ENGINEERING MATHEMATICS

MSC 91B84 **DOI: 10.14529/jcem230404**

THE ALL-RUSSIAN CLASSIFIERS ARE AN IMPORTANT TOOL FOR DESCRIBING AND MODELING TYPES OF ECONOMIC ACTIVITY

N. V. Poletaeva[1]**,** nexochuxa.08@mail.ru**,**
N. S. Kolotova[1]**,** kolotovans@susu.ru
[1]South Ural State University, Chelyabinsk, Russian Federation

The article analyzes the construction of a time series of data obtained using two all-Russian classifiers of economic activities: ACTEA-2007 (data from 2005 to 2016) and ACTEA2 (data from 2017 to 2022). The research was carried out by types of activities in the field of information technology, the encoding of which varies greatly in these classifiers. The series are studied using Python program code. For the generated time series, statistical models were constructed using regression analysis and a forecast for two years ahead was made. When checking the possibility of applying one obtained equation to two samples, the Chow test is used. The research showed that it is possible to make a qualitative statistical model using various regression equations based on the given time series: in this article, polynomial and piecewise linear models are considered.

Keywords: statistical modeling; the all-Russian classifier of economic activities; regression analysis; polynomial; piecewise broken model.

Introduction

The All-Russian Classifier of Types of Economic Activities (ACTEA) is used to systematize the diverse economic activities in the Russian Federation. In order to compare the domestic and foreign economies, the ACTEA classifier is harmonized with the foreign counterpart, the Statistical Classification of Economic Activities in the European Union (NACE). Since 2003, the activity has been described by the ACTEA-2001 classifier, then the ACTEA-2007 [1] classifier was introduced, which was replaced in 2016 by ACTEA2 [2]. The ACTEA-2001 and ACTEA-2007 classifiers have very few differences, but the ACTEA2 classifier has undergone significant changes with the emergence of new types of activities.

This article analyzes the changes in the turnover of organizations (excluding small businesses, budget organizations, banks, insurance and other financial and credit organizations) by type of economic activity from 2005 to 2022 in Russia, and also examines the practice of modeling and forecasting indicators by type of economic activity using the ACTEA-2007 and ACTEA2 classifiers. The forecast requires data for recent years, that is, data on the classifier ACTEA2. But information on this classifier has been collected since 2017, so the data series consists of only 6 values. Therefore, the series needs to be supplemented with data collected according to the ACTEA-2007 classifier.

The research of the time series is carried out using Python program code. For this purpose, an activity has been selected, the encoding of which in ACTEA-2007 and ACTEA2 is very different. In ACTEA-2007, class 72 «Software development and consulting in this area» , located in section K «Real estate transactions, rental and

provision of services» , was divided into two groupings in section J «Information and communication activities» of the classifier ACTEA2:

– computer software development, consulting services in this field and other related services (class 62);

– information technology activities (class 63).

1. Computer Software Development, Consulting Services in this Field and Other Related Services

The algorithm of the program code is considered on the example of a number of data: turnover by type of activity «Development of computer software, consulting services in this area and other related services» .

1.1. Preliminary Analysis

Formation of a time series of the turnover of organizations by type of economic activity from 2005 to 2022 in Russia. The source of information for the reserch is the Unified Interdepartmental Information and Statistical System (UIISS) [6]:

– turnover of organizations (excluding small businesses, budget organizations, banks, insurance and other financial and credit organizations) by 2016 (thousand rubles, the value of the indicator for the year) - data collected by the ACTEA-2007 classifier from 2005 to 2016;

– turnover of organizations (excluding small businesses, budget organizations, banks, insurance and other financial and credit organizations) since 2017 (thousand rubles, the value of the indicator for the year) - data collected according to the ACTEA2 classifier from 2017 to 2022.

In order to build a time series of the turnover of organizations by type of economic activity from 2005 to 2022 of the type of activity «Development of computer software, consulting services in this area and other related services» since 2005, we will form this series using the transition key and the corresponding ACTEA-2007 codes.

1. Subclass 62.0 «Development of computer software, consulting services in this area and other related services» (ACTEA2) corresponds to the code of subclass 72.2 «Software development and consulting in this area» (ACTEA-2007).

2. Group 62.02 «Advisory activities and work in the field of computer technology» switched from two groups by key:

   – 72.22 «Other software development and consulting activities in this area» ;

   – 72.10 «Consulting on computer hardware» .

3. Group 62.09 «Activities related to the use of computer technology and information technology, other» corresponds to code 72.60 «Other activities related to the use of computer technology and information technology» .

Thus, to get a series in ACTEA-2007, which can be extended in ACTEA2, you need to add up the data for groupings 72.10, 72.2 and 72.60 (Fig. 1). The possibility of combining samples into one will be discussed after the Chow test [4, p.59], for which it is necessary

to obtain the equation of the model. At this stage, we will combine two samples into one conditionally, the force series will be called a *TurnoverSFact*. The graph of the resulting series with a length of 18 values without missing data is shown in Figure 2.

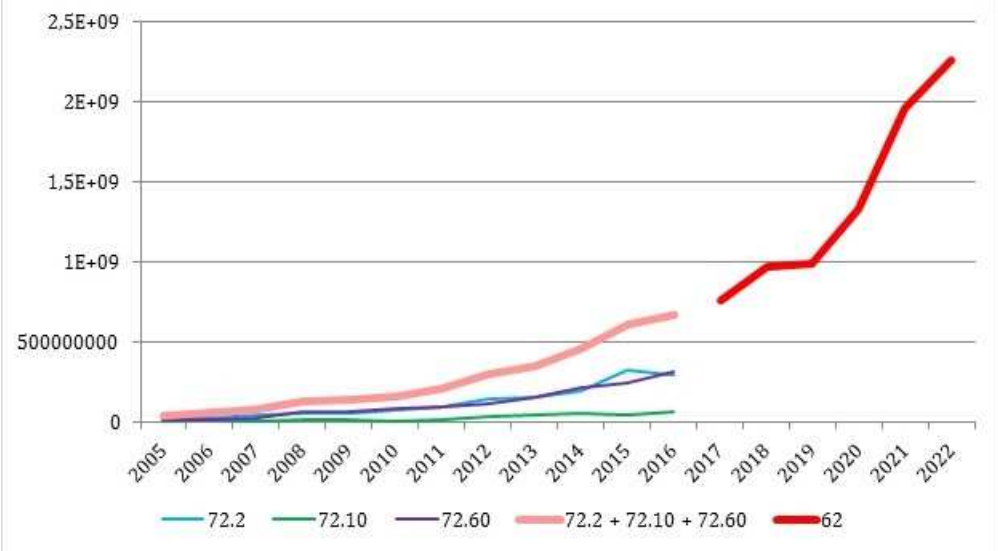

Fig. 1. Graph of the turnover of organizations by type of activity «Development of computer software, consulting services in this area and other related services» , two rows

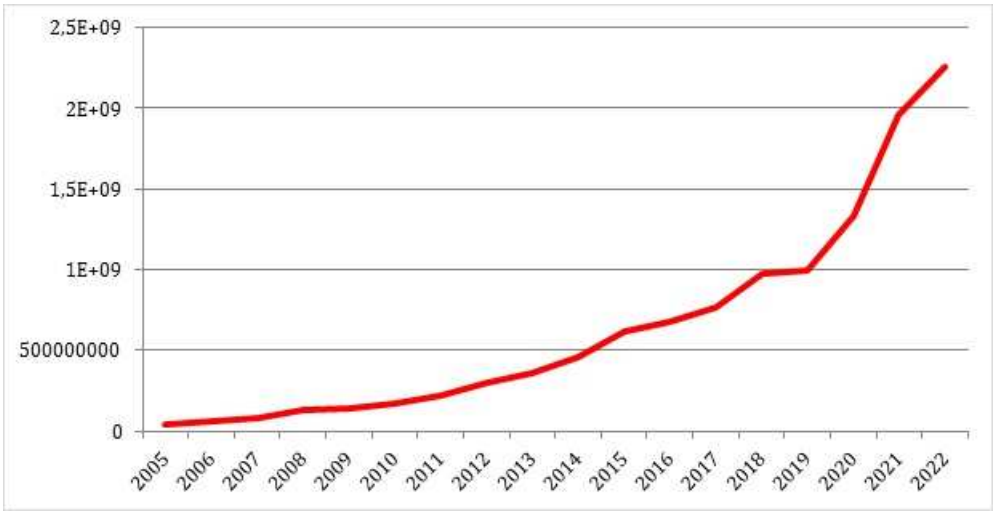

Fig. 2. Graph of turnover changes of organizations by type of activity «Development of computer software, consulting services in this area and other related services» , one row

We check a number of data for uniformity and the presence of a trend [5, p. 190]. Checking the time series for uniformity by the Irwin method [5, p. 189] showed that the series is homogeneous. There is no anomaly in the transition from ACTEA-2007 to ACTEA2 (2017 level).

Next, we identify the presence of a trend in the time series – the Foster–Stewart method [4, p. 74], we get the result: $t_{cr} = 1.740$, $t_{obser} = 3.279$, $t_{cr} < t_{obser}$. If the actual

value of the t-statistics is greater than the critical one, then the hypothesis H_1 - about the presence of a trend in the original time series - is accepted.

So, the time series «Computer software development, consulting services in this field and other related services» is homogeneous and tends.

1.2. Regression Analysis

After the regression analysis , the following equation was obtained:

$$Y = 51512299.36500939 + 3786137.226382116t^2 + 0.09318905929371174t^8,$$

where Y is a dependent variable reflecting changes in the turnover of organizations by type of activity «Computer software development, consulting services in this area and other related services» .

The model has the following characteristics:

– approximation error is equal to 8.750% ($< 10\%$);

– the coefficient of determination is 0.990 (> 0.7), that is, the model values are 99.0% close to the empirical ones;

– the Fisher criterion is equal to 750.0, greater than the critical value of the Fisher criterion of 3.55, as well as a p-value equal to 9.28×10^{-16}, significantly less than the significance of α, therefore the regression equation is significant;

– the p-significance of the coefficients is comparable to the given value of α equal to 0.05: at a constant – 0.089 (> 0.05), insignificant; at t^2 – 0.000 (< 0.05), significant; at t^8 – 0.000 (< 0.05), significant.

– the Jarque–Bera statistic is 0.301, greater than the given significance α, so the residuals are distributed normally;

– the Durbin–Watson criterion is equal to 2.272, is in the range from 1.39 to 2.61, this indicates that the residuals are independent.

The Chow test [4, p. 59] showed that the samples can be combined into one: $F_{cr} = 3.88, F_{obser} = 1.158, F_{cr} > F_{obser}$.

The significance of the constant 0.089 slightly exceeds the significance of α, but this is acceptable. The remaining coefficients and the equation are significant.

1.3. Checking the Quality of the Model

The model meets the criteria of adequacy [4, p.79]:

1. At the significance level of 0.05, the hypothesis of randomness of the residuals is accepted, since the maximum length of the series is $L_{max} = 5$, the number of series is $V = 10$, $L_{max} < 3.3(\lg(n+1)$ and $V > (n + 1 - 1.96((n-1)^{0.5}))/2$.

2. Shapiro–Wilk statistics [3, p. 26]: $stat = 0.91$, $p = 0.099$, $p \geq \alpha$, the distribution of a number of residues is normal.

3. The condition that the average of the residuals is equal to zero is also fulfilled: $t_{cr} = 1.740$, $t_{obser} = 1.671 \cdot 10^{-14}$, $t_{cr} > t_{obser}$.

4. The Darbin–Watson criterion of 2.272 is in the range from 1.39 to 2.61, therefore, the hypothesis of the independence of the residues is accepted.

The actual (*TurnoverSFakt*) and model (*TurnoverSModel*) values with a forecast for 2023 and 2024 are shown on the graph (see Figure 3). We obtained a qualitative model using a polynomial equation.

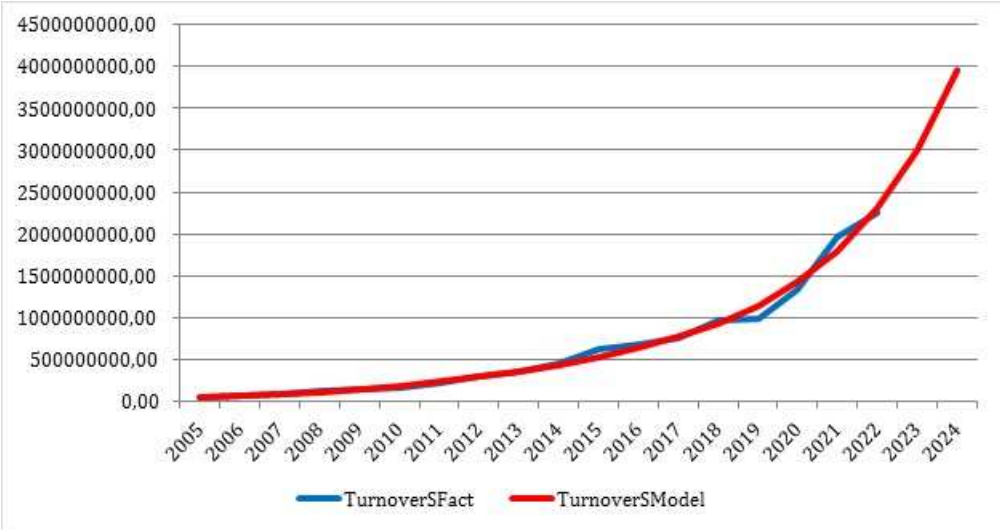

Fig. 3. Graph of the turnover of organizations by type of activity «Development of computer software, consulting services in this area and other related services»

2. Information Technology Activities

We conduct a similar reserch for a number of «Information technology activities» , which includes:

1. Subclass 63.1 «Data processing activities, provision of information placement services, activity of portals in the Internet information and communication network» corresponds to the code of subclass 72.3 «Data processing» (ACTEA-2007).

2. Subgroup 63.11.1 «Activities for the creation and use of databases and information resources» was transferred from group 72.4 «Activities for the creation and use of databases and information resources, including Internet resources» .

3. According to the transition key, subclass 63.9 «Activities in the field of other information services» corresponds to code 92.40 «Activities of information agencies» (ACTEA-2007, section On «Provision of other communal, social and personal services»).

Therefore, to get a complete series, let's call the *TurnoverITFact*, in ACTEA-2007, equivalent to the series in ACTEA2, you need to add up the data for groupings 72.3, 72.4 and 92.40 (see Figure 4). The resulting time series is homogeneous and has a tendency. But since 2018, the behavior of the trend is changing, for this reason it is difficult to build a model that would pass the Chow test. Having calculated the F-statistics of this distribution, we get the result: $F_{cr} = 3.88, F_{obser} = 13.320, F_{cr} < F_{obser}$. Therefore, a piecewise linear model will be optimal for this data series [5, p. 205]. The model is described by two linear equations:

1) from 2005 to 2017: $Y = -4621834.492307716 + 14731125.55934066t$;
2) from 2018 to 2022: $Y = -138112476.3999997 + 139394908.68000004t$.

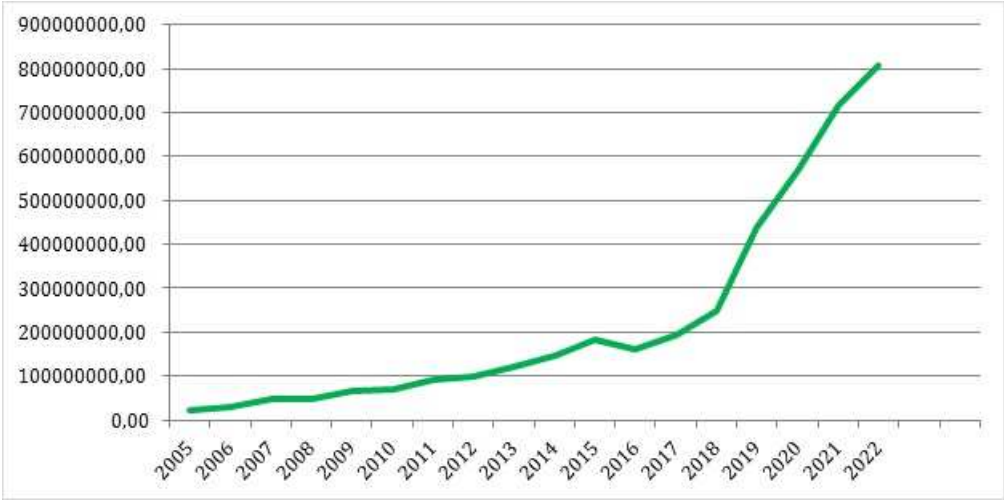

Fig. 4. Graph of the change in turnover of organizations by type of activity «Information technology activities»

Characteristics of the first model:
– approximation error 12.938%($> 10\%$);
– significance of the equation $3.17 \times 10^{-9}(< \alpha)$;
– significance of the coefficients: constant – $0.517(> \alpha)$, t – $0.000(< \alpha)$;
– coefficient of determination is 0.963 (> 0.7).
Characteristics of the second model:
– approximation error 4.952% ($< 10\%$);
– significance of the equation 0.000656 ($< \alpha$);
– significance of coefficients: constant – 0.021 ($< \alpha$), t – 0.001 ($< \alpha$);
– coefficient of determination is 0.987 (> 0.7).

The models have been obtained of sufficient quality. The general model was obtained with an approximation error of 10.707%, a determination coefficient of 0.994 and meets the quality criteria:

1. At the significance level of 0.05, the residuals are considered random: the maximum length of the series, $L_{max} = 6$, the number of series, $V = 7$: $L_{max} < (3.3(lg(n+1))$ and $V > ((n + 1 - 1.96((n-1)^{0.5}))/2$.

2. Shapiro-Wilk statistics: $stat = 0.96$, $p = 0.686$, $p \geq \alpha$, the distribution of a number of residues is normal.

3. The equality of the mean residuals to zero: $t_{cr} = 1.740$, $t_{obser} = 0.000945$, $t_{cr} > t_{obser}$.

4. Since the Durbin-Watson criterion 2.148 is in the range from 1.39 to 2.61, the residuals are independent.

The actual and model values are shown in Figure 5. The predicted values for 2023 and 2024 are calculated using the second equation.

Conclusions

The analysis showed that combining the data collected by different classifiers ACTEA-2007 and ACTEA2 into one time series makes it possible to compile a qualitative statistical

model using various regression equations: in this article, a polynomial and piecewise linear model are considered. As a result of the study, a mathematical model of the dynamics of turnover of organizations by type of economic activity from 2005 to 2022 was compiled and a forecast for 2023 and 2024 was made.

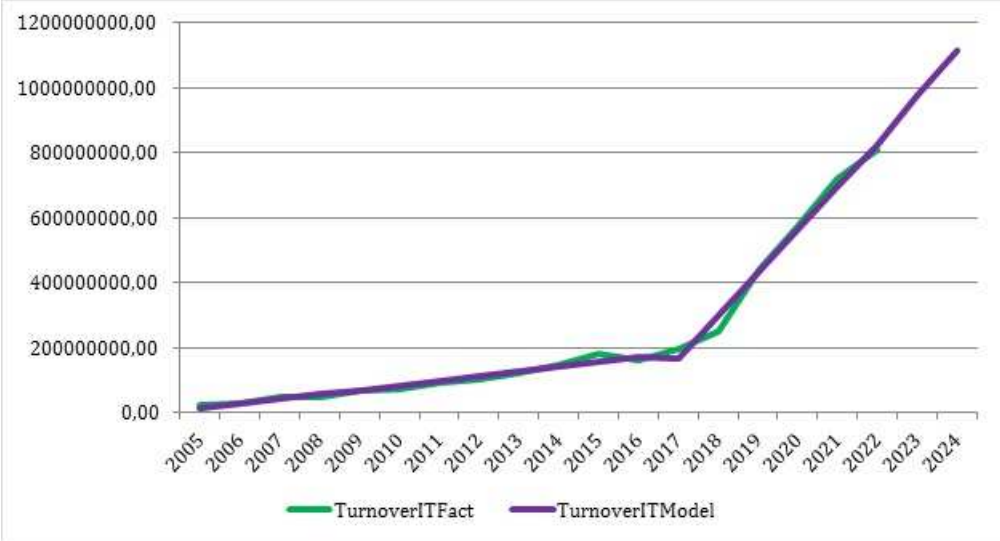

Fig. 5. Graph of changes in turnover of organizations by type of activity «Information technology activities» , actual and model values

References

1. The All-Russian Classifier of Types of Economic Activity (ACTEA OK 029-2007 (KDES Ed.1.1)): Decree of the Government of the Russian Federation of November 10, 2003, no. 677. – URL: http://www.consultant.ru (in Russian) (date of access: 11/30/2023)
2. The All–Russian Classifier of Types of Economic Activity (ACTEA OK 029-2014 (KDES Ed.2)): Decree of the Government of the Russian Federation of November 10, 2003, no. 677. – URL: http://www.consultant.ru (in Russian) (date of access: 11/30/2023)
3. Lemeshko A.Y. [*Criteria for Checking the Deviation of the Distribution from the Normal Law*]. – URL: https://ami.nstu.ru/˜headrd/seminar/publik_html/ guid _normal_tets.pdf (in Russian) (date of access: 11/30/2023)
4. Shanchenko N.I. [*Lectures on Econometrics*]. Ulyanovsk, UlSTU, 2008. (in Russian)
5. Kamornikov S.F., Kamornikov S.S. [*Econometrics*]. Moscow, Integration, 2012. (in Russian)
6. Unified Interdepartmental Information and Statistical System (UIISS) [Electronic resource] – URL: http:// rosstat.gov.ru/emiss (in Russian) (date of access: 11/30/2023

Natalia V. Poletaeva, Master Student, Department of Mathematical and Computer Modeling, South Ural State University (Chelyabinsk, Russian Federation), nexochuxa.08@mail.ru

Nadezhda S. Kolotova, PhD (Econom), Department of Mathematical and Computer Modeling, South Ural State University (Chelyabinsk, Russian Federation), kolotovans@susu.ru

Received October 31, 2023

УДК 519.2 **DOI: 10.14529/jcem230404**

ОБЩЕРОССИЙСКИЕ КЛАССИФИКАТОРЫ – ВАЖНЕЙШИЙ ИНСТРУМЕНТ ДЛЯ ОПИСАНИЯ И МОДЕЛИРОВАНИЯ ВИДОВ ЭКОНОМИЧЕСКОЙ ДЕЯТЕЛЬНОСТИ

Н. В. Полетаева, Н. С. Колотова

В статье проведен анализ построения временного ряда данных, полученных с помощью двух общероссийских классификаторов видов экономической деятельности: ОКВЭД-2007 (данные с 2005 по 2016 года) и ОКВЭД2 (с 2017 по 2022 года). Исследование выполнено по видам деятельности в области информационных технологий, кодировка которых в этих классификаторах сильно различается. Ряды изучены при помощи программного кода Python. Для сформированных временных рядов построены статистические модели с помощью регрессионного анализа и выполнен прогноз на два года вперед. При проверке возможности применения одного полученного уравнения к двум выборкам используется тест Чоу. Исследование показало, что по приведенным временным рядам возможно составить качественную статистическую модель с применением различных регрессионных уравнений: в данной статье рассмотрены полиномиальная и кусочно-линейная модели.

Ключевые слова: статистическое моделирование; общероссийский классификатор видов экономической деятельности; регрессионный анализ; полином; кусочно-ломанная модель.

Литература

1. Общероссийский классификатор видов экономической деятельности (ОКВЭД ОК 029–2007 (КДЕС Ред.1.1)): Постановление Правительства Российской Федерации от 10 ноября 2003г. № 677 – URL: http://www.consultant.ru. (дата обращения: 30.11.2023)
2. Общероссийский классификатор видов экономической деятельности (ОКВЭД ОК 029–2014 (КДЕС Ред.2)): Постановление Правительства Российской Федерации от 10 ноября 2003г. № 677 – URL: http://www.consultant.ru (дата обращения: 30.11.2023)
3. Лемешко, А.Ю. Критерии проверки отклонения распределения от нормального закона / А.Ю. Лемешко. – URL: https://ami.nstu.ru/ headrd/seminar/publik_html/ guid_normal_tets.pdf (дата обращения: 30.11.2023)
4. Шанченко, Н.И. Лекции по эконометрике / Н.И. Шанченко. – Ульяновск: УлГТУ, 2008.

5. Каморников, С.Ф. Эконометрика / С.Ф. Каморников, С.С. Каморников. – Москва: Интеграция, 2012.
6. Единая межведомственная информационно-статистическая система (ЕМИСС) – URL: http:// rosstat.gov.ru/emiss (дата обращения: 30.11.2023)

Полетаева Наталья Васильевна, магистрант, кафедра математического и компьютерного моделирования, Южно-Уральский государственный университет (г. Челябинск, Росиийская Федерация), nexochixa.08@mail.ru

Колотова Надежда Самуиловна, кандидат экономических наук, кафедра математического и компьютерного моделирования, Южно-Уральский государственный университет (г. Челябинск, Росиийская Федерация), kolotovans@susu.ru

Поступила в редакцию 31 октября 2023 г.