Krassimir Markov, Vitalii Velychko,
Lius Fernando de Mingo Lopez, Juan Casellanos
(editors)

New Trends
in
Information Technologies

I T H E A
SOFIA
2010

Krassimir Markov, Vitalii Velychko, Lius Fernando de Mingo Lopez, Juan Casellanos (ed.)
New Trends in Information Technologies
ITHEA®
Sofia, Bulgaria, 2010
ISBN 978-954-16-0044-9

First edition

Recommended for publication by The Scientific Concil of the Institute of Information Theories and Applications FOI ITHEA

This book maintains articles on actual problems of research and application of information technologies, especially the new approaches, models, algorithms and methods of membrane computing and transition P systems; decision support systems; discrete mathematics; problems of the interdisciplinary knowledge domain including informatics, computer science, control theory, and IT applications; information security; disaster risk assessment, based on heterogeneous information (from satellites and in-situ data, and modelling data); timely and reliable detection, estimation, and forecast of risk factors and, on this basis, on timely elimination of the causes of abnormal situations before failures and other undesirable consequences occur; models of mind, cognizers; computer virtual reality; virtual laboratories for computer-aided design; open social info-educational platforms; multimedia digital libraries and digital collections representing the European cultural and historical heritage; recognition of the similarities in architectures and power profiles of different types of arrays, adaptation of methods developed for one on others and component sharing when several arrays are embedded in the same system and mutually operated.

It is represented that book articles will be interesting for experts in the field of information technologies as well as for practical users.

General Sponsor: Consortium FOI Bulgaria (www.foibg.com).

Printed in Bulgaria

Copyright © 2010 All rights reserved

© 2010 ITHEA® – Publisher; Sofia, 1000, P.O.B. 775, Bulgaria. www.ithea.org ; e-mail: info@foibg.com

© 2010 Krassimir Markov, Vitalii Velychko, Lius Fernando de Mingo Lopez, Juan Casellanos – Editors

© 2010 Ina Markova – Technical editor

© 2010 For all authors in the book.

® ITHEA is a registered trade mark of FOI-COMMERCE Co.

ISBN 978-954-16-0044-9

C\o Jusautor, Sofia, 2010

FLOOD RISK ASSESSMENT BASED ON GEOSPATIAL DATA

Nataliia Kussul, Sergii Skakun, Andrii Shelestov, Yarema Zyelyk

***Abstract**: The problem statement of disaster risk assessment, based on heterogeneous information (from satellites and in-situ data, and modelling data) is proposed, the problem solving method is grounded and considers its practical use for risk assessment of flooding in Namibia. The basis of the method is the ensemble approach to the heterogeneous data analysis with the use of the data fusion techniques and evaluation the probability density function of a natural disaster using this method.*

***Keywords**: risk assessment, natural disasters, geospatial data, remote sensing, data fusion, ensemble data processing, probability density function, parametric statistics, maximum likelihood classifier, neural network classifier*

***ACM Classification Keywords**: I.5 PATTERN RECOGNITION - I.5.1 Models – Neural nets; G.1 NUMERICAL ANALYSIS - G.1.8 Partial Differential Equations - Inverse problems; F. Theory of Computation - F.1.1 Models of Computation - Probabilistic computation; G.4 MATHEMATICAL SOFTWARE - Parallel and vector implementations; H. Information Systems - H.3 INFORMATION STORAGE AND RETRIEVAL - H.3.5 Online Information Services; I.4 IMAGE PROCESSING AND COMPUTER VISION - I.4.6 Segmentation - Pixel classification; I.4.8 Scene Analysis - Sensor fusion; J. Computer Applications - J.2 PHYSICAL SCIENCES AND ENGINEERING - Earth and atmospheric sciences*

Introduction

Changes in climate are caused numerous natural disasters: floods, droughts, heavy snowfall, forest fires, etc., bringing great damage to the economy of individual countries and entire regions. In recent years, to monitor natural disasters are increasingly using the geospatial data of different nature: the satellite images and products (such as digital relief model, land use maps), as well as two-dimensional or three-dimensional modelling data (particularly meteorological or hydrological models). The monitoring result of such information use has become digital maps or multi-layered geospatial data, greatly facilitating the decision-making process relevant authorities. Such information can be used not only for the mapping of disaster areas during or after the event itself, but also at other stages of the disaster cycle - including for the construction of risk maps that illustrate the probability of occurrence and the damage that can be they caused.

To problem of operational services creating for natural disaster risk assessment in Europe the project SAFER (http://www.emergencyresponse.eu/site/FO/scripts/myFO_accueil.php?lang=EN) of GMES program (Global Monitoring for Environment and Security) is devoted. The French Space Agency CNES is actively developing an approach to risk assessment of infectious diseases caused by the spread of insect vectors caused by floods in Africa (http://www.redgems.org/spip.php?rubrique4). However the risk assessment techniques used today in operating systems are often too simplistic and are not based on a fairly well developed mathematical apparatus for the average risk assessment that was developed for the problem of the quality estimating of the functional dependencies recovery based on empirical data and used in the statistical learning theory [Vapnik, 1995; Vapnik, 1998; Haykin, 1999; Bishop, 2006].

The problem statement of disaster risk assessment, based on heterogeneous information (from satellites, in-situ data, and modelling data) is proposed in this paper, the problem solving method is grounded and considers its practical use for risk assessment of flooding in Namibia.

Existing approaches to assessment of the natural disaster risk on the basis of geospatial information

In a variety of subject areas (economics, public health, and financing activities) the general concept of risk is determined by roughly the same. "Risk is a combination of the likelihood of an occurrence of a hazardous event or exposure(s) and the severity of injury or ill health that can be caused by the event or exposure(s)" (OHSAS 18001:2007 — Occupational Health and Safety Management Systems Requirements Standard). In general mathematically risk *R* often simply defined as a function *f* of disaster probability *p* and expected loss *l* (http://www.wired.com/science/planetearth/magazine/17-01/ff_dutch_delta?currentPage=3):

$$R = f\left(p, l\right). \tag{1}$$

In the statistical decision theory risk function to estimate $\delta(x)$ of parameter θ (using the classifier or decision rule), calculated on the basis of observation *x* of parameter θ, is defined as the expected value of the loss function *L* [Christian, 2007]

$$R(\theta, \delta(x)) = \int L(\theta, \delta(x)) f(x / \theta) dx . \tag{2}$$

In [Jonkman et al, 2003] contains a detailed review of metrics for determining the risk of an individual, as well as social, economic and other risks associated with natural disasters. However, in general, the risk is described as a function of the probability of damage. For example, a simple measure of social risk is the expected number of victims per year, calculated by the formula

$$E(N) = \int_0^{\infty} x f_N(x) dx , \tag{3}$$

where $f_N(x)$ is probability density function of the number of victims per year.

Another example of the risk function [Piers, 1998] is a function of aggregated weighted risk (AWR), defined by the relation

$$AWR = \iint_A IR(x,y) h(x,y) dxdy , \tag{4}$$

where $IR(x, y)$ is the risk of a disaster (so-called individual risk) in the position with coordinates (x, y), $h(x, y)$ is the number of houses on location (x, y), and *A* is area, for which the AWR is determined.

The following sections of paper will be formalized concept of disaster risk on the basis of heterogeneous geospatial information, and will state the risk assessment method and will identify the data sources.

Problem statement of disaster risk assessment on the basis of heterogeneous geospatial information and method of its solution

The aggregated expected risk of disaster consequences (the aggregated expected losses) in the area *A* will by called the value

$$R_A = \iint_A r(x,y) dxdy , \tag{5}$$

where $r(x,y)$ is the individual expected risk of disaster consequences z (individual expected losses) at the point (x,y) calculated as the mathematical expectation of damage consequences function $h_{xy}(z)$ in the location (x,y)

$$r(x,y)=\int_0^\infty h_{xy}(z)p_{xy}(z)\,dz\,, \tag{6}$$

where $p_{xy}(z)$ is probability density function of the disaster z at the point (x,y) being evaluated on the basis of joint analysis of heterogeneous geospatial data. One method of estimating of the probability density function $p_{xy}(z)$ and the damage consequences function is determined by the type of disaster and will be described below.

The probability density function $p_{xy}(z)$ of the disaster is determined by various environmental factors and weather conditions that may be directly or indirectly measured by in-situ and remote-sensing methods, or obtained through modelling.

For example, the likelihood of spring flooding is determined by snow watersupplies, snowmelt intensity, air temperature and rainfall in snowmelt period in the given area and upstream, as well as by soil structure, its degree of freezing and by other factors. Expected runoff volume and water level in the river can be estimated using hydrologic models with assimilation into model satellite data and in-situ measurements data.

To recover the probability density function can be used well-developed at the statistical learning theory [Vapnik, 1995; Vapnik, 1998; Haykin, 1999; Bishop, 2006] method of the empirical functional minimizing in the problem of the average risk minimizing. Constructing an empirical functional is based on an approximation of an unknown probability density that is included in the average risk functional type of (6) (in our case), some of the empirical density restored on the basis of measurement data and use it instead of the unknown density of the empirical functional. Next, we need not solve the minimizing problem of this empirical functional of average risk, as it is done in the classical theory of dependencies recovery from empirical data. In our case, interest is only the problem of the probability density $p_{xy}(z)$ reconstructing from sample data, which is classical problem of mathematical statistics. Having restored the probability density, we can estimate the individual risk according to (6) as the expected value of the disaster damage, we can estimate the aggregated expected risk of natural disaster consequences (the aggregated expected losses) in the area *A* in accordance with (5) and we can use the information about the risk to decide on measures, which reduce the damage of the disaster consequences.

The problem of reconstructing the probability density in the class of continuous functions is reduced to an ill-posed problem of numerical differentiation of probability distribution function [Vapnik, 1995; Vapnik, 1998]. It can be solved using the non-parametric methods (such as Parzen's method, the method of ordered risk minimization using the covariance matrix of correlated measurement errors), which take into account the ill-posed problem and rely on the statistical theory of regularization [Vapnik, 1995; Vapnik, 1998]. However, in cases where there is a priori information about the unknown probability density, we can avoid the ill-posed formulation of this problem. For example, if the restored probability density is known up to a finite number of parameters, the problem of its recovery from empirical data is correct, and for its solutions can be used effective methods of parametric statistics [Vapnik, 1995; Vapnik, 1998]. So in a class of average risk minimizing problems associated with the classification problem (pattern recognition learning), the recovery of the unknown parameters in the probability density $p_{xy}(z)$ may by performed, in particular in [Vapnik, 1995; Vapnik, 1998], using various methods of parametric statistics (depending from the problem context): Bayesian approximations method; the best unbiased approximations method, maximum likelihood method. In addition, as models for evaluating of the probability density function of a natural disaster can be used different regression models or model in the form of a black box (such as neural network, kernel methods, other methods of machine learning theory, etc.) [Vapnik, 1995; Vapnik, 1998; Haykin, 1999; Bishop, 2006].

To estimate disaster risk probability we should analyze (classify) information from different sources with different time and space resolution. For joint analysis of such information methods of data fusion are used [Mitchell, 2007]. Density of disaster probability is estimated via fusion not row data but mainly information of higher levels of data processing [Das, 2008]. We propose following general scheme for estimation of disaster probability density $p_{xy}(z)$ (Fig. 1).

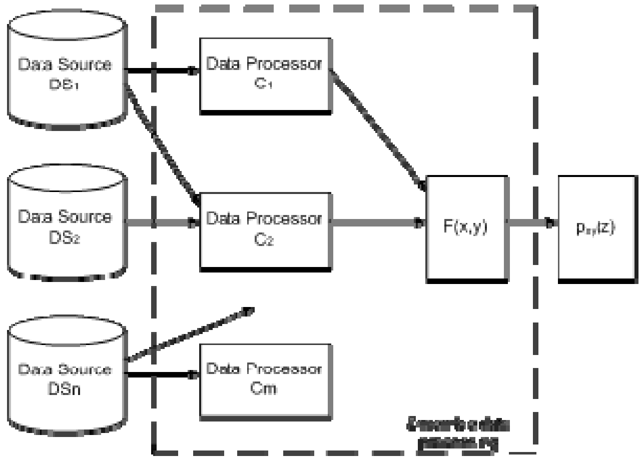

Figure 1. Estimation the density of disaster probability $p_{xy}(z)$ on the base of heterogeneous information

Blocks DS_i, i=1,…, n in Fig. 1 represent different data sources — satellite data, in-situ observations and modelling data. Blocks C_i, i=1,…, m provide data processing and higher level information acquisition. In general case $m \neq n$, because data from one source could be processed by several classifier and vice versa. So data fusion is provided already by classifiers. Concrete type of each classifier C_i, i=1,…, m is determined by input data specifics. For example, satellite data processing includes several kinds of preprocessing (reprojection, geocoding, atmosphere and geometric correction and so on) before „thematic processing“. So each processor C_i, i=1,…, m could provide several levels of data transformation, but for clearness we will not explicitly specify the levels.

Generally speaking, each processor C_i, i=1,…, m is a special decision rule or classifier (so called weak or component classifier) analyzing data from one or several sources. Thus classifiers C_i, i=1,…, m form an ensemble of experts (or „strong“ classifier) and their decisions are integrated within single decision F with correspondent weights α_i

$$F(x,y) = \sum_{i=1}^{m} \alpha_i C_i(x,y) . \tag{7}$$

Ensemble of classifiers is shown in Fig. 2. Such complex method of data processing provides more accurate estimation of heterogeneous information then each of “weak” classifiers [Jaakkola, 2006]. Note that the process of combining simple “weak” classifiers into one “strong” classifier is analogous to the use of kernels to go from a simple linear classifier to a non-linear classifier. The difference is that here we are learning a small number of highly non-linear features from the inputs rather than using a fixed process of generating a large number of features from the inputs as in the polynomial kernel. To improve the accuracy of classification we can use boosting technique [Kotsiantis and Pintelas, 2004], reduced to the estimation of a loss function and minimizing loss by adding new weak classifiers. Such an approach allows us to optimize the number of classifiers and complexity of the model.

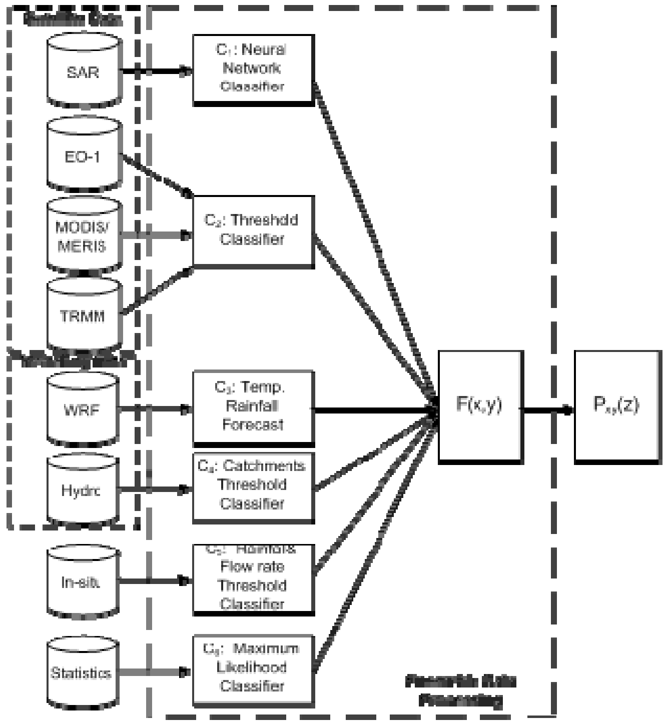

Figure 2. Estimation of $p_{xy}(z)$ for Namibia

Case-study: flood risk assessment based on heterogeneous data

As a case study we consider the Namibia Sensor Web Pilot Project - A case study on Integrated Flood modeling, forecasting, mapping and Water-related disease management. The purpose of this project is to integrate remote sensing into a flood and water-related disease modeling, monitoring, and early warning and decision support system. This international project was initiated by Ministry of Agriculture, Water and Forestry (MAWF) and Ministry of Health and Social Services (MHSS) of Namibia; United Nations Platform for Space-based Information for Disaster and Emergency Response (UN-SPIDER); Ukraine Space Research Institute (USRI); NASA/GSFC; NOAA/National Environmental Satellite Data and Information Service (NESDIS); German Aerospace Center (DLR); and Committee on Earth Observing Satellites (CEOS) Working Group on Information Systems and Services (WGISS). The overall project framework is shown in Fig. 3.

The following data sets are used for flood risk assessment within a joint project of UN-SPIDER, NASA, DLR, NOAA and Space Research Institute NASU-NSAU:

- Satellite imagery:
  - synthetic-aperture radar: Envisat/ASAR
  - optical: EO-1, MODIS (Terra and Aqua)
  - TRMM
- Modelling data:
  - meteorological data (numerical weather prediction)
  - hydrological data (river catchments)

- In-situ observations and river gauges:
  - o rainfall and river flowrateS
- Statistical data:
  - o Statistical information on floods for previous years.

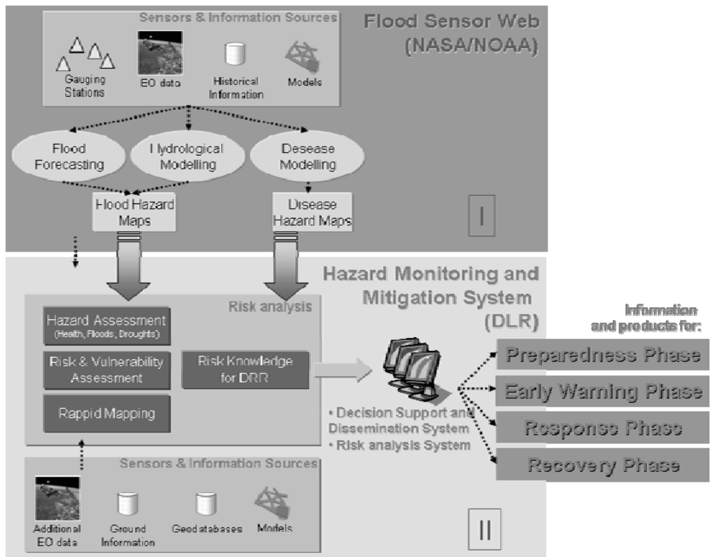

Figure 3. Overview of proposed framework

Flood mapping from satellite imagery. We use both microwave and optical satellite data to estimate the flood extent. We use intelligent computation techniques to derive flood mask from satellite imagery [Skakun, 2009; Kussul et al., 2008a; Kussul et al., 2008b]:

- Envisat/ASAR (within ESA Category-1 project): medium spatial resolution (150 m): products are delivered within 24h after image acquisition; high spatial resolution (30 m): products are delivered on demand.
- RADARSAT-2 (within the International Charter "Space and Major Disasters" and Disaster Working Group of GEO): high spatial resolution (3 to 30 m).

For cloud-free days we acquire data from optical sensors:

- Envisat/MERIS: medium spatial resolution (300 m);
- Terra and Aqua/MODIS: medium spatial resolution (250 m – 1 km);
- NASA EO-1: high spatial resolution (30 m).

Products are delivered in KML (for Google Earth), GeoTiff, WMS and others. An example is depicted in Fig. 4.

We use data from joint mission of NASA and JAXA Tropical Rainfall Measuring Mission (TRMM) to monitor rainfall rate (Fig. 5).

We have also setup a Website that shows all the flood products that were derived from Envisat/ASAR imagery (Fig. 6).

Meteorological data. We run Weather Research and Forecast (WRF) numerical prediction model to obtain forecast of meteorological parameters.

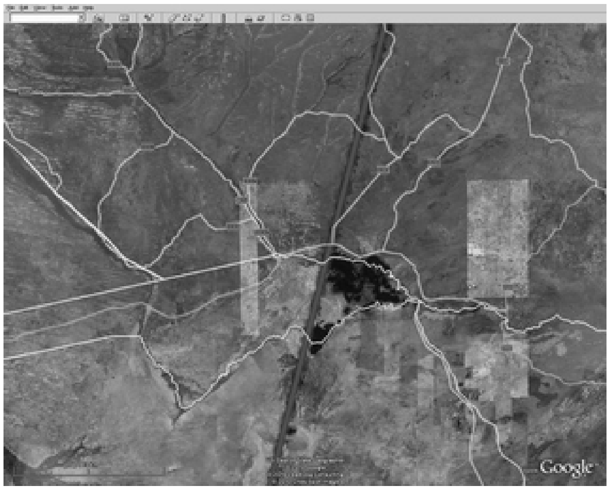

Figure 4. Flood mask for Katima-Mulilo region in Namibia derived from Envisat/ASAR, 03.03.2010

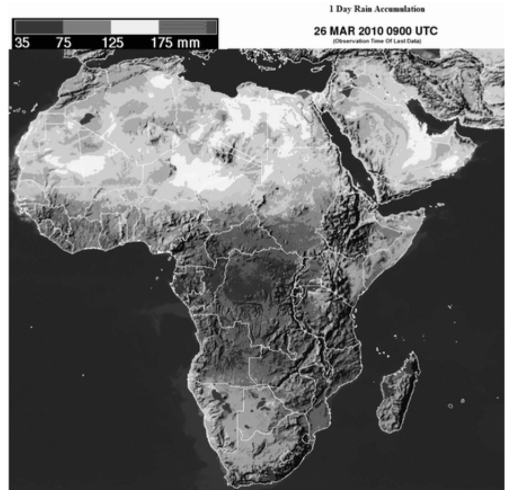

Figure 5. TRMM observations

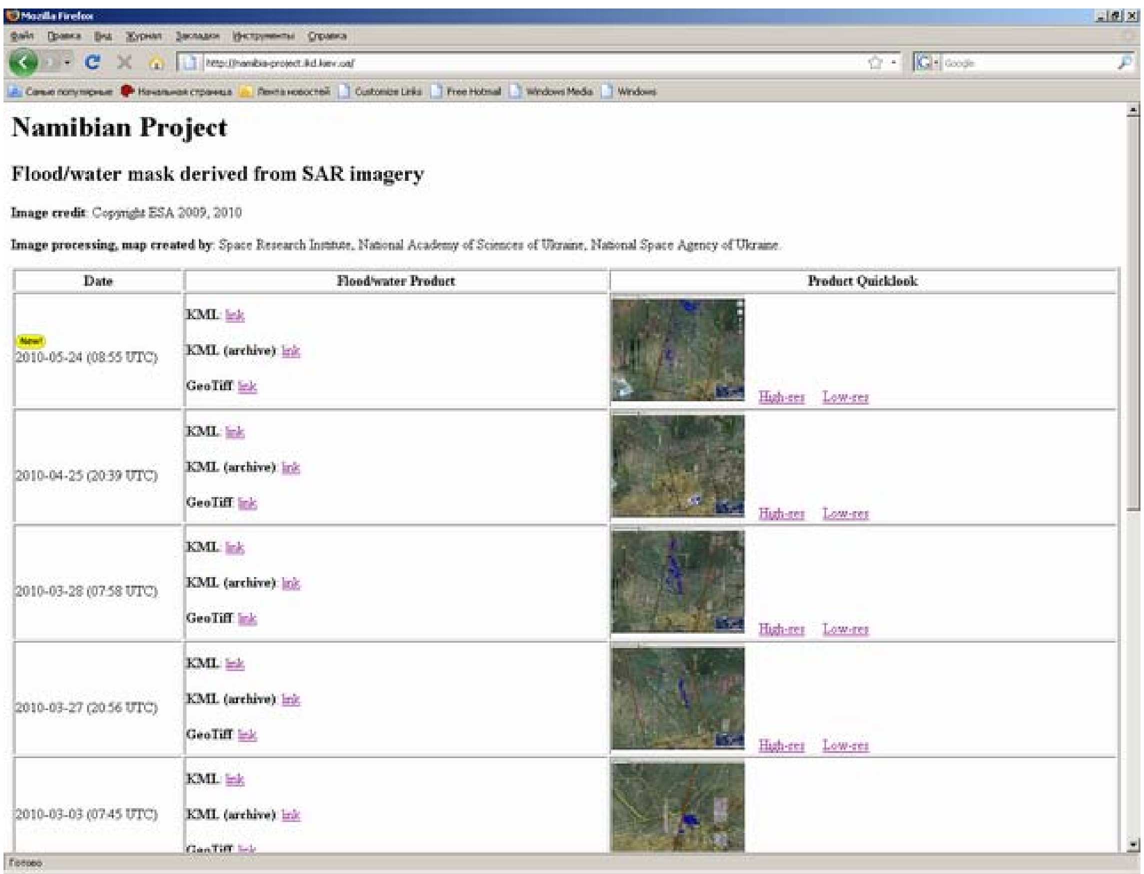

Figure 6. The list of flood products derived from Envisat/ASAR data for Namibia.

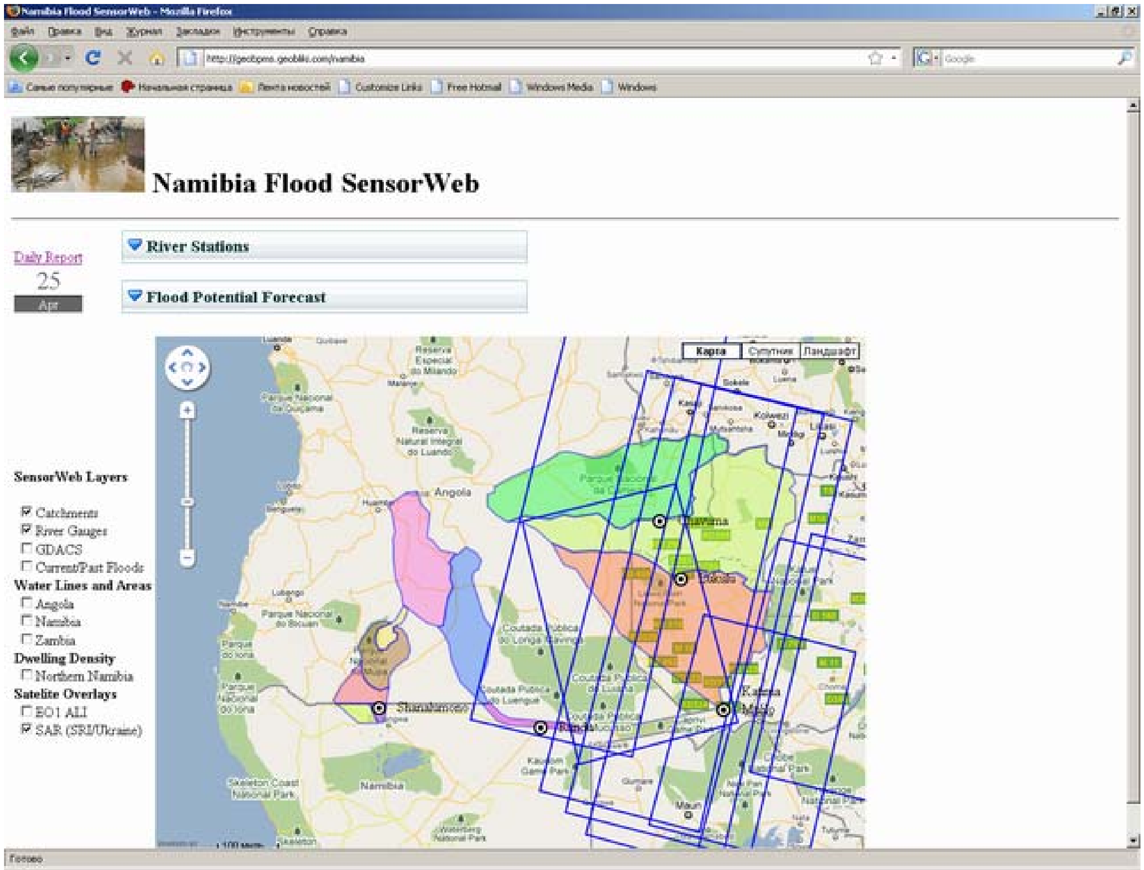

Figure 7. Web portal of the Namibia Sensor Web Pilot Project

River catchments. Data on river catchments are provided by the Ministry of Agriculture, Water and Forestry of Namibia. For each are, both archive and current data on rainfall and river flowrate are provided.

Statistical data. Statistical data of previous floods are derived from MODIS flood products that are provided by the Dartmouth Flood Observatory. These data are available from 1999.

For global flood detection we use data provided by the Joint Research Centre (JRC) of the European Commission [Groeve and Riva, 2009].

Web portal of the Namibia Sensor Web Pilot Project was established that integrates all the data and products derived by the participants (Fig. 7). These data are integrated using the ensemble approach proposed in this paper to provide flood risk assessment and are delivered to the end-users.

Conclusions

We proposed a unique approach to flood risk assessment using heterogeneous geospatial data acquired from multiple sources. This approach is based on statistical learning theory and incorporates an ensemble of classifiers for estimating probability density of the emergency. The advantage of the proposed approach is higher accuracy of risk assessment while using optimal model complexity. This approach is used within the Namibia Sensor Web Pilot Project for flood assessment using geospatial data of different nature. We plan to extend our approach to estimate the concrete risk category such as financial, social, economic, etc.

Acknowledgments

This work is partly supported by STCU project #4928 and Project № M/72-2008 of National Ministry of Education “Development of system for complex remote sensing data processing using Grid-technologies. The investigations are carried out within International Pilot Sensor Web Project for Flood Monitoring in Namibia.

Bibliography

[Vapnik, 1995] Vapnik, V. The Nature of Statistical Learning Theory. - New York: Springer Verlag, 1995

[Vapnik, 1998] Vapnik, V. - Statistical Learning Theory. New York: Wiley, 1998.

[Haykin, 1999] Haykin, S. Neural Networks. A comprehensive Foundation. — New Jersey: Prentice Hall, 1999.

[Bishop, 2006] Bishop C.M. Pattern Recognition and Machine Learning. - New York: Springer Science+Business Media, 2006. – 738 p.

[Christian, 2007] Robert, Christian (2007). The Bayesian Choice (2nd ed.). New York: Springer. doi:10.1007/0-387-71599-1. MR1835885. ISBN 0-387-95231-4.

[Jonkman et al, 2003] S.N. Jonkman et al. An overview of quantitative risk measures for loss of life and economic damage // Journal of Hazardous Materials A99 (2003). P. 1–30.

[Piers, 1998] M. Piers, Methods and models for the assessment of third party risk due to aircraft accidents in the vicinity of airports and their implications for societal risk// In: R/E/ Jorissen, P.J.M.Stallen (Eds.), Quantified Societal Risk and Policy Making, Kluwer Academic Publishers, Dordrecht, 1998.

[Mitchell, 2007] H. B. Mitchell, Multi-sensor Data Fusion – An Introduction (2007) Springer-Verlag, Berlin, ISBN 9783540714637.

[Das, 2008] S. Das, High-Level Data Fusion (2008), Artech House Publishers, Norwood, MA, ISBN 9781596932814 and 1596932813.

[Jaakkola, 2006] Tommi Jaakkola, course materials for 6.867 Machine Learning, Fall 2006. MIT OpenCourseWare(http://ocw.mit.edu/), Massachusetts Institute of Technology.

[Kotsiantis and Pintelas, 2004] Kotsiantis S., Pintelas P. Combining Bagging and Boosting, International Journal of Computational Intelligence, Vol. 1, No. 4 (324-333), 2004.

[Skakun, 2009] Skakun S. A neural network approach to flood mapping from satellite imagery // Scientific Papers of Donetsk National Technical University “Informatics, Cybernetics and Computer Science”. - 2009. - Vol. 10(153). - P. 92-100. (in Ukrainian)

[Kussul et al., 2008a] Kussul N., Shelestov A., Skakun S. Intelligent Computations for Flood Monitoring// International Book Series "Advanced Research in Artificial Intelligence" (ed. Markov K., Ivanova K., Mitov I.), 2008, number 2, pp.48-54.

[Kussul et al., 2008b] Kussul N., Shelestov A., Skakun S. Grid System for Flood Extent Extraction from Satellite Images // Earth Science Informatics. - 2008. - 1(3-4). - P. 105-117.

[Groeve and Riva, 2009] De Groeve, T., P. Riva, 2009. Global real-time detection of major floods using passive microwave remote sensing. Proceedings of the 33rd International Symposium on Remote Sensing of Environment, Stresa, Italy, May 2009.

Authors' Information

Nataliia Kussul *– Prof., Deputy Director, Space Research Institute NASU-NSAU, Glushkov Prospekt 40, build. 4/1, Kyiv 03680, Ukraine;*
e-mail: inform@ikd.kiev.ua

Major Fields of Scientific Research: 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.

Andrii Shelestov *– Prof., Senior Scientist, Institute of Cybernetics NASU, Glushkov Prospekt 40, build. 4/1, Kyiv 03680, Ukraine;*
e-mail: inform@ikd.kiev.ua

Major Fields of Scientific Research: 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.

Sergii Skakun *– Senior Scientist, Space Research Institute NASU-NSAU, Glushkov Prospekt 40, build. 4/1, Kyiv 03680, Ukraine;*
e-mail: serhiy.skakun@ikd.kiev.ua

Major Fields of Scientific Research: Remote sensing data processing (optical and radar), image processing, Grid computing, Sensor Web, neural networks, analysis of computer system's users behavior.

Yarema Zyelyk *– Prof., Leading Scientist, Space Research Institute NASU-NSAU, Glushkov Prospekt 40, build. 4/1, Kyiv 03680, Ukraine;*
e-mail: adapt@ikd.kiev.ua

Major Fields of Scientific Research: estimation and control under uncertainty, dynamics of structural formations and magnetic fields on the Sun, the analysis and prediction of the temporal series and fields, information systems and technologies