Forecasting indicators of territorial entity based on phenomenological models of collective behavior

Indicadores de previsión de la entidad territorial basados en modelos fenomenológicos de comportamiento colectivo

Abstract
The article is aimed at solution of the urgent scientific task of creating new methods of forecasting analytics that allow obtaining more accurate and reliable forecasts of the socioeconomic development of countries, regions and cities. In contrast to traditional methods it is achieved by simultaneously taking into account both the dynamic patterns of changes in the states of individual objects, and the statistical patterns of their collective behavior. It is shown that such forecasting methods can be based on the use of phenomenological approaches and techniques in relation to the description of statistical data and states of objects in multidimensional spaces of social and economic indicators. This scientific direction refers to the current section of system dynamics and sociophysics and allows to formulate a fundamentally new approach to modeling and forecasting the development of territorial entities, based on the application of the concept of continual fields of empirical measures and various models of object state spaces. On the basis of considering statistical distributions of joint events observation of a set of indicators there have been proposed dependencies for predicting the states of objects. As an implementation of the developed method, forecasting of some indicators of the development of cities of Russia was carried out.

Key words: countries, regions and cities, forecasting indicators, state spaces, models of collective behavior, phenomenological methods.

Resumen
El artículo se ocupa de la posibilidad de que el grupo de expertos científico de la creación de nuevos métodos de predicción de los análisis que permita obtener una mayor exactitud y una buena prevision del desarrollo socioeconómico de países, regiones y ciudades. En contraste con los métodos tradicionales, se realiza mediante la realización de una cuenta con los cambios dinámicos en los estados de los objetos individuales y los patrones de sus comportamientos. Se muestra que los métodos de predicción se pueden basar en el uso de fenómenos tropicales y técnicos relacionados con la descripción de estadísticas estadísticas y estados de los objetos en los niveles multidimensional de los criterios sociales y económicos. Esta técnica de dirección se refiere a la sección actual del sistema dinámico y sociofísica y permite formular un nuevo enfoque básico para modelar y pronosticar el desarrollo de territorios territoriales basados en la aplicación del concepto de los contingentes de medidas de medida y varios modelos de estado espacio. El método de pronóstico de algunos indicadores del desarrollo de las ciudades de Rusia fue realizado.

Keywords: países, regiones y ciudades, indicadores de pronóstico, estados, modelos de comportamiento del comportamiento, fenómenos fenomenológicos.
Resumo

O artigo tem como objetivo a solução da urgente tarefa científica de criar novos métodos de previsão de análises que permitam previsões mais precisas e confiáveis do desenvolvimento socioeconômico de países, regiões e cidades. Em contraste com os métodos tradicionais, ele é alcançado levando em conta simultaneamente os padrões dinâmicos de mudanças nos estados dos objetos individuais e os padrões estatísticos de seu comportamento coletivo. Mostra-se que tais métodos de previsão podem ser baseados no uso de abordagens e técnicas fenomenológicas em relação à descrição de dados estatísticos e estados de objetos em espaços multidimensionais de indicadores sociais e econômicos. Esta direção científica refere-se à secção actual da dinâmica e sociophysics sistema e permite a formulação de uma abordagem fundamentalmente nova para modelar e prever o desenvolvimento de entidades territoriais, com base na aplicação do conceito de campos contínuos de medidas empíricas e vários modelos de estado de objeto espaços. Com base nas distribuições estatísticas da observação de eventos conjuntos do conjunto de indicadores, foram propostas dependências para prever os estados dos objetos. Como uma implementação do método desenvolvido, a previsão de alguns indicadores do desenvolvimento de cidades da Rússia executou-se.

Palavras-chave: países, regiões e cidades, indicadores de previsão, espaços de estado, modelos de comportamento coletivo, métodos fenomenológicos.

Introduction

The basis of strategic planning for the development of countries, regions and cities is the forecasting of the social economic state of the objects on a given forecast horizon. In this direction, there are many topical challenges, among which we can distinguish the following: forecasting the states and trends of evolutionary development of objects; complex evaluation and ranking of objects on a set of indicators for the purpose of the irrelative comparison; scenario forecasting of development processes for making management decisions.

A perspective way of developing predictive methods is connected with new approaches to modeling social processes and objects that can use natural scientific methodology. The solution of such a problem can be achieved on the basis of using predictive methods that will be implemented at high levels of understanding information. This way involves the use of knowledge about the phenomenological laws of collective behavior of objects, based on an analysis of the totality of indicators, and characteristic events and their probabilities.

So, the objective of this work is the development of new methods for forecasting the social economic development of countries, regions and cities based on the use of phenomenological approaches to the processing and analysis of statistical information and the use of event models of collective behavior.
reduced to the study of the processes of the development of objects in multidimensional spaces under various influences and certain limiting conditions. For example, the main research in this area is related to the development of methodology, mathematical apparatus and software solutions for forecasting complex social and social systems and managing collective behavior. In mathematical support, the emphasis is not made on the use of methods of mathematical statistics, but on the use of category theory and functors, methods of pattern recognition, neural computing, machine learning, computational mathematics, digital intelligence technologies and the construction of various models of collective interaction and the development of objects (Lillo, 2008; Newman, 2011).

Studies in the field of system dynamics and sociophysics are being conducted constantly, since it is recognized that in these sciences important fundamental results can be obtained in the field of modeling and forecasting of social processes (Slovohotov, 2012; Averin, 2014). However, the analysis of existing publications indicates that at present there is no fundamental theory that would characterize the socio-economic development of countries, regions and cities and that would allow using phenomenological methods for processing and analyzing multivariate statistical data.

**Principles, Hypotheses and Methods**

Core of methods of forecasting, realized on high levels of information understanding, lies in the fact that algorithms of analysis, processing and specification of data are initially based on consistent patterns of formation of processes under study. By consistent patterns it is typically meant logical, phenomenological, and other regularities, which are represented as certain theory applied to the class of systems under study.

Perspective ideas of forecasting are related to the phenomenological approaches of researching systems based on actual data (Slovohotov, 2012; Averin, 2014). Practice shows, that it is possible to form temporal data arrays for a whole range of systems which transforms and develops in time. Usually such data arrays have structure of tables in form “objects-indicators”, and the corresponding number of tables (t) is ordered in time, for instance: years, months, etc. All statistical data of state and development of cities, regions and countries are related to creation of temporal databases. State of object in multidimensional space of states is determined by complex of values of object’s indicators, which parameters are formed in specific moment of time (Zviagintseva, 2016)

Put the case, that for r objects of one class in temporal data arrays some quantitative information about n indicators \( p_k \) \((k=1, \ldots, n)\) is contained, and it characterizes different attributes of current objects. Any array of n variables for parameters of attributes sets n-dimensional space of states \( \Omega^p \), where \( p=(p_1, p_2, \ldots, p_n) \), \( p \in \Omega^p \). Points of this space correspond to n-dimensional sets of values of all variables \( p_1, p_2, \ldots, p_n \). So any state in n-dimensional space at each moment of time would be reflected as multidimensional point \( M = M(p_1, p_2, \ldots, p_n) \), while process of transforming of state in time as some curve.

Basic principles used for processing quantitative information are consisted in the following.

The continual principle of representation of quantitative information in space is laid in basis of methods of describing the states of objects \( \Omega^p \). According to this principle hypothetical environment in the form of space of states is counted as uninterrupted, structure less, while each element of space is related to all neighbor elements taking into account continual consistent patterns. It allows to determine experimental data as certain discrete sample from continuous hypothetical environment of an infinite number of states for objects of the same class. This sample consists of experimental points \( M_i \), their total amount is equal to r \((i=1,2,\ldots, r)\).

The second principle is based on hypothesis that discrete data forms certain “image” in continual space. It is considered, that this “image” can be described on the basis of estimates of the statistical probabilities of events, characterizing position of each experimental point relative to the whole group of objects under study (the whole space of points \( M \)). It is supposed, that the equation of objects’ states in the form of certain empirical measure \( W = W(p_1, p_2, \ldots, p_n) \) can be constructed, while each state in the space \( \Omega^p \) is characterized by mutual event \( A_i \) (indicative event) of simultaneous observation of indicators \( p_1, p_2, \ldots, p_n \). The hypothesis is based on the assumption that empirical measure W can be related to probability of this event. With this
approach, the modelling object is the state of objects (countries, regions, cities, etc.), which can be characterized by equation of states, equivalent for the whole space \( \Omega^\prime \).

The third principle is related to possibility of algorithmic determination of statistical probability of events \( A_i \) by direct counting. This statistical probability \( w \) can be named as probability of object’s state (certain area near the point \( M \)). This probability \( w \) is determined by relative frequencies of indicative events \( A_i \) by splitting the whole space \( \Omega^\prime \) on multidimensional parallelepipeds \( V_k^\beta \), based on the specified number of intervals of grouping \( \theta \) for each variable \( p_i \) (usually the same for all \( p_i \)). After this, relative frequencies of events are counted. Statistical probability of state in the specific space \( V_k^\beta \) taken in the form of cumulative relative frequencies of the events under study. The corresponding algorithms used for the direct calculation of the probabilities \( w \) are given in (Averin, G., Konstantinov, I. et al. 2015).

The fourth principle is related to the hypothesis of description of continuum regularities of space \( \Omega^\prime \) by means of simple model functions, which will reflect the accepted metric of the state space \( \theta = \theta (p_1, p_2, \ldots, p_n) \).

We will formulate several postulates that allow us to construct a model of collective behavior of objects. These postulates will be presented in the form of:

1. Let the state space \( \Omega^\prime \) each point of \( M \) corresponds to a real number \( W (-\infty < W <= +\infty \) or \( 0 <= W <= +\infty \)), which we call the empirical measure of the observed condition;
2. The figure \( W(M) \) is a function of the point and forms a scalar field that is continuous at \( \Omega^\prime \).

To construct the model, let’s suppose that the scalar field of the value \( W = W(M) \) can be analytically described. For this, let’s assume that in region \( \Omega^\prime \) one can specify a continuous function \( \theta (p_1, p_2, \ldots, p_n) \) in the metric space, which will form the mathematical model. This allows us to build another scalar field, which will be the basis for modeling. Based on this, we formulate a hypothesis.

3. Let the state space \( \Omega^\prime \) scalar field variables \( W \) and \( \theta \) be clearly linked. If the process \( M \) is carried out in the neighborhood of any point \( I \), then for the process line \( I \) the ratio \( dW = c \theta d\theta \) is true, where \( c \) - empirical values, which are the functions of the process.

The metric of the space \( \Omega^\prime \) can be represented as a functional dependency with respect to all \( n \) parameters: multiplicative, exponential, expertise, etc. dependencies, or in the form of various measures of similarity: Euclidean, Manhattan, power distance, Chebyshev distance, Minkowski etc. In this paper, we assume that it is possible to establish an unequivocal link between the empirical measure \( W \) and the probability of state \( w \).

Then we choose an arbitrary point \( M \) in region \( \Omega^\prime \). We assume that the process of changing of some object’s state is taking place near this point. Then the elementary change of the empirical measure \( W \) can be represented as:

\[ dW = \left( \frac{\partial W}{\partial \theta} \right)_{p_1, p_2, \ldots, p_n} dp_1 + \left( \frac{\partial W}{\partial \theta} \right)_{p_1} dp_2 + \ldots + \left( \frac{\partial W}{\partial \theta} \right)_{p_n} dp_n \]

(1)

\[ dW = c_1 \frac{\partial \theta}{\partial p_1} dp_1 + c_2 \frac{\partial \theta}{\partial p_2} dp_2 + \ldots + c_n \frac{\partial \theta}{\partial p_n} dp_n \]

(2)

where \( \theta_k \) – unit vectors directed along the coordinate axes \( p_1, p_2, \ldots, p_n \) of the state space \( \Omega^\prime \); \( c_i = c_i(p_1, p_2, \ldots, p_n) \).

The metric \( \theta \) in the field \( \Omega^\prime \) can be represented as a functional dependency belonging to the classes homogeneous or multiplicative functions. In work (Averin, G., Zviagintseva, A. et al. 2015), it is established that under these conditions for construction of models it is possible to use quasilinear multidimensional partial differential equations of the first order which are connected with differential forms of Pfaff of the type (2).

For example, let’s take a measure of relative changes as a metric \( \theta \) for describing the state space as a model function

\[ \theta = (p_1, p_2, \ldots, p_n)(p_{10}, p_{20}, \ldots, p_{n0}) \]

where \( p_{k0} \) are the values of \( p_k \) values for the accepted reference state. It can be shown that in this case for equation (2) there is an integrating divisor in the form of the accepted function \( \theta \). Substituting this function in (2) and dividing this equation by \( \theta \), we obtain:

\[ ds = \frac{dW}{\theta} = c_1 \frac{dp_1}{p_1} + c_2 \frac{dp_2}{p_2} + \ldots + c_n \frac{dp_n}{p_n} \]

(3)

Integrating equation (3) in a neighborhood of an arbitrary state of \( M \), and considering the
magnitude of the conditional constant \( c_n \), we will present the general integral in the form:

\[
s - s_0 = c_1 \ln \left( \frac{p_1}{p_{10}} \right) + c_2 \ln \left( \frac{p_2}{p_{20}} \right) + \ldots + c_n \ln \left( \frac{p_n}{p_{n0}} \right)
\]

(4)

We define \( s \) as an entropy based on analogies with thermodynamics and information theory. Entropy is a characteristic function of the space state of the system \( \Omega^2 \). As shown in the parametric representation in the source, the entropy is the arc length of the vector line of a certain direction field generated by the scalar field of the empirical measure \( W \). This conclusion follows from the form of the Pfaffian equation (2) and its general solutions. It is known that the corresponding field of directions has the form:

\[
\Gamma(p_1, p_2, ..., p_n) = \left( \frac{p_1}{n c_1} e_1 + \frac{p_2}{n c_2} e_2 + \ldots + \frac{p_n}{n c_n} e_n \right)
\]

(5)

Vector lines of this field will be determined by ordinary differential equations:

\[
n c_1 \frac{dp_1}{p_1} = n c_2 \frac{dp_2}{p_2} = \ldots = n c_n \frac{dp_n}{p_n} = dW
\]

(6)

from where it is easy to get the expression (3) for the entropy of the state.

For a vector \( \mathbf{r} \), there is a family of surfaces orthogonal to entropy lines \( s \). The equation of such surfaces will be determined from the scalar product \( (\mathbf{\Gamma} \cdot \mathbf{t}) = 0 \), where \( \mathbf{t} = e_1 dp_1 + e_2 dp_2 + \ldots + e_n dp_n \) is a vector lying in the tangent plane to the original surface, from where:

\[
\frac{p_1}{c_1} dp_1 + \frac{p_2}{c_2} dp_2 + \ldots + \frac{p_n}{c_n} dp_n = 0.
\]

(7)

The Pfaffian form on the left side of equation (7) with constant \( c_i \) values in the neighborhood of an arbitrary point \( M \) is a complete differential, so the potential of the state space \( P(p_1, p_2, ..., p_n) = C \) can be determined from equation (7):

\[
P = \frac{1}{2} \left( \frac{p_1^2 - p_{10}^2}{c_1 p_{10}^2} + \frac{p_2^2 - p_{20}^2}{c_2 p_{20}^2} + \ldots + \frac{p_n^2 - p_{n0}^2}{c_n p_{n0}^2} \right)
\]

(8)

The potential in the reference state is assumed to be zero \( (P_0 = 0) \). Entropy \( s \) and potential \( P \) can be used as criteria for predicting the collective behavior of objects in space \( \Omega^2 \). An important feature is that the values of \( s \) and \( P \) are functions of the state under the conditions of existence of the scalar field of the empirical measure \( W \). Change of these functions depends only on the initial and final state of the object and does not depend on the path of its transition between these States.

In this type of theory, it is important to choose the measure \( W \) to characterize the States of objects, as well as the development of a system for determining this value. Assume that this value is uniquely related to the probability of state \( w \), which is algorithmically evaluated for the available data.

Since entropy in the form (4) is a general Pfaffian form integral (2) with an integrating divisor \( \theta \), every integrating factor of equation (2) is given by the formula \( \mu = \phi(s)/\partial \), where \( \phi \) is a continuous differentiable function. In this case we can write the general relationship between the integrals of the Pfaffian form in the form \( dw = \phi(s)ds \). Due to the fact that \( \phi \) function can be selected arbitrarily, it can be selected in such a way that it satisfies one of the known probability distributions of random variables, for example, normal. For this purpose, the density of the normal distribution can be used

\[
dw = \left( \frac{1}{\sqrt{2\pi}} \right) \exp \left( - \frac{\omega^2}{2} \right) d\omega.
\]

The value \( \omega \) is called the probit of probability and is usually denoted \( \text{Prob} \).

All that has been said above makes it possible to propose the following method for processing discrete data on the probabilities of joint events. First, the probability values of events are determined from the available temporal data. Further, on the known \( w \) probability values there are determined probit-values (\( \text{Prob} \)), according to the experimental data the components of the entropy function (4) is determined as a logarithm values and parameters, and by multiple regression method \( c_i \) values are defined, which characterize the average trend in the formation states of the studied objects. The quality of the phenomenological model obtained is estimated by statistical methods.

Following the above mentioned, it can be seen that the proposed method is closely related to the logic of constructing the theory of thermodynamics, since the phenomenologically determined values of \( c_i \) are initially introduced. Therefore, by analogy, it is possible to use the existing mathematical apparatus in modeling the processes.
**An example of forecasting socio-economic indicators of Russian cities.** The information of the database of the Federal Service for Statistics in the form of a sample of data on 63 indicators for 159 cities in Russia with a population of more than 100 thousand people have been used for forecasting of object states. The selection scope comprised of 120 thousand experimental observations (2013–2015).

The method of forecasting the social and economic state of cities will be considered by using the example of compiling a medium-term forecast for two indicators: the volume of goods and services of industrial production $p_1$, mln. rub. and retail trade turnover $p_2$, mln. rub.

Equations of state of cities for each year in the period 2003–2015 were determined on the basis of the available data. Further the values of the coefficients $c_t$ in time were predicted from the obtained data for the 2020 year:

$$
\text{Prob} = -4.684 + s; \quad s = 0.279 \ln \left( \frac{p_1}{p_{10}} \right) + 0.526 \ln \left( \frac{p_2}{p_{20}} \right) .
$$

(9)

Indicators for evaluation were relative to the minimum values of urban indicators in 2003, which are: $p_{10} = 150$ mln. rub.; $p_{20} = 117$ mln. rub.

To implement the study of scenarios for the development of objects an empirical measure $\Delta W = \theta \cdot \Delta s$ was used, which allows to predict the level of impact of the socio-economic environment on the forecasting object. This value will have its own characteristics for each set of indicators.

The medium-term forecast for the development of Russian cities was represented by three variants of scenarios. The first option (the evolutionary scenario) assumed that the development of cities takes place with account of the characteristic trends of changes in indicators that have developed in 2013–2015. The second option (the progressive scenario) assumed that the development of the entire group of cities occurs when there is a general growth trend in relation to the indicators of the state of objects, estimated under the evolutionary scenario. In this case, the values of $W$ were assumed to be 10% greater than the corresponding value determined in the evolutionary scenario. In its turn, the depressive scenario (the third variant) suggested the tendency of decreasing indicators in relation to the adopted evolutionary development of cities (the $W$ values were assumed to be 10% less).

In the forecasting process, extrapolation of values of $W$ for five years was carried out for each city, based on the available data on its values in the period 2003–2015 (Fig. 1).

![Fig. 1. Example of extrapolation of the W value for some Russian cities](image-url)

When analyzing it is important not only to assess the indicators of the state of objects, but also to determine the changes in these indicators in different processes of urban development. To this end, we introduce for the observed process the value of $c_t$, equal to the ratio of the increment of the measure $\Delta W$ to the increment $\Delta \theta$ of the:

$$
\frac{\Delta W}{\Delta \theta} = \frac{W - W_s}{\theta - \theta_s}; \quad c_t = 0.459 + 0.350 \ln (W_s).
$$

(10)
where \( W, \theta \) – values of variables for 2015; \( W_i, \theta_i \) – the values of the corresponding variables for the previous years. After 2010, this trend is true for the vast majority of cities. The development of 15 cities out of 159 is not determined by the trend (10). The correlation coefficient of equation (10) was 0.98, the results of data processing are shown in figure 2.

For different periods of time, the values of \( \theta \) were predicted from the values of the empirical measure \( W \) (Fig. 2), taking into account (10). The obtained values of these variable for 2020 allow to estimate the entropy of the state of objects. Further, using the equation of state (9) and the accepted metric \( \theta = p_1 p_2/(p_1 + p_2) \), the forecast values of the volume of goods and services of industrial production \( p_1 \) and the turnover of retail trade \( p_2 \) are found.

The obtained results have made it possible to fulfill the forecast of the indicators under the accepted scenarios for the development of cities for the period of preemption to 2020.

![Fig. 2. The value of the quantity \( c_r \) as a function of the value of \( W \).](image)

<table>
<thead>
<tr>
<th>Russian cities</th>
<th>Socio-economic indicators for scenarios (mln. rub.):</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Evolutionary</td>
</tr>
<tr>
<td>Belgorod</td>
<td>71359</td>
</tr>
<tr>
<td>Vladimir</td>
<td>83077</td>
</tr>
<tr>
<td>Voronezh</td>
<td>185690</td>
</tr>
<tr>
<td>Kuga</td>
<td>179893</td>
</tr>
<tr>
<td>Kursk</td>
<td>88950</td>
</tr>
<tr>
<td>Oryol</td>
<td>56407</td>
</tr>
<tr>
<td>Ryazan</td>
<td>205660</td>
</tr>
<tr>
<td>Yaroslavl</td>
<td>178421</td>
</tr>
<tr>
<td>Murmansk</td>
<td>94906</td>
</tr>
<tr>
<td>Volgograd</td>
<td>560126</td>
</tr>
<tr>
<td>Rostov-na-Donu</td>
<td>400327</td>
</tr>
<tr>
<td>Ufa</td>
<td>470852</td>
</tr>
<tr>
<td>Permian</td>
<td>524790</td>
</tr>
<tr>
<td>Orenburg</td>
<td>137153</td>
</tr>
<tr>
<td>Ekaterinburg</td>
<td>453937</td>
</tr>
<tr>
<td>Chelyabinsk</td>
<td>512296</td>
</tr>
<tr>
<td>Irkutsk</td>
<td>163444</td>
</tr>
<tr>
<td>Omsk</td>
<td>935485</td>
</tr>
<tr>
<td>Tomsk</td>
<td>117704</td>
</tr>
</tbody>
</table>

Table 1. The values of \( p_1 \) and \( p_2 \) in 2020 for Russian cities.
The error in determining indicators was: for the volume of goods and services of industrial production $p_1$ - the average 9.8%, the maximum 27.5%; for the turnover of retail trade $p_2$ – the average 9.2%, the maximum 23.5%.

Similar results were obtained for some other indicators characterizing the state and development of cities.

Conclusions and Prospects

So, it is shown that with sufficient volume of experimental data it is possible to construct phenomenological models for certain classes of socio-economic systems. The proposed approach allows us to develop mathematical models for solving prediction problems based on taking into account the patterns of collective behavior of objects. The above practical examples of creating models for assessing the state and development of countries, regions and cities indicate a sufficient theoretical study of the proposed prognostic methods.

References


