Descriptive Models of System Dynamics

G.N. Zholtkevych[1]([✉]), K.V. Nosov[1], Yu.G. Bespalov[1], L.I. Rak[1],
E.V. Vysotskaya[2], Y.B. Balkova[1], and V.K. Kolomiychenko[2]

[1] V. N. Karazin Kharkiv National University, 4, Svobody Sqr.,
Kharkiv 61022, Ukraine
g.zholtkevych@gmail.com, k.nosov@karazin.ua
[2] Kharkiv National University of Radio-Electronics,
Science ave., 14, Kharkiv 61166, Ukraine

Abstract. Nowadays, while performing their investigations, researchers often face large datasets requiring fast processing for analysis and drawing adequate conclusions. Data mining, statistical methods and big data analytics provide an impressive arsenal of tools allowing scientists to address those tasks. However, in some cases investigators need techniques enabling on the base of relatively simple and cheap measurements of easily accessible parameters to build useful and meaningful concepts of an area of research.

In our paper two classes of dynamical models aimed at revealing between-component relationships in natural systems with feedback are presented. The idea of the both models follows from the frameworks of theoretical biology and ecology, where pairwise interactions between the parts of a system are regarded as background of the system's behavior. Both deterministic and stochastic cases are considered, that allow us to determine the direction of pairwise relationships in deterministic case and the direction and strength of relationships in stochastic one.

Keywords: Dynamical systems · Deterministic models · Stochastic models · Markov chains

1 Introduction

The world of scientific research has being immersing into an extraordinary information explosion over past decades, accompanied by the rapid growth in the use of Internet and the number of connected computers worldwide. We see a rate of increase in data growth that is faster than at any period throughout history. Enterprise application and machine-generated data continue to grow exponentially, challenging experts and researchers to develop new innovative techniques to evaluate hardware and software technologies and to develop new methods of big data study.

The problems raise during data acquisition and preliminary exploration, when the amount of data requires us to make decisions, often in an *ad hoc* manner, on importance and interpretability of data. Besides, much data today

are not natively in structured format, have gaps and are incomplete. Hence, data analysis, organization, retrieval, and modeling are foundational challenges. Finally, presentation of the results and its interpretation by non-technical domain experts is crucial to extracting actionable knowledge.

Our study is devoted to a well-known problem of revealing conditions of stability in natural systems providing long and steady development and existence of systems. Today, there is a large amount of online big data collections comprising datasets taken from different branches of biology, health sciences, ecology etc. As examples we can mention Data Centre of International Council for the Exploration of the Sea (includes hundreds of thousands marine biology related datasets) and PhysioNet at Massachusetts Institute of Technology, a huge collection of datasets of diverse physiologic signals and open-source software for study of such data.

The problem of homeostasis and stability in the living organisms community or natural systems is closely related to the problem of dynamic stability. The practical aspect of this problem is connected to the disturbance in stability of systems, that is often accompanied, for example, by outbreaks in number or biomass of species.

The study of stability in communities or natural systems is closely connected to investigation of relationships that determine the dynamic features of a system, i.e. relationships between system's parameters having influence on the system dynamics.

For decades systemic methods, for example, based on the Shannon index of diversity, have been used for studying the relationships between the structure and stability of systems. Generalizing many such approaches, Margalef [1] states that "the ecologist sees in any measure of diversity an expression of the possibilities of constructing feedback systems, or any sort of links, in a given assemblage of species". Similar ideas were therefore presented in studying the structure of correlation pleiads, cluster analysis and other statistical techniques to establish such relationships for investigating similar problems.

Despite different approaches to revealing between-component relationships, in biology and ecology there is a general approach for presenting such relationships on the base of the following pairwise relationships: $(+, +)$, $(-, -)$, $(-, +)$, $(-, 0)$, $(+, 0)$, $(0, 0)$. In such a way they usually denote pairwise, or paired, relationships between two components of a system. This means that two components interact each other according to the symbols presented in the corresponding relationship. E.g., $(-, +)$ means that the first component takes benefits by interacting with another, while the second suffers from the first. A quantitative measure of effects derived by the relationships are introduced in the corresponding sections. For the multi-component systems, this set of relationships exhausts all possible pairwise inter-component relationships categorized by the type of effect and have been thoroughly studied in biology and ecology [2–4]. Therefore, in the current paper the analysis of the relationships structure is based on the idea of regarding the objects (say, living organisms in a community) as the components of a system between which the mentioned pairwise relationships are possible.

This allows us to present the structure of relationships in an explicit form of relationships between the system's components.

It should be noted, that mentioned relationships cannot be always revealed with the help of statistical methods. For example, correlation analysis is initially used for estimation of a relationship between two variables, but it covers only statistical relation and cannot reveal a cause-effect relationship [5].

There are statistical methods (structural relation modeling, analysis of path and adjacent techniques), which are devoted to revealing between-component relationships (and other tasks as latent variables' analysis) and can be used for causality analysis [6–8, 10]. But these methods express the relationships of a system in the terms of regression coefficients and not in the form of paired relationships. So interpretation of results of an analysis is occasionally difficult (e.g. while studying the relationships between feedback system and homeostasis in a community) and requires additional assumptions.

The models suggested in the paper tend to express the component relations in an explicit, easy-to-understand form based on pairwise relationships. Besides, intra-component relations are allowed. The models except the structure of relationships also reveal the dynamics of the system, deterministic for one case and probabilistic in another, that enables to observe the changes of the system's states over time. These advantages determine topicality and importance of the study presented herein.

2   Theory

Below, we present two dynamical models developed for revealing between-component relationships on the base of observations of real natural system.

First model has a deterministic dynamic, finite number of states and discrete time. As it is described in [11] at length, here we describe the model in short.

The second model has stochastic nature and will be describe in the paper in details.

Both models have a common background, so we begin with its description and later will go to specific properties of each models.

We assume that a natural system to be modelled comprises N components, which can be denoted by $A_1, A_2, \ldots, A_N$. Each component has a nature intrinsic to the system, for example, the number of animals or amount of biomass of different species etc. It is assumed that values of each component are integer numbers 1, 2, \ldots, K, i.e. each component may be at K levels. The value 1 means a minimum amount of a component, the value K means maximum, i.e. a component value varies from 1 to K.

The system develops in discrete time and the moments of time are denoted $t = 0, 1, \ldots$. So, the value of the component A_i at the moment of time $t = 0, 1, \ldots$ are numbers $A_i(0)$, $A_i(1)$, \ldots.

Next properties of a system are different for deterministic and stochastic cases, so we shall describe them separately.

2.1   Deterministic Model Revealing the Direction
of Between-Component Relationships

We begin with deterministic case discussed, as mentioned, in [11] and was named the Discrete model of dynamical systems with feedback. For the deterministic system its state at the moment $t + 1$ is fully determined by the state at the moment t.

If the system at the moment t is in the state $A_1(0), A_2(0), \ldots, A_N(0)$, all the following states can be written as the trajectory, where each column is a state at a corresponding moment of time:

$$
\begin{pmatrix}
A_1(0) & A_1(1) & A_1(2) & \ldots \\
A_2(0) & A_2(1) & A_2(2) & \ldots \\
\vdots & \vdots & \vdots & \ldots \\
A_N(0) & A_N(1) & A_N(2) & \ldots
\end{pmatrix}.
\tag{1}
$$

In the theory of dynamical systems [12], such a system is called a free dynamical system with discrete time. The system has only finite number of states, so there exists a positive integer $\mathcal{T}$, called a period of the trajectory, for which the conditions of periodicity hold

$$
\begin{pmatrix}
A_1(s) \\
A_2(s) \\
\vdots \\
A_N(s)
\end{pmatrix}
=
\begin{pmatrix}
A_1(s + \mathcal{T}) \\
A_2(s + \mathcal{T}) \\
\vdots \\
A_N(s + \mathcal{T})
\end{pmatrix},
$$

for enough large s.

Taking into account the periodicity, we extract the minor

$$
\begin{pmatrix}
A_1(s) & A_1(s+1) & \ldots & A_1(s + \mathcal{T} - 1) \\
A_2(s) & A_2(s+1) & \ldots & A_2(s + \mathcal{T} - 1) \\
\vdots & \vdots & \ddots & \vdots \\
A_N(s) & A_N(s+1) & \ldots & A_N(s + \mathcal{T} - 1)
\end{pmatrix}
\tag{2}
$$

from (1) presenting full description of the system's dynamics.

Now we introduce the concept of relationships between components. Let $\Omega = \{-, 0, +\}$—three-entries set. A relationship between specified components A_i and A_j is determined as an entry from the set $\Omega \times \Omega$ and denoted by $\Lambda(A_i, A_j) = (\omega_1, \omega_2)$, where $\omega_1 \in \Omega$, $\omega_2 \in \Omega$. If $\Lambda(A_i, A_j) = (\omega_1, \omega_2)$, it means that:

– if $\omega_1 = \{-\}$, then increasing the value of A_j will determine the decrease of the value of A_i.
– if $\omega_1 = \{0\}$, then the A_j doesn't influence the value of the component A_i.
– if $\omega_1 = \{+\}$, then large values of the A_j will raise the value of the A_i.

The relationship Λ is antisymmetric in the following sense: $\Lambda(A_i, A_j) = (\omega_1, \omega_2)$ implies $\Lambda(A_j, A_i) = (\omega_2, \omega_1)$. It is also assumed, that inner relationships (self-relationship $\Lambda(A_i, A_i)$) are symmetric—$(0,0)$, $(-,-)$, and $(+,+)$ for any A_i.

Assume that all the relationships $\Lambda(A_j, A_i)$ between all pairs (A_j, A_i) of components $A_1, A_2, \ldots, A_N$ are given. For each A_j and each $(s, u) \in \Omega \times \Omega$, let $L_j(s, u) = \{A_i | \Lambda(A_j, A_i) = (s, u)\}$ (the set of components, with which A_j has the relationship (s, u)). We can express the relationships between the components by the following relationships' matrix

$$\begin{bmatrix} & A_1 & A_2 & \ldots & A_N \\ A_1 & (\omega_1, \omega_1) & & & \\ A_2 & (\omega_2, \omega_1) & (\omega_2, \omega_2) & & \\ \vdots & \vdots & \vdots & \ddots & \\ A_N & (\omega_N, \omega_1) & (\omega_N, \omega_2) & \ldots & (\omega_N, \omega_N) \end{bmatrix}. \tag{3}$$

Taking into account the antisymmetric property, the entries above main diagonal in (3) are omitted, since they can be recovered by the lower triangular part of matrix.

Let $\varkappa = \{1, 2, \ldots, K\}$ be the set of the states of the components and $N_j(s, u)$ is the number of components in the set $L_j(s, u)$, $j = 1, 2, \ldots, N$, $(s, u) \in \Omega \times \Omega$. A transition from the state $(A_1(t), A_2(t), \ldots, A_n(t))$ to the state $(A_1(t+1), A_2(t+1), \ldots, A_n(t+1))$ is described by N transition functions F_j, each of which defines the mapping

$$\varkappa^{N_j(+,+)+N_j(+,0)+N_j(+,-)+N_j(-,+)+N_j(-,0)+N_j(-,-)} \mapsto \varkappa.$$

This mapping symbolically is expressed by the formula

$$\begin{aligned} A_j(t+1) = F_j(&A_k(t) \in L_j(+,+), A_k(t) \in L_j(+,0), \\ &A_j(t) \in L_k(+,-), A_k(t) \in L_j(-,+), \\ &A_k(t) \in L_j(-,0), A_k(t) \in L_j(-,-)), \ j = 1, 2, \ldots, N, \end{aligned} \tag{4}$$

where $A_k(t) \in L_j(+,+)$, $A_k(t) \in L_j(+,0)$, $\ldots$ are the values of $A_k(t)$ belonging to $L_j(+,+)$, $L_j(+,0)$, $\ldots$ respectively.

The transition function, introduced by equation (4), is quite natural in its structure. Given component A_j is influenced only by those components, which indeed influence A_j, i.e. the components from the sets $L_j(+, \omega)$ and $L_j(-, \omega)$ for any $\omega \in \Omega$.

Two Types of Relationships, Intrinsic to Natural Systems. The formula (4) presents a general form of transition of the system from the state at the moment t to the state at $t+1$.

For more detailed description of the dynamics of a natural system, numerical experiments and procedures of system identification one needs to specify explicit form of the mappings.

We introduced two approaches based on the concepts of biological interactions: the weight functions' approach and the approach based on principles of Justus von Liebich's law.

Define the following functions on the set $\varkappa$: $\mathrm{Inc}(A) = \min\{K, A+1\}$, $\mathrm{Dec}(A) = \max\{1, A-1\}$.

The system dynamics according to the weight functions' approach. First we define the type of dynamics, which takes into account the weighted sum of all $A_j(t)$ (including $A_i(t)$) for calculating the value of the component A_i at the moment $t+1$.

As we defined above, for each j ($j = 1, 2, \ldots, N$) and each pair $(s, u) \in \Omega \times \Omega$ there exists the set $L_j(s, u)$ with $N_j(s, u)$ entries. Assume that $\varphi_{j,1}^{\langle s,u \rangle}(\cdot)$, $\varphi_{j,2}^{\langle s,u \rangle}(\cdot)$, $\ldots$, $\varphi_{j,N_j(s,u)}^{\langle s,u \rangle}(\cdot)$ are to be the functions of interactions of those components, with which the A_j has relationships (s, u). The functions are defined on the discrete set $\varkappa$ and have the following properties:

1. $\varphi_{j,k}^{\langle +,+ \rangle}(\cdot)$, $\varphi_{j,k}^{\langle +,0 \rangle}(\cdot)$, $\varphi_{j,k}^{\langle +,- \rangle}(\cdot)$ are increasing functions.
2. $\varphi_{j,k}^{\langle -,+ \rangle}(\cdot)$, $\varphi_{j,k}^{\langle -,0 \rangle}(\cdot)$, $\varphi_{j,k}^{\langle -,- \rangle}(\cdot)$ are decreasing functions.
3. $\varphi_{j,k}^{\langle s,u \rangle}(1) = 0$ for any $(s, u) \in \Omega \times \Omega$.

We also introduce the numbers $\delta_j > 0$ ($j = 1, 2, \ldots, N$) which can be called thresholds of sensitivity.

For the system's state at the moment t, the following value is calculated

$$
\begin{aligned}
d_j = & \sum_{A_k \in L_j(+,+)} \varphi_{j,k}^{\langle +,+ \rangle}(A_k(t)) + \sum_{A_k \in L_j(+,0)} \varphi_{j,k}^{\langle +,0 \rangle}(A_k(t)) \\
& + \sum_{A_k \in L_j(+,-)} \varphi_{j,k}^{\langle +,- \rangle}(A_k(t)) + \sum_{A_k \in L_j(-,+)} \varphi_{j,k}^{\langle -,+ \rangle}(A_k(t)) \\
& + \sum_{A_k \in L_j(-,0)} \varphi_{j,k}^{\langle -,0 \rangle}(A_k(t)) + \sum_{A_k \in L_j(-,-)} \varphi_{j,k}^{\langle -,- \rangle}(A_k(t)).
\end{aligned} \tag{5}
$$

The value of the component $A_j(t+1)$ is calculated as follows

1. if $d_j \geq \delta_j$, then $A_j(t+1) = \mathrm{Inc}(A_j(t))$;
2. if $d_j \leq -\delta_j$, then $A_j(t+1) = \mathrm{Dec}(A_j(t))$;
3. if $-\delta_j < d_j < \delta_j$, then $A_j(t+1) = A_j(t)$.

Now, the meaning of introduced transition functions can be explained in clear way. For example, the functions $\varphi_{j,k}^{\langle -,+ \rangle}(\cdot)$ ($k = 1, 2, \ldots, N_j(-, +)$) reflects the influence upon the component A_j by components in the set $L_j(-, +)$, which are related with A_j by relationship $(-, +)$. The greater the influence (i.e. the greater values of $A_i(t)$ from the set $L_j(-, +)$), the lower the values of d_j.

The influence of other components, with which A_j has other relationships, is "weighted" in similar way. If the cumulative influence of components, interacted with A_j and expressed by Eq. (5), exceeds the threshold δ_j, then the value of A_j is changed by unit.

The threshold δ_j clearly influenced the dynamics of the system in the following way: the greater δ_j, the greater absolute value of the weighted sum d_j required for overcoming this δ_j for changing the value of A_j. So if δ_j is very large, the system becomes very inert.

The dynamics based on the Liebig's law of the minimum. Next approach is based on principles of Justus von Liebich's law (Liebig's law of the minimum) and essentially differs from first approach, which is basically additive.

Assume that the system of relationships between A_1, A_2, ..., A_N is given. For defining the system's dynamics, let's introduce two constant matrices, C and C^* of size $N \times N$. The transition function is based on the following algorithm.

Suppose the system in state $(A_1(t), A_2(t), \ldots, A_N(t))$ at time t and A_j is an arbitrary fixed component. Let i runs from 1 to N, and u means arbitrary entry of the set Ω.

1. If for the current i the equality $\Lambda(A_j, A_i) = (-, u)$ holds true, assume that

$$f_i = \begin{cases} -1, & \text{if } A_i(t) \geq c^*_{ji}, \\ 0, & \text{if } c_{ji} + 1 \leq A_j(t) \leq c^*_{ji} - 1, \\ 1, & \text{if } A_j(t) \leq c_{ji}. \end{cases}$$

The specific value of u doesn't matter because only the influence on A_i from the side of A_j matters.

2. If for the current i the equality $\Lambda(A_j, A_i) = (+, u)$ holds true, assume

$$f_i = \begin{cases} -1, & \text{if } A_j(t) \leq c_{ji}, \\ 0, & \text{if } c_{ji} + 1 \leq A_j(t) \leq c^*_{ji} - 1, \\ 1, & \text{if } A_j(t) \geq c^*_{ji}. \end{cases}$$

3. If for the current i the equality $\Lambda(A_j, A_i) = (0, u)$ holds true then it is assumed that $f_i = 0$.

After the cycle termination, the sequence f_1, f_2, ..., f_N is obtained. The value $A_j(t+1)$ is calculated according to the following rule:

$$A_j(t+1) = \begin{cases} \text{Dec}(A_j(t)), & \text{if } \min_{1 \leq i \leq N} f_i = -1, \\ A_j(t), & \text{if } \min_{1 \leq i \leq N} f_i = 0, \\ \text{Inc}(A_j(t)), & \text{if } \min_{1 \leq i \leq N} f_i = 1. \end{cases} \tag{6}$$

Applying this algorithm for each $j = 1, 2, \ldots, N$, the system's state at the moment $t+1$ is calculated.

The meaning of transition from t to $t+1$ can be explained clearly. Suppose, a given component A_j has relationship $(+, -)$ with this current component A_i (see the algorithm). According to the relationship $(+, -)$, large values of A_i should decrease A_j. Indeed, according to item 1 of the algorithm, if $A_i(t) \geq c^*_{ji}$ (in other words, when $A_i(t)$ is "large enough"), $f_i = -1$ and, according to (6), A_j would decrease if $A_j(t) > 1$. Other cases of transition work in similar way.

The System Identification Based on the Data of Observation. When we deal with real data, as a rule, we don't observe their dynamics explicitly. Often real data are unordered in time in contrast to data used for time series modeling. So we don't observe any dynamism described by the relationship (3), by the trajectory (1) or the minor (2).

Usually, the result of observation is represented by the table:

$$\tilde{M} = \begin{pmatrix} C_{11} & C_{12} & \ldots & C_{1B} \\ C_{21} & C_{22} & \ldots & C_{2B} \\ \vdots & \vdots & \ddots & \vdots \\ C_{N1} & C_{N2} & \ldots & C_{NB} \end{pmatrix}, \tag{7}$$

where columns correspond to cases and rows correspond to components (N components, B cases). We emphasize unordered character of the data above, i.e. these is no time order between the cases in the table $\tilde{M}$.

Here we describe a principle allowing to reveal the system relationships of above mentioned type on the basis of the observation table $\tilde{M}$.

This algorithm determines inter- and intra-component relationships, which are as close as possible to relationships, which form matrix (2) in some sense.

Assume that the relationships structure is given. In that case for the initial state $(A_1(0), A_2(0), \ldots, A_N(0))$ and for the given sets $L_1(u, s)$, $L_2(u, s)$, $\ldots$, $L_N(u, s)$, $u \in \Omega$, $s \in \Omega$ the minor (2) can be calculated. Let

$$P = \begin{pmatrix} 1 & r_{12} & \ldots & r_{1N} \\ r_{12} & 1 & \ldots & r_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1N} & r_{2N} & \ldots & 1 \end{pmatrix}$$

is to be the correlation matrix (Pearson or Spearman) between the rows of the minor (2). Also, for the table $\tilde{M}$, the correlation matrix of its rows can be calculated:

$$\tilde{P} = \begin{pmatrix} 1 & \rho_{12} & \ldots & \rho_{1N} \\ \rho_{12} & 1 & \ldots & \rho_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1N} & \rho_{2N} & \ldots & 1 \end{pmatrix}.$$

Introduce the measure of distance between the correlation matrices P and $\tilde{P}$

$$D(P, \tilde{P}) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (r_{ij} - \rho_{ij})^2. \tag{8}$$

Set the problem of minimization $D(P, \tilde{P})$ by all possible vectors of initial states $(A_1(0), A_2(0), \ldots, A_N(0))$ and all allowable sets $L_j(s, u)$, $s \in \Omega$, $u \in \Omega$ for all j

$$D(P, \tilde{P}) \mapsto \min \quad \begin{array}{l} \text{(by all initial states \&} \\ \text{by all allowable sets } L_j(s, u)). \end{array} \tag{9}$$

The meaning of this problem can be explained in the following way. Suppose, a process in some natural system is cyclical with the trajectory (2). There is no possibility to observe the dynamics of this trajectory, i.e. a full length cycle. The observations are taken from the system at random moments of time t from s

to $s + \mathcal{T} - 1$ with equal probability. When an observation is taken, the column $(A_1(t), A_2(t), \ldots, A_N(t))^T$ from (2) is attached to the table of observations. In other words, the columns of table of observations M are obtained from (2) by equiprobable choice of columns.

The stated problem means the search for such relationships between components, that the minor (2) is to be as close as possible to the table of observations regarding the measure (8).

The following theorem proved in [11] shows that this problem is well-grounded in probabilistic sense.

Theorem 1. *If the table of observations $\tilde{M}$ is obtained from the minor (2) by equiprobable choice of columns, then the Pearson correlation matrix of the observations table $\tilde{P}$ converges to the correlation matrix of minor P (in probability)*

$$\lim_{B \to \infty} \rho_{ij} = r_{ij}, \quad i = 1, 2, \ldots, N, \ j = 1, 2, \ldots, N.$$

The same result takes place for the Spearman correlation matrix as well.

2.2   Additive Stochastic Model of Between-Component Relationships

Our another model is also described by the set of components $A_1, A_2, \ldots, A_N$ taking discrete values 1, 2, ..., K.

But, in contrast to the first one, the second model introduces into consideration not only direction of relationships (in fact, for the first model we considered three direction—negative, neutral, and positive), but also a strength of relationships.

The structure of relationships between the components $A_1, A_2, \ldots, A_N$ is described by the following relationships matrix

$$\mathcal{M} = \begin{pmatrix} m_{1,1} & m_{1,2} & \ldots & m_{1,N} \\ m_{2,1} & m_{2,2} & \ldots & m_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ m_{N,1} & m_{N,2} & \ldots & m_{N,N} \end{pmatrix}.$$

Any entry $m_{i,j}$ reflects the strength and direction of influence of the component A_j upon the component A_i. The direction of influence is expressed by the sign of the value $m_{i,j}$ (may be $-, 0, +$) and the strength—by modulus of $m_{i,j}$ and varies from 0 to 1. So, $-1 \leq m_{i,j} \leq 1$ for each i, j. The influence of the component A_i on A_j is expressed by $m_{j,i}$. It is easy to see, that the relationship between the components A_i on A_j is described by the pair $(m_{i,j}, m_{j,i})$, which is in some way similar to the relationship (ω_1, ω_2) introduced for the first model.

Now describe the dynamics of transition from the state of the system at the moment t to the state at the next moment $t + 1$, i.e.

$$(A_1(t), A_2(t), \ldots, A_N(t)) \mapsto (A_1(t+1), A_2(t+1), \ldots, A_N(t+1)).$$

As for the weight functions' approach, we assume, that the set of functions $\psi_{i,j}(\cdot)$, $(i, j = 1, 2, \ldots, N)$ reflecting relationships between all pairs of components, including inner relationships, are given.

All the functions $\psi_{i,j}(\cdot)$ have the following properties:

1. $\psi_{i,j}(\cdot)$ are defined on the set $\varkappa$;
2. $\psi_{i,j}(1) > 0$;
3. $\psi_{i,j}(\cdot)$ are increasing functions on $\varkappa$.

It should be noted that the property 2 of functions $\psi_{i,j}(\cdot)$ is different from corresponding property for functions $\psi_{i,j}^{\langle s,u \rangle}(\cdot)$, which requires $\psi_{i,j}^{\langle s,u \rangle}(1) = 1$.

Also assume that a positive number δ playing the a role of threshold, is given.

Let the system be in the state $(A_1(t), A_2(t), \ldots, A_N(t))$. For each pair of indices i, j ($i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, N$) define the random variable $\xi_{i,j}$ as follows

$$\xi_{i,j} = \begin{cases} \psi_{i,j}(A_j(t))\text{sign}(m_{i,j}) & \text{with probability } |m_{i,j}| \\ 0 & \text{with probability } 1 - |m_{i,j}|. \end{cases}$$

Then we introduce the set of N random variables

$$d_i = \sum_{j=1}^{N} \xi_{i,j}, \ i = 1, 2, \ldots, N.$$

Using the set $(d_1, d_2, \ldots, d_N)$, it's possible to calculate the set of probabilities (p_i^-, p_i^0, p_i^+) for each i as follows

$$\begin{aligned} p_i^+ &= P(d_i \geq \delta), \\ p_i^0 &= P(-\delta \leq d_i \leq \delta), \\ p_i^- &= P(d_i \leq -\delta), \end{aligned}$$

for each i from 1 to N. This definition implies $p_i^- + p_i^0 + p_i^+ = 1$.

For each i, the transition from the state at the moment t to the state at $t+1$ is defined by

$$A_i(t+1) = \begin{cases} \text{Dec}(A_i(t)) & \text{with probability } p_i^-, \\ A_i(t) & \text{with probability } p_i^0, \\ \text{Inc}(A_i(t)) & \text{with probability } p_i^+. \end{cases}$$

That is, at the moment $t+1$ the value of A_i can be increased by 1, remains the same or decreased by 1 with probabilities p_i^-, p_i^0, p_i^+ correspondingly.

Applying this rule for each i, the probabilities of transition from any appropriate state $(A_1(t), A_2(t), \ldots, A_N(t))$ can be calculated.

It can be proved, that if each row of the matrix $\mathcal{M}$ include both negative and positive entries, we obtain the Markov chain with K^N states $A_1(t), A_2(t), \ldots, A_N(t)$ $(A_i \in \varkappa)$. Besides, this chain is regular, so there a unique steady-state stochastic vector $\mathbf{w}$.

Now the reasons for the problem can be explained. We assume, that a natural system is described by this model, and the probability of being of the system in states converges to the entries of the vector $\mathbf{w}$ (as the system acts as a regular Markov chain). Using the states $A_1 A_2, \ldots, A_N$ and the steady-state vector $\mathbf{w}$, we can calculate a weighted Pearson correlation matrix [13] between the components.

Describe this step at length. All states of the system can be written in the table

$$
\begin{bmatrix}
A_1 & A_2 & \ldots & A_{N-2} & A_{N-1} & A_N \\
1 & 1 & \ldots & 1 & 1 & 1 \\
1 & 1 & \ldots & 1 & 1 & 2 \\
1 & 1 & \ldots & 1 & 2 & 1 \\
1 & 1 & \ldots & 1 & 2 & 2 \\
1 & 1 & \ldots & 2 & 1 & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
K & K & \ldots & K & K & K
\end{bmatrix}
\tag{10}
$$

having K^N rows and N columns.

Write out the steady-state vector of the Markov chain in the form

$$
\mathbf{w} = (w_1, w_2, \ldots, w_{K^N}),
$$

where the entry w_k corresponds to k-th state in the table (10).

Taking $\mathbf{w}$ as weights, calculate the weighted Pearson correlation matrix between the columns of the table (10). Denote such the matrix by $R_{\mathbf{w}}$.

We suppose, that the explicit dynamics of our natural system is not available. In other words, we cannot observe time series of states, but can record a state of the system at random moments of time. These observations are collected in the observation table $\tilde{M}$ having N variables and B cases (after B observations). Let the Pearson correlation matrix between rows of (7) is denoted by $\tilde{R}$.

Theorem 2. *If the observation table $\tilde{M}$ is obtained according to the way described above, then*

$$
\tilde{R} \to R_{\mathbf{w}}(\text{in probability if } B \to \infty),
$$

where $\tilde{R} \to R_{\mathbf{w}}$ means entry-wise convergence.

Proof. Omitted for brevity sake.

Introduce the measure of proximity for the matrices R and $\tilde{R}$

$$
D(R_{\mathbf{w}}, \tilde{R}) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (\tilde{R}_{i,j} - [R_{\mathbf{w}}]_{i,j})^2.
\tag{11}
$$

The Theorem 2 means that the sample observation matrix consistently represents a true dynamics, which is not observed explicitly. This result is used as a base for identifications of entries of the relationships matrix $\mathcal{M}$. In any investigations there is only a finite number of observations (B is finite). Therefore we

can try to calculate transition probabilities of the Markov chain, that provide the best approximation of a true correlation matrix by a sample matrix in the sense of the measure (11).

According to this approach, $\mathcal{M}$ is obtained by resolving the following optimization problem

$$D(R_{\mathbf{w}}, \tilde{R}) \mapsto \min \text{ by entries } m_{i,j}.$$

In fact, we find the relationships matrix $\mathcal{M}$, which makes the modelled correlation matrix as close as possible to the observe correlation matrix.

3   Case Studies

In this section we present three examples of application of developed models to various natural and technical systems.

3.1   Factors Determining Users Activity in Social Networks

First example concerns analysis of system factors affecting activity of users of social networks playing an important role in modern culture [14,15]. The structure of relationships between the components of the system for two states of the Internet-forum on fantasy literature were calculated and compared. This comparison aimed at reveal system aspects of forum visiting in two periods. One state can be regarded as "low-performance", other as "high-performance" according to number of written fanfictions (also abbreviated as fan fics, fanfics) of visitors at the site dedicated to the cycle of novels of Joanne Rowling about Harry Potter (snapetales.com). The period of first half of December 2010 is regarded as "high-performance", the second period of the first half of December 2014 is called "low-performance". For these two periods a statistically significant difference according to Student t-test ($p < 0.05$) in overall average number of visits per day was also detected.

The fanfictions were divided into 4 categories according to their length—fanfictions of small, large, and medium size; the last, fourth category includes fanfictions not related to the novels about Harry Potter.

The following values were taken as the components of the system reflecting the authors activity

- the number of *small size* fanfictions per day related to the cycle of novels about Harry Potter (denoted by **MIN**);
- the number of *large size* fanfictions per day (**MAX**);
- the number of *medium size* fanfictions (**MID**);
- the number of fanfictions not related to the cycle of novels about Harry Potter, based on another literary works (**OTHER**).

For the "high-performance" and "low-performance" periods, the structure of relationships were built. We identified the models using the Pearson correlation matrix and the approach on the base of von Liebig law, with $K = 3$ levels of

components values. The structure of relationships for both period is presented in Figs. 1 and 2 correspondingly. The graphs in the figures present the structure of relationships. Nodes of the graphs correspond to the components in the systems as defined above. Edges of the graphs with embedded ovals on them present pairwise relationships revealed in the model. For example, $\Lambda(\mathbf{MIN}, \mathbf{MID}) = (+, -)$, that is presented on the graph by the corresponding oval.
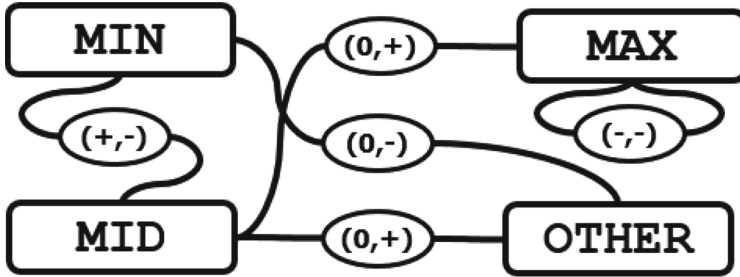

Fig. 1. The structure of relationships for "high-performance" period. Rounded rectangulars present the components of the system, the ovals present the relationships between the components.
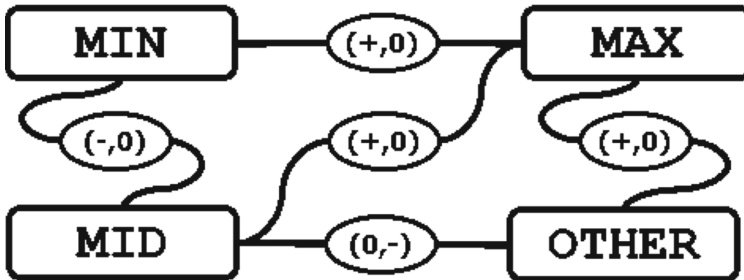

Fig. 2. The structure of relationships for "low-performance" period.

Comparing the graphs in Figs. 1 and 2 shows a system-forming role of the component **MID** for the "high-performance" period, in which **MID** positively affected the other three components. This affect disappeared in the "low-performance" period together with loss a stabilizing mechanism through the relationship $(+, -)$ between **MID** and **MIN** supporting a dynamic equilibrium of the system.

These results are consistent with empirically established ideas about significant positive role of fanfictions of medium size (**MID**) in a functioning of social networks of this category and their close relation to short-sized fanfictions

(**MIN**) representing a reaction of the most dynamic part of users. Differences in role of **OTHER** correspond to significance of "offtopic" as an index of deterioration in work of dedicated web-sites.

3.2   Relations Between the Anthropometric Parameters of Adolescents with Cardiovascular Disorders

Our next case concerns the system relations of anthropometric parameters of adolescents suffering diseases of cardiovascular system.

Anthropometry is important in school health care, in particular, for determining the factors of predisposition of adolescents to cardiovascular disorders. At the same time, among other drawbacks of currently used anthropometric methods they often refer to insufficient use of systematic approach, among other things, in description of regularities in formation of body's proportions in the individual development of adolescents.

Here we present a demo of the application of the deterministic model developed above for this purpose, built on the material of adolescents anthropometry with arterial hypertension and other forms of cardiovascular disorders. Body compositions related to overweight plays an important role in development of arterial hypertension. Taking that in account, the models for four components were built: hip circumference, waist circumference, chest circumference, and shoulder breadth divided by height of a subject. The Spearman correlation and Liebig's approach with $K = 3$ levels of components were used in modeling.

Comparison of these graphs revealed a different role of such anthropometric parameters as the hip circumference for the two group of adolescents under investigation. In the group with disorders different from arterial hypertension high values of hip circumference increase other three components. Simultaneously, shoulder breadth negatively affects hip circumference, that should form a proportion of male's future body perceiving by subconscious as harmonious on the base on evolutionary history and recognized as such by modern physiology and medicine—the proportions of male "triangle" directed beneath by edge. The
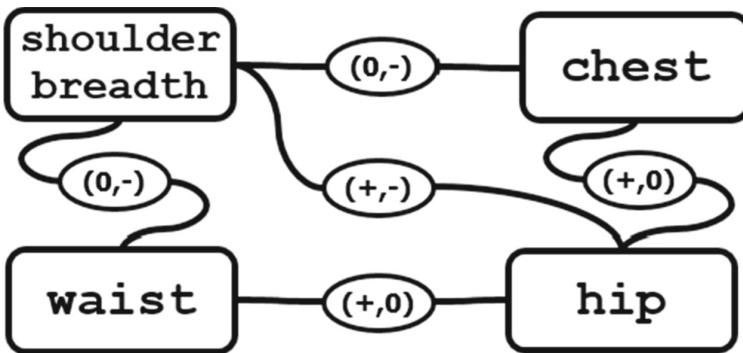

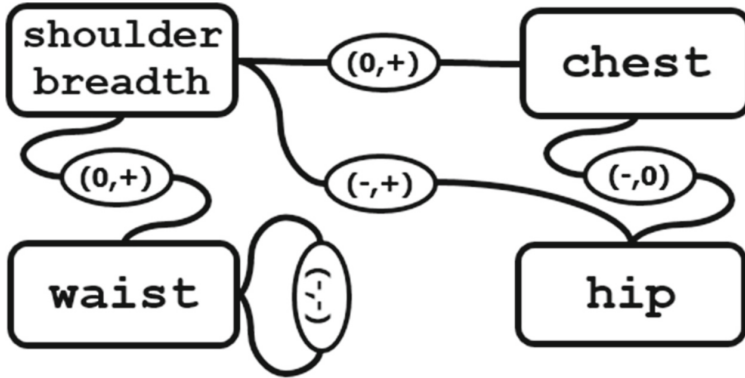**Fig. 3.** The structure of relationships for adolescents arterial hypertension.

Fig. 4. The structure of relationships for adolescents without arterial hypertension.

structure of relationships in the group with hypertension prevents the formation of such a standard and associates with the accumulation of a depot fat in certain parts of a human body: relatively high values of the hip circumference negatively affects shoulders breadth and chest circumference, not directly affecting waist circumference, on which shoulders breadth positively influences.

These results, regarded by authors as preliminary, do not contradict known facts about the impact of anthropometric parameters on the risk of development of hypertension in adolescents groups.

3.3 Factors Influencing the Efficiency of Industrial Fishery at North Sea

Our last case concerns the issues on industrial fishery of Atlantic cod (*Gadus morhua*) at North Sea. The fishery of the cod plays an important role in the economy of several countries and provokes considerable interest to using mathematical models in industrial ichthyology describing large fluctuations of catching [18] (well-known example of this kind is collapse of the Atlantic northwest cod fishery in 1992).

As the demo the additive stochastic model of relationships structure between dimensional parameters of cod populations was considered. The average fish body length ($\mathbf{L}$), the difference between Upper Length Bound and Lower Length Bound ($\mathbf{vL}$), the average stomachs weight ($\mathbf{M}$), and the average weight of preys of cod ($\mathbf{dM}$) were taken as components of the model. Additive stochastic models were built according to data of International Council for the Exploration of the Sea for two years (1984 and 1989) preceding to rapid changes of CPUE (the catch per unit effort). We used the model with $K = 4$ levels of components values. The results are shown in Table 1.

In the matrix corresponding to 1984, which precedes significant (till 1990) decrease of catching yield, there are large (above 0.85) negative effects of high values of $\mathbf{vL}$ on $\mathbf{M}$ and $\mathbf{dM}$. That is, increasing the diversity of dimensional

Table 1. Structure of relationships between dimensional parameters **L**, **vL**, **M**, and **dM** of the model for years 1984 and 1989.

1984					1989				
	L	vL	M	dM		L	vL	M	dM
L	−0.936	−0.379	0.893	0.581	**L**	−0.843	0.137	0.953	0.059
vL	0.325	−0.737	−0.405	0.519	**vL**	0.963	−0.721	−0.997	0.494
M	−0.184	−0.868	0.969	−0.081	**M**	−0.882	−0.054	0.788	0.224
dM	0.016	−0.941	0.999	−0.028	**dM**	0.091	0.066	0.941	−0.995

characteristics of the cod population, that improves the consumption possibilities of forage reserve by the cod, leads to exhaustion of food resources (reducing the number of available preys) and deterioration of preys quality (reducing the average size of forage organisms), and results in deterioration of food supply of the cod, that lowers the values of **M** and **dM**.

In the matrix corresponding to 1989, which precedes sharp increase of CPUE, recorded a year later, in 1991, there exist extremely small (below 0.07) negative effects of high values of **vL** on **M** and **dM**. In this case, the increasing diversity of sizes, that enhances abilities of consumption of forage reserve, does not lead to exhaustion and deterioration of the latter. This result of modeling explains differences described above in the dynamics of catching in accordance with modern concepts of industrial ichthyology.

Presented results bring hope for the possibility of developing methods of forecasting the cod catching with use of the stochastic models, built on the base of actual material on the size structure of the population. The data of material can be obtained among others by remote methods with the help of low-cost means and relatively little effort, and even from commercial reports.

4   Conclusion

In the paper we followed the established framework in model development, appropriated for natural sciences. Typical approach in development, among others, comprises the data selection, specification of assumptions and simplifications, selection of a mathematical modeling framework, estimation of parameter values, model diagnostics, model validation, model refinements and model application. It's clear, that all these stages of building mathematical models for biological systems are too complicated, but the most difficult task among them is the model parameters' estimation for identifying structure in the underlying biological networks.

The models presented in the paper are created for description of biological and ecological systems, based on pairwise relationships characterized by the direction (positive, negative, or neutral) for the both models and by strength varied from 0 to 1 in the stochastic model only.

The problem of parameters estimation is a true challenging problem for both models and requires development of special algorithms of numerical optimization. For example, if the system has N components and the number of levels is to be assumed K, for the first deterministic model the number of initial states is equal to K^N and the number of possible relationships' structures is equal to 3^{N^2}. For solving the stated optimization problem (9), one should built the minor (2) with use of an initial state and a relationships' structure, calculate correlation matrix P and calculate the distance (8). So, the exhaustive search of both initial states and relationships' structures in total gives us $K^N 3^{N^2}$ variants, that is a huge number for even moderate N and K.

The case studies presented in the paper, considered by the authors as preliminary and illustrating, offer the prospects of applications of proposed models.

The results of modeling of system aspects in anthropometry of adolescents present the approaches to use of this simple and cheap method for identifying the risk groups of the progress of arterial hypertension. These approaches may be applied in school medicine and, if necessary, in extreme situations for mass screening as well.

The study of system factors of performance of the web-site dedicated to fiction about characters from original works about Harry Potter, due to use of components of the system, that are invariant to the content of the web-site, may have a broader meaning in analysis of the social networks performance.

The model of the cod population as a whole does not contradict known facts on the role of fish size and state of a forage reserve in the population dynamics. At the same time, these results reveal some promises and can be used in the development of approximate methods for prediction of populations of commercial fish with use of relatively simple and inexpensive methods of data acquisition, even with use of commercial reports concerning the assortment of fish products.

References

1. Margalef, R.: Perspectives in Ecological Theory. Chicago University Press, Chicago (1968)
2. Lidicker Jr., W.Z.: A clarification of interactions in ecological systems. Bioscience **29**(8), 475–477 (1979)
3. Abrams, P.A.: On classifying interactions between populations. Oecologia **73**(2), 272–281 (1987)
4. Wootton, J.T., Emmerson, M.: Measurement of interaction strength in nature. Ann. Rev. Ecol. Evol. Syst. **36**, 419–444 (2005)
5. Shipley, B.: Cause and Correlation in Biology. Cambridge University Press, Cambridge (2004)
6. Spirtes, P., Glymour, C.N., Scheines, R.: Causation, Prediction, and Search. MIT Press, Cambridge (2000)
7. Pearl, J.: Causality: Models, Reasoning, and Inference. Cambridge University Press, Cambridge (2009)
8. Bollen, K.A., Pearl, J.: Eight myths about causality and structural equation models. In: Morgan, S.L. (ed.) Handbook of Causal Analysis for Social Research, pp. 301–328. Springer, Dordrecht (2013). https://doi.org/10.1007/978-94-007-6094-3_15

9. Elwert, F.: Graphical causal models. In: Morgan, S.L. (ed.) Handbook of Causal Analysis for Social Research, pp. 245–273. Springer, Dordrecht (2013). https://doi.org/10.1007/978-94-007-6094-3_13

10. Menzies, P.: Causal models, token causation, and processes. Philos. Sci. **71**(5), 820–832 (2004)

11. Zholtkevych, G.N., Bespalov, Y.G., Nosov, K.V., Abhishek, M.: Discrete modeling of dynamics of zooplankton community at the different stages of an antropogeneous eutrophication. Acta. Biotheor. **61**(4), 449–465 (2013)

12. Kalman, R.E., Falb, P.L., Arbib, M.A.: Topics in Mathematical System Theory. McGraw-Hill, New York (1969)

13. Blalock, H.M.: Social Statistics. McGraw-Hill, New York (1972)

14. Boyd, D., Ellison, N.: Social network sites: definition, history, and scholarsh. J. Comput. Mediated Commun. **13**(1), 210–230 (2007)

15. Di Caprio, D., Santos-Arteaga, F.J., Tavana, M.: Self-organizing social networks by preference similarity and the networking capacity of their users. Procedia Comput. Sci. **83**, 496–503 (2016)

16. Chen, M., Mao, S., Liu, Y.: Big data: a survey. Mobile. Netw. Appl. **19**, 171–209 (2014)

17. Cole, T.J., Bellizzi, M.C., Flegal, K.M., Dietz, W.H.: Establishing a standard definition for child overweight and obesity worldwide: international survey. BMJ Clin. Res. **320**, 1–6 (2000)

18. ICES: Report of the Benchmark Workshop on Arctic Stocks (WKARCT), 26–30 January 2015. ICES Headquarters, Denmark. ICES CM 2015/ACOM:30 (2015)