On the Methodology of Satellite Data Utilization in Multi-Modeling Approach for Socio-Ecological Risks Assessment Tasks: A Problem Formulation

Yuriy V. Kostyuchenko

Scientific Centre for Aerospace Research of the Earth
National Academy of Sciences of Ukraine
55-b, O. Honchar Street, Kiev, 01601, Ukraine
E-mail: yuriy.v.kostyuchenko@gmail.com

Abstract
The problem of satellite data utilization in multi-modeling approach for socio-ecological risks assessment is formally defined. Methodology of natural geo-systems modeling for variables and indicators selections is described. Observation, measurement and modeling data utilization method in the framework of multi-model approach is described. Methodology and models of risk assessment in framework of decision support approach are defined and described. Brief conclusions on efficiency of the described methodology are proposed. Proposed methodology can applied for wide range of risk-related tasks, such as natural and technological disaster monitoring, air-water-soil pollution control, crop productivity control, etc.

Keywords- Remote sensing, Indicators, Geo-system model, Variables, Decision making, Risks assessment.

1. Introduction
Since its origin, remote sensing demonstrates a stormy evolution (Elachi and Van Zyl, 2006; Lillesand et al., 2014). The common methodology of remote sensing is also being developed, and its basic idea is grounded on models of signals formation, i.e., models of individual indicators and methods for their monitoring (Elachi and Van Zyl, 2006). Conclusions on the development of the individual processes in studied natural systems usually are built on such monitoring (Elachi and Van Zyl, 2006; Qiu et al., 2007).

Modern understanding of ecosystems is more comprehensive (Suter, 2016). We understand the complex interactions between the processes and phenomena, can simulate the feedbacks in multi-agent environment, model the integrated dynamics of the processes, and predict the behavior of multi-component systems. Remote sensing now can and should become a source of information about behavior of the variables in these complex, interlinked models. Remote sensing should be tool not only for monitoring, but also for predictions and forecasts.

However, it will require a change of methodology of remote sensing applications, new methods of processing and interpretation of data (Campbell and Wynne, 2011). In particular, we need a new foundation to select the sets of interrelated indicators, based on the models.

Using of such indicators will also require other approaches to risk assessment. These approaches should not be based on assessments of deviations of observed values from the mean, and will be focused on making optimal decisions in a complex multi-component and multi-physics environment (Ermoliev et al., 2012).

This paper is aimed to the formal task definition of the utilization of satellite data in the multi-model approach for socio-ecological risks assessment tasks.

2. Data Utilization: Methodology of Modeling, Selection of Variables
First of all, we need consider the problem of the models application to selection of the optimal set of remote sensing indicators in risk assessment tasks (Kostyuchenko et al., 2015).

At the initial stage the process of forecast generation should be based on the set of initial assumptions, captured in a vector x ($x = (x^1, x^2,...,x^s)$). These could be a priori assumptions, observed or measured values.

The further step is the modeling: model recalculation these values into a group of core hydrological, bio-physical, and climatological series (with $F(x, \varepsilon, y) = 0$, $(x, \varepsilon) \to y$), collected in a vector y. next, basing on the information from the pair *(x,y)*, we calculate values for a list of parameters, grouped into what referred as the vector of satellite indicators based models: $z = (z^1, z^2,...,z^s)$ (with $z^s = g^s(x, y)$, $(x, y) \to z^s$).

As the result of the integrated modeling we gets parameters summarized in the triplet *(x,y,z)*. This combined vector is a starting point for the modeling of socio-economic, socio-ecological and risk parameters (Fig. 1).

Usually, we consider a group of *S* satellite data based models, labeled $s \in \{1,2,...,s\}$. Each of these equations is such that the endogenous variables z^s can be obtained as an explicit mapping of the core variables $z^s = g^s(x, y)$.

Therefore the satellite model might be conceptually presented by time-series of *(x,y,z)*, which will determine the behavior of z_t^s, with a residual term μ_t^s :

$$z_t^s = f(x_t, x_{t-1},...,x_{t-L}, y_t, y_{t-1},...,y_{t-L}, z_{t-T}^s, \mu_t^s) \tag{1}$$

As z_t^s is the only variable on the left-hand side of the equation, the relationship is unidirectional: from *(x,y)* to *z*. This simple time-series satellite data based model formally also allows for no interactions with other satellite variables nor any feedback between z_t^s and the core assumptions in *(x,y)* (Campbell and Wynne, 2011), so we can use both methodology utilizing as separate as well interlinked indicators.
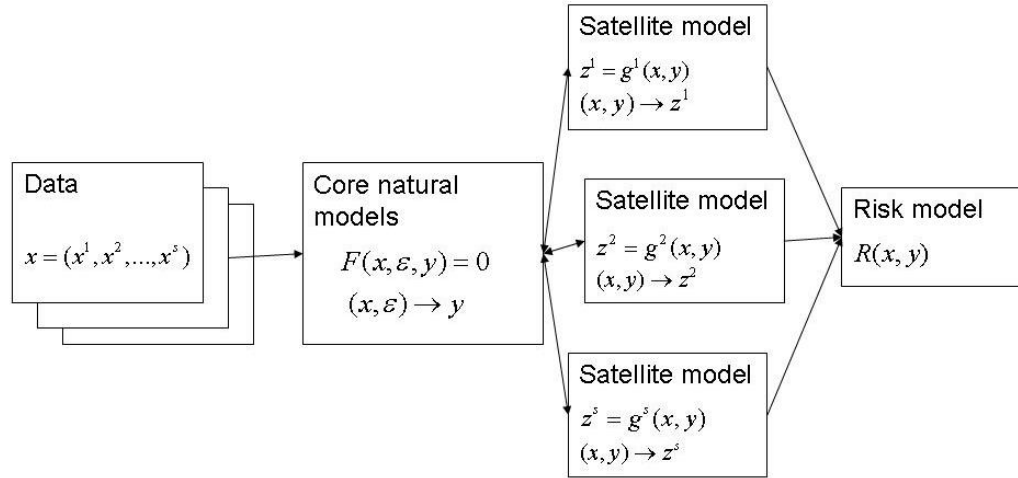
Fig. 1. Multi-model approach to socio-ecological risks assessment

Traditional time-series models such as autoregressive moving average models are good examples of satellite data based models. This representation may include, for example, autoregressive lags and/or moving average components. A number of standard methodologies (for example, Box-Jenkins method, or nonlinear autoregressive exogenous model) are followed to find the most usable model of the data-generating process for a given risk metric Z_t:

$$Z_t = c + \sum_{l=0}^{N} \beta_l X_{t-l} + \sum_{l=0}^{P} \rho_l Y_{t-l} + \sum_{l=1}^{L} \partial_l Z_{t-l} + \sum_{k=0}^{K} \theta_k \varepsilon_{t-k} \tag{2}$$

where Z_t is a satellite variable, X_t is a row vector of initial exogenous core variables, Y_t is a row vector of the layer of core parameters series, and ε_t is the value of the stochastic error term. The parameters c, β, ρ, ∂ are unknown and should be estimated.

However, including autoregressive terms in the model often results in a muted impact of core drivers on a target variable. Thus, it is a common and recommended practice is to exclude autoregressive terms from the supplementary variable equations (Engle and Russell, 1998). Therefore, depending of risk metric Z_t and of type of supplementary variable may be applied different form of equation (2). For example, for analysis of climate related risk an approach based on copulas utilization may be used (Kostyuchenko et al., 2013).

A key aspect of satellite model development is variable selection to identify which core drivers best explain the dynamic behavior of the studied socio-ecological risk variable (Kostyuchenko et al., 2013). In accordance with modern principles of Earth sciences, our approach toward variable selection is based on a combination of ecology, climatology, hydrogeology, hydrology, and geostatistics as consideration of the statistical properties of the estimated model (Kostyuchenko, 2015).

Models built using pure data-mining techniques or principles such as machine learning, neural network, etc., though they may fit the existing data well, are more likely to fail in a changing external environment because they lack theoretical underpinnings. The best analytical and prediction models employ a combination of statistical rigor with physical principles (Kostyuchenko et al., 2013). Hence, our models combine geo-ecological models with statistical optimization (Fig. 2). Models built this way have an additional benefit of ease of interpretation.
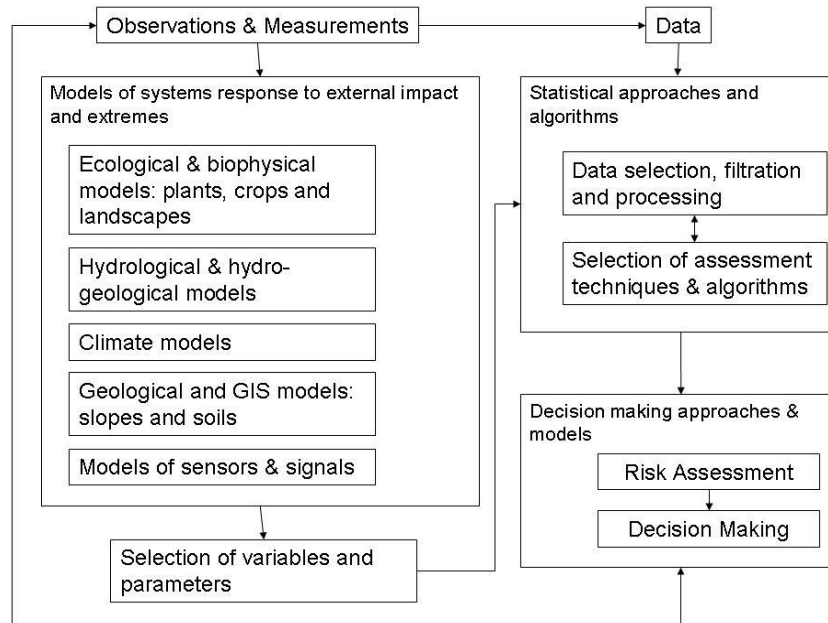

Fig. 2. Data utilization in framework of multi-modeling approach for socio-ecological risks assessment

The satellite model development process consists of selecting optimal exogenous drivers X_t, Y_t in equation (2) from a set of potential drivers. Once the final model is selected and estimated, the conditional dynamic forecasts of Z_t are generated given the sets of final parameters estimates and the forecasts of the core variables from the first stage (Fig. 2). The final step is to validate the final model in and out of sample.

The procedure of selecting optimal drivers is the following. Potential drivers are identified based on relevant theory and ensuring with calibration measurements or a priory assumption. This ensures that we obtain the most robust and predictive model available from the tested variables. To avoid model over-fitting, uncorrelated core drivers are preferably selected. The selected drivers should be significant at a conventional level and have the analyzed parameters of distribution. To obtain a required distribution a regularization procedure should be applied (Kostyuchenko, 2015). The final models selected by the search procedure are reviewed for consistency with initial assumptions.

So, the problem is the selection of variables for each model type $(x, \varepsilon) \rightarrow y$, the search of the relevant type of formal relationship $(x, y) \rightarrow z^s$ between the physical and observable variable parameter, and development of the total distribution for each type of risk investigated (2). There also should be separately considered the problem of regularization of initial distributions of variables (Kostyuchenko et al., 2015).

After we obtain distributions of parameters that determine the state of the system, we need to estimate the distribution of risk and make the management decisions.

3. Models of Risk Assessment and Decision Support
Basing on the obtained sets of indicators, the methodology of risk assessment, based on the optimal decisions can be proposed (Ermoliev et al., 2012).

In the framework of task of risk assessment and risk management, as opposed to the classic example of the resource management, reducing or non-increasing of losses should be used as quantitative characteristics. Variables that affect the characteristics of the management system (or decision-making system) can be or we can influence (controlled) and independent from our influence (unmanageable). Thus, the controlled variables are the parameters of decision-making under the influence of information (input data) on the behavior of unmanageable variables. Analysis of the effectiveness of full process of collecting, processing, interpretation of information about the system studied, decision making, analysis of system's response to decisions may be considered as part of the "information – response" formalization.

Such formalization can be made as follows (Schlaifer and Raiffa, 1961): define as $I(x,y,z)$ (presumably stochastic) information obtained from direct (field) measurements, observations and model forecasts; $H_I(i|\theta)$ - is probability distribution function and where θ – is a state of the studied natural object or system. In general, the state of the system cannot be determined with certainty, and thus we should define the appropriate probability distribution $p(\theta)$ and distribution $H_I(i|\theta)$, which describing a priori incompleteness of information available about the studied system.

Making and implementation of management decisions will formalize as a response to incoming information as decision function $d(I)$. Classical approach assuming that in the case of certain specific strategy of decision making in conditions of constant state of natural systems, θ, or with a defined change this state, the losses are defined as $l(d(I), \theta)$. For decision function d expected losses or risks associated with the development of dangerous processes, connected with the management decisions based on the information received, can be described as:

$$R(I,d) = R\big[H_I(\bullet),d),d\big] = \iint l(d(i),\theta)dH_I(i|\theta)p(\theta)d\theta \tag{3}$$

This risk is minimized by optimal decision function $d*$, entitled as Bayes decision function, and it is determined by information I:

$$R(I(z^s),d*) = \min_{d(i)} \iint l(d(i),\theta(x,y))dH_I(i|\theta)p(\theta)d\theta \tag{4}$$

So, the risk is a simple functional of decision function. It is important that the minimization of losses requires an intention to completeness of information on the studied system, i.e. definition of θ states, for each of which can be defined a solution $\{a\}$ (which build up the set of possible decisions or administrative actions A).

Let's consider the realization of the set of data (information obtained from direct measurements (field), observations and model forecasts) i^*, which optimizes the decision function d^* and minimizes appropriate risk, therefore nominally makes information (I) formally completed (I^*).
From a formal viewpoint, $I^* \equiv I$ (information is nominally full), if there exists a function $\phi(i)$, that when $H_I(i|\theta) \neq 0$, $\theta = \phi(i)$. In other words, completeness formally means that there is a single state of studied natural object or system that meets all of the set of data - a separate realization of information I. Here it means that we have to develop a set of models $(x, \varepsilon) \rightarrow y$ that have operated the set of parameters $(x, y) \rightarrow z^s$, which can be controlled by certain technological tools within the framework of a sustainable methodology of measurement. In our case, these requirements correspond to the data and methods of satellite observation of the Earth's surface z_t^s.

In the case described an optimal decision function may be defined as follows:

$$d*(i^*) = b , \; l(b, \phi(i^*)) = \min_a l(a, \phi(i^*))$$ (5)

Then, under optimal decision function and nominally complete information about the studied system, the risk will be defined as:

$$R(I^*, d^*) = \int \min_{a \in A} l(a, \theta) p(\theta) d\theta$$ (6)

For each case should be described models aimed to analysis of behavior the distribution of $H_I(i|\theta)$ та $p(\theta)$, and so to the determination of the realization of i^* of set I. As an optimal decision function in this approach may be used stochastic (Kopachevsky et al., 2016; Ram and Manglik, 2016), Bayesian (Kostyuchenko et al., 2012), neural network (Joshi, 2016) or fuzzy operators, (Kostyuchenko et al., 2016) depending on the task, data availability and properties of their distributions.

Considered complex of analytical models is aimed to the calculation of a unique set of parameters that should be obtained from determined observation systems, using defined tools of processing and interpretation of data. Equation (6) allows to estimate the distributions of risk of disasters and also to develop a basement for a system of risk management decision-making.

4. Conclusions
Modeling of geo-systems should be an integral part both of remote sensing interpretation methods, as well as of the risk assessment systems based on remote sensing data utilization. It requires of increased level of our knowledge in the field of Earth sciences, as well by increased

requirements in the area of decision-making. New challenges define new methodological requirements.

Firstly, the methodology proposed allows to expand the problem definition of using the satellite observations in tasks of socio-ecological security. In addition to traditional statistical analysis directed to surface change detection, it is possible to analyze and predict state of the studied systems, basing on the models of geo-systems.

This certainly expands the scope and sphere of application of approach, and could positively affect the reliability of the results obtained through the using of different sources of data.

Second, the proposed methodology includes feedbacks between management decisions and the systems state. Thus, it is postulated that the state of the system depends on the observer: risks depend on the decision made and management impacts (past, current and planned) to the system.

This could positive affecting to the effectiveness of management decisions and to the quality of risk assessment (Ermoliev and Winterfeldt, 2012).

References

Campbell, J. B., & Wynne, R. H. (2011). *Introduction to remote sensing*. Guilford Press.

Elachi, C., & Van Zyl, J. J. (2006). *Introduction to the physics and techniques of remote sensing* (Vol. 28). John Wiley & Sons.

Engle, R. F., & Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, *66*(5), 1127-1162.

Ermoliev, Y., & von Winterfeldt, D. (2012). Systemic risk and security management. In *Managing Safety of heterogeneous systems* (pp. 19-49). Springer Berlin Heidelberg.

Ermoliev, Y., Makowski, M., & Marti, K. (2012). Robust management of heterogeneous systems under uncertainties. In *Managing Safety of Heterogeneous Systems* (pp. 1-16). Springer Berlin Heidelberg.

Ermoliev, Y., Makowski, M., & Marti, K. (Eds.). (2012). *Managing safety of heterogeneous systems: decisions under uncertainties and risks* (Vol. 658). Springer Science & Business Media.

Joshi, R. (2016). Artificial Neural Network (ANN) based empirical interpolation of precipitation. *International Journal of Mathematical, Engineering and Management Sciences*, *1*(3), 93–106.

Kopachevsky, I., Kostyuchenko, Y. V., & Stoyka, O. (2016). Land use drivers of population dynamics in tasks of security management and risk assessment. *International Journal of Mathematical, Engineering and Management Sciences*, *1*(1), 18-24.

Kostyuchenko, Y. V. (2015). Geostatistics and remote sensing for extremes forecasting and disaster risk multiscale analysis. In *Numerical Methods for Reliability and Safety Assessment* (pp. 439-458). Springer International Publishing. DOI 10.1007/978-3-319-07167-1_16.

Kostyuchenko, Y. V., Bilous, Y., Kopachevsky, I., & Solovyov, D. (2013). Coherent risk measures assessment based on the coupled analysis of multivariate distributions of multisource observation data. In *Proceedings of 11-th International Probabilistic Workshop*, pp.183-192.

Kostyuchenko, Y. V., Kopachevsky, I., Yuschenko, M., Solovyov, D., Marton, L. A. S. Z. L. O., & Levynsky, S. (2012). Spectral reflectance indices as indirect indicators of ecological threats. *Sustainable Civil Infrastructures–Hazards, Risk, Uncertainty, Phoon, K. K., Beer, M., Quek, S. T. & Pang, S. D. (Editors), Research Publishing, Singapore, pp.*557-562.

Kostyuchenko, Y. V., Movchan, D., Artemenko, I., & Kopachevsky, I. (2016). Stochastic approach to uncertainty control in multiphysics systems: modeling of carbon balance and analysis of ghg emissions using satellite tools. In *Mathematical Concepts and Applications in Mechanical Engineering and Mechatronics*, ed. by Mangey Ram and J. Paulo Davim, IGI Global, USA, 2017, pp. 350-378, DOI: 10.4018/978-1-5225-1639-2.ch017.

Kostyuchenko, Y. V., Movchan, D., Kopachevsky, I., & Bilous, Y. (2015, November). Robust algorithm of multi-source data analysis for evaluation of social vulnerability in risk assessment tasks. In *SAI Intelligent Systems Conference (IntelliSys)*, 2015 (pp. 944-949). IEEE, DOI: 978-1-4673-7606-8/15.

Kostyuchenko, Y. V., Yuschenko, M., Movchan, D., Zagorodny, A. G., & Yermoliev, Y. M. (2013). Regional risk analysis based on multisource data statistics of natural disasters. In *Integrated Modeling of Food, Energy and Water Security Management for Sustainable Social, Economic and Environmental Developments*, 229-238.

Lillesand, T., Kiefer, R. W., & Chipman, J. (2014). *Remote sensing and image interpretation*. John Wiley & Sons.

Qiu, H. L., Sanchez-Azofeifa, A., & Gamon, J. A. (2007). *Ecological applications of remote sensing at multiple scales. In Functional Plant Ecology*, Second Edition. CRC Press.

Ram, M., & Manglik, M. (2016). Stochastic biometric system modelling with rework strategy. *International Journal of Mathematical, Engineering and Management Sciences*, *1*(1), 1-17.

Schlaifer, R., & Raiffa, H. (1961). *Applied statistical decision theory*. Harvard Univ. Press.

Suter II, G. W. (2016). *Ecological risk assessment*. CRC press.