# PROBABILISTIC APPROACHES TO ECOLOGICAL MODELING

TOMÁŠ HLÁSNY

National Forest Center – Forest Research Institute, T. G. Masaryka 22, 960 92 Zvolen, The Slovak Republic
e-mail: hlasny@nlcsk.org

Abstract

Hlásny T.: Probabilistic approaches to ecological modeling. Ekológia (Bratislava), Vol. 25, Supplement 1/2006, p. 66–75.

Current technologies provide comprehensive foundations for knowledge-based landscape management, modeling of dynamic landscape processes, various approaches to the multicriteria decision making for optimal land allocation, and temporal predictive and retrospective analyses of various landscape states and properties. Common denominator of these approaches is that require handling a substantial amount of data, which is often associated with uncertainties and utilizing expert knowledge. In this contribution we focused on certain aspect of probabilistic modeling with particular importance for landscape ecology. The contribution provides theoretical foundations of the theory of regionalized variables and some parts of Bayesian Theory in the view of ecological modeling. These approaches have been practically demonstrated in the biotopes diversity evaluation case study and the results have been compared. In general, the theory of regionalized variables focuses on spatial variability evaluation and on predicting values at unrecorded positions, when the source data can be measured directly, whilst the Bayesian approach provides modeling capabilities when only the indirect proofs on the phenomenon under consideration is available.

*Key words*: ecological modeling, geostatistics, Bayesian probability theory, biotopes diversity

## Introduction

Recent years have brought intensive development of all branches of natural sciences as had never been noticed before. This primarily relates to the accelerated development of computer science progressively integrating into all fields of research. In the field of landscape ecology emerging cross-cutting areas facilitated intensive development of new approaches to ecological modeling, comprehensive spatio-temporal analyses and to effective multicriteria decision techniques, especially for the optimal land allocation.

In practice we distinguish variety of approaches to modeling – stochastic and deterministic, (Isaaks, Srivastava, 1989), dynamic or static (Turner et al., 2001), mechanistic, process-based (Goodchild et al., 1993), descriptive and prescriptive approaches (Tomlin, 1990) and many others. Special kind of modeling, significantly important for landscape ecology, is a spatial modeling, where spatial dimension plays a crucial rule. In landscape ecology we often use different kinds of models – models approximating real biodiversity state, models describing landscape suitability for certain ways of exploitation, dynamic models describing how organisms behave under different conditions, or models of various landscape states and properties. Common denominator of these approaches is the term of model – by definition, the abstraction or simplification of reality that helps to generate testable hypotheses, which can be used to guide field studies by exploring conditions that cannot be manipulated in the field (Turner et al., 2001). This means the biodiversity model provides information about certain aspects of landscape structure generalized by means of diversity index, landscape suitability model abstracts away from the real suitability depending on number of factors involved, predictive models provide generalized information on expected temporal behaviour of chosen landscape features and so forth. With regard to elusive landscape complexness, integrity and multidimensional nature, this facilitates all analyses and explorations in a feasible way. In this regard Harvey (1969, p. 448) stated that "In reality any system is infinitely complex and we can only analyze some system after we abstracted from the real system".

### Probabilistic and deterministic modeling

**Probabilistic modeling** deals with the statistical probability of occurrence of certain phenomenon. There are two important streams we focused on – **geostatistical modeling**, which provide, among others, specific tools for spatial variability analysis and predicting values at unrecorded locations, and **Bayesian modeling**, which derives the probability of occurrence different phenomena by means of a set of indirect proofs. An important asset of these approaches is an ability to involve uncertainties of both source data and individual analyses. Probabilistic modeling is a counterpart of **deterministic approach,** which is based on the exact knowledge of the most desirable information brought to bear on the problem of modeling. By definition, there are three conditions that must be satisfied to make a prediction in the deterministic realm – flawless models are required to characterize the event of interest, assumption must be honored and all model parameters must be known (Olea, 1999). Purely deterministic model has no stochastic part producing different results under stable conditions and allow reasonable extrapolation beyond the available sampling.

Probabilistic models of any kind consist of two parts – so-called deterministic and stochastic components. Deterministic component unambiguously describes the most typical features, or behaviour of the system under consideration. In the view of spatial modeling, this might be analytically expressed by a polynomial function of certain degree. A general term according to Ripley (1981) is as follows:

$$f(x,y) = \sum_{r+s<p} a_{rs} x^r y^s \quad \text{which is used to minimize} \quad \sum_{1}^{N} \{z(x_i) - f(x_i)\}^2$$

where $p$ expresses the order of the polynomial, $z(x_i)$ are available data points, $r$ and $s$ are respective coefficients (Fig. 9). On the contrary, the stochastic part expresses the random deviations from the deterministic trend – it originates in certain randomness of natural systems and often-unpredictable influence of anthropogenic activities and natural hazards.

## Geostatistical approach

In this part we focused on the theory that has been established in 60th, primarily for mining purposes. At the beginnings, this approach had been termed the theory of regionalized variables (Matheron, 1971), recently it has been changed onto the popular term of geostatistics (e.g. Journel, Huijbregts, 1978). The main concept is that all processes and structures are viewed as results of a random function $z(x)$ that generates values at the points $x_i$ over the considered spatial domain $D$. The random function consists of a set of spatially autocorrelated random variables $z(x_i)$ (biodiversity values, soil pH, contamination etc.). The set of random variables has a probability distribution function that describes relative probability of occurrence for a range of possible values. Since we conceptualized mean and variance of known source sample data set as the realizations of a random function, we might assume that both random variables and source data have the same mean and variance. Under the assumption of random function stationarity, these parameters along with distribution function provide basis for respective modeling – we are able to predict the probability of occurrence certain values at certain positions.

Recently ecologists have begun to implement two geostatistical methods – variography, which is one way to model spatial dependence, and kriging, which provides the estimates for unrecorded locations (Rossi et al., 1992). The **variography** focuses on the evaluation how spatial variability develops at specific distances and by specific directions (e.g. Errikson, Šiška, 2000; Hlásny, 2005). The importance of this procedure outlined for example Pielou (1977), who stated that ecological analysis normally includes the investigations of the dispersions and patterns in association between different species at different places and at different times – patterns that reflects spatial dependence.

A basic tool for the spatial variability analysis – the variogram is in the terms of random functions theory defined as the integral over the squared differences of random variables $z(x_i)$ for a given lag $h$ (separative distance). Thus,

$$G_R(h) = \frac{1}{2|D(h)|} \int_{D(h)} (z(x+h) - z(x))^2 \quad \text{with expectation} \quad \gamma(h) = E[G_R(h)]$$

that gives the exactly defined theoretical (model) variogram over the domain $D$ (Wackernagel, 1998; Matheron, 1971).

The **kriging** provides wide set of approaches to the estimation or prediction values at unrecorded locations. The prediction strictly follows the results of variographic analysis, hence all the information about the spatial variability are involved. In the terms of random functions theory, geostatistical estimation is viewed as an outcome of random process created by weighted linear combination of other random variables (Journel, Huijbregts, 1978; Isaaks, Srivastava, 1989). A general form is

$$Z(x) = \mu(x) + \varepsilon(x)$$

where $\mu(x)$ is a mean function and $\varepsilon(x)$ is a random error process, having $E(\varepsilon(x)) = 0$. The most frequently used ordinary kriging is given as:

$$Z(x_0) = \sum_{i=1}^{k} \lambda_i Z(x_i) \qquad \text{subject to} \qquad \sum_{i=1}^{k} \lambda_i = 1,$$

where $Z(x_i)$ is $k^{th}$ random variables inside $D$, $\lambda$ are respective weights. The weights $\lambda$ are computed with respect to variogram behavior, ensuring the prediction is unbiased (the sum of weight is equal to 1) and it has the smallest mean square error of prediction (prediction variance). This results in a smooth surface of the phenomenon considered. Besides, as we can predict values at unrecorded locations, we can also derive the prediction variance that expresses the uncertainty of resultant model (although, geostatistical simulation provide more effective approach to this).

## Bayesian modeling

The second approach we briefly outline is the Bayesian Probability Theory, which provides unique tools for modeling spatial phenomena. This is an extension of Classical Probability Theory, which allows us to combine new *evidences* about a hypothesis along with prior knowledge to arrive at an estimate of the likelihood that the hypothesis is true. A general concept is

$$p(h/e) = \frac{p(e/h).p(h)}{\sum_i p(e/h_1).p(h\ )},$$

where *p(h/e)* expresses the probability of the hypothesis being true given the probability (*posterior probability*), *p(e/h)* the probability of finding that evidence given the hypothesis being true (*conditional probability*) and *p(h)* expresses the probability of the hypothesis being true regardless of the evidence (*prior probability*) (Bernardo, Smith, 1994; Jaynes, 1996). The term of conditional probability refers to the set of proofs or supportive evidences – the probabilities, which are interpreted as a *degree of belief*. Important concept is that beliefs are always subjective, so that they are not an objective property of some physical setting, but they are conditional to the prior assumptions and experiences (Cox, 1946). This approach is particularly useful, when direct measurement is impossible, or hardly feasible. In such cases, the sets of indirect proofs are to be involved to assess the probability of occurrence considered phenomenon or event.

## Case study

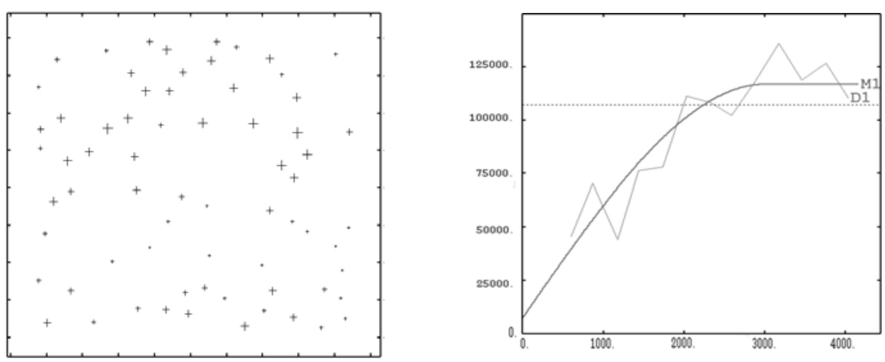
In this part we have compared the Bayesian and geostatistical approach to the biotopes diversity evaluation in the Štiavnicke Bane (Central Slovakia) research area. Consequently, respective deterministic and stochastic components have been analyzed. As source data the results of field biotopes diversity measurement, model of landscape heterogeneity at the terrain morphometry level (according to Hlásny, 2003) and research area road network along with settled area border have been used. Biotopes diversity values are expressed in the Shannon's index units with the base of logarithm *e*. Landscape heterogeneity model at the terrain morphometry level has been derived by means of moving window analysis (e.g. Ripley, 1981) from the original morphometry model.

## Geostatistical approach

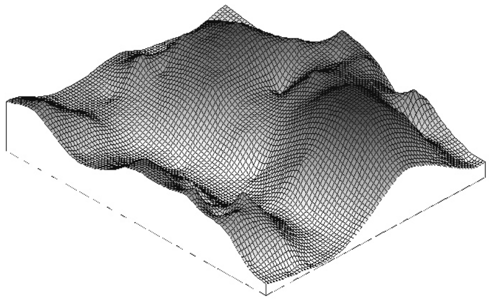
To model the biotopes diversity spatial structure in the realm of geostatistics, all the measured samples have been viewed as an outcome of a random process. On the bases of the set of known random function realizations, respective covariance function has been derived. Source data distribution and covariance function can be seen in the Figs. 1, 2. As can be seen, the data are significantly autocorrelated up to the distance of 280 m. For the sake of simplicity, the isotropical behaviour has been supposed. Furthermore, by means of linear weighted combination of known realization of a random function, with respect to the variogram behaviour, all the values at unrecorded locations have been predicted (Fig. 3).

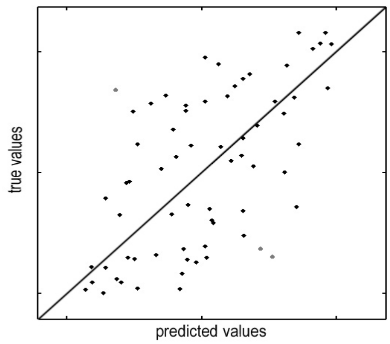
As can be seen, the highest values follow the central part, what relates to the main communication course and also to the highly diversified terrain morphometry (Fig. 6).

To asses the reliability (accuracy) of derived model the cross-validation procedure according to Wackernagel (1998) has been accomplished. The cross-plot of observed vs.




Figs 1, 2. In the field measured realizations of a random function (biotopes diversity values) and its covariance (expressed by spherical variogram with nugget effect). Symbols size expresses the magnitude of Shannon's index.



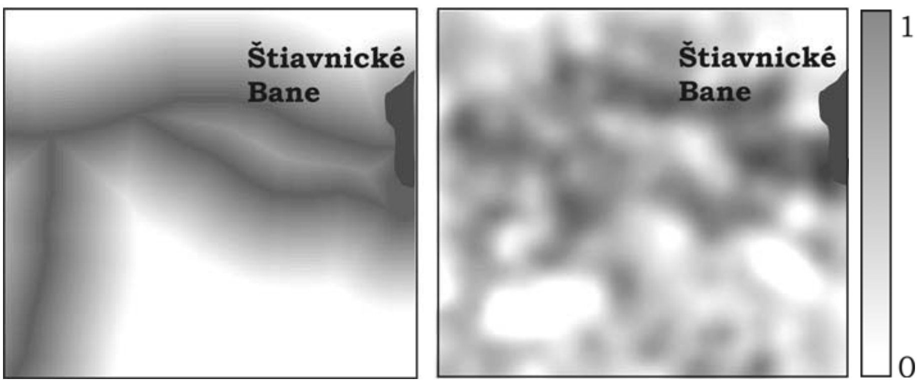



Figs 3, 4. Biotopes diversity model predicted on the bases of source data set in the Fig. 1 and respective scatter-plot of true vs. predicted values (*cross-validation procedure*).

predicted values might be seen in the Fig. 4. Respective standardized error expressed in the standard deviation units is 1.13 with the mean value of 0.031.

## Bayesian approach

In spite of the fact that this approach is primarily aimed at modeling hidden phenomena, on the bases of indirect proofs, we use the above problem to compare this solution with the geostatistical approach. For the sake of simplicity we designed very simple hypothesis – higher biodiversity values relate to the road network and to the settled area (*prior probability*), and to the highly




Figs 5, 6. The model of distance from settled area and the model of landscape heterogeneity at the terrain morphometry level (according to Hlásny, 2003) derived from original morphometry by means of moving window analysis.

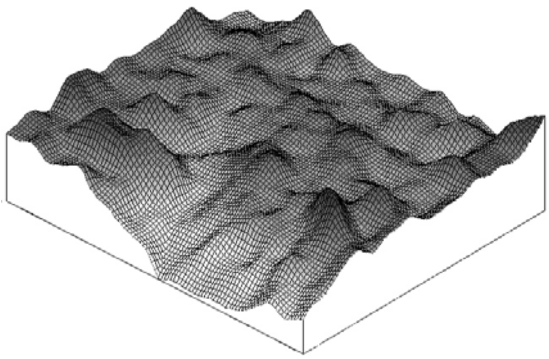


Fig. 7. The model of the posterior probability of high diversity values occurrence.

diversified terrain (*supportive evidence, conditional probability*). This assumption is based on previous research, where the correlation coefficient of biotopes diversity vs. distance from the village was *–0.417* and of biodiversity vs. morphometry *0.711*. The *posterior probability* of higher biotopes diversity values occurrence is based on the combination of prior probability and available evidences by means of the Bayesian rule above.

In the context of spatial modeling all these components can be represented by so-called probability surfaces (Figs 5, 6), which provide the information on respective probabilities taken on by all the positions of considered spatial domain. Prior and conditional probability models are expressed in the range of 0–1, although the value of 1 does not need to be reached, if the uncertainty due to the lack of confidence in data or hypothesis is to be expresses.

As can be seen, derived probability model is quite different from that derived using kriging. In general, indirect proofs such as morphometry heterogeneity and distance from the settled area and from the road network provide too unreliable evidences to models such a complex phenomenon as the biotopes diversity. In spite of these facts, there are certain features, which can be found to be similar. The most remarkable one is a higher values occurrence in the central zone and low values at the edges of the study area. These similarities have also affected the correlation coefficient magnitude of these two models that is of 0.397.

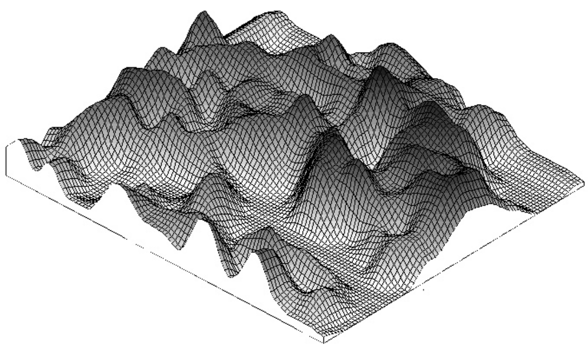


Fig. 8. The model of the posterior probability of high diversity values occurrence.

In order to synthesize the results of both introduced approaches, the easy feasible approach is to combine all the available data sources – to use the geostatistical model as the prior probability, and distance from the settled area and from the road network, and terrain morphometry heterogeneity as the supportive evidences. The

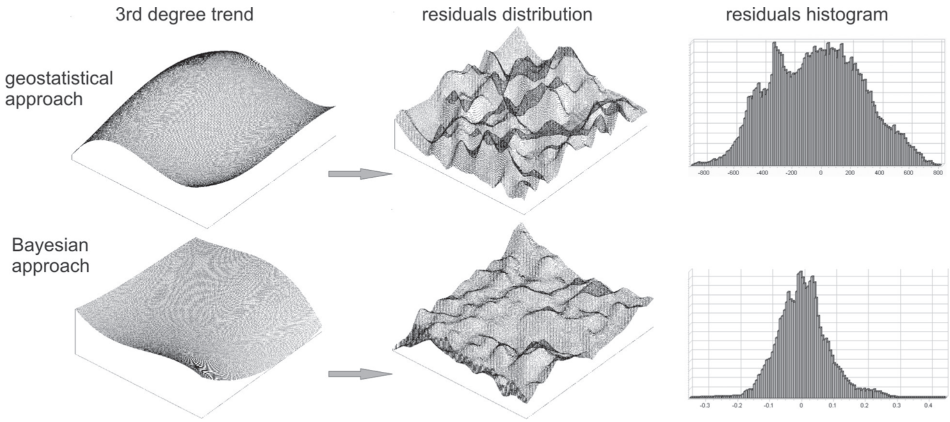



Fig. 9. The figure shows the trend components of both geostatistical and Bayesian model, respective residual structure and residual's histograms.

result of this synthesis can be seen in the Fig. 8. As can be seen, the model reflects the features of all prior and conditional probability models – the course of high values in the central zone typical of both geostatistical and Bayesian model, accompanied by slightly decreasing trend, along with remarkable fragmentation of central zone typical of terrain morphometry.

## Stochastic and deterministic components

The decomposition of derived models into the deterministic and stochastic parts brings unique sight on their structure and allows inferring theories on their genesis. To decompose geostatistical and Bayesian models, the 3rd degree trend have been extracted. Furthermore, the trend models have been subtracted from the original models to investigate visually the residuals distribution (Fig. 9).

As can be seen, there are quite remarkable differences between respective trend models, although certain common features can be found. The trend of geostatistical model shows the course of highest diversity values passing from E to W, whilst the Bayesian model values significantly decrease towards the SE part. As far as geostatistical model, this behaviour undoubtedly relates to the proved decreasing trend of diversity values from the settled area outwards, whilst in the case of Bayesian model this behaviour might reflect insufficient source data, or in general, incorrectly designed hypothesis. As far as residuals component, this might be used to determine areas with significant proportion of stochasticity, or on the contrary, those following the overall trend.

## Conclusion

In the paper we outlined the bases of geostatistics and Bayesian Probability Theory, two diametrically different streams, providing capabilities for the modeling of different ecological phenomena and processes. The approaches to spatial dependence analysis, predictions at unrecorded locations and treating different supportive evidences as conditional probabilities to support given hypotheses were described. A simple analysis aimed at modeling biotopes diversity using these approaches was carried out. We proved the effectiveness of selected statistical and mathematical tools integrated into a GIS environment for landscape-ecological modeling and analyses. The difficulties associated with the introduced modeling relate mainly to proper design of underlying hypothesis, requiring a great portion of experience.

From the methodological view, there is a lot of derived forms of the approaches above, developed to achieve more exhaustive data analysis – the kriging might deal with multivariate data, provide non-linear estimations, or to work in a 3-dimensional space. Analogously, the Bayesian Theory might be extended into the popular Decision Theory, which is much more flexible in dealing with conditional probabilities. Furthermore, all these approaches can be extended into the temporal dimension – to provide temporal predictions, analyze time series stochastic and deterministic components, or to evaluate the causal aspect of respective time series behaviour. All these approaches help understanding the variety of natural phenomena and processes around in order to move the scope of the knowledge of landscape.

*Translated by the author*

### References

Bernardo, J.M., Smith, A.F., 1994: Bayesian Theory. Wiley and Sons, New York, 586 pp.

Cox, R.T., 1946: Probability, frequency and reasonable expectation. American Journal of Physics, *14*, 1, p. 1–13.

Erriskon, M., Šiška, P., 2000: Understanding Anisotropy Computations. Mathematical Geology, *32*, 6, p. 683–700.

Goodchild, M.F., Parks, P.O., Steyaert, L.T., 1993: Environmental Modeling with GIS. Oxford University Press, 520 pp.

Harvey, D., 1969: Explanation in Geography. London, Edward Arnold, 521 pp.

Hlásny, T., 2003: Landscape heterogeneity as a measure of landscape system entropy. Ekológia (Bratislava), *22*, Suppl. 2/2003, p. 130–140.

Hlásny, T., 2005: The geostatistical concept of spatial dependence for geographical applications (in Slovak). Geografický Časopis, *54*, 2, p. 97–116.

Isaaks, H.E., Srivastava, R.M., 1989: Introduction to Applied Geostatistics. Oxford university Press, 560 pp.

Jaynes, E.T., 1996: Probability Theory. The logic of science, http://bayes.wustl.edu

Journel, A.G., Huijbregts, CH.J., 1978: Mining Geostatistics. Academic Press London, 600 pp.

Matheron, G., 1971: The Theory of Regionalised Variables and its Applications. Les Cahiers du Centre Morphologie Mathématique de Fontaineblau, 209 pp.

Olea, R.A., 1999: Geostatistics for Engineers & Earth Scientists. Kluwer Publishers, 304 pp.

Pielou, E.C., 1977: Mathematical Ecology. John Wiley and Sons, New York, USA, 386 pp.

Rossi, R., Mullad, A., Journel, A., Franz, E., 1992: Geostatistical Tools for Modeling and Interpreting Ecological Spatial Dependence. Ecological Monographs 629, Ecological Society of America, p. 277–314.
Ripley, B.D., 1981: Spatial Statistics. Wiley and Sons, Inc., New York, 252 pp.
Tomlin, C.D., 1990: GIS and Cartographic Modeling. New Jersey, Prentice Hall, 249 pp.
Turner, M., Gardner, H.R., O'Neill, V.R., 2001: Landscape Ecology in Theory and Practice. Patterns and Process, Springer-Verlag, New York, 404 pp.
Wackernagel, H., 1998: Multivariate Geostatistics. Springer Verlag, Berlin, 291 pp.

*Received 18. 11. 2003*

Hlásny T.: **Pravdepodobnostné prístupy k ekologickému modelovaniu.**

Súčasné technológie vytvárajú komplexné základy pre manažment krajiny založený na syntéze množstva čiastkových informácií, modelovanie dynamických procesov v krajine, rozličné prístupy k multikritériovému rozhodovaniu pre optimálne využitie územia alebo pre časové prediktívne a retrospektívne analýzy rozličných stavov a vlastností krajiny. Spoločným menovateľom týchto prístupov je potreba spracovávania mimoriadne veľkých objemov údajov, ktoré sú často zaťažené istou mierou neurčitosti a vyžadujú integráciu expertných znalostí. V príspevku sme sa zamerali na určité aspekty pravdepodobnostného modelovania s aplikáciami v krajinnej ekológii. Uvádzame základy teórie regionalizovaných premenných a Bayesovej teórie v kontexte krajinno-ekologického modelovania. Tieto prístupy sme prakticky demonštrovali pomocou prípadovej štúdie hodnotenia diverzity krajiny na úrovni biotopov a výsledky sme porovnali. Vo všeobecnosti sa teória regionalizovaných premenných zameriava na hodnotenie priestorovej variability analyzovaného systému a predikciu hodnôt na nezmapovaných pozíciách v prípadoch, keď sa zdrojové údaje merajú priamo, kým využitie Bayesovej teórie vytvára silný nástroj krajinno-ekologického modelovania v prípadoch, keď o charaktere analyzovaného systému sú dostupné len nepriame dôkazy.