Представлена дескриптивна динамічна модель бінарних даних, що дозволяє по вихідним спостереженнями з порушеним часовим порядком відновити вихідний порядок на підставі принципу парсимонії. Модель застосована для знаходження системних колориметричних параметрів, використовуваних для обробки зображень скупчень токсичних ціанобактерій на основі аналізу компонентів RGB-моделі цифрової фотографії

Ключові слова: дескриптивні моделі, динамічні системи, бінарні дані, парсимонія, інтелектуальний аналіз даних, скупчення токсичних ціанобактерій

Представлена дескриптивная динамическая модель бинарных данных, позволяющая по исходным наблюдениям с нарушенным временным порядком восстановить исходный порядок на основании принципа парсимонии. Модель применена для нахождения системных колориметрических параметров, используемых для обработки изображений, скоплений токсических цианобактерий на основе анализа компонентов RGB-модели цифровой фотографии

Ключевые слова: дескриптивные модели, динамические системы, бинарные данные, парсимония, интеллектуальный анализ данных, скопления токсических цианобактерий

DEVELOPMENT OF THE DESCRIPTIVE BINARY MODEL AND ITS APPLICATION FOR IDENTIFICATION OF CLUMPS OF TOXIC CYANOBACTERIA

K. Nosov
PhD, Research Fellow
Scientific research unit*
E-mail: k.nosov@karazin.ua

G. Zholtkevych
Doctor of Technical Sciences, PhD, Professor
Department of Theoretical and Applied Computer Science*
E-mail: g.zholtkevych@karazin.ua

M. Georgiyants
MD, Professor
Department of Pediatrics Anesthesiology and Intensive Therapy
Kharkiv Medical Academy of Postgraduate Education
Amosova str., 58, Kharkiv, Ukraine, 61176
E-mail: eniram@bigmir.net

O. Vysotska
Doctor of Technical Sciences, Professor**
E-mail: olena.vysotska@nure.ua

Y. Balym
Doctor of Veterinary Science, Professor
Department of Reproductology
Kharkiv State Zooveterinary Academy
Academichna str., 1, Malaya Danylivka, Dergachi district,
Kharkiv region, Ukraine, 62341
E-mail: yubalym8@gmail.com

A. Porvan
PhD**
E-mail: andrii.porvan@nure.ua
*V. N. Karazin Kharkiv National University
Svobody sq., 4, Kharkiv, Ukraine, 61022
**Department of Biomedical Engineering
Kharkiv National University of Radio Electronics
Nauky ave., 14, Kharkiv, Ukraine, 61166

1. Introduction

In analytics of raw data, in particular, in big data analytics, they often recognize three types of analysis – descriptive, predictive, and prescriptive. What are the differences between three main types of analytics?

Usually, the majority of raw data doesn't offer a lot of value in its unprocessed state. By applying the appropriate tools, we can pull powerful insights from the bulk of numbers, properties, characteristics.

Receiving the data in hand from the experiment or Internet, you can begin doing certain analysis and reveal useful

information. Scientists and data analysts have an impressive set of tools appropriate for data of different type. First, in order to take a look at data, the researcher must develop or select a model. Such a model allows him/her to capture-on-the-fly the main properties and relations underlying the data under consideration. Once the model is built, the researcher can predict.

Sometimes descriptive analysis is regarded as the simplest class of analytics that allows you to condense big data into smaller, more useful pieces of information. That step makes raw data more suitable for human consumption with the information derived from the data.

Predictive analytics is the next step up in data reduction. It utilizes a variety of statistical, modeling, data mining, and machine learning techniques to study data, thereby allowing analysts to make predictions about the future. Predictive analysis can forecast what might happen in the future because all predictive analytics are probabilistic in nature.

The emerging technology of prescriptive analytics goes beyond descriptive and predictive models by recommending one or more courses of action – and showing the likely outcome of – each decision. Prescriptive analytics not only anticipates what will happen and when it will happen, but also why it will happen. Further, prescriptive analytics suggests decision options on how to take advantage of a future opportunity or mitigate a future risk and shows the implication of each decision option.

The authors have previously proposed a number of descriptive models of system dynamics to describe the behavior of complex natural and technical systems [1]. The idea underlying these models is to present the system as a set of interacting components. In fact, each component corresponds to a certain part or property of the system. For example, in an animal community, a component is the number, or biomass, or density of a species from the community. For such a system, the model of between-component interactions can be built, and properties of interactions and dynamics can be identified by raw data obtained from the real (natural) system.

The development and expansion of the arsenal of mathematical tools for implementation of the systematic approach for tackling biosafety issues have always been the topical problem. Mathematical and information models describing the structure and stability of biological systems have the utmost importance. This problem is at the interface of cybernetics, mathematical modeling, computational and theoretical biology.

The clumps of biomass of toxic cyanobacteria, arising during their mass development in water reservoirs, create a threat to biosafety. These threats are related to the quality of potable water for animals through the release of rotting and dead toxins containing dead organics in coastal areas.

There is a wide range of veterinary medicine's solutions related to the toxicity of cyanobacteria, as well as the problems of environmental safety for humans [3]. In this connection, the monitoring of rising and movement of these clumps is one of the urgent problems not only in ecology, but also in veterinary medicine [4].

The topicality of this problem has recently increased to a large extent as a result of increasing anthropogenic pressure on natural systems in the whole and, particularly, on water bodies, as well as due to the influence of factors of the global climate change. Thus, cyanobacteria's biomass outbreaks occur in significant water areas in not only terrestrial water reservoirs, but also in seas, e.g., the Baltic Sea. The monitoring of these environmental violations requires remote (aerospace) methods, which use complicated equipment – satellite color scanners. However, as noticed in [5], the full picture of "bloom" recorded on satellite images of the water surface can be reproduced by neither experimental no theoretical models due to the complexity of relations between relevant factors of this phenomenon.

At the same time, as it was shown in [6], certain practically significant system aspects of the performance of plant communities can be remotely registered using mathematical modeling. Initial factual data, obtained by the methods that directly record only parameters of the RGB model of digital photography with the help of a relatively simple, widely used and not-expensive equipment, can be used [7].

Such a possibility is provided by new classes of mathematical models, developed with the authors' participation, called the discrete models of dynamical systems (DMDS). In particular, it was shown that the dynamics of some colorimetric parameters of crops of cultivated plants can be described by the mathematical model that is very similar to the so-called marginal model of succession [7]. The use of DMDS, in this case, provides the effect of increasing the information content while using relatively rough input colorimetric parameters due to the ability for modeling the relationships between these parameters.

2. Literature review and problem statement

Binary data arise when a particular response variable of interest can take only two values, say, {0, 1}. For example, to understand socioeconomic processes, economists often need to analyze individuals' binary decisions (whether to make a particular purchase, participate in the labor force, obtain a college degree, see a doctor, migrate to a different country, or vote in an election).

There are a few statistical methods for modeling binary data based on expression of relations between independent variables and a binary response. Apparently, logistic regression is the most popular model among the well-known ones [8]. It enables to calculate the probability of a value for a binary dependent variable using a number of independent variables in the interval or ordinal scale [9]. In addition to calculating the probability, logistic regression allows one to estimate the parameters of the model, calculate the quality of the statistical data, test hypotheses about the values of the parameters, etc. [10].

This fairly simple model has numerous extensions. They can be performed in various directions.

One of the directions is multilevel models for binary outcomes. Multilevel or clustered data consist of units of analysis at a lower level nested within units of analysis at a higher level [11, 12]. For this type of analysis, it is assumed that the hierarchy of units of analysis is natural. Thus, for individuals in a certain hierarchy, there is a tendency to be more similar in their characteristics than in a similar sample of individuals chosen at random from the population [13, 14]. Family relations are a good example of the natural hierarchy because we would expect that the children of the same parents to be similar in many important ways. Students nested in schools would be another typical hierarchical data structure. Often hierarchies reflect individual social differentiation, for example, when individuals with the similar ability are grouped

in selected schools. In other cases, nesting may less reflect the individual characteristics and may arise at random. For example, in clinical trials, we often find experimental research conducted at several randomly chosen hospitals or among randomly selected groups of subjects. Multilevel models, for example, can address clustered data, repeated measures or longitudinal data [15, 16].

When repeated measurements are conducted on the same individuals, a hierarchy is established with individuals at level two and measurement occasions at level one. These data are often referred to simply as longitudinal data. For data of such a type, a huge number of effective models have been developed. For example, in [17], for prospective studies with binary responses, they suggest using the Poisson's regression, which enables to build the models with correlated binary responses originating in longitudinal or cluster randomized trials. In [18], on the basis of the likelihood ratio, a new unified model was suggested. This model extends the traditional marginal regression models that describe consecutive and long-term follow-ups of binary variables.

An important method of revealing causal relations through statistical techniques is Structural Equation Modelling (SEM), which has been increasingly used in applied research. However, there is a great obstacle for its wider use in handling categorical, in particular, binary, variables. None the less, there are several more or less successful attempts to extend SEM to binary data. For example, in [19] they use Yule's transformation on the basis of odds ratios to approximate the matrix of Pearson's correlation coefficients. In [20], a SEM's extension for predicting a dependent binary variable (so called non-linear mixed model, NLMM) is suggested.

There are a few approaches to causal inference and structure learning of Bayesian networks studied in statistics and artificial intelligence [21, 22]. Most of them derive candidate causal structures from the observed data set by assuming acyclicity of causal dependencies. In these models, they use the information up to the second-order statistics of observed variables and narrow down the candidate directed acyclic graphs by using some constraints and/or scoring functions.

Meanwhile, the methods and models, both included in and beyond the review are hard-to-apply to some data structures. In this respect, the following general methodological remark can be made. Evidently, one should not expect that exhaustive set of models will ever be developed that adequately will describe all possible types of objects and systems – natural, technical, social, etc. The tasks a researcher is solving may require the development and study of models that cannot be reduced up to any of previously encountered.

In this study, we suggest a model of dichotomous data that solves the problem of processing a dataflow, which is formed as the series of observations of a certain dynamical system with discrete states. Additionally, the time order of observations can be disturbed, in contrast to the data similar to time series. Such a dataflow can arise, for example, when results of monitoring of the system's states are delivered via different channels, if measurements of these states are conducted asynchronously and in other similar cases. As far as the authors know, such models for dichotomous data have not been developed yet.

Consider the data from the dichotomous scale. In this case, the variables can take only 2 values from the set $B=\{0, 1\}$. Besides 0 and 1, these values can be denoted by true/false, male/female, etc. The order between the values is not established.

Let the system comprises N components, denoted by A_1, A_2,..., A_N; each component takes values from the set B. The system is dynamical and has discrete time. Hence, its state at the moment t can be denoted by $(A_1(t), A_2(t),..., A_N(t))$, where each $A_i(t){\in}B$. Also assume, that the system's state at the moment $t+1$ is determined in full by the state at the moment t.

Initial data are presented in the observation table composed of M cases (in rows) and N columns:

$$A = \begin{pmatrix} A_1 & A_2 & \cdots & A_N \\ a_{1,1} & a_{1,2} & \cdots & a_{1,N} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M,1} & a_{M,2} & \cdots & a_{M,N} \end{pmatrix}, \qquad (1)$$

where any $a_{ij}{\in}B$.

Each column corresponds to a corresponding component.

We assume, that

1) in the table (1), each row is a state of the system at a certain moment of time;

2) the table (1) includes all or at least the major part of available system's states;

3) all rows in (1) are unique (that is, $M{\leq}2^N$).

The following assumption is that (1) presents the whole cycle of the dynamical system, but the sequence of moments may be out of order. It is assumed, that the dynamic system described above consisting of N components generates a number of observations in the form of the matrix (1).

3. The aim and objectives of the study

The aim of the study is to develop a model that allows one to construct a descriptive model of the dynamical system from data. This enables to describe relationships between components and restore the dynamics of the system, in particular, the correct time order of observations.

To achieve the aim, the following objectives are stated:

– to recover the true order of the moments or the true order of rows using the observation matrix (1);

– to recover the rules of dynamics of the given system from the observation matrix (1);

– to develop the approach that enables more efficient identification of localization of clumps of toxic cyanobacteria with the use of digital images. This approach should be based on the analysis of weighted oriented graphs that reflect the dynamics of colorimetric parameters of snapshots of cyanobacterial clumps.

4. The descriptive models of binary data

Suppose, that the initial data in the form of the observation table (1) are given.

For addressing the task, we apply the *parsimony* principle: the simpler the model, the more correct it is. In contrast to the above-mentioned models developed by the authors, we don't minimize a discrepancy between the model's dynamics and observed data, but solve rather an unsupervised optimization problem.

Define a *K*-ary binary function as a map

$$\underbrace{B \times B \times \ldots \times B}_{K} \xrightarrow{F} B.$$

In this definition, *K* can take the value 0. In other words, the function *F* can take a constant value from B, 0 or 1. As the *K*-ary binary function is a map into B, it defines the two disjoint sets of series from zeros and ones of a length of *K*. The first set is the series mapped by the function *F* into 0, the second one – into 1. If $M=2^N$, a unique truth table for the map *F* can be composed. But, if $M<2^N$, the truth table is not unique.

Let us start with the second task.

Suppose, the table *A* is already a regular cycle. That is, its first row ($a_{1,1}$, $a_{1,2}$,..., $a_{1,N}$) is the state of the system at the moment *t*=1, the second – at the moment *t*=2, the *M*-th row – at *t*=*M*. After the last moment, the cycle is being repeated starting from the first row, as the system is finite-state. As mentioned above, the state at the moment *t*+1 is uniquely determined by the state *t*.

Now try to establish the relationships between the components. For example, consider the component A_1. It is needed to build such a binary function that performs the mapping (by rows):

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,N} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M-1,1} & a_{M-1,2} & \cdots & a_{M-1,N} \\ a_{M,1} & a_{M,2} & \cdots & a_{M,N} \end{pmatrix} \xrightarrow{F_1} \begin{pmatrix} a_{2,1} \\ a_{3,1} \\ \vdots \\ a_{M,1} \\ a_{1,1} \end{pmatrix}.$$

This means that the function *F* maps the system's state at *t*=1, written in the first row, in $a_{2,1}$ (the state of the component A_1 at *t*=2); the system's state at *t*=2 – into the component A_1's state at *t*=3 and so on.

Because all the rows of the table *A* are different, we can obtain a trivial solution of the problem using the *N*-ary function as F_1. But this mapping sometimes may be implemented by a m_1-ary function (on some m_1 components), where $m_1<N$.

The function F_1 constructed in such a manner is only defined on the entries from **B** available in (1).

For example, consider the table below (*N*=3, *M*=4)

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

If it is needed to find the mapping F_1, we easily can do this: $F_1: B \rightarrow B$ depends on A_2, $F_1(0)=0$, $F_2(1)=1$. So, $m_1=1$.

For a given component A_i, the minimal arity σ_i of the map built according to the described approach is called the *degree of dependency*. One σ_i can be provided by several components' subsets (of the same size).

Introduce the average degree of dependence for all components:

$$\sigma = \frac{1}{N} \sum_{i=1}^{N} \sigma_i, \tag{2}$$

where σ is the measure of the simplicity of the relations determined by the model. The smaller σ, the easier the relation,

the more adequate the models are, according to the interpretation of the principle of parsimony used.

If the rows in (1) are out of time order, we face the first task of revealing the true timeline sequence of observations. One needs to select a permutation of rows provided the minimal value of (2).

The upper boundary of the number of binary maps F_1, F_2,..., F_N to be used in solving the problem is as follows

$$N_{\max} = (M-1)! \times N \times 2^N,$$

where *M* is the number of rows' permutation, *N* is the number of components, and 2^N is the maximum number of sets to be viewed for each map F_k.

Development of effective algorithms for integer optimization appropriate for large *M* and *N* is an important and challenging problem.

5. Application of the theory to image processing regarding biosafety issues

The obtained theoretical results could have practical applications related to the identification of disturbance of bioproduction processes that create certain biosafety threats. A significant example of such a disturbance is, as already mentioned, the effect of spots of bloom resulting from eutrophication on biosafety of water consumption for watering domestic animals.

To identify and determine a location of clumps of cyanobacteria on the surface of water, a set of strategies that determine transitions between the states of clumps will be used. The parameters of the sets of strategies can be expressed in the terms of directly measured "primary" colorimetric parameters (CPs), as well as "secondary", system colorimetric parameters (SCPs). The "primary" colorimetric parameters include the values of the RGB model and their derivatives expressed by elementary functions. The SCPs can be, e. g., the evenness of the "primary" values and/or the range of their variations.

Within the framework of the present work, the abilities of the obtained results are shown by the example of digital image processing of clumps of toxic cyanobacteria called spots of bloom.

Consider clumps of cyanobacteria as a dynamic system in the state of dynamic equilibrium characterized by changing the values of its parameters within a certain cycle. To be more precise, the image of a clump, divided into a number of parts (rectangles in this case), is considered as a system comprising four components described below.

Using the RGB color model, for each triangle, the average value of 4 parameters R/(R+G+B), G/(R+G+B), (R+G)/(R+G+B) and R/G were calculated.

Using a set of triangles as a sample, the 1st and 3d quartiles of this sample were calculated. Then, initial values of 4 parameters were encoded. If for a rectangle, the average of the corresponding parameters lies between the 1st and 3d quartiles of this parameter, the new value is assumed to be 1, otherwise – 0. As a result of the encoding, we obtain the observation table with 4 columns and the numbers of rows equal to the number of rectangles in the division of the image.

The values of the components equal to 1 are hereinafter referred to as stable values (SVs); and, the values equal to 0 – unstable values (UVs).

The reference image, from which the set of strategies was derived, is shown in Fig. 1. The original image was taken from [23].

The set of strategies obtained by applying the model to the reference image is shown in Fig. 2–5. It shows the strategies of transition of system's components to SVs and UVs as the four weighted digraphs, one graph per each component of the system. A vertex with outgoing edges (at the top of each digraph) includes the lists of the components that define the strategies:

– the transitional strategies for the states of the parameter R/(R+G+B) (Fig. 2);

– the transitional strategies for the states of G/(R+G+B) (Fig. 3);

– the transitional strategies for the states of (R+G)/(R+G+B) (Fig. 4);

– the transitional strategies for the states of R/G (Fig. 5).

Fig. 1. The reference image (the selected rectangle in the center), on the basis of which a set of strategies was derived
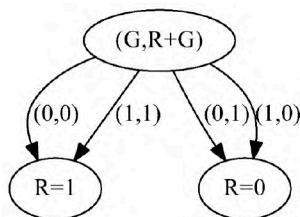

Fig. 2. The strategies for transition of system's components in the state R to SVs and UVs

The image was processed with the use of the "resynchronization" method. The method is based on the assumption that fairly large parts of the image of the modeled plant community change their color parameters according to the unique cycle.

For example, in Fig. 5, the list of strategies includes three components: G/(R+G+B), (R+G)/(R+G+B), R/G. The denominator of the components is omitted for brevity. Weights of edges correspond to the values of the components defined by the strategies' list. The vertices with entering edges correspond to SVs (1) and UVs (0) of the corresponding component.
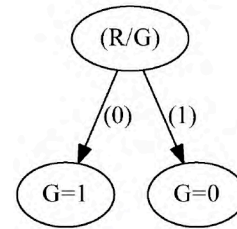

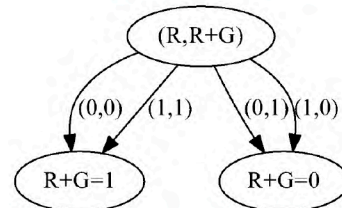Fig. 3. The strategies for transition of system's components in the state G to SVs and UVs

Fig. 4. The strategies for transition of system's components in the state (R+G) to SVs and UVs
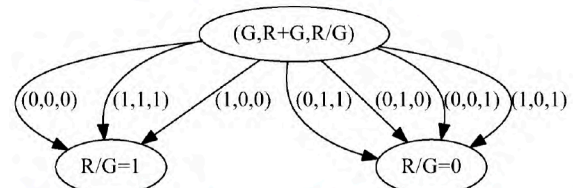

Fig. 5. The strategies for transition of system's components in the state R/G to SVs and UVs

On the basis of analysis of graphs constructed for each of the colorimetric parameters, the system colorimetric parameters (SCPs) can be built. A system colorimetric parameter expresses a certain feature of all the strategies, due to which the system oscillates.

In Fig. 2–5, there is only one case of the simplest set of strategies related to the component G/(R+G+B) (Fig. 3). For the three other graphs, the sets of strategies that lead to SVs and UVs differ in the degree of evenness. Non-strictly speaking, the evenness is defined as a number of identical values, zeros and ones, on the same positions in the strategies. In cases, for which the strategy is determined by two parameters, the degree of evenness is proportional to the modulus of the difference between them. That is, when both parameters are equal to 0 or 1 simultaneously, the difference will be equal to zero. But in two such cases (R/(R+G+B), (R+G)/(R+G+B)), the use of such the parameter of evenness results in a trivial solution, when the SCPs are reduced to one colorimetric parameter.

For the graph corresponding to R/G (Fig. 5), we have a more complicated case.

In this case, it also seems advisable to use a paired index of evenness, which includes two parameters out of three ones affecting the dynamics of R/G. In this case, the set of 3 strategies corresponding to these SVs includes only the balanced values of UVs (3 cases). And in the set of 4 strategies corresponding to UVs, there is only one unique balanced strategy.

$$\xi = \left| \frac{R+G}{R+G+B} - \frac{R}{G} \right|. \tag{3}$$

Thus, the transitions from SVs to UVs are characterized by a significant difference of this index of evenness.

Accordingly, we can expect a wider value range of this index on the image of clumps in comparison with the surrounding water. One can see that the index does not directly describe the dynamics of the system in the process of its oscillations, but characterizes this dynamics rather implicitly.

This assumption was verified using the standard deviation of the index. The result of the test is shown in Fig. 6–8.
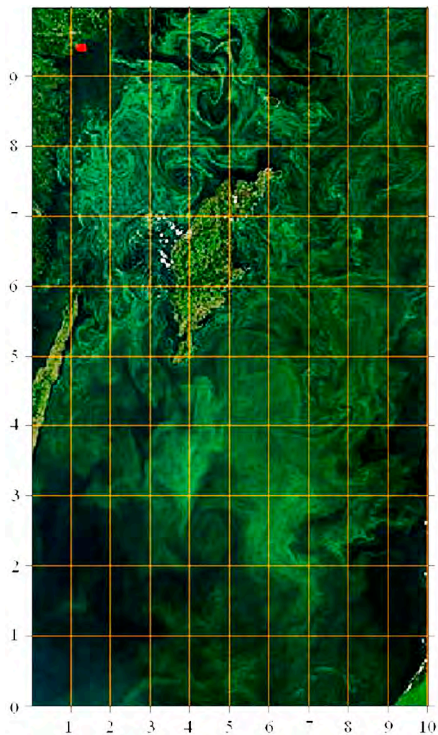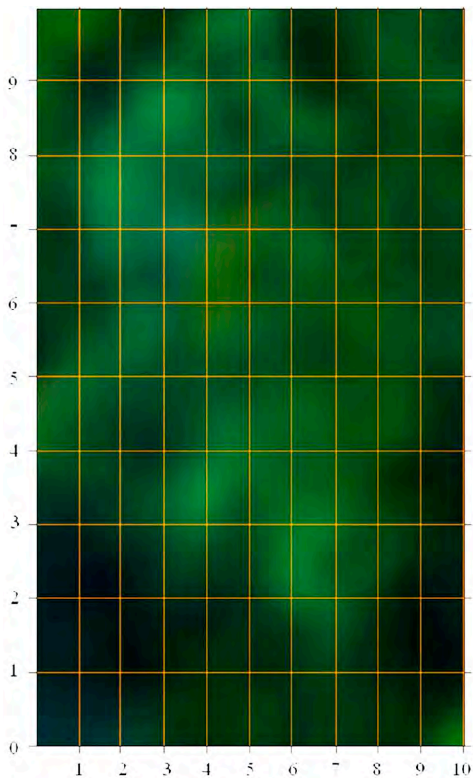

Fig. 6. Original noiseless image

Fig. 7. Original image with noise added

The digital noise was added to the original image (Fig. 7). The noise simulates a deterioration in visibility conditions, for example, the presence of fog, cloudiness, and precipitation. The noised image is shown in Fig. 7. Then the image with the noise was split into rectangles by a grid shown in all the illustrations (Fig. 6–8). Each rectangle of this partition, in turn, was divided into (several dozen of) smaller triangles. For any of the smaller triangles, the values of R, G, B were obtained from the image and the paired index of evenness (3) was calculated. The sample comprising the values of this index for each initial rectangle was used to calculate the standard deviation. The resulting image after the processing is shown in Fig. 8. Image processing and calculation were performed in Matlab software with Image Processing Toolbox installed.
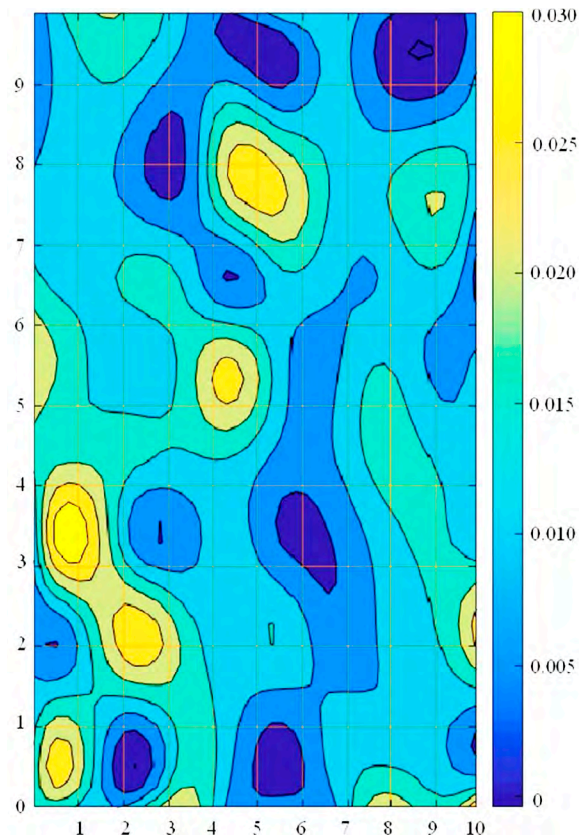

Fig. 8. The result of processing of the noised image with the intensity scale

Fig. 6 shows that the clumps of cyanobacteria on the initial image are not clearly visible and practically invisible on the noised image (Fig. 7), but some additional details are visible on the processed image (Fig. 8). Therefore, it can be assumed that clumps of cyanobacteria can be identified by a certain indicator of the range of the index of evenness according to (3).

6. Discussion of results and application of the discrete binary dynamic model

The descriptive model of system dynamics for binary data was developed. The model enables to reconstruct the time order from an unordered set of observations and describe the dynamics that generates the initial data. The model is an autonomous deterministic dynamical system comprising the set of binary components.

The identification of the model by observable data is based on the principle of parsimony. The essence of this principle is the following. The simpler the description of the system's dynamics (in a certain sense determined by a numerical measure), the more the system is appropriate. Thus, the problem of integer optimization, minimizing the specified measure of simplicity for system identification from data should be solved. The upper bound of the number of binary mappings, which are candidates for solving the optimization problem, is obtained. This number grows rapidly with the increase of dimension of the problem, and, even for initial data of moderate size, the problem becomes computationally expensive. This feature of the model restricts to a certain extent its applicability and states the problem of development of effective computational algorithms for system identification from data.

The efficiency of the model was demonstrated by the example of revealing the clumps of toxic cyanobacteria on the surface of the Baltic Sea. The reference image was used as a source of initial data; on its basis, the binary model was built. Then, the properties of the dynamics of the binary model were used for calculation of the index of evenness of the systemic colorimetric parameters. The index was applied to the processing of the snapshot with added digital noise that simulated unfavorable observation conditions. The resulting noised image of the snapshot contained details that were virtually invisible in this image before processing.

This model can be used for descriptive analysis of an incoming binary dataflow, when the researcher needs information on relationships that generate the data. For example, if the data come from channels with time displacement error, asynchrony, then the time order of the initial data may be disturbed and its recovery is needed.

The model can be expanded in several directions. The most obvious extension is the results' application to nominal data. The principle of parsimony underlying the identification of the model from the data can be applied in a number of different ways. In the paper, we used the simplest measure of dependence based on the arity of mappings that generate the model's dynamics. For example, if the process generating a binary dataflow essentially depends on the amount of computer memory, it is reasonable explicitly to take into account the sizes of series that form transitional mappings.

7. Conclusions

1. For the measure introduced during solving the stated problem, the true time order is calculated as a time order that minimizes this measure among all available time orders. Thus, the problem of recovery of the true order was solved.

2. The numerical measure, which enables to calculate the complexity of system dynamics for an arbitrary time order of observations (true or not), is introduced. This measure is based on the arity of the mappings that determine the dependence of the current value of each component on the preceding values of all components. Thus, the problem of restoring the law of system dynamics is solved.

3. The abilities of the dynamic binary model for selection of colorimetric parameters for identification of clumps of toxic cyanobacteria have been demonstrated with the use of the digital satellite image from the NASA site. The demonstration consisted of several steps:

– identification of the model from the reference image;

– derivation of the index that reflects the systemic properties of the dynamics;

– use of the index to process the image with the digital noise simulating unfavorable visibility conditions.

References

1. Zholtkevych, G. N. Discrete Modeling of Dynamics of Zooplankton Community at the Different Stages of an Antropogeneous Eutrophication [Text] / G. N. Zholtkevych, Y. G. Bespalov, K. V. Nosov, M. Abhishek. // Acta Biotheoretica. – 2013. – Vol. 61, Issue 4. – P. 449–465. doi: 10.1007/s10441-013-9184-6

2. Zholtkevych, G. Descriptive Models of System Dynamics [Text] / G. Zholtkevych, K. Nosov, Yu. Bespalov, E. Vysotskaya et. al. // Proceedings of the 12th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. – 2016. – Vol. 1614. – P. 57–72.

3. Carmichael, W. W. Health Effects of Toxin Producing Cyanobacteria: «The CyanoHABS» [Text] / W. W. Carmichael // Human and Ecological Risk Assessment: An International Journal. – 2001. – Vol. 7, Issue 5. – P. 1393–1407. doi: 10.1080/20018091095087

4. Zohary, T. Hyperscums and the population dynamics of Microcystis aeruginosa [Text] / T. Zohary, R. D. Roberts // Journal of Plankton Research. – 1990. – Vol. 12, Issue 2. – P. 423–432. doi: 10.1093/plankt/12.2.423

5. Karabashev, G. S. Spektral'nye priznaki cveteniya cianobakterij v Baltijskom more po dannym skanera MODIS [Text] / G. S. Karabashev, M. A. Evdoshenko // Sovremennye problemy distancionnogo zondirovaniya Zemli iz kosmosa. – 2015. – Vol. 12, Issue 3. – P. 158–170.

6. Vysotska, O. V. Using of margalef succession model in remote detection technologies for indications of human impact on vegetation cover [Text] / O. V. Vysotska, Yu. G. Bespalov, A. I. Pecherska, D. A. Parvadov // Radioelektronni i komp'uterni sistemi. – 2016. – Vol. 2, Issue 76. – P. 15–19.

7. Vysotskaya, E. V. Unmasking the soil cover's disruption by modeling the dynamics of ground vegetation parameters [Text] / E. V. Vysotskaya, G. N. Zholtkevych, T. A. Klochko, Yu. G. Bespalov, K. V. Nosov // Visnyk Natsionalnoho Tekhnichnoho Universytetu Ukrayiny «KPI». Seriya – Radiotekhnika. Radioaparatobuduvannya. – 2016. – Vol. 64. – P. 101–109.

8. Shulika, B. Control over grape yield in the North-Eastern region of Ukraine using mathematical modeling [Text] / B. Shulika, A. Porvan, O. Vysotska, A. Nekos, A. Zhemerov // Eastern-European Journal of Enterprise Technologies. – 2017. – Vol. 2, Issue 3 (86). – P. 51–59. doi: 10.15587/1729-4061.2017.97969

9. Fortunato, S. de Menezes. Data classification with binary response through the Boosting algorithm and logistic regression [Text] / F. S. De Menezes, G. R. Liska, M. A. Cirillo, M. J. F. Vivanco // Expert Systems with Applications. – 2017. – Vol. 69. – P. 62–73. doi: 10.1016/j.eswa.2016.08.014

10. Pierola, A. An ensemble of ordered logistic regression and random forest for child garment size matching [Text] / A. Pierola, I. Epifanio, S. Alemany // Computers & Industrial Engineering. – 2016. – Vol. 101. – P. 455–465. doi: 10.1016/j.cie.2016.10.013

11. Ghattas, B. Clustering nominal data using unsupervised binary decision trees: Comparisons with the state of the art methods [Text] / B. Ghattas, P. Michel, L. Boyer // Pattern Recognition. – 2017. – Vol. 67. – P. 177–185. doi: 10.1016/j.patcog.2017.01.031

12. Yamamoto, M. Clustering of multivariate binary data with dimension reduction via L1-regularized likelihood maximization [Text] / M. Yamamoto, K. Hayashi // Pattern Recognition. – 2015. – Vol. 48, Issue 12. – P. 3959–3968. doi: 10.1016/j.patcog.2015.05.026

13. Katahira, K. How hierarchical models improve point estimates of model parameters at the individual level [Text] / K. Katahira // Journal of Mathematical Psychology. – 2016. – Vol. 73. – P. 37–58. doi: 10.1016/j.jmp.2016.03.007

14. Erkan Ozkaya, H. An assessment of hierarchical linear modeling in international business, management, and marketing [Text] / H. Erkan Ozkaya, C. Dabas, K. Kolev, G. T. M. Hult, S. H. Dahlquist, S. A. Manjeshwar // International Business Review. – 2013. – Vol. 22, Issue 4. – P. 663–677. doi: 10.1016/j.ibusrev.2012.10.002

15. Van Oirbeek, R. Assessing the predictive ability of a multilevel binary regression model [Text] / R. Van Oirbeek, E. Lesaffre // Computational Statistics & Data Analysis. – 2012. – Vol. 56, Issue 6. – P. 1966–1980. doi: 10.1016/j.csda.2011.11.023

16. Asar, Ö. Flexible multivariate marginal models for analyzing multivariate longitudinal data, with applications in R [Text] / Ö. Asar, Ö. İlk // Computer Methods and Programs in Biomedicine. – 2014. – Vol. 115, Issue 3. – P. 135–146. doi: 10.1016/j.cmpb.2014.04.005

17. Zou, G. Y. Extension of the modified Poisson regression model to prospective studies with correlated binary data. [Text] / G. Y. Zou, A. Donner // Statistical methods in medical research. – 2013. – Vol. 22, Issue 6 – P. 661–670. doi: 10.1177/0962280211427759

18. Schildcrout, J. S. Marginalized models for moderate to long series of longitudinal binary response data [Text] / J. S. Schildcrout, P. J. Heagerty // Biometrics. – 2007. – Vol. 63, Issue 2. – P. 322–331. doi: 10.1111/j.1541-0420.2006.00680.x

19. Kupek, E. Beyond logistic regression: structural equations modelling for binary variables and its application to investigating unobserved confounders. [Text] / E. Kupek // BMC Medical Research Methodology. – 2006. – Vol. 6, Issue 1. doi: 10.1186/1471-2288-6-13

20. Rhemtulla, M. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions [Text] / M. Rhemtulla, P. É. Brosseau-Liard, V. Savalei // Psychological methods. – 2012. – Vol. 17, Issue 3. – P. 354–373. doi: 10.1037/a0029315

21. Pearl, J. Causal inference in statistics: An overview [Text] / J. Pearl // Statistics Surveys. – 2009. – Vol. 3. – P. 96–146. doi: 10.1214/09-ss057

22. Bollen, K. A. Eight myths about causality and structural equation models [Text] / K. A. Bollen, J. Pearl. –Handbooks of Sociology and Social Research, 2013. – P. 301–328. doi: 10.1007/978-94-007-6094-3_15

23. LANCE-MODIS Collection 6 [Electronic resource]. – Available at: https://lance3.modaps.eosdis.nasa.gov