $k = 1,\ldots,d$, with length d. As a rule, the value τ is named as forestalling. The set of every possible all prehistory, that have length d denote as D_X, and the set of every possible all sets forestalling denote as D_Y. Let us understand a prediction decision function as a f mapping of the D_X set on the D_Y set, i. e. $f: D_X \to D_Y$, $\dim D_X = d$, $\dim D_Y = \tau$. Model's construction f of prediction is defined by decision function class Φ.

If a simultaneous distribution is known than optimal decision function, constructing predict to time $t+\tau$, is conditional average of distribution $E(\eta_{t+\tau} \mid \eta_{t-d},\ldots,\eta_t)$. In order to solve this problem it is necessary to restore conditional distribution. However, that way is not practical because we have not enough size of sample in applied tasks. Therefore, it is possible to offer a different depending on specified suggestions targets setting (concerning properties of random process) and the different methods (concerning decision function class) of their decision accordingly.

At present time it was developed many method for prediction depended of time random process (probabilistic characteristics of process are not changed on time). Its methods are based on constructing several models, which usually use some suggestion. For example, if we want to do long-time prediction than the best offer (concerning error variance value) is global model, if we want to do short-term prediction, than it is local model [Bezruchko and Smirnov, 2003]. Note that most of models accomplish solitary prediction. It is next time $t + \Delta t$ or at time moment $t + \tau\ \Delta t$, $\tau = 2,\ldots,N-d$.

We propose model, that accomplishes prediction on all forestalling term τ, in other words, to time moments $t + k\ \Delta t$, $k = 1,\ldots,\tau$. That prediction allows to take one decision function (structure of model) and to do simultaneously several predictions on future by one prehistory.

For that problem statement, it is important to researcher several steps:

– detection time moment of changing model's structure (criterion of adaptive);

- optimization of prehistory length d;
- optimization of forestalling term τ .

In order to solve these items we will use class of logical decision function. We will consider two ways:

a) when the information about event is kept safe in the process, and
b) when it is kept safe in depending process.

We will perform the primary ideas of these ways in following paragraphs.

✓ Analysis unvariate time series problem

Let us we have univariate time series { $x(t)$ } of any random process $\eta(t)$. It is necessary to problem of constructing predict function f by empirical data, which is presented as terminal points N for given prehistory length d and forestalling term τ. We will construct decision function from the logical decision function Φ_M by sample data, which is made from points of discrete time series. Procedure of building data table $v = \{v_x, v_y\}$ depends of problem statement and of data generally.

For example, it may be

a) Shift of prehistory window step-by-step on time series,
b) Shift of prehistory window to some position on time series,
c) Building prehistory window from the series points, that is positioned on some distance.

In addition, we can consider some combination of items indicated above. Visual illustration of univariate time series and principle of building sample table are presented on Figure 163.

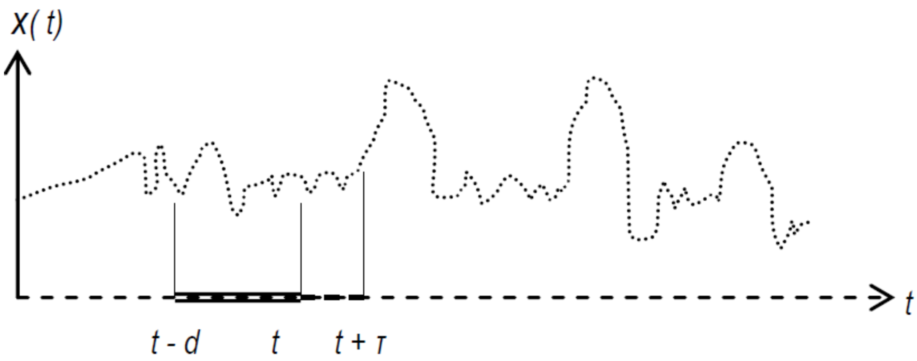


Figure 163. Analysis univariate time series

Not lose commonality let us consider, that $\Delta t = 1$, then prehistory table is built as $v_x = \{x_{kj}\} = \{x(j+k-1)\}$, were $j = R-d+1,\ldots,R$, $d \le R \le N-\tau$, $k = 1,\ldots,N-R-\tau$, and forestalling term table (future predictions) is builder as $v_y = \{x_{kj}\} = \{x(j+k)\}$, were $j = R+1,\ldots,R+\tau$, $k = 1,\ldots,N-R-\tau$, for case (a) as above. With the help of data table $v = \{v_x, v_y\}$ of the size $N-R-\tau$ we will construct sample decision function $\bar{f}$ from the class Φ_M by the MLRP-method. So we have that choose of optimal length d^* of prehistory corresponds to choose of informative characteristic subset. A choose of optimal forestalling term length τ^* will be correspond to definition of likely problem size (complexity) for a given sample size.

We define a time moment of changing model's structure (adaptive) as a time moment $t^* = t_{R-d}$ for which the condition $|F(\bar{f}) - F^*| \geq h$ is carry out, were the value F^* is threshold value of model quality, h is admissible value of deviation for established quality.

Below we will consider logical decision function class Φ_M and its properties. We will define criterion of quality F(f) for decision function f.

✓ **The prediction with respect to other time series**

This paragraph is devoted to detection of correlation between two univariate time series. That problem statement is well known and is commonly applied for solving practice problem [Bezruchko and Smirnov, 2003]. However for the most part methods indicate some power of correlation for the given time point.

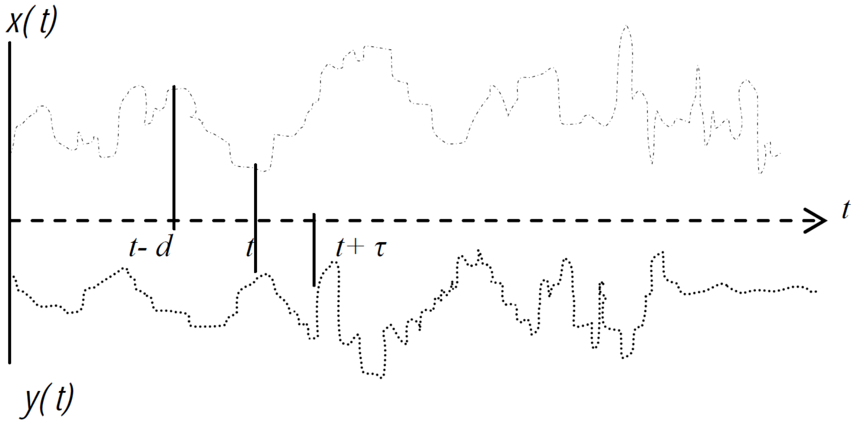


Figure 164. Two dependent time series

The method to attack is founded on constructing so function for that definitional domain is assigned in domain of realization one time series $\{x(t)\}$, and value domain (domain of prediction point $t+\tau\ \Delta t$, $\tau = 1,\ldots,N-d$.) is assigned in domain of realization other time series $\{y(t)\}$. It is supposed that with respect to other process. Visual illustration of two dependent time series and principle of building sample table are presented on Figure 164.

The data table is constructed by principle like above. The power of correlation f is defined by quality value $F(f)$ a) on the learning sample and b) on the control sample. We will construct sample decision function $\bar{f}$ (logical-and-probabilistic correlation) from the logical decision function class Φ_M by the MLRP-method like above.

10.4 Application MLRP-method to prediction multivariate time series problem

10.4.1 Application for analysis some of problems in climate

This paragraph is devoted to some of practical problem from hydrological area. It consists in the prediction of the columbine $(k=1)$, transmitting across Oby riverbed, the average monthly temperature $(k=2)$, and the atmospheric precipitates $(k=3)$ by like hydrometeor data (in the course of the 86 years) in control post of the city Kolpashevo of Novosibirsk region in Russia. In order to construct decision function of prediction variable system (y_1, y_2, y_3) in April by variable system in the course November, December, and January the average monthly data was worked up in the course of November $(i=1)$, December $(i=2)$, January $(i=3)$, and April $(j=1)$ in control post. So target setting provides potential for detecting possible shallowness or hydramnios in Oby lade in order to alarm about extremal hydrological situation.

We have considered several versions of this target setting and provided resolutions by MLRP-method. The sample decision functions were constructed by learning data {$x_k(t_{k_i+12})$, $y_k(t_{k_j})$} of the size 76.

First version – prediction of the columbine, the temperature and the atmospheric precipitates in April (three variables) by columbine data, temperature data and the atmospheric precipitates data for November, December and January (nine variables). We have constructed resolution so as logical probabilistic function, presented below.

Variable declaration:

X1 – the average monthly temperature for November,

X2 – the average monthly temperature for December,

X3 – the average monthly temperature for January,

X4 – the average monthly atmospheric precipitates for November,

X5 – the average monthly atmospheric precipitates for December,

X6 – the average monthly atmospheric precipitates for January,

X7 – the average monthly columbine for November,

X8 – the average monthly columbine for December,

X9 – the average monthly columbine for January,

Y1 – the average monthly temperature for April,

Y2 – the average monthly atmospheric precipitates for April,

Y3 – the average monthly columbine for April.

Logical regularity:

1. IF (X4=(45.000, 182.700]),
   THEN (Y1=[-4.200, 1.700] & Y2=[17.000, 60.000] & Y3=[873.000, 7390.000])
   The estimate of probability is 0.90000.
   Number of sample points (N) is 10.
2. IF (X7=[553.000, 1320.000] & X3=(-21.600, 22.300] & X4=[2.000, 31.200]),
   THEN (Y1=[-4.500, 2.500] & Y2=[4.100, 18.000] & Y3=[1010.000, 4890.000])
   The estimate of probability is 1.00000.
   Number of sample points (N) is 5.
3. IF (X4=(31.200, 36.000] & X7=[553.000, 2140.000]),
   THEN (Y1=[-2.100, 3.000] & Y2=[7.800, 46.000] & Y3=[1400.000, 5420.000])
   The estimate of probability is 1.00000.

Number of sample points (N) is 9.

4. IF (X4=(36.000, 45.000] & X7=[553.000, 2140.000]),
THEN (Y1=[-3.700, 1.000] & Y2=[14.000, 34.100] & Y3=[834.000, 3950.000])
The estimate of probability is 0.90000.
Number of sample points (N) is 10.

5. IF (X1=[-31.000, -11.600] & X7=(1320.000, 20700.000] & X3=(-21.600, 22.300] & X4=[2.000, 31.200]),
THEN (Y1=[-4.200, 4.000] & Y2=[7.800, 33.600] & Y3=[1870.000, 5110.000])
The estimate of probability is 1.00000.
Number of sample points (N) is 7.

6. IF (X1=(-11.600, 22.300] & X7=(1320.000, 20700.000] & X3=(-21.600, 22.300] & X4=[2.000, 31.200]),
THEN (Y1=[-6.300, 1.800] & Y2=[13.000, 34.000] & Y3=[1120.000, 4850.000])
The estimate of probability is 0.90909.
Number of sample points (N) is 11.

7. IF (X1=[-31.000, -8.800] & X7=(2140.000, 20700.000] & X4=(31.200, 45.000]),
THEN (Y1=[-4.700, 0.400] & Y2=[18.000, 34.000] & Y3=[856.000, 3410.000])
The estimate of probability is 1.00000.
Number of sample points (N) is 6.

8. IF (X1=(-8.800, 22.300] & X7=(2140.000, 20700.000] & X4=(31.200, 45.000]),
THEN (Y1=[-4.000, 1.900] & Y2=[2.000, 27.000] & Y3=[1530.000, 3270.000])
The estimate of probability is 1.00000.
Number of sample points (N) is 6.

9. IF (X1=[-31.000, -13.100] & X3=[-31.000, -21.600] & X4=[2.000, 31.200]),
THEN (Y1=[-2.600, 1.300] & Y2=[12.500, 33.600] & Y3=[894.000, 6560.000])
The estimate of probability is 1.00000.
Number of sample points (N) is 5.

10. IF (X1=(-13.100, 22.300] & X3=[-31.000, -21.600] & X4=[2.000, 31.200]),
THEN (Y1=[1.400, 3.800] & Y2=[5.400, 35.000] & Y3=[2110.000, 7060.000])
The estimate of probability is 1.00000.
Number of sample points (N) is 7.

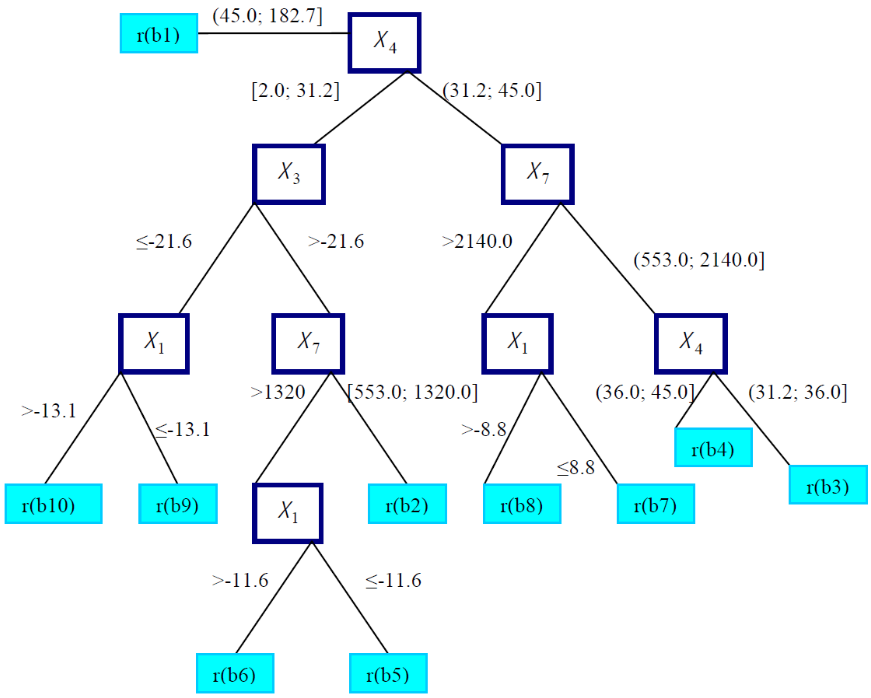
Value of decision function quality $F(\overline{f_{10}})$ is 0.9957355517.

Visual illustration of decision tree for that logical regularity and decision example on top $r(b\ 6)$ are presented on Figure 165 with complexity M = 10, $N^* = 3$.

Estimation of quality criterion (probability estimation of veritable decision by rule $\bar{f}$) was taken by control sample of the size 10 (it is last years of time series) and it was equal 0.8 that is satisfactory result.

For consecutive constructing decision three (the decision function from logical decision function may be presented by dichotomous count) four important features was choose from initial nine features. In compliance with construction decision function the average monthly temperature in November (x_1) and in January (x_2), the atmospheric precipitates in November (x_4), the columbine in November had the most influence on prediction quality.

At consecutive construction of a tree of decisions four significant signs from nine on which fragmentation has been done have been allocated. According to the constructed rule it has appeared, that the greatest influence on quality of the forecast is rendered by monthly average temperature in November (X1) and January (X3), by deposits in November (X4) and by a reservoir in November (X7).



$$r(b6) = "Y_1 \in [-6.3; 1.8]" \wedge "Y_2 \in [13.0; 34.0]" \wedge "Y_3 \in [1120.0; 4850.0]"$$

Figure 165. The decision tree of hydrological situations

Second version – prediction of meteorological values (the temperature and the atmospheric precipitates) in April (three variables) by corresponding data for November, December and January.

The minimum number of laws is 10; the minimum number of points in node is four.

First logical regularity:

1. IF (X5=(25.100, 56.000] & X4=[10.000, 33.000]),
   THEN (Y1=[-6.300, 3.000] & Y2=[13.000, 19.300])

   The estimate of probability is 1.00000.

   Number of sample points (N) is 6.

2. IF (X4=[10.000, 16.700] & X5=[8.700, 25.100]),
   THEN (Y1=[-5.800, 4.000] & Y2=[7.800, 33.600])

   The estimate of probability is 1.00000.

Number of sample points (N) is 6.

3. IF (X5=(18.300, 25.100] & X4=(16.700, 33.000]),
THEN (Y1=[-0.600, 2.500] & Y2=[4.100, 31.000])
The estimate of probability is 1.00000.
Number of sample points (N) is 6.

4. IF (X5=[8.700, 23.000] & X4=(33.000, 68.000]),
THEN (Y1=[-4.200, 0.100] & Y2=[17.000, 32.000])
The estimate of probability is 1.00000.
Number of sample points (N) is 7.

5. IF (X5=(23.000, 27.000] & X4=(33.000, 68.000]),
THEN (Y1=[-1.700, 1.900] & Y2=[2.000, 42.000])
The estimate of probability is 1.00000.
Number of sample points (N) is 6.

6. IF (X5=[8.700, 12.400] & X4=(16.700, 33.000]),
THEN (Y1=[-3.000, 3.000] & Y2=[7.900, 33.600])
The estimate of probability is 1.00000.
Number of sample points (N) is 5.

7. IF (X5=(12.400, 18.300] & X4=(16.700, 33.000]),
THEN (Y1=[-2.000, 3.700] & Y2=[5.400, 34.500])
The estimate of probability is 1.00000.
Number of sample points (N) is 9.

8. IF (X3=(-18.400, -11.000] & X5=(27.000, 56.000] & X4=(33.000, 68.000]),
THEN (Y1=[-3.100, 2.900] & Y2=[23.000, 60.000])
The estimate of probability is 1.00000.
Number of sample points (N) is 4.

9. IF (X6=[3.800, 21.000] & X3=[-31.000, -18.400] & X5=(27.000, 56.000] & X4=(33.000, 68.000]),
THEN (Y1=[-3.800, 1.900] & Y2=[7.700, 52.300])
The estimate of probability is 1.00000.
Number of sample points (N) is 7.

10. IF (X6=(21.000, 47.000] & X3=[-31.000, -18.400] & X5=(27.000, 56.000] & X4=(33.000, 68.000]),
THEN (Y1=[-2.100, -0.300] & Y2=[7.800, 46.000])
The estimate of probability is 1.00000.
Number of sample points (N) is 4.

Value of decision function quality $F(\bar{f}_{10})$ is 0.8372715792.

Visual illustration of decision tree for that logical regularity is presented on Figure 166 with complexity M = 10, $N^* = 4$.

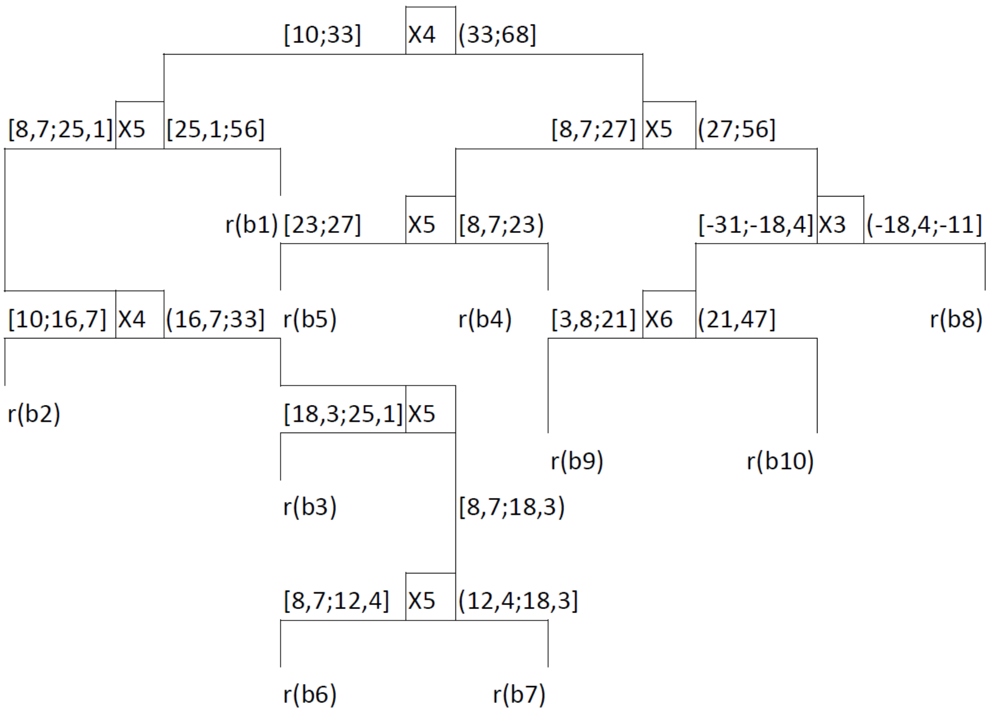


Figure 166. First decision tree of meteorological situations

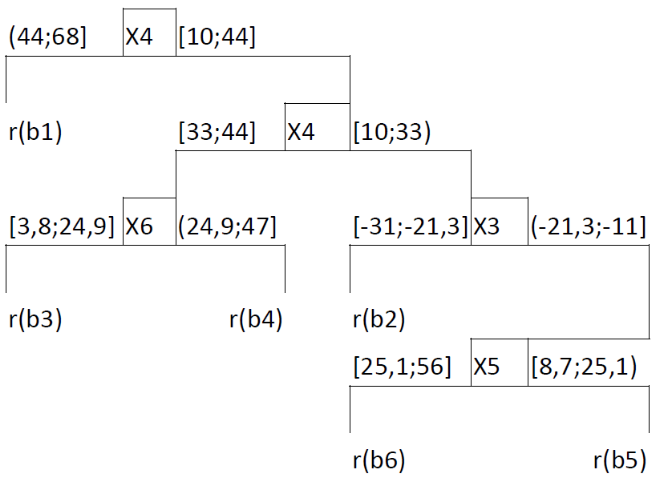


Figure 167. Second decision tree of meteorological situations

Forecast decision function has been constructed on training sample of volume 60. The estimation of criterion of quality is received on control sample of volume 26 (last twenty six years) and is equal 0.53, thus the error on the control at forecasting only monthly average temperature is equal 0.73, at forecasting only deposits – 0.69. Result visualization of the decision is presented in the form of a tree of decisions on Figure 167.

In addition the solving tree of complexity 6 (the maximum number of nodes of a tree of decisions) with the minimum number of points in node equal 4 has been received.

Second logical regularity:

1. IF (X4=(44.000, 68.000]),
   THEN (Y1=[-8.800, 1.700] & Y2=[15.000, 60.000])

The estimate of probability is 1.00000.

Number of sample points (N) is 11.

2. IF (X3=[-31.000, -21.300] & X4=[10.000, 33.000]),
THEN (Y1=[-3.000, 3.900] & Y2=[5.400, 35.000])

The estimate of probability is 1.00000.

Number of sample points (N) is 16.

3. IF (X6=[3.800, 24.900] & X4=(33.000, 44.000]),
THEN (Y1=[-3.800, 1.900] & Y2=[2.000, 37.400])

The estimate of probability is 1.00000.

Number of sample points (N) is 16.

4. IF (X6=(24.900, 47.000] & X4=(33.000, 44.000]),
THEN (Y1=[-4.700, 2.900] & Y2=[15.000, 46.000])

The estimate of probability is 1.00000.

Number of sample points (N) is 9.

5. IF (X5=[8.700, 25.100] & X3=(-21.300, -11.000] & X4=[10.000, 33.000]),
THEN (Y1=[-5.800, 4.000] & Y2=[4.100, 34.000])

The estimate of probability is 1.00000.

Number of sample points (N) is 19.

6. IF (X5=(25.100, 56.000] & X3=(-21.300, -11.000] & X4=[10.000, 33.000]),
THEN (Y1=[-6.300, -2.800] & Y2=[13.000, 25.000])

The estimate of probability is 1.00000.

Number of sample points (N) is 5.

Value of decision function quality $F(\bar{f}_6)$ is 0.7169073023.

Forecast decision function has been constructed on training sample of volume 76. The estimation of criterion of quality is received on control sample of volume 10 and is equal 0.5, thus the error on the control at forecasting only monthly average temperature is equal 0.6, at forecasting only deposits – 0.8. Result visualization of the decision is presented in the form of a tree of decisions on Figure 167.

According to the constructed rules it has appeared, that the greatest influence on quality of the forecast of deposits and temperature in April is rendered by monthly average temperature in January (X3), by deposits in November, December, January (X4, X5, X6).

Demonstration of solving problems from the considered applied areas shows efficiency of the offered methods of construction of logical-and-probabilistic models and can be applied in many natural-science areas.

10.4.2 Application for analysis some of problems in seismology

In this section, we consider the problem of detection of statistical regularities in the analysis of seismological data presented time series. The task of study of the earthquake as a physical process is quite complicated at least because of the earthquake originated in the deep bowels of the earth, not accessible to direct observation and measurement. Analysis of spatial-temporal course of seismicity is one of the promising directions in the modern seismology. Study of the regularities that occur before, during or after the earthquake, provides an opportunity to get closer to the mystery of earthquakes and, possibly, give some predictions of their occurrence or migration [Nikonov, 2006]. We were interested in testing of the algorithm MLRP on seismic data to test the hypothesis of logical-probabilistic regularities between seismic events occurring in two geographical areas.

The approach proposed to solve this problem, based on the construction of predictive function f: $D_x \rightarrow D_y$ from the class Φ_M of logical decision functions. The area of definition D_x is given in the space of realizations of a time series (x (t)), and the area of values D_y (forecasts) – in the space of realizations of another time series (y (t)), the complexity of the class of decision functions defined by the parameter M. A series {x (t)} is formed from the values of maximum magnitude of the earthquakes for every half of day in zone I, a series {y (t)} – the values of maximum magnitude earthquakes for every half of day in zone II, geographically distinguished from the Zone I.

In the data table v = {v_x,v_y} of volume N – d – τ, d – length of the background, τ – the period of advance, we will build a sample decision function $\bar{f}$ of the class Φ_M by algorithm MLRP. Selecting the optimal length d* of prehistory corresponds to choosing an informative subset of features, the choice of the optimal length of the period of advance τ* corresponds to definition of appropriate dimension (complexity) of the problem for a fixed amount of training.

Two zones of equal area, located on a small geographically remote from each other (the first zone presented by the territory between -10° and 0° the latitude and between 150° and 160° the longitude, the second – from -20° to -10° the latitude and from 160° to 170° the longitude) were arbitrarily selected from the data presented in the directory ftp://www.ncedc.org/pub/catalogs/cnss//

For d = 10 we have a set of variables X_i, i = 1,..., 10, whose values are defined as the maximum magnitude of events for the corresponding half of day rounded to the integer value to transfer to the nominal space of variables). Let τ = 2, ie, the variables Y_i, i = 1, 2, refer to the following day for X_{10}. A decision tree for M = 3 and time series of 10000 samples per period of 5000 days, starting from 1970, was constructed.

1. IF ($X_8 \in$ {4, 5, 6, 10}),

THEN ($Y_1 \in$ {0, 5} & $Y_2 \in$ {0, 4, 5, 6})

The estimate of probability is 0,902.

Number of sample points N_1 is 1748.

2. IF ($X_6 \in$ {0, 4, 5, 6, 8, 10} & $X_8 \in$ {0, 1, 2, 3, 7, 8, 9}),

THEN ($Y_1 \in$ {0, 4, 5, 6} & $Y_2 \in$ {0, 5})

The estimate of probability is 0,906.

Number of sample points N_2 is 8092.

3. IF ($X_6 \in$ {1, 2, 3, 7, 9} & $X_8 \in$ {0, 1, 2, 3, 7, 8, 9}),

THEN ($Y_1 \in$ {0, 2, 5, 9} & $Y_2 \in$ {0, 5, 6})

The estimate of probability is 0,900.

Number of sample points N_3 is 160.

Value of decision function quality is 0,839.

Unfortunately, the accuracy of seismic event with a given magnitude in a given class of decision functions is poor. The reason may be either too rough rounding of data and the presence of a large number of events with a magnitude of zero. Nevertheless, it is interesting result, which consists in the fact that relatively short time events (variables X_6, X_8) were separated with the choice of informative subspace of variables.

10.5 Discussion

In this chapter, two ways to solving analysis univariate time series problem were considered. It was founded on the MLRP-method constructing logical-and-probabilistic model for prediction heterogeneous variables system. The idea's approach and ways of realizations was formulated here.

The decision was constructed by MLRP-method from the logical decision function class. It allows taking optimization parameters as such prehistory length and forestalling term for univariate time series. Our method has very priority properties for investigation of very complexity processes. It is: a) easily interpreted decision is constructed; b) the given class allows to process heterogeneous experimental data; c) it has a small measure of complexity; d) it allows to work in the presence of admissions in empirical tables and to receive adequate decision at small volumes of sample.

Very actual practice problems from hydrological and seismology domains are presented here by MLRP-method. We have received quite good results. Results of seismology problem allow us to understand process of geographic migration of seismic zones and perhaps even to detect earthquake precursors. Results of hydrological problem show us the most likely signs, which have most influence on the forecast of meteorological and hydrological features. Demonstration of solving problems from the considered applied areas shows efficiency of the offered methods of construction of logical-and-probabilistic models and can be applied in many natural-science areas.

We want to note that proposed approach to joint analyses of some time series can be have more than enough applications. We can to solve problem of statistically important correlation detection between processes arising on the most distant region. Method presented here depends on statistical information about studied object and can be used in many domains of data mining. We think that many actual problems of monitoring and better understanding our environment can be solved by our method.

We think that our investigations are included to GMES-project properly and we look forward to collaboration with colleagues from other countries. We have science contact with colleagues from Bulgaria GMES and we hope to join efforts to solve general problems. Our knowledge and science experience will be beneficial to GMES project.

11

Polyhedral Coherent Risk Measures and their Application to Investment Decisions Support under Catastrophic Flood Risks

11.1 Introduction and some definitions

The results of managerial decisions are influenced by numerous factors, parameters, and processes that impart them a stochastic, or even, to a certain extent, an uncertain character. It is caused by as the fundamental impossibility to describe fully enough and to forecast future processes in advance, as the stochastic nature of some factors and parameters of these processes (for example, fluctuations in financial markets, natural cataclysms caused by the coincidence of some conditions and phenomena, and so on). This concerns in full degree various decisions in economic and financial fields, which are sensitive enough to numerous risks. Similar properties have a special significance in development and realization of important (global) decisions with long-term consequences, because appropriate uncertainties and risks essentially increase in a long-term prospect. Therefore, the development of a methodology for decision-making under risk and uncertainty, and appropriate models and methods of search of optimal (effective) decisions take a critical meaning in the present conditions of global interferences of various processes.

As a rule, risk arises in those cases where there is a possibility (probability) of adverse consequences (damages, losses), hence, where considered processes cannot be described by deterministic values. Besides, risk can be caused by essentially stochastic nature of parameters and characteristics of processes, or by the impossibility to predict and describe them in advance (uncertainty), or by imposing of both specified circumstances.

In various applications, especially in financial and economic ones, the choice of concept of efficiency is transparent, as a rule, it is a profit, return, utility, etc. At the same time, a question how to estimate risk remains methodologically difficult and ambiguous. Certain functions which describe risk quantitatively are called as risk functions, or, according to terminology [Artzner et al, 1999], as risk measures. What kind should be a risk measure? An inappropriate choice of such measure can lead to inconsistent decision-making results, to difficulties in searching optimal decisions for studied problems, etc.

In the process of development of the theory and applications to support of economic and financial decisions in conditions of risk and uncertainty, different functions were used as risk measures: a dispersion (deviation) [Markowitz, 1959], a semi-deviation [Ogryczak and Ruszczynski, 1999], VaR (Value-at-Risk) [Jorion, 1996], an expected absolute deviation [Konno and Yamazaki, 1991], and others.

However, none of them is perfect. For instance, the dispersion is a traditional measure in the theory of errors for the estimation of a deviation from the mean; however, for financial area measures characterizing downside deviations from some level (loss) are more suitable. Therefore, the dispersion [Markowitz, 1959] estimating both a downside deviation (losses), and an upside deviation (profit), is not an adequate risk measure for financial and economic problems. The same concerns the semi-deviation [Ogryczak and Ruszczynski, 1999] and the expected absolute deviation [Konno and Yamazaki, 1991].

Currently, VaR is the most popular risk measure in financial applications [Jorion, 1996]; it came there from the insurance area and was propagated by the RiskMetrics methodology. The measure has simple and clear interpretation for risk-managers; however, it has drawbacks as well. In particular, it is not subadditive that can lead to the following paradox: a portfolio diversification can increase its risk (in terms of VaR as the risk measure). Besides, it ignores risks of high losses. Therefore, in 2000 the Basel committee on bank supervision did not recommend to use VaR, as well as dispersion, for measuring risks.

Later, in [Artzner et al, 1999] four axioms have been formulated, to which from a theoretical point of view a risk function should satisfy to claim to be a successful measure of risk, and the corresponding class of risk functions was called as the class of coherent risk measures (CRM). We remind that such functions should be: 1) translation invariant; 2) subadditive; 3) positively homogeneous; 4) monotonous.

Then in [Rockafellar and Uryasev, 2000; 2002], CVaR (Conditional VaR) risk measure was proposed, which, on the one hand, is interpreted as integrated VaR (integral of the appropriate tail distribution), and, on the other hand, belongs to the class of CRM. Remarkable properties of this measure allow to reduce portfolio optimization problems with CVaR objective or/and constraint functions to linear programming problems (LPP).

Later, in [Kirilyuk, 2003; 2004a; 2004b], a class of polyhedral coherent risk measures (PCRM) was introduced, which, on the one hand, is a subset of CRM class (hence it possesses all theoretically attractive properties); on the other hand, it allows to reduce various portfolio optimization problems with such risk measures to LPP. The class contains CVaR and some other important risk measures as special cases.

Let us notice that a considerable interest to CVaR in the financial literature and attempts to consider it as reasonable alternative to VAR currently are observed. However, a question of a choice of the risk measure for concrete problem settings remains open. Use in important problems not only one risk measure, but also a whole set of measures, for instance, in constraints on risk levels which are described by these risk measures, looks quite reasonable.

In this paper we review some results of the PCRM theory from [Kirilyuk, 2003-2007], consider their applications to decision making support in conditions of risk, and develop numerical methods for searching optimal decisions. As a particular application, an investment decisions making under catastrophic flood risks is considered (section 5).

Let there be a certain amount of money (resources), which need to be distributed among several instruments (financial assets, branches, programs, etc.) in such a way that the total result obtained was optimal with respect to efficiency–risk ratio. Accordingly, such instruments are called as the portfolio components, and their total result is called as the portfolio result. Such problems often arise in financial and economic applications.

We consider decision-making models guided definitely by the scenario analysis. Suppose that there is some set of developed scenarios, which allow describing price evolution for each portfolio component. For example, this can be done from the analysis of corresponding prehistory (a statistical data analysis) and a certain forecast of future events. In case when probabilities of the developed scenarios are known, it is clear that distributions of all portfolio components are known as well. Accordingly to terminology [Knight, 1921], we say then that decisions are made in the condition of risk.

A situation, when scenario probabilities are not known but some of their estimates are available, is more difficult and requires certain additional methodological efforts. In this case, the situation is characterized as partially uncertain incomplete information on distributions).

Let us consider now discrete random variables (r.v.) taking n values, which are called scenarios. Then each scenario i = 1, …, n has some probability $p_i > 0$, that is, a vector of scenario probabilities $p_0 = (p_1^0, ..., p_n^0), p_i^0 > 0, i = 1, ..., n, \sum_1^n p_i^0 = 1$ is given, and a r.v. X is characterized by its distribution $x = (x_1, …, x_n) \in R^n$ on these scenarios, hence, it is identified with this n-dimensional vector.

Consider a function of the following form

$$\rho(x) = \max\{E_p[-X] / p \in P\} \quad (1)$$

where $E_p[-X] = \langle -x, p \rangle$ is the mean value of r.v. (–X) with a discrete probability measure p, and P is a convex closed set of probability measures.

Let's remind that functions of form (1) which have a certain sense and where the set P is described in the form of a convex hull of a finite number of points, are called by the polyhedral coherent risk measure (PCRM) [Kirilyuk, 2003]. More exactly, if P is a set of the following form

$$P = co\{p_i: i=1,…,k\},$$

Alternatively, it is equivalent,

$$P = \{p: B\, p \le c,\ p \ge 0\} \quad (2)$$

where B and c are a matrix and a vector of appropriate dimensions, the relations (1)–(2) unambiguously describe a polyhedral coherent risk measure.

Note that as P is a set of probability measures, its description in the form of (2) should include the condition $\sum_{i=1}^{n} p_i = 1$, i.e., $\sum_{i=1}^{n} p_i \le 1, -\sum_{i=1}^{n} p_i \le -1$.

Hence, the first two rows of the matrix B are (1, …., 1) and –(1, …., 1), and the first two components of the vector c are 1 and –1.

Consider known risk measures, which satisfy this PCRM definition. Because of all of them are set in form of (1), for the full description of a concrete measure it is enough to describe the set P from relation (2).

1) WCR is the worst-case risk on all distribution of a r.v. [Artzner et al, 1999]. In the case, $WCR(x) = \max\{-x_i: i = 1,…, n\}$. Then, as it is easy to see, the set P has the following form

$$P_{WCR} = \{p = (p_1,…, p_n): p_i \ge 0,\ i = 1,…, n,\ \sum_1^n p_i = 1\}.$$

2) $CVaR_\alpha$ is the conditional average of losses on α-tail of a r.v. distribution (Conditional VaR) [Rockafellar and Uryasev, 2000;, 2002]. It has the form of (1), where appropriate set P is described as

$$P_{CVaR_\alpha} = \{p = (p_1,…, p_n): p_i \le p_i^0 / \alpha,\ p_i \ge 0,\ i = 1,…, n,\ \sum_1^n p_i = 1\},$$

where $p_0 = \left(p_1^0,\ldots,p_n^0\right)$ is the vector of initial scenario probabilities.

3) WCE $_\alpha$ is the worst conditional expectation of r.v. [Artzner et al, 1999]. Appropriate set P is

$$P_{WCE_\alpha} = \text{co}\{(p_1,\ldots, p_n) / \text{ for } \textstyle\sum_1^m p_{i_j}^0 > \alpha\ ,\ p_{i_j} = p_{i_j}^0 / \sum_1^m p_{i_j}^0, j \le m; p_{i_j} = 0, j > m\ \}.$$

4) SCRM is the spectral coherent risk measure [Acerbi, 2002]. As it shown in [Kirilyuk, 2006],

$$SCRM(x) = \sum_1^m \lambda_j CVaR_{\alpha_j}(x)$$

therefore,

$$P_{SCRM} = \{p = (p_1,\ldots,p_n): p_i \le \left(\sum_1^m (\lambda_j / \alpha_j)\right) p_i^0,\ p_i \ge 0,\ i=1,\ldots,n, \sum_1^n p_i = 1\},$$

where $p_0 = \left(p_1^0,\ldots,p_n^0\right)$ is the vector of initial scenario probabilities.

As it shown in [Kirilyuk, 2004], the following measures are PCRM as well:

5) δ_S (x; r) is a measure on the semideviation from the mean value of a r.v. [Ogryczak and Ruszczynski, 1999], where

$$\delta_S(x; r) = -E[x] + r\, E[(E[x] - x)^+];$$

6) δ_A (x; r) is a measure on the absolute deviation from the mean value of a r.v. [Ogryczak and Ruszczynski, 1999], where

$$\delta_A(x; r) = -E[x] + r\, E[\ |x - E[x]|\ ].$$

These measures can be presented in form of relations of (1)–(2) as well.

Besides, the PCRM class is invariant relatively to the following operations [Kirilyuk, 2004]: 1) convex combination; 2) maximum functions; 3) infimal convolution.

Thus, it is possible to generate new such risk measures by application of the specified operations to representatives of the class, and it is easy to describe the obtained measures, having specified appropriate sets P from (1). It means that the class PCRM is wide enough and includes all (known for authors) coherent risk measures.

Now in PCRM terms we consider models of support of portfolio optimal decisions on efficiency-risk (return-risk) ratio. These models are put in the following forms. Let a certain sum (set) of money (resources) which need to be distributed on tools (financial assets, branches, programs, other) which are called as portfolio components, is given. It is necessary to distribute it so that the total result obtained thus was optimal on efficiency-risk ratio. The following sections of the paper are devoted to the problem.

11.2 Models of support of optimal portfolio decisions on return-risk ratio in conditions of risk (known distributions)

Let distribution of efficiency (return) of portfolio component z_j, j=1,…, k be described by matrix H of n×k dimension where a j column describes the distribution of j-th component. A vector u = (u_1, …, u_k) which describes a portfolio structure, it is considered as a variable, where $\sum_1^k u_i = 1$, $u_i \ge 0$, i=1,…,k. It is necessary to find such portfolio structure u, which optimizes a total portfolio result on a efficiency-risk ratio. In case of known component distributions (in our problem statement it is described by matrix H of component distributions according to scenarios and by the known vector of probabilities of scenarios $p_0 = (p_1^0,...,p_n^0)$), as an efficiency indicator the average efficiency, and as a risk measure some PCRM are considered. It is clear that, the more the average efficiency and the less the portfolio risk level, the more preferable it to be for a decision maker (DM). However, these characteristics are much interconnected and, as a rule, improving one of them, simultaneously we worsen the second one. Suppose that a DM has certain knowledge about what risk levels are acceptable and what efficiency levels are desirable. He optimizes one of these criteria under constraints on another. On this way, the following two interconnected problem statements are possible.

Risk measure minimization problem under average efficiency guaranteed. Fix the lower admissible value of average efficiency $E_p[Hu]$ which the investigated portfolio should guarantee, by the value μ in the form of constraints, and minimize its risk measure $\rho(Hu)$:

$$\min_{\substack{\sum_1^n u_i=1, u_i\ge 0 \\ E_{p_0}[Hu]\ge\mu}} \rho(Hu) \tag{3}$$

Average efficiency maximization problem under risk measure constrained. Fix some level of risk measure $\rho(Hu)$ which the investigated portfolio should not exceed, by the value σ in the form of constraints, and maximize its average efficiency $E_p[Hu]$:

$$\max_{\substack{\sum_1^n u_i=1, u_i\ge 0 \\ \rho(Hu)\le\sigma}} E_{p_0}[Hu] \tag{4}$$

Let remind the results obtained in [Kirilyuk, 2004], relatively to reduction of these problems to LPP.

Theorem 1. A solution of portfolio problem (3)–(1) is the component u of a solution (v, u) of the following LPP:

$$\min_{\substack{(v,u) \\ -B^T v - Hu \le 0 \\ Au\le b \\ v\ge 0, u\ge 0}} \langle c, v \rangle \tag{5}$$

Theorem 2. A solution of portfolio problem (4), (1)–(2) is the component u of a solution (v, u) of the following LPP:

$$\max_{\substack{(v,u) \\ -B^T v - Hu \le 0 \\ \langle c,v\rangle\le\sigma \\ \sum_1^n u_i=1 \\ v\ge 0, u\ge 0}} \langle H^T p_0, u \rangle \tag{6}$$

Where

$$A = \begin{pmatrix} 11...1 \\ -1-1...-1 \\ -p_0^T H \end{pmatrix}, b = \begin{pmatrix} 1 \\ -1 \\ -\mu \end{pmatrix}, B = \begin{pmatrix} 11...1 \\ -1-1...-1 \\ B_0 \end{pmatrix}, c = \begin{pmatrix} 1 \\ -1 \\ c_0 \end{pmatrix}.$$

It is easy to see that the matrix B_0 and the vector c_0, which define some PCMR, are supplemented with technical restrictions from units and a minus of units (the first two rows) and in the form of B and c accordingly are used in LPP (5) and (6).

It is important sometimes that under the average efficiency maximization, a number of constraints on risk measures, which guarantees certain reliability of obtained solutions within the bounds of some concept of risk management, should be satisfied. For instance, it is important for a DM that certain risk measures should not be exceeded by some critical levels. In the case, problem (4) contains m constraints on risk measures (PCRM), it has the following form

$$\max_{\substack{\sum_1^n u_i=1, u_i\geq 0 \\ \rho_i(Hu)\leq\sigma_i, i=1,\ldots,m}} E[Hu] \tag{7}$$

where ρ_i (.), i =1, …, m are appropriate risk measures:

$$\rho_i(x) = \max\{E_p[-X] \,/\, B_i\, p \leq c_i,\ p \geq 0\} \tag{8}$$

Then the result formulated in the form of the following theorem takes place**.**

Theorem 3. A solution of portfolio problem (7)–(8) is the component u of a solution (v_1, v_2, …, v_m, u) of the following LPP:

$$\max_{\substack{(v_1,\ldots,v_m,u) \\ -B_1^T v_1 - Hu\leq 0 \\ <c_1, v_1>\leq\sigma_1 \\ \ldots\ldots\ldots\ldots \\ -B_m^T v_m - Hu\leq 0 \\ <c_m, v_m>\leq\sigma_m \\ \sum_1^n u_i = 1 \\ v_1\geq 0,\ldots,v_m\geq 0, u\geq 0}} < H^T p_0, u > \tag{9}$$

Let us remark that in this case, appropriate LPP has more large dimension (in proportion to the number of constraints on risk measures), than the one in the conditions of theorem 2. That is, a charge for performance of all formulated constraints is the essential increase in dimension of the problem, which should be solved.

Let us notice that the PCRM concept supposes certain generalizations, which allow operating with a wider class of risk measures, without losing thus the basic remarkable property of this class as a possibility of reduction of portfolio optimization problems on return-risk (efficiency-risk) ratio to LPP. Besides, the mathematical technique, which is used, becomes complicated not essentially, and within certain technical details only [Kirilyuk, 2006].

11.3 PCRM in conditions of partial uncertainty

Essentially other looks a situation at which a matrix H of the component distributions in accordance with scenarios remains known and a vector of scenario probabilities $p_0 = (p_1^0, \ldots, p_n^0)$ is not. As a rule, identification of probabilities of the future scenarios is essentially more a challenge, than a development of such scenarios of future events. Especially, it is important for so-called rare events [Ermoliev et al, 2000a,b]. For example, how many years should to fix flooding to guarantee, that a flooding of a certain capacity occurs only once in hundred years?

Nevertheless, it is clear that certain information on probabilities is available. Moreover, sometimes it is possible to estimate these scenario probabilities in certain bounds, or, more mathematically, in the form of an inclusion: $p_0 \in P$, where P is a polyhedron. However, at once such situation adds principal difficulties to the problem, because of it is impossible to operate even with average values (mathematical expectations) under unknown probabilities p_0.

Therefore, in [Kirilyuk, 2006; 2007] the following mathematical technique has been proposed.

Together with a risk function in the PCRM form

$$\rho(x) = \max\{E_p[-X] / p\in P\} \tag{10}$$

an efficiency (return) functional is considered as well

$$g(x) = \max\{E_p[X] / p\in P\} \tag{11}$$

where P is a convex closed polyhedral set of probability measures, which is described in the form of (2):

$$P =\{p: B\, p \le c,\ p \ge 0\}= co\{p_i: i=1,\dots,k\} \tag{12}$$

Let's notice that, generally speaking, sets P from (10) and (11) can be different (various risk measures ρ_i (.) can use different sets P_i), however they are identical in the most simple case when these sets are interpreted as an estimation on scenario probabilities in the form $p_0 \in P$.

It is easy to see that

$$\rho(x) = \max\{E_p[-X] / p\in P\} = -\min \{E_p[X] / p\in P\} \tag{13}$$

and together with (11), the average value of a r.v. X can be estimated from above and from below by the interval $[-\rho(x), g(x)]$, since

$$-\rho(x) = \min \{E_p[X] / p\in P\} \le E_p[X] \le \max\{E_p[X] / p\in P\} = g(x).$$

Hence, it is clear that in terms of these criteria under decision-making it is necessary to shift the interval $[-\rho(x), g(x)]$ to the right, increasing both its ends (if X describes "positive" outputs: efficiency, return, others). Or, in terms of efficiency and risk, it is necessary to maximize the efficiency functional g(.) and to minimize the risk functional ρ(.).

Consider now certain aspects of construction of risk measures on set P_0 that estimates a vector of scenario probabilities p_0, and the appropriate mathematical technique for it from [Kirilyuk, 2008].

Address now to examples of risk measures from the PCRM class considered in section 1. As it is easy to see, sets of probability measures P for all them from the relation (2) depends on the vector of scenario probabilities p_0. That is, the set P from (2), generally speaking, is some set-valued map (s.v.m.) on the vector p_0.

Consider now a construction, in which set of probability measures P for design of risk measures ρ (.) is described by some s.v.m. a(.) on a vector of scenario probabilities p_0 with closed convex images:

$$\rho(x) = \sup\{E_p[-X] / p\in a(p_0)\} \tag{14}$$

As it is easy to see, since the image of a(.) is the subdifferential of the function ρ(.) accurate within sign, a(.) has been named the subdifferential s.v.m. [Kirilyuk, 2008].

Consider now a construction of risk measures for the case of partial uncertainty, that is, such construction on set P_0 that estimates a vector of scenario probabilities p_0. Let now there be some CRM, given on a known vector of scenario probabilities p_0 in the form of (14).

Definition 1. The risk measure induced by the initial CRM and the uncertainty set P_0, the following function is called as

$$\rho(x; P_0) = \sup\{E_p[-X] \,/\, p \in P(P_0)\} \quad (15)$$

where

$$P(P_0) = \overline{co}\left(a(P_0)\right), \; a(P_0) = \bigcup_{p_0 \in P_0} a(p_0) \quad (16)$$

Here co means the convex hull, $\overline{M}$ means the closure of set M. Obviously, ρ (x;P_0) is a CRM through its construction.

Remind that a s.v.m. a (.) is called as convex in a range of definition dom a, if

$$a(\lambda p_1 + (1-\lambda)p_2) \supseteq \lambda a(p_1) + (1-\lambda)a(p_2) \;\forall p_1, p_2 \in \mathrm{dom}\, a \;\forall \lambda \in (0,1).$$

Proposition 1. If P_0 is convex set, and a subdifferential s.v.m. a(.) of an initial risk measure ρ(.) is a convex s.v.m., then appropriate set P has the following form

$$P(P_0) = \overline{a(P_0)} \quad (17)$$

Definition 2. We will say that s.v.m. a(.) is quasilinear in a range of definition dom a, if

$$a(\lambda p_1 + (1-\lambda)p_2) = \lambda a(p_1) + (1-\lambda)a(p_2) \;\forall p_1, p_2 \in \mathrm{dom}\, a \;\forall \lambda \in (0,1) \quad (18)$$

Proposition 2. Subdifferential s.v.m. for risk measures WCR, $CVaR_\alpha$ and SCRM are quasilinear.

From quasilinear properties of s.v.m. a(.), as it is easy to see, the following proposition follows at once.

Proposition 3. If P_0 is a polyhedral set, i.e. it is described in the form of (12), and a subdifferential s.v.m. of an initial risk measure ρ(.) is a quasilinear s.v.m., the set from (17) has the following form

$$P(P_0) = \mathrm{co}\left\{a(p_i^0), i = 1,...,k_0\right\} \quad (19)$$

Corollary 3. If in conditions of proposition 3 an initial risk measure is a PCRM, i.e. its subdifferential s.v.m. has polyhedral images in the extreme points of P_0:

$$a(p_i^0) = \mathrm{co}\left\{p_j(p_i^0), j = 1,...,m(p_i^0)\right\}, i = 1,...,k_0$$

then set from (19) has the following form

$$P(P_0) = \mathrm{co}\left\{p_j(p_i^0), j = 1,...,m(p_i^0), i = 1,...,k_0\right\}. \quad (20)$$

Corollary 4. For PCRM with a quasilinear subdifferential s.v.m. a(.), the composition operation is invariant on the class:

$$\rho(.) \equiv \rho_1 \circ \rho_2(.) \Leftrightarrow \rho(x) = \sup\{E_p[-X] : p \in a_1(a_2(p_0))\}.$$

Corollary 5. Corollaries 3 and 4 take place for WCR, $CVaR_\alpha$ and SCRM.

11.4 Models of support of optimal portfolio solution in conditions of partial uncertainty (incomplete information on scenario probabilities)

Turn now to the formulation of portfolio optimization problems on efficiency-risk ratio in terms of functionals of efficiency g(.) and risk ρ(.). Let, as before, distributions of efficiencies of portfolio components z_j, j=1,…, k be described by a matrix H, and a vector u = (u_1, …, u_k) which describes portfolio structure, be considered as a variable.

Risk measure minimization problem under average efficiency guaranteed. Begin from the minimization problem of the portfolio risk measure ρ(·) under guaranteed return values g(·)≥ g_0. The problem is formulated as follows

$$\min_{\substack{\sum u_i=1, u\geq 0 \\ g(Hu)\geq g_0}} \rho(Hu) \qquad (21)$$

Average efficiency maximization problem under risk measure constrained. Connected with the previous task, the maximization problem of the portfolio efficiency g(·) under constraints on its risk measure of risk in the form ρ (·) ≤ ρ_0 can be formulated in the following form

$$\max_{\substack{\sum u_i=1, u\geq 0 \\ \rho(Hu)\leq \rho_0}} g(Hu) \qquad (22)$$

Consider now possibilities of a reduction of portfolio optimization problems (21) and (22) to certain sequences of LPP obtained in [Kirilyuk, 2008].

Theorem 4. A solution of problem (21) is the part u of a solution (u, v) of the following problem

$$\min_{1\leq i\leq k}\{ \min_{\substack{(u,v) \\ \sum u_i=1, u\geq 0, v\geq 0 \\ -B^T v - Hu\leq 0 \\ <Hu, p_i>\geq g_0}} <c, v>\} \qquad (23)$$

where p_i, i =1, …, k are the extreme points of set P from (12) and denotations:

$$\min_{\substack{(u,v) \\ \sum u_i=1, u\geq 0, v\geq 0 \\ -B^T v - Hu\leq 0 \\ <Hu, p_i>\geq g_0}} <c, v> = +\infty$$

if constraints of the subproblem are not fulfilled.

Theorem 5. A solution of problem (22) is the part u of a solution (u, v) of the following problem

$$\max_{1\leq j\leq k}\{ \max_{\substack{(u,v) \\ \sum u_i=1, u\geq 0, v\geq 0 \\ -B^T v - Hu\leq 0 \\ <c, v>\leq \rho_0}} <Hu, p_j>\} \qquad (24)$$

where p_j, j = 1, …, l are the extreme points of set P from (12).

Remark on the case of different sets P which can be used to define functionals of risk measure ρ(.) and efficiency g(.) from (10) and (11) respectively. In theorem 4, p_i from LPP (16) designate the extreme points of set P for the risk measure ρ(.) from (10), and in the theorem 5, p_j from a problem (24) are the extreme points of set P for the efficiency functional g(.) from (11).

If there are various risk measures which keep the form of functional (10), but use various sets P_i, for example, for the purpose of more conservative behavior of a DM, etc (see the interpretation of risk measures in [Kirilyuk, 2008]), the situation can demand the efficiency functional maximization under constraints on these risk measures. The content of similar functionals in some sense can guarantee some robustness of the obtained solutions relatively to risk.

So, let there be the efficiency functional (11) and some risk measures

$$\rho i(x) = \max\{Ep[-X] / p \in Pi \} \qquad (25)$$

where

$$P_i = \{p: B_i\, p \le c_i,\ p \ge 0\} = \text{co}\,\{p^i_j,\ i = 1,\dots, s_i\},\ 1 \le i \le m, \tag{26}$$

which should be taken into account in a process of search of optimal solutions.

Consider the return functional maximization problem g(·) under constraints of m risk measures $\rho_i(\cdot)$ from (25)–(26) by values ρ_i^0 , $1 \le i \le m$ accordingly which is formulated as follows

$$\max_{\substack{\sum u_j = 1, u \ge 0 \\ \rho_1(Hu) \le \rho_1^0 \\ \dots\dots\dots \\ \rho_m(Hu) \le \rho_m^0}} g(Hu) \tag{27}$$

The following theorem takes place.

Theorem 6. A solution of problem (27)–(25) is the part u of a solution $(u, v_1, \dots, v_m)$ of the following problem

$$\max_{1 \le j \le m} \{ \max_{\substack{(u, v_1, \dots, v_m) \\ \sum u_i = 1, u \ge 0, v \ge 0 \\ -B_1^T v_1 - Hu \le 0 \\ < c_1, v_1 > \le \rho_1^0 \\ \dots\dots\dots \\ -B_m^T v_m - Hu \le 0 \\ < c_m, v_m > \le \rho_m^0 \\ v_1 \ge 0, \dots, v_m \ge 0}} < u, H^T p_j > \} \tag{28}$$

where p_j, $1 \le j \le m$ are the extreme points of set P from (12).

Various generalizations of these problem statements when, for example, various variants of efficiency functionals $g_i(\cdot)$ can be given and it is necessary to minimize risks under guaranteed values of these functionals, are possible as well. Similar problem statements were considered in [Kirilyuk, 2006].

At last, make the following two useful remarks. First, in the case, when the set P of efficiency functional g(.) is described as $P = \{p_0\}$, that is functional $g(.) = E_{p_0}[X]$ for known scenario probabilities p_0, as it is easy to see, theorems 1–3 immediately follow from theorems 4–6 respectively.

Secondly, the weighed sum of the upper and lower estimations of functionals g(.) and $-\rho(\cdot)$ can be quite a reasonable variant of the optimization criterion (so-called Gurwitz's criterion). Besides, sometimes it has sense to consider so-called convolution of criteria in a case of use of a certain set of different risk measures. Then it is easy to obtain the statements similar to theorems formulated above which reduce appropriate portfolio optimization problems to certain sequences of LPP [Kirilyuk, 2006].

It is quite transparent also, development of appropriate mathematical technique on the case of multicomponent efficiency criteria looks. Then it is natural to apply already stated results under a convolution of these criteria. According to author viewpoint, there is not any principal problem in a reformulation of these results for more difficult case of the search of weak portfolio optimums: optimization of a one of criteria under constraints on the guaranteed values of others, under ranging of criteria by their importance, etc. It is clear that the value of necessary calculations for this grow, however it is purely technical problems.

11.5 Decision making on investments allocation under catastrophic flood risks

Investments into the objects planned or located in valleys and near rivers, often appear more favorable, than investments into objects remote from rivers. It can be explained by a smaller slope of a landscape, the greater fertility of lands, availability of water as a resource, presence of infrastructure, roads, affinity to consumers and manufacturers of goods or services that imply smaller cost of construction and faster recoupment of projects. Volume of investments can take both continuous and discrete values. However, in decision-making on investments it is necessary to take into account risk of possible losses from flooding and risk of significant losses from catastrophic flooding, and so to provide mitigation measures against these risks, for example, such as diversification and insurance of the investments.

A similar problem arises before an insurance company aspiring geographically to diversify insurance contracts in areas subjected to catastrophic flooding. On one hand, the closer to water, the higher insurance tariffs, but on the other hand the greater risk of big dependent insurance claims.

Another example is an investment into structural flood mitigation measures (construction of dikes, pools, etc.) characterized by more or less certain costs and by saved different property values under possible future floods.

One of a complex methodological problems is that decisions are made today and the implementation money are spent today, but a catastrophic flood may happen as in the nearest future as not to happen at all. The problem is to compare today's expenses with losses in uncertain and possibly rather distant future. In the applied hydrology, this problem is considered for many years [USACE, 1992; 2000], [RAUFDRD, 2000]. Each decision (plan, portfolio of subdecisions, insurance coverage) under uncertainty is characterized by a spectrum (distribution) of future outcomes and the problem is to select decision corresponding to the most preferable outcome distribution. This problem setting assumes the existence of certain order (stochastic dominance) in the space of outcome distributions and has basically theoretical significance. In practice, decision making under uncertainty is made by means of some functionals (expected utility or profit functions, variance and other risk measures and etc.) on outcome distributions and thus defining corresponding order in the space of distribution. There is a number of results connecting these functionals with (first or second order) stochastic dominance [Ogryczak and Ruszczynski, 2001]. In financial theory and practice, many decision-making problems concern to the structure of financial portfolios with random return. In an aggregate form these portfolios are characterized by two criteria, mean and variance of return, and thus as a risk measure the variance is used [Markowitz, 1959]. A more natural risk measure is the down-side (risk) deviation of return from its mean value [Konno and Yamazaki, 1991]. Variance or down-side variance of a decision outcome is an adequate risk measure if random outcomes are grouped around mean outcome. However, if the decision maker concerns on large losses he has to take into account characteristics of tails of outcome distributions. In this context tail related risk measures are now used such as quantile, value at risk, conditional value at risk and others [Jorion, 1996], [Rockafellar and Uryasev, 2000, 2002], [Pflug and Romisch, 2007].

Decision making problem under uncertainty assumes modeling decisions and uncertainty structure. Decisions may include discrete and continuous components, and uncertainty may be represented either by discrete tree of possibilities (scenarios) with associated probabilities, or through random simulations. In most cases, this problem includes at least two-criteria, for example, expected outcome and some associated risk measures. Each of these criteria is a specific functional defined on outcome distribution, for instance, some expected value, conditional value at risk, etc. Remark that investment decision making problems under flood risk belong exactly to the latter case, small and frequent floods need not be taken into account, but a considerable risk is connected with catastrophic rare floods and thus with the tail of the decision outcome distribution. In this case, as a measure of risk one can take, for example, a mean value of losses from rare catastrophic floods. The arising stochastic programming problems admit a variety of modeling and solution techniques [Ruszczynski and Shapiro, 2003], in some cases they can be reduced to large scale (mixed-integer) linear programming problems [Rockafellar and Uryasev, 2000] and in case of very large (or continuum) number of scenarios adaptive Monte Carlo technique may be useful [Ermoliev et al, 2000a,b].

The goal of the project "Integrated system for hazardous flood modeling and risks reduction: case study for Tisza (Ukraine), Riony (Georgia) rivers" (2005-2007, Glushkov Institute of Cybernetics, Institute of mathematical machines and systems (Kiev, Ukraine), Tbilisi State University (Georgia) and Science and Technology Center in Ukraine (STCU)) was to develop contemporary tools to support non-structural measures for flood mitigation at mostly exposed to hazardous floods rivers Tisza (Ukraine) and Rioni (Georgia), estimating the risks for the insurance and investment allocation in the areas affected by catastrophic floods. Accordingly, the main project objectives were the following:

To develop a methodology and a prototype computer system for optimal investments allocation and optimal insurance coverage in the areas, exposed to risk of hazardous flooding, on the basis of the up-to-date technologies of stochastic risk optimization and numerical flood mapping;

To implement the developed methods and prototype software for the watersheds of Ukrainian part of Tisza basin and Rioni river basin, Georgia, providing by this way to the regional and national authorities the possibilities for mapping the hazardous floods and promotion of future investment activities and developments of insurance coverage system in Tisza and Rioni basins and, after the verification of the software and proposed methods, in other river basins of Ukraine and Georgia.

11.5.1 Structure of the modeling framework

The specifics of flood risks – unpredictable timing of flood occurrences, absence of spatially explicit information about potentially inundated areas and related losses, long-term flood re-occurrence patterns, complex dependencies between structural and financial flood defense and damage sharing measures, socio-economic heterogeneities of various agents (such as individuals, farmers governments, and insurers) – all these call for adequate model based approaches integrating socio-economic, topographic, geophysical, policy related data and knowledge for evaluation of flood prevention and reduction structural and financial measures.

Decision making under flood risks requires study of the influence of decisions on probability and propagation of floods, impacts on economies losses and on their ability to recover after floods. In turn, this requires integral (system) approach to flood modeling from their beginning, propagation, up to impact upon economic objects. At each stage of modeling one has to take into account inevitable uncertainties in data and knowledge, in particular a lack of data, uncertainty in the structure and parameters of processes and their models, stochasticity, and uncertainty of decisions outcomes. Adequate modeling and treatment of these uncertainties is a key issue for making sound decisions under risk of catastrophic flood losses.

Typically, flood management framework combines geographically explicit data on property values in the region with a stochastic flood risk scenario generator to give estimates of potential flood losses.

In the project this idea is implemented in the form of blocks (modules): (1) discharge/rainfall scenario generation block; (2) a river flow/rainfall-runoff module that assesses water flows in the region; (3) inundation module for evaluation of inundated areas; (4) property/vulnerability module that incorporates damage curves (structural and agricultural) for loss estimation; (5) decision support block.

In more details each module is described as follows.

✓ Discharge/rainfall scenario generation block

There exists three basic ways to generate stochastic input data for flood modeling [Blokhinov, 1974], [Kuchment and Gelfan, 1993; 2002], [Cameron et al, 1999]:

- sampling from historical data;
- discharge generation by means of maximal monthly/decade discharge distribution and standard hydrograph;
- runoff generation by means of stochastic extreme weather (rainfall) conditions simulation and their transformation into discharges by rainfall-runoff model.

In the project, all three options are utilized.

In particular, Institute of Mathematical Machines and Systems of the National Academy of Sciences of Ukraine procured observation data on catastrophic flood occurrences in Zacarpattye in 1998 and 2001 on Tisza river and its tributaries. These data have a form of time series on water level at a certain water gauge station. By means of specific for each site "stage-discharge" curves, these data are transformed into a hydrograph (discharge as a function of time). Since there are a rather detailed observation data on 1998 and 2001 floods in Zacarpattye these floods serve as a reliable examples for river flow model identification and input generation. Input hydrographs of different probability were simulated by scaling 1998 Tisza flood by means of maximal site discharge distribution. Remark that it is rather difficult adequately to estimate the particular floods of 1998 and 2001.

Georgian school of stochastic hydrology [Svanidze, 1977], [Grigolia, 1994] developed original methods for hydrological time series generation and for maximal discharge distributions. In particular, it was proposed to use in hydrological modeling four parametric Jonson’s distribution with a bounded support. This method was tested on more than 200 world largest rivers and was recommended for application by "International handbook on basic hydrological characteristics calculation" (Leningrad, 1984; Paris, 1987). In the project, this methodology was utilized by Tbilisi state University for calculation of the maximal monthly discharges of different probabilities for Rioni river.

For flash floods modeling, it is possible to use maximal rainfall distributions with an appropriate within month allocation for stochastic input generation for rain-fall-runoff model. In the project, this approach was used at a certain site in a Tisza river valley.

The scenario generation block includes also the description of possible events leading to destruction of hydrological constructions. Each such event forms a separate flood scenario.

Each flood scenario generation method contains a large number of different uncertainties starting from incompleteness of our knowledge and models to errors in parameters and random factors estimates. So an important problem is the evaluation of the robustness of modeling conclusions with respect to these uncertainties. An example of uncertainty contemporary treatment is given in US Army Corps of Engs flood damage analysis system HEC-FDA [HEC-FDA, 1997].

✓ River flow block

In the river, flow module two physical processes are considered: rainfall-runoff due to precipitation and water flow in the river (channels) and the river dynamics process.

✓ Rainfall-runoff submodel

Simulation of a rainfall-runoff process is made based on the TOPKAPI model methodology. TOPKAPI belongs to the class of the so-called physically-based rainfall-runoff models and was developed by Prof. Todini in mid 1990th [Todini, 1995], [Ciarapica and Todini, 2002]. The model is based on the idea of combining the kinematic approach with the topography of the basin described by means of a lattice of square cells, generally increasing in size with the scale of the problem, over which the model equations are integrated. Each cell represents a computational node for the physical characteristics of the model, namely the mass balance and the momentum balance. The flow paths and slopes are evaluated from the DEM, according to a neighborhood relationship based on the principle of minimum energy, namely the maximum elevation difference which takes into account the links between the active cell and the eight surrounding cells connected along the edges or vertices; the active cell is assumed to be connected downstream with a sole cell. At present, the model version developed by UCEWP is structured around three modules, which represent the soil component, the overland flow component and flow through the drainage network respectively. The present version of the TOPKAPI model does not account for water percolation towards the deeper soil layers and for their contribution to the discharges; this will be introduced as an additional model layer in the future.

In the project rainfall-runoff model TOPKAPI-IPMMS (version by Institute of Mathematical Machines and Systems Problems, [Kivva and Zheleznyak, 2005]) was calibrated for the watersheds of Tisza river and its tributary Uzh. As input for the model precipitation time series registered at the closest weather stations were used.

✓ River dynamics submodel

A river dynamics submodel is developed to perform calculation of the flow for the specific river. Water flow in open channels is simulated using unsteady flow 1-D model within MIKE-11 software package developed by DHI Water & Environment, Denmark [Mike-11, 2004]. The basis objects of river network are

- a branch – simplest part of a river channel;
- a computational gridpoint – an element of a grid, in which flow related variables are calculated for every computational time step;
- a node or a junction – a connector of branches into more complex river network.

Therefore, each branch contains several computational gridpoints, and several branches are connected with nodes into network.

The mathematical model of a branch is based on Saint-Venant system of partial differential equations of 1-D flow mass and momentum conservation, which is solved numerically for every computational gridpoint on a branch. Additionally, in every node the mass balance and level balance equations are written. In a perfect case the information on river network structure and the cross-sectional profiles, representing water capacity of a gridpoint, is extracted from a detailed DEM, but in many cases, it is supplemented with manual river measurements. Additional information on structures along the river channel is introduced. The module transforms dynamics of input discharges into the flow dynamics using a representation of conservational laws. Due to this it is possible to get water levels and discharges along the river for the whole period of modeling.

✓ Inundation block

The calculated by the RIVER module water levels combined with DEM information are further used by the inundation module. For the calculated flood event dynamics, it is possible to reconstruct two types of maps:

- inundation maps that show the depth of standing water. These can relate either to a certain time or to the maximum level throughout the flood event;
- duration maps that represents for how long the water is standing on a floodplain.

For example, the module calculates inundation zones, in which the inundation level was 0-2 meters, 2-4 meters and more than 4 meters. Duration maps show zones, which were covered by water for less than 12 hours, 12-24 hours, 24-48 hours, and more than 48 hours. The maps can be generated to represent dynamics of inundation, say every 3 hours of flood event, but also the maximum inundated area can be estimated. Combination of inundation and duration maps gives time-depth-area detailization, which is used directly in the VULNERABILITY module for estimation of losses caused by a flood.

The module can be utilized within a Monte Carlo framework, giving on the output inundation/duration maps corresponding to flood events of different return period, for example 50-year flood, 100-year flood, 200-year flood, 500-year flood etc. All the computations are made within MikeGIS package (DHI software), built as a project in ArcView 3.x GIS software.

✓ Vulnerability block

Combination of inundation and duration maps with so-called vulnerability curves gives immediately estimation of losses. This task is performed by property/vulnerability module. The Vulnerability block produces estimates of losses for a given pattern of flood. Potential structural damages or losses associated with flood can be calculated in relative values for each type of the building in the region. Here losses are described as a certain decrease in percentage of the whole property value. This approach is especially applicable for the cases where detailed spatial information on property distribution is not available now, but can be obtained later. Vulnerability block utilizes vulnerability curves (Depth-Damage functions) – functions that represent losses depending on severity of catastrophe event and the type of structure. This can be agricultural losses depending on the inundation time, the crop and the time of the year; losses in building, depending on the depth and duration of flood as well as building material (wood, concrete, brick etc.). Usually, vulnerability curves are derived from historical observations, and are available for the modeling purposes. There are a number of works on the utilization of the DD functions; see [USACE, 1992; 2000], [Merz et al. 2004]. A nice discussion on the state-of-the-art of the problem is presented in [Messner and Meyer, 2005]. In the most papers, duration factor is not taken into account, there are only several approaches that consider a change of DD curve with increase of flood duration [Penning-Rowsell and Chatterton, 1977], [Penning-Rowsell et al, 2003]. In some cases, there is no detailed GIS information on structures' types within the modeling region. In this case, loss estimation can be done in relative or percentage terms. For example, in a certain sub-area of the region due the losses for wooden houses to flood event would be 50%, for brick houses – 40%, and for concrete houses with pillars – 10%. Once the GIS distribution of house types becomes available, the produced relative losses can be easily converted into absolute ones.

One of the direct applications of the Vulnerability block is estimation of effectiveness of a flood mitigation structural measure in terms of reduction of associated losses in "what-if" scenarios. From the other hand, as the Integrated System naturally supports Monte Carlo approach, the losses probability distribution information becomes available in terms of mean estimates, histograms etc.

Furthermore, once potential flood losses can be described in financial terms different financial oriented applications can be implemented, for example, investment allocation, insurance coverage planning, catastrophe fund design etc. The broad number of applications appears starting with estimation of the optimal premium for the fund, estimation of the current financial policies etc.

Options of input, editing and maintaining catalogue of the so-called "depth-damage relationships" were implemented in HEC-FDA (1997) system.

In the project own "depth-damage relationships" database was developed based on Excel and damage calculation module.

✓ Decision support block

Decision support block integrates information on the goals and constraints of agents that are involved in catastrophe management and may potentially suffer, share or mitigate the losses. These are households, farmers, local and central governments, flood defense offices, city planners, insurers, investors, financial markets, etc. The methodology has been already tested for the analysis of flood and seismic risks in Italy, Hungary, Poland, Russia, Japan, Ukraine [Baranov, 1999], [Galambos et al, 2000], [Amendola et al, 2000], [Ermoliev et al, 2000a,b], [Ermolieva et al, 2003], [Ermolieva and Ermoliev, 2005], [Linnerooth-Bayer and Amendola, 2003]. For example, procedures for flood reduction plans evaluation are implemented in US Army Corps of Engs HEC-FDA system [HEC-FDA, 1997].

In the project a special subsystem for investment/financial/insurance decision support was developed. The subsystem was designed to plan and evaluate structural and non-structural actions against flood damages to take into account a complex interaction between flood scenarios, river topography and flood countermeasures.

✓ Model based flood damage estimation

Within the proposed framework, we estimate potential flood-induced damages and explore feasible financial mechanisms to share the losses between the stakeholders. Implementation of the Vulnerability module is directed to estimate potential losses due to a flood. Once inundation maps are constructed, they can be directly combined with the Depth-Damage Functions (DD functions, DD curves) showing dependence of the structural damages caused by flooding. DD functions are derived by statistical analysis of losses measured in the past. The curves take into account material of the structure, depth of the flood and implicitly include the temporal flood pattern for the region in which they were estimated, see [USACE, 1992; 2000], [Merz et al, 2004].

Potential structural damages or losses associated with flood can be calculated in relative values for each type of the building in the region. Here losses are described as a certain decrease in percentage of the whole property value. This approach is especially applicable for the cases where detailed spatial information on property distribution is not available now, but can be obtained later.

One of the direct applications of the Vulnerability module is estimation of effectiveness of a flood mitigation structural measure in terms of reduction of associated losses in "what-if" scenarios. From the other hand, as the Framework naturally supports Monte Carlo approach, the losses probability distribution information becomes available in terms of mean estimates, histograms etc.

Furthermore, once potential flood losses can be described in financial terms different financial oriented Framework extensions can be implemented, for example, Insurance module, Catastrophe Fund module etc. The broad number of applications appears starting with estimation of the optimal premium for the Fund, estimation of the current financial policies etc.

Monte Carlo simulation of the Framework makes possible to move from "what-if" scenarios (passive changes) to optimization of financial instruments of flood regulation (active changes).

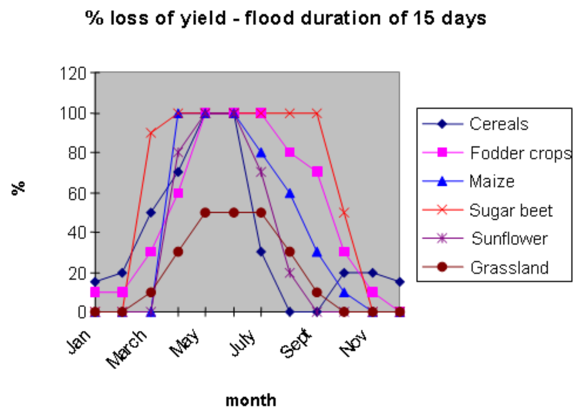


Figure 168. Functions "moment of flooding – yield losses"

✓ Catalogue of functions "flood moment – yield losses"

Losses for agricultural crops are determined basically by moment and duration of flood and are calculated by means of empirical functions "moment and duration of flood – percentage losses for yield". A number of such functions for fixed flood duration are presented on Figure 168.

✓ Catalogue of functions "depth – damage to structure"

For calculation of flood losses functions "depth – percentage losses to a structure" are used. A number of such functions is presented on Figure 169.

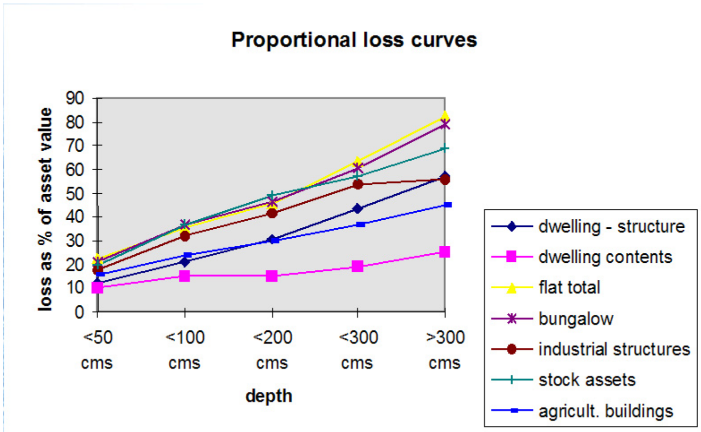


Figure 169. Graphs of percent property losses dependent on structure inundation depth

✓ Methodology for optimal investment allocation

Optimal investment allocation problem is set as follows: choose an investment plan (portfolio) $u=(u_1,\ldots,u_n)$, where component u_i designate the volumes of investments in i-th object, in such a way that the obtained investment portfolio is optimal with respect to return – risk criteria.

Let for a given river, a set (ω_s, p_s), $s=1,\ldots,S$, of flood scenarios has been developed, where ω_s is the description of scenario s, for instance as a hydrograph (an input for modeling), p_s is the probability of scenario s. Assume that each scenario is mapped into flood zone indexed s. According to flood zones, and classifications of investment objects, percent damages to every object i can be found for every scenario s. Let every object i is described by investment return θ_i and potential percent damage $l_i(\omega_s)$ for every scenario s: $i = 1,...,n$, $s = 1,...,S$. Thus, return distribution of every object i looks like: $(\theta_i(\omega_s) - l_i(\omega_s), p_s), s = 1,...,S$, and return distribution of total portfolio is described as:

$$x(u,\omega_s) = \sum_{i=1}^{n}(\theta_i(\omega_s) - l_i(\omega_s))u_i \quad (29)$$

where variable ω_s describes realizations of corresponding scenario s with probability p_s, $s = 1,...,S$.

Under insurance the return function $x(u,\omega)$ of insured portfolio looks like

$$X(u,v,\omega_s) = x(u,\omega_s) + g(u,v,\omega_s) \quad (30)$$

where

$$g(u,v,\omega_s) = -\sum_{i=1}^{n}\pi_i(v_i) + \sum_{i=1}^{n}\min\{l_i(\omega_s)u_i, v_i\} \quad (31)$$

and $\pi_i(v_i)$ designates the insurance premium for compensation of damages with a maximum compensation level v_i.

Then the following optimization problems can be considered:

1) minimization of a risk measure under a guaranteed level of the return expectation;

2) maximization of expected return under constrains on a risk measure level, where as risk measure it is proposed to use the polyhedral coherent risk measures (PCRM).

Another possible setting [Norkin, 2006] is to make such decisions that give maximal outcome in normal conditions (no catastrophe scenario) and limit losses in abnormal scenarios (catastrophes of different severity).

The class of polyhedral coherent risk measures (PCRM) contains the following risk measures (see [Kirilyuk, 2004a,b; 2008]): 1) risk of an inexact estimation of scenario probability; 2) worst case risk (WCR); 3) conditional loss expectation on α-tail distribution ($CVaR_\alpha$); 4) worst conditional expectation (WCE_α); 5) spectral coherent risk measure (SCRM); 6) measure, based on semi-deviation (absolute deviation) on expected return etc.

Portfolio optimization problems on return-risk ratio **1**) and **2**) are described as

$$F(u,v) = \max_{Bp\le c,\, p\ge 0;\, \sum_{s=1}^{S} p_s x(u,v,\omega_s)\ge\mu;\, \sum_{i=1}^{n} u_i=\bar{u}.} \left\{-\sum_{s=1}^{S} x(u,v,\omega_s)p_s\right\} \to \min_{u\ge 0, v\ge 0};$$

$$G(u,v) = \sum_{s=1}^{S} p_s x(u,v,\omega_s) \to \max_{u\ge 0, v\ge 0},$$

$$\max_{Bp\le c,\, p\ge 0}\left\{-\sum_{s=1}^{S} p_s x(u,v,\omega_s)\right\} \le \sigma,$$

$$\sum_{i=1}^{n} u_i = \bar{u},\ u_i \ge 0.$$

Here, depending on problem setting (with or without insurance) instead of $x(\cdot)$ either formula (29), or (30)-(31) are used, and matrix B and vector c describe the chosen risk measure.

Concerning mathematical methods**,** these problems were reduced to corresponding linear programming (LP) problems [Kirilyuk, 2004a,b, 2008]. The obtained result allows solving very large problems by standard LP technique by means of available software.

Besides, optimal insurance coverage models (without and with reinsurance) were proposed. The difference with respect to optimal investment allocation and optimal insurance coverage problems consists in constrains on portfolio variables: $u_i \geq 0, \sum_{i=1}^{n} u_i = \overline{u},$ for the first problem, and $u_i \in [0,1]$ or $u_i \in \{0,1\}$ – for the second one. Concerning optimum insurance coverage models with reinsurance, they are not convex and require special mathematical technique.

- ✓ **Planning of investments in commercial objects and countermeasures in the areas of catastrophic flood risks**

Let index $i = 1,...,n$ marks position (including geographical location and type of the project) of possible allocation of objects of investments. With every position some cost of investments c_i, profitableness θ_i on a unit of cost of object i under normal operational conditions and relative losses l_{ij} in catastrophic scenario j are tied-up. In this setting different types of objects of investing which are located in one geographical place have different indexes i. An investor can fully invest a project, or not to build it at all. A total volume of investments is limited by C. Let us introduce a Boolean variable $x_i \in \{0,1\}$, which means that at $x_i = 1$ object i is built (invested) and at $x_i = 0$ the object is not built. Let $\{\lambda_j\}$ is an admissible rate of losses of investments under catastrophic scenario j. Then the task of making optimal decisions reads as [Norkin, 2006]:

To maximize over $x = \{x_i\}$ function

$$F(x) = \sum_i \theta_i c_i x_i$$

subject to constraints

$$G_j(x) = \sum_i l_{ij} c_i x_i \leq \lambda_j \sum_i c_i x_i \; ; \; j = 1,...,m \; ; \; \sum_i c_i x_i \leq C \; ; \; x_i \in \{0,1\}, \; i = 1,...,n \, .$$

More general mathematical model (nonlinear discrete optimization), which takes into account a possibility of countermeasures, has the following form:

To maximize over investment plans $x = \{x_i \in \{0,1\}, i = 1,...,n\}$ and countermeasures $y = \{y_s \in \{0,1\}, s = 1,...,k\}$ profitableness function in normal conditions

$$F(x,y) = \sum_i \theta_i c_i x_i - \sum_s d_s y_s$$

subject to constraints on relative losses in the case of a catastrophic event

$$\sum_i l_{ij}(y) c_i x_i + \sum_s d_s y_s \leq r_j \sum_i c_i x_i \, , \; j = 1,...,m;$$

and constraints on available resources

$$\sum_i c_i x_i + \sum_s d_s y_s \leq C \, ,$$

where θ_i is index of profitableness of object i;

c_i is cost of object i;

d_s is cost of countermeasure s;

$l_{ij}(y)$ is a coefficient of losses of investments in an object i in scenario j under condition of implementation of plan of countermeasures y;

C is a total volume of investments;

$r_j,\ j=1,...,m,$ is a rate of relative risk.

✓ Algorithm of the problem solution

1. For the plan of countermeasures y and every catastrophic scenario j to model a corresponding flood and to build inundation maps.
2. For every potential object i, plan of countermeasures y and scenario j to calculate inundation levels and coefficients of losses $l_{ij}(y)$.
3. For every feasible plan of countermeasures y to find the optimum plan of investments $x(y)$ and a corresponding profitableness $F(x(y),y)$.
4. To find the plan of countermeasures y^*, which corresponds to the maximal profitableness $F(x(y^*),y^*)=\max_{y:\sum_s d_s y_s} F(x(y),y)$.

✓ A continuous investments allocation task in the risky agricultural areas

Planning of agricultural production in the areas of risky agriculture (back-waters of the rivers, non-irrigated droughty territories, mountain slopes) is an actual task. Thus there always is more costly alternative to develop a production in the protected or irrigated territories. A problem consists in the choice of sound compromise between a costly and reliable technology and cheap, but risky one. Let there is only one type of investments, for example, in sowing of agricultural culture of certain kind. Let x_i designate sowing area in region i, c_i is a cost of growing of the culture on a unit square in region i, θ_i is the productivity of the culture in region i at normal conditions, l_{ij} is an expertly determined coefficient of losses of sowing areas in region i at catastrophic scenario j, m_{ij} is an experimental coefficient of losses of the productivity in region i at catastrophic scenario j, b_i is the maximal sown area in region i, λ_j are admissible relative losses of sowing areas for scenario j, μ_j are admissible relative losses of the total harvest for scenario j, I is a total volume of investments. Then the task of investments allocation reads as follows [Norkin, 2006]:

To maximize over $x=\{x_i\}$ a function

$$F(x)=\sum_i \theta_i x_i$$

Subject to constraints

$F_j(x)=\sum_i m_{ij}\theta_i x_i \le \mu_j \sum_i \theta_i x_i,\ j=1,...,m;$

$G_j(x)=\sum_i l_{ij}x_i \le \lambda_j \sum_i x_i,\ j=1,...,m;$

$\sum_i c_i x_i \le I,\ 0\le x_i \le b_i,\ i=1,...,n.$

✓ Software for decision making under catastrophic flood risks

Besides hydrological modeling, the proposed modeling framework allows estimating potential flood damages for regional economies and private investors; it can help to select investment allocation plan, insurance and reinsurance arrangements in a catastrophic flood risk zone. For this, flood damage models and flood loss mitigation (investment diversification, insurance, and reinsurance) models are developed for study regions. In particular, V.M.Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine developed a catalogue of "depth-damage" functions, and a software Decision Support subSystem "Catastrophic Flood Risk Manager" (DSS CFRM) for investment/insurance decision support in a flood prone area. The developed software supports data preparation and solution of a number of tasks related to optimal investment allocation (investment portfolio), insurance coverage (insurance portfolio) planning and counter measures selection by calculation and analysis of the risk-return relationships. The inputs to the subSystem are data on potential commercial objects/projects and their inundation levels of different probabilities obtained because of flood scenarios generation and simulation by means of hydrological models of the modeling framework. The outputs of the subsystem are dependences of decision outcomes and measure of risk. In the decision support subsystem contemporary approaches to decision making under catastrophic risks are implemented, which numerically are reduced to solution of a number of large scale linear and mixed integer nonlinear programming problems. The decision support subsystem operates as a standalone MS Windows application with help, diagnostics, and graphics facilities; data for the subsystem are prepared or imported through Excel, output results are presented in a graphical form in output windows and are also put into Excel tables.

For solution of such problems within the modeling framework, the following preparation steps are fulfilled:

By means of the scenario generation block a number of discharge scenarios of different annual exceedance probabilities (p = 0.5, 0.2, 0.1, 0.04, 0.02, 0.01, 0.004, and 0.002) are formed;

By means of the river flow block for each flood scenario and a structural flood counter measure (if such measures are potentially planned), flood wave propagation is simulated and maximal water stage profile along a given river section is obtained;

By means of the inundation block water stage profiles are transformed into inundation maps of different annual probabilities for the site of interest;

By means of geo-information system ArcView the site inundation maps are viewed, a set of places of existing or potential structures is indicated and structure inundation levels corresponding to different flood scenarios are calculated;

By means of Excel the existing or potential structure inventory is updated with structure information (structure occupancy type and attributes);

By means of the vulnerability block (bank of depth-damage curves) and structure characteristics percent damage to structure value and to its content is calculated.

As a result, coefficients of the value loss for each structure, each flood scenario, and each counter flood measure are obtained. This information is an input data for setting and solution of a variety of decision-making problems under catastrophic flood risks.

For the analysis of decision making problems a special subsystem "CATASTROPHIC FLOOD RISKS MANAGER" was developed.

Decision Support System CATASTROPHIC FLOOD RISKS MANAGER (DSS CFRM) is a standalone computer system designated for decision support under catastrophic flood risks. Decision making concerns careful resources allocation and countermeasures planning accounting for trade-offs between costs, benefits and possibility of catastrophic losses. The subsystem implements decision-making methodology, described in [Norkin, 2006, 2007], [Ermoliev et al, 2000, 2001], [Kirilyuk, 2003-2008].

DSS CRM (Version 1.0) supports solution of the following tasks:

1. Selection of objects in a flood prone area for insuring (and levels of insurance) to get maximum total premium under reasonable exposure to flood claims risks.
2. Selection of a reinsurance level and objects in a flood plain for insuring to get maximum total revenue accounting for high reinsurance tariffs and flood claims risks.
3. Selection of potential projects in a flood plain to invest to get maximum total revenue under reasonable exposure to flood damage risks.
4. Selection of flood mitigation measures and potential projects in a flood plain to finance to get maximum total revenue under reasonable exposure to catastrophic flood damage risks.
5. Minimization of a certain (coherent) portfolio risk measure subject to a bound (from below) on the portfolio means revenue.
6. Maximization of a financial portfolio expected return subject to a bound (from above) on a certain (polyhedral coherent) portfolio risk measure and guaranteed mean revenue.
7. Maximization of a financial portfolio expected return subject to several constraints (from above) on certain (polyhedral coherent) portfolio risk measures.

The class of polyhedral coherent risk measures (PCRM)contains the following risk measures (see [Kirilyuk, 2004a,b, 2008]: 1) risk of an inexact estimation of scenario probability; 2) worst case risk; 3) conditional loss expectation on α-tail distribution; 4) worst conditional expectation; 5) spectral coherent risk measure; 6) measure, based on semi-deviation (absolute deviation) from expected return etc.

Concerning solution technique**,** these problems were reduced to mixed integer linear and nonlinear programming problems.

Tasks data are prepared and edited in Excel, results of tasks solution are put in Excel files and also in output windows. The data describe characteristics of possible decisions, normal and catastrophic scenarios, volumes of presently available resources, and acceptable levels of their deficits, parameters of risk measures. Some (loss) data are calculated within the system based on objects/projects data, scenarios hitting (inundation) levels and catalogue of "hitting factor-damage functions" also are accessible for analysis in Excel.

The results represent dependences of decisions outcomes as functions of some risk parameter. As a risk parameter can serve a level of reserve deficit, level of reinsurance, value of some specific risk measure.

Interactive algorithm of solution of any task consists of the following steps:

- selection of a task;
- data/example/template selection and preparation/editing in Excel;
- attempts to solve the task and correction of (reaction on) errors in data;
- visual analysis of results in monitor windows and through Excel charts;

– change in data/parameters and repeated solutions.

The look of the decision support system interface is presented on Figure 170.

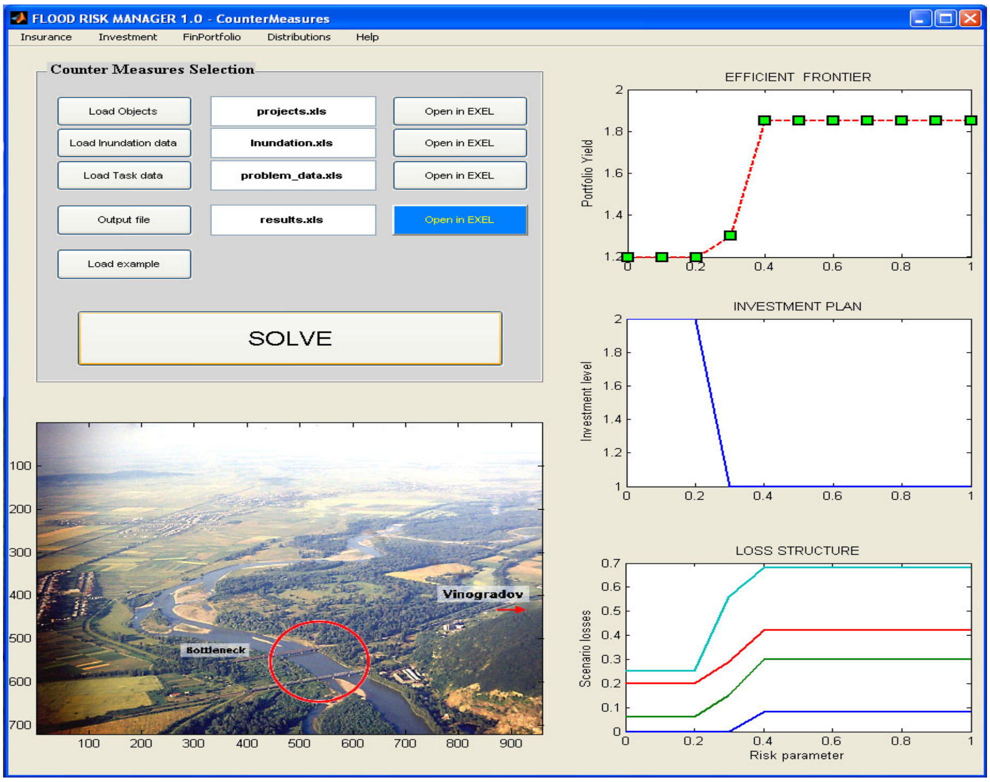


Figure 170. The look of the decision support system interface

12 Techniques for Robust Bayesian Estimation

12.1 Introduction

The sharp increase in global demand for energy caused by the emergence of such major industrial economies as China and India accelerated approaching of time when sources of oil and natural gas will be exhausted. The alternative to those fuels are nuclear energy, coal, and renewable sources. Although coal is not a limiting source of energy today, it is highly polluting. Renewal energy sources are "clear" types of energy, but they are still largely undeveloped. Therefore, only nuclear energy can be a real alternative to fossil fuels today.

Unfortunately, the Three Mile Island and Chernobyl accidents caused the adverse public opinion that has practically brought nuclear programs in many countries to a standstill. On the other hand, these accidents caused an upheaval in nuclear safety analyses. Until the accident at Three Mile Island in 1979, the USA Atomic Energy Commission (now the Nuclear Regulatory Commission (NRC)), has generally regulated Nuclear Power Plants (NPP) based on deterministic approaches, which consider a set of challenges to safety and determine how those challenges should be mitigated.

The Three Mile Island accident has clear demonstrated how important it is to quantify both the probability and the consequences of an accident. These are the key quantities that should be a basis in regulatory and safety decisions. Besides, in the early 1960s F. R. Farmer introduced another fruitful idea that predetermined future investigations in nuclear safety analyses, namely the idea of reactor safety based on the reliability of consequence-limiting equipment [NUREG/CR-2300, 1983]. These ideas substantially changed the character of the nuclear safety analyses resulting in a systematic and comprehensive methodology to evaluate Severe Accident Risks known as Probabilistic Risk Assessment (PRA).

First, PRA was systematically applied in 1975 for the study of core meltdown accidents in two commercial nuclear power plants. Results of this study were reported in the Reactor Safety Study (RSS) [WASH-1400, 1975]. This study identified a collection of accident sequences potentially capable of resulting in significant releases of radioactive material from a NPP, estimated the radiological consequences of these events, and the probability of their occurrence, using a fault tree/event tree approach. This study indicated that the probabilities of such accidents were higher than previously believed but that the offsite consequences were significantly lower. The study was peer-reviewed by the "Lewis Committee" in 1977, which approved the methodology, but cautioned that the risk Figures were subject to large uncertainty. That led to an NRC policy statement on the use of PRA [WASH-1400, 1979].

To support increasing application of PRA techniques within the industry and the regulatory process, the PRA Procedures Guide [NUREG/CR-2300, 1983] was developed. This document was intended to provide an overview of the risk-assessment field and to identify acceptable techniques for the systematic assessment of the risk from nuclear power plants. This guidebook was very useful to support PRAs that were performed in the 1980 and 1990 and in fact, much of it is still relevant today.

In 1988, the NRC requested information on the assessment of severe accident vulnerabilities by each licensed nuclear power plant [NUREG-1150, 1990]. The assessment could be done with either PRA or other approved means. In response, virtually all licensees indicated that they intended to perform PRAs in their assessments. Results of such assessments in five commercial nuclear power plants of different design was summarizes in NUREG-1150 [NUREG-1150, 1990]. This document provides improved PRA techniques and perspectives on how the results may be used by the NRC staff in carrying out its safety and regulatory responsibilities. To this day, the NUREG-1150 PRAs represents the NRC's most outstanding contribution to the state-of-the-art of PRA.

Probabilistic risk assessment involves developing a set of possible accident sequences and determining their outcomes. Accident sequences are sequences of events, which challenge the safety of the NPP. These sequences are launched by initiating events, and are modeled with event trees, and fault trees (system models). After occurrence of an initiating event, some standby safety systems must be demanded, and other, normally operating, systems must continue operating to prevent or mitigate serious undesirable consequences. Event trees depict initiating events and combinations of systems successes and failures. Fault trees depict ways in which the system failures represented in the event tree can occur.

To provide a quantitative assessment of the risk from accidents in NPPs, accident sequences must be quantified. In order to quantify the frequencies of the accident sequences delineated in the event trees, failure rates are assigned to each system model and frequencies are assigned to each initiating event. Combining the appropriate system success and failure models with each class of initiating events yields a logical representation of each accident sequence.

For a system to fail to perform its mission, several components must either fail or be unavailable. The logic events in system models that represent these failures or modes of unavailability are called basic events. Initiating events and basic events are modeled as resulting from random processes. These models have one or more parameters that should be estimated using available data. In this paper we consider only probability models based on exponential distribution. Detailed description of the standard probability models for each event and statistical methods for estimating parameters of these models can be found in the handbook NUREG/CR-6823 [NUREG/CR-6823, 2003].

Initiating events are commonly modeled with Poisson process defined by a rate of frequency λ. This model is appropriate, if simultaneous (common-cause) events do not occur. Traditionally, PRA includes to the set of initiating events such events that directly affect NPP safety, for instance, unplanned reactor trips, plant-centered loss-of-offsite power, and through-wall leak. The handbook [NUREG/CR-6823, 2003] includes also some failures occurred in a continuously running systems (such as sensor/transmitter failures) to the set of initiating events, following the viewpoint of probability modeling, in which the important fact is not the consequence of the events, but the way that they occur randomly in time. Failures occurred in a continuously running system are quickly revealed and removed. In contrast to this type of failures, failures to run during mission are failures of components or systems that do not run continuously. They are demanded to start when initiating event occur, and then they should run during some mission time. If they fail during the mission, they are nonrepairable, that is, they cannot be repaired or replaced quickly.

In PRA, components failures in both standby and continuously running systems are modeled by exponential distribution with the single parameter λ. This distribution is inherently associated with the Poisson process. Namely, the time to first failure and the times between successive failures follow an exponential distribution and the number of failures in a fixed time interval follows a Poisson distribution.

In order to use these models for quantification of accident sequences, the reliability parameters should be estimated. These estimates are propagated through logical relations to produce an estimated frequency of the undesirable end state, such as core damage. Uncertainties in the parameter estimates must be also quantified and propagated through the PRA models to obtain the uncertainty in the final estimate of risk.

Statistical techniques that are usually used for estimation of reliability parameters are based on two main approaches: the classical sampling theory methods and the Bayesian approach. The sampling theory methods are the preferred methods, if sufficient plant-specific data are available. In this case, risk estimates derived from this information reflect the actual plant experience. However, sampling theory methods are inappropriate for treating scarce data samples, which are usually available when reliability parameters of highly reliable equipment are estimated. For some nuclear power plant components, no failures were registered during many reactor-years of commercial operation. In these cases, the use of the sampling theory methods in estimation procedures leads to significant errors and makes it meaningless to use existing methods of probabilistic analysis.

One of the ways to overcome this shortcoming of the sampling theory is to utilize not only raw data specific for current investigation of equipment reliability but incorporate also other relevant information available from other data sources. Some possible sources of additional data are operating data in different environments; engineering judgments and personnel experience; and operating experiences with similar equipment. Several generic data sources are currently available and used throughout the nuclear PRA industry as sources of additional data. As a rule, generic data sources represent compilations of raw data, which have been collected directly from various NPPs. In generic databases, the raw data have been statistically analyzed and generic values of reliability parameters have been estimated. Uncertainties in the parameters estimates are presented in form of the four characteristics of probability distributions: the mean, the median, 5th percentile, and 95th percentile of the distributions. Unfortunately, the sampling theory methods are inappropriate for incorporating this relevant information.

The weak place of the Bayesian approach is to select a true distribution function. Incorrect selection of a prior distribution can result to significant errors and affect the confidence to results of a safety investigation in PRA.

Here we, following to Robbins [Robbins, 1951] and Berger [Berger, 1984], take into account that Bayesian procedures actually deal with a family of priors Γ instead of a single prior distribution. The uncertainty in the priors generates uncertainty in the posterior values of interest (e.g. mean, variance) described by the range of the posterior values as the prior varies in Γ. Finding lower and upper bounds for this range, we can evaluate sensitivity of the posterior values to the choice of the prior from the class Γ.

To find exact lower and upper bounds for the range of the posterior values as the prior varies in Γ one should solve optimization problems with linear-fractional functional with respect to distribution functions (priors). Detailed mathematical statements of such problems are presented in [Golodnikov et al, 2004].

Methods of optimization in class of distribution functions has a long story and begins from works of P.L.Chebyshev, who developed a basic inequality of probability theory called Chebyshev's inequality, and used the latter inequality to give a very simple and precise demonstration of the generalized law of large numbers [Enc.Britannica, 2009]. A.A.Markov developed ideas of his teacher in [Markov, 1898]. Detailed information about methods of optimization in class of distribution functions can be found in fundamental works of M. G. Krein, and A. A. Nudelman [Krein and Nudelman, 1977] and S. Karlin, and V. Studden [Karlin and Studden, 1966]. These works investigated only linear functionals with respect to distribution functions and utilized analytical tools.

The first work suggested computer based approach to optimization in class of distribution functions was the paper of Y. Ermoliev [Ermoliev, 1970]. Detailed description of idea underlying this work, namely, use of modification of the revised simplex method, was presented in [Ermoliev, 1976]. These works stimulated development of numerical methods of optimization in class of distribution functions in two directions.

The first approach considers the original problem of optimization of some functional under several constraints on other functionals as the optimization problem of infinite dimension. First, it was supposed that functionals are linear with respect to distribution functions. Using the idea of limit extremal problems [Ermoliev and Nurminski, 1973], the original problem of infinite dimension is approximated by the sequence of linear programming sub-problems, which can be solved by standard simplex methods [Golodnikov and Stoikova, 1978a], [Golodnikov, 1979b]. This approach was a basis for building algorithms for optimizing linear-fractional functionals [Golodnikov and Stoikova, 1978b], [Golodnikov, 2007] and nonlinear functionals [Golodnikov, 1982] with respect to distribution functions. Using the idea of ε- Quasigradient Method [Nurminski and Zhelikhovski, 1977], numerical method was developed for solving minimax problem in which the "inner" problem of maximization is linear-fractional functional with respect to distribution functions [Golodnikov, 1979a]. Actually, this was the first algorithm for building robust Bayesian estimates. The next works suggested methods for building robust Bayesian estimates, [Lavine, 1991], [Lavine et al, 1993] has appeared more than ten years later.

The second approach is based on ideas of duality. In the works of Y. Ermoliev, A. Gaivoronski, and C. Nedeva [Ermoliev and Nedeva, 1982], [Ermoliev et al, 1985] the original linear problem of infinite dimension is reduced to dual problem which is formulated as finite dimensional minimax problem without concavity of "inner" problem of maximization. A vast amount of work has been done on minimax problems but virtually all of the existing numerical methods fail if the inner problem is nonconcave. To overcome this difficulty the work [Ermoliev et al, 1985] adopts an approach based on stochastic optimization techniques.

In [Golodnikov et al, 2007] a new approach to selecting the Gibbs distribution in models of objects to be recognized is proposed. This approach proposes to determine the lower and upper bounds for probabilities of the object under study. The distance between these bounds may be used as a measure of error in pattern recognition problems.

This chapter analyzes statistical techniques that are usually used for estimation of reliability parameters. It compares two main approaches: the classical sampling theory methods and the Bayesian approach. It is known that sampling theory methods are inappropriate for treating scarce data samples. In contrast to sampling theory methods, the Bayesian approach allows naturally to incorporate data from various sources in reliability parameters estimates by considering each source as a sample from the same population. However, the justification of a prior distribution frequently is a practical difficulty in the application of the Bayesian approach.

In cases when small datasets of past reliability data are available, it is desirable to estimate how far the calculated Bayesian estimate is from the true Bayesian estimate. Therefore, when only partial prior information is available it is necessary to search upper and lower bounds for Bayesian estimates which can be derived for any prior distribution satisfying the given partial prior information. The chapter considers models for searching such bounds.

12.2 Methods of estimating parameters in risk models based on classical sampling theory

One of the important steps of the currently used methodology of risk assessment is estimating parameters characterizing the equipment reliability (failure rates and probability of demand-related failures). The accuracy of parameter estimates obtained at this step significantly affects the accuracy of ultimate result: risk assessment of the object.

Statistical techniques that are usually used for estimation of reliability parameters are based on two main approaches: the classical sampling theory methods and the Bayesian approach. The classical approach allows obtaining satisfactory estimates when detailed data concerning each component are available.

12.2.1 Sample methods of point estimation of reliability parameters

Let $f(x;\theta)$ be a probability density function of the random value X, where θ is an unknown parameter to be estimated. Consider method of point estimation of this parameter. Let $X_1, X_2, \ldots, X_n$ be a random sample from $f(x;\theta)$. The likelihood function for this random sample is the joint probability density function of $X_1, X_2, \ldots, X_n$:

$$L(\theta; x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i;\theta),$$

which is considered as a function of θ.

The maximum likelihood estimate of parameter θ is the value $\hat{\theta}$ such that $L(\hat{\theta}; x_1, x_2, \ldots, x_n) \geq L(\theta; x_1, x_2, \ldots, x_n)$ for every value of θ. The maximum likelihood estimate is a function of the observed random sample $x_1, x_2, \ldots, x_n$, i.e. it is random.

Method of moments equates the distribution moments $\mu'_j = E(X^j)$ to the sample moments

$$m'_j = \frac{\sum_{i=1}^{n} X_i^j}{n} \text{ for } j = 1, 2, \ldots, k.$$

Using the method one can easily obtain the estimate for parameter p of the binomial distribution, which provides model for failures occurring when demand occurs. According to this model, the probability P that x failures occur when n demands occur is calculated as follows:

$$P(x) = \frac{n!}{(n-x)!\,x!} p^x (1-p)^{n-x} \tag{1}$$

The estimator for this probability is [Martz and Waller, 1991]:

$$\hat{p} = \frac{x}{n} \tag{2}$$

12.2.2 Sample methods of interval estimation of reliability parameters

Practice of implementation of sample estimators in investigating reliability parameters shows that in many cases it is not enough to consider only point estimators. In cases when medium or small samples are available, values of estimators may vary too greatly from experiment to experiment. For this reason, they cannot be used as a stable estimator. That is why in addition to point estimators of reliability parameters one should also consider their confidence intervals as a measure of uncertainty.

Two-sided confidence interval for estimator ε of level $1-\gamma$ is a random interval $(\varepsilon_1(x), \varepsilon_2(x))$ such that

$$P\{\varepsilon_1(x) < \varepsilon < \varepsilon_2(x)\} \geq 1-\gamma .$$

Consider confidence intervals for the binomial model widely used in the probabilistic safety analysis.

Lower and upper bounds, $\varepsilon_1(x), \varepsilon_2(x)$, of two-sided confidence interval for estimator of parameter p in the binomial model is calculated as follows [Martz and Waller, 1991]:

$$\varepsilon_1(x) = \frac{x}{x + (n-x+1)F_{1-\gamma/2}(2n-2x+2, 2x)} \tag{3}$$

$$\varepsilon_2(x) = \frac{(x+1)F_{1-\gamma/2}(2x+2, 2n-2x)}{(n-x) + (x+1)F_{1-\gamma/2}(2x+2, 2n-2x)} \tag{4}$$

where $F_{1-\gamma/2}(n_1, n_2)$ is the $(1-\gamma/2)$-quantile of the $F-$distribution with (n_1, n_2) degrees of freedom.

12.2.3 Basic shortcoming of the sample methods of estimation of reliability parameters

However, sampling theory methods are inappropriate for treating scarce data samples, which are usually available when reliability parameters of highly reliable equipment are estimated. For some nuclear power plant components, no failures were registered during many reactor-years of commercial operation. In these cases, the use of the sampling theory methods in estimation procedures leads to significant errors and makes it meaningless to use existing methods of probabilistic analysis.

One of the ways to overcome this shortcoming of the sampling theory is to utilize not only raw data specific for current investigation of equipment reliability but incorporate also other relevant information available from other data sources. Some possible sources of additional data are operating data in different environments; engineering judgments and personnel experience; and operating experiences with similar equipment. Unfortunately, the sampling theory methods are inappropriate for incorporating this relevant information.

12.3 Bayesian approach to parameters estimation. Analysis and methods of investigation

In contrast to sampling theory methods, the Bayesian approach allows naturally to incorporate data from various sources in reliability parameters estimates by considering each source as a sample from the same population. While in the classical approach the parameters of underlying life distribution are assumed unknown constants to be determined, in the Bayesian approach the parameters are considered to be random values to which is assigned a prior probability density function.

The basic elements of the Bayesian approach are the sampling model $f(x_i \mid \theta)$ – the conditional probability density function of random value X_i which depends on parameter θ; the prior model $H(\theta)$ - the prior probability distribution of parameter θ with the density $h(\theta)$; the posterior model $G(\theta \mid x)$ – the posterior probability distribution of parameter θ given sample data x with the density $g(\theta \mid x)$, and the loss function $L(\theta,\delta)$ for the estimator δ.

12.3.1 Choice of a prior distribution function

The prior probability density function $h(\theta)$ accumulates the totality information available concerning the parameter θ prior to the observation of the sample data x. The prior information, which is transformed into a prior distribution, can be obtained from objective or subjective sources. Objective prior information includes operating data in different environments, observational data from experiments conducted earlier, and operating experiences with similar equipment. Subjective prior information may include the assessor's personal experiences and judgments, and design information.

In cases when no prior information concerning the parameter θ is available or such information is very limited, non-informative priors are selected. Such choice is based on the "Principle of Insufficient Reason" which was first stated by H. Jeffreys in [Jeffreys, 1961]. This principle requires the distribution on the finitely many events to be uniform unless there is some definite reason to consider one event more probable than another does.

Ordinarily, one deals with cases when only partial prior information concerning the parameter θ is available which is insufficient to specify exactly a prior distribution. In such cases, there is a possibility of incorporation of additional subjective unsuitable information during transformation of the partial prior information to a full prior.

Transformation of the partial prior information to a full prior can be carried out by using various techniques. The most popular techniques use the partial prior information to specify a particular member of a chosen family of distributions. According to this technique, a prior can be obtained by selecting a distribution family for the prior and fitting its parameters based on estimated moments or quantiles of the prior. For example, the Reactor Safety Study assesses upper and lower values of the failures rates of components and then fits a lognormal distribution by using these values as the 95-th and 5-th percentiles of the distribution [Mosleh and Apostolakis, 1982]. This example demonstrates that selection of the lognormal distribution family for the prior incorporates additional subjective information, which does not belong to the prior information, into the full prior distribution.

For transformation of the partial prior information into a full prior which is as non-informative as possible subject to partial prior information, the maximum-entropy method can be used [Jaynes, 1968], [Deeley et al, 1970], [Savchuk and Martz, 1994]. According to this method, the prior probability density function $h(\theta)$ of the parameter θ is selected such, which maximizes the Shannon-Jaynes entropy [Savchuk and Martz, 1994]:

$$I(g) = -\int h(\theta)\ln(h(\theta))d\theta \to \max \qquad (5)$$

among all those distributions which satisfy the given constraint:

$$S(h(\theta)) = 0, \qquad (6)$$

where $S(h(\theta))$ is functional constraint on $h(\theta)$, which reflects given partial prior information.

Frequently, assessor selects a prior distribution that possesses some useful properties to simplify mathematical inference of Bayesian estimates. Conjugate prior distributions are important class of such distribution functions. A conjugate prior distribution $h(\theta)$ for a given sampling distribution $f(x \mid \theta)$ is such that the posterior distribution $g(\theta \mid x)$ and the prior $h(\theta)$ are members of the same family of distributions [Martz and Waller, 1991]. For example, widely used choice of a prior distribution in the Bayesian estimating parameter p in the binomial model is the beta distribution with parameters x_0 and n_0:

$$h(p; x_0, n_0) = \frac{\Gamma(n_0)}{\Gamma(x_0)\Gamma(n_0 - x_0)} p^{x_0 - 1}(1-p)^{n_0 - x_0 - 1}, \quad n_0 > x_0 > 0. \tag{7}$$

With such prior probability density function, the posterior probability density function of p given sample data x is [Martz and Waller, 1991]:

$$g(p \mid x; x_0, n_0) = \frac{\Gamma(n + n_0)}{\Gamma(x + x_0)\Gamma(n + n_0 - x - x_0)} p^{(x + x_0) - 1}(1-p)^{(n + n_0 - x - x_0) - 1} \tag{8}$$

which coincides with the beta probability density function with parameters $x + x_0$, $n + n_0$. Thus, the beta-distribution is the conjugate prior distribution for the binomial failure model.

12.3.2 Bayesian methods of point estimation of reliability parameters

For a quadratic loss function, the Bayesian estimating procedure is to minimize the posterior risk given by

$$E\left[a(\theta - \hat{\theta})^2 \mid x\right] = \int a(\theta - \hat{\theta})^2 g(\theta \mid x) d\theta \tag{9}$$

This risk is minimized when

$$\hat{\theta} = E(\theta \mid x) = \int \theta \, g(\theta \mid x) d\theta \tag{10}$$

Thus, for a quadratic loss function, the Bayesian point estimator coincides with the posterior mean of the posterior distribution.

If in Bayesian point estimating parameter p for the binomial failure model a beta distribution (7) is used as prior probability density function, and x failures occur when n demands occur, then, the posterior probability density function of p given sample data x is the beta probability density function (8) with parameters $x + x_0$, $n + n_0$, and the Bayesian point estimator, which is the posterior mean, is [Martz and Waller, 1991]:

$$\hat{p} = E(p \mid x; x_0, n_0) = \frac{x + x_0}{n + n_0}. \tag{11}$$

12.3.3 Bayesian methods of interval estimation of reliability parameters

In Bayesian estimation, unknown reliability parameter is considered to be a random value ε to which is assigned a prior probability density function $h(\varepsilon)$. For given sample data x the posterior probability density function $g(\varepsilon \mid x)$ of ε can be obtained. In Bayesian approach, analog of classical two-sided confidence interval is a $(1-\gamma)$ level two-sided Bayes probability interval $(\varepsilon_1, \varepsilon_2)$ for ε given by

$$P\{\varepsilon_1(x) < \varepsilon < \varepsilon_2(x)\} = \int_{\varepsilon_1}^{\varepsilon_2} g(\varepsilon \mid x) d\varepsilon \geq 1-\gamma \tag{12}$$

It can be found by solving the following two equations:

$$\Pr(\varepsilon \leq \varepsilon_1 \mid x) = \int_0^{\varepsilon_1} g(\varepsilon \mid x) d\varepsilon = \frac{\gamma}{2}, \tag{13}$$

and

$$\Pr(\varepsilon \geq \varepsilon_2 \mid x) = \int_{\varepsilon_2}^{1} g(\varepsilon \mid x) d\varepsilon = \frac{\gamma}{2} \tag{14}$$

for the lower limit ε_1 and the upper limit ε_2.

Consider the binomial failure model. Assume that a prior distribution of parameter p is the beta distribution with parameters x_0 and n_0, and x failures occur when n demands occur. Then, the lower limit $\varepsilon_1(x)$ and the upper limit $\varepsilon_2(x)$ of $(1-\gamma)$ level two-sided Bayes probability interval $(\varepsilon_1, \varepsilon_2)$ for parameter p are calculated as follows [Martz and Waller, 1991]:

$$\varepsilon_1(x) = \frac{x + x_0}{x + x_0 + (n + n_0 - x - x_0)F_{1-\gamma/2}(2n + 2n_0 - 2x - 2x_0, 2x + 2x_0).} \tag{15}$$

$$\varepsilon_2(x) = \frac{(x + x_0)F_{1-\gamma/2}(2x + 2x_0, 2n + 2n_0 - 2x - 2x_0)}{n + n_0 - x - x_0 + (x + x_0)F_{1-\gamma/2}(2x + 2x_0, 2n + 2n_0 - 2x - 2x_0).} \tag{16}$$

where $F_{1-\gamma/2}(n_1, n_2)$ is the $(1-\gamma/2)$-quantile of the $F-$distribution with (n_1, n_2) degrees of freedom.

12.3.4 Methods of fitting parameters of a prior distribution function

When the prior distribution is assumed a member of a certain parametric family of distributions the problem of fitting its parameters arises. It is unlikely that expert judgment can be quantified in terms of statement, which gives directly values of parameters of the prior distribution. However, the assessors are able to give more or less accurate values of quantiles and moments of the prior distribution in quantifying their beliefs. Frequently these characteristics of the prior distribution are used to specify a particular member of a chosen family of distributions. This can be easily done when the characteristics and the parameters of the distribution are related through a simple analytic relation [Martz and Waller, 1991].

Unfortunately, there are no simple analytic relations between the quantiles and parameters of the gamma and beta distributions, which are widely used as the prior in Bayesian estimation of reliability parameters [Martz and Waller, 1991]. For this reason graphical techniques [Waller et al, 1977], [Weiler, 1965] or tables of the incomplete gamma or beta functions [Pearson, 1957], [Pearson, 1968] are used. The existing graphical techniques cover a very limited range of quantiles or the parameters. Tables for determining the beta prior are presented in [Waterman et al, 1976] for cases when 1) the mean and 5th quantile or 2) the mean and the 95th quantile are given. The paper [Mosleh and Apostolakis, 1982] presents techniques that facilitate estimation of the parameters of a gamma or beta distribution when two percentiles or one percentile and the mean value are known. The problem of determination of the parameters of a beta distribution for the same types of available prior information is also considered in [Duran and Booker, 1988].

When there is observational data from past experiments, it can be used to estimate parameters of the prior distribution, which is assumed a member of a certain parametric family of distributions. Two basic methods are usually used for this purpose: the method of matching moments and method of maximum likelihood based on the marginal distribution [Martz and Waller, 1991]. The method of moments is to equate the sample moments to their expected values and solve for the parameters. The method of maximum likelihood chooses estimators that maximize the marginal likelihood function.

Consider estimators for parameters x_0 and n_0 of a beta distribution. Assume that we have results obtained in N binomial life test experiments conducted previously with the same or similar items. Let p_j denote the unknown value of the parameter p in j th experiment. We assume that the underlying sequence $p_1, p_2, \ldots, p_N$ is statistically independent realizations of a random value, which has beta distribution with constant x_0 and n_0 values for the sequence of tests [Martz and Waller, 1991]. Let for the j th experiment x_j failures occur out of sample size n_j. The classical estimator for the parameter p in the j th experiment is $\hat{p}_j = x_j / n_j$. Then, the unweighted moment estimators for n_0, x_0 are [Martz and Waller, 1991]:

$$\begin{aligned} \hat{n}_0 &= \frac{N(\overline{p}_u - m_u^2)}{N m_u^2 - K\overline{p}_u - (N-K)\overline{p}_u^2}, \\ \hat{x}_0 &= \hat{n}_0 \overline{p}_u, \end{aligned} \tag{17}$$

where

$$\overline{p}_u = \sum_{j=1}^{N} \frac{\hat{p}_j}{N}, \qquad m_u^2 = \sum_{j=1}^{N} \frac{\hat{p}_j^2}{N}, \qquad K = \sum_{j=1}^{N} n_j^{-1}. \tag{18}$$

The marginal maximum likelihood estimates for parameters n_0, x_0 can be obtained by numerical solving the following system of equations [Martz and Waller, 1991]:

$$\begin{aligned} &\sum_{j=1}^{N} \sum_{i=0}^{x_j-1} \left(\frac{1}{x_0 + i} \right) - \sum_{j=1}^{N} \sum_{i=0}^{n_j - x_j - 1} \left(\frac{1}{n_0 - x_0 + i} \right) = 0, \\ &\sum_{j=1}^{N} \sum_{i=0}^{n_j - x_j - 1} \left(\frac{1}{n_0 - x_0 + i} \right) - \sum_{j=1}^{N} \sum_{i=0}^{x_j-1} \left(\frac{1}{x_0 + i} \right) = 0. \end{aligned} \tag{19}$$

12.3.5 Empirical Bayesian point estimation of reliability parameters

In order to avoid the necessity of identifying the unknown prior distribution empirical Bayesian estimation was suggested in [Robbins, 1955]. The difference between empirical Bayesian and ordinary Bayesian is that the former does not make explicit the form of the prior distribution in order to make possible a Bayesian solution. Instead, the empirical Bayesian method depends on existence of prior information in the form of past estimates of either parameter in question or some close variation of it [Tillman et al, 1982], [Tsocos and Canavos, 1972].

Empirical Bayesian procedures consists of 1) methods that attempt to approximate the Bayesian estimator without explicitly estimating the unknown prior distribution [Clemmer and Krutchkoff, 1968], [Lin, 1972], [Martz and Krutchkoff, 1969], [Miyawasa, 1961], [Nichols and Tsokos, 1972], and 2) methods in which the unknown prior distribution is explicitly estimated [Bennet and Martz, 1972], [Lemon and Krutchkoff, 1969], [Maritz, 1966; 1970].

One of the widely used empirical Bayesian estimators is the Copas estimator [Copas, 1972]. Consider a sequence of binomial sampling experiments in which the j th experiment results in x_j failures in n_j trials [Martz and Waller, 1991]. Let $y_j = x_j / n_j$. It is assumed that N-1 previous experiments have been conducted prior to the current or N-th experiment. For convenience let $x = x_N$, $n = n_N$, $y = x/n$ and $p = p_N$. Let μ and σ^2 denote the prior mean and variance, respectively. Then, the Copas estimator [Copas, 1972] is given by

$$\widehat{p}_C = (y + W_N \mu_{1N})/(1 + W_N), \tag{20}$$

where

$$\mu_{1N} = \sum_{i=1}^{N} y_i / N, \; W_N = [\mu_{1N}(1-\mu_{1N}) - \sigma_{1N}^2]/n\sigma_{1N}^2, \tag{21}$$

$$\sigma_{1N}^2 = \max[0, \{\sum_{i=1}^{N}(y_i - \mu_{1N})^2 - k\mu_{1N}(1-\mu_{1N})\}/(N-k)], \; k = \sum_{i=1}^{N} n_i^{-1} \tag{22}$$

Another empirical Bayesian estimator of binomial parameter is the Lemon and Krutchkoff estimator [Lemon and Krutchkoff, 1969]:

$$\widehat{p}_L = [\sum_{i=1}^{N} \widetilde{p}_i^{x+1}(1-\widetilde{p}_i)^{n-x}]/[\sum_{i=1}^{N} \widetilde{p}_i^{x}(1-\widetilde{p}_i)^{n-x}], \tag{23}$$

where $\widetilde{p}_i$ is given by

$$\widetilde{p}_i = [\sum_{j=1}^{N} y_j^{x_i+1}(1-y_j)^{n_i-x_i}]/[\sum_{j=1}^{N} y_j^{x_i}(1-y_j)^{n_i-x_i}], \quad N \geq 2, \tag{24}$$

and

$$y_j = x_j / n_j. \tag{25}$$

12.3.6 Basic shortcomings of Bayesian and empirical Bayesian estimation of reliability parameters

A major shortcoming of the Bayesian approach is related to the ability of the investigator to select a true distribution function. In cases when no prior information concerning the parameter θ is available, the more appropriate selection is a non-informative prior distribution or the distribution, which maximizes the Shannon-Jaynes entropy [Jaynes, 1968], [Deeley et al, 1970]. Difficulties of implementation of the Bayesian procedures of estimation arise when the partial prior information is available.

Most techniques of identification of a prior distribution considered above consist of two stages. At the first stage, the family of distributions is selected. At the second stage, a particular member of this family is specified by means of estimating its parameters based on the available prior information. The second stage can be considered as justified and formalized since it is based on the available prior information. In contrast to the second stage, choice of the family of distribution functions on the first stage depends on a subjective factor. At this stage, additional subjective information can be introduced into decision procedure, which may not belong to the available prior information.

This shortcoming of the Bayesian approach is usually overcome by using empirical Bayesian decision process that does not require accurately estimate a prior distribution function. Empirical Bayesian decision procedure is an efficient tool for combining existing sets of either reliability data or reliability parameter estimates from various sources. One of the advantages of such methods is their asymptotic optimality. However, the rate of convergence of the empirical Bayesian risk to the minimum Bayesian risk can be quite slow. Therefore, in cases when small datasets of reliability data or reliability parameter estimates are available, the accuracy of the approximation of the Bayesian estimator by the empirical Bayesian estimators is never really known.

Errors in Bayesian estimating reliability parameters resulted from incorrect selecting a prior distribution can be significant and affect the confidence to results of a safety investigation of ecologically dangerous objects.

12.4 Statement of the problem of robust estimating reliability parameters

One of the ways to avoid arbitrariness in selecting prior distribution based on partially available prior information is to use Γ-minimax approach that was originally proposed by Robbins [Robbins, 1951] and was investigated by Berger [Berger, 1984]. The Γ-minimax approach takes into account that Bayesian procedures actually deal with a family of priors Γ instead of a single prior distribution. Any distribution from the class Γ has the same reasons to be used as a prior distribution. The uncertainty in the priors generates uncertainty in the posterior values of interest (e.g. mean, variance) described by the range of the posterior values as the prior varies in Γ.

This range is used as a measure of robustness [Vidakovic, 2000], [Ruggeri and Sivaganesan, 2000], [Carota and Ruggeri, 1994]. A small range means that uncertainty in the priors does not heavily affect the posterior value, and it is robust with respect to the choice of the prior from the class Γ. Conversely, a large range means that the posterior value is sensitive to the choice of the prior, i.e. it is not robust. In the latter case, the statistician should pay attention to the least favorable prior in Γ and use Γ-minimax approach to develop robust estimates.

In the Γ-minimax approach, the estimate that minimizes the supremum of the cost functional over distributions in Γ is selected as a robust [Vidakovic, 2000]. In case of scarce prior information, the class Γ is large and result obtained by using this approach is close to the result in minimax approach. In other case when complete prior information is available, the class Γ contains a single prior and the Γ-minimax approach is equivalent to the ordinary Bayesian approach.

The natural choices of the cost functional are

1. Bayes risk

$$r_H(\delta(x)) = \int_{\Theta}\int_{X} L(\theta,\delta(x)) f(x \mid \theta)\, dx\, dH(\theta) \tag{26}$$

2. posterior risk

$$\phi_H(\delta(x)) = \int_\Theta L(\theta,\delta(x))dG(\theta \mid x) = \frac{\int_\Theta L(\theta,\delta) f(x \mid \theta) dH(\theta)}{\int_\Theta f(x \mid \theta) dH(\theta)} \tag{27}$$

3. posterior mean of parameter θ given sample data x

$$\widehat{\theta}_H = \frac{\int_\Theta \theta\, f(x \mid \theta) dH(\theta)}{\int_\Theta f(x \mid \theta) dH(\theta)} \tag{28}$$

From the point of view of the statistician a prior distribution, which maximizes Bayes (26) risk or posterior risk (27), is the least favorable prior in Γ since it corresponds to the worst estimation accuracy. For any other prior from class Γ the estimation accuracy can be only better.

According to conservative approach it is naturally to assume that the "true" prior always coincides with the least favorable prior in Γ. Therefore, the estimator that minimizes Bayes risk or posterior risk under the least favorable prior is robust.

The cost functional (28) corresponds to a Bayesian estimator when a squared-error loss function is used in Bayesian inference. Calculating values of Bayes estimates $\widehat{\theta}_H$ for any prior $H \in \Gamma$ we obtain the range of possible values (θ_*, θ^*). Suppose that upper bound θ^* corresponds to some distribution $H^* \in \Gamma$, i.e. $\theta^* = \widehat{\theta}_{H^*}$. From the point of view of risk theory, the conservative assumption concerning a prior is that H^* is the "true" prior. Use of any other Bayesian estimate $\widehat{\theta}_H \in (\theta_*, \theta^*)$, which corresponds to any other distribution $H \in \Gamma$, can only diminish the assessment of failure probability and thus promote obtaining more optimistic result of probabilistic safety assessment. For this reason, the upper bound θ^* of the range of possible values (θ_*, θ^*) of posterior mean is appropriate for utilization as a robust Bayesian estimate in estimating reliability parameters for probabilistic safety assessment.

If the cost functional is posterior risk then, the problem of seeking robust Bayesian estimate given sample data x is reduced to the following stochastic minimax problem:

$$\min_\delta \sup_{H \in \Gamma} \phi_H(\delta, x) = \min_\delta \sup_{H \in \Gamma} \frac{\int_\Theta L(\theta,\delta) f(x \mid \theta) dH(\theta)}{\int_\Theta f(x \mid \theta) dH(\theta)} \tag{29}$$

The objective functional in (29) is the linear-fractional functional with respect to distribution functions. The inner problem in (29) is the problem of optimization of the linear-fractional functional in the space of distribution functions that belong to class Γ.

If the cost functional is posterior mean of parameter θ given sample data x then, the problem of seeking robust Bayesian estimate is reduced to the following stochastic programming problem of maximization of linear-fractional objective functional over distribution functions from class Γ:

$$\sup_{H\in\Gamma} \widehat{\theta}_H = \sup_{H\in\Gamma} \frac{\int_\Theta \theta\, f(x\mid\theta)dH(\theta)}{\int_\Theta f(x\mid\theta)dH(\theta)} \tag{30}$$

12.5 Statement of the problem of seeking lower and upper bounds for Bayesian estimates of reliability parameters

In order to investigate sensitivity of Bayesian estimates with respect to selection of a prior distribution from class Γ it is necessary to find low and upper bounds for the range of possible values of the cost functional of interest. According to the Bayesian approach, if $H(\theta)$ is the "true" prior, the quality of the Bayesian point estimator $\widehat{\theta}_H(x)$ is measured in terms of the Bayesian risk $r(\hat{\theta}_H(x))$ (26), or the posterior risk $\phi_H(\widehat{\theta}_H(x))$ (27). Now consider the case when instead of a single prior $H(\theta)$, class Γ of such distribution functions is available. In this case the value $r_H(\widehat{\theta}_H(x))$ or $\phi_H(\widehat{\theta}_H(x))$ cannot be used for characterization of the quality of the Bayesian point estimator $\widehat{\theta}_H(x)$. For this purpose the range $(r_*(\widehat{\theta}_H(x)), r^*(\widehat{\theta}_H(x)))$ of possible values of the Bayesian risk or the range $(\phi_*(\widehat{\theta}_H(x)), \phi^*(\widehat{\theta}_H(x)))$ of possible values of the posterior risk, obtained by varying prior in Γ, are more appropriate.

In estimating reliability parameters for probabilistic safety assessment it is also of interest to know how wide is the range (θ_*, θ^*) of possible values of posterior mean of parameter θ given sample data x (Bayesian estimate), as the prior varies in Γ.

The problem of calculating low $r_*(\widehat{\theta}_H(x))$ or upper $r^*(\widehat{\theta}_H(x))$ bounds for the range $(r_*(\widehat{\theta}_H(x)), r^*(\widehat{\theta}_H(x)))$ of possible values of the Bayesian risk associated with the Bayesian point estimator $\widehat{\theta}_H(x)$ is reduced to the following optimization problems in the space of distribution functions:

$$r_*(\widehat{\theta}_H(x)) = \inf_{G\in\Gamma} \int_\Theta \int_X L(\theta, \widehat{\theta}_H(x)) f(x\mid\theta) dx dG(\theta) \tag{31}$$

and

$$r^*(\widehat{\theta}_H(x)) = \sup_{G\in\Gamma} \int_\Theta \int_X L(\theta, \widehat{\theta}_H(x)) f(x\mid\theta) dx dG(\theta) \tag{32}$$

In (31), (32) objective functionals are linear with respect to distribution functions.

The problem of calculating lower $\phi_*(\widehat{\theta}_H(x))$ or upper $\phi^*(\widehat{\theta}_H(x))$ bounds for the range $(\phi_*(\widehat{\theta}_H(x)), \phi^*(\widehat{\theta}_H(x)))$ of possible values of the posterior risk associated with the Bayesian point estimator $\widehat{\theta}_H(x)$ is reduced to the following optimization problems in the space of distribution functions:

$$\phi_*(\widehat{\theta}_H(x)) = \inf_{G\in\Gamma} \frac{\int\limits_\Theta L(\theta,\widehat{\theta}_H(x)) f(x\mid\theta)dG(\theta)}{\int\limits_\Theta f(x\mid\theta)dG(\theta)} \tag{33}$$

and

$$\phi^*(\widehat{\theta}_H(x)) = \sup_{G\in\Gamma} \frac{\int\limits_\Theta L(\theta,\widehat{\theta}_H(x)) f(x\mid\theta)dG(\theta)}{\int\limits_\Theta f(x\mid\theta)dG(\theta)} \tag{34}$$

In (33) and (34) objective functionals are linear-fractional with respect to distribution functions.

The problem of calculating lower θ_* or upper θ^* bounds of the range (θ_*, θ^*) of possible values of posterior mean of parameter θ given sample data x (Bayesian estimate), is reduced to the following minimization or maximization of linear-fractional objective functional (posterior mean) over distribution functions from class Γ :

$$\theta_* = \inf_{G\in\Gamma} \frac{\int\limits_\Theta \theta\, f(x\mid\theta)dG(\theta)}{\int\limits_\Theta f(x\mid\theta)dG(\theta)} \tag{35}$$

and

$$\theta^* = \sup_{G\in\Gamma} \frac{\int\limits_\Theta \theta\, f(x\mid\theta)dG(\theta)}{\int\limits_\Theta f(x\mid\theta)dG(\theta)} \tag{36}$$

Now consider the problem of seeking two-sided Bayes probability interval for the parameter θ. We assume that the parameter θ is one-dimensional and $\theta \in [a,b]$. In accordance with the Bayesian approach (13), (14), symmetric $(1-\gamma)$ level two-sided Bayes probability interval for the parameter θ can be calculated by solving the following two equations:

$$\Pr(\theta \le \theta_* \mid x) = \int_a^{\theta_*} dG(\theta\mid x) = \frac{\int_a^{\theta_*} f(x\mid\theta)dH(\theta)}{\int_a^b f(x\mid\theta)dH(\theta)} = \frac{\gamma}{2} \tag{37}$$

and

$$\Pr(\theta \ge \theta^* \mid x) = \int_{\theta^*}^{b} dG(\theta\mid x) = \frac{\int_{\theta^*}^{b} f(x\mid\theta)dH(\theta)}{\int_a^b f(x\mid\theta)dH(\theta)} = \frac{\gamma}{2} \tag{38}$$

for the lower limit θ_* and the upper limit θ^*. Thus

$$\Pr(\theta_* \le \widehat{\theta} \le \theta^* \mid x) = 1-\gamma .$$

Now consider the case when available prior information is insufficient to accurately specify a single distribution function. Instead of a single prior $H(\theta)$ it defines the class Γ of such distribution functions, use of any of which as a prior does not contradict to available prior information. In this case, if for any distribution $H(\theta) \in \Gamma$ one determines symmetric $(1-\gamma)$ level two-sided Bayesian probability interval $(\theta_*(H), \theta^*(H))$ by solving equations (37), (38), then, there is no warranty that unknown parameter belong to it with the probability $1-\gamma$. In this case the generalized symmetric $(1-\gamma)$ level two-sided Bayes probability interval (θ_*, θ^*) for the parameter θ can be determined by solving the following optimization problems:

$$\theta_* \to \sup \tag{39}$$

subject to

$$\int_a^{\theta_*} dG(\theta \mid x) = \frac{\int_a^{\theta_*} f(x \mid \theta) dH(\theta)}{\int_a^{b} f(x \mid \theta) dH(\theta)} \le \frac{\gamma}{2} \tag{40}$$

for any $H(\theta) \in \Gamma$,

and

$$\theta^* \to \inf \tag{41}$$

subject to

$$\int_{\theta^*}^{b} dG(\theta \mid x) = \frac{\int_{\theta^*}^{b} f(x \mid \theta) dH(\theta)}{\int_a^{b} f(x \mid \theta) dH(\theta)} \le \frac{\gamma}{2} \tag{42}$$

for any $H(\theta) \in \Gamma$.

The problems (39)-(40) and (41)-(42) can be expressed in the following equivalent form:

$$\theta_* \to \sup \tag{43}$$

subject to

$$\sup_{H \in \Gamma} \frac{\int_a^{\theta_*} f(x \mid \theta) dH(\theta)}{\int_a^{b} f(x \mid \theta) dH(\theta)} \le \frac{\gamma}{2} \tag{44}$$

and

$$\theta^* \to \inf \tag{45}$$

subject to

$$\inf_{H\in\Gamma} \frac{\int_a^{\theta^*} f(x \mid \theta) dH(\theta)}{\int_a^b f(x \mid \theta) dH(\theta)} \geq 1 - \frac{\gamma}{2} \qquad (46)$$

12.6 Classes of distributions

Methods of solution the problems that were stated in sections (3), (4) depend on types of classes Γ of distributions. Commonly the following classes Γ are used in applications [Berger, 1994].

1. class of distributions of given shape, for example,
   Γ = {all symmetric, unimodal priors};
2. class of distributions that have given system of quantiles
   $\Gamma = \{H(\theta) : \alpha_i \leq \Pr(\theta \in \Theta_i) \leq \beta_i,\ i = 1, \ldots, m\}$;
3. class of distributions that have given system of moments;
4. contamination class
   $\Gamma = \{H = (1-\varepsilon)H_0 + \varepsilon F,\ F \in Q\}$,
   where H is a base prior, Q – is the allowed class of contaminations;
5 mixture classes
   $\Gamma = \{g(\theta) = \int g(\theta \mid \alpha) dG(\alpha),\ G \in Q\}$

In our investigations, we consider the class of distribution functions that can be expressed by means of the following linear constraints:

$$\int_\Theta dH(\theta) = 1 \qquad (47)$$

$$\begin{aligned} &\int_\Theta f_i(\theta) dH(\theta) \leq d_i, \\ &i = 1,2,\ldots,m_1, \\ &\int_\Theta g_j(\theta) dH(\theta) = \mu_j, \\ &j = 1,\ldots,m_2, \end{aligned} \qquad (48)$$

where functions $f_i(\theta),\ i = 1,2,\ldots,m_1,\ g_j(\theta),\ j = 1,2,\ldots,m_2,$ can be discontinuous.

In case when parametric space is one-dimensional, $\Theta = [a,b]$, and functions $g_j(\theta) = \theta^j$, $j = 1,2,\ldots,m_2$, this class consists of all distribution functions that have given constraints on the first m moments. If $f_i(\theta),\ i = 1,2,\ldots,m$, are indication functions for some intervals of the parametric space $\Theta = [a,b]$, then this class consists of all distribution functions that have given constraints on the quantiles.

12.7 Discussion

The approach considered in this paper overcomes one of the shortcomings in the modern methodology of evaluation of Severe Accident Risks on Nuclear Power Plants. Particularly, it provides techniques for quantification of uncertainty in estimates of reliability parameters resulted from insufficient data.

This approach was used in [Golodnikov et al, 2009], [Golodnikov, 2009] for conducting sensitivity analysis of Bayesian estimates of reliability parameters with respect to selection of a prior distribution in two widely used failure models: exponential and binomial.

For selected data we calculated lower and upper bounds for Bayesian estimates of Centrifugal pump failure rate and Isolation valves failure probability per demand assuming that only two quantiles are known about prior distribution.

Basing on results obtained we drew the following conclusions.

1. In the case of Centrifugal pump failure rate estimation (exponential model), Bayesian estimates of failure rates are essentially sensitive to the selection of a prior distribution function.
2. In the case of the Isolation valves failure probability per demand estimation (binomial model), strength of sensitivity of Bayesian estimates of failure probability to selection prior distribution depends on the number of failures occurred. When number of failures was less or equal 1 we observed slight sensitivity. Therefore, in this case any distribution, which has the same two quantiles, may be used as a prior. When number of failures exceeded 1, we observed rapid increasing of the sensitivity with the increase of number of failures.

13
Application of Information Theories to Safety of Nuclear Power Plants

13.1 Introduction

It is widely recognized that safety of nuclear power plants is a problem of great relevance for society. If it is not properly managed, the increase in power and complexity of the plants can lead to a catastrophic release of energy and dangerous materials and pollute the environment. To prevent this to happen, the nuclear community always put significant efforts into finding new ways to improve safety.

In fact, the nuclear accidents in Chernobyl and in Three Mile Island lead to initiation of extensive research activities within the nuclear society. A consensus quickly emerged that the breakdowns could not be explained exclusively from the perspective of technological failures but indeed required new, holistic views on safety. A solution seemed to be a strong focus on human attitudes and behaviors. The concept of safety culture, first introduced by the International Atomic Energy Agency (IAEA), rapidly became increasingly popular [IAEA, 2002].

These efforts notwithstanding, incidents continued to occur indicating that important aspects of plants' safety remain unsolved. In fact,

- in 2000, at the Davis Besse nuclear power plant, an extensive material degradation was detected in an area around a nozzle of the reactor pressure vessel. Commentators have suggested that a minor additional propagation of the crack would have lead to the rupture of the pressure vessel. Obviously, the plant owners did not possess accurate and updated information of current condition of the reactor pressure vessel [NUREG/BR-0353, 2008];
- in 2006, at the Forsmark nuclear power plant, a sudden disruption from an external power supply lead to the failure of the house turbine operation, the internal battery secured power supply, and of 2 of the 4 emergency diesel generators. Luckily, the 2 diesel generators that started were able to supply sufficient power to the cooling system of the reactor core, and, thus, were able to maintain the whole system in functioning conditions. This dangerous incident has resulted from errors in modifications of old components, caused by insufficient understanding of the consequences of introduced changes [Wikdahl, 2006].

These examples indicated that the established safety frameworks were in need of further improvement if the occurrence of severe accidents is to be prevented. Recognizing this need of further improvement, the IAEA has recently declared a strong focus on knowledge management and built a new knowledge management group for assisting member states in associated questions [IAEA, 2004; 2007].

However, the scope of knowledge management is known to be broad and overlap, for instance, information management, information theories, artificial intelligence, systems theories, synergetic, informatics, etc. What all of these theories and sciences have in common is the information orientation. With other words, the category of information plays the fundamental role here whilst other categories as data, knowledge, intelligence, etc. can be derived from the category of information. Therefore, a review of main information theories and interpretations seems to be a reasonable start.

13.2 Review of main information theories and interpretations

The concept of information acquires meaning only with respect to the context within which it is used. Specifically, different definitions of information can be provided depending on whether information is:

1. Used as a mean for regulation and control.
2. Transferred (communication).
3. Alternatively, generated (acquisition).

13.2.1 Information as a tool for regulating

Information as a tool for regulating (controlling, steering) activities was emphasized first within the science of cybernetics. This perspective is often denoted as functional (also known as cybernetic, external, active, or relation-based). Norbert Wiener, who may be considered as the father of cybernetics, claimed that all goal-oriented actions of human beings are based on information [Wiener, 1948]. An opposite, structural (also known as attributive, or internal) perspective believes that information mirrors an objectively existing diversity in the reality [Ursul, 1970], [Ashby, 1952; 1956]. The structural perspective emphasizes that information is an overall property of the reality from its simplest forms to the human brain or complex engineered facilities.

That is to say, the concept of information is the bi-polar concept that arrives into two shapes: a functional perspective and a structural perspective. Both perspectives are necessary for an effective and rigorous management of safety. From the functional point of view, all decisions on safety must to be based on accurate and updated information. From the structural point of view, safety of an engineered facility is determined by its structural organization, i.e. its components, subsystems, systems, and connections between them.

A measure of regulation of complex systems was proposed by another prominent figure in the field of cybernetics, Ross Ashby, who stated, "only variety can destroy variety" [Ashby, 1952; 1956]. The meaning of this statement, which is also known as the Law of requisite variety, is that the survival of a system depends on the regulator’s ability to master the diversity of external impacts and to block the flow of information. When essential variables start going outside the acceptable range, the regulator must take actions until the essential variables are stabilized and the safe condition is reached. A notion of essential variables (also known as order parameters within the science of synergetic) denotes those variables that govern the entire system. Ashby’s view of information as the variety agrees with the structural perspective on information.

Information as an instruction (algorithm, program) is emphasized within a non-probabilistic approach, also known as algorithmic information theory, developed by Andrej Kolmogorov [Kolmogorov, 1965]. Kolmogorov considers information as an instruction that has to be executed in order to transform a system from state A into state B. The larger the difference between the states A and B, the longer (more complex) the transformation instruction.

Although an exact mathematical measure of Kolmogorov's complexity has not yet been provided by the community of mathematicians, the idea itself allows fruitful discussions on how to manage instructional information, which is contained in norms, standards, instructions, and other types of documentation. Another insight that follows is that successful accomplishment of the goal is dependent on the quality in a program, which describes what measures need to be taken to achieve the goal.

13.2.2 Information as a transferred message

Information is regarded as a transferred message within a communication model, which was proposed by Claude Shannon [Shannon, 1948]. This model includes at least the following elements: a source (a sender), a channel, and a receiver. The channel is always noisy, and noise leads to the loss or misinterpretation of information. Ashby [Ashby, 1952; 1956] was first to point out that the distinction between message and noise depends on what the receiver regards as important. The receiver tends to ignore information that does not promote the achievement of his goals and objectives. Furthermore, the receiver cannot understand information that considerably exceeds his background knowledge. For this reason, all concepts involving information communication should be formulated accounting first and foremost the information acquired by the receiver, while the information sent by the source should play a lesser role, as pointed out by David Harrah in his model of rational communication [Harrah, 1956; 1957].

Illustrativeness of Shannon's and Harrah's communication models helps to understand the role of individual objectives, values, and knowledge for information perception and making choices. It is therefore increasingly important for top managers to clarify for all the employees the overall goal and values of the entire organization. In case of conflicts between the subjective objectives of the individual employees and the overall goal and values of the entire organization, the latter have precedence.

As mentioned above, Shannon's communication model [Shannon, 1948], considers information as a message that is transferred from a source to a receiver via a channel. The information content of that message was defined by Hartley [Hartley, 1928], Shannon [Shannon, 1948], and Wiener [Wiener, 1948] as the uncertainty that can be eliminated upon reception of the message. Ralph Hartley [Harrah, 1956] appears to be first with providing an explicit mathematical way to determine the information content of an event in the simplest case in which an event has N possible outcomes with equal probability of occurrence $p = (1/N)$:

$$I = -\log p = \log N$$

Note that according to this interpretation, it is possible to speak about information only when several alternatives are available. For situations of certainty (determinism), a number of available alternatives shrink to 1 and information content shrinks to 0. For situation of complete randomness, all alternatives have equal probability and information content goes to maximum. The smaller probability of a chosen alternative to happen, the larger the uncertainty it removes and thus the larger amount of information it generates.

The great advance of information theories lies in highlighting relationships between information, uncertainty, presence of alternatives, choice, and, in the end, decision making. A decision, i.e. a choice of one among available alternatives, may need to be made although the available information is insufficient. Each choice is thus associated with a risk of making wrong decision and resulting unwanted consequences. Looking from that perspective, insufficient information can be interpreted in terms of the existence of several alternatives to act. In case of decision making by a group of people, a lack of consensus indicates that more information is needed in order to clarify the best choice.

Hartley's equation was limited to the simplest case of complete randomness and later Shannon proposed a more general equation [Shannon, 1948]. The following mathematical expression gives the average information, which is available when the knowledge about an alternative is expressed by the probabilities p_i. Information content is then measured by averaging over n groups:

$$I = -\sum_{i=1}^{n} p_i \log p_i$$

This mathematical expression is identical to that of entropy, as it is defined in statistical mechanics, and plays a fundamental role in several applications of information theories. Of interest to this review, Jaynes [Jaynes, 1957] suggested the information content as a fundamental quantity from which the probability values, p_i, weighting the possible outcomes of that event, could be recovered. To this end, the values p_i must be chosen so that information content has a maximum constrained by the available knowledge about a given event. This procedure is known as a principle of maximum entropy.

Prescription that follows from the principle of maximum entropy is to use the probability distribution, which maximizes the information content with respect to the available knowledge. This procedure allows making the least biased conclusions under situations, when available information is not sufficient for making certain conclusions. Figuratively speaking, one needs to choose a broad probability distribution that comprises all known events.

13.2.3 Information generation addressed by a dynamical information theory

Information generation is addressed by a dynamical information theory [Chernavsky, 2004], which has recently being developed within the frame of science of synergetic. A central notion of information value estimates to what extend the information helps to accomplish goals and objectives of the user.

Information without a meaning has certainly no value. Therefore, it is in many cases convenient to study information value and meaning at the same time, using semantic-pragmatic information theories. The fundamental works of Bar-Hillel and Carnap [Bar-Hillel and Carnap, 1953] suggested using logic probability to measure semantic information. Logic probability describes to which degree a hypothesis has been confirmed and, from the practical point of view, resembles probabilistic equations of Shannon and Wiener.

The equation for estimate of information value, V, was proposed in an early work of Alexander Harkevich [Harkevich, 1960]:

$$V = \log \frac{p_1}{p_0} = \log p_1 - \log p_0$$

where p_0 and p_1 - are probabilities of goal accomplishment before (a priori) and after (a posteriori) the information has been acquired.

Disinformation leads to decrease in probability of goal accomplishment and information value becomes negative. In the opposite situation, when the goal has actually been accomplished, the value of information becomes equal maximal information content for the system:

$$V = \log \frac{p_1}{p_0} = \log p_1 - \log p_0 = \log 1 - \log \frac{1}{N} = 0 - (-\log N) = \log N = I$$

In the above equation, the posteriori probability p_1 is equal 1 because the goal has been accomplished. The a priori probability p_0 is equal 1/N because under the conditions of limited knowledge all alternatives are considered being equally probable.

13.3 A proposal on how some information theories and interpretations can be applied to safety of nuclear plants

It needs to be emphasized that the nuclear community including the nuclear operators, the government regulators, the international organizations such as the IAEA, and other involved actors, has always had a strong focus on safety and a strive for continuous improvement. However, the nuclear power plants belong to a class of complex dynamic systems, which can be difficult to fully overview, understand, and control. This work suggests using information theories and interpretations to solve some safety issues of nuclear plants, in particular those associated with nuclear containments, maintenance schedule, configuration management, and incident explanation.

13.3.1 Configuration management

In fact, the amount of information that is handled at nuclear power plants increases steadily with time. The amount of information at the operational start was quite limited and well structured as so-called safety analysis reports (SAR). With time, the plants undergo modifications and the old components and systems are being replaced by new items. Changes in technological processes, variations in environmental parameters, new-employments, and other factors contribute to the need to update the information.

The vital role of information for safety management of nuclear power plants was recognized by the IAEA [IAEA-TECDOC-1335, 2003] within the concept of configuration management. The IAEA structures the information into following categories:

- the documentation of the entire life-time of the nuclear power plant comprising its design, manufacturing, construction, pre-operational testing, operation, maintenance, testing, and further modification;
- the information contained in safety standards, codes, norms, etc.;
- the personnel files and work instructions.

In all three cases, the IAEA requires information to be complete and accurate. However, measures of information completeness and accuracy are not discussed. Similarly, potential problems arising from excess of information are not addressed, disregarding the fact that extraction of relevant information from abundant sources is as problematic as the lack of information. Crucial issues connected with quantifying the amount of transferred information and assessing the impact of the receiver's background knowledge on the successful completion of this process do not receive the much-needed attention. Finally, the need for time optimization of information generation is mentioned in wordings like "right information at right time", but no indication is offered on appropriate strategies to achieve this objective. An additional remark, which demands consideration by any satisfactory approach to safety management but is not addressed adequately by the IAEA's document, concerns the IAEA's requirement for information be clear. It should be stressed that, although clarity is a necessary requirement for achieving safety, it does not suffice. In fact, clear information is not necessarily true, or, alternatively, disinformation can also be clearly communicated.

It is to be noted that according to the communication theories of Shannon and Harrah the information that is received by the receiver is not the same information that has been sent by the source. Information acquired by the receiver is conditioned to the receiver's background knowledge as well as subjective goals, objectives and values. That is why all measures (such as information completeness, accuracy, clarity) must be estimated from the perspective of receivers (users) of information.

13.3.2 Degradation of nuclear containments

It is known that containments in nuclear power plants constitute the last barrier between the dangerous radiation and the environment. In case of a nuclear accident, the containment is supposed to confine the radiation and, by doing so, to protect people and the environment. However, containments were constructed for decades ago and are subject to a long-term ageing deterioration. The original design lifetime of containments has been exceeded in many cases [Österberg, 2004]. At the same time, the established testing and inspection practices are limited mostly to visual inspections of containments' accessible surfaces and pressure tests, and are not capable of providing needed information about the state and safety of containments.

From the informational-functional point of view, all conclusions on containments' safety must be based on the accurate and updated information. In this regard, two major questions must be answered: "What material parameters are needed to be measured to assess the state and safety of containments?" and "What methods shall be used to perform needed measurements?"

Here, the earlier mentioned concept of "essential variables" by Ashby (corresponds to a more recent term is "order parameters" in the science of synergetic) is useful. Though the containments are complex structures, composed of a reinforced and pre-stressed concrete and a steel liner vessel, it might be sufficient to use a few essential variables that govern the state and safety of the entire structure. The previous study [Österberg, 2004] has indicated that measurements of four major variables as concrete strength, concrete fracture toughness, pre-stressing force, and corrosion of steel members provides a solid ground for overall assessment of containments state and safety.

Once essential variables have been identified, one needs to decide what method to use for the measurements. It is known that concrete structures may be tested by means of various methods such as taking cores, Schmidt hummer, visual inspections, radar, radiography, fiber-optical method, acoustic emission, and other destructive and non-destructive (NDT) testing methods. The information approach offers an appropriate tool for choosing the best method with regard to methods' informativeness. According to the previous study [Österberg, 2004] the best (the most informative) method seems to be a quantitative acoustic emission (QAE) NDT method. The QAE method can reliably and independently:

- monitor the entire reinforced concrete structure under the specific conditions of nuclear power plants;

– reveal specific defects or combination of defects in the entire structure and differentiate between undamaged and damaged parts of the structure;
– distinguish between different types of defects;
– assess flaws in terms acceptable for fracture mechanics analysis. Particularly, the real stress state in the pre-stressed concrete structure.

An additional remark concerns the algorithmic information and its impact on containments' safety. The first generation of containments in the Swedish nuclear power plants was designed and constructed in 60s and 70s, when there were no specific standards or norms available for the task. The second generation of containments was designed and constructed in 80s after the American standards were developed and published. Because of the insufficient instructional (algorithmic) information at the time of design and construction, the old containments are very probably not as strong as the new containments [Österberg, 2004].

13.3.3 Optimization of maintenance schedule

Each nuclear power plant contains several thousands of components, most of which are needed to be maintained over time. A question that arises is "How to determine an optimal time point for maintenance activities?" Support for the optimization of maintenance schedule is provided by the dynamical information theory. As previously discussed, the information theories highlight the intimate relationship between decision-making and information.

In case of maintenance, a maintenance engineer must choose (decide) when to perform maintenance activities. The overall goal of the maintenance is to find defects, if any, and to repair them in order to restore the state of the facility. If maintenance activities are scheduled too late, a defect will likely cause failure of the component. Then the decision to repair becomes obvious, the posterior probability p_1 goes to 1, and the information value V goes to maximum:

$$V = \log(p_1 / p_0) = \log(1 / p_0) = V_{\max}$$

At the same time, the information content I goes to zero. Once the component has broken down, the choice becomes obvious. That is it to say, this is the situation of certainty with one outcome:

$$I = \log N = \log 1 = 0$$

If maintenance activities are scheduled too early, the defect will very probably not be detected, which means that available alternatives of the component's state (the presence of defect alternatively the absence of defect) have equal probabilities. In this situation of uncertainty the information content goes to maximum:

$$I = \log N = I_{\max}$$

At the same time, the information value V goes to zero because the maintenance actions do not increase probability of goal-achievement:

$$V = \log(p_1 / p_0) = \log 1 = 0$$

To sum up, when choosing an optimal time point for maintenance, one needs to consider and maximize both information content and value. In many cases, the optimal time point for maintenance will lie close to the end of the lifetime, when the defect if large enough to be reliably detected, but is not so large that it can cause a sudden failure.

13.3.4 Incident explanation

It has almost become a common practice for nuclear regulators to blame poor safety culture, each time a degradation is observed in a nuclear plant. For instance, after the well known incident in Forsmark in July 2006 [Wikdahl, 2006], the regulator explained the incident by deficiencies in the plants' safety culture. As a result, the following program of corrective measures had a strong focus on safety culture and associated questions as attitudes to safety, existence of written instructions, etc.

This work believes that the focus on safety culture is necessary but not sufficient for explanation of occurred incidents and prevention of new incidents to happen. Indeed, one needs to look from the information perspective and to identify and correct deficiencies in information acquisition, information communication, knowledge creation, decision-making, and other stages in information processes.

In case of the Forsmark incident, it is known that the incident was initiated by a severe distortion in the external power grid. This distortion propagated far into the plant and lead to the failure of several power supply systems and safety systems. Actually, the employees of the plant lacked experiences of similar distortions and could not foresee possibility of such distortions to happen. Furthermore, the incident revealed that the employees of the plant did not fully realize that recently performed modifications of old electrical equipments have changed systemic interactions of the plant. In particular, the sensitivity to distortions increased or with other words, vulnerability of the plant degraded.

Actually, modification works and exchange of old equipments take place in several operating nuclear plants. As a result, the nuclear community may face situations when decisions need to be taken when the available information is not sufficient to take certain decisions. Therefore, it is important to highlight the role of information aggregation and the need to consider all types of knowledge including practical experiences, modeling results, expert judgments, calculations, tests, etc. for making plausible decisions under conditions of uncertainty.

13.4 Discussion

From the review of main information theories and perspectives, it emerges that information can be understood:

- as a message which is transferred from a source to a receiver via a communication channel;
- as an instruction (program, algorithm) which, once it is carried out, allows the transformation of a system from state A to state B;
- as uncertainty regarding the present state or evolution of a system, which can be reduced upon reception of the message;
- to mirror the diversity of reality (structural perspective);
- to provide the ground for regulation, control, steering, and other goal-directed activities (functional perspective).

Furthermore, information can be quantified, and therefore measured. The relevance of its potentially fundamental role as a tool to verify whether transfer of crucial knowledge has properly occurred cannot be overestimated. In addition, as it is emphasized in the communication models, issues concerning completeness, accuracy, and clarity of information should mainly concern the information acquired by the receiver. The role of background knowledge, objectives, and values of the receiver is an additional point of concern, which emerges from this review. Finally, a mathematical model has been developed which allow current information to be updated without running the risk of unjustified bias in favor of a particular alternative/outcome.

The diversity in interpretations of the concept of information mirrors potential instances in which problems for the safe operation of nuclear plants may arise. The promotion of a proper understanding this concept within the specific context in which it is used, and of the mathematical tools developed within information theories, will hopefully help preventing accidents to occur in the future.

PART III. Technological Aspects

14 Growing Pyramidal Networks

14.1 Intelligent systems memory structuring

Formation of intelligent systems memory structure needs to be done simultaneously with perception of information and under the impact of the information perceived and already stored. The memory structure reflects the information perceived. Information structuring is an indispensable function of memory. [Gladun, 2003]

The main processes of structuring include formation of associative links by means of identifying the intersections of attributive representations of objects, hierarchic regulation, classification, forming up generalized logical attributive models of classes, i.e. concepts.

Under real conditions of information perception there is often no possibility to get whole information about an object at once (for example, because of faulty foreshortening or lighting during the reception of visual information). That is why the processes of memory formation should allow for the possibility of "portioned" construction of objects models and class models by parts.

In different processes of information processing, objects are represented by one of the two means: by a name (convergent representation) or by a set of meanings of attributes (displayed representation). The structure of memory should provide convenient transition from one representation to another.

Systems, in which the perception of new information is accompanied by simultaneous structuring of the information stored in memory, are called **self-structured** [Gladun et al, 2008]. Self-structuring provides a possibility of changing the structure of stored in memory data during the process of the functioning because of interaction between the received and already stored information.

The building of self-structured artificial systems had been proposed to be realized on the basis of networks with hierarchical structures, named as "***growing pyramidal networks***" (**GPN**) [Gladun et al, 2008]. The theory as well as practical application of GPN was expounded in a number of publications [Gladun, 1987], [Gladun, 1994], [Gladun, 2000], [Gladun and Vashchenko, 2000].

Pyramidal network is a network memory, automatically tuned into the structure of incoming information.

Unlike the neuron networks, the adaptation effect is attained without introduction of a priori network excess. Pyramidal networks are convenient for performing different operations of associative search. Hierarchical structure of the networks, which allows them to reflect the structure of composing objects and gender-species bonds naturally, is an important property of pyramidal networks. The concept of GPN is a generalized logical attributive model of objects' class, and represents the belonging of objects to the target class in accordance with some specific combinations of attributes. By classification manner, GPN is closest to the known methods of data mining as decision trees and propositional rule learning.

GPN realization has following stages:

- building the structure of a network for some initial set of objects, assigned by attributive descriptions;
- training the structure, with a purpose to allocate its elements, allowing classifying all objects of the initial set;
- recognizing the belonging to some class of objects of certain object, which not belongs to initial set of objects.

The growing pyramidal networks respond to the main requirements to memory structuring in the artificial intelligent systems [Gladun, 2003]:

- in artificial intelligent systems, the knowledge of different types should be united into net-like structure, designed according to principles common for all types of knowledge;
- the network should reflect hierarchic character of real media and in this connection should be convenient for representation of gender-type bonds and structures of composite objects;
- obligatory functions of the memory should be formation of association bonds by revealing intersections of attributive object representations, hierarchic structuring, classification, concept formation;
- within the network there should be provided a two-way transition between convergent and displayed presentations of objects.

The research done on complex data of great scope showed high effectiveness of application of growing pyramidal networks for solving analytical problems. Such qualities as simplicity of change introduction the information; combining the processes of information introduction with processes of classification and generalization; high associability makes growing pyramid networks an important component of forecasting and diagnosing systems, especially in the area of GMES. The applied problems, for solving of which GPN were used are: forecasting new chemical compounds and materials with the indicated properties, forecasting in genetics, geology, medical and technical diagnostics, forecasting malfunction of complex machines and sun activity, etc. Special kind of applications is aimed to support intelligent data processing in GMES.

14.1.1 GPN theoretical foundations

The word "object" here and further is understood in a broad sense is there can be a real physical object, some process, a situation, etc.

The model of classes of the objects, used for the decision of tasks of classification, diagnostics and forecasting, should include all the most important attributes describing a class. The model also should display for this class the characteristic logical connections between essential attributes. Therefore, the basic attention concentrates on formation of the generalized logical multivariate models of objects classes. Such models, in fact, are the concepts that correspond to the classes of objects [Voyshvillo, 1967], [Gorskii, 1985].

The concept is usually defined in logic as an idea that reflects essence of objects. Most of used concepts are the result of generalization of attributes that characterizes the objects of the class.

The concept "attribute" can be used for characterizing the objects and can be used in such logic operations as extraction, recognition, identification, etc. It is necessary to note, that separation of attributes on essential and unessential is conditional and depends on problems for which decision they are used.

✓ Concept

From a philosophical point of view, the concept consists of two parts – extensional and intensional:

- extensional part covers all instances belonging to this concept;
- intensional part includes all the properties that representative to these instances.

The connections between instances and their attributes play an important role in determining the hierarchical relationship between concepts and attributes.

The set of the instances, generalized in concept, constitute the volume of defined concept. In system of knowledge, the concepts play a role of base elements for composing propositions and other logic forms of thinking. Transition from a sensual step of cognition to abstract thinking, actually, means transition from reflection of the world in the form of perception and presentations to its reflection in concepts.

Classification, generalization, structuring of the perceived information, its inclusion in system of knowledge, is carried out based on an available set of concepts. In these processes two basic functions of concepts: recognition and production of the elements models of the world, in which the bearer of knowledge operates, are realized. Recognition processes became for a long time object of research and automation while production of models for the present is little-investigated problem. Production of models plays the important role in creative activity. Only by a concrete definition of concepts we can create (for example, to draw) images of concrete houses, trees, cars, etc. Production of elements models of the world underlies designing of the engineering objects.

Attributes belonging to concept by their role in realization of the basic functions of concept divide on two types – disjunctive and unified:

- disjunctive attributes are attributes, which do not occur or occur seldom in concepts volume. These attributes are most effective at realization of recognition functions;
- unified attributes are those attributes, which are inherent in all or many elements of concept volume refer to, but they can be widespread and outside of concept volume. Without these attributes, the production of elements models of the world is impossible.

For example, for all birches such attributes, as presence of a trunk, roots, and crown are characteristic. It their unifying attributes which are inherent in all trees. A well-known disjunctive attribute of birches is white color of a bark.

The degree of detailed elaboration of model created based on concept depends on the purpose of a task. The model of the bridge created at the decision of a task "to draw a bridge", essentially differs on a degree of detailed elaboration from the model of the bridge created at the decision of a task "to design the bridge".

The success of the decision of the problems including production of models depends on that, how much used concepts correctly and full characterize corresponding classes of elements of the world.

Now, it is possible to give more constructive definition of concept, more suitable by consideration of information-technical aspects of problems of formation and processing of concepts.

Concept – an element of the knowledge system, representing the generalized model of some class of instances.

In processes of recognition and production of models, the concept is used as logic function of the attributes, having the value "true" for instances from volume of concept and value "false" in other cases.

✓ System of concepts

The set of concepts included in the system of knowledge, will be called the system of concepts of knowledge bearer.

Systems of concepts are hierarchical, as a rule. Volumes of concepts of all levels of hierarchy, except for bottom, are formed by consolidation of volumes of some concepts of lower levels. For example, the volume of concept "fruit" unites volumes of concepts "apple", "pear", etc.

Systems of concepts are dynamical. The structure of concepts varies because of interaction of their bearers with an environment, and during the decision of problems.

At each moment of time, the state of system of concepts reflects individual experience of its bearer. Therefore, separate concepts and systems of concepts in the whole are subjective.

Any system of concepts by virtue of the discreteness, limitation of structure of concepts, imperfection of separate concepts cannot reflect variety and a continuity of the real world. Volumes of the concepts which have been not introduced "by definition", as a rule, have no precise dividing boundaries. There are many of the transitive forms that complicate carrying out of conditional boundaries between volumes of concepts. There are many transitional forms that complicate definition of conditional boundaries between volumes of concepts.

Because of incompleteness of the world mapping in concepts system, and of concepts subjectivity, univocal identification of elements of the world based on concepts system often appears inconvenient or even impossible. Therefore, volumes of many concepts can be considered as fuzzy sets. Each bearers of concepts system possesses the membership function, which, thus, has subjective character.

✓ Inductive formation of concepts

From logic structure point of view, the concepts are categorized as:

- conjunctive concepts, which can be described by conjunction of attributes;
- disjunctive concepts, which can be described by disjunction of conjunctions of attributes;
- concepts with the exclusive attributes, reflecting absence of some attributes in the instances, which belongs to concept volume.

The concepts, included in everyday practice, usually are conjunctive. More complex logic structure is characteristic for the concepts, formed in the research process.

In this case, the logical complexity of the concepts usually arises following circumstances:

- the attributes space is incorrectly chosen;
- training set is incomplete reflects specificity of concept volume;
- the volume of formed concept consists of instances that are vastly different from each other.

Consider a task of inductive formation of concepts for not intersected sets of objects $V_1, V_2, \ldots, V_n$, , each set represents some class of objects with known properties. Let L – the set of objects used as training set. All the objects of set L are represented by sets of attribute values. Relations $L \cap V_i \neq \varnothing$ and $V_i \not\subset L \left(i = 1,2,\ldots,n\right)$ take place. Each object from set L corresponds to one of set V_i. It is necessary to generate n concepts by analysis L. The amount of these concepts must be sufficient for correct recognition of belongings of anyone $l \in L$ to one of set V_i.

In forming the concept corresponding to set V_i, the objects of training set included in V_i, are considered as examples of set V_i, and the objects, not included in V_i, – as counterexamples of set V_i.

Each concept, generated on the basis of training set, is approximation to real concept. The proximity of concepts depends on representativeness of training set, i.e. on the detailization of the concept volume peculiarities.

Problem of inductive formation of concepts is similar to the problem of learning pattern recognition. And in both cases as a result of learning a model of a class of objects is constructed. At formation of concepts stronger requirements are made to this model (concept). It must provide not only recognition, but also the opportunity to generate models of concrete objects. In this regard, the model should be reflected attributive, structural, and logical characteristics of objects.

The training set usually has the tabular form. The rows of the table correspond to the set of objects properties, columns – to attributes. Names of classes are specified in a special column. The concept, which is formed because of the analysis of the training set, is usually described by a logical expression in which the variables are the names of the attributes values.

Known methods of formation of concepts [Gladun, 1987] [Bongard, 1967], [Vagin, 1988], [Gladun and Vashchenko, 1995], [Pospelov, 1986], [Gladun and Rabinovich, 1980], [Gladun and Vaschenko, 1995], [Michalski et al, 1986], [Piatetsky-Shapiro and Frawley, 1991] as a matter of fact are methods of controlled choice of the attributes values that characterize the classes of objects. The choice can be simplified due to use of adequate representation of the analyzed information.

14.1.2 Requirements to the methods of concepts formation

During the work, following requirements to the methods of concepts formation are revealed:

- for increasing the reliability of the diagnosis or the forecast, it is necessary to consider dependence of the defined variable from combinations of known attributes, i.e. to consider joint simultaneous influence of attributes. The formed concept should reflect such dependences;
- depending on a choice of a method of training for concepts of the same class of objects various logic descriptions can be received. Naturally, there is a question on quality of logic models. The best results of application of concepts for classification, diagnostics, and forecasting, as a rule, correspond to more generalized concepts, i.e. concepts that are described by more simple logic expressions. Degree of complexity of logic expression can be estimated by the number of its variables. The method of concepts training should provide formation of as more as possible simple concepts;
- choice operations, such as a choice of values of properties, objects, combinations of values of signs, etc., prevail in processes of knowledge mining. It is a combinatory problem. The volume and time of choice operations quickly grows with increase amount of data. This effect of "information explosion" blocks practical application of many methods. In this connection, there is a necessity for use the network structures reducing amount of search operations at realization of processes of knowledge mining.

14.1.3 Requirements to the network structure used for the knowledge mining

The key enabler of increase of search operations efficiency is use of network structures for modeling environments in which problems solving. Orientation to real applied environments essentially raises a level of requirements to network models. We shall define the features of real environments rendering strong influence on processes of the problems solving:

- multicoupling. Real environments usually include many objects connected by a lot of relations;
- heterogeneity. For real environments the variety of objects and relations is characteristic;
- hierarchy. In real environments it is necessary to operate with the compound objects representing compositions of more simple objects;
- dynamism. Real environments are usually subject to frequent changes.

Given the above features of real environments, we formulate requirements to the network structure, representing the environment.

- the network should possess the developed associative properties, i.e. to provide effective performance of various search operations;
- the network should reflect the hierarchy of real environments and therefore should be convenient to represent genus-species relations and structures of composite objects;
- in a network the means limiting zones of search by time, spatial or substantial criteria should be stipulated, i.e. the network should provide selectivity of search on a time, spatial or meaningful context;
- at construction of a network the classes of objects and situations should be formed; input of the new information into the network must be accompanied by the classification process;
- the network should allow parallel execution of search operations.

14.2 Pyramidal networks

The set forth above requirements are answered to the full with pyramidal networks [Gladun et al, 2008].

A growing pyramidal network (GPN) is an acyclic oriented graph having no vertexes with a single incoming arc. Examples of the pyramidal networks are shown in figures below. Vertices having no incoming arcs are referred to as ***receptors***. Other vertices are named ***conceptors***. The ***subgraph*** of the pyramidal network that contains vertex **a** and from all the vertices that belong to subgraph there are paths to vertex **a** is named the pyramid of vertex **a**. The set of vertices contained in the pyramid of vertex **a** is referred to as the ***subset*** of vertex **a**. The set of vertices reachable by paths from vertex **a** is named the ***superset*** of vertex **a**. The set of vertices that are connected with paths to vertex **a**, is referred to its superset.

In subset and superset of the vertex, ***0-subset*** and ***0-superset*** are allocated, consisting of those vertices, which are connected to it directly.

When the network is building, the input information is represented by sets of attributes values describing some objects (materials, states of the equipment, a situation, illness etc.). Receptors correspond to values of attributes. In various tasks, they can be represented by names of properties, relations, states, actions, objects or classes of objects. Conceptors correspond to descriptions of objects in general and to crossings of descriptions and represent GPN vertices.

14.2.1 Building of GPN

Initially the network consists only of receptors. Conceptors are formed as a result of algorithm of construction of a network. After input of object attribute description, corresponding receptors switch to a state of excitation. The process of excitation propagates through the network. A conceptor switches into the state of excitation if all vertices of its 0-subset are excited. Receptors and conceptors retain their state of excitation during all operations of network building.

Let $\mathbf{F_a}$ be the subset of excited vertices of the 0-subset of vertex **a**; **G** be the set of excited vertices in the network that do not have other excited vertices in their supersets. New vertices are added to the network by the following two rules:

Rule A1. If vertex **a**, that is a conceptor, is not excited and the power of set $\mathbf{F_a}$ exceeds 1, then the arcs joining vertices of set $\mathbf{F_a}$ with the vertex **a** are liquidated and a new conceptor is added to the network which is joined with vertices of set $\mathbf{F_a}$ by incoming arcs and with the vertex **a** by an outgoing arc.

The new vertex is in the state of excitation. Rule A1 is illustrated in Figure 171(a,b). According to the Rule A1, the condition for adding a new vertex to the network is a situation, when certain network vertex is not completely excited (at least two vertices of 0-subset are excited). Figure 171(a) shows a fragment of network in some initial state. Receptors 4, 5 switch to a state of excitation, the network switches to state (b), and a new vertex appears – a new conceptor. Receptors 2, 3 switch to a state of excitation additionally. The network switches to state (c).

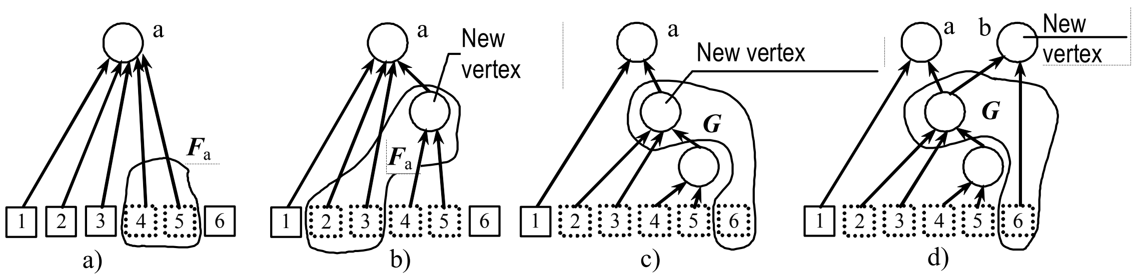


Figure 171. Pyramidal network building

New vertices are inserted in 0-subset of vertices, which are not completely excited. New vertices correspond to intersection of object descriptions, represented by incoming arches. Once new vertices have been introduced into all network sections where the condition of rule A1 is satisfied, rule A2 is applied to the obtained network fragment, concluding the object pyramid building.

Rule A2. If the power of set G exceeds 1 element, a new conceptor is added to the network, which is joined with all vertices of set G by incoming arcs.

The new vertex is in the state of excitation. Rule A2 is illustrated in Figure 171(c,d). The state (d) was obtained after the excitation of receptors 2-6.

In applying the Rule A1 the main cross-linking relation is a relation of intersection of receptor set, excited by input of the object description and other sets of receptors included into pyramid, recently formed by conceptors. Rule A2 concludes the building of pyramid, which represents complete description of the introduced object.

✓ Properties of a pyramidal network

Let us note some properties of pyramidal networks.

Depending on applied area in which networks are used, the receptor can represent the attribute value, the elementary fact from the description of a situation, value of an economic parameter, a symptom of illness, the letter, a word, etc. Conceptors correspond to descriptions of objects, situations, realizations of processes or the phenomena, words, phrases, plans, and also crossings of descriptions.

The pyramidal network is the network memory, which is automatically customized on structure of the input information. Optimization of the information representation due to adaptation of network structure to structural features of entrance data is as a result reached. In addition, unlike neural networks, the effect of adaptation is reached without introduction of aprioristic redundancy of a network.

Pyramidal networks are convenient for execution of various operations of associative search. For example, it is possible to select all the objects that contain a given combination of attribute values by tracing the paths that outgo from the network vertex corresponding to this combination. To select all the objects whose descriptions intersect with the description of a given object it is necessary to trace the paths that outgo from vertices of its pyramid. Rules A1, A2 establishes associative proximity between objects having common combinations of attribute values.

Hierarchical organization is an important property of pyramidal networks. This provides a natural way for reflecting the structure of composite objects and generic-species interconnections.

The algorithm of a network building provides an automatic establishment of associative affinity between objects based on common elements and their descriptions. All the processes connected with a network building at processing one description are localized in rather small part of a network – a pyramid corresponding to this description. The important property of semantic networks of pyramidal structure is their hierarchy allowing naturally mapping structure of compound objects and genus-species relations.

Conceptors of the network correspond to combinations of attribute values that define separate objects and conjunctive classes of objects. By introducing the excited vertices into the object pyramid, the object is referred to classes, which descriptions are represented by these vertices. Thus, during the network building the conjunctive classes of objects are formed, the classification of objects is performed without a teacher. Classifying properties of pyramidal network are vital for modeling environments and situations.

Profitability is also the advantage of pyramidal networks, because identical combinations of attributes values several objects are represented in network by one common pyramid.

In a pyramidal network, the information is stored by its mapping in structure of a network. The information on objects and classes of objects is presented by vertex ensembles (pyramids) distributed on all network. Entering of the new information causes redistribution of links between vertexes, i.e. change of network structure.

Certainly, the full advantages of pyramidal networks are appeared at their physical realization supposing parallel distribution of signals on a network. The important property of a network as means of storage of the information is that the opportunity of parallel distribution of signals is combined with an opportunity of parallel reception of signals on receptors.

This property appears useful at applications of pyramidal networks in robotic systems, the automated systems of scientific researches, systems of the automated designing. Conversion from converged representation of objects (conceptors) to expanded (sets of receptors) is performed by scanning pyramids in top-down and down-top directions.

✓ Concept formation in a pyramidal network

Training GPN consists in formation of the structures representing concepts, on a basis of attributive descriptions of the objects incorporated into classes with known properties.

Concept is an element of knowledge system, representing generalized logic attributive model of objects class. This model is used in processes of objects recognition. The set of objects generalized in concept is its volume.

The combinations of attributes allocated in ready-built pyramidal network, representing descriptions of objects of training set, are used as "a building material", a basis of further logic structure of concept.

Let L be the pyramidal network representing all of training set objects. For formation of concepts $A_1, A_2, \ldots, A_i, \ldots, A_n$ corresponding to sets $V_1, V_2, \ldots, V_i, \ldots, V_n$, pyramids of all objects of training set are scanned in order. The vertices of scanned pyramid during its scanning are considered excited. Special vertices in network are identified in order to recognize objects from the concept volume. They are referred to as check vertices of a certain concept. At performance of inductive generalization, it is natural that the most important attribute or the combination of attributes describing group of objects – concept A_i, those vertex from pyramid A_i, which meet in pyramid A_i more often. Such attributes (or their combinations) are necessary for noting as check vertices. Check vertices are used in the further at decision about belonging of a new object to the concept. If in a pyramid of concept A_i there are some vertices, which include into equal quantity of objects from the given concept volume, it is natural to choose from the given vertex such, which unites maximal quantity of attributes (receptors) from a concept pyramid. This vertex defines the most typical combination of attributes of the objects incorporated into concept. In selecting the check vertexes, two characteristics of network vertices are used:

- $\{m_1, \ldots, m_i, \ldots m_n\}$, where m_i $(i = 1,2,\ldots,n)$ is a number of objects of volume of concept A_i, which pyramids include the given vertex;
- k is the number of receptors in the pyramid of this vertex.

For receptors $k = 1$. While scanning, the pyramid is transformed by the following rules:

Rule B1. If in the pyramid of an object from concept volume A_i, the vertex, having the largest k among all the vertices with the largest m_i, is not a check vertex of concept A_i, then it is marked as a check vertex of the concept A_i.

The rule allows existence several vertexes among the excited vertexes with identical m_i, exceeding m_i of other excited vertexes. If in group of the vertexes having largest m_i, values k of all vertexes are equal, any of vertexes can be marked as check vertex of concept A_i.

The rule B1 is illustrated in Figure 172. In this situation, vertex 6 is selected as check vertex, because it has the largest k among vertices with the largest m_i (6, 13, 14). Values m_i are shown inside symbols of vertices.

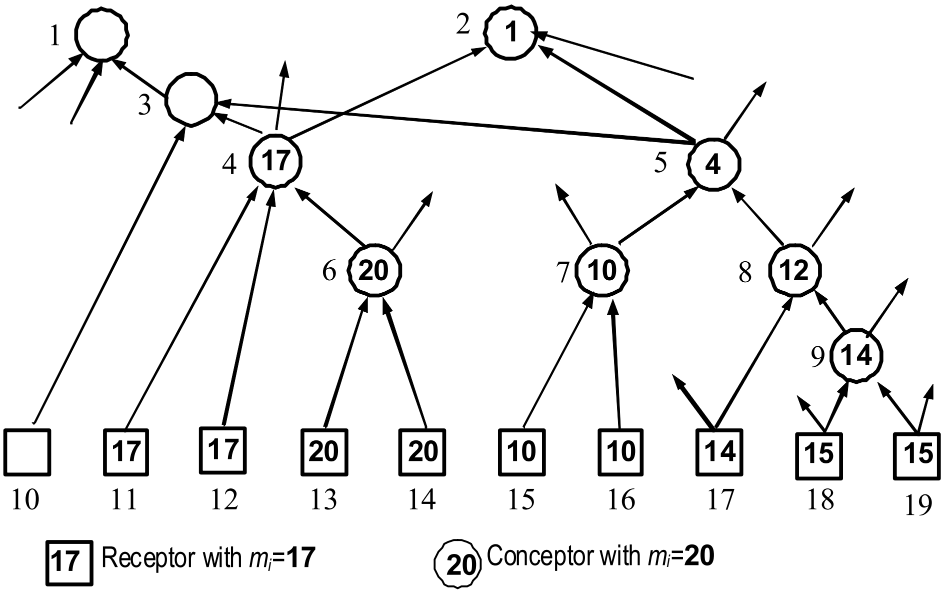


Figure 172. Forming of pyramidal growing network – rule B1

Rule B2. If the pyramid of an object from concept volume A_i contains check vertices of other concepts whose supersets do not contain excited check vertices of concept A_i, then in each of these supersets the vertex, having the largest k among all excited vertices with the largest mi, is marked as a check vertex of concept A_i.

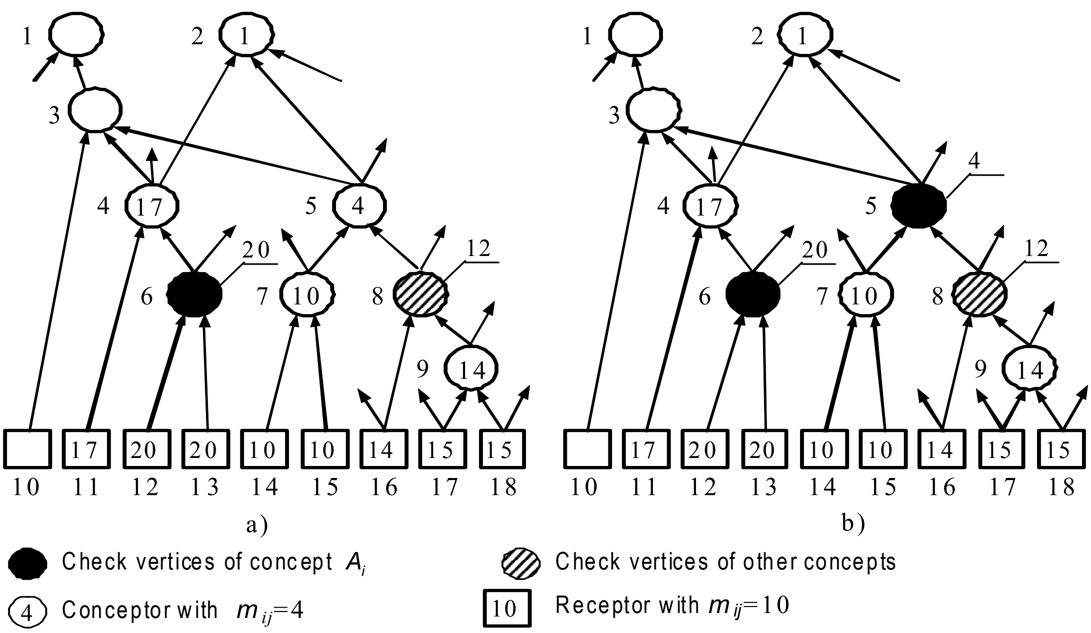


Figure 173. Forming of pyramidal growing network – rule B2

According to this rule the excitation of the pyramid of vertex 2 (Figure 173a) on the condition, that it represents an objects from concept volume A_i, results in choosing vertex 5 as the check vertex of concept A_i (Figure 173b). By check vertexes we select the most typical (having the largest m_i) combinations of attribute values, belonging to objects from concept volume. For example, selecting the vertex 8 as a check vertex means selection of combination of value attributes, corresponding to receptors 17, 18, 19.

If at least one new check vertex appears while scanning objects of the training set, i.e. conditions of Rules B1 or B2 have been performed once at least, the training set is rescanned. The algorithm stops if during the scanning of the training set no new check vertex appears.

14.2.2 Recognition on basis of GPN

The task of recognition is based on the following rule.

Certain object belongs to the concept volume A_i if its pyramid has check vertexes A_i and does not contain check vertices of any other concept not having excited check vertices of concept A_i concept in their supersets. If this condition does not hold for any of the concepts, the object is referred to as unrecognized.

The execution time of the above algorithm is always finite. If the volumes of the formed concepts $V_1, V_2, \ldots, V_i, \ldots V_n$ do not intersect, than after execution the algorithm the recognition rule completely divides the training set into subsets $L_i = V_i \cap L\left(i = 1 \ldots n\right)$

The formed concepts are represented in the network as ensembles of check vertices.

There is an algorithm [Gladun, 1987] of composing the logic descriptions of concepts, formed in the network as a result of the training process, described above. The formed logical expression contains logical relations, represented by allocation of check vertices, describing the concepts in the network, defining different classes of objects.

For example, the concept, presented on Figure 173b check vertices with numbers 5 and 6, by following expression is described:

$$(12\wedge 13)\vee(14\wedge 15)\wedge\neg(16\wedge 17\wedge 18).$$

The analytical tasks, such as diagnostics or forecasting, can be reduced to the task of classification, i.e. to belongings the research object to a class of objects, with a property characteristic or a set of properties significant for prognosis

Classification of new objects is performed by comparing the attribute descriptions with the concept, defining a class of predictable or diagnosing objects. Objects can be classified by evaluating the value of the logical expressions that represent corresponding concepts. The variables, corresponding to the attribute values of the recognized object, set 1, other variable set 0. If the entire expression possesses the value 1 which means the object is included into volume of concept.

In concept, which is formed by algorithm, the general essential attributes of objects from volume of concept and logic relations between attributes are reflected. Unifying attributes are allocated as a result of performance of rule B1. At performance of rule B2 disjunction attributes are allocated.

An important distinction of a method of concepts formation in growing pyramidal networks is the possibility to introduce in concepts the so-called excluding attributes which do not correspond to objects of a researched class. As a result, the formed concepts have more compact logic structure, which allows increasing the accuracy of diagnosis or forecasting. In logic expression the excluding attributes are presented by variables with negation.

All search operations in growing pyramidal network are limited to rather small fragment of a network, which includes an object pyramid and vertices directly linked to it. As a result, we have a possibility solve practical analytical problems based on large-scale data.

14.3 Program complex CONFOR

Methods for solution of regularities discovery tasks based on pyramidal networks, and methods of using of the retrieved regularities for decision making described in the previous section are implemented in program complex CONFOR (Abbreviation of CONcept FORmation). In the case of decision-making in risk management, the described objects are assigned to specific disasters and / or emergent situations. This makes it possible to apply universal approach of growing pyramidal networks to analysis of attributive risk management and disaster emergencies.

Let us consider briefly basic functions and structure of a program complex.

14.3.1 Architecture of CONFOR

The basic functions of program complex CONFOR are:

- discovery of regularities (knowledge) inherent to data;
- using of the retrieved regularities for object classification, diagnostics and prediction.

Main program unit and interrelations between subsystems are presented on Figure 174.

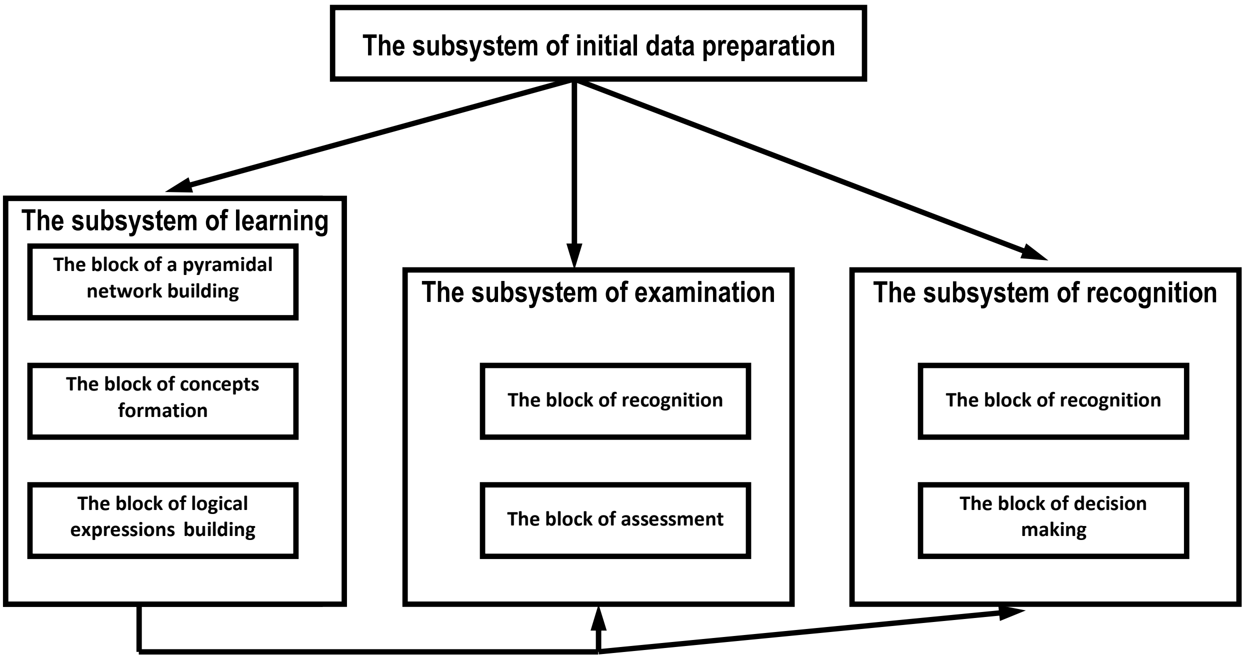


Figure 174. The structure of program system CONFOR

The program complex CONFOR includes following subsystems:

1. Subsystem of initial data preparation.
1. Learning subsystem.
2. Examination subsystem.
3. Recognition subsystem.

14.3.2 Initial data preparation subsystem

The subsystem of data preparation can be functionally extracted in the separate block though it is an integral part of subsystems of training, examination, and recognition. The subsystem realizes multitude of the operations, allowing entering attributive descriptions of situations directly from the screen or from the previously prepared text file.

The subsystem will transform the entered information in the form of training, examination, or recognition set, to internal representation, which is used for construction or the analysis of a pyramidal network.

Objects of the training, examination, or recognition set are represented by attribute descriptions, i.e. by sets of attribute values.

Before operating with the system, it is necessary to:

- choose the set of attributes that will be used for describing investigated objects;
- specify for every attribute a set of its values. Attribute values can be given in numerical, Boolean or nominal scales. Numerical attributes must be discretized, that is divided into subintervals;
- describe objects using the chosen attribute values, that is for every object specify its value for every attribute;
- specify for every object (of the training and examination set) the class name that the object belongs to.

The entered attribute descriptions of situations are displayed on the screen in the form of the table, owing to what it is convenient for supervising and editing.

Input attribute descriptions of situations from the screen is provided with a following set of functions: Add Column, Add Row, Delete Column, Delete Row, Rename Column.

Input attribute descriptions of situations from the previously prepared text file is carried out by means of function "Import". The test file can be prepared outside of a tool by means of the text editor in the form of a *.txt-file according to strictly certain syntax: the first line – heading; lines-descriptions of objects further follow; each line comes to an end with a symbol "line feed"; elements of a line are separated by a symbol "a comma (,)"; unknown values of attributes fall, but the divider (,) is put, i.e. empty value is designated as ","; for a designation of a fractional part of a quantitative attribute the "point (.)" is used.

After input of descriptions of situations (from the screen or from a file), the columns containing a name of object (situation) and a name of a class (type of situation) should be marked in appropriate way. With this purpose, next functions are used: to choose a column "Object"; to choose a column "Class".

The specified functions expand opportunities of a tool as research system, allowing investigating the same training set from the various points of view, easily changing an attribute, which serves as a name of a class.

14.3.3 Learning subsystem

Subsystem of learning realizes the discovery of the regularities that characterize the class of emergency. Input data for a subsystem of learning is served training set, which includes examples of the situations, described as various types (classes). Each class of situations should be presented by a quantity of examples; sufficient that based on them the regularity describing the given class has been allocated. Examples of situations are represented by set of attributive descriptions. Attributive description of a situation should include a name of a situation, a class to which the situation belongs, and a set of attributes values, which characterize a situation.

Output data for a subsystem of learning is served regularities, which characterize classes of disasters situations.

Process of learning consists of following stages:

- representation of initial data in the form of a pyramidal network;
- concepts formation on the basis of a pyramidal network;
- construction of the logic expressions corresponding generated concepts.

✓ Block of pyramidal network building

The block of a pyramidal network building realizes the first stage of process of training when internal representation of training set objects will be transformed to a pyramidal network.

As the basic process at regularities extraction is search of combinations of the attributes values describing groups of similar objects, as a rule, it is necessary to look through repeatedly objects training set that with growth of objects number and of number of attributes values leads to "information explosion". Representing of data in the form of a pyramidal network allows to avoid this danger, first of all, due to features of algorithm of the network building providing allocation of the common combinations of attributes values during input of objects in a network, and also due to associativity and hierarchy of the network. Only two viewings of training set are necessary for full a network building.

The pyramidal network is dynamic structure, which is restructured depending on the incoming information. The network consists of two types vertices: receptors and conceptors. Receptors correspond to values of attributes. In various tasks, they can be represented by names of properties, relations, states of the equipment, a situation, actions, objects, or classes of objects. Receptors have no input connections.

Conceptors correspond to crossings of objects descriptions and to descriptions of objects in general and represent GPN vertices. Conceptors of the first type is named as intermediate vertices and have input and output connections. Conceptors of the second type – the main vertices, which can have output connections only in that case when the description of the object, corresponding the given main vertex, is a part of the description of some other object.

The main process of a pyramidal network building, at which in a network the common combinations of attributes values (all intermediate vertices are building) are fixed, is realized even at the first viewing of objects descriptions of training set. At the second viewing, the network only is corrected to provide unequivocal representation of each object in the form of a separate pyramid (with one main vertex). At network building objects of training set are entered into a network serially and thus the description of the next object is compared to those objects of training set which already have been entered into a network. Comparison is carried out by tracing output connections of receptors from the description of considered object in a direction to the main vertices of pyramids. Tracing is carried out by consecutive transition from the analysis in vertex of lower level of a network to the analysis of higher level.

During tracing it is formed and corrected TBC (the table of target communications) in which pairs analyzed units in the form of "the subordinated vertex – the subordinating vertex" are fixed. The "excited" fragment of a network, which is already constructed fragment of a network, which includes receptors from the description of considered object, is as a result allocated.

Thus are fixed both completely excited vertices of a network, and partially excited. Vertex at which all input connections conduct from receptors from the description of selected object is considered as completely excited. At partially excited vertex from receptors from the description of selected object, two input connections conduct, at least. Other input connections can conduct from receptors, which correspond to the values of attributes not inherent in selected object.

If on a way of tracing of receptors output connections there are completely excited vertex, it means, that these vertex should be included in a pyramid of selected object, and process of tracing proceeds. At detection of partially excited vertex process of tracing in this direction is stopped. Process of tracing finally stops, when partially excited vertices and the "highest" completely excited vertices are revealed and fixed all. "Highest" completely excited vertices are named such completely excited vertices from which do not conduct output connections to other completely excited vertices.

As to the beginning of the second viewing in a network intermediate and main vertices are already constructed all, at this stage also calculation m_i is carried out for each vertex (m_i is the number of excitation for each i-th class). Number of excitation of vertex for some class to equally number of objects of the given class in which pyramids there is this vertex.

- ✓ **Block of concepts formation**

The block of concepts formation realizes process of regularities discovery proper. This block on the basis of the analysis of the constructed pyramidal network selects from the combinations of attributes values most essential to everyone class and determines connections between them. Vertices, which correspond to the selected combinations, were named control vertices.

The formed concepts are represented in the network structure as ensembles of check vertices, which belong to corresponding classes.

Process of concepts formation is carried out by consecutive updating formed concepts on each object of training set.

In spite of the fact that during updating training set is looked through some times, this process is carried out much more quickly, than process of a network building because each time the area of viewing is limited only to a pyramid of object. Pyramids are looked through "from top to down", i.e. in a direction from the main vertices to receptors. Process ends, when there are no conditions for occurrence of new check vertices, i.e. objects of training set are divided completely.

- ✓ **Block of logical expressions building**

The block of logical expressions building makes it possible to represent the generated concepts in the form of logic expressions. Each logical expression, which corresponds to concept, has as operands of attributes value by means of which situations were described, and as operators – conjunction, disjunction, and negation. Representation of concept in the form of logical expression is evident, is well interpreted, and can be used by the expert for the analysis with the purpose of deeper understanding of regularities, which are inherent in a object domain. Logical expressions' building is carried out consistently for each class, which objects are presented in training set. Generated logic expressions are written in a text file, which can be screened by a special command, or is read by means of a text editor.

After adjustment for the next class all check vertices of the given class are analyzed. Check vertices of a considered class are named by "positive" check vertices, all check vertices of other classes are considered "negative" in relation to a considered class. Construction of logic expression begins with ordering positive check vertices of a considered class in decreasing order m – their excitation numbers for the given class. Each "positive" control unit is a basis for formation of a corresponding disjunctive member of the logic expression, representing concept of the considered class.

Formation of each disjunctive member begins with even not considered check vertices with the greatest m. First of all in a text file the is written number m which corresponds to number of objects of the training set belonging the given class and containing in its description receptors, corresponding the chosen check vertex. Then in a text file the receptors of the pyramid of the selected node are written out, for which the pyramid is scanned in the direction "from top to down". In a text file receptors are bound by symbol of conjunction (&). Such conjunction we shall name base conjunction for a formed disjunctive member.

Further so-called conjunction-exceptions form for what in superset of positive check vertices search of the nearest negative check vertices is carried out. Nearest negative checked vertices are the vertices on a path to which from positive check vertices, there are no other check vertices. As negative checked vertices the vertices belonging to any other class are considered.

If those are not present in a network, in a text file the symbol of a disjunction (V) enters and formation of the next disjunctive member begins. If in superset of positive check vertex there are negative check vertices, formation of conjunction-exception proceeds by tracing out the receptors entering into a pyramid of negative check vertex, but without taking into account receptors which have entered into a pyramid of positive check vertex.

Written out receptors are united by the symbol of conjunction (V), undertake in brackets and join with earlier generated part of logical expression through symbols of conjunction (&) and negation (¬). The analysis of negative check vertices and formation of corresponding conjunction-exceptions also is carried out in order decreasing m, describing a class corresponding negative check vertex. In this case formation of a disjunctive member comes to end after the analysis of all check vertices which are negative in relation to the given positive check vertex.

After all positive check vertices of considered class are analyzed; changeover to formation of logic expression for a next class is carried out.

14.3.4 Examination subsystem

The subsystem of examination is intended for testing quality of a tool training and quality of training set. The quality of tool training depends from:

- qualities of tools used for training;
- a material for training, i.e. structure of objects of training set and a manner of their description.

As the algorithm of concepts formation based on a pyramidal network provides 100% division of learning set, testing of quality of tool training is reduced to check quality of training set by recognition of examination set objects. Comparison of objects recognition results of examination set to the information on a real accessory of situations to classes allows to judge about quality of a complex training.

The subsystem of examination includes following blocks:

- **block of recognition analyzing situations**, which do not enter into training set but for which their accessory to one of investigated classes is known;
- **block of an assessment**, which gives out the information on quantity of correct, wrong and uncertain answers of a subsystem.

The important feature of realizable process of recognition is the opportunity to give out uncertain answers when recognizable objects contain in the description a combination of receptors, characteristic simultaneously for different classes, or when recognizable objects are not similar to objects of training sample. A large number of incorrect and uncertain answers of a subsystem demonstrate the necessity continuation of the learning by improving of training set.

Process of recognition can be carried out both based on the analysis of the trained pyramidal network, and by means of the constructed logic expressions.

The block of an assessment compares with the results received at recognition of objects of training set, with the information on a real accessory of objects to classes and gives out a percentage of correct, wrong, and uncertain answers of a subsystem.

14.3.5 Recognition subsystem

The subsystem of recognition realizes second of the basic functions of a complex, namely, use of the regularities allocated in a learning stage for classification of new situations and outputting of the control decision on elimination of an unforeseen contingency.

The subsystem of recognition consists of following blocks:

- **block of recognition**, which allows to classify a new situation to one or another class;
- **block of decision making**, offering to the operator the recommendation at the choice of the operating decision with the purpose of normalization the contingency.

For recognition in a subsystem is used the same block, as in a subsystem of examination. The block of decision-making, as well as the block of assessment of examination results, can give out both exact and uncertain answers. As each class of objects of training set represents the contingency, identification of a new contingency is unequivocally connected with sequence of control actions on its normalization. Identification of an accessory of a new contingency to some class is the operator prompting. Operator makes the final decision on a choice of actions on normalization of a situation.

In case of the uncertain answer the subsystem gives out the additional information on that, how much distinguished situation is similar to the situations corresponding different classes, or absolutely not similar to situations from training set.

For an estimation of a similarity degree the function of confidence is used. This function is calculated based on the analysis of conjunctions involved in the recognition of this situation. Function of confidence reflects a percentage parity of conjunctions informativity, describing the regularities of different classes of situations.

14.4 Discussion

The main characteristic of the pyramidal networks is the possibility to change their structure according to structure of the incoming information. Unlike the neural networks, the adaptation effect is attained without introduction of a priori network excess. Pyramidal networks are convenient for performing different operations of associative search. Hierarchical structure of the networks, which allows them to reflect the structure of composing objects and gender-species bonds naturally, is an important property of pyramidal networks. The concept of GPN is a generalized logical attributive model of objects' class, and represents the belonging of objects to the target class in accordance with some specific combinations of attributes. By classification manner, GPN is closest to the known methods of data mining as decision trees and propositional rule learning.

15
Multi-dimensional Information Spaces as Memory Structures for Intelligent Data Processing in GMES

15.1 Memory management

Memory management is a complex field of computer science. Over the years, many techniques have been developed to make it more efficient [Ravenbrook, 2010]. Memory management is usually divided into three areas: hardware, operating system, and application, although the distinctions are a little fuzzy. In most computer systems, all three are present to some extent, forming layers between the user's program and the actual memory hardware:

- **Memory management at the hardware level** is concerned with the electronic devices that actually store data. This includes things like RAM and memory caches;
- **Memory in the operating system** must be allocated to user programs, and reused by other programs when it is no longer required. The operating system can pretend that the computer has more memory than it actually does, and that each program has the machine's memory to itself. Both of these are features of virtual memory systems;
- **Application memory management** involves supplying the memory needed for a program's objects and data structures from the limited resources available, and recycling that memory for reuse when it is no longer required. Because in general, application programs cannot predict in advance how much memory they are going to require, they need additional code to handle their changing memory requirements.

Application memory management combines two related tasks:

- **Allocation**: when the program requests a block of memory, the memory manager must allocate that block out of the larger blocks it has received from the operating system. The part of the memory manager that does this is known as the allocator;
- **Recycling**: when memory blocks have been allocated, but the data they contain is no longer required by the program, the blocks can be recycled for reuse. There are two approaches to recycling memory: either the programmer must decide when memory can be reused (known as manual memory management); or the memory manager must be able to work it out (known as automatic memory management).

The progress in memory management gives the possibility to allocate and recycle not directly blocks of the memory but structured regions or fields corresponding to some types of data. In such case we talk about corresponded "***access methods***".

15.2 Access methods

The Access Methods (AM) have been available from the beginning of the development of computer peripheral devices. As many devices so many possibilities for developing different AM there exist. In the beginning, the AM were functions of the Operational Systems Core or so called Supervisor, and were executed via corresponding macro-commands in the assembler languages [Stably, 1970] or via corresponding input/output operators in the high level programming languages like FORTRAN, COBOL, PL/I, etc.

The establishment of the first databases in the sixties of the previous century caused gradually accepting the concepts "physical" as well as "logical" organization of the data [CODASYL, 1971], [Martin, 1975]. In 1975, the concepts "access method", "physical organization" and "logical organization" became clearly separated.

In the same time, Christopher Date specially remarked:

"The Database Management System (DBMS) does not know anything about:

a) how physical records (blocks) are disposed;
b) how the stored fields are integrated in the records (nevertheless that in many cases it is obvious because of their physical disposition);
c) how the sorting is realized (for instance it may be realized on the base of physical sequence, using an index or by a chain of pointers);
d) how the direct access is realized (i.e. by index, sequential scanning or hash addressing).

This information is a part of the structures for data storing but it is used by the access method but not by the DBMS" [Date, 1975].

Every access method presumes an exact organization of the file, which it is operating with and is not related to the interconnections between the files, respectively – between the records of one file and that in the others files. These interconnections are controlled by the physical organization of the DBMS.

Therefore, in the DBMS we may distinguish four levels:

1. Access methods at the core (supervisor) of the operation system.
2. Specialized access methods which upgrade these at the core of the operating system.
3. Physical organization of the DBMS.
4. Logical organization of the DBMS.

During the eighties, the total growing of the research and developments in the computers' field, especially in image processing, data mining and mobile support cause impetuous progress of establishing convenient "spatial information structures" and "spatial-temporal information structures" and corresponding access methods. From different points of view, this period has been presented in [Ooi et al, 1993], [Gaede and Günther, 1998], [Arge, 2002], [Mokbel et al, 2003], [Moënne-Loccoz, 2005]. Usually the "one-dimensional" (linear) AM are used in the classical applications, based on the alphanumerical information, whereas the "multi-dimensional" (spatial) methods are aimed to serve the work with graphical, visual, multimedia information.

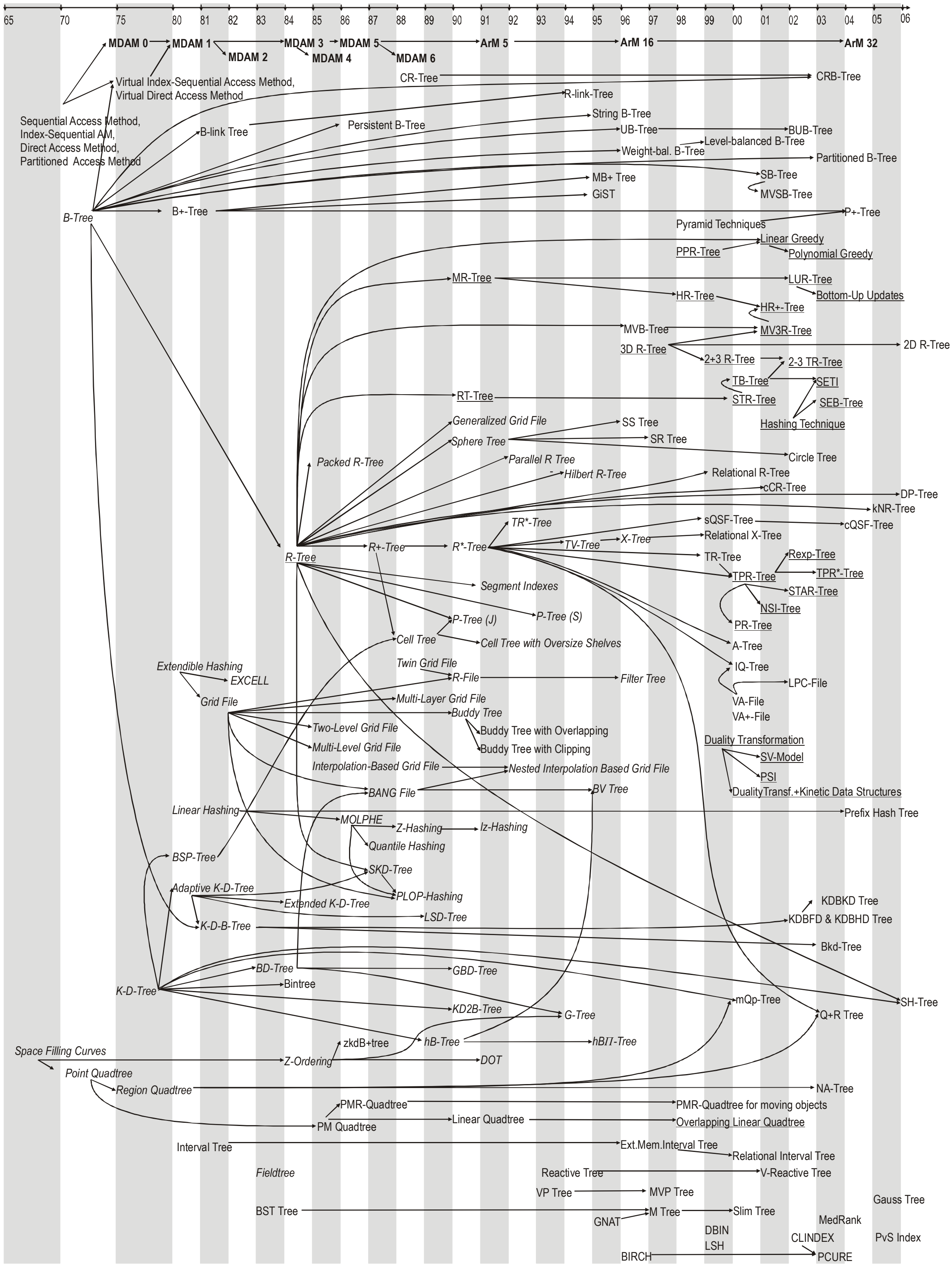


Figure 175. Genesis of the Access Methods and their modifications
extended variant of [Gaede and Günther, 1998] and [Mokbel et al, 2003] presented in [Markov et al, 2008]

15.2.1 Interconnections between raised access methods

Maybe one of the most popular analyses is given in [Gaede and Günther, 1998]. The authors presented a scheme of the genesis of the basic multi-dimensional AM and theirs modifications. This scheme firstly was proposed in [Ooi et al, 1993] and it was expanded in [Gaede and Günther, 1998]. An extension in direction to the multi-dimensional spatio-temporal access methods was given in [Mokbel et al, 2003].

The survey [Markov et al, 2008] presents a new variant of this scheme (Figure 175), where the new access methods, created after 1998, are added. A comprehensive bibliography of corresponded articles, where the methods are firstly presented is given.

15.2.2 The taxonomy of the access methods

From the point of view of the served area, the access methods, presented on Figure 175, may be classified as follows (Figure 176): One-dimensional AM; Multidimensional Spatial AM; Metric Access Methods; High Dimensional Access Methods; and Spatio-Temporal Access Methods.

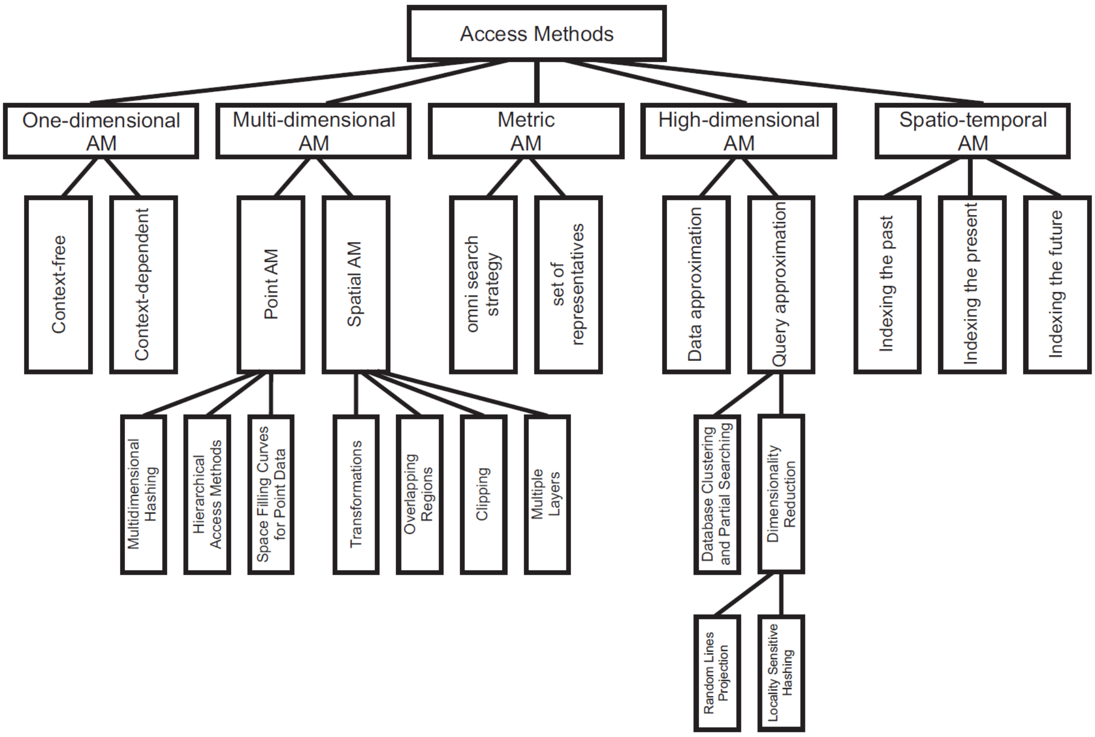


Figure 176. Taxonomy of the Access Methods

✓ One-dimensional access methods

One-dimensional AM are based on the concept "record". The "record" is a logical sequence of fields, which contain data eventually connected to unique identifier (a "key"). The identifier (key) is aimed to distinguish one sequence from another [Stably, 1970]. The records are united in the sets, called "files". There exist three basic formats of the records – with fixed, variable, and undefined length.

In the ***context-free methods,*** the storing of the records is not connected to their content and depends only on external factors – the sequence, disk address, or position in the file. The necessity of stable file systems in the operating systems does not allow a great variety of the context-free AM.

There are three main types well known from sixties and seventies: *Sequential Access Method (SAM); Direct Access Method (DAM) and Partitioned Access Method (PAM)* [IBM, 1965-68].

The main idea of the ***context-depended AM*** is that a part of the record is selected as a key, which is used for making decision where to store the record and how to search it. This way the content of the record influences the access to the record.

Historically, from the sixties of the previous century on, the attention is directed mainly to this type of AM. Modern DBMS are built using context-depended AM such as: unsorted sequential files with records with keys; sorted files with fixed record length; static or dynamic hash files; index file and files with data; clustered indexed tables [Connolly and Begg, 2002].

✓ Multidimensional spatial access methods

Multidimensional Spatial Access Methods are developed to serve information about spatial objects, approximated with points, segments, polygons, polyhedrons, etc. The implementations are numerous and include traditional multi-attributive indexing, geographical and/or GMES information systems, and spatial databases, content indexing in multimedia databases, etc.

From the point of view of the spatial databases, the access methods can be split into two main classes of access methods – Point Access Methods and Spatial Access Methods [Gaede and Günther, 1998].

Point Access Methods are used for organizing multidimensional point objects. Typical instances are traditional records, where every attribute of the relation corresponds to one dimension. These methods can be separated in three basic groups:

- Multidimensional Hashing (for instance Grid File and its varieties, EXCELL, Twin Grid File, MOLPHE, Quantile Hashing, PLOP-Hashing, Z-Hashing, etc);
- Hierarchical Access Methods (includes such methods as KDB-Tree, LSD-Tree, Buddy Tree, BANG File, G-Tree, hB-Tree, BV-Tree, etc.);
- Space Filling Curves for Point Data (like Peano curve, N-trees, Z-Ordering, etc).

Spatial Access Methods are used for working with objects, which have an arbitrary form. The main idea of the spatial indexing of non-point objects is to use an approximation of the geometry of the examined objects as more simple forms. The most used approximation is Minimum Bounding Rectangle (MBR), i.e. minimal rectangle, which sides are parallel of the coordinate axes and completely include the object. There exist approaches for approximation with Minimum Bounding Spheres (SS Tree) or other polytopes (Cell Tree), as well as their combinations (SR-Tree).

The usual problem when one operates with spatial objects is their overlapping. There are different techniques to avoid this problem. From the point of view of the techniques for the organization of the spatial objects, Spatial Access Methods can be split in four main groups:

- Transformation – this technique uses transformation of spatial objects to points in the space with more or less dimensions. Most of them spread out the space using space filling curves (Peano Curves, z-ordering, Hilbert curves, Gray ordering, etc.) and then use some point access method upon the transformed data set;
- Overlapping Regions – here the data sets are separated in groups; different groups can occupy the same part of the space, but every space object associates with only one of the groups. The access methods of this category operate with data in their primary space (without any transformations) eventually in overlapping segments. Methods which use this technique includes R-Tree, R-link-

Tree, Hilbert R-Tree, R*-Tree, Sphere Tree, SS-Tree, SR-Tree, TV-Tree, X-Tree, P-Tree of Schiwietz, SKD-Tree, GBD-Tree, Buddy Tree with overlapping, PLOP-Hashing, etc.;

- Clipping – this technique uses the clipping of one object to several sub-objects, which will be stored. The main goal is to escape overlapping regions. However, this advantage can lead to the tearing of the objects, extending the resource expenses, and decreasing the productivity of the method. Representatives of this technique are R+-Tree, Cell-Tree, Extended KD-Tree, Quad-Tree, etc.;
- Multiple Layers – this technique can be considered as a variant of the techniques of Overlapping Regions, because the regions from different layers can overlap. Nevertheless, there exist some important differences: first – the layers are organized hierarchically; second – every layer splits the primary space in a different way; third – the regions of one layer never overlaps; fourth – the data regions are separated from the space extensions of the objects. Instances for these methods are Multi-Layer Grid File, R-File, etc.

✓ Metric access methods

Metric Access Methods deal with relative distances of data points to chosen points, named anchor points, vantage points or pivots [Moënne-Loccoz, 2005]. These methods are designed to limit the number of distance computation, calculating first distances to anchors, and then finding the searched point in a narrowed region. These methods are preferred when the distance is highly computational, as e.g. for the dynamic time warping distance between time series. Representatives of these methods are: Vantage Point Tree (VP Tree), Bisector Tree (BST-Tree), Geometric Near-Neighbor Access Tree (GNNAT), as well as the most effective from this group – Metric Tree (M-Tree) [Chavez et al, 2001].

✓ High dimensional access methods

Increasing the dimensionality strongly aggravates the qualities of the multidimensional access methods. Usually these methods exhaust their possibilities at dimensions around 15. Only X-Tree reaches the boundary of 25 dimensions, after which this method gives worse results then sequential scanning [Chakrabarti, 2001].

The exit of this situation is based on the data approximation and query approximation in sequential scan. These methods form a new group of access methods – High Dimensional Access Methods.

Data approximation is used in VA-File, VA+-File, LPC-File, IQ-Tree, A-Tree, P+-Tree, etc.

For query approximation, two strategies can be used:

- Examine only a part of the database, which is more probably to contain the resulting set – as a rule these methods are based on the clustering of the database. Some of these methods are: DBIN, CLINDEX, PCURE;
- Splitting the database to several spaces with fewer dimensions and searching in each of them. Here two main methods are used:
    1. Random Lines Projection. Representatives of this approach are MedRank, which uses B+-Tree for indexing every arbitrary projection of the database, and PvS Index, which consist of combination of iterative projections and clustering.

2. Locality Sensitive Hashing, which is based on the set of local-sensitive hashing functions [Moënne-Loccoz, 2005].

✓ **Spatio-temporal access methods**

The Spatio-Temporal Access Methods have additional defined time dimensioning [Mokbel et al, 2003]. They operate with objects, which change their form and/or position during the time. According to position of time interval in relation to present moment, the Spatio-Temporal Access Methods are divided as follow:

- ***indexing the past***. These methods operate with historical spatio-temporal data. The problem here is the continuous increase of the information over time. To overcome the overflow of the data space two approaches are used – sampling the stream data at certain time position or updating the information only when data is changed. Spatio-temporal indexing schemes for historical data can be split in three categories:
  - the first category includes methods that manage spatial and temporal aspects into already existing spatial methods;
  - the second category can be explained as snapshots of the spatial information in each time instance;
  - the third category focuses on trajectory-oriented queries, while spatial dimension lag on second priority.

  Representatives of this group are: RT-Tree, 3DR-Tree, STR-Tree, MR-Tree, HR-Tree, HR+-Tree, MV3R-Tree, PPR-Tree, TB-Tree, SETI, SEB-Tree;
- ***indexing the present***. In contrast to previous methods, where all movements are known, here the current positions are neither stored nor queried. Some of the methods, which answer the questions of the current position of the objects are 2+3R-Tree, 2-3TR-Tree, LUR-Tree, Bottom-Up Updates, etc.;
- ***indexing the future***. These methods have to answer the questions about the current and future position of a moving object – here are embraced the methods like PMR-Quadtree for moving objects, Duality Transformation, SV-Model, PSI, PR-Tree, TPR-Tree, TPR*-tree, NSI, VCIR-Tree, STAR-Tree, R^{EXP}-Tree.

The survey of the access methods suggests that the context-free multi-dimensional access methods practically are not available. A step in developing such methods is the Multi-domain Information Model and corresponding Multi-domain Access Method introduced in [Markov, 1984] [Markov, 2004]. It will be outlined later in this chapter.

15.3 Multi-dimensional numbered information spaces

A simple idea is coming from this reasoning: "Why do we have no possibility to use "external memory arrays" with one, two, three, and more dimensions?" [Markov, 2004]. Additionally, by analogy with the Relation Model [Codd, 1970], we may want to have multidimensional relations. Such theory still does not exist, but it is possible to be realized.

The proposed external memory structure is based on the numbering as a main approach. The idea consists in replacing the (symbol or real; point or interval) values of the objects' attributes with integer numbers of the elements of corresponding ordered sets. This way, each object will be described by a vector of integer values, which may be used as the co-ordinate address in the multi-dimensional information space.

In other words, the process of replacing the names by numbers permits the use of mathematical functions and address vectors for accessing the information instead of search engines.

This type of memory organization is called "Multi-dimensional numbered information spaces". Its advantages have been demonstrated in many practical realizations during more than twenty-five years [Markov, 1984], [Markov, 2004], [Markov, 2005]. In recent years, this kind of memory organization has been implemented in the area of intelligent systems memory structuring for several data mining tasks and especially in the area of association rules mining.

15.3.1 Multi-Domain Information Model (MDIM)

The independence of dimensionality limitations is very important for developing new intelligent systems aimed to process high-dimensional data. To achieve this we need information models and corresponding access method to cross the boundary of the dimensional limitations and to obtain the possibility to work with information spaces with practically unlimited number of dimensions. The first step is to establish context free multi-dimensional models and based on it to develop high-level context depended applications. Examining the state of the art in this area shows that the context-free multi-dimensional information models and access methods practically are not available. One attempt in this direction is establishing the **Multi-Domain Information Model** (**MDIM**) [Markov, 2004] and the corresponding Multi-domain Access Method. Their possibilities for operating with context-free multidimensional data structures will be presented below.

15.3.2 Basic structures of MDIM

Basic structures of MDIM are basic information elements, information spaces, indexes and metaindexes, and aggregates. The definitions of these structures are given below:

✓ Basic information element

The basic information element (BIE) of MDIM is an arbitrary long string of machine codes (bytes). When it is necessary, the string may be parceled out by lines. The length of the lines may be variable.

Let the universal set ***UBIE*** be the set of all BIE.

Let $\boldsymbol{E}_1$ be a set of basic information elements:

$$\boldsymbol{E}_1 = \{e_i \mid e_i \in \boldsymbol{UBIE}, i=1,\ldots, m_1\}.$$

Let μ_1 be a function, which defines a biunique correspondence between elements of the set $\boldsymbol{E}_1$ and elements of the set $\boldsymbol{C}_1$ of positive integer numbers:

$$\boldsymbol{C}_1 = \{c_i \mid c_i \in \boldsymbol{N}, i:=1,\ldots, m_1\},$$

$$\text{i.e. } \mu_1 : \boldsymbol{E}_1 \leftrightarrow \boldsymbol{C}_1.$$

The elements of $\boldsymbol{C}_1$ are said to be numbers (co-ordinates) of the elements of $\boldsymbol{E}_1$.

✓ Information spaces

The triple $\mathbf{S_1} = (\mathbf{E_1}, \mu_1, \mathbf{C_1})$ is said to be a ***numbered information space of range 1*** *(one-dimensional or one-domain information space).*

The triple $\mathbf{S_2} = (\mathbf{E_2}, \mu_2, \mathbf{C_2})$ is said to be a ***numbered information space of range 2*** iff $\mathbf{E_2}$ is a set whose elements are numbered information spaces of range 1 and μ_2 is a function which defines a biunique correspondence between elements of $\mathbf{E_2}$ and elements of the set $\mathbf{C_2}$ of positive integer numbers:

$$\mathbf{C_2} = \{c_j \mid c_j \in \mathbf{N}, j:=1,\ldots,m_2\},$$

$$\text{i.e. } \mu_2 : \mathbf{E_2} \leftrightarrow \mathbf{C_2}.$$

The triple $\mathbf{S_n} = (\mathbf{E_n}, \mu_n, \mathbf{C_n})$ is said to be a ***numbered information space of range n*** *(n-dimensional or multi-domain information space)* iff $\mathbf{E_n}$ is a set whose elements are information spaces of range **n-1** and μ_n is a function which defines a biunique correspondence between elements of $\mathbf{E_n}$ and elements of the set $\mathbf{C_n}$ of positive integer numbers:

$$\mathbf{C_n} = \{c_k \mid c_j \in \mathbf{N}, k:=1,\ldots,m_n\}, \text{ i.e. } \mu_n : \mathbf{E_n} \leftrightarrow \mathbf{C_n}.$$

Every basic information element "e" is considered as an ***information space*** $\mathbf{S_0}$ of range 0. It is clear that the information space $\mathbf{S_0} = (\mathbf{E_0}, \mu_0, \mathbf{C_0})$, is constructed in the same manner as all others:

- the machine codes (bytes) b_i, i=1,…,m_0 are considered as elements of $\mathbf{E_0}$,
- the position p_i (natural number) of b_i in the string e is considered as co-ordinate of b_i, i.e. $\mathbf{C_0} = \{p_k \mid p_k \in \mathbf{N}, k:=1,\ldots,m_0\}$,
- function μ_0 is defined by the physical order of b_i in e and we have:

$$\mu_0 : \mathbf{E_0} \leftrightarrow \mathbf{C_0}$$

In this way, the string $\mathbf{S_0}$ may be considered as a set of ***sub-elements (sub-strings)***. The number and length of the sub-elements may be variable. This option is very helpful but it closely depends on the concrete realizations and it is not considered as a standard characteristic of MDIM.

The information space $\mathbf{S_n}$, which contains all information spaces of a given application is called the ***information base*** of range **n**. Usually, the concept information base without indication of the range is used as generalized concept to denote all available information spaces.

✓ Indexes and metaindexes

The sequence $\boldsymbol{A} = (c_n, c_{n-1}, \ldots, c_1)$ where $c_i \in \boldsymbol{C_i}$, i=1,…,n is called the ***multidimensional space address*** of range **n** of a basic information element. Every space address of range **m**, **m<n**, may be extended to space address of range **n** by adding leading **n-m** zero codes. Every sequence of space addresses $\boldsymbol{A_1}, \boldsymbol{A_2}, \ldots, \boldsymbol{A_k}$, where **k** is arbitrary positive number, is said to be a ***space index***.

A special kind of space index is the ***projection***, which is the analytically given space index. There are two types of projections:

- ***hierarchical projection*** – where the top part of coordinates is fixed and the low part vary for all possible values of coordinates, where non-empty elements exist;
- ***arbitrary projection*** – in this case, it is possible to fix coordinates in arbitrary positions and the other coordinates vary for all possible values of coordinates, where non-empty elements exist.

Every index may be considered as a basic information element, i.e. as a string, and may be stored in a point of any information space. In such case, it will have a multidimensional space address, which may be pointed in the other indexes, and, this way, we may build a hierarchy of indexes. Therefore, every index, which points only to indexes, is called ***meta-index***.

The approach of representing the interconnections between elements of the information spaces using (hierarchies) of meta-indexes is called ***polyindexation***.

✓ Aggregates

Let **G** = {$\mathbf{S}_i$ | i=1,..,m} be a set of numbered information spaces.

Let $\tau=\{\nu_{ij} : \boldsymbol{S}_i \rightarrow \boldsymbol{S}_j \mid i=const, j=1,\ldots m\}$ be a set of mappings of one "main" numbered information space $\mathbf{S}_i \subset \mathbf{G}$, i=const, into the others $\mathbf{S}_j \subset \mathbf{G}$, j=1,…m, and, in particular, into itself.

The couple: **Đ** = (**G**, τ) is said to be an "***aggregate***".

It is clear we can build **m** aggregates using the set **G** because every information space $\mathbf{S}_i \subset \mathbf{G}$, j=1,…,m, may be chosen to be a main information space.

15.3.3 Operations in the MDIM

After defining the information structures, we need to present the operations, which are admissible in the model.

It is clear that the operations are closely connected to the defined structures.

In MDIM, we assume that **all** information elements of **all** information spaces **exist**. If for any $\mathbf{S}_i : \mathbf{E}_i=\varnothing \wedge \mathbf{C}_i=\varnothing$, than it is called ***empty.*** Usually, most of the information elements and spaces are empty. This is very important for practical realizations.

✓ Operation with basic information elements

Because of the rule that all the structures given above must exist, we only need two operations: (1) updating and (2) getting the value of BIE.

For both types of operations, we need two service operations: (1) getting the length and (2) the positioning in the BIE.

Updating, or simply – writing the element, has several modifications with obvious meaning: writing a BIE as a whole, appending/inserting in a BIE, cutting/replacing a part of a BIE and deleting a BIE.

There is only one operation for getting the value of a BIE, i.e. **Read** a portion from a BIE starting from given position. We may receive the whole BIE if the starting position is the beginning of BIE and the length of the portion is equal to the BIE length.

✓ Operation with spaces

With a **single space,** we may do only one operation – clearing (deleting) the space, i.e. replacing all BIE of the space with empty BIE – $\varnothing$. After this operation, all BIE of the space will have zero length. Really, the space is cleared via replacing it with empty space.

With **two spaces**, we may provide two operations with two modifications both: (1) copying and (2) moving the first space in the second.

The modifications concern how the BIE in the recipient space are processed. We may have: copy/move with clear and copy/move with merge.

The "clear" modifications first clear the recipient space and after that provide a copy or move operation.

The merge modifications may have two types of processing: destructive or constructive. The ***destructive merging*** may be "conservative" or "alternative". In the conservative approach, the recipient space BIE remain in the result if it is with none zero length. In the other approach – the donor space BIE remain in the result. In the ***constructive merging*** the result is any composition of the corresponding BIE of the two spaces.

Of course, the move operation deletes the donor space after the operation.

✓ Operation with indexes and metaindexes

The indexes are the main approach for describing the interconnections between the structures.

We may receive the space address of the **next** or **previous, empty** or **non-empty** elements of the space starting from any given co-ordinate. This corresponds to the processing of given hierarchical projections.

Analogically, we may receive the space address of the **nextproj** or **previousproj non-empty** elements of the space for the current address in operation with a given arbitrary projection.

The possibility to count the number of non-empty elements of a given projection is useful for practical realizations.

The operations with indexes are based on usual logical operations between sets. The difference from usual sets is that the information spaces are built by the interconnection between two main sets: the set of co-ordinates and the set of information elements.

This way the operations with indexes may be classified in two main types: context free and context depended operations.

The context free operations defined in the MDIM are based on the classical logical operations – intersection, union, and supplement, but these operations are notso trivial. Because of the complexity of the structure of the information spaces, these operations have at least two principally different realizations based on:

- co-ordinates;
- information elements.

The operations based on co-ordinates are determined by the existence of the corresponding space information elements. Therefore, the values of the co-ordinates of the existing information elements determine the operations.

In the other case, the existing BIE values determine the logical operations.

In both cases, the result of the logical operations is an index.

The context depended operations need special realizations for concrete purposes.

The main information operation is creating the indexes and meta-indexes. This may be very complicated processes and could not be given in advance. The main purpose of the MDIM is to provide the possibility for access to the practically unlimited information space and easy approach for building interconnection between its elements. The goal of the concrete applications is to build tools for creating and operating with the indexes and meta-indexes and to implement these tools in the realization of user requested systems.

For instance, such tools may realize the transfer from one structure to another, information search, sorting, making reports, more complicated information processing, etc. The information operations can be grouped into four sets corresponding to the main information structures: basic information elements, information spaces and index or meta-index structures.

✓ Operations with aggregates

Theory of aggregates may be assumed as an extension of the Relation theory because the relation in the sense of the model of Codd [Codd, 1970] may be represented by the aggregate. It is easy to see that if the aggregation mappings are one-one mappings it will be relation in the sense of the model of Codd. Therefore, we may say that the aggregate is a more universal structure than the relation and the operations with aggregates include those of relation theory. The relation algebra is a very good starting point to understand the algebra of aggregates. The new element is that the mappings of different aggregates may be not one-one mappings. This field is not investigated until now.

✓ Multi-domain access method ArM 32

The program realization of MDIM is called Multi-Domain Access Method. For a long period, it has been used as a basis for the organization of various information bases. There exist several realizations of MDIM for different hardware and/or software platforms. The most resent one is the FOI Archive Manager – ArM. One of the first goals of the development of ArM was to represent the digitalized military defense situation, which is characterized by a variety of complex objects and events, which occur in the space and time and have a long period of variable existence. The great number of layers, aspects, and interconnections of the real situation may be represented only by information space hierarchy. In addition, the different types of users with individual access rights and needs insist on the realization of a special tool for organizing such information base. Over the years, the efficiency of ArM is proved in wide areas of information service of enterprise managements and accounting. Organizing the datum in appropriate multi-dimensional information space model permits omitting the heavy work of creating of OLAP structures [Markov, 2005].

The newest ArM Version No.9, called ArM32, is developed for MS Windows, and realizes the proposed algorithms.

The ArM32 elements are organized in numbered information spaces with variable ranges. There is no limit for the ranges of the spaces. Every element may be accessed by a corresponding multidimensional space address (coordinates) given via coordinate array of type cardinal. At the first place of this array, the space range needs to be given. Therefore, we have two main constructs of the physical organizations of ArM32 – numbered information spaces and elements.

In ArM32, the length of the string may vary from 0 up to 1G bytes. There is no limit for the number of strings in an archive but their total length plus internal indexes could not exceed 4G bytes in a single file.

The main ArM32 operations with basic information elements are: ArmRead (reading a part or a whole element); ArmWrite (writing a part or a whole element); ArmAppend (appending a string to an element); ArmInsert (inserting a string into an element); ArmCut (removing a part of an element); ArmReplace (replacing a part of an element); ArmDelete (deleting an element); ArmLength (returns the length of the element in bytes).

The operations over the spaces are: DelSpace (deleting the space); CopySpace and MoveSpace (copying/moving the first space in the second in the frame of one file); ExportSpace (copying one space from one file the other space, which is located in other file).

The operations, aimed to serve the hierarchical projections are: ArmNextPresent, ArmPrevPresent, ArmNextEmpty, ArmPrevEmpty. For arbitrary projections the operations are: ArmNextProj and ArmPrevProj.

The operations, which create indexes, are: ArmSpaceIndex (returns the space index of the non-empty structures in the given information space; ArmProjIndex (gives the space index of basic information elements of a given hierarchical or arbitrary projection).

The service operations for counting non-empty elements or subspaces are correspondingly: ArmSpaceCount (returns the number of the non-empty structures in given information space); ArmProjCount (gives the number of elements of given (hierarchical or arbitrary) projection).

The ArM32 logical operations defined in the multi-domain information model are based on the classical logical operations – intersection, union, and supplement, but these operations are not so trivial. Because of complexity of the structure of the spaces these operations have at least two principally different realizations based on codes of the information spaces' elements and on contents of those elements.

The ArM32 information operations can be grouped into four sets corresponding to the main information structures: elements, spaces, aggregates, and indexes. Information operations are context depended and need special realizations for concrete purposes. Such well-known operations are for instance transferring from one structure to another, information search, sorting, making reports, etc.

At the end, there exist several operations, which serve information exchange between ArM32 archives (files) such as copying and moving spaces from one to another archive.

15.4 Discussion

We need to discuss shortly the main concept we use – the information space. Its main structure is an ordered set of numbered information elements. These elements may be information spaces or terminal elements. Of course, the hierarchical structures are well known. The new aspect of this model is the possibility to connect elements from different spaces and levels of the hierarchy using poly-indexation and in this way to create very large and complex networks with a co-ordinate hierarchical basis.

The variety of interconnections is the characteristic, which permits us to call the ordered set of numbered information elements "Information Space". In the information space, different information structures may exist at the same time in the same set of elements. In addition, the creation and destruction of the link's structures do not change the basic set of elements. The elements and spaces always exist but, in any cases, they may be "empty". At the end, the possibility to use coordinates is good for well-structured models where it is possible to replace search with addressing.

Summarizing, the advantages of the numbered information spaces, are:

- the possibility to build growing space hierarchies of information elements;
- the great power for building interconnections between information elements stored in the information base;
- the practically unlimited number of dimensions (this is the main advantage of the numbered information spaces for well-structured tasks where it is possible "to address, not to search");
- the possibility to create effective and useful tools, in particular for association rule mining.

Below, we show the advantages of using such memory structuring in the field of association rule mining in GMES.

15.5 Data mining analysis environment "PaGaNe"

The authors of this chapter form an international joint research group that work out a design of a data mining analysis environment called "PaGaNe". It contains variety of data mining algorithms, such as association rule miners, class association rule (CAR) algorithms, etc. [Mitov et al, 2009a/b].

The main specificity of PaGaNe is using of the advantages of multi-dimensional numbered information spaces [Markov, 2004], given by the access method ArM 32, such as:

- the possibility to build growing space hierarchies of information elements;
- the great power for building interconnections between information elements stored in the information base;
- the possibility to change searching with direct addressing in well-structured tasks.

The PaGaNe approach is a successor of the main ideas of GPN, such as hierarchical structuring of memory that allows reflecting the structure of composing instances and gender-species connections naturally, convenience for performing different operations of associative search. The recognition is based on reduced search in the multi-dimensional information space hierarchies.

An important idea of the approaches, used in PaGaNe, is replacing the symbol values of the objects' features with integer numbers of the elements of corresponding ordered sets. This way each instance or pattern can be represented by a vector of integer values, which may be used as co-ordinate address in corresponded multi-dimensional information space.

Here we will stop our attention on MPGN algorithm (abbreviation from "**M**ulti-layer **P**yramidal **G**rowing **N**etworks of information spaces"), which is kind of CAR algorithm that use advantages of numbered information spaces in order to overcome bottlenecks of exponential growth of combinations between patterns in the training stage, as well as quickly finding the potential answer in the recognition stage. The main goal is to extend the possibilities of network structures by using a special kind of multi-layer memory structures called "pyramids", which permits defining and realizing of new opportunities.

15.6 CAR algorithm MPGN

15.6.1 Coding convention

Usually in classification tasks rectangular data sets are used, every one of which is a set of instances $\mathbf{R} = \{R^i, i \in 1,...,r\}$. Each instance represent a set of attribute-value pairs $R = \{C = c, A_1 = a_1,..., A_n = a_n\}$. Because in the rectangular data sets the positions of class and attributes are fixed, the instances are written as vectors, which contains only values of attributes: $R = (c, a_1,..., a_n)$.

Every instance has the same quantity of attributes, but some of the values may be omitted. First attribute is the class attribute denoted c; other attributes are input attributes, denoted a_k.

Attribute positions of a given instance, which can take arbitrary values from the attribute domain, are denoted as "-".

Thus each instance (record) is presented as: $R = (c, a_1,..., a_n)$; where n is the number of attributes (feature space dimension), $c \in \mathrm{N}$; $a_k \in \mathrm{N}$ or $a_k = "-"$, $k \in [1,...,n]$.

More precisely, the class values and attribute values receive values, which are natural numbers from 1 to some maximal value (specific for each attribute), i.e. $c \in [1,...,M_c]$, $a_k \in [1,...,M_{a_k}]$.

Pattern is denoted P and has similar structure as instances. A pattern is a subset of an instance.

$$\left.\begin{array}{l} P = (c, a_1, ..., a_n) \\ R = (c, b_1, ..., b_n) \\ a_i = b_i \text{ or } a_i = "-" \end{array}\right\} P \subseteq R .$$

For example $P = (2,3,2,-,2) \subseteq R^1 = (2,3,2,1,2)$.

Every instance is a pattern but not every pattern is an instance, because the attributes with arbitrary values in the pattern may be more than in the instance.

In $P = (c, a_1, ..., a_n)$ usually (c) is called head of a pattern and $(a_1, ..., a_n)$ is called its body.

The cardinality of one pattern is defined as number of "non-arbitrary" attribute values:

$$|P| = \text{number of } a_k \neq "-" ;\ |P| \leq n .$$

For the set of patterns $\mathbf{P} = \{P^i, i \in 1,...,m\}$ we can define maximal cardinality as maximum of cardinalities of patterns in the set.

$$MaxCard(\mathbf{P}) = \max_{i \in 1,...,m} \left|P^i\right| .$$

The intersection between P^i and P^j is the result of matching of these patterns.

$$P^i \cap P^j = (c^l, a_1^l, ..., a_n^l):\ c^l = \begin{cases} c^i : c^i = c^j \\ "-" : c^i \neq c^j \end{cases} \text{ and } a_k^l = \begin{cases} a_k^i : a_k^i = a_k^j \\ "-" : a_k^i \neq a_k^j \end{cases} .$$

If $\left|P^i \cap P^j\right| > 0$ and $c^i = c^j$, then $P^i \cap P^j$ is a pattern, called generalized pattern of P^i and P^j.

For example, the generalized pattern of $R^1 = (2,3,2,1,2)$ and $R^3 = (2,3,2,2,2)$ is $P = (2,3,2,-,2)$.

The contradiction between two patterns means that they have equal attributes (equal bodies) but belong to different classes (different heads).

P^i contradicts to P^j if $P^i = (c^i, a_1, ..., a_n)$ and $P^j = (c^j, a_1, ..., a_n)$, but $c^i \neq c^j$.

The query instance Q (or only query) is similar to pattern, but the class value is unknown. It is denoted $Q = (?, b_1, ..., b_n)$.

The intersection percentage between pattern P and query Q is calculated as:

$$IntersectionPercentage(P, Q) = 100 * \frac{|P \cap Q|}{|P|} .$$

The intersection percentage is 100% in the case when P became a subset of pattern Q', which has the head of the pattern P and the body of the query Q.

$$\left.\begin{array}{l} P = (c, a_1, ..., a_n) \\ Q = (?, b_1, ..., b_n) \end{array}\right\} \text{ if } P \subseteq Q' = (c, b_1, ..., b_n) \text{ then } IntersectionPercentage(P, Q) = 100\% .$$

The support of a pattern P in a data set $\mathbf{R} = \{R^i, i \in 1,...,r\}$ is the number of instances for which P became their subset.

$Supp(P,\mathbf{R}) = \text{number of } R^i : P \subseteq R^i,\ R^i \in \mathbf{R}, i \in 1,...,r$.

15.6.2 Training process

The training process in MPGN consists of:

- preprocessing phase;
- generalization phase;
- pruning phase.

✓ Preprocessing phase

MPGN deals with instances and patterns separately for each class. This allows MPGN algorithm to be realized for using on parallel computers.

The preprocessing phase is aimed to convert the learning set in a standard form for further steps. It consists of:

- discretization of numerical attributes [Mitov et al, 2009b];
- numbering the values of attributes.

After discretization and the juxtaposing positive integers to nominal values, the instances are converted to numerical vectors.

✓ Generalization phase

The process of generalization is a chain of creating the patterns of upper layer as intersection between patterns from lower layer until new patterns are generated. For each class, the process starts from the layer 1 that contains the instances of training set. Patterns, generated as intersections between instances of the training set are stored in layer 2. Layer N is formed by patterns generated as intersections between patterns of the layer N-1. This process continues until no intersections are possible.

During generalization, for every class a separate pyramidal network structure is built. The process of generalization creates "vertical" interconnections between patterns from neighborhood layers. These interconnections for every pattern are represented by a set of "predecessors" and a set of "successors".

The predecessors' set of a concrete pattern contains all patterns from lower layer, which were participated in the process of its generalization. This means that if different intersections generate one and the same pattern than all patterns, which participate in the intersection are united as predecessors of resulting pattern.

The predecessors sets for instances of layer one are empty.

The successors' set of a concrete pattern contains the patterns from upper layer, which are created from it.

The successors' sets of patterns on the top of the pyramid are empty. These patterns are called "vertexes" of the corresponded pyramids.

One pattern may be included in more than one pyramid, but the vertex pattern belongs only to one pyramid.

It is possible any pyramid to contain only one instance.

✓ Pruning phase

Pruning phase is the process of iterative analysis of vertex patterns of all pyramids from different classes and removing all contradictory vertex patterns. The pruning algorithm is displayed on diagram, showed in Figure 177.

As a result, some of the most general patterns are deleted, because the vertexes with the same bodies were available in other classes (and they are deleted). The primary pyramids are decomposed to several sub-pyramids with lower number of layers.

Finally, the received vertexes of the pyramids not contradict to vertexes of pyramids of other classes.

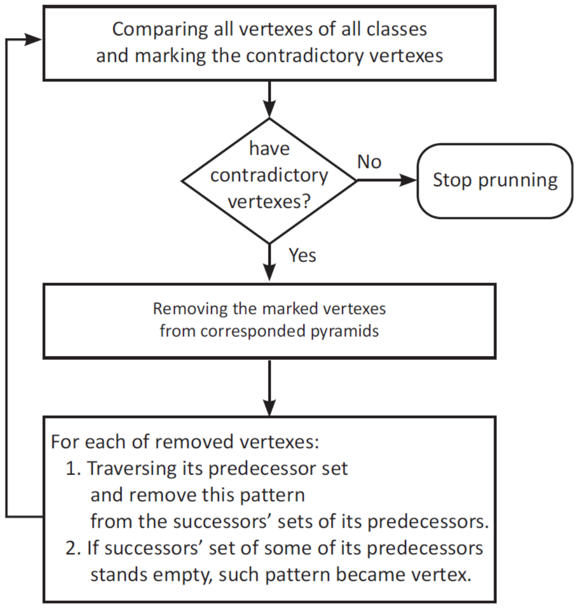


Figure 177. The MPGN pruning algorithm

15.6.3 Recognition process

The instance to be recognized is given by the values of its attributes $Q = (?, b_1, ..., b_n)$. Some of the values may be omitted. If some attributes are numerical, the values of these attributes are replaced with the number of corresponded discretized interval, where the value belongs. The categorical attributes also is recoded with the corresponded number values.

The recognition process consists of two main stages:

– creating recognition set for every class separately;

– analyzing resulting recognition sets from all classes and making decision which class to be given as answer.

✓ Creating recognition set for every class

In this stage each class is processed separately.

The goal is for each class to create the recognition set, which contains all patterns with maximal cardinality that have 100% intersection percentage with the query.

The process starts from the vertexes of all pyramids that belong to examined class. Using the predecessor sets of the patterns in the recognition set each pattern is replaced with the set of their predecessor that have 100% intersection percentage with the query, if such set is not empty. After lighting the recognition set keeping only patterns with maximal coverage, the process is iteratively repeated down to the layers until no new patterns became in the recognition set.

Figure 178 shows the block-scheme of this process.

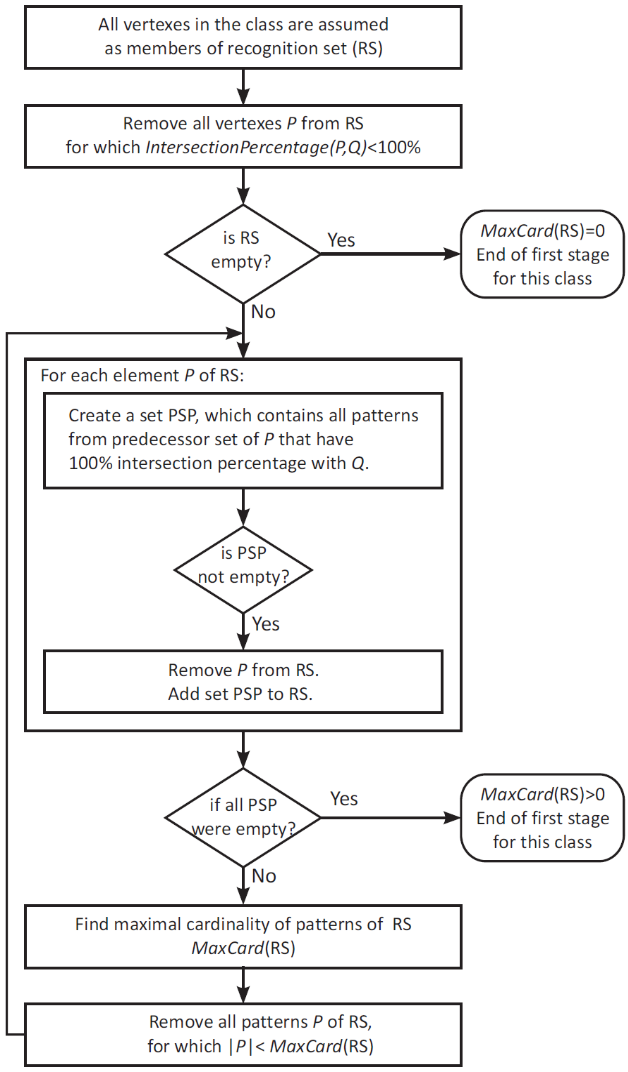


Figure 178. MPGN - Creating recognition set for one class

The process of creation of recognition set for each class also can be realized for using on parallel computers.

✓ Analyzing results and make final decision

The result of the first stage of the processing are the recognition sets for all classes $\mathrm{RS}c$, $c \in [1,\dots,M_c]$, which contain the patterns with maximal cardinality for this class $MaxCard(\mathrm{RS}c)$, $c \in [1,\dots,M_c]$ that have 100% intersection percentage with Q.

The goal is to find the class, which contains the patterns in its recognition set with highest cardinality.

For this purposes, first the maximum of all maximal cardinalities of the recognition sets of all classes is discovered.

$$MaxC = \max_{c \in [1,\dots,M_c]} MaxCard(\mathrm{RS}c)$$

The best case is when only one class has such maximal cardinality $MaxC$. Then, this class is given as answer.

The worst case is when $MaxC = 0$, which means that no recognition sets exists. In this case, if there is no instance, which body is equal to the query, the answer is that the request is not recognized and additional processing needs to be provided on the base of some other algorithms.

In the case when several classes have maximal cardinality $MaxC$ the number of instances with 100% intersection percentage with the query is find and the ratio between this number and all instances in the class is calculated. The class with maximal ratio is given as answer. The reason is that bigger ratio is received because the rule is more inherent to this class.

The full description of algorithm is shown in Figure 179.

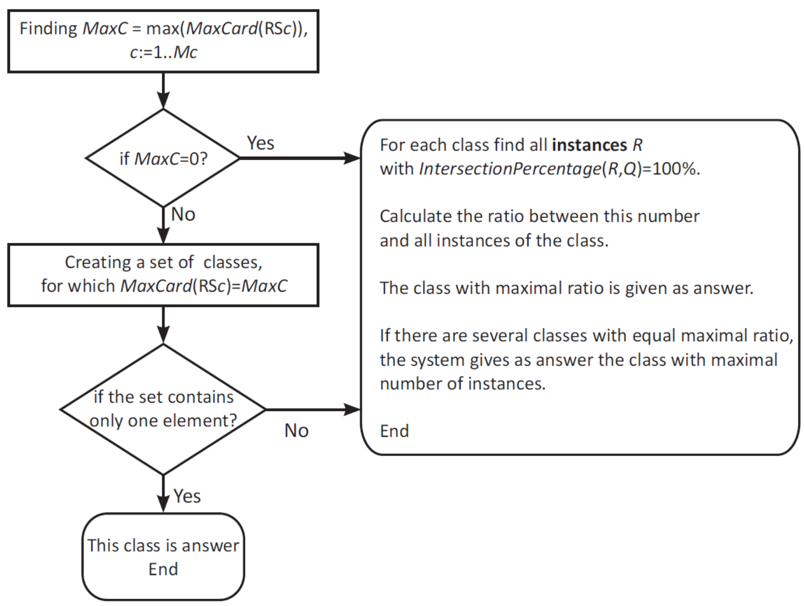


Figure 179. MPGN - Analyzing resulting recognition sets and making decision which class to be given as answer

This stage makes analysis of the results from previous step of all classes simultaneously. The parallelization of this process is not explicitly shown (it is not inherent to algorithm). The parallelization can be made in software realization during the process of comparing the results from classes between each other.

15.6.4 Experiments with Forestfires data set

Forest fires are major environmental issue, creating economical and ecological damage while endangering human lives. Fast detection is a key element for controlling such phenomenon. To achieve this, one alternative is to use automatic tools based on local sensors, such as provided by meteorological stations. In effect, meteorological conditions (e.g. temperature, wind) are known to influence forest fires and several fire indexes, such as the forest Fire Weather Index (FWI), use such data. In [Cortez and Morais, 2007] authors explore a data mining approach to predict the burned area of forest fires. Five different data mining techniques, e.g. Support Vector Machines (SVM) and Random Forests, and four distinct feature selection setups (using spatial, temporal, FWI components and weather attributes), were tested on real-world data collected from the northeast region of Portugal. The best configuration uses a SVM and four meteorological inputs (i.e. temperature, relative humidity, rain and wind) and it is capable of predicting the burned area of small fires, which are more frequent. Such knowledge is particularly useful for improving firefighting resource management (e.g. prioritizing targets for air tankers and ground crews).

The *Forestfires data set* is available in the UC Irvine Machine Learning Repository [UCI MLR, 2011]: http://archive.ics.uci.edu/ml/datasets/Forest+Fires. The data can be used to test regression, feature selection or outlier detection methods.

The goal of the task is to predict the burned area of forest fires, in the northeast region of Portugal, by using meteorological and other data.

The dataset contains 517 instances. There are used 12 real attributes without missing values. Class values are two types – "no" (48% of instances) or "fire" (52% of instances).

Here is given explanation of each attribute [Cortez and Morais, 2007]:

1. X: x-axis spatial coordinate within the Montesinho park map: 1 to 9.
2. Y: y-axis spatial co-ordinate within the Montesinho park map: 2 to 9.
3. Month: month of the year: 'jan' to 'dec'.
4. Day: day of the week: 'mon' to 'sun'.
5. FFMC: FFMC index from the FWI system: 18.7 to 96.20.
6. DMC: DMC index from the FWI system: 1.1 to 291.3.
7. DC: DC index from the FWI system: 7.9 to 860.6.
8. ISI: ISI index from the FWI system: 0.0 to 56.10.
9. Temp: temperature in Celsius degrees: 2.2 to 33.30.
10. RH: relative humidity in %: 15.0 to 100.
11. Wind: wind speed in km/h: 0.40 to 9.40.
12. Rain: outside rain in mm/m2 : 0.0 to 6.4.

For the experiments, we randomize Forestfires data set and split in ratio 3:1. Thus, the learning set contains 388 instances and examining set contains the rest 129 instances. We chose to use experiments with supplied examining set in order to ensure comparableness of the results between classifiers in different environments (PaGaNe and Weka).

✓ Results of training process of MPGN with Forestfires

The result from training process of MPGN for chosen learning data set is given in Table 11.

Table 11. Results from training process of MPGN for Forestfires data set

	Class "No"	Class "Fire"
Number of layers for class	5	5
Number of vertexes for class	1	1
Number of non-contradictory vertexes after pruning	8	515

Graphical representations of the pyramids for each class (before pruning) are given in Figure 180 (class="No") and Figure 181 (class="Fire"). It is clearly seen that one pyramid for each class is formed.

After pruning (the process of clearing the contradictory vertexes), the pyramid of class "No" is decomposed to 8 pyramids, and respectively the pyramid of class "Fire" is divided to 515 pyramids.

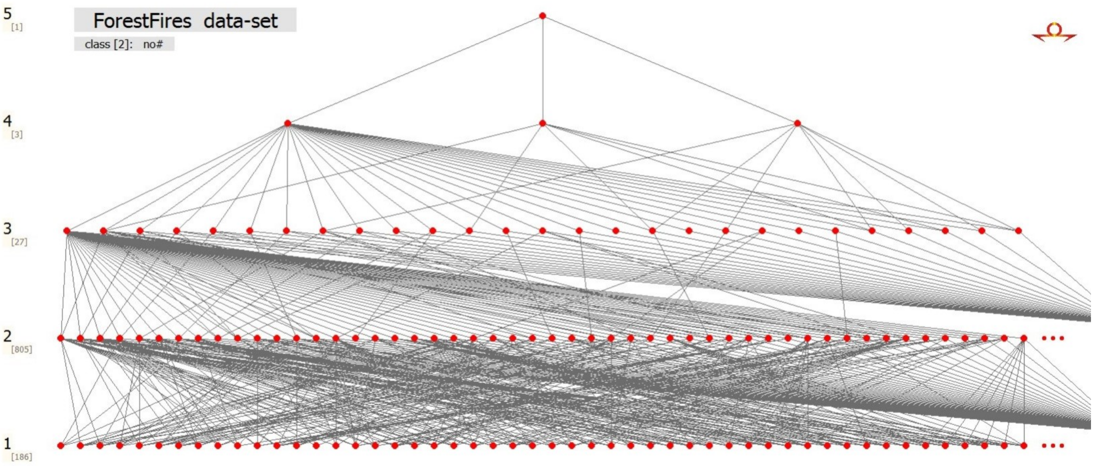


Figure 180. Pyramid for Forestfires data set, class="No"

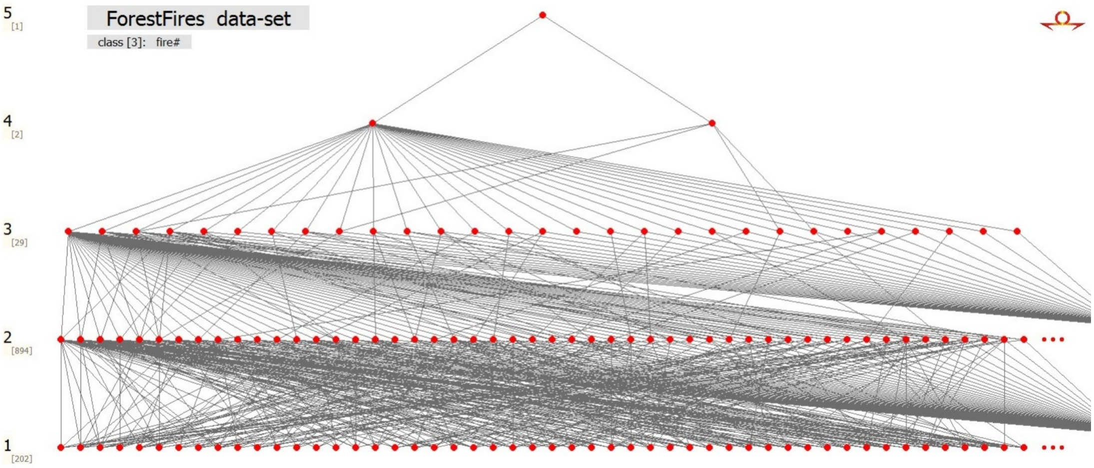


Figure 181. Pyramid for Forestfires data set, class="Fire"

✓ Results of recognition and comparison with other classifiers

We have made the comparison between classification accuracy of MPGN and other classifiers, realized in WEKA [Witten and Frank, 2005] for chosen data set.

For this comparison we have use best representatives from different groups of classifiers in WEKA (the classifiers that have showed best results within the group).

The chosen classifiers from groups, sorted in decreasing order of accuracy in the group, are as follows:

- Bayes – AODE (Aggregating One-Dependence Estimators), HNB (Hidden Naive Bayes), NaiveBayes;
- Functions – SMO (Sequential Minimal Optimization), SimpleLogistic, MultilayerPerceptron;
- Lazy – IB1, IBk, KStar;
- Meta – AdaBoostM1, LogitBoost, MultiScheme;
- Rules – JRip, ConjunctiveRule, OneR;
- Trees – ADTree, LADTree, J48.

The total number of the instances in the examining set was 129 instances.

In Table 12 the results of classification accuracy of MPGN and chosen representatives of WEKA classifiers for Forestfires data set are given.

Here is shown not only the global accuracy, received by classifiers, but also how they recognize corresponded classes, which is also important. For instance, ConjunctiveRule and MultiScheme have overall classification accuracy 52.71%, but this is as result that these classifiers give as answer only class="Yes". Other classifiers as classifiers in the Bayes group give not good accuracy (they are at the end of the ranking), but recognize relatively uniformly the instances of two classes.

Table 12. Results of recognition for Forestfires data set with MPGN and WEKA classifiers

Classifiers classifier model (supplied test set)	***Correctly classified instances***	**Accuracy (%)**	**Correctly classified "class=No" (%)**	**Correctly classified "class=Fire" (%)**
MPGN				
MPGN	*74*	57.36	57.38	57.35
Bayes				
AODE	*66*	51.16	44.26	57.35
HNB	*66*	51.16	52.46	50.00
NaiveBayes	*60*	46.51	44.26	48.53
Functions				
SMO	*71*	55.04	49.18	60.29
SimpleLogistic	*70*	54.26	65.57	44.12
MultilayerPerceptron	*67*	51.94	19.67	80.88
Lazy				
IB1	*69*	53.49	67.21	41.18
IBk	*68*	52.71	62.30	44.12
KStar	*66*	51.16	50.82	51.47
Meta				
AdaBoostM1	*72*	55.81	14.75	92.65
LogitBoost	*70*	54.26	65.57	44.12
MultiScheme	*68*	52.71	0.00	100.00
Rules				
JRip	*77*	59.69	26.23	89.71
ConjunctiveRule	*68*	52.71	0.00	100.00
OneR	*65*	50.38	47.54	52.94
Trees				
ADTree	*83*	64.34	36.07	89.71
LADTree	*78*	60.47	24.59	92.65
J48	*67*	51.94	19.67	80.88

Figure 182 shows overall classification accuracy of Forestfires data set for examined classifiers. Different colors of bars correspond to the groups, in which classifiers belongs. As we can see best results are given from ADTree and LADTree. In contrary, classifiers from Bayes group have not so good overall accuracy. MPGN showed relatively good results and it is on the fourth position.

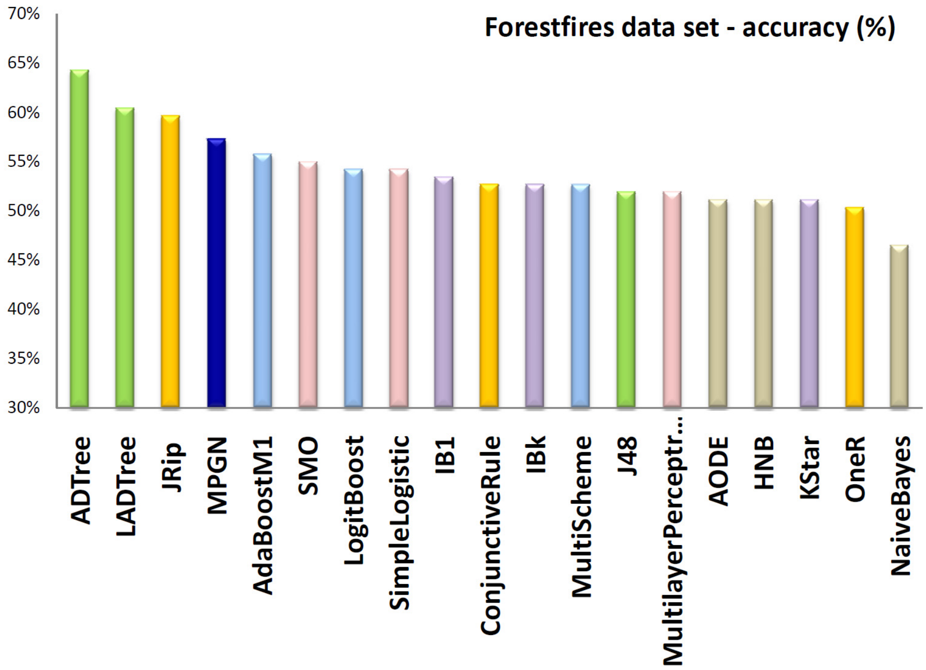


Figure 182. Classification accuracy of Forestfires data set given by MPGN and selected Weka classifiers

As we can see on Figure 183 it is not enough one classifier to have good overall accuracy. Keeping awareness commands in permanent readiness is also not so good situation. From first five classifiers only MPGN gives relatively good ratio on recognition two classes. In the other four classifiers prevailing answers are class="Fire".

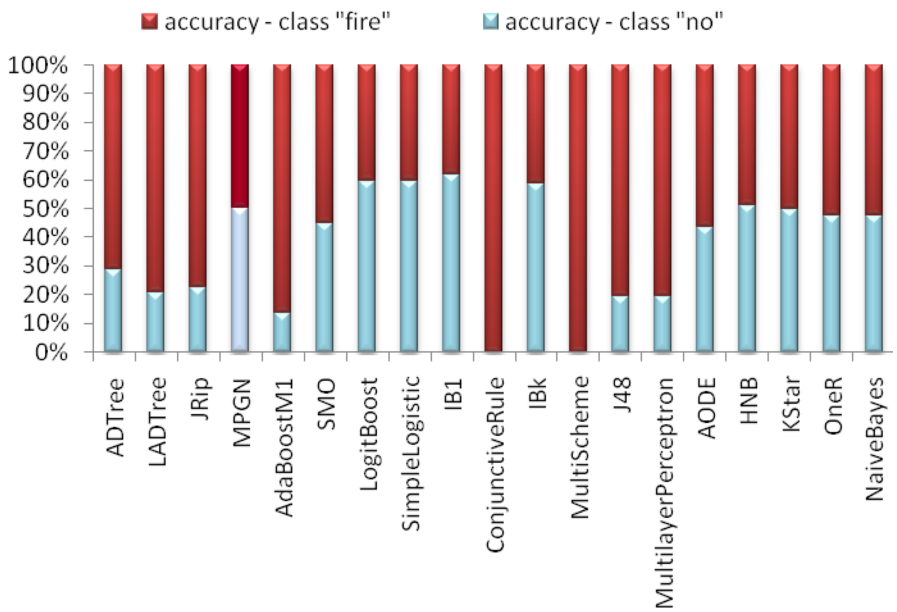


Figure 183. The ratio between recognized instances from class "No" versus class "Fire"

15.7 Discussion

The Memory Data Structures (MDS) and corresponded Access Methods (AM) had been available from the beginning of the developing the computer devices. As many devices there exists so many possibilities for developing different MDS and AM we have.

The main goal of creation of data mining environment PaGaNe was to show the advantages of using the multi-dimensional numbered information spaces as a model for memory structuring in the intelligent systems.

The main idea consists in replacing the (symbol or real; point or interval) values of the objects' attributes with integer numbers of the elements of corresponding ordered sets. This way each object is described by a vector of integer values, which is used as co-ordinate address in the multi-dimensional information space.

In other words, replacing the names by numbers permits the using of mathematical functions and address vectors for accessing the information.

ArM32 engine supports multithreaded parallel access to the information base in real time. Very important characteristic of ArM32 is possibility not to occupy space for empty structures (elements or spaces).

The Data Mining Environment PaGaNe is the first realization based on the Multi-dimensional numbered information spaces. The good results received by its systems show a new way for building intelligent systems.

The "class association rules" (CAR) algorithms have their important place in the family of classification algorithms. The advantages of associative classifiers can be highlighted in several very important directions, such as: very efficient training; possibility to deal with high dimensionality; no assumptions for the independence of attributes; very fast classification and the result is easily understandable by humans. The latter two advantages make CAR algorithms an irreplaceable assistant in the processes of disaster risk management, where fast reaction and reliability of the systems are crucial. The realization of discussed MPGN classifier, based on such structures, and conducted experiments with it show promising results in the area of using such paradigm in the field of disaster risk management.

Conclusion

In this collective scientific monograph, we have presented several important aspects of Intelligent Data Processing in Global Monitoring for Environment and Security. The book was structured in fifteen chapters collected in three parts, regarding practical, theoretical, and technological aspects.

The Part I, "Practical aspects", was devoted to describe main features of the disasters and corresponded risk management as well as several approaches for possible solutions. In particular, we have discussed:

- examples of different disasters and approaches to manage corresponded risk (Chapter 1. Managing Risk and Safety);
- different approaches to multi-source data integration for the solution of complex applied problems using Grid computing technologies for, in particular flood mapping, and vegetation state estimation using satellite, modeling and in-situ data. (Chapter 2. High-performance Intelligent Computations for Environmental and Disaster Monitoring);
- results from the investigations accomplished by the CERGOP-2 project, based on GNSS (Global Navigation Satellite Systems) campaigns. (Chapter 3. Investigation of Geodynamics of Central and Eastern Europe, Balkan Peninsula and Bulgaria);
- Bulgarian and Ukrainian experience in development of measurement devices and an organization of express and continuous field electronic measurements for evaluation of main environment parameters. (Chapter 4. Intelligent Tools for Environment Monitoring: Features and Applications);
- machine learning techniques for intelligent data processing in environmental and security-related monitoring employing gamma spectroscopy. (Chapter 5. Intelligent Gamma-Ray Data Processing for Environmental Monitoring);
- automation in the field of the risk management, based on data processing of remote space monitoring of the spatially-distributed natural and technogenic objects for timely detection maintenance, diagnostics and the development predicate of the dangerous phenomena and emergencies. (Chapter 6. Acquisition, Processing and Analysis of Space Images at Risks Management of Natural and Technogenic Emergencies).

The Part II, "Theoretical aspects" has presented several important theoretical results, which are received by the authors:

- using of effective methods and algorithms of estimation, comparison, generalization and decision making under uncertainty; using of contemporary methods of basic earth and water parameters measuring, specifically aerospace microwave radiometry; combining these effective means in the big Regional Monitoring System. (Chapter 7. Elaboration of Geoinformation Regional Monitoring Environmental System ("GERMES-I") Enriched by Artificial Intelligence Instruments);
- situational modeling for decision making in emergency microsituation as a base part of risks management. (Chapter 8. Microsituation Concept in GMES Decision Support Systems);

- analysis of different file objects, information attacks, and protection methods most frequently used in Software Critical Infrastructures is made on the base of available information of National Laboratory of Computer Virology at the Bulgarian Academy of Sciences for accomplished attacks in Bulgaria, Balkan Peninsula, and southeast Europe. (Chapter 9. Methods and Means for Protection of Software Critical Infrastructures);
- application of the heterogeneous variables system prediction method to solving the time series analysis problem with respect to the sample size. (Chapter 10. The MLRP-method for Analysis of Some Problems in Climate and Seismology);
- results of the PCRM theory, consider applications to decision making support in conditions of risk, and develop numerical methods for searching optimal decisions. (Chapter 11. Polyhedral Coherent Risk Measures and their Application to Investment Decisions Support under Catastrophic Flood Risks);
- models for searching the upper and lower bounds for Bayesian estimates which can be derived for any prior distribution satisfying the given partial prior information. The chapter considers. (Chapter 12. Techniques for Robust Bayesian Estimation);
- quantitative measures of information content and value; completeness, accuracy, and clarity as attributes of information acquired by the receiver; suggestions on how to use interpretations and mathematical tools developed within the information theories to maintain and improve safety of nuclear power plants. (Chapter 13. Application of Information Theories to Safety of Nuclear Power Plants).

The Part III, "Technological aspects" was aimed to pay attention to technological approaches, which have shown very good results in the practice:

- an universal approach to analysis of attributive risk management and disaster emergencies: growing hieratical network structures for memory organization called Growing Pyramidal Neural Networks (GPN) and Program complex CONFOR (abbreviation of CONcept FORmation) to be used in problems solving in GMES. (Chapter 14. Growing Pyramidal Networks);
- intelligent support of decision making in GMES based on “Multi-dimensional Numbered Information Spaces” and "class association rules" (CAR) algorithms implemented in the Data mining analysis environment "PaGaNe" and especially in MPGN classifier aimed to be used in the field of disaster prediction. (Chapter 15. Multi-dimensional Information Spaces as Memory Structures for Intelligent Data Processing in GMES).

Now, it is clear, the Intelligent Data Processing in the Global Monitoring for Environment and Security is very important for humanity to survive in the variety of disasters and technogenic collapses. All aspects (practical, theoretical, and technological) need to be intensively investigated. Let remember the example from introduction about "the local" and "the total" time. The Global Monitoring means "the total" monitoring of the world around us. It is impossible without intelligent systems.

We hope the presented ideas will cause interest and further collaboration in the area of intelligent data processing in GMES. For us, this monograph was nice possibility to integrate our potential, independently of real or intellectual boundaries. Finishing this book we do not finish our collaboration – several international projects have started, other are at the preparing stage. In addition, uniting a team of sixty authors, working two years together, and using modern electronic communication systems, gave us new skills and knowledge to manage natural and artificial risk, to work for security of our world.

Bibliography

[21st Century, 1987] Infrastructure for the 21st Century, Washington, D.C., National Academy Press, 1987

[Acerbi, 2002] Acerbi C. Spectral Measures of Risk: a Coherent Representation of Subjective Risk Aversion. Journal of Banking and Finance, 26(7), 2002, pp.1505–1518.

[Achlioptas, 2003] Achlioptas D. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. J. Comput. System Sci., 66(4), 2003, pp.671-687.

[Agarwal et al, 2001] Agarwal R., Aggarwal C., Prasad V. A tree projection algorithm for generation of frequent item-sets. In Journal of Parallel and Distributed Computing, Volume 61, Issue 3, 2001, pp.350-371.

[Agrawal and Srikant, 1994] Agrawal R., Srikant R. Fast Algorithms for Mining Association Rules, In Proc. of the 20th Int. Conference on Very Large Databases, Santiago, Chile, 1994, pp.487-499.

[Agrawal et al., 1993] Agrawal R., Imieliński T., Swami A. Mining Association Rules Between Sets of Items in Large Databases. In Proc. of the ACM SIGMOD Int. Conf. on Management of Data, Washington, DC, 1993, pp.207-216.

[Aichhorn et al, 2008] Aichhorn C., Stangl G., Krauss, S. A velocity Field for Romania and Bulgaria. In: Reports on Geodesy No.1 (84), Eds. Sledzinski, J., Warsaw University of Technology, Warsaw, 2008, pp.17-22.

[Akaike, 1974] Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 1974, pp.716-723.

[Alekseev, 2010] Alekseev D. About earthquake on Haiti warned two years ago. In: www.mk.ru. 2010.

[Amari and Cichocki, 1998] Amari S., Cichocki A. Adaptive blind signal processing – neural network approaches. Proceedings IEEE,. 86, 1998, pp.1186-1187.

[Amari et al, 2002] Amari S., Hyvarinen A., Lee S.Y., Lee T.W., Sanchez V.D. Blind signal separation and independent component analysis. Neurocomputing, 49(12), 2002, pp.1-5.

[Amendola et al, 2000] Amendola A., Ermoliev Y., Ermolieva T., Gitits V., Koff G., Linnerooth-Bayer J. A Systems Approach to Modeling Catastrophic Risk and Insurability, Natural Hazards Journal, 21, Issue 2/3, 2000, pp.381-393.

[Ao and Gelman, 2009] Ao S., Gelman L. Advances in Electrical Engineering and Computational Science, 2009, pp. 277-288

[Arge, 2002] Arge L. External Memory Data Structures. Part 4, chapter 9 in Handbook of Massive Datasets. J. Abello, P.M. Pardalos, M.G.C. Resende (eds), Kluwer Academic Publishers, 2002, pp. 313-357.

[Artzner et al, 1999] Artzner P., Delbaen F., Eber J.-M., Heath D. Coherent Measures of Risks. Mathematical Finance, 1999, 9, pp.203–227.

[Ashby, 1952] Ashby R. Design for a brain. Chapman and Hall, London, 1952.

[Ashby, 1956] Ashby R. An introduction to cybernetics. J Wiley, New York, 1956.

[Aslanyan and Castellanos, 2007] Aslanyan L., Castellanos J. Logic based Pattern Recognition – Ontology content (1), Information Theories and Applications, ISSN 1310-0513, Sofia, Vol. 14, N. 3, 2007, pp.206-210.

[Aslanyan and Ryazanov, 2008] Aslanyan L., Ryazanov V. Logic Based Pattern Recognition – Ontology Content (2), Information Theories and Applications, ISSN 1310-0513, Sofia, Vol. 15, N. 4, 2008, pp.314-318.

[Aslanyan and Sahakyan, 2009] Aslanyan L., Sahakyan H. Chain split and computation in practical rule mining, Information Science and Computing, International book series no. 8., Classification, forecasting, data mining, 2009, pp.132-135.

[Aslanyan et al, 2002] Aslanyan L., Margaryan K., Sahakyan H. Data analysis algorithms in network protection systems, Proceedings of the 3rd Int. Conf. Digital information processing and control in extreme situations, 28-30 May, Minsk, 2002, pp.221-225.

[Aslanyan et al, 2003] Aslanyan L., Castellanos J., Mingo F., Sahakyan H., Ryazanov V. Algorithms for Data Flows, X-th International Conference Knowledge-Dialogue-Solution, 16-26.06.2003, Varna.

[Aslanyan et al, 2007] Aslanyan L., Arsenyan I., Karakhanyan V. Learning Schemes and Pattern Recognition, Computer Science and Information Technologies Conference, Yerevan, September, 2007.

[Babenko, 1984] Babenko V. A modeling of visual situation is in the trainers of transport vehicles, Kiev, 1984, 16p.

[Bach and Jordan, 2002] Bach F.R., Jordan M.I. Kernel independent component analysis. Journal of Machine Learning Research, 3, 2002, pp.1-48.

[Bacour et al, 2002] Bacour C., Jacquemoud S., Tourbier Y. et al. Design and analysis of numerical experiments to compare four canopy reflectance models. Ibid, V. 79, 2002, pp.72-83.

[Bacour et al, 2006] Bacour C., Baret F., Béal D., Weiss M., Pavageau K. Neural network estimation of LAI, fAPAR, fCover and LAI×Cab, from top of canopy MERIS reflectance data: principles and validation. Ibid, 105, 2006, pp.313–325.

[Banerji, 1978] Banerji R.B. A Language for Pattern Recognition. Pattern Recognition, No 1, 1978.

[Baranov, 1999] Baranov S. Generator of Earthquake Scenarios and Estimation of Damage for a Seismic Region (Toscana, Italy), IIASA interim report IR-99-047, Laxenburg, Austria, 1999.

[Bar-Hillel and Carnap, 1953] Bar-Hillel Y., Carnap R. Semantic information. British Journal of science, N 4, 1953, pp.147-157.

[Bayardo, 1997] Bayardo R. Brute-force mining of high-confidence classification rules. In 3rd Int. Conf. on Knowledge Discovery and Data Mining (KDD’97), 1997, pp.123-126.

[BBC, 2010] Icelandic volcanic ash alert grounds UK flights. In: news.bbc.co.uk. 2010.

[Belgium, 2010a] In the Belgium trains 18 persons were killed. In: www.meadeast.ru. 2010.

[Belgium, 2010b] Sad result of the train crash in Belgium. In: www.baltinfo.ru. 2010.

[Bennet and Martz, 1972] Bennet G.K., Martz H.F. A Continuous Empirical Bayes Smoothing Technique. Biometrika, 59, 1972, pp.361-368.

[Berger, 1984] Berger J.O. The Robust Bayesian Viewpoint (with discussion). In Robustness in Bayesian Z. Statistics, (J.Kadane, ed.). Amsterdam: North Holl, 1984.

[Berger, 1994] Berger J.O. An Overview of Robust Bayesian Analysis (with comments). Test, 3, № 1, 1994, pp.5-124.

[Berikov, 2002] Berikov V.B. An approach to the evaluation of the performance of a discrete classifier. Pattern Recognition Letters. Vol. 23 (1-3), 2002, pp.227-233.

[Bezruchko and Smirnov, 2003] Bezruchko B.P., Smirnov D.A. The modern modeling by time series. http://www.nonlinmod.sgu.ru/doc/review.pdf, 2003.

[Billings and Hovgaard, 1999] S. Billings, J. Hovgaard. Modeling detector response in airborne gamma-ray spectrometry. Geophysics, 64(5), 1999, pp.1378-1392.

[Billings et al, 2003] Billings S.D., Minty B.R., Newsam G.N. Deconvolution and spatial resolution of airborne gamma-ray surveys. Geophysics, 68(4), 2003, pp.1257–1266.

[Bilous and Kobzar, 2008] Bilous N., Kobzar G. Automatic ECG analysis for Preliminary and Detailed diagnostics based on scale-space representation. International Journal Information Technologies and Knowledge Vol.2, 2008, pp.53-60.

[Bilous and Kobzar, 2009] Bilous N., Kobzar G. Fast curvature scale space calculation technique for real-time or time-limited shape recognition. Scientific and Technical Journal Artificial Intelligence №3 (09), Donetsk: Science and Education, 2009.

[Bilous et al, 2008] Bilous N., Bondarenko M., Kobzar G., Krasov A., Rogozyanov. Normal ECG recognition for express-diagnostics based on scale-space representation and dynamic matching. International Book Series Information Science and Computing №5. Intelligent Technologies and Applications, 2008, pp.47-53.

[Bishop, 1996] Bishop C. Neural Networks for Pattern Recognition. Oxford University Press, 1996.

[Blokhinov, 1974] Blokhinov E.G. Probability Distribution of River Runoff Discharges. Moscow, Nauka, 1974.

[Blyth, 2006] Blyth A., Ec2nd 2005: Proceedings of the First European Conference on Computer Network Defence, Springer, 2006, 102 p.

[Bodon and Ronyai, 2003] Bodon F., Ronyai L. Trie: an alternative data structure for data mining algorithms. In Mathematical and Computer Modelling, Volume 38, Number 7, 2003, pp.739-751.

[Bodon, 2003] Bodon F. A Fast APRIORI Implementation. IEEE ICDM Workshop on Frequent Itemset Mining Implementations (FIMI'03), Melbourne, Florida, USA, 2003.

[Bongard, 1967] Bongard M. The problems of knowledge formation. Nauka-Moscow, 1967, 320 pp. (in Russian)

[Borisenko et al, 2004] Borisenko V.P., Kolodyajny V.V., Ponomarev Y.V.. Development methodology and implementation experience of Ukrainian distributed gas-transport system. In the book: Pipline Energy Systems. Development and Functionin Control. Novosibirsk: Science, 2004.

[Borisenko et al, 2008a] Borisenko V., Kluk B., Ponomarev Y., Starovoytov A. The Complex Unified Evolutionary Approach to the Multilevel Distributed Control System of Gas Transport Company.In: International Book series Information Science and Computing.- Number 4.- Suppl. to the Int. J. Information Technologies and Knowledge V.2, Sofia, Bulgaria, 2008.

[Borisenko et al, 2008b] Borisenko V.P., Medvedeva L.M., Borisenko T.I. Basic aspects of automatic control system development for business-processes of oil and gas enterprises. In: ACS and automatic devices. № 144, 2008.

[Borodin et al, 1996] Borodin L., Krapivin V., Long B. Application of the GIMS technology to the Aral-Caspian aquageosystem monitoring. Problems of Environment and Natural Resources, 10, 1996, pp 46-61, (in Russian).

[Botev et al, 2006] Botev E., Slejko D., Bressan G., Bragato P., Glavcheva R. On the Geodynamics of Bulgarian Lands through Seismological Data. Geodynamics of the Balkan Peninsula. Monograph. Report on Geodesy, Warsaw University of Technology, No. 5 (80), 2006, pp.149-168.

[Box and Dgenkins, 1974] Box Dj., Dgenkins G. Analysis time series. Prediction and direction. Publ. Moskow, Wold, 1974, 242 p.

[Breismeister, 2000] Breismeister J.F. MCNP (4C) A General Monte Carlo N-Particle Transport Code, report: LA-13709-M, Los Alamos National Laboratory, 2000.

[Brent and Luk, 1985] Brent R.P., Luk F.T.. The solution of singular-value and symmetric eigenvalue problems on multiprocessor arrays/ SIAM J. Sci. and Stat. Comput., 6, 1985, pp.69-84.

[Brent, 1988] Brent R.P. Parallel algorithms for digital signal processing. In: Proc. of NATO Advance Study Institute on Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms, Leuven, Belgium, 1988, pp.93-110.

[Bugriy, 2004] Bugriy A.N., Gusyatin V.M., Ostroushko A.P. Determination of pixel colour parameters subject to physical characteristics of light sources// Herald of National technical university KHPI. Collection of scientific papers. Subject collection: Automobile and tractor industry, №24, Kharkov: NTU KHPI, 2004, pp.13-18.

[Burchfiel et al, 2006] Burchfiel B., King R., Todosov A., Kotzev V., Dumurdzanov N., Serafimofski T., Nurce B. GPS results for Macedonia and its importance for the tectonics of southern Balkan extensional system, Tectonophys. 413, (3-4), 2006, pp.239-248.

[Buser et al, 1987] Buser O., Butler M., Good W. Avalanche forecast by the nearest neighbors method. In: IAHS, Publ. № 162, 1987, pp557–569.

[Camenisch, 2009] Camenisch J. Inetsec 2009 – Open Research Problems in Network Security: Ifip Wg 11.4 International Workshop, Zurich, Switzerland, April 23-24, 2009, Revised Selected Papers, Springer, 2009, p.54.

[Cameron et al, 1999] Cameron D.S., Beven K.J., Tawn J., Blazkova S., Naden P. Flood Frequency Estimation by Continuous Simulation for a Gauged Upland Catchment (with uncertainty), J. Hydrol., 219, 1999, pp.169-187.

[Canberra] In situ gamma spectroscopy systems for soil and surface activity measurements http://www.canberra.com/literature/973.asp

[Caporali et al, 2007] Caporali A., Becker M., Fejes I., Gerhatova L., Ghitau D., Grenerczy G., Hefty J., Medak D., Milev G., Mojzes M., Mulic M., Nardo A., Pesec P., Rus T., Simek J., Sledzinski J., Solaric M., Stangl G., Vespe F., Virag G., Vodopivec F., Zablotskyi F. Geokinematics of Central Europe: new insights from the CERGOP-2/Environment Project. Reports on Geodesy No.2 (83), Eds. Sledzinski, J., Warsaw University of Technology, Warsaw, 2007, pp.7-46.

[Caporali et al, 2008] Caporali A., Aichhorn C., Becker M., Fejes I., Gerhatova L., Ghitau D., Grenerczy G., Hefty J., Krauss S., Medak D., Milev G., Mojzes M., Mulic M., Nardo A., Pesec P., Rus T., Simek J., Sledzinski J., Solaric M., Stangl G., Vespe F., Virag G., Vodopivec F., Zablotskyi F. Geokinematics of Central Europe: New insights from the CERGOP-2/Environment project, J. Geodyn., 45, 2008, pp.246-256.

[Caporali et al, 2009] Caporali A., Aichhorn C., Barlik M., Becker M., Fejes I., Gerhatova L., Ghitau D., Grenerczy G., Hefty J., Krauss S., Medak D., Milev G., Mojzes M., Mulic M., Nardo A., Pesec P., Rus T., Simek J., Sledzinski J., Solaric M., Stangl G., Stopar B., Vespe F., Virag G. Surface kinematics in the Alpine-Carpathian-Dinaric and Balkan region inferred from a new multi-network GPS combination solution, Tectonophysics 474, 2009, pp.295-321.

[Carlson and Ripley, 1998] Carlson T., Ripley D. On the relation between NDVI, fractional vegetation cover, and leaf area index. Remote Sensing of Environment 62, 1998, pp.241-252.

[Carminati et al, 2004] Carminati E., Doglioni C., Carrara G., Dabovski C., Dumurdjanov N., Gaetani M., Georgiev G., Mauffret A., Sartori R., Seranne M., Scrocca D., Scionti V., Torelli L., Zagorchev I., Argnani A. A lithospheric cross section through the central and eastern Mediterranean region. Transmed: 32th International Geological Congress, Florence, Italy, 2004.

[Carota and Ruggeri, 1994] Carota C., Ruggeri F. Robust Bayesian analysis given priors on partition sets. Test, 3, No.2, 1994, pp.73-86.

[Ceccato et al, 2002] Ceccato P., Gobron N., Flasse S., et al. Designing a spectral index to estimate vegetation water content from remote sensing data: part 1 Theoretical approach. Remote Sensing of Environment, 82, 2002, pp.188-197.

[Cendrowska, 1987] Cendrowska, J. "PRISM: An algorithm for inducing modular rules", International Journal of Man-Machine Studies, 1987, 27, 349-370

[CEOSDMSG, 2001] Committee on Earth Observation Satellites Disaster Management Support Group: The Use of Earth Observing Satellites for Hazard Support: Assessments and Scenarios, Final Report; National Oceanic and Atmospheric Administration, Department of Commerce, USA, 2001.

[Chakrabarti, 2001] Chakrabarti K. Managing Large Multidimensional Datasets Inside a Database System. Phd Thesis, University of Illinois at Urbana-Champaign. Urbana, Illinois, 2001.

[Chandra et al, 2009] Chandra, P., Bensky, A., Hurley, C., Rackley, S., Wireless security, Newnes, 2009, p.313.

[Chavez et al, 2001] Chavez E., Navarro G., Baeza-Yates R., Marroquin J. Searching in Metric Spaces. ACM Computing Surveys, 33(3), 2001, pp.273–321.

[Chernavsky, 2004] Chernavsky D. Synergetic and information. URSS, Moscow, 2004.

[Chichocki and Amari, 2002] Chichocki A., Amari S. Adaptive Blind Signal and Image Processing, Wiley, 2002, 586 pp.

[Christoskov et al, 2006] Christoskov L., Solakov D., Simeonova S. Seismicity of Bulgaria. Geodynamics of the Balkan Peninsula. Monograph. Report on Geodesy, Warsaw University of Technology, No. 5 (80), 2006, pp.127-136.

[Churumova, 2009] Churumova D. FBI suspects Russians of the theft millions dollars from Citibank. In: www.rb.ru. 2009.

[Ciampa, 2008] Ciampa M. Security+ Guide to Network Security Fundamentals, Cengage Learning, 2008, p.17.

[Ciampa, 2009] Ciampa M., Security Awareness: Applying Pracitical Security in Your World, Cengage Learning, 2009, p.25.

[Ciarapica and Todini, 2002] Ciarapica L., Todini E. TOPKAPI: a model for the representation of the rainfall-runoff process at different scales. Hydro. Proc., 16 (2), 2002, pp.207-229.

[Clarke, 2009] Clarke, J., SQL injection attacks and defense, Syngress, 2009, pp.170-171.

[Clemmer and Krutchkoff, 1968] Clemmer B.A., Krutchkoff R.G. The Use of Empirical Bayes Estimation in a Linear Regression Model. Biometrika, 55, 1968, pp.525-534.

[CODASYL, 1971] Codasyl Systems Committee. Feature Analysis of Generalized Data Base Management Systems. Technical Report, May, 1971.

[Codd, 1970] Codd E. A Relational Model of Data for Large Shared Data Banks. Communications of the ACM, 13 (6), 1970, pp.377–387.

[Coenen and Leng, 2004] Coenen F., Leng P. An evaluation of approaches to classification rule selection. In IEEE Int. Conf. on Data Mining (ICDM'04), 2004, pp.359-362.

[Cohen and Kalbaugh, 2008] Cohen A., Kalbaugh G. E. ESI Handbook: Sources, Technology and Process, Aspen Publishers Online, 2008, p. 5-8.

[Cohen, 1995] Cohen W. Fast Effective Rule Induction. In Proceedings of the Twelfth International Conference on Machine Learning, Lake Taho, California, Morgan Kauffman, 1995.

[Combal et al, 2002] Combal B., Baret F., Weiss M., et al. Retrieval of canopy biophysical variables from bidirectional reflectance using prior information to solve the ill-posed inverse problem. Ibid, 84, 2002, pp.1-15.

[Comon, 1994] Comon P. Independent component analysis: a new concept. Signal Processing, 36, 1994, pp.287-314.

[Connolly, Begg, 2002] Connolly T.M., Begg C.E. Database Systems. A Practical Approach to Design, Implementation, and Management. Third Edition. Addison-Wesley Longman, Inc. – Pearson Education Ltd., 2002.

[Copas, 1972] Copas J.B. Empirical Bayes Methods and the Repeated Use of a Standard. Biometrika, 59, 1972, pp.349-360.

[Corbley, 1999] Corbley K. "Red River Growers Turn to Satellite to Manage Nitrogen". Modern Agriculture: The Journal for Site-Specific Crop Management, 1999. http://www.eomonline.com/modernagsite/archives/Corbley.html

[Cortez and Morais, 2007] P. Cortez and A. Morais. A Data Mining Approach to Predict Forest Fires using Meteorological Data. In J. Neves, M. F. Santos and J. Machado Eds., New Trends in Artificial Intelligence, Proceedings of the 13th EPIA 2007 – Portuguese Conference on Artificial Intelligence, December, Guimaraes, Portugal, pp. 512-523, 2007. APPIA, ISBN-13 978-989-95618-0-9.

[Craig et al, 1999] Craig M., Dickson B., Rodrigues S. Correcting aerial gamma-ray survey data for aircraft altitude. Exploration Geophysics, 30, 1999, pp.161-166.

[Craig, 1993] Craig M. The point spread function for airborne radiometry. Math. Geol., 25, 1993, pp.1003-1013.

[Cunjian et al, 2001] Cunjian Y., Yiming W., Siyuan W., Zengxiang Z., Shifeng H. Extracting the flood extent from satellite SAR image with thesupport of topographic data. Proc. of International Conference on Information Technology and Information Networks (ICII 2001). Beijing, China. Volume 1, 2001, pp.87-92.

[Dale and Lewis, 2009] Dale N., Lewis J. Computer Science Illuminated, Jones & Bartlett Learning, 2009, p.367.

[Date, 1975] Date C.J. An Introduction to Database Systems. Addison-Wesley Inc. 1975.

[Davenport and Harris, 2007] Davenport Th. H., J. G. Harris. Competing on Analytics: The New Science of Winning. Harvard Business School Press, Boston, Massachusetts. ISBN-13:978-1-4221-0332-6, 2007. See also online by the same authors: An Introduction to Data Mining: Discovering Hidden Value in your Data Warehouse. http://www.thearling.com/text/dmwhite/dmwhite.htm

[Davidov, 2007] A. Davidov, Digital sygnal processing, 2007.

[De Chiara et al, 2006] De Chiara G., Bovolin V., Migliaccio M. Remote sensing technique to estimate the water surface of artificial reservoirs Villani – Problems and potential solutionsp IEEE GOLD Remote Sensing Conference, 2006.

[De Meijer, 2007] de Meijer R.J. Method and system for detecting a property of a pavement by measuring gamma-radiation. US Patent 7235780, http://www.freepatentsonline.com/7235780.html, 2007.

[Deeley et al, 1970] Deeley J.J., Tierney M.S., Zimmer W.J. On the Usefulnes of the Maximum Entropy Principle in the Bayesian Estimation of Reliability. IEEE Transactions on Reliability, R-19, 1970, pp.110-115.

[DEGREE, 2008] Project: Dissemination and Exploitation of Grids in Earth Science, https://www.eu-degree.eu/

[Denning, 1997] Denning D.E. An Intrusion Detection Model. In: IEEE Trans. Software Eng., vol. 13, no. 2. 1997.

[Depaire et al, 2008] Depaire B., Vanhoof K., Wets G. ARUBAS: An Association Rule Based Similarity Framework for Associative Classifiers. IEEE International Conference on Data Mining Workshops, 2008, pp.692-699.

[Desbarats and Killeen, 1990] Desbarats A.J., Killeen P.G. A least-squares inversion approach to stripping in gamma-ray spectral logging. Nuclear Geophysics, 4(3), 1990, pp.343-352.

[DeVore, 1998] DeVore R.A. Nonlinear approximation. Acta Numerica, 7, 1998, pp.51–150.

[Dickson and Taylor, 1998] Dickson B., Taylor G. Noise reduction of aerial gamma-ray surveys. Exploration Geophysics 29, 1998, pp.324-329.

[Dickson and Taylor, 2000] Dickson B., Taylor G. Maximum noise fraction (MNF) method reveals detail in aerial gamma-ray surveys. Exploration Geophysics, 31, 2000, pp.73-77.

[Dickson, 2004] Dickson B.L. Recent advances in aerial gamma-ray surveying. Journal of Environmental Radioactivity 76 (1-2), 2004, pp.225–236.

[Dictionary, 2010] Online Etymology Dictionary, Douglas Harper, Historian, http://dictionary.reference.com/browse/infrastructure

[Dingledine and Golle, 2009] Dingledine R., Golle F. Financial Cryptography and Data Security. 13th International Conference, FC 2009, Accra Beach, Barbados, 23-26.02.2009, Revised Selected Papers, 2009, pp. 238-255.

[Docter et al, 2009] Docter, Q., Dulaney, E., Skandier, T., CompTIA A+ Complete Study Guide: Exams 220-701 (Essentials) and 220-702 (Practical Application), John Wiley and Sons, 2009, pp. 942-944.

[Donoho et al, 2004] Donoho D.L., Elad M., Temlyakov V. Stable Recovery of Sparse Overcomplete Representations in the Presence of Noise. Technical report, Department of Statistics, Stanford University, 2004.

[Drineas et al, 2007] Drineas P., Mahoney M.W., Muthukrishnan S., Sarlos T. Faster least squares approximation. Tech. Rep. 0710.1435, 2007.

[Dubinskaya, 2009] Dubinskaya I. Cyber attack at Citibank. In: www.voanews.com, 2009.

[Dubois et al, 1997] Dubois D., Prade H., Sabbadin R. Decision Under Qualitative Uncertainty with Sugeno Integrals: An Axiomatic Approach. Proc. of IFSA'97, Vol. 2, Prague, 1997, pp.144-147.

[Dubrawsky, 2009] Dubrawsky I. CompTIA Security+: Exam SYO 201, Study Guide and Prep Kit, Syngress, 2009, p.189 and p.432.

[Duran and Booker, 1988] Duran B.S., Booker J.M. A Bayes Sensitivity Analysis when Using the Beta Distribution as a Prior. IEEE Trans. Reliability, 37, No. 2, 1988, pp.239-247.

[Durand, 1993] Durand Y, Brun E., Merindol L., Guyomarc H., Lesaffre B., Martin E. A meteorological estimation of relevant parameters for snow models. Ann. Glaciol.,18, 1993, pp.65–71.

[Dyachenko et al, 2007] Dyachenko O., Kuzemin A., Lyashenko V, Toroyev A. Situational model of creating avalanche and non- avalanche micrositiations. In: Data registration, storing and processing, Vol. 9, N2, 2007, pp.33-41 (in Russian).

[EalarmS, 2010] Information from the official site of EalarmS system. In: www.elarms.org. 2010

[Enc.Britannica, 2009] Pafnuty Lvovich Chebyshev Encyclopædia Britannica. 2009. Encyclopædia Britannica Online. 30 Aug. 2009 <http://www.britannica.com/EBchecked/topic/108214/Pafnuty-Lvovich-Chebyshev>.

[Engl et al, 2000] Engl H.W., Hanke M., Neubaer A. Regularization of Inverse Problems. Kluwer Academic Publishers, Dordrecht, 2000, 321 p.

[Erjavec, 2009] Erjavec J., NATEF Standards Job Sheets Area A2, Cengage Learning, 2009, p.22.

[Ermoliev and Nedeva, 1982] Ermoliev Y., Nedeva C. Stochastic Optimization Problems with Partially Known Distribution Functions. CP-62-60. Laxenburg, Austria: International Institute for Applied Systems Analysis, 1982.

[Ermoliev and Nurminski, 1973] Ermoliev Yu., Nurminski E. Limit extremal problems. Kibernetika, 4, 1973.

[Ermoliev et al, 1985] Ermoliev Yu., Gaivoronski A., and Nedeva S. Stochastic Optimization Problems with Incomplete Information on Distribution Functions. SIAM J. Control. Optim., 23 (5), 1985, pp.697-716.

[Ermoliev et al, 2000a] Ermoliev Y.M., Ermolieva T.Y., Amendola A., MacDonald G., Norkin V.I. A System Approach to Management of Catastrophic Risks. Eur. J. Oper. Res, 122, 2000, pp.452-460.

[Ermoliev et al, 2000b] Ermoliev Y.M., Ermolieva T.Y., MacDonald G., Norkin V.I. Stochastic Optimization of Insurance Portfolios for Managing Exposure to Catastrophic Risks, Annals of Operations Research 99, 2000, pp.207-225.

[Ermoliev, 1970] Ermoliev Y. Methods for Stochastic Programming in Randomized Strategies. Moscow: Kibernetika, Vo1.1, 1970, pp.3-7.

[Ermoliev, 1976] Ermoliev, Y. Methods of Stochastic Programming, Nauka, Moscow, 1976 (in Russian).

[Ermolieva and Ermoliev, 2005] Ermolieva T., Ermoliev Y. Catastrophic Risk Management: Flood and Seismic Risks Case Studies. In: Applications of Stochastic Programming, Eds. S.W.Wallace and W.T. Ziemba, MPS-SIAM Series on Optimization, Philadelphia, PA, USA, 2005.

[Ermolieva et al, 2003] Ermolieva T., Fischer G., Obersteiner M. Integrated Modeling of Spatial and Temporal Heterogeneities and Decisions Induced by Catastrophic Events, IIASA Interim Report IR-03-023, International Institute For Applied Systems Analysis, Laxenburg, Austria, 2003.

[Esl, 2009] Esl I., Introduction to Computer Science, Pearson Education India, 2009, pp.431-434.

[Feret et al, 2008] Feret J.B., François C., Asner G.P., et al. PROSPECT-4 and 5: advances in the leaf optical properties model separating photosynthetic pigments. Ibid, 112, 2008, pp.3030-3043.

[Forex, 2005] The tornado Katrina influences Forex and oil market. In: www.optima-finance.ru. 2005.

[Foster and Kesselman, 2004] Foster I., Kesselman C. The Grid: Blueprint for a New Computing Infrastructure. 2nd Edition, Morgan Kaufmann, 2004.

[Frank and Asuncion, 2010] Frank, A., Asuncion, A. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science. 2010.

[Fritz, 1997] Fritz W. Intelligent Systems and their Societies. e-book: http://www.intelligent-systems.com.ar/intsyst/intsyst.htm. Buenos Aires, Argentina, 1997.

[FTC, 2010] FTC, Case 1:05-cv-00330-SM Document 22 Filed 04/19/2006, http://www.ftc.gov/os/caselist/0423205/060504modifiedpi.pdf

[Fuhn, 1998] Fuhn P. An overview of avalanche forecasting models and methods. In: Oslo, NGI, Pub.N 203, 1998, pp.19–27.

[Fusco et al, 2003] Fusco L., Goncalves P., Linford J., Fulcoli M., Terracina A., Terracina G. Putting Earth-Observation on the Grid. ESA Bulletin, 114, 2003, pp.86-91.

[Gaede and Günther, 1998] Gaede V., Günther O. Multidimensional Access Methods. ACM Computing Surveys, 30(2), 1998.

[Gal et al, 2001] Gal O., Izaca C., Jean F., Lain´e F., L´eveque C., Nguyena A.. CARTOGAM a portable gamma camera for remote localization of radioactive sources in nuclear facilities. Nucl. Instr. and Meth. A, 460, 2001, pp.138–145.

[Galambos et al, 2000] Galambos I., Ermoliev Y., Ermolieva T. Flood Risk Management Policy in the Upper Tisza Basin. A System Analytical Approach. Modeling report, IIASA, Stockholm University and the Hungarian Academy of Sciences, 2000.

[Galatenko, 1999] Galatenko A. Active audit. In: Jet-info, 1999.

[Gao, 1996] Gao B.C. NDWI – a normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment, 58, 1996, pp.257–266.

[GeoEye-1, 2009] www.satimagingcorp.com, GeoEye-1 Satellite Imagery, Sensor Specifications. http://www.satimagingcorp.com/satellite-sensors/geoeye-1.html.

[GEOSS, 2010] The Global Earth Observation System of Systems http://www.earthobservations.org/geoss.shtml.

[Gerashenko, 2009] Gerashenko E. Sberbank's clients pay twice. In: infox.ru, 2009 (in Russian).

[Gerke and Heipke, 2001] Gerke M., Heipke C., Straub B. Building Extraction from Aerial Imagery Using a Generic Scene Model and Invariant Geometric Moments, Proc. IEEE/ISPRS Workshop on Remote Sensing and Data Fusion over Urban Areas, 2001, pp.85-89.

[Germany, 2009] Because of the Spanish payment operator in Germany were withdrawn over hundred thousand bank cards. In: www.xaker.ru. 2009.

[GES, 2009] The accident of the Sayano-Shyshenskay GES. In: www.interfax.ru. 2009.

[Gladun and Rabinovich, 1980] Gladun V., Rabinovich Z. Formation of the World Model in Artificial Intelligence Systems. Machine Intelligence, 9, Ellis Herwood Ltd., Chichester, 1980, pp.299-309.

[Gladun and Vashchenko, 1995] Gladun V., Vaschenko N. Local Statistical Methods of Knowledge Formation. Cybernetics and System Analysis, v.31, N2, 1995, pp.207-217 (in Russian).

[Gladun and Vashchenko, 2000] Gladun V.P., Vaschenko N.D. Analytical Processes in Pyramidal Networks. Int. Journal Information Theories and Applications, Vol.7, No.3, 2000, pp.103-109.

[Gladun et al, 2008] Gladun V., Velichko V., Ivaskiv Y. Selfstructurized Systems. International Journal Information Theories and Applications. FOI ITHEA, Sofia, Vol.15,N.1, 2008, pp.5-13.

[Gladun, 1987] Gladun V.P. Planning of Solutions. Kiev, Naukova Dumka, 1987, 168 p, (in Russian).

[Gladun, 1994] Gladun V.P. Processes of New Knowledge Formation. Sofia, SD Pedagog 6, 1994, 192 p, (in Russian).

[Gladun, 2000] Gladun V.P. Partnership with Computers.. Man-Computer Task-oriented Systems. Kiev, Port-Royal, 2000, 120 p, (in Russian).

[Gladun, 2003] Gladun V.P. Intelligent Systems Memory Structuring. Int. Journal Information Theories and Applications, Vol.10, No.1, 2003, pp.10-14.

[GMES, 2010] The European Earth Observation Programme (GMES). http://ec.europa.eu/gmes/index_en.htm

[Golodnikov and Stoikova, 1978a] Golodnikov A.N., Stoikova, L.S. A Numerical Method to Estimate Certain Functionals Characterizing Reliability. Cybernetics, 14, No.2, 1978, pp.228-234, (transl. from Kibernetika, Kiev, No.2, 73-77,1978).

[Golodnikov and Stoikova, 1978b] Golodnikov A.N., Stoikova L.S. Determination of the Optimum Period of Preventive Replacement on the Basis of Information on Mathematical Expectation and Time Variance of the System Failure-Free Operation Time. Cybernetics, 14, No.3, 1978, 431-440, (translated from Kibernetika, Kiev, No.3, 110-118, 1978).

[Golodnikov et al, 2004] Golodnikov A., Knopov P., Pepelyaev V. Estimation of Reliability Parameters Under Incomplete Primary Information. Theory and Decision, 57, 2004, pp.331-344.

[Golodnikov et al, 2007] Golodnikov A.N., Knopov P.S., Pepelyaev V.A. Some approaches to Pattern Recognition Problems. Cybernetics and Systems Analysis, Volume 43, Issue 6, 2007, pp. 810-821.

[Golodnikov et al, 2009] Golodnikov A., Knopov P. Pepelyaev V. Investigation of Bayesian Estimates for Binomial Failure Models. In Simulation and Optimization Methods in Risk and Reliability Theory. Nova Science Publishers Inc, 2009, pp.173-220.

[Golodnikov, 1979a] Golodnikov A.N. Minimax Approach to Bayes Estimation. In Operations Research (Models, Systems, Solutions), Academy of Science of the USSR, Calculating Center, Moscow, 7, 1979, pp.36-41 (in Russian).

[Golodnikov, 1979b] Golodnikov A.N. Search for the Extremum of a Linear Functional in a Class of Distribution Functions Satisfying Linear Constraints of the Inequality Type. Akad. Nauk Ukrain. SSR Inst. Kibernet. Preprint, 13:7-14, 53,. Stochastic optimization models, 1979.

[Golodnikov, 1982] Golodnikov A.N. Numerical method of minimizing convex functionals in the class of distribution functions satisfying nonlinear constraints. Cybernetics and Systems Analysis (Historical Archive), Volume 18, Number 3, 1982, pp.377-383 (Translated from Kibernetika, No. 3, pp. 93–97, May–June, 1982.)

[Golodnikov, 2007] Golodnikov A.N. Search for the upper bound of Bayesian estimates of the parameter in an exponential failure model with two known quantiles of a priori distribution function. Cybernetics and Systems Analysis, Volume 43, Issue 1, 2007, pp.73-84.

[Golodnikov, 2009] Golodnikov A.N. Investigation of Sensitivity of Bayesian Estimates for Exponential Failure Models. In Simulation and Optimization Methods in Risk and Reliability Theory. Nova Science Publishers Inc, 2009, pp.221-238.

[Gorskii, 1985] Gorskii D. Generalization and knowledge. Mysl-Moscow, 1985, 208 pp. (in Russian)

[Grabner, 2008] Grabner H., Nguyen T.T., Gruber B., Bischof H., On-line boosting-based car detection fron aerial images. ISPRS Journal of Photogrammetry and Remote Sensing, 2008.

[Grabovsky, 2009] Scientists: the snowfall in Saint-Petersburg was not an anomaly. In: www.gazeta.spb.ru. 2009.

[Green et al, 1988] Green A., Berman M., Switzer P., Craig M. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. IEEE Transactions on Geoscience and Remote Sensing, 26(1), 1988, pp.65-74.

[Gribonval et al, 2006] Gribonval R., Figueras R.M., Ventura P. Vandergheynst. A simple test to check the optimality of sparse signal approximations. Signal Processing, 86(3), 2006, pp.496-510.

[Grigolia, 1994] Grigolia G. Selection and analysis of the observation data to calculation of maximum discharge. Tbilisi, TSU, 1994.

[Grigoriev et al, 2006] Grigoriev D., Hirsch E.A., Pervyshev K. A Complete Public-Key Cryptosystem, ECCC Report TR06-046, April, 2006.

[Grunvald and Vitanyi, 2003] P.D. Grunvald, P.M.B. Vitanyi. Kolmogorov complexity and information theory. With an interpretation in terms of questions and answers. J.Logic, Language and Information, 12(4), 497-529, 2003.

[Gunn, 1978] Gunn P.J. Inversion of airborne radiometric data. Geophysics, 43, 1978, pp.133–143.

[Guoa et al, 2004] Guoa W., Gardnera R.P., Todd A.C. Using the Monte Carlo – Library Least-Squares (MCLLS) approach for the in vivo XRF measurement of lead in bone. Nuclear Instruments and Methods in Physics Research, A 516, 2004, pp.586–593.

[Gusyatin, 2000] Gusyatin V.M., Ostroushko A.P., Filimonchuk M.A., Yankovsky A.A. Special graphic processor for the visualization systems. Author's certificate of application № 2000031738 Ukraine, IPC G06F7/00 / 28.03.2000.

[Gusyatin, 2002] Gusyatin V.M., Yankovsky A.A. The computing system of flying machines visual control.: Patent № 56876A, Application №2002108289, IPC 7 B64F1/18, G06F7/548 / 18.10.2002

[Haarbrink et al, 2007] Haarbrink R., Shutko A., Novichikhin E., Sidorov I., Milenova L., Vassilev V. Rapid Multi-Sensor System for Effective Risk Analyses. Conference on Geomatics in support of the Common Agriculture Policy. Madrid, Spain, 2007.

[Haiti, 2010a] New powerful earthquake occurred in Haiti. In: www.rbc.ru, 2010.

[Haiti, 2010b] Earthquake on Haiti wasn't a surprise for scientists. In: top.rbc.ru, 2010.

[Haldar and Aravind, 2009] Haldar S., Aravind A. Operating Systems, Pearson Education India, 2009, pp. 373-374.

[Halko et al, 2009] Halko N., Tropp J.A., Martinsson P.G. Finding structure with randomness: Stochastic algorithms for constructing approximate matrix decompositions. In: ACM Report 2009-05, Caltech, 2009.

[Hall, 1979] Hall J.A. Signal and noise at the image transfer, In book: Semiconductor Image Sensors. Ed. P. Yespers, F. Van de Vile and M. White: transl. from English, M.: Mir, 1979, 573 p.

[Han et al, 2000] Han J., Pei J., Yin Y. Mining Frequent Patterns without Candidate Generation. Proc. 2000 ACM-SIGMOD Int. Conf. on Management of Data, 2000, pp.1-12.

[Hansel, 1966] Hansel G. Sur le nombre des functions booleennes monotones de n variables. C.R. Acad. Sci. Paris, 262, serie A, 1966, p.1088.

[Hansen and O'Leary, 1993] Hansen P.C., O'Leary D.P. The use of the L-curve in the regularization of discrete ill-posed problems. SIAM J. Sci. Comput. 14, 6, 1993, pp.1487-1503.

[Hansen and Yu, 2001] Hansen M., Yu B. Model selection and minimum description length principle. J. Amer. Statist. Assoc, 96, 2001, pp.746-774.

[Hansen, 1998] Hansen P.C. Rank-deficient and discrete ill-posed problems. Numerical Aspects of Linear Inversion. SIAM, Philadelphia, 1998, 247 p.

[Haritonov, 2010] Haritonov S. Everyone around shouted that it was the end of the world. In: www.utro.ru, 2010.

[Harkevich, 1960] Harkevich A. On information value. Cybernetics problems, N 4, Physmathgiz, Moscow, 1960.

[Harrah, 1956] Harrah D. A model of communication. Philosophy of science, Vol 23, N 4, 1956, pp.333-342.

[Harrah, 1957] Harrah D. The psychological concept of information. Philosophy and Phenomenological Research, Vol 18, N 2, 1957, pp.242-249.

[Hartley, 1928] Hartley R. Transmission of information. Bell system technical journal, N 7, 1928, pp.335-363.

[Haykin, 1999] Haykin S. Neural Networks: A Comprehensive Foundation. Upper Saddle River, New Jersey: Prentice Hall, 1999.

[HEC-FDA, 1997] HEC-FDA (Hydrologic Engineering Center Flood Damage Analysis) User's manual. – Hydrologic Engineering Center, US Army Corps of Engineers, Davis, CA, 1997.

[Hefty et al, 2009] Hefty J., Gerhatova L., Becker M., Drescher R., Stangl G., Krauss S., Caporali A., Liwosz T., Kratochvil R. Long-Term Densification of Terrestrial Reference Frame in Central Europe as the Result of Central Europe Regional Geodynamic Project 1994-2006, in H. Drewes (Ed.), Geodetic Reference Frames, IAG Symposium Munich, Germany, 9-14.10.2006, International Association of Geodesy Symposia Volume 134, Springer Berlin-Heidelberg 2009, pp.149-154.

[Hendriks et al, 2001] Hendriks P.H., Limburg J., de Meijer R.J. Full-spectrum analysis of natural gamma-ray spectra. J Environ Radioact, 53(3), 2001, pp.365-380.

[Hendriks et al, 2002] Hendriks P.H., Maucec M., de Meijer R.J. MCNP modeling of scintillation-detector gamma-ray spectra from natural radionuclides. Appl Radiat Isot., 57(3), 2002, pp.449-457.

[Hinz, 2004] Hinz S., 2004. Detection of vehicles and vehicle queues in high resolution aerial images. Photogrammetrie – Fernerkundung – Geoinformation (PFG) 3/04, 2004, pp.201-213.

[Hollenstein, 2007] Hollenstein C. GPS deformation field and geodynamic implications for the Hellenic plate boundary region. Zurich, 2007.

[Holt and Chung, 2002] Holt J.D., Chung S.M. Mining association rules using inverted hashing and pruning. Information Processing Letters Archive, Volume 83, Issue 4, 2002, pp.211-220.

[Horritt, 1999] Horritt M.S. A statistical active contour model for SAR image segmentation. Image and Vision Computing, 17, 1999, pp.213-224.

[Horritt, 2006] Horritt M.S. A methodology for the validation of uncertain flood inundation models. J of Hydrology, 326, 2006, pp.153-165.

[Hovgaard and Grasty, 1997]. Hovgaard J., Grasty R.L. Reducing statistical noise in airborne gamma-ray spectra through spectral component analysis. In Gubins, A.G., Ed. Proceedings of Exploration 97: Fourth Decennial Conference on Mineral Exploration, 1997, pp.753- 764.

[Hyvarinen, 1999] Hyvarinen A. Fast and robust fixed-point algorithms for Independent Component Analysis. IEEE Trans. On Neural Networks, 10(3), 1999, pp.626-634.

[IAEA, 2002] Safety culture in nuclear installations. Guidance for use in the enhancement of safety culture. IAEA, Vienna, 2002.

[IAEA, 2003] International Atomic Energy Agency IAEA-TECDOC-1363 2003 Guidelines for radioelement mapping using gamma-ray spectrometry data, 2003.

[IAEA, 2004] IAEA General Conference, IAEA, Vienna, 2004.

[IAEA, 2007] Knowledge management now seen as a priority. IAEA, Nuclear news, Vienna, Sept.2007.

[IAEA-TECDOC-1335, 2003] Configuration management in nuclear power plants. IAEA-TECDOC-1335, Wien, 2003.

[IBM, 1965-68] IBM System/360, Disk Operating System, Data Management Concepts. IBM System Reference Library, IBM Corp. 1965, Major Revision, Feb.1968.

[IEOS,2005] Images of the Earth and Outer Space: application examples: Moscow, LTD Engineering and Technological Center ScanEcs, 2005.

[Ilgum and Kemmerer, 1995] Ilgum K., Kemmerer R.A. State Transition Analysis: ARule-Based Intrusion Detection System. In: IEEE Trans. Software Eng., vol. 21, no. 3. 1995.

[Inokuchi et al, 2003] Inokuchi A., Washio T., Motoda H. Complete mining of frequent patterns from graphs: Mining graph data. In Machine Learning, Volume 50, 2003, pp.321-354.

[Italy, 2009a] Italian places of interest suffered during the earthquake. In: www.travel.ru. 2009.

[Italy, 2009b] Strong earthquake in Italy killed 150 people, more than 50 000 people lost their houses. In: rus.newsru.ua. 2009.

[Jaynes, 1957] Jaynes E. Information theory and statistical mechanics. Physical review, Volume 106, N4, 1957, pp.620-630.

[Jaynes, 1968] Jaynes E. Prior Probabilities. IEEE Transactions on System Science and Cybernetics, SSC-4, 1968, pp.227-241.

[Jeffreys, 1961] Jeffreys H. Theory of Probability, (3rd ed.), Clarendon Press, Oxford, 1961.

[Jogova, 2010] Jogova N. Banks block clients' plastic cards because of data mass leak. In: www.rb.ru. 2010.

[Johnson and Lindenstrauss, 1984] Johnson W.B., Lindenstrauss J. Extensions of Lipschitz mappings into a Hilbert space. Contemporary Mathematics, 26,1984, pp.189-206.

[Jorion, 1996] Jorion P.H. Value at Risk: A New Benchmark for Measuring Derivatives, New York: Irwin Professional Publishers, 1996, 284 p.

[Karkischenko, 1998] Karkischenko A., Lepskiy A., Bezuglov A. An approach for vector and analytical image contour representation. Proceeding to All-Russian Scientific and Technical conference Intellection SAPR–97, 1998, p. 107-112.

[Karlin and Studden, 1966] Karlin S., Studden W.J. Tchebycheff Systems: With Applications in Analysis and Statistics. Wiley Interscience, New York, 1966.

[Kartalopoulos, 2009] Kartalopoulos S. Security of Information and Communication Networks, Wiley-IEEE. 2009, pp.24-25.

[Katrina, 2005] Inside the hurricane Katrina. In: http://www.1tv.ru/documentary/fi=6149. 2005 (in Russian).

[Kaufman, 2010] Kaufman S. Iceland Volcano Has Global Economic Impact. In: www.america.gov. 2010

[Kautsky and Hirsch, 1931] Kautsky H., Hirsch A. Neue Versuche zur Kohlenstoffassimilation. Natur wissenenschaften, 1931.

[Kautsky and Hirsch, 1934] Kautsky H., Hirsch A. Das Fluoreszenzverhalten gruner Pflanzen. Biochem Z, 1934, pp.422–434.

[Kazakov, 2010] Kazakov K. We pay for the leak of our data. In: http://www.ves.lv/article/112941, 2010 (in Russian)

[Kirilyuk, 2003] Kirilyuk V.S. On coherent risk measures and portfolio optimization problem. Teoriya optimalnth rishen, Kyiv: V.M.Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine, 2, 2003, pp.111–119.

[Kirilyuk, 2004a] Kirilyuk V.S. On a class of polyhedral coherent risk measures. Kibernetika I sistemnyi analiz, N.4, 2004, pp.155–167.

[Kirilyuk, 2004b] Kirilyuk V.S. On one generalization of a polyhedral coherent risk measure. Teoriya optimalnth rishen, Kyiv: V.M.Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine, N.3, 2004, pp.48-55.

[Kirilyuk, 2006] Kirilyuk V.S. Optimal Decisions in Conditions of Risk on the Basis of Technique of Set-Valued Maps.– Dissertation for Doctor of Sciences' Degree in Informatics and Cybernetics, Kyiv, 2006, 307 p. (in Ukrainian).

[Kirilyuk, 2007] Kirilyuk V.S. Polyhedral Coherent Risk Measures, Portfolio Optimization and Investment Allocation Problems, IIASA Interim Report, 2007, IR-07-030, Laxenburg, Austria, 2007, 21 p.

[Kirilyuk, 2008] Kirilyuk V.S. Polyhedral Coherent Risk Measures and Portfolio Investment Optimization. Kibernetika i sistemnyi analiz, N.2, 2008, pp.120–133.

[Kivva and Zheleznyak, 2005] Kivva S.L., Zheleznyak M.I. Two-Dimensional Modeling of Rainfall Runoff and Sediment Transport in Small Catchment Areas. International Journal of Fluid Mechanics Research, 32/6, 2005, pp.703-717.

[Knight, 1921] Knight F.H. Risk, Uncertainty and Profit, Houghton Miffin: Boston, 1921, 381 p.

[Knyazikhin et al, 1998] Knyazikhin et al. Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data, J. Geophys. Res., 1998.

[Kobzar, 2008] Kobzar G.A. Curvature Inter-Scale Space model for geometric object shape representation. Scientific and Technical Journal Artificial Intelligence №1 (08), Donetsk: Science and Education, 2008, pp.153– 165.

[Kogan et al, 2004] Kogan F., Stark R., Gitelson A., Adar E., Jargalsaikhan L., Dugrajav C., Tsooj S.. Derivation of Pasture Biomass in Mongolia from AVHRR-based Vegetation Health Indices. Int. J. Remote Sens, 25(14), 2004, pp.2889-2896.

[Kohonen, 1995] Kohonen T. Self-Organizing Maps. Series in Information Sciences, Vol. 30. Springer, Heidelberg, 1995.

[Kolev et al, 1986] Kolev N., Kirkova Y. et al. Physical methods and technical devices for evaluation of soil moisture, Intern. Agrophysics, 1, 1986, pp.107-114.

[Kolmogorov, 1965] Kolmogorov A. Three approaches to the definition of information content. Problems of information communication, N 1, Moscow, 1965, pp.3-11.

[Kondratyev et al, 2002] Kondratyev K., Krapivin V., Phillips G. Global Environmental Change: Modelling and Monitoring. Springer-Verlag, Berlin, 2002, pp. 243-258.

[Konno and Yamazaki, 1991] Konno H., Yamazaki H. Mean Absolute Deviation Portfolio Optimization Model and Its Application to Tokyo Stock Market. Management Science, 37, 1991, pp. 519–531.

[Kopp et al, 2007] Kopp P., Petiteville I., Shelestov A., Li G.. Wide Area Grid (WAG). In: Proc. The 7th Ukrainian Conference on Space Research, National Flight and Control Center, Evpatoria, Ukraine, 2007, p.209.

[Kotzev et al, 2006] Kotzev V., Nakov R., Georgiev Tz., Burchfiel B.C., King R.W. Crustal motion and strain accumulation in western Bulgaria, Tectonophys. 413 (3-4), 2006, pp.127-145.

[Kovalev, 1988] Kovalev A., Tarasov U. Texture on the arbitrary oriented flat surfaces. Autometry. N.6, 1988, pp.46-51.

[Kovalev, 1991] Kovalev A., Tarasov U. Increase of texture images clearness on flat surfaces. Autometry, N.3, 1991, pp.3-9.

[Krapivin and Phillips, 2001] Krapivin V., Phillips G. A remote-sensing based expert system to study the Aral-Caspian aquageosystem water regime. Remote Sensing of the Environment, 75, 2001, pp.201-215.

[Krapivin et al, 1996] Krapivin V., Long B., Rochon G., Hicks D. A global simulation model as a method for estimation of the role of regional area in global change. Proc. of the Second Ho Chi Minh City Conf. on Mechanics, 24-25.09.1996, pp.68-69.

[Krasovsky, 1995] Krasovsky A.A. Base principles of aviation trainers, M.: Knowledge, 1995, 303 p.

[Kravchenko, 2009] Kravchenko A. Neural networks method to solve inverse problems for canopy radiative transfer models. Cybernetics and System Analysis, N 3, 2009, pp.159-172 (in Russian)

[Krein and Nudelman, 1977] Krein M., Nudelman A. The Markov Moment Problem and Extremal Problems. Trans. Math. Monographs, American Mathematical Society, Providence, RI, 1977.

[Krissilov A. and V. Krissilov, 2005] Krissilov A., Krissilov V. Creation of Aim-oriented vector model applying to the estimation of complicate socio-ecological objects. Monography "Methods of environm. problems solving", ed. Prof. L. Melnik, Sumy, "Kozatskij Val", 2005 (inRussian).

[Krissilov et al, 2000] Krissilov A., Krissilov V., Shutko A. Decision Making Procedures that Operate with Dependent Features and Their Environmental Applications. Proc. of 5th Int. Conf. ITA 2000, Vol.7, No 4, 2000.

[Krissilov et al, 2001] Krissilov A., Krissilov V., Shutko A., B Blyukher. Monitoring Tasks in Touristics Area: Problems, Measuring and Forming the Geoinformation Monitoring System (GIMS). Journ. of Env. Protect. and Ecology, Special Issue, Sofia and Thessaloniki, 2001.

[Krissilov et al, 2007] Krissilov A., Shutko A., Baryshnikov I., Kostenyuk B. Forming of Geo-Information Environmental Monitoring System for management and harmonious use of ecological Belts surrounding large cities. Proc of International Workshop on Environmental Problem in Metropolitan Cities; Istanbul, Turkey, June 2007.

[Krissilov V. and A. Krissilov, 2000] Krissilov V., Krissilov A. High-Quality Decision Making by Aim-Oriented Modelling. Proc. of 19th International Conference on Fuzzy Sets Theory and Its Applications (NAFIPS), USA, Florida, 2000.

[Krissilov V. et al, 1998] Krissilov V., Krissilov A., Tarasenko R. Transformation of Object Feature Space Under the Goal of Evaluation. Proc. of Conf. "Inform. Processing and Management of Uncertainty in Knowledge-Based Systems" IPMU'98, Paris, 1998.

[Krissilov, 1962] Krissilov A. Synthesis method of reading computer device. Inventor's Certificate № 152248, Inv. Bull. № 24, M., 1962; (co-auth. – M. Gliklikh, G. Poddubnyj, rus).

[Krissilov, 1984] Krissilov A. Algorithms for the Quantitatively Grounded Decisions in Situations under Uncertain Conditions. Znanije, Kijev, 1984 (in Russian).

[Krissilov, 1999] Krissilov A. Towards a new economic-ecological order for the Black Sea region: organizing, socio-economic and technical aspects of international geoinformation monitoring system. Proc. of Int. Leadership Seminar "Using today's scientific knowledge for the Black Sea area's development tomorrow", Mamaia, Romania, 1999; pp.65-76.

[Kuchment and Gelfan, 1993] Kuchment L.S., Gelfan A.N. Dynamic stochastic Models of River Runoff Generation, Nauka, Moscow, 1993.

[Kuchment and Gelfan, 2002] Kuchment L.S., Gelfan A.N. Estimation of Extreme Flood Characteristics Using Physically Based Models of Runoff Generation and Stochastic Meteorological Input, Water International, Vol. 27, 2002, pp.77-86.

[Kuncheva, 2004] Kuncheva L. Combining Pattern Classifiers: Methods and Algorithms. Willey, ISBN: 978-0-471-21078-8, 2004, 376 pp.

[Kuramochi and Karypis, 2001] Kuramochi M., Karypis G. Frequent subgraph discovery. In Proceedings of the 1st IEEE International Conference on Data Mining (ICDM'01), 2001, pp.313-320.

[Kussul et al, 2007] Kussul N., Lupian E., Shelestov A., Skakun S., Tischenko Yu., Hluchy L. Flood extent extraction using data from different sources, J. of Automation and Information Sciences, Issue 6, 2007, pp. 117-126.

[Kussul et al, 2008a] Kussul N., Shelestov A., Skakun S. Grid System for Flood Extent Extraction from Satellite Images. Earth Science Informatics, 1(3-4), 2008, pp.105-117.

[Kussul et al, 2008b] Kussul N., Shelestov A., Skakun S., Kravchenko O. Data Assimilation Technique For Flood Monitoring and Prediction. International Journal on Information Theory and Applications, 15(1), 2008, pp.76-84.

[Kuzemin and Lyashenko, 2006] Kuzemin A., Lyashenko V. Fuzzy set theory approach as the basis of analysis of financial flows in the economical security system in: International Journal Information Theories and Applications, 13/1, 2006, pp.45–51.

[Kuzemin and Lyashenko, 2007a] Kuzemin A., Lyashenko V. Procedure of formalization of the indices of banks stable functioning in comparative estimates of their development. In: International Journal Information Technologies and Knowledge, V1/N2, 2007, pp. 175–181.

[Kuzemin and Lyashenko, 2007b] Kuzemin A., Lyashenko V. Probabilistic and multivariate aspects of construction of the models and procedures for prediction of the avalanche-dangerous situations initiation in: Fifth International Conference Information Research and Applications i.TECH 2007, 26-30.06.2007, Varna, Bulgaria, Vol.2., 2007, pp.284–288.

[Kuzemin and Toroev, 2006] Kuzemin A., Toroev A. Mobile means of control and prediction of avalanche climate using information conversion in acoustic range. in: IDRC, DAVOS, Vol. 2, 2006, pp.291–294.

[Kuzemin et al, 2005] Kuzemin A., Lyashenko V., Bulavina E., Torojev A. Analysis of movement of financial flows of economical agents as the basis for designing the system of economical security (general conception) In: Third international conference Information research, applications, and education. Varna, Bulgaria, 2005, pp.204–209.

[Kuzemin et al, 2007a] Kuzemin A., Lyashenko V., Toroyev A., Klymov I. Developing an expert system for situational analysis of avalanche danger. In: Fifth International Conference Information Research and Applications i.TECH 2007, 26-30.06.2007, Varna, Vol. 2, 2007, pp.294–297.

[Kuzemin et al, 2007b] Kuzemin A, Lyashenko V., Fastova D. Interpretational model for analyzing the environment of the avalanche climate. In: Data registration, storing and processing, Vol. 9, N1, 2007, pp.27-34.

[Larman, 2004] Larman C. Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design and Iterative Development, 3rd Edition, Prentice Hall, 2004, 736 p.

[Latvia, 2010a] In Service state of incomes of Latvia there was an unprecedented information leak. www.regnum.ru. 2010.

[Latvia, 2010b] Guilty of information leak in Latvia will be punished. In: www.riga-lv.com. 2010.

[Lavine et al, 1993] Lavine M., Wasserman L., Wolpert R. Linearization of Bayesian Robustness Problems. Journal of Statistical Planning and Inference, 37, 1993, pp.307-316.

[Lavine, 1991] Lavine M. Sensitivity in Bayesian Statistics: the Prior and the Likelihood. Journal of the American Statistical Association, 86, 1991, pp.396-399.

[Lazebnik, 2005] Lazebnik S., Schmid C., Ponce J., A sparse texture representation using local affine regions. IEEE Trans. Pattern Anal. Mach. Intell. 27, 2005, pp.1265-1278.

[Lbov and Starceva, 1994] Lbov G.S., Starceva N.G. Complexity of Distributions in Classification Problems. Proc. RAS, Vol 338, No 5, 1994, pp.592-594.

[Lbov and Starceva, 1999] Lbov G.S., Starceva N.G. Logical Decision Functions and Questions of Statistical Stability. Inst. Of Mathematics, Novosibirsk, 1999.

[Lbov and Stupina, 1999] Lbov G.S., Stupina T.A. Some Questions of Stability of Sampling Decision Functions, Pattern Recognition and Image Analysis, Vol 9, 1999, pp.408-415.

[Lbov and Stupina, 2002] Lbov G.S., Stupina T.A. Performance criterion of prediction multivariate decision function. Proc. of international conference Artificial Intelligence, Alushta, 2002, pp.172-179.

[Lbov and Stupina, 2003] Lbov G.S., Stupina T.A. To statistical stability question of sampling decision function of prediction multivariate variable. Proc. of the seven international conference PRIP'2003, Minsk, Vol. 2, 2003, pp.303-307.

[Lei et al, 2008] Lei Z., Li D., Fang T. Vehicle detection in high-resolution satellite imagery using sift features and support vector machines. ISPRS Journal of Photogrammetry and Remote Sensing, 2008.

[Lemon and Krutchkoff, 1969] Lemon G.H., Krutchkoff R.G. An Empirical Bayes Smoothing Technique. Biometrika, 56, 1969, pp.361-365.

[Leonov, 1977] Leonov U.P. Theory of statistical decisions and psychophysics, M.: Science, 1977, 228 p.

[Li et al, 2001] Li W., Han J., Pei J. CMAR: Accurate and efficient classification based on multiple class-association rules. In IEEE Int. Conf. on Data Mining (ICDM'01), 2001, pp.369-376.

[Li et al, 2004] Li Y., Cichocki A., Amari S. Analysis of sparse representation and blind source separation. Neural Computation, 16(6), 2004, pp.1193–1204.

[Liang, 2004] Liang S. Quantitative Remote Sensing of Land Surfaces, Wiley, Inc., 2004, 534 p.

[Liang, 2008] S. Liang (ed.). Advances in Land Remote Sensing. Springer, 2008.

[Lin, 1972] Lin P.E. Rates of Convergence in Empirical Bayes Estimation Problems. Discrete Case. Annals of the Institute of Statistical Mathematics, 24, 1972, pp.319-325.

[Linderberg, 1994] Linderberg T. Scale Space Theory in Computer Vision, Kluwer Academic Publishers, 1994, 440 p.

[Linnerooth-Bayer and Amendola, 2003] Linnerooth-Bayer J., Amendola A. (Eds.) Special Issue on Flood Risks in Europe, Risk Analysis, Vol. 23, Issue 3, 2003.

[Liu and Todini, 2002] Liu Z., Todini E. Towards a comprehensive physically-based rainfall-runoff model, Hydrology and Earth System Sciences, 6(5), 2002, pp.859–881.

[Liu et al, 1998] Liu B., Hsu H., Ma Y. Integrating classification and association rule mining. In 4th Int. Conf. on Knowledge Discovery and Data Mining (KDD'98), 1998, pp.80-86.

[Liu et al, 2003] Liu G., Lu H., Yu J., Wang W., Xiao X. AFOPT: An Efficient Implementation of Pattern Growth Approach. In Workshop on Frequent Itemset Mining Implementations (FIMI' 03) in Conjunction with IEEE ICDM'03, 2003.

[Lloyd, 1975] Lloyd J.M. Thermal Imaging Systems, New York: Plenum Press, 1975, 456 p.

[Lukashin, 2003] Lukashin Y.P. Adaptive short time prediction methods for time series. Publ. Moskow, Finances and Statistika, 2003, 416 p.

[Malicki and Skierucha, 1989] Malicki M., Skierucha W. A manually controlled TDR soil moisture meter operating with 300 ps rise-time needle pulse. Irrrigation Science, 10, 1989.

[Mallat and Zhang, 1993] Mallat S., Zhang Z. Matching pursuit with time frequency dictionaries. IEEE Trans. Signal Proc., 41(12), 1993, pp.3397-3415.

[Mallows, 1973] Mallows C.L. Some comments on Cp. Technometrics, 15(4), 1973, pp.661-675.

[Maritz, 1966] Maritz J.S. Smoothed Empirical Bayes Estimation for One-Parameter Discrete Distributions. Biometrika, 53, 1966, pp.417-429.

[Maritz, 1970] Maritz J.S. Empirical Bayes Methods, Methuen, London, 1970.

[Markman et al, 2003] Markman A.B., Rachkovskij D.A., Misuno I.S., Revunova E.G. Analogical reasoning techniques in intelligent counterterrorism systems. International Journal Information Theories and Applications, Sofia, Bulgaria, 10(2), 2003, pp.139-146.

[Markov et al, 2006] Markov Kr., Kr. Ivanova, I. Mitov. Basic Structure of the General Information Theory. IJ ITA, Vol.14, No.: 1, 2006. pp.5-19.

[Markov et al, 2008] Markov Kr., Ivanova Kr., Mitov I., Karastanev S. Advance of the Access Methods. International Journal Information Technologies & Knowledge, Volume 2, Number 2, 2008, pp. 123-135

[Markov et al, 2009] Markov Kr., Kr. Ivanova, I. Mitov. Theory of Infos. Int. Book Series "Information Science & Computing" – Book No: 13. Intelligent Information and Engineering Systems, ITHEA, Sofia, 2009, pp.9 16.

[Markov, 1898] Markov A.A. About limit values of integrals in connection with interpolation. Zapiski Akademii nauk, VI, VIII, № 5, 1898.

[Markov, 1984] Markov Kr. A Multi-domain Access Method. Proceedings of the International Conference on Computer Based Scientific Research. Plovdiv, 1984, pp.558-563.

[Markov, 2004] Markov Kr. Multi-Domain Information Model. International Journal "Information Theories and Applications", Vol. 11, No: 4, 2004, pp.303-308.

[Markov, 2005] Markov Kr. Building Data Warehouses Using Numbered Multidimensional Information Spaces. Int. Journal "Information Theories and Applications", Vol. 12, No. 2, 2005, pp. 193-199.

[Markowitz, 1959] Markowitz H.M. Portfolio Selection, Efficient Diversification of Investment, Wiley: NY, 1959, 344 p.

[Martin, 1975] Martin J. Computer Data-Base Organization. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1975.

[Martinez and Le Toan, 2007] Martinez J.M., Le Toan T. Mapping of flood dynamics and spatial distribution of vegetation in the Amazon floodplain using multitemporal SAR data. Remote Sensing of Environment, 108, 2007, pp.209-223.

[Martz and Krutchkoff, 1969] Martz H.F., Krutchkoff R.G. Empirical Bayes Estimation in a Multiple Linear Regression Model. Biometrika, 56, 1969, pp.367-374.

[Martz and Waller, 1991] Martz, H.F., Waller, R.A. Bayesian Reliability Analysis. Krieger publishing company, Malabar, Florida, 1991.

[Mathers and Crowley, 1904] S. L. MacGregor Mathers, A. Crowley, The Goetia: The Lesser Key of Solomon the King (1904). 1995 reprint: ISBN 0-87728-847-X.

[Mauring and Smethurts, 2005] Mauring E., Smethurts M. Reducing noise in radiometric multi-channel data using noise-adjusted singular value decomposition (NASVD) and maximum noise fraction (MNF). Geological Survey of Norway, Report no.:2005.014, 2005.

[Mayer, 1999] Mayer H. Automation Object Extraction from Aerial Imagery–A Survey Focusing on Buildings, Computer vision and image understanding, col. 74, no.2, 1999, pp.138-149.

[Mazurok et al, 1994] Mazurok B.S., Rogkov A.F., Salnykov U.A., Tessen U.E., Unry P.I. Generation of teksturirovannykh surfaces and specialized effects in the systems Albatross, Autometry, N6, 1994, pp.57-61.

[Merz et al, 2004] Merz B., Kreibich H., Thieken A., Schmidtke R. Estimation uncertainty of direct monetary flood damage to buildings; Natural Hazards and Earth System Sciences, 4, 2004, pp.153–163.

[Messner and Meyer, 2005] Messner F., Meyer V. Flood damage, vulnerability and risk perception – challenges for flood damage research; UFZ discussion paper, 2005.

[Metoffice, 2010] http://www.metoffice.gov.uk/aviation/vaac/data/VAG_140948.png

[Michalski et al, 1986] Michalski S., Carbonell G., Mitchell M.(eds) Machine Learning, an Artificial Intelligence Approach.-Morgan Kaufmann, San Mateo, California, v.1,2, 1986.

[Mike-11, 2004] Mike-11 Reference Manual. DHI Water and Environment, 2004 (http://www.dhisoware.com/mikell)

[Mikolajczyk and Schmid, 2002] Mikolajczyk K., Schmid C. An affine invariant interest point detector, ECCV, 2002, pp.128-142.

[Milev and Dabovski, 2006] Milev G., Dabovski Ch. (Eds) Geodynamics of the Balkan Peninsula. Monograph. Report on Geodesy, Warsaw University of Technology, No. 5 (80), 2006, 650 p.

[Milev and Vassileva, 2003] Milev G., Vassileva K. Report on the Project Activities of Bulgaria Warsaw university of technology, Institute of geodesy and geodetic astronomy. Reports on Geodesy. No 3 (66), 2003, pp.129-138.

[Milev and Vassileva, 2004] Milev G., Vassileva K. The Project Cergop-2 and its Realization for the Balkan Peninsula. Report on the Serbian Geological Society. Belgrade, 3.03.2004, p.13.

[Milev and Vassileva, 2009] Milev G., Vassileva K. Development of the Global navigation satellite systems and their application in Bulgaria. Technospere, 3(7), 2009, pp.29-40 (in Bulgarian)

[Milev et al, 2005] Milev G., Vassileva K., Dimitrov N. Bulgarian Geodynamics Cergop-2 Activities – Vienna Progress Report 2005. Reports on Geodesy, Warsaw University of Technology, No 2(73), 2005, pp.41-48.

[Milev et al, 2009] Milev G., Vassileva K., Milev I. Development and application of DGNSS in Bulgaria. Proceedings of the Int. Symposium on GNSS, Space-based and Ground-based Augmentation systems and applications, Berlin, Germany, 2009, pp.73-77.

[Miller and Gregory, 2009] Miller L., Gregory P. CISSP for Dummies. For Dummies, 2009, p.76.

[Minty and Hovgaard, 2002] Minty B.J., Hovgaard J. Reducing noise in gamma-ray spectrometry using spectral component analysis. Exploration Geophysics, 33, 2002, pp.172-176.

[Minty and McFadden, 1998] Minty B., McFadden P. Improved NASVD smoothing of airborne gamma-ray spectra. Expl. Geophys., 29, 1998, pp.516–523.

[Minty, 1992] Minty B. Airborne gamma-ray spectrometric background estimation using full spectrum analysis: Geophysics, 57, 1992, pp.279–287.

[Misuno et al, 2005] Misuno I.S., Rachkovskij D.A., Slipchenko S.V. Vector and distributed representations reflecting semantic relatedness of words. Mathematical Machines and Systems, Issue 3, 2005, pp.50–66 (in Russian).

[Mitov et al, 2009a] Mitov I., Ivanova Kr., Markov Kr., Velychko V., Vanhoof K., Stanchev P. "PaGaNe" – A Classification Machine Learning System Based on the Multidimensional Numbered Information Spaces. In World Scientific Proceedings Series on Computer Engineering and Information Science, No:2, 2009, pp. 279-286.

[Mitov et al, 2009b] Mitov I., Ivanova Kr., Markov Kr., Velychko V., Stanchev P., Vanhoof K. Comparison of Discretization Methods for Preprocessing Data for Pyramidal Growing Network Classification Method. Int. Book Series "Information Science & Computing" – Book No: 14. New Trends in Intelligent Technologies, Sofia, 2009, pp. 31-39.

[Mitov et al, 2010] Mitov I., Kr. Markov, Kr. Ivanova. The Intelligence. Third International Scientific Conference "Informatics in the Scientific Knowledge". University Publishing House, VFU "Chernorizets Hrabar", 2010. ISSN: 1313-4345. pp. 7-13.

[Miyawasa, 1961] Miyawasa K. An Empirical Bayes Estimator of the Mean of a Normal Population. Bulletin of the International Statistical Institute, 39, 1961, pp.181-188.

[Moënne-Loccoz, 2005] Moënne-Loccoz N. High-Dimensional Access Methods for Efficient Similarity Queries. Technical Report N:0505, University of Geneva, Computer Vision and Multimedia Laboratory, May 2005.

[Mokbel et al, 2003] Mokbel M.F., Ghanem T.M., Aref W.G. Spatio-temporal Access Methods. IEEE Data Engineering Bulletin, 26(2), 2003, pp.40-49.

[Morishita and Sese, 2000] Morishita S., Sese J. Transversing itemset lattices with statistical metric pruning. In PODS ’00: Proc. of the 19th ACM SIGMOD-SIGACT-SIGART symposium on Principles of Database Systems, New York, 2000, pp.226-236.

[Morozov, 1984] Morozov V.A. Methods for Solving Incorrectly Posed Problems, Springer Verlag, New York, 1984.

[Mosleh and Apostolakis, 1982] Mosleh A., Apostolakis G. Some Properties of Distributions Useful in the Study of Rare Events. IEEE Transactions on Reliability, R-31, 1982, pp.87-94.

[Nakagawa et al, 2002] Nakagawa M., Shibasaki R., Kagawa Y. Fusing stereo linear CCD image and laser range data for building 3d urban model, ISPRS Symposium Geospatial Theory, Processing and Applications Ottawa, Canada, July 9-12, 2002.

[Nakov et al, 2006] Nakov R., Kotzev V., Burchfiel B., King R. Crustal motion in Bulgaria based on geological and GPS data. Geodynamics of the Balkan Peninsula. Monograph. Report on Geodesy, Warsaw University of Technology, No. 5 (80), 2006, pp.205-212.

[Nemenyi, 1963] Nemenyi P. Distribution-free multiple comparisons. PhD thesis, Princeton University, 1963.

[Newman et al, 2008] Newman R., Lindsay R., Maphoto K., Mlwilo N., Mohanty A., Roux D., de Meijer R., Hlatshwayo I. Determination of soil, sand and ore primordial radionuclide concentrations by full-spectrum analyses of high-purity germanium detector spectra. Applied Radiation and Isotopes, 66, 2008, pp.855–859.

[Newman, 2009] Newman R. Computer Security: Protecting Digital Resources, ones & Bartlett Learning, 2009, p.49 and pp.58-59.

[Newsru, 2010] The volcanic eruption in Iceland paralyzed Europe. In: www.newsru.ru. 2010.

[Nguyen et al, 1997] Nguyen H.T., Kreinovich V., Bouchon-Meuiner B. Soft Computing Explains Heuristic Numerical Methods in Data Processing and in Logic Programming. AAAI Technical Report FS-97-04.

[Nichols and Tsokos, 1972] Nichols W.G., Tsokos C.P. Empirical Bayes Point Estimation in a Family of Probability Distributions, International Statistical Review, 40, 1972, pp.146-161.

[Niedermeier et al, 2000] Niedermeier A., Romaneessen E., Lenher S. Detection of coastline in SAR images using wavelet methods. IEEE Transactions Geoscience and Remote Sensing, 38(5), 2000, pp.2270-2281.

[Nielsen et al, 1983] Nielsen D., Tlotson P., Viera S. Analising field-measured soil water properties, Agricultural Water Management, 6,, 1983, pp.93-109.

[Nikonov, 2006] Nikonov A.A. Macroseisms… The past, The present, Forecast. Moscow, ComKniga, 2006, 192 p.

[Niyogi and Girosi, 1996] Niyogi P., Girosi F. On the relationship between generalization error, hypothesis complexity and sample complexity for Radial Basis Functions. Neural Computation, 8, 1996, pp.819-842.

[Nocquet and Calais, 2003] Nocquet J.-M., Calais E. Crustal velocity field of Western Europe from permanent GPS array solutions, 1996-2001, Geophys. J. Int., 154, 2003, pp.72-88.

[Norkin, 2006] Norkin V.I. On measurement and profiling of catastrophic risks. Cybernetics and Systems Analysis, Vol. 42, No.6, 2006, pp.839-850 (translated from Kibernetika i Sistemnyi Analiz, 2006, № 6, P.80–94).

[Norkin, 2007] Norkin V.I. Self-insurance of an investor under repeating catastrophic risks. Cybernetics and systems analysis, Vol. 43, No. 3, 2007 (Translated from Kibernetika i Sistemnyi Analiz, 2007, N 3, P. 74-83).

[NUREG/BR-0353, 2008] Davis Besse reactor pressure vessel head degradation. US NRC, NUREG/BR-0353, rev 1, August 2008.

[NUREG/CR-2300, 1983] U. S. Nuclear Regulatory Commission, PRA Procedures Guide, A Guide to the Performance of Probabilistic Risk Assessments for Nuclear Power Plants, Final Report, Vol. 1-2, NUREG/CR-2300, January 1983.

[NUREG/CR-6823, 2003] U. S. Nuclear Regulatory Commission, NUREG/CR-6823, Handbook of Parameter Estimation for Probabilistic Risk Assessment Sandia National Laboratories, September 2003.

[NUREG-1150, 1990] U. S. Nuclear Regulatory Commission, Severe Accident Risks: An Assessment for Five U.S. Nuclear Power Plants, NUREG-1150, December 1990.

[Nurminski and Zhelikhovski, 1977] Nurminski E.A., Zhelikhovski A.A. $\mathcal{E}$ - Quasigradient Method for Solving Nonsmooth External Problems, Cybernetics, Vol. 13, 1, 1977, 109-114.

[Ogryczak and Ruszczynski, 1999] Ogryczak W., Ruszczynski A. From Stochastic Dominance to Mean-Risk Models: Semideviation as Risk Measures. European Journal of Operation Research, 116, 1999, pp.33–50.

[Ogryczak and Ruszczynski, 2001] Ogryczak W., Ruszczynski A. On consistency of stochastic dominance and mean-semideviation models, Mathematical Programming 89, 2001, pp.217-282.

[Ooi et al, 1993] Ooi B.C., Sacks-Davis R., Han J. Indexing in Spatial Databases. Technical Report, 1993.

[Ortec, 2010] Gamma Spectroscopy. http://www.ortec-online.com/Solutions/gamma-spectroscopy.aspx

[Osborne, 2006] Osborne M. How to Cheat at Managing Information Security, Syngress Publishing, 2006, p.190.

[Österberg, 2004] Österberg E. Revealing of age-related deterioration of reinforced concrete containments in nuclear power plants – Requirements and NDT methods. The licentiate research thesis, The royal institute of technology, Stockholm, 2004.

[Ostroushko, 2004] Ostroushko A.P., Gusyatin V.M., Bugriy A.N. Approach to description of surface reflecting properties in the tasks of ray-tracing image synthesis. Radio-electronic and computer systems. Scientific and technical magazine, – № 3 (7), Kharkov: KHAI, 2004, pp.15-18.

[Outpost, 2010] Intrusion detection system FORTPOST 1.3. In: www.rnt.ru. 2010.

[Özel and Güvenir, 2001] Özel S., Güvenir H. An Algorithm for Mining Association Rules Using Perfect Hashing and Database Pruning. In Proceedings of the Tenth Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN'2001), 2001, pp.257-264.

[Papadimitriou et al, 2000] Papadimitriou C.H., Raghavan P., Tamaki H., Vempala S. Latent semantic indexing: A probabilistic analysis. J. Comput. System Sci., 61, 2000, pp.217-235.

[Paparoditis, 1988] Paparoditis N., Cord M., Jordan M., Cocquerez J.-P. Building Detection and Reconstruction from Mid- and High-Resolution Aerial Imagery, Computer vision and image understanding, vol 72, no2, Nov.1998, pp.122–142.

[Park et al, 1995] Park J., Chen M., Yu P. An effective hash based algorithm for mining association rules. In ACM SIGMOD International Conference on Management of Data, Volume 24, Issue 2, 1995, pp.175-186.

[Pearson, 1957] Pearson K. Tables of the Incomplete Gamma Function. Biometrika, University College, London, 1957.

[Pearson, 1968] Pearson K. Tables of the Incomplete Beta Function, University Press, Cambridge, 1968.

[Pei et al, 2001] Pei J., Han J., Lu H., Nishio S., Tang S., Yang D. Hmine: Hyper-structure mining of frequent patterns in large databases. In Proceedings of IEEE Int. Conf. Data Mining, 2001, pp.441-448.

[Penning-Rowsell and Chatterton, 1977] Penning-Rowsell E.C., Chatterton J.B. The benefits of flood alleviation: a manual of assessment techniques (The blue manual). Aldershot, UK: Gower Technical Press, 1977.

[Penning-Rowsell et al, 2003] Penning-Rowsell E., Johnson C., Tunstall S., Tapsell S., Morris J., Chatterton J., Coker A., Green C. The Benefits of flood and coastal defence: techniques and data for 2003. Flood Hazard Research Centre, Middlesex University, 2003.

[Perry et al, 1988] Perry S.G., Frasers A.B., Thomson D.W., Norman J.M. Indirect sensing of plant canopy structure with simple radiation measurements, Agricult. and Forest Meteorology, 42, 1988, pp.255-278.

[Pflug and Romisch, 2007] Pflug G.Ch., Romisch W. Modeling, Measuring and Managing Risk, NJ: World Scientific, 2007, 286 p.

[Piatetsky-Shapiro and Frawley, 1991] Piatetsky-Shapiro G., Frawley W. (eds) Knowledge Discovery in Databases. AAAI Press, Menlo Park, California, 1991.

[Plag et al, 1998] Plag H.-P., Ambrosius B., Baker T., Beutler G., Bianco G., Blewitt G., Boucher C., Davis J., Degnan J., Johansson J., Kahle H.-G., Kumkova I., Marson I., Mueller S., Pavlis E., Pearlman M., Richter B., Spakman W., Tatevian S., Tomasi P., Wilson P., Zerbini S. Scientific objectives of current and future WEGENER activities. Tectonophysics 294, 1998, pp.177-223.

[Plant and Murrell, 2007] Plant R., Murrell S. The Executive's Guide to Information Technology, Cambridge University Press 2007, p.51, p.256.

[Plato, 1981] Plato. The State. Science and Art (Nauka I izkustvo), Sofia, 1981 (in Bulgarian).

[Ponomarev, 2006] Ponomarev Y.V., Himko M.P., Datsuk A.V., Frolov V.A. Supervisory service engineer manual. Kiev-Kharkov, UCEBOPnaftogaz, 2006, 291 p.

[Popov, 1971] Popov A., Krissilov A. et al. On Automation of Medical Diagnostic Procedures. Proc. of IFIP-71 Congress, TA-7, Nederland, 1971, pp.841-847.

[Pospelov, 1986] Pospelov D. The situational control. Theory and practice. Nauka-Moscow, 1986, 278 pp. (in Russian)

[Prosvetov, 2005] Prosvetov G. Econometrics. SPb, RDL, 2005, 104 p.

[Qingyuan et al, 2005] Qingyuan Z., Xiangming X., Braswell B., et al. Estimating light absorption by chlorophyll, leaf and canopy in a deciduous broadleaf forest using MODIS data and a radiative transfer model. Remote Sensing of Environment, 99, 2005, pp.357-371.

[Quinlan and Cameron-Jones, 1993] Quinlan J., Cameron-Jones R. FOIL: A midterm report. In Proc. of European Conf. On Machine Learning, Vienna, Austria, 1993, pp.3-20.

[Quinlan, 1993] Quinlan J. C4.5: Programs for Machine Learning. M. Kaufmann, San Mateo, CA, 1993.

[Rabe, 2009] Rabe G. Computer Safety, Reliability, and Security. Proceedings of 28th International Conference, SAFECOMP 2009, Hamburg, Germany, Springer, 2009, pp. 91-94.

[Rachkovskij and Kussul, 2001] Rachkovskij D.A., Kussul E.M. Binding and normalization of binary sparse distributed representations by context-dependent thinning. Neural Computation, 13(2), 2001, pp.411-452.

[Rachkovskij, 2001] Rachkovskij D.A. Representation and processing of structures with binary sparse distributed codes. IEEE Transactions on Knowledge and Data Engineering, 13(2), 2001, pp.261-276.

[Rachkovskij, 2004] Rachkovskij D.A. Some approaches to analogical mapping with structure sensitive distributed representations. Journal of Experimental and Theoretical Artificial Intelligence, 16(3), 2004, pp.125-145.

[Radhamani and Rao, 2007] Radhamani G., Rao R., Web Services Security and E-business, Global. 2007, p.115, p.25

[Rak et al, 2005] Rak R., Stach W., Zaïane O., Antonie M.-L. Considering re-occurring features in associative classifiers. In Advances in Knowledge Discovery and Data Mining, volume 3518/2005 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2005, pp.240-248.

[Ramachandran et al, 2006] Ramachandran, R. Rushing, J. Li, X. Kamath, C. Conover, H. Graves, S. Bird's-eye view of data mining in geosciences. In: SPECIAL PAPERS- GEOLOGICAL SOCIETY OF AMERICA. NUMB 397, pages 235-248. Boulder, Colo.; Geological Society of America, USA. 2006. ISSN: 0072-1077

[Ramos et al, 2007] Ramos F., Dickson B., Kumar S. Denoising aerial gamma-ray surveying through non-linear dimensionality reduction. Journal of Field Robotics, 24(10), 2007, pp.849-861.

[Raudis, 1976] Raudis Sh.Yu. Limited Samples in Classification Problems, Statistical Problems of Control, Vilnus: Inst. Of Mathematics and Computer Science, vol. 18, 1976, pp.1-185.

[RAUFDRD, 2000] Risk Analysis and Uncertainty in Flood Damage Reduction Studies. Committee on Risk-Based Analysis for Flood Damage Reduction, G.B. Baecher, Chair, and J.W. Jacobs. – Washington, D.C.: The National Academies Press, 2000.

[Ravenbrook, 2010] Ravenbrook – software engineering consultancy, 2010, http://www.ravenbrook.com/

[RBS, 2009] Break-in without a crow. In: www.rinhed.ru. 2009.

[Rees, 2001] Rees W.G. Physical Principles of Remote Sensing, Cambridge University Press, 2001.

[Reilinger et al, 2006] Reilinger, R., McClusky S., Vernant P., Lawrence S., Ergintav S., Cakmak R., Ozener H., Kadirov F., Guliev I., Stepanyan R., Nadariya M., Hahubia G., Mahmoud S., Sakr K., ArRajehi A., Paradissis D., Al-Aydrus A., Prilepin M., Guseva T., Evren E., Dmitrotsa A., Filikov S., Gomez F., Al-Ghazzi R., Karam G. GPS constraints on continental deformation in the Africa-Arabia-Eurasia continental collision zone and implications for the dynamics of plate interactions, Journal of geophysical research, Vol. 111, B05411, 2006.

[Revunova and Rachkovskij, 2005] Revunova E.G., Rachkovskij D.A. Building a linear model of gamma-ray spectrum under noise. In Proc. Intern. Workshop on Inductive Modelling (IWIM-05), Kiev, Ukraine, 1, 2005, pp.250–254, (in Russian).

[Revunova and Rachkovskij, 2009s] Revunova E.G., Rachkovskij D.A. Randomized algorithms for solving discrete ill-posed problems. Submitted to IJITA in 2009.

[Revunova, 2005a] Revunova E.G. Comparison of model selection criteria in the approximation tasks with natural basis. Mathematical Machines and Systems, Issue 3, 2005, pp.116–125 (in Russian).

[Revunova, 2005b] Revunova E.G. Signal mixtures separation on the basis of minimal description length principle. Computer tools, networks and systems, Issue 4, 2005, pp.86-93 (in Russian).

[Revunova, 2005c] Revunova E.G. Two approaches to signal mixtures separation on the basis of linear modeling. System technologies, Issue 6(41), 2005, pp.124-148 (in Russian).

[Revunova, 2007a] Revunova E.G. Information technology of data analysis by linear modeling under uncertainty. Adaptive automatic control systems, Issue 11(31), 2007, pp.72-80 (in Russian).

[Revunova, 2007b] Revunova E.G. Blind source separation based on minimum description length. In Intern. Workshop on Inductive Modelling (IWIM-07), Prague, 1, 2007, pp.318–321.

[Revunova, 2008] Revunova E.G. Information technology and methods for analysis of signal mixtures based on the models linear with respect to parameters. PhD thesis, Kiev, Ukraine, 2008, 184 p, (in Russian).

[RH, 2009] The Russian hackers broke into American bank. In: http://www.finansmag.ru/news/59952. 2009. (in Russian).

[Rissanen, 1978] Rissanen J. Modeling by shortest data description. Automatica, 14, 1978, pp.465-471.

[Rissanen, 2002] Rissanen J. Lectures on Statistical Modeling Theory, http://www.cs.tut.fi/~rissanen/, 2002.

[Ristau and Moon, 2001] Ristau J.P., Moon W.M. Adaptive filtering of random noise in 2-D geophysical data. Geophysics, 66, 2001, pp.342–349.

[Robbins, 1951] Robbins H. Asymptotically Sub-Minimax Solutions to Compound Statistical Decision Problems. In: Proc. Second Berkeley Symp. Math Stat. Probab1, University of California Press, Berkeley, 1951.

[Robbins, 1955] Robbins H. An Empirical Bayes Approach to Statistics. Proc. 3rd Berkeley Symp. Prob. and Stat., 1, 1955, pp.157-163.

[Rockafellar and Uryasev, 2000] Rockafellar R.T., Uryasev S. Optimization of Conditional Value-at-Risk. Journal of Risk, 2, 2000, pp.21–42.

[Rockafellar and Uryasev, 2002] Rockafellar R.T., Uryasev S. Conditional Value-at-Risk for General Loss Distribution. Journal of Banking and Finance, 26, 2002, pp.1443–1471.

[Rokhlin and Tygert, 2008] Rokhlin V., Tygert M. A fast randomized algorithm for overdetermined linear least-squares regression PNAS September, 105(36), 2008, pp.13212-13217.

[Romanov et al, 2007] Romanov V., Fedak V., Galelyuka I., Sarakhan Ye., Skrypnyk O. Portable Fluorometer for Express-Diagnostics of Photosynthesis: Principles of Operation and Results of Experimental Researches. Proceeding of the 4th IEEE Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS'2007, Dortmund, 2007, pp. 570–573.

[Rose, 1973] Rose A. Vision: human and electronic, New York: Plenum Press, 1973, 197 p.

[Ruff, 2006] T. Ruff. Nuclear terrorism. http://energyscience.org.au/FS10 Nuclear Terrorism.pdf Tilman, 2006.

[Ruggeri and Sivaganesan, 2000] Ruggeri F., Sivaganesan S. On a Global Sensitivity Measure for Bayesian Inference. Sankhya: The Indian Journal of Statistics, 62, Series A, Pt. 1, 2000, pp.110-127.

[Ruszczynski and Shapiro, 2003] Ruszczynski A., Shapiro A. (Eds) Stochastic Programming, Amsterdam: Elsvier, 2003, 688 p.

[Ruvr, 2010] Volcanic ash cancels 500 flights in Europe. In: English.ruvr.ru. 2010.

[Saint-Petersburg, 2009a] Saint-Petersburg is in snow, the city is close to a collapse. In: www.oreand.ru. 2009.

[Saint-Petersburg, 2009b] Collapse in Petersburg: inhabitants move by means of shovels. In: rus.ruvr.ru. 2009

[Sampey, 1986] Sampey H. Operating Manual, ARAX International Corp., Vanderbilt, 1986.

[Sarknas, 2006] Sarknas P. Pro Asp.net 2.0 E-Commerce in C# 2005, Apress, 2006, p.34

[Sarlos, 2006] Sarlos T. Improved approximation algorithms for large matrices via random projections. In Proc. 47th Ann. IEEE Symp. Foundations of Computer Science (FOCS), 2006, pp.143-152.

[Savchuk and Martz, 1994] Savchuk V.P., Martz H.F. Bayes Reliability Estimation Using Multiple Sources of Prior Information: Binomial Sampling. IEEE Transactions on Reliability, 43, 1994, pp.138-144.

[Scheuren et al, 2008] Scheuren J.-M., le Polain de Waroux O., Below R., Guha-Sapir D., Ponserre S. Annual Disaster Statistical Review – The Number and Trends 2007. Center for Research of the Epidemiology of Disasters (CRED). Jacoffsaet Printers, Melin, Belgium, 2008.

[Schlaifer, 1979] Schlaifer R. Analysis of Decisions under Uncertainty. McGraw Hill, 1979.

[Schlesinger and Hlavach, 2002] M.I. Schlesinger, V. Hlavach. Ten Lectures on Statistical and Structural Pattern Recognition. Series: Computational Imaging and Vision, Vol. 24. Springer, 2002, 544 p., ISBN: 978-1-4020-0642-5.

[Seber, 1977] Seber G.A.F. Linear Regression Analysis. John Wiley and Sons, 1977.

[Sempau et al, 2003] Sempau J., Fernandez-Varea J.M., Acosta E., Salvat F. Experimental benchmarks of the Monte Carlo code PENELOPE. Nucl. Instr. Meth. in Phys. Res. B207, 2003, pp.107-23.

[Serebryakov, 2008] Serebryakov V. Regional Space Monitoring Center. SOVZOND company, 2008.

[Shannon, 1948] Shannon C. A mathematical theory of communication. The Bell system technical journal, N 27, 1948, pp.379-423, pp.623-656.

[Shaw, 2006] Shaw W., Cybersecurity for SCADA Systems, PennWell Corp., 2006, p. 194

[Shelestov et al, 2006] Shelestov A.Yu., Kussul N.N., Skakun S.V. Grid Technologies in Monitoring Systems Based on Satellite Data. J. of Automation and Information Science, 38(3), 2006, pp.69-80.

[Shelestov et al, 2008] Shelestov A., Kravchenko O., Ilin M.. Distributed visualization systems in remote sensing data processing GRID, International Journal Information Technologies and Knowledge, 2/1, 2008, pp.76-82.

[Shen et al, 2005] X. Shen, Y. Gao, C. Ding, R. Archambault. Lightweight Reference Affi nity Analysis. In: Proceedings of the 19th ACM International Conference on Supercomputing. ICS'05, June 20-22, Boston, MA, USA. ACM 1-59593-167-8/06/2005. pp. 131-140

[Shutko et al, 1995] Shutko A., Haldin A., Novichikhin E., Milshin A., Golovachev S., Grankov A., Mishanin V., Jackson T., Logan B., Tilley G., Ramsey III E., Pirchner H. Microwave radiometers and their application in field and aircraft campaigns for remote sensing of land and water surfaces. Proc. of IGARSS'95, 1995.

[Shutko et al, 1997] Shutko A., Haldin A., Novichikhin E., Yazerian G., Chukhray G., Vorobeichik E., Agura V., Kalashnik S., Sarkisjants V., Sklonnaja N., Logan B., Ramsey III E. Application of microwave radiometers for wetlands and estuaries monitoring. Proc. 4th Intern. Conf. on Remote Sensing for Marine and Coastal Environment, Vol. I, 1997.

[Shutko et al, 1998] Shutko A., Stepanov V., Krissilov A. et al. Dangerous hydrological situations monitoring in Ukrainian Black Sea region: problems, organization, effectiveness. Odessa, Inst. MPEER, "Consulting", 1998; pp.2-32 (in Russian).

[Shutko et al, 2007] Shutko A., Haldin A., Krapivin V., Novichikhin E., Sidorov I, Haarbrink R., Georgiev G., Pampaloni P., Krissilov A. Microwave radiometry in monitoring and emergency mapping of wares seepage and dangerously high groundwaters. Journal of Telecommunications and Information Technology. No. 1, 2007.

[Shutko, 1982] Shutko A. Microwave radiometry of lands under natural and artificial moistening. IEEE Transactions on Geosciences and Remote Sensing, vol. GE-20, 1982, pp.18-26.

[Shutko, 1987] Shutko A. Remote sensing of waters and land via microwave radiometry (The principles of method, problems feasible for solving, economic use). Proc. Study Week on Remote Sensing and Its Impact on Developing Countries, Pontifical Academy of Sciences, Vatican City, 1987, pp.413-441 (awarded by Pontifical Academy' Gold Medal).

[Singh, 1995] Singh V.P. Computer Models of Watershed Hydrology. Water Resources Publications, Littleton Co, 1995, p.1144.

[Sisalem et al, 2009] Sisalem D., Floroiu J., Kuthan J., Abend U., Schulzrinne H. SIP Security. John Wiley and Sons, 2009, pp.229-230.

[Slipchenko and Rachkovskij, 2009s] Slipchenko S.V., Rachkovskij D.A. Analogical mapping using similarity of binary distributed representations. Submitted to IJITA in 2009.

[SMC, 2009] Space Monitoring Center (SMC). – www.sovzond.ru/solutions/4844.html].

[Soifer, 1996] Soifer V. Computer processing of images. Part 2. Methods and algorithms. Soros Educational Journal №3. V. Soifer, 1996.

[Soifer, 2003] Soifer V., Gashnikov V., Glumov V., Popov S., Chernov V. Image processing methods, 2003, 784 p.

[Spain, 2010] Last year's vacation brought blocked bank cards to tourists. In: www.travel.ru. 2010.

[SRI NASU-NSAU, 2010] http://inform.ikd.kiev.ua/index.php?path=/en/index

[Stably, 1970] Stably D. Logical Programming with System/360. New York, 1970.

[Stangl and Bruyninx, 2006] Stangl G., Bruyninx C. Recent monitoring of crustal movements in the eastern Mediterranean. The use of GPS measurements. In: The Adria Microplate: GPS Geodesy, Tectonics and Hazards, Eds. Pinter, N., G. Grenerczy, J. Weber, S. Stein, D. Medak, Springer, Dordrecht, 2006, pp.169-181.

[Stangl et al, 2008a] Stangl G., Aichhorn C., Krauss S. Potential Networks and New Velocity Estimations for South-East Europe and the Orient, WEGENER Symposium Darmstadt, 2008.

[Stangl et al, 2008b] Stangl G., Aichhorn C., Krauss S. Comparing velocity estimations from permanent time series and CEGRN epoch campaigns. Reports on Geodesy No.1 (84), Eds. Sledzinski, J., Warsaw University of Technology, Warsaw, 2008, pp.23-30.

[Startseva,1995] Startseva N.G. Estimation of Convergence of the Expectation of the Classification Error Probability for Averaged Strategy, Proc. Ross. RAS, vol. 341, no. 5, 1995, pp.606-609.

[Starzev, 2009] Starzev S.. The earthquake in Italy left 29 000 people without homes. In: eco.rian.ru. 2009.

[Stepanov et al, 2001] Stepanov V., Krissilov A., Isakov M., Shutko A., Mishev D.. Project Proposals and Developing the Institutional Forms of Activity in a Frames of the International Black Sea Collaborations. Ecologia and Industry, v.3, #1-3, Sofia, 2001.

[Stepanov et al, 2003] Stepanov V., Krissilov A., Shutko A., Coleman T., Kostenyuk B. Consideration on using evaluation program methods, simulation and remotely sensed data for marine pollution control. Proc. of Int. Congress of World Oceanology Federation, devoted to 30 Anniv. of Intern. Oceanology Institute, Kiev, Ukraine, 2003.

[Stupina and Lbov, 2006] Stupina T.A., Lbov G.S. Application of the multivariate prediction method to time series. International Journal ITHEA, Vol 13, No 3, 2006, pp.278-285.

[Stupina, 2005] Stupina T.A. Estimation of quality removal for prediction multivariate heterogeneous variable problem. Proceeding of the Russian conference Mathematical methods of pattern recognition, Moscow, 2005, pp.209-212.

[Stupina, 2006] Stupina T.A. Recognition of the Heterogeneous Multivariate Variable. Proceeding of the international conference, (KDS'2006), Varna (Bulgaria), Vol 1, 2006, pp.199-202.

[Sugiyama and Ogawa, 2001] Sugiyama M., Ogawa H. Subspace information criterion for model selection. Neural Computation, 13(8), 2001, pp.1863-1889.

[Svanidze, 1977] Svanidze G.G. Mathematical modeling of hydrological series, Gidrometeoizdat, Leningrad, 1977.

[Tarantola, 2005] Tarantola A. Inverse problem theory problem and methods for model parameter estimation. SIAM, 2005.

[Temlyakov, 2003] Temlyakov V.. Nonlinear methods of approximation. Foundations of Comp. Math., 3, 2003, pp.33-107.

[Terzopoulos, 1987] Terzopoulos D. Snakes: Active Contour Models Int. Journal of Computer Vision. N1. D. Terzopoulos, M. Kaas, A. Witkin., 1987, p.331.

[Tikhonov and Arsenin, 1977] Tikhonov A.N., Arsenin V.Y. Solution of ill-posed problems. V.H. Winston, Washington, 1977.

[Tillman et al, 1982] Tillman F.A., Kuo W.H., Grosh D.L. Bayesian Reliability and Availability – a Review. IEEE Trans. Reliability, 31, No. 4, 1982, pp.362-372.

[Todini, 1995] Todini E. New trends in modeling soil processes from hill-slope to GCMS Scales – The role of water and the hydrological cycle in global change, edited by H. R. Oliver, S. A. Oliver, NATO ASI Series I: Global Environmental Change, 31, 1995, pp.317-347.

[Tonoyan, 1976] Tonoyan G.P. Chain decomposition of n dimensional unit cube and reconstruction of monotone Boolean functions, JVM&F, v. 19, No. 6, 1976, pp.1532-1542.

[Toubon et al, 2006] Toubon H., Boudergui K., Pin P., Nohl B., S. Lefevre, M. Chiron. New methodology for source location and activity determination in preparation of repairing or decommissioning activities. IRPA 2006, Paris, France, 2006.

[Tropp and Gilbert, 2007] Tropp J.A., Gilbert A.C. Signal recovery from random measurements via orthogonal matching pursuit. IEEE Transactions on Information Theory, 53(12), 2007, pp.4655-4666.

[Truhanov, 2009] Truhanov A. Russian hackers theft tens million dollars form Citibank. In: www.cnews.ru. 2009.

[Tsocos and Canavos, 1972] Tsocos C.P., Canavos G.C. Bayesian Concepts for the Estimation of Reliability in the Weibull Life-Testing model. International Statistical Institute Review, 40, 1972, pp.153-160.

[Two Crows Corp., 2005] Introduction to Data Mining and Knowledge Discovery, Third Edition. 2005. ISBN: 1-892095-02-5. Two Crows Corporation, Potomac, MD 20854 (U.S.A.).

[Tygert, 2009] Tygert M. A fast algorithm for computing minimal-norm solutions to underdetermined systems of linear equations. arXiv:0905.4745, May 2009.

[Uchitomi and Mine, 1988] Uchitomi S., Mine K. Intermittent diagnostics of the thermal process by means of attention subset diagnosis based on the Tree-Root Structure, Proceedings of the XI World Congress of IMEKO, Houston, 1988, pp.271-278.

[UCI MLR, 2011] UC Irvine Machine Learning Repository. http://archive.ics.uci.edu/ml/index.html Last visited 15.01.2011

[UNCOSA, 2003] United Nations Coordination of Outer Space Activities: Coordination of outer space activities within the United Nations system: programme of work for 2003 and 2004 and future years. United Nations Committee on the Peaceful Uses of Outer Space, A/AC.105/792, 2003.

[UN-SPIDER, 2010] http://www.un-spider.org

[Ursul, 1970] Ursul A. Information: a philosophical study. Berlin, Dietz, 1970.

[USACE, 1992] USACE (U.S. Army Corps of Engineers) 1992: Guidelines for risk and uncertainty analysis in water resources planning, Institute for Water Resources, IWR Report 92-R-1, Fort Belvoir, VA, 1992.

[USACE, 2000] USACE (U.S. Army Corps of Engineers) 2000: Risk Analysis and Uncertainty in Flood Damage Reduction Studies.

[Vagin, 1988] Vagin V. Deduction and generalization in the decision-making systems. Nauka-Moscow, 1988, 383 p. (in Russian)

[Vapnik and Chervonenkis, 1970] Vapnik V., Chervonenkis A. Theory of Pattern Recognition, Moscow: Nauka, 1970.

[Vassileva, 2004] Vassileva, K. Results of the CEGRN'03/BULREF'03 GPS Campaign, Reports on Geodesy, Warsaw University of Technology, No 4(71), 2004, pp.208-217.

[Vassileva, 2009a] Vassileva K. Processing and Analysis of GPS Data for Balkan Peninsula Permanent Stations. Proceedings of the International symposium on modern technologies, education and professional practice in geodesy and related fields, Sofia, 05-06.11.2009, pp.28-39.

[Vassileva, 2009b] Vassileva K. Velocity analysis of Balkan Peninsula Permanent Stations from GPS solutions. Proceedings of the International symposium on modern technologies, education and professional practice in geodesy and related fields, Sofia, 05-06.11.2009, pp.40-48.

[Vavilov, 1981] Vavilov S.I. Eye and sun, M.: Science, 1981, 125 p.

[Verhoef et al, 2007] Verhoef W., Xiao Q., Jia L., et al. Unified optical-thermal four-stream radiative transfer theory for homogeneous vegetation canopies. IEEE Transactions on Geoscience and Remote Sensing, V. 45, 2007, pp.1808–1822.

[Vidakovic, 2000] Vidakovic B. Gamma-Minimax: A Paradigm for Conservative Robust Bayesians. In: Robustness of Bayesian Inference. Editors Rios and Ruggeri, Springer-Verlag, Lecture Notes in Statistics 152, 2000, pp.241-259.

[Voloshin, 2005] Voloshin A. About decision-making problems in social-economic systems. In: XI-th International Conference Knowledge-Dialogue-Solution, Varna, V.1., 2005, pp.205–212.

[Voronov, 2009] Voronov V. Unmanned Aerial Vehicles at LAAD 2009 Exhibition, http://www.uav.ru/articles/LAAD-2009_report.pdf

[Voyshvillo, 1967] Voyshvillo E. The Concept. MGU-Moscow, 1967, 285 p. (in Russian)

[Wagner et al, 2007] Wagner W., Pathe C., Sabel D., Bartsch A., Kuenzer C., Scipal K. Experimental 1 km soil moisture products from ENVISAT ASAR for Southern Africa, ENVISAT and ERS Symposium, Montreux, Switzerland, 23-27.04.2007.

[Wahba, 1990] Wahba G. Spline models for observation data. CBMS-NSF Regional Conference Series in Applied Mathematics, 59, SIAM, Philadelphia, 1990, 169 pp.

[Wai-Wong et al, 2007] Wai-Wong K., Gedeon T., Fung C., JACIII Vol.11 No.3, 2007, pp.259-260.

[Waller et al, 1977] Waller R.A., Johnson M.M. Waterman M.S., Martz, H.F. Gamma Prior Distribution Selection for Bayesian Analysis of Failure Rate and Reliability. in: Nuclear Systems Reliability Engineering and Risk Assessment, J.B. Fussel, G.R. Burdick, Eds., SIAM, Philadelphia, 1977.

[Wang, 2009] Wang J. Computer network security: theory and practice. Springer, 2009, pp. 20-21.

[WASH-1400, 1975] U.S. Nuclear Regulatory Commission (USNRC), Reactor Safety Study-An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants, WASH-1400 (NUREG-75/014), Oct.1975.

[WASH-1400, 1979] U. S. Nuclear Regulatory Commission, NRC Statement on Risk Assessment and the Reactor Safety Study Report (WASH-1400) in Light of the Risk Assessment Review Group Report, January 18, 1979.

[Waterman et al, 1976] Waterman M.S., Martz H.F., Waller R.A. Fitting beta prior distributions in Bayesian reliability analysis. Los Alamos Scientific Laboratory report LA-6395-MS, Jul., 1976.

[Weiler, 1965] Weiler H. The use of incomplete beta functions for prior distributions in binomial sampling. Technometrics, 7, 1965, pp.335-347.

[Wiener, 1948] Wiener N. Cybernetics: control and communication in the animal and the machine. New York, 1948.

[Wikdahl, 2006] Wikdahl C.E. Forsmark incident on the 25th of July 2006. Analysis group at KSU, N 4, Sweden, 2006.

[Witten and Frank, 2005] Witten I., E. Frank. Data Mining: Practical Machine Learning Tools and Techniques. 2nd Edition, Morgan Kaufmann, San Francisco, ISBN 0-12-088407-0, 2005. http://www.cs.waikato.ac.nz/ml/weka/, visited on 15.01.2011.

[Witten and MacDonald, 1988] Witten, I. H. & MacDonald, B. A. (1988). Using concept learning for knowledge acquisition. International Journal of Man-Machine Studies, 27, (pp. 349-370).

[Wrembel and Koncilia, 2007] Wrembel R., Koncilia C. Data Warehouses and OLAP: Concepts, Architectures, and Solutions, Idea Group Inc (IGI), 2007, p.16.

[Yan and Han, 2002] Yan X., Han J. gSpan: Graph-based structure pattern mining. In Proceedings of the 2nd IEEE International Conference on Data Mining (ICDM'02), 2002, pp.721-724.

[Yaremchuk, 2007] S.Yaremchuk. IDS guards the network perimeter. In: Hacker, №98. 2007.

[Yin and Han, 2003] Yin X., Han J. CPAR: Classification based on predictive association rules. In Proc. of the SIAM Int. Conf. on Data Mining., San Francisco, CA: SIAM Press, 2003, pp.369-376.

[Yuan and Huang, 2005] Yuan Y., Huang T. A Matrix Algorithm for Mining Association Rules. Lecture Notes in Computer Science, Volume 3644, (Sep 2005), 2005, pp.370-379.

[Zabulonov and Revunova, 2006] Zabulonov Yu.L., Revunova E.G. Hardware systolic method of distribution density function reconstruction for surface contaminations by inverse problem solution. Modelling and information technologies, Issue 36, 2006, pp.127-131 (in Russian).

[Zabulonov et al, 2004a] Zabulonov Yu.L., Lisichenko G.V., Revunova E.G.. Determining the extent of radioactive contamination of objects in non-fixed geometry. Modelling and information technologies, 27, 2004, pp.95-101 (in Russian).

[Zabulonov et al, 2004b] Zabulonov Yu.L., Lisichenko G.V., Revunova E.G. Express-analysis of radionuclide composition by approximation approach to the spectrum reconstruction. Geochemistry and ecology. Issue 10, 2004, pp.32-38 (in Russian).

[Zabulonov et al, 2005] Zabulonov Yu.L., Lisichenko G.V., Revunova E.G. Increase of radionuclides identification reliability at the analysis of the low intensity radiation fields. In Proceedings of Sevastopol Institute of Nuclear Energy and Industry, Issue 16, 2005, pp.114-124 (in Russian).

[Zabulonov et al, 2006] Zabulonov Yu.L., Korostil Yu.M., Revunova E.G. Optimization of inverse problem solution to obtain the distribution density function for surface contaminations. Modelling and information technologies, 39, 2006, pp.77-83 (in Russian).

[Zabulonov et al, 2009a] Zabulonov Yu.L., Lisichenko G.V., Revunova E.G., Odukalets L.A.. Increasing the accuracy of radionuclide identification by matching pursuit. Modelling and information technologies, 53, 108-114, 2009 (in Russian).

[Zabulonov et al, 2009b] Zabulonov Yu.L., Lisichenko G.V., Korostil Yu.M., Revunova E.G. Identifying sources of radiation using independent component analysis. Modelling and information technologies, 52, 73-79, 2009 (in Russian).

[Zagoruiko, 1976] Zagoruiko N. Applied Methods of Data and Knowledge Analysis. Novosibirsk, 1976 (in Russian).

[Zaïane and Antonie, 2002] Zaïane O., Antonie M.-L. Classifying text documents by associating terms with text categories. In Proc. of the Thirteenth Australasian Database Conference (ADC'02), 2002, pp.215-222.

[Zaïane and Antonie, 2005] Zaïane O., Antonie, M.-L. On pruning and tuning rules for associative classifiers, In Proc. of Int. Conf. on Knowledge-Based Intelligence Information & Engineering Systems (KES'05), 2005, pp.966-973.

[Zaki et al., 1997] Zaki M., Parthasarathy S., Ogihara M., Li W. New Algorithms for Fast Discovery of Assoctiation Rules. In 3rd International Conference on Knowledge Discovery and Data Mining, 1997, pp.283-286.

[Zarco-Tejada et al, 2003] Zarco-Tejada et al. Water content estimation in vegetation with MODIS reflectance data and model inversion methods. Remote Sensing of Environment, 85, 2003, pp.109-124.

[Zerby, 1963] Zerby C.D. A Monte Carlo calculation of the response of gamma-ray scintillation counters. Methods in Computational Physics, eds. B. Alder, S. Fernbach and M. Rotenberg, Academic Press, New York, 1, 1963, pp.89-134.

[Zhao and Nevatia, 2003] Zhao T., Nevatia R. Car detection in low resolution aerial images. Image and Vision Computing 21, 2003, pp. 693–703.

[Zielosko and Wakulicz-Deja, 2005] Zielosko B., Wakulicz-Deja A. Intelligent Data Processing in Distributed Internet Applications. In: Advances in Soft Computing, Volume 31, 2005, pp.585-591.

[Zimmermann and De Raedt, 2004] Zimmermann A., De Raedt L. CorClass: Correlated association rule mining for classification. In Discovery Science, volume 3245/2004 of Lecture Notes in Computer Science, Springer, 2004, pp.60-72.

Authors' Information

ASLANYAN Levon

Professor, Dr., Institute for Informatics and Automation Problems, NAS Armenia

P.Sevak St. 1, Yerevan-14, Armenia

Main research areas: classification, forecasting, data mining.

e-mail: lasl@ipia.sci.am

BECKER Matthias

Professor, Dr., Darmstadt University of Technology

Petersenstr. 30, 64287 Darmstadt, Germany

e-mail: becker@ipg.tu-darmstadt.de

BILOUS Nataliya

Ph.D., head of ITLCVS laboratory, professor of Kharkov National University of Radio Electronics, Lenin Prospect 14, Kharkiv 61166, Ukraine

Main research areas: image processing, segmentation and recognition, building of computer vision system, computer testing and learning systems.

e-mail: Bilous@kture.kharkov.ua

BOJILOV Venko

Application engineering, Business modeling: General Insurance (Bulstrad PLC BG); INSPIRE transposition: Consultancy (ASDE BG)

Main research areas: complex process analyzes; complex functional models; risk management & mitigation; Earthquake risk management.

BONDARENKO Michael

PhD, professor, corresponding member of NAS of Ukraine, rector of Kharkov National University of Radio Electronics, research leader of ITLCVS laboratory, Lenin Prospect 14, Kharkiv 61166, Ukraine

Main research areas: artificial intelligence.

BORISENKO Victor

PhD, Head of department, Research & Design Institute for Gas Transportation, Marshala Koneva Street 16, Kharkiv 61004, Ukraine;

Main research areas: modern information technologies, corporate information system, enterprise database, complex system simulation, programming methodology.

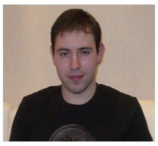
e-mail: vborisenko@itransgaz.com

BUGRIY Andriy

PhD, Senior Lecture, Kharkiv National University of Radio and Electronics (KHNURE), Lenin Prospect 14, Kharkiv 61166, Ukraine;

Main research areas: models, methods and algorithms of 3D images synthesis, structures of the digital visualization systems.

e-mail: abugriy@kture.kharkov.ua



BURAK Pavel

Petersburg State Transport University, department of Informatics and Information safety; 190031, Russia, Saint-Petersburg, Moskovskij prospect, 9

e-mail: wistle3b@gmail.com

DEPAIRE Benoit

Universiteit Hasselt; Campus Diepenbeek; Department of Applied Economic Sciences; Research Group Data Analysis & Modelling

Wetenschapspark 5; bus 6; BE-3590 Diepenbeek; Belgium

Main research areas: data mining and knowledge discovery

e-mail: benoit.depaire@uhasselt.be

DIMITROVA Rumyana

Master student, University of National and World Economy – Sofia, Bulgaria;

e-mail: dimitrova_rumyana@yahoo.com

GALELYUKA Igor

Senior research fellow, PhD, V.M. Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine, Kiev, Ukraine

e-mail: galib@gala.net

GLADUN Victor

Professor, V.M. Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine, Kiev, Ukraine

GOLODNIKOV Alexandr

Senior research fellow, PhD, V.M. Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine, Kiev, Ukraine

e-mail: al_gol2006@ukr.net

HAARBRINK Roland

Director of the Dutch company "MIRAMAP" for radiometric mapping (Techn. Inc. of European Space Ag.), Netherlands

Main research areas: Earth observation, multispectral remote sensing of the environment, microwave radiometry, MCW mapping.

E-mail: rhaarbrink@yahoo.com

HRUSHA Volodymyr

Research fellow, V.M. Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine, Kiev, Ukraine

e-mail: vhrusha@gmail.com

ILINA Elena

The Swedish radiation safety authority, Department of nuclear safety, System assessments

Solna Strandväg 122, 17 154 Solna, Sweden

e-mail: elena.ilina@ssm.se

IVANOVA Krassimira

Researcher, Institute of Mathematics and Informatics, BAS

Acad. G.Bontchev St., bl.8, Sofia-1113, Bulgaria

Main research areas: image processing, data mining, knowledge retrieval.

e-mail: kivanova@math.bas.bg

IVANOVA Krassimira B.

Assistant professor, University of National and World Economy – Sofia, Bulgaria;
e-mail: krasy78@mail.bg

IVANOVA Natalia

Ph.D.(Tech), Associated Professor of Petersburg State Transport University, department of Informatics and Information Safety

190031, Russia, Saint-Petersburg, Moskovskij prospect, 9

e-mail: natali_iv@rambler.ru

KIRILYUK Vladimir

Leading researcher, PhD, ScD, V.M. Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine, Kiev, Ukraine

e-mail: vkirilyuk@i.ua

KNOPOV Pavel

PhD, ScD, Prof., V.M. Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine, Kiev, Ukraine

e-mail: Knopov1@yahoo.com

KOBZAR Gleb

Researcher of ITLCVS laboratory, Kharkov National University of Radio Electronics, Lenin Prospect 14, Kharkiv 61166, Ukraine;

Main research areas: image processing and shape based recognition, computer graphics, and computer vision systems.

e-mail: spwar.3dm@gmail.com

KOLEV Nikola

Doctor of Sciences, Professor, Institute of Soil Science "Nikola Poushkarov", Sofia, Bulgaria

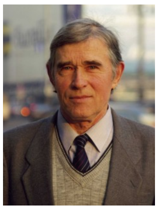
e-mail: nvkolev@mail.bg

KOROBULINA Olga

Post Graduate Student of Petersburg State Transport University, department of Informatics and Information safety

190031, Russia, Saint-Petersburg, Moskovskij prospect, 9

e-mail: Olga_korobulina@list.ru



KRAPIVIN Vladimir

Prof., Doct. of Sci., Head of Department, Institute of Radio engineering and Electronics, Russian Academy of Sciences, Acad. Vvedenskij Sq.1, 141190 Moscow region, Fryazino, Russia

Main research areas: development and application of a geo-information monitoring system technology for remote sensing.

e-mail: vfk@ms.ire.rssi.ru

KRAVCHENKO Oleksii

PhD Student, Space Research Institute NASU-NSAU, Glushkov Prospekt 40, build. 4/1, Kyiv 03680, Ukraine.

Main research areas: inverse modeling, pattern recognition, remote-sensing data processing, parallel algorithms, GRID computing.

KRISSILOV Anatolij

PhD, Assoc. Prof., Senior Researcher, Institute for Information Technologies of Odessa State Academy of Refrigeration; Corresponding member and adviser of the Ukrainian Ecological Academy of Sciences, Odessa, Ukraine.

Main research areas: decision-making, artificial intelligence, pattern recognition, geoinformation system, regional and urban socio-economic development.

e-mail: adkrissilov@list.ru

KRISSILOV Victor

Prof., Doct. of Sci., Chair of System Programming Department, Odessa State Polytechnic University, Odessa, Ukraine

Main research areas: artificial intelligence, information processing formalization, quality and technology assessment, neuron networks.

E-mail: krissilovva@mail.ru

KUSSUL Nataliia

Deputy Director, Space Research Institute NASU-NSAU, Glushkov Prospekt 40, build. 4/1, Kyiv 03680, Ukraine

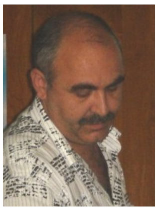
Main research areas: grid computing, design of distributed software systems, parallel computations, intelligent data processing methods, Sensor Web, neural networks, satellite data processing, risk management and space weather.

e-mail: inform@ikd.kiev.ua

KUZEMIN Alexandr

Doctor of Sciences, Prof. of Information Department, Kharkov National University of Radio Electronics, Head of IMD, (Ukraine)

e-mail: kuzy@kture.kharkov.ua



LYASHENKO Vyacheslav

Senior research Fellow of problem-research laboratory "Transfer-Informational Technologies in the Management of Emergent situations", Research Department, Kharkov National University of Radio Electronics, Kharkov, Ukraine

MAKIVSKY Dmitry

Engineer of ITLCVS laboratory. Kharkov National University of Radio Electronics, Lenin Prospect 14, Kharkiv 61166, Ukraine

Main research areas: image processing, segmentation and recognition, building of computer vision system.

e-mail: mak.spectrum@gmail.com

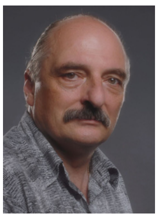
MARKOV Krassimir

Assoc. Prof., PhD, Institute of Mathematics and Informatics, BAS,

Acad. G.Bontchev St., bl.8, Sofia-1113, Bulgaria

Main research areas: data bases, multimedia data bases, information systems, data mining, knowledge retrieval, theoretical information research.

e-mail: markov@foibg.com



MILENOV Kristian

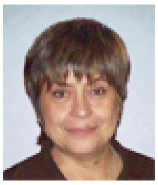
Architect, Assoc. Prof., Executive director of Agency for Sustainable Development and Eurointegration-ECOREGIONS, Sofia; member of GEO HLWG.

Main research areas: sustainable development, sustainable urban planning, energy efficiency & environmental architecture, RES, eco-auditing, GIS, remote sensing applications, risk and security management

MILENOV Pavel

National Expert at the Joint Research Centre of the European Commission in Ispra, Italy; Member of the staff of the Agency for Sustainable Development and Eurointegration; one of the founders of Remote Sensing Application Center.

Main research areas: GIS and remote sensing, land cover/land use mapping, image processing, land management and urban planning, infrastructure monitoring and control of the EU agriculture subsidies using satellite imagery.



MILENOVA Lyudmila

Executive director of Remote Sensing Application Center – ReSAC; member of Bulgarian Aerospace Agency.

Main research areas: remote sensing and GIS in decision making for agricultural and environmental management, land cover/land use, soil and forest inventory, water resources, environmental hazards, urban planning, infrastructure.

MILEV Georgi

Professor, Dr., Corresponding Member of Bulgarian Academy of Sciences

Institute of Water Problems of BAS

Bulgaria, 1113 Sofia, Acad. G. Bonchev Str., Bl. 1

e-mail: milev@bas.bg

MILEV Ivo

Dr., Technet – rail 20-10 GmbH

Berlin, Germany

e-mail: Ivo.milev@technet-rail.de

MITOV Iliya

PhD Student of the Institute of Mathematics and Informatics, BAS and Hasselt University, Belgium

Main research areas: data mining, knowledge retrieval, data bases, information systems, ERP-systems

e-mail: mitov@foibg.com

NENOV Martin

PhD,Institute of Soil Science "Nikola Poushkarov",

Sofia, Bulgaria

e-mail: nenov.martin@gmail.com

NICKOLOV Eugene

Prof., DSc, PhD, Eng, National Laboratory of Computer Virology, Bulgarian Academy of Sciences, Regular lecturer at New Bulgarian University

Main research areas: Computing Systems, Computer Virology, Information Security

e-mail: eugene@nlcv.bas.bg

NORKIN Vladimir

V.M. Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine, Leading research fellow, Doctor of Sciences, Kiev, Ukraine

e-mail: norkin@i.com.ua

NOVICHIKHIN Eugenij

Leading Researcher, Institute of Radio engineering and Electronics, Russian Academy of Sci., Acad. Vvedenskij Sq.1, 141190 Moscow region, Fryazino, Russia

Main research areas: development and implementation of radio-physical methods for soil moisture assessment, microwave radiometry and mapping, Earth observation.

E-mail: epnov@mail.ru

OSTROUSHKO Andrii

PhD, Associate Professor, Kharkiv National University of Radio and Electronics (KHNURE), Lenin Prospect 14, Kharkiv 61166, Ukraine;

Main research areas: computer graphics, ray tracing, digital visualization systems.

e-mail: osa@kture.kharkov.ua

PALAGIN Aleksandr

Doctor of technical sciences, Professor,

Academician of National Academy of Sciences of Ukraine,

Depute-director of V.M. Glushkov's Institute of Cybernetics of National Academy of Sciences of Ukraine, Prospect Akademika Glushkova 40, Kiev–187, 03680, Ukraine

e-mail: palagin_a@ukr.net

PEPELYAEV Vladimir

PhD, ScD., V.M. Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine, Kiev, Ukraine

e-mail: Emk160@gmail.com

POLIMIROVA Dimitrina

PhD, Research Associate, National Laboratory of Computer Virology, Bulgarian Academy of Sciences, Part-time lecturer at New Bulgarian University

Main research areas: Computer virology, Information Security of File objects, Methods of Compression

e-mail: polimira@nlcv.bas.bg

RACHKOVSKIJ Dmitri

PhD, DSc, Leading Research Scientist, International Research and Training Center for Information Technologies and Systems, NAS and MES of Ukraine; Pr. Acad. Glushkova, 40, Kiev, 03680, Ukraine; e-mail: dar (at) infrm.kiev.ua; http://dnipt.irtc.org.ua/

Main research areas: Computational and Artificial Intelligence

RADKOV Radko

Dipl.Eng. Project manager in IT development and RS/GIS technology programs in STALKER Ltd, ASDE and ReSAC.

Main research areas: map analyses, fast track services for disaster management, creation of reliable database, control, and revision by space imagery, land cover mapping, and development of GIS based system for monitoring linked with control devices for dynamic data provision, satellite imagery orthorectification, DEM creation.

REVUNOVA Elena

PhD, Research Scientist, International Research and Training Center for Information Technologies and Systems, NAS and MES of Ukraine; Pr. Acad. Glushkova, 40, Kiev, 03680, Ukraine;

Main research areas: Signal Processing

e-mail: egrevunova (at) gmail.com; http://dnipt.irtc.org.ua/

ROMANOV Volodymyr

Doctor of Sciences, Professor, Head of Department in V.M. Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine

e-mail: VRomanov@i.ua, dept230@insyg.kiev.ua

SAHAKYAN Hasmik

PhD, Leading Researcher, Institute for Informatics and Automation Problems, NAS Armenia

P.Sevak St. 1, Yerevan-14, Armenia

Main research areas: data mining, discrete mathematic.

e-mail: hasmik@ipia.sci.am

SARAKHAN Yevgeniya

Research fellow, PhD, V.M. Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine

e-mail: sarakhan2006@ukr.net

SHELESTOV Andrii

Senior Scientist, Space Research Institute NASU-NSAU, Glushkov Prospekt 40, build. 4/1, Kyiv 03680, Ukraine;

Main research areas: Grid computing, intelligent methods of estimation and modeling, satellite data processing, Sensor Web, neural networks, multi-agent simulation and control, software engineering, distributed information systems design.

e-mail: inform@ikd.kiev.ua

SHUTKO Anatolij

Prof., Doct. of Sci., Head of Laboratory, Institute of Radio engineering and Electronics, Russian Acad. of Sci.; leading researcher in radio-constructing concern "Vega", Moscow, Russia. Corr. member and adviser of the Ukrainian Ecological Acad. of Sci.

Main research areas: Earth observation, multispectral remote sensing of the environment, development and implementation of radio-physical methods for soil moisture assessment, microwave radiometry, development of a geo-information monitoring system and advanced technologies for remote sensing.

E-mail: anshutko@hotmail.com

SIDOROV Igor

Leading researcher in Radio-Constructing Concern "Vega", Moscow, Russia.

Main research areas: microwave radiometry, new MCW technique elaboration, multispectral remote sensing.

E-mail: igor__sidorov@mail.ru

SKAKUN Sergii

Senior Scientist, Space Research Institute NASU-NSAU, Glushkov Prospekt 40, build. 4/1, Kyiv 03680, Ukraine

Main research areas: remote sensing data processing (optical and radar), image processing, grid computing, sensor web, neural networks, analysis of computer system's users behavior.

e-mail: serhiy.skakun@ikd.kiev.ua

STANGL Guenter

Dr., Space Research Institute, Austrian Academy of Sciences

Schmiedlstrasse 6, A-8042 Graz, Austria

e-mail: guenter.stangl@oeaw.ac.at

STUPINA Tatyana

Senior research fellow, PhD, Trofimuk Institute of Petroleum-Gas Geology and Geophysics of SBRAS, Novosibirsk, Russia

e-mail: stupinata@ipgg.nsc.ru

VANHOOF Koen

Professor, Universiteit Hasselt; Campus Diepenbeek; Department of Applied Economic Sciences; Research Group Data Analysis & Modelling.

Wetenschapspark 5; bus 6; BE-3590 Diepenbeek; Belgium

Main research areas: data mining, knowledge retrieval.

e-mail: koen.vanhoof@uhasselt.be

VASSILEVA Keranka

Assoc. Prof., Dr., Central Laboratory of Geodesy, Bulgarian Academy of Sciences

Bulgaria, 1113 Sofia, Acad. G. Bonchev Str., Bl. 1

E-mail: ker@bas.bg

VELYCHKO Vitalii

Senior research fellow, PhD, V.M.Glushkov Institute of Cybernetics of NAS of Ukraine

Prosp. Akad. Glushkov, 40, Kiev-03680, Ukraine

Main research areas: data mining, knowledge retrieval, NLP processing.

e-mail: velychko@aduis.com.ua

ITHEA® International Scientific Society

ITHEA International Scientific Society (**ITHEA ISS**) is aimed to support growing collaboration between scientists from all over the world.

ITHEA ISS a successor of the international scientific co-operation organized within 1986-1992 by international workgroups (**IWG**) researching the problems of databases and artificial intelligence. As a result of tight relation between these problems in 1990 in Budapest appeared the international scientific group of Data Base Intellectualization (**IWGDBI**) integrating the possibilities of databases with the creative process support tools. The leaders of the IWGDBI were Prof. Victor Gladun (Ukraine) and Prof. Rumyana Kirkova (Bulgaria). Starting from 1992 until now the international scientific co-operation has been organized by the Association of Developers and Users of Intellectualized Systems (**ADUIS**), Ukraine. It has played a significant role for uniting the scientific community working in the area of the artificial intelligence.

To extend the possibilities for international scientific collaboration in all directions of informatics by wide range of concrete activities, in 2002 year, the Institute for Information Theories and Applications FOI ITHEA (**IITA FOI ITHEA**) has been established as an international nongovernmental organization. IITA FOI ITHEA is aimed to support international scientific research through international scientific projects, workshops, conferences, journals, book series, etc. The achieved results are remarkable. IITA FOI ITHEA became worldwide known scientific organization. One of the main activities of the IITA FOI ITHEA is building the ITHEA International Scientific Society aimed to unite researches from all over the world who are working in the area of informatics.

Till the beginning of 2011 year, the **ITHEA International Scientific Society** was joined by more than **2700** members from **48** countries all over the world: *Armenia, Azerbaijan, Belarus, Brazil, Belgium, Bulgaria, Canada, China, Czech Republic, Denmark, Egypt, Estonia, Finland, France, Germany, Greece, Hungary, India, Iran, Ireland, Israel, Italy, Japan, Jordan, Kyrgyz Republic, Latvia, Lithuania, Malaysia, Malta, Mexico, Moldova, Netherlands, Peru, Poland, Portugal, Romania, Russia, Scotland, Senegal, Serbia, Montenegro, Sweden, Spain, Sultanate of Oman, Turkey, UK, Ukraine, and USA.*

ITHEA Publishing House (**ITHEA PH**) is the official publisher of the works of the ITHEA ISS. The scope of the books of the ITHEA ISS covers the area of Informatics and Computer Science. ITHEA PH welcomes scientific papers and books connected with any information theory or its application.

ITHEA ISS has two International Journals, established as independent scientific printed and electronic media, published by ITHEA Publishing House:

- **International Journal "Information Theories and Applications"** (**IJ ITA**), since 1993
- **International Journal "Information Technologies and Knowledge"** (IJ ITK) since 2007

Every year, ITHEA ISS organizes many scientific events called **"ITHEA ISS Joint International Events of Informatics" (ITA**). For the first 10 years of the XXI century, ITHEA ISS has organized **64** International Conferences and workshops and has published **44** scientific books.

ITHEA ISS has organized its scientific events in collaboration with:

- Association of Developers and Users of Intelligent Systems (Ukraine)
- V.M.Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine
- Dorodnicyn Computing Centre of the Russian Academy of Sciences (Russia)
- Institute of Linguistics, Russian Academy of Sciences (Russia)
- Institute of Mathematics and Informatics, BAS (Bulgaria)
- Institute of Mathematics of SD RAN (Russia)
- Alexandru Ioan Cuza University (Romania)
- Astrakhan State University (Russia)
- BenGurion University (Israel)
- Kazan State University (Russia)
- Kharkiv National University of Radio Electronics (Ukraine)
- Moscow State Linguistic University (Russia)
- New Bulgarian University (Bulgaria)
- Rzeszow University of Technology (Poland)
- Taras Shevchenko National University of Kiev (Ukraine)
- The University of Zadar (Croatia)
- Universidad Politecnica de Madrid (Spain)
- University of Calgary (Canada)
- University of Hasselt (Belgium)

as well as many other scientific organizations. For more information: www.ithea.org.

All journals and books of ITHEA ISS are edited by the Institute of Information Theories and Applications FOI ITHEA, in collaboration with the leading researchers from the: *Institute of Cybernetics "V.M.Glushkov", NASU* (Ukraine)*; Institute of Mathematics and Informatics, BAS* (Bulgaria)*; University of Hasselt* (Belgium)*; Natural Computing Group (NCG) of the Technical University of Madrid* (Spain)*; Astrakhan State Technical University* (Russia)*; Taras Shevchenko National University of Kiev* (Ukraine)*; University of Calgary* (Canada)*; VLSI Systems Centre, Ben-Gurion University* (Israel).

The great success of ITHEA International Journals, Books, and Conferences belongs to the whole of the ITHEA International Scientific Society.

More information may be obtained from www.ithea.org.

February 2011

ADUIS

Association of Developers and Users of Intelligent Systems

ADUIS consists of about one hundred members including ten collective members. The Association was founded in Ukraine in 1992. The main aim of ***ADUIS*** is to contribute to the development and application of the artificial intelligence methods and techniques. The efforts of scientists engaged in ***ADUIS*** are concentrated on the following problems: expert system design; knowledge engineering; knowledge discovery; planning and decision making systems; cognitive models designing; human-computer interaction; natural language processing; methodological and philosophical foundations of Artificial Intelligence.

Association has long-term experience in collaboration with teams, working in different fields of ***research and development***. Methods and programs created in Association were used for revealing regularities, which characterize chemical compounds and materials with desired properties. Some thousands of high precise prognoses have been done in collaboration with chemists and material scientists of Russia and USA.

Association can help ***businesspersons*** to find out conditions for successful investment taking into account region or field peculiarities as well as to reveal user's requirements on technical characteristics of products being sold or manufactured.

Physicians can be equipped with systems, which help in diagnosing or choosing treatment methods, in forming multi-parametric models that characterize health state of population in different regions or social groups.

Sociologists, politicians, managers can obtain the Association's help in creating generalized multi-parametric "portraits" of social groups, regions, enterprise groups. Such "portraits" can be used for prognostication of voting results, progress trends, and different consequences of decision making as well.

Association provides a useful guide in technical diagnostics, ecology, geology, and genetics.

ADUIS has at hand a broad range of high-efficiency original methods and program tools for solving analytical problems, such as knowledge discovery, classification, diagnostics, and prognostication.

ADUIS unites the creative potential of highly skilled scientists and engineers

Since 1992, ***ADUIS*** holds regular conferences and workshops with wide participation of specialists in AI and users of intelligent systems. The proceedings of the conferences and workshops are published in scientific journals. ***ADUIS*** cooperates through its foreign members with organizations that work on AI problems in Russia, Belarus, Moldova, Georgia, Bulgaria, Czech Republic, Germany, Great Britain, Hungary, Poland, etc. ***ADUIS*** is the collective member of the European Coordinating Committee for Artificial Intelligence (ECCAI).

www.aduis.com.ua

For contacts: V.M.Glushkov Institute of Cybernetics; National Academy of Science of Ukraine;

Prospect Akademika Glushkova, 40, 03680 GSP Kiev-187, Ukraine;

Phone: (+38 044) 5266429; Fax: (+38 044) 5263348; E-mail: info@aduis.com.ua