ISSN 2587-814X (print), ISSN 2587-8158 (online)
Russian version: ISSN 1998-0663 (print), ISSN 2587-8166 (online)

Vol. 17 No. 3 – 2023

BUSINESS INFORMATICS

HSE Scientific Journal

Publisher:
HSE University

The journal is published quarterly

The journal is included into the list of peer reviewed scientific editions established by the Supreme Certification Commission of the Russian Federation

Address:
26-28, Shabolovka Street,
Moscow 119049, Russia

Tel./fax: +7 (495) 772-9590 *28509
http://bijournal.hse.ru
E-mail: bijournal@hse.ru

Circulation:
English version – 100 copies,
Russian version – 100 copies,
online versions in English and Russian – open access

Printed in HSE Printing House
44, build.2, Izmaylovskoye Shosse,
Moscow, Russia

CONTENTS

ABOUT THE JOURNAL

Business Informatics is a peer reviewed interdisciplinary academic journal published since 2007 by HSE University, Moscow, Russian Federation. The journal is administered by HSE Graduate School of Business. The journal is issued quarterly, in English and Russian.

The mission of the journal is to develop business informatics as a new field within both information technologies and management. It provides dissemination of latest technical and methodological developments, promotes new competences and provides a framework for discussion in the field of application of modern IT solutions in business, management and economics.

The journal publishes papers in the following areas: modeling of social and economic systems, digital transformation of business, innovation management, information systems and technologies in business, data analysis and business intelligence systems, mathematical methods and algorithms of business informatics, business processes modeling and analysis, decision support in management.

The journal is included into the list of peer reviewed scientific editions established by the Supreme Certification Commission of the Russian Federation.

The journal is included into Scopus, Web of Science Emerging Sources Citation Index (WoS ESCI), Russian Science Citation Index on the Web of Science platform (RSCI), EBSCO.

The journal is distributed both in printed and electronic forms.

EDITORIAL BOARD

ABOUT HSE UNIVERSITY

Consistently ranked as one of Russia's top universities, HSE University is a leader in Russian education and one of the preeminent economics and social sciences universities in Eastern Europe and Eurasia.

Having rapidly grown into a well-renowned research university over two decades, HSE sets itself apart with its international presence and cooperation.

Our faculty, researchers, and students represent over 50 countries, and are dedicated to maintaining the highest academic standards. Our newly adopted structural reforms support both HSE's drive to internationalize and the groundbreaking research of our faculty, researchers, and students.

Now a dynamic university with four campuses, HSE is a leader in combining Russian educational traditions with the best international teaching and research practices. HSE offers outstanding educational programs from secondary school to doctoral studies, with top departments and research centers in a number of international fields.

Since 2013, HSE has been a member of the 5-100 Russian Academic Excellence Project, a highly selective government program aimed at boosting the international competitiveness of Russian universities.

ABOUT HSE GRADUATE SCHOOL OF BUSINESS

HSE Graduate School of Business was created on September 1, 2020. The School will become a priority partner for leading Russian companies in the development of their personnel and management technologies.

The world-leading model of a 'university business school' has been chosen for the Graduate School of Business. This foresees an integrated portfolio of programmes, ranging from Bachelor's to EMBA programmes, communities of experts and a vast network of research centres and laboratories for advanced management studies. Furthermore, HSE University's integrative approach will allow the Graduate School of Business to develop as an interdisciplinary institution. The advancement of the Graduate School of Business through synergies with other faculties and institutes will serve as a key source of its competitive advantage. Moreover, the evolution and development of the Business School's faculty involves the active engagement of three professional tracks at our University: research, practice-oriented and methodological.

What sets the Graduate School of Business apart is its focus on educating and developing globally competitive and socially responsible business leaders for Russia's emerging digital economy.

The School's educational model will focus on a project approach and other dynamic methods for skills training, integration of online and other digital technologies, as well as systematic internationalization of educational processes.

At its start, the Graduate School of Business will offer 22 Bachelor programmes (three of which will be fully taught in English) and over 200 retraining and continuing professional development programmes, serving over 9,000 students. In future, the integrated portfolio of academic and professional programmes will continue to expand with a particular emphasis on graduate programmes, which is in line with the principles guiding top business schools around the world. In addition, the School's top quality and all-encompassing Bachelor degrees will continue to make valuable contributions to the achievement of the Business School's goals and the development t of its business model.

The School's plans include the establishment of a National Resource Center, which will offer case studies based on the experience of Russian companies. In addition, the Business School will assist in the provision of up-to-date management training at other Russian universities. Furthermore, the Graduate School of Business will become one of the leaders in promoting Russian education.

The Graduate School of Business's unique ecosystem will be created through partnerships with leading global business schools, as well as in-depth cooperation with firms and companies during the entire life cycle of the school's programmes. The success criteria for the Business School include professional recognition thanks to the stellar careers of its graduates, its international programmes and institutional accreditations, as well as its presence on global business school rankings.

DOI: 10.17323/2587-814X.2023.3.7.23

Short-term forecasting of electricity prices using generative neural networks

Andrej S. Kaukin
E-mail: kaukin@iep.ru

Pavel N. Pavlov
E-mail: pavlov@ranepa.ru

Vladimir S. Kosarev
E-mail: kosarev-vs@ranepa.ru

Russian Presidential Academy of National Economy and Public Administration
Address: 82, Vernadskogo Prospect, Moscow, 119571, Russia

Abstract

This article studies the predictive abilities of the generative-adversarial neural network approach in relation to time series using the example of price forecasting for the nodes of the Russian free electricity market for the day ahead. As a result of a series of experiments, we came to the conclusion that a generative adversarial network, consisting of two models (generator and discriminator), allows one to achieve a minimum of the error function with a greater generalizing ability than, all other things being equal, is achieved as a result of optimizing the static analogue of the generative model – recurrent neural network. Our own empirical results show that with a near-zero mean square error on the training set, which is demonstrated simultaneously by the recurrent and generative models, the error of the latter on the test set is lower. The adversarial approach also outperformed alternative reference models in out-of-sample forecasting accuracy: a convolutional neural network adapted for time series forecasting and an autoregressive linear model. Application of the proposed approach has shown that a generative-adversarial model with a given universal architecture and a limited number of explanatory factors, subject to additional training on data specific to the target node of the power system, can be used to predict prices in market nodes for the day ahead without significant deviations.

Keywords: electricity market, day-ahead market, time series, generative neural network, recurrent neural network

Citation: Kaukin A.S., Pavlov P.N., Kosarev V.S. (2023) Short-term forecasting of electricity prices using generative neural networks. *Business Informatics*, vol. 17, no. 3, pp. 7–23. DOI: 10.17323/2587-814X.2023.3.7.23

Introduction

A two-level electricity and capacity market operates on the territory of Russia. The day-ahead market is the main platform where industrial producers purchase electricity and capacity. The Day Ahead Market (DAM) is a system where competitive selection of price bids from suppliers and buyers of electricity is carried out a day before its actual delivery to consumers with the determination of prices and supply volumes for each hour of the day[1]. Marginal pricing is carried out on the DAM. In other words, the price is determined on the basis of the balance of supply and demand and applies to all participants in this market.

The level of electricity prices directly affects the costs of companies. The availability of an accurate forecast of electricity prices is necessary for planning the operating activities of enterprises representing energy-intensive industries, as well as for building financial models. The non-linear nature of electricity price dynamics makes their forecasting a difficult empirical task[2].

There are many empirical works by both domestic and foreign researchers where neural network methods were used to predict the prices of the free electricity market and demonstrated their superiority over linear models. However, recently there is evidence that the use of a generative adversarial networks (GAN[3]) approach can further improve the accuracy of a neural network built and optimized specifically for solving a specific problem.

GAN is both a representative of statistical forecasting methods and simulation methods[4]. An inexperienced expert using a set of available predictors (a generative neural network at the beginning of training) makes predictions about the price of electricity in a

[1] Competitive selection of applications is carried out by a commercial operator ATS. The price of DAM is determined for each node of two price zones: the first price zone includes the territories of the European part of Russia and the Urals (Central, North-Western with the exception of territories belonging to non-price zones, Southern, North Caucasus, Volga and Ural Federal districts); the second is the territory of Siberia (Siberian Federal District).

[2] Significant volatility, a complex structure of seasonality (annual, intra-weekly, intraday seasonality), as well as the presence of frequent emissions are a characteristic feature of the prices of the free electricity market. The elections can be explained by both abnormal situations in the energy system and the conjuncture of consumer industries, which makes it impossible to build a complete structural model of the market

[3] The symbiosis of two neural networks: a generator and a discriminator. The generator is designed directly for making forecasts. The discriminator is designed to distinguish real data from the generator forecast and stimulate the generator to learn more efficiently.

[4] Simulation modeling methods are usually used to represent the structure and connections of elements of complex economic systems, where it is explicitly necessary to take into account the interactions of many agents [1–3].

particular node, and another neural network (discriminator) learns to distinguish them from real data. Networks stimulate each other in the process of learning to more and more accurately solve the individual task assigned to them. Training continues until the expert (generator) forecasts reach the required accuracy in accordance with the selected metric.

The purpose of this work is to test the methodology for constructing generative adversarial networks (GAN) to predict per-node electricity prices in Russia on the day-ahead market, as well as to conduct a comparative analysis of the quality of forecasts based on the GAN methodology and forecasts obtained by alternative (reference) methods.

The following tasks were solved in the work presented here to achieve the goal of the study. First, approaches to electricity price forecasting were summarized, including those using generative adversarial neural networks. Secondly, a description of the data characterizing the wholesale electricity market in Russia was given. Thirdly, methodological approaches to the development of a generative-adversarial neural network model for predicting the price of electricity at the level of individual nodes of the energy system in Russia were characterized. Fourth, an overview of the empirical results of the study was provided, comparing the results of electricity price forecasting using the Generative Adversarial Neural Network (GAN) methodology and the main benchmarks, including the basic recurrent network, convolutional neural network and autoregressive model (ARIMA), widely used in modeling time sequences in the field of economics and finance [4–6]. A summary of the results of the study is given in the conclusion.

1. Analysis of the literature

1.1. Classical approaches to forecasting electricity prices on DAM

Various economic-mathematical and simulation models were used to solve the problem of forecasting electricity prices. For example, domestic scientists in [1] used a simulation model based on CGE (general economic equilibrium model) to search for optimal rates of tariff growth in the electric power industry in the regions of the Russian Federation in the regulated market segment. The system of equations developed by the authors took into account the interaction of many economic agents: consumer, producer, importers and exporters of electricity and the state.

The system of simultaneous equations was used in [7] to predict free market prices. The authors modeled spot electricity prices in Scandinavia using a model based on 29 equations, which took into account functional relationships between climatic factors, snow cover development, river water content (hydroelectric power plants are the main source of electricity in Scandinavia) and power system parameters. In [8], a structural model of spot electricity prices for New England in the United States was developed, taking into account functional relationships between fuel prices, as well as electricity demand and the availability of generating capacities.

The simulation approach is an effective tool for modeling the economy based on an analytical representation of the interaction of various agents and takes into account the physical processes that affect the economy. For example, in [2], a simulation model was presented for predicting the dynamics of oil production by wells, taking into account the implemented and planned geological and technical measures at each well. In another case, a simulation approach was used to develop a digital twin of a TV production plant [3]. However, the use of a simulation approach for forecasting, in the case of node prices for DAM, seems difficult, since it requires detailed information about the topology of the energy network, its parameters and operating conditions.

Economic and mathematical approaches in terms of time series models are more common for free market price forecasting. They can be generalized into two groups: statistical and machine learning methods. Statistical methods usually include additive econometric models [9]. For example, in [5, 6, 10], various variations of autoregressive models (ARIMA, ARMAX, AGARCH) were used to predict free market prices. More and more works began to appear with the development and popularization of machine learning methods, where linear econometric models are compared with such representatives of machine learning methods as support vector models, gradient boosting and neural networks [9, 11, 12] , turned out to be less accurate in short-term price forecasting.

Many researchers have turned to neural networks to predict the prices of the free electricity market [12–14]. Most researchers, as of 2020 [15], preferred the multilayer perceptron[5] (MLP). Domestic researchers have also repeatedly used this architecture to predict DAM prices. For example, Maryasin and Lukashova used MLP with two hidden layers to forecast free electricity prices in the Yaroslavl region. Zolotova and Dvorkin [16] in their study proposed to use a perceptron with 8 neurons in the hidden layer to predict the hourly equilibrium price index of the first price zone.

Many works exist in the foreign literature using other architectures that have proven themselves in the problems of time series forecasting in other areas. For example, a combination of convolutional and recurrent neural networks was used to predict prices and demand for electricity in [14, 17]. Article [14] shows that this architecture has justified itself in many areas where forecasting required the extraction of both temporal and spatial characteristics of time series. The authors of the study [12] proposed using a convolutional neural network with extended convolutions[6] to predict prices on the wholesale electricity market in the Canadian city of Ontario. However, recently there is evidence that the use of a generative adversarial approach can improve the result of a network of any architecture, if this neural network is used as a generator in a GAN [18].

1.2. Generative-adversarial neural networks

1.2.1. General characteristics

The basic theory of generative adversarial networks with examples of practical use can be found in Nikolenko's monograph [19]. A simple generative adversarial network consists of two artificial neural networks that interact with each other in turn. The first is a generator. It spawns objects in the data space. The second is the discriminator. It learns to distinguish objects generated by the generator from real examples from the training sample.

The generator must learn to trick the discriminator, and the discriminator must correctly distinguish between generated examples and real ones. This is the adversarial component in the interaction of two networks. The above description in terms of game theory is a minimax optimization problem, which can be written as Equation 1:

$$\min_G \max_D V(D,G), \text{ where } \max V(D,G) = \\ = E_{x\sim p_{data}(x)}\left[\log D(x)\right] + E_{z\sim p_z(z)}\log\left(1 - D\left(G(z)\right)\right), \quad (1)$$

where $D(x)$ – functional form of the discriminator;

$G(z)$ – functional form of the generator;

$p_z(z)$ – data distribution generated by the generator;

$p_{data}(x)$ – distribution of actual data.

In practice, the functional forms of the discriminator and generator can be any architecture of neural networks. The solution of the minimax problem provides alternate training of the generator with fixed weights of the discriminator and the discriminator with fixed weights of the generator.

Goodfellow and Benji [20] first described and put into practice adversarial networks in 2014. Sub-

[5] Multilayer perceptron is an artificial neural network that is characterized by several layers of input nodes connected in the form of a directed graph between the input and output layers.

[6] Adaptation of a convolutional network for time series forecasting, which allows taking into account a wide range of history when forecasting.

sequently, their idea was widely applied in practice. Adversarial networks have made it possible to achieve significant results in such areas as image generation from a text description [21], the creation of drugs [22], the generation of pseudo-realistic time sequences with the preservation of distribution highlights [23], etc.

For example, in the case of generating an image by description, the text is converted into a numerical feature space using a recurrent encoder, and then these features are used as a condition in the GAN that generates the image. As a result, for example, a person's face in a photograph can first be displayed in the feature space, and then the age feature can be changed and a new image generated. Thus, it is possible to artificially "age" or "rejuvenate" a person.

To create a new drug, the researchers from [22] used an adversarial autoencoder to generate molecules that can be promising candidates for creating new drugs based on them.

Time series are a unique object for generative modeling. In [23], it is noted that time series forecasting models, such as classical or neural network autoregressions, are inherently deterministic. Generative models, in turn, allow us to add an element of randomness to the neural network output.

1.2.2. Application of GAN for forecasting the electricity market

There are a number of examples of the use of generative-adversarial networks for forecasting the electricity market in foreign research practice. For example, in [24], the authors tested the generative-adversarial network model on two data sets: electricity consumption at the level of an individual household and the dynamics of the exchange rate. As a result, the generative-adversarial model in both experiments surpassed in accuracy its deterministic equivalent – a generative neural network that was trained independently.

In [25], researchers propose a model based on generative-adversarial networks for predicting node-by-node prices of a part of the US energy system. The neural network model uses spatial-temporal correlations between historical prices at nodes and accepts historical prices ordered into a three-dimensional tensor as input data. This tensor consists of a series of time-ordered matrices. In turn, each matrix is actually a map of node prices while preserving the spatial location of nodes. The task of the generative model in this case was to generate a new matrix with forecasted node-by-node prices for electrical energy. The basic model was trained to make a forecast for an hour ahead.

In [26], a generative adversarial network is used to forecast wholesale electricity prices with an interval of 30 minutes for the Australian energy market. Unlike previous works, the authors do not build a point estimate of the price, but an interval one. The generative network allowed the authors to obtain predictive intervals covering rare and extreme observations more accurately than alternative stochastic models.

Following the approach described in [24], we use a two-step procedure to develop our own generative-adversarial model. At the first step, we develop and optimize a recurrent neural network to solve the problem of predicting the price of electricity in a random node of the power system. In the second step, we incorporate the resulting neural network into the GAN as a generator and check the stability of the model on a subset of nodes. This approach allows you to narrow the search space of the GAN architecture to a discriminator and makes it possible to test the hypothesis that it is possible to improve the performance of the underlying neural network by including it in the GAN architecture.

2. Data

The analysis is carried out on the basis of hourly reports of the Trading System Administrator on equilibrium prices in the largest nodes of the energy system [27]. The database covers the period from April 13, 2019 to December 31, 2022 and contains information about 7215 nodes in 66 regions of the Russian Federation. The dynamics of averaged prices for all nodes of the energy system of the Russian Federation is shown in *Fig. 1*.

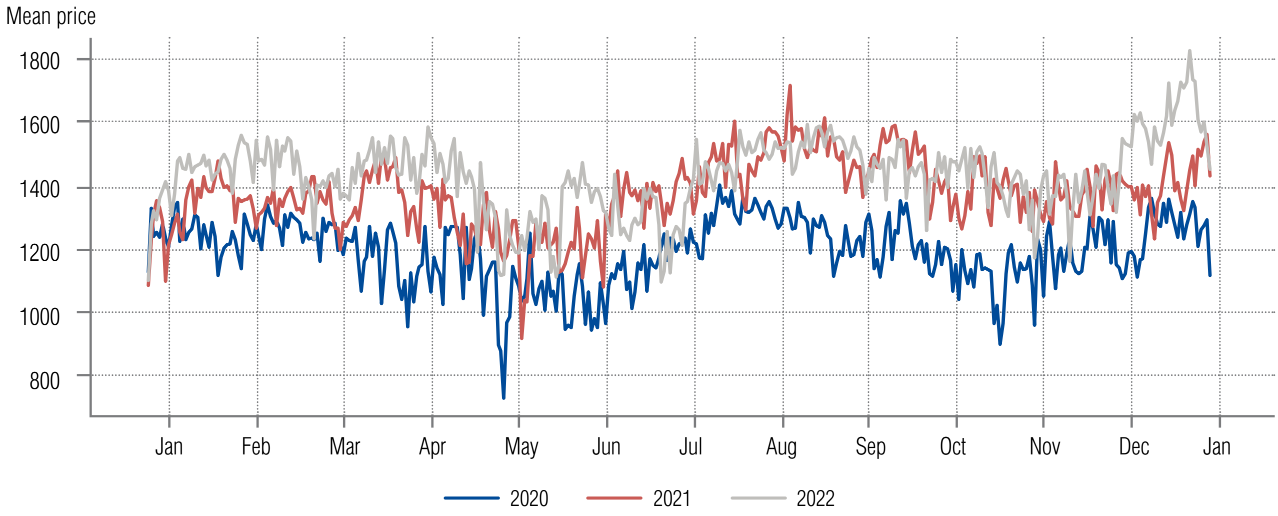


Fig. 1. Average prices for DAM for all nodes of the energy system of the Russian Federation.

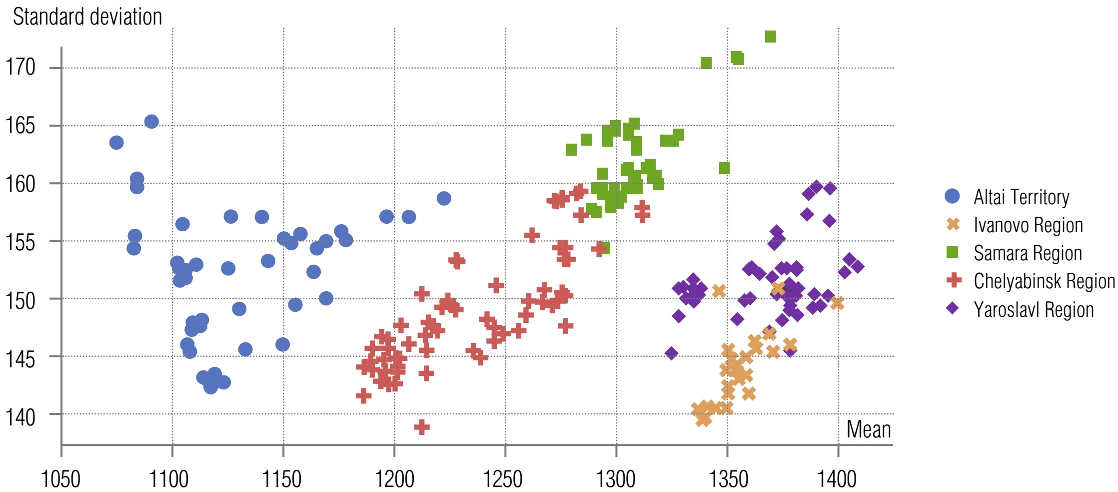


Fig. 2. Clustering of statistical characteristics of nodes within one region.

It is possible to note a tendency to spatial correlation within one region despite the stochastic behavior of each of the node price series. An illustration of this is the clustering of statistical characteristics of prices: the average and standard deviation for the period under review on the example of some subjects of the Russian Federation (*Fig. 2*).

There are a number of features that need to be taken into account when forming a list of explanatory variables:

1. Significant autocorrelation of prices. This phenomenon is described in detail in the work of Zolotova and Dvorkin [16] at the level of price zones. Our own analysis showed that in particular cases, at the node level, the picture is generally similar.

2. Annual seasonality. The seasonal component changes synchronously in most regions of Russia. Price growth is observed in the third quarter, which is associated with the dynamics of the all-Russian production cycle and the beginning of the heating period (the reasons for the annual seasonality are analyzed in more detail in the work of Prokhorova et al. [28]). The exceptions are the Krasnoyarsk Territory and the Irkutsk Region (see *Fig. 3*), in which, on the contrary, there is a decrease in prices in the 3rd quarter, which may be due to the availability of relatively cheap electricity from hydroelectric power plants in the regions.

3. The behavior of prices within the week at the regional level has some relatively constant frequency: in most regions on Monday and Friday prices are at the maximum level, on Saturday – at the minimum. The exceptions are the Irkutsk region and the Republic of Buryatia, where prices on Saturday are on average the highest. However, during the working week, the dynamics may differ from region to region. On public holidays, there is a significant decrease in prices in all regions, with the exception of a number of regions of the North Caucasian Federal District (*Fig. 4*).

4. Nonlinear dependence on temperature. The above-mentioned study by Prokhorova et al. [28] noted the need to take into account both steady changes in temperature dynamics throughout the year affecting the annual seasonality in electricity consumption, and weather factors that take into account deviations from the norm.

3. Research methodology

Analysis of price dynamics in the nodes of the energy system of the Russian Federation made it possible to reveal the presence of spatial autocorrelation. The statistical characteristics of prices differ more and more as the nodes move away, which is associated with changing conditions of supply and demand. For this reason, a predictive model can be specific to:

a) node and take into account the spatial lag (spatial autocorrelation),

b) region and have a multiple output – a forecast simultaneously in all nodes of the regional energy system, where the format of the input data involves taking into account geographic connectivity.

In the second case (option b), a significant number of observations over time is required to obtain a mod-

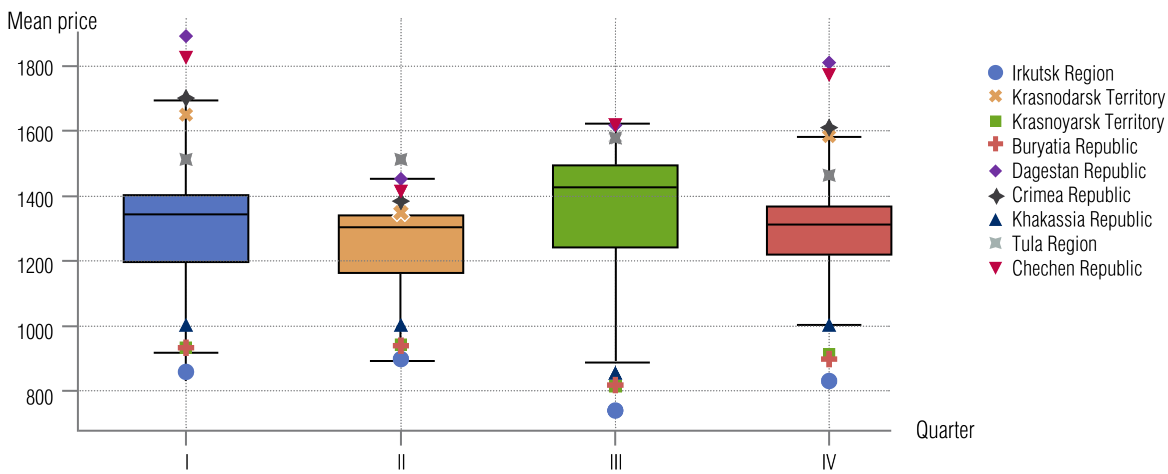


Fig. 3. Diagram of the range of average DAM prices depending on the quarter (the calculation was made for all regions in the sample, emissions are shown for individual regions of the Russian Federation).

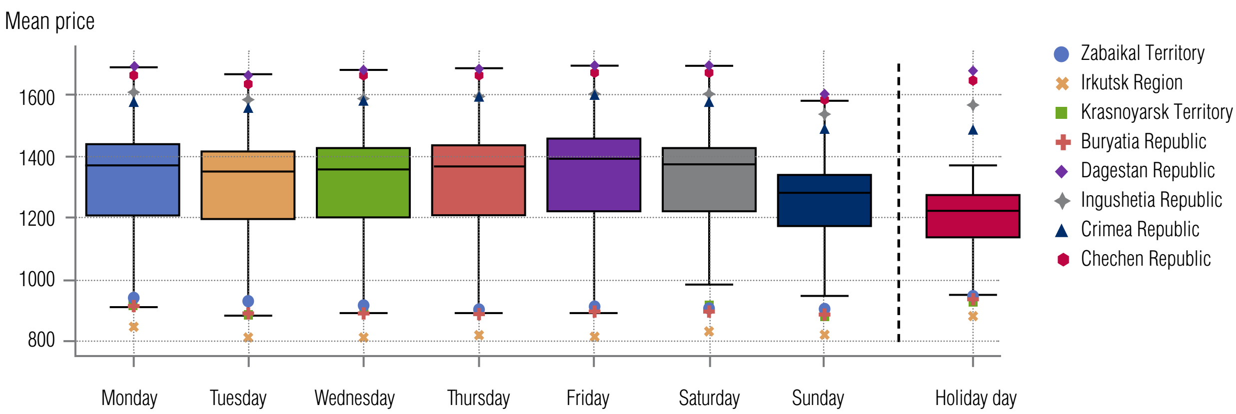


Fig. 4. Diagram of the range of averaged DAM prices by days of the week and holidays (the calculation is carried out for all regions in the sample; emissions are shown for individual regions of the Russian Federation).

el with a high generalizing ability. Unfortunately, the limitations of the available data retrospective on the website of the Trading System Administrator do not allow training such a model. Therefore, in our study, we will focus on the first option (option a), which involves building a universal neural network architecture, which, if trained on node-specific data will allow us to predict prices for any individual node of the energy system without significant error variance.

Taking into account the above data features, the following factors will be used among the explanatory variables:

- lagged values of electricity prices in the target node;
- lag values of electricity prices of the three nearest nodes within the regional network;
- dummy variables for weekends and holidays;
- average daily temperature in the region;
- temperature deviation from moving average.

It is important to note that the node model can also include locality-specific factors, for example, the water level of rivers, which is relevant for the regions of the Urals and Siberia [16], the capacities of industrial consumers, and others. However, the lack of statistics in the required context does not provide this opportunity.

As noted above, at the first step of modeling, a static recurrent neural network is formed which is optimized, and its stability is checked at random nodes of the power system. At the second step of the modeling, a GAN is formed based on the existing architecture of a static recurrent neural network, which is used as a generator. After optimizing the discriminator architecture and GAN hyperparameters, the accuracy of the prediction results of the constructed model is compared with alternative benchmarks.

Initially, all models are trained in parallel on 5 random nodes in each region[7]. Thus, each model is sequentially trained at 330 nodes[8], and the average error and its variance obtained from the test data are used to compare the quality of the models. Accordingly, data preprocessing for all nodes is unified: the sample is divided into training, validation and test samples in the ratio of 80%, 10% and 10%; data standardization and tensor transfor

[7] In order to save computing resources, a limited number of nodes are used for testing the model. This is because each region has a different number of nodes with full data coverage (up to 546 nodes).

[8] 5 nodes · 66 regions.

mation in the format of sliding windows from the original time series (the dimension of the time window is a hyperparameter).

3.1. Building a static generator model (the first step of modeling)

The static model is a two-layer recurrent neural network GRU[1] with 55 cells in the first layer and 20 in the second[2]. The first layer at the output preserves the dimension of the data in time for the next recurrent layer. The second layer transmits a vector of dimension 3 as output data. Thus, the initial matrix of input data of dimension (7 x 8), where 7 is the size of the time window, 8 is the number of explanatory variables, the model maps into a space of dimension (1 x 3). The sum of the elements of this vector is a forecast of the price of electricity for one day ahead (the elements of the vector are summed taking into account the weights, the values of which are selected during the training of the model). The hyperparameters of the model (the size of the time window that determines the number of time lags for all variables, the number of elements of the training sample used to calculate one iteration of gradient descent) were determined empirically as a result of iteration. The criterion was the standard error on the test sample[3].

3.2. Building a GAN (second modeling step)

The generative-adversarial model is a generator and discriminator connected in series (*Fig. 5*). A previously defined model of a recurrent neural network is used as a generator. A noise vector is added to the input data matrix as a separate factor, which prevents the generator from retraining and allows adding a stochastic element to the output of the model.

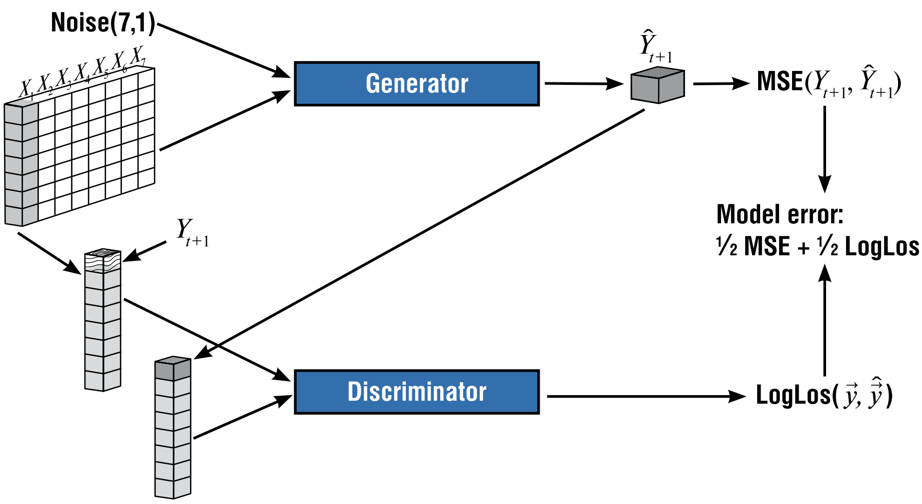


Fig. 5. The device of a generative-adversarial model used to build a forecast of node-by-node electricity prices on the DAM.

[1] Gated Recurrent Units is a type of recurrent neural network designed to model time sequences.

[2] The number of neurons in each layer, the type of activation function and the gradient descent step were determined as a result of enumeration on a given set using the KerasTuner neural network optimization package.

[3] Experiments to determine the underlying architecture were made on the data of a randomly selected node. Later, the model was further trained for each node individually.

The discriminator architecture is a convolutional neural network with one hidden layer that accepts dimension 7 as input. The first 6 elements of the vector are the actual price values in retrospect, and the last element is the generator forecast or the actual price, depending on the template being submitted to the input. The discriminator learns to classify incoming data: (X, Y) – "True" or 1 and $(X, G(X))$ – "False" or 0. The discriminator learns using the error function (Equation 2), which is based on cross-entropy (BCA – binary cross entropy), Equation 3:

$$L_D = l_{bce}(D(X,Y), 1) + l_{bce}(D(X,G(X)), 0), \quad (2)$$

$$l_{bce} = -(y \cdot \log(p) + (1 - y) \cdot \log(1 - p). \quad (3)$$

To train the generator, a two-component error function is used, consisting of cross–entropy and mean squared error (MSE – mean squared error), Equation 4.

$$L_G = \lambda_1 l_{bce}(D(X,G(X)), 1) + \lambda_2 \sum(Y - G(X))^2. \quad (4)$$

The cross-entropy metric is calculated by assigning the "True" label to the vector with the predicted price value. This is necessary to mislead the discriminator and allows the generator to pick up weights in the learning process that complicate the task of training the discriminator in the subsequent iteration, which introduces an element of "competition" into the learning process. It is important to note that theoretically adding the MSE component to the error function is not strictly necessary [29]. We are adding MSE for faster generator convergence and shorter training time.

As mentioned above, the discriminator and generator are trained alternately until the specified accuracy of the price forecast is achieved. This provides a solution to the original minimax problem. A general description of the entire learning process is shown in pseudocode (*Table 1*).

More detailed information about the algorithm of generative adversarial network learning is contained in [30, 31].

Thus, the discriminator learns to distinguish generated samples from real data in the learning process. At the next iteration of training, the generator strives to improve its forecast so that the probabilistic output of the model is closer to the "Truth" with fixed discriminator weights. This approach allows you to change the gradient trajectory in the learning process and come to a different optimum in comparison with the static model, where only the root-mean-square error is used [32].

The results of the GAN model are compared with "reference" alternative models: ARIMA autoregression, basic recurrent neural network, two-layer convolutional neural network (CNN). The architecture of the latter is based on the model proposed in [33].

4. Empirical results

We trained the discriminator five times more than the generator during each training iteration. This allows the discriminator to be somewhat ahead of the generator in accuracy and distinguish the generated data from the real ones. *Fig. 6* shows the process of learning a generative-adversarial model for some random node.

The left axis shows the total error of the model, which is the weighted sum of the error of the discriminator and the generator (mse + logloss). The value of the discriminator error at each training iteration (logloss) is shown on the right axis.

In order to test the stability of the proposed model, the following experiment was conducted: 5 nodes of the power system were randomly selected in each region. The only selection criterion is the absence of gaps in the data on the analyzed time horizon. Next, a set of explanatory variables was formed for all nodes in accordance with a unified procedure, data was preprocessed, and then the previously saved generative model was further trained on the data of each of the 5 nodes. Based on the results of the experiment, the error of the electricity price forecast was calculated on a test sample (in parallel, training and error calculation were carried out for a static generator model and other benchmarks).

Table 1.

General algorithm of generative-adversarial neural network training

Setting conditions: gradient descent step ρ_D, ρ_G; error function weights parameters for the generative model λ_1, λ_2; random initialization of weights in discriminator and generator models.

While (until the algorithm converges):

Discriminator Training (*D*):

Getting M samples from a training sample: X^{train}:

$(X, Y) = (X^1, Y^1), \ldots, (X^m, Y^m) \subset X^{train}$

Stochastic gradient descent step and updating of weights *D* at fixed weights *G*::

$$W_D = W_d - \rho_D \sum_{i=1}^{M} \frac{\partial L_D\left(X^{(i)}, Y^{(i)}\right)}{\partial W_D}$$

Generator Training (*G*):

Getting *M* new samples from X^{train}:

$(X, Y) = (X^1, Y^1), \ldots, (X^m, Y^m) \subset X^{train}$

Stochastic gradient descent step and updating of weights *G* at fixed weights *D*:

$$W_G = W_G - \rho_G \sum_{i=1}^{M} \frac{\partial L_G\left(X^{(i)}, Y^{(i)}\right)}{\partial W_G}$$

End While

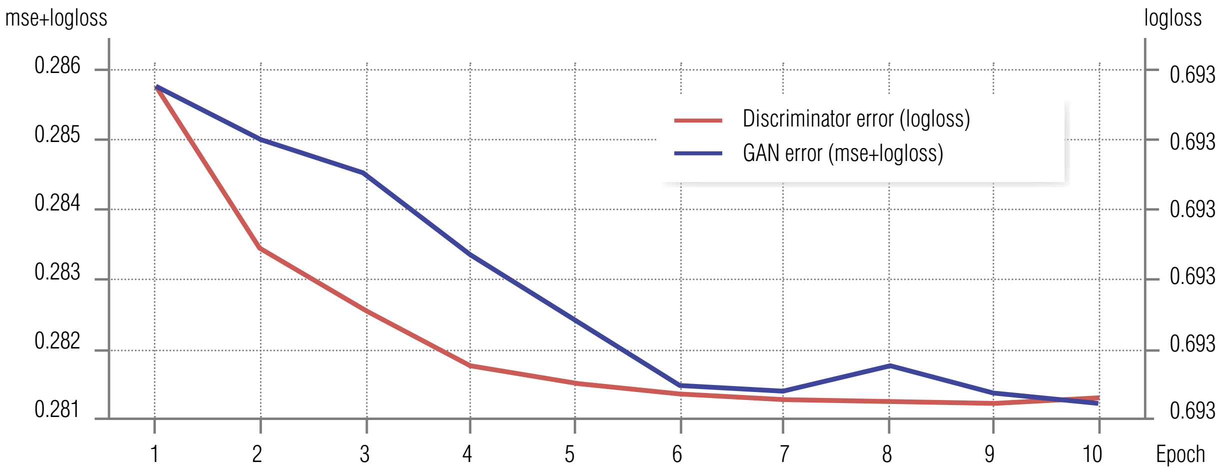


Fig. 6. The value of errors of the generative-adversarial model in the learning process on a training sample.

The results of the experiment showed that the generative-adversarial model demonstrates the smallest average error on the test sample and the minimum spread of error values for individual nodes of the power system. *Table 2* shows the average errors on the test sample of the models under consideration and their standard deviations.

In [25], the authors associate the superiority of GAN with the ability to provide the required gradient for optimizing the generator during training: the gradient directed by the discriminator allows one to achieve a wider minimum than, other things being equal, is achieved as a result of optimizing the static analogue of the generative model. Our own empirical results also support this thesis, since with a near-zero error on the training set achieved simultaneously by a static generator model and a GAN, the error of the latter on the test set is lower.

The average errors of the GAN model and their standard deviations are visualized in *Fig. 7*: the color gradation on the conditional map of regions corresponds to the value of the average error calculated for five randomly selected nodes in the corresponding region of the Russian Federation (see the left part of *Fig. 7*) and the variance of average errors (see the right part of *Fig. 7*).

The model demonstrates the best results in most of the regions of the Central Federal District, the Northwestern Federal District and the Volga Federal District. In turn, the model demonstrates the largest average error in such regions as the Republic of Tatarstan, Bashkiria, a number of regions of Siberia and the Caucasus. The high variance of the average error is also manifested mainly in these regions. The unsatisfactory performance of the model in these regions seems to be due to the insufficiency of a set of explanatory factors, which in all cases was standard. It is possible that energy bridges with other countries connected with the domestic energy system are important for pricing in border regions; for the central regions of Siberia, the water level of rivers is an important factor, since a significant share of generation falls on hydroelectric power plants.

It is also important that the test part of the sample (from August 19, 2022 to December 31, 2022) accounts for a period with a clear structural shift – a sharp increase in electricity prices at the end of 2022 in regions such as Irkutsk, Tyumen and Tomsk Regions, Krasnoyarsk Krai, Stavropol Krai, Krasnodar Krai. The sharp increase in the price trend in this case was due to conjunctural factors, the influence of which could not be present in the training sample.

Table 2.

Comparative table of averaged errors and their spread when applying the model to five random nodes in the analyzed regions

Indicator		Error on the test sample			
		GAN	GRU	CNN	ARIMA
MAE	mean	0.0497	0.062	0.0681	0.0724
	std	0.0002	0.0009	0.0063	0.0092
MAPE	mean	0.092	0.0973	0.1055	0.1114
	std	0.0005	0.0018	0.0113	0.0242
MSE	mean	0.0046	0.0059	0.0073	0.0082
	std	0	0.0002	0.0013	0.0034

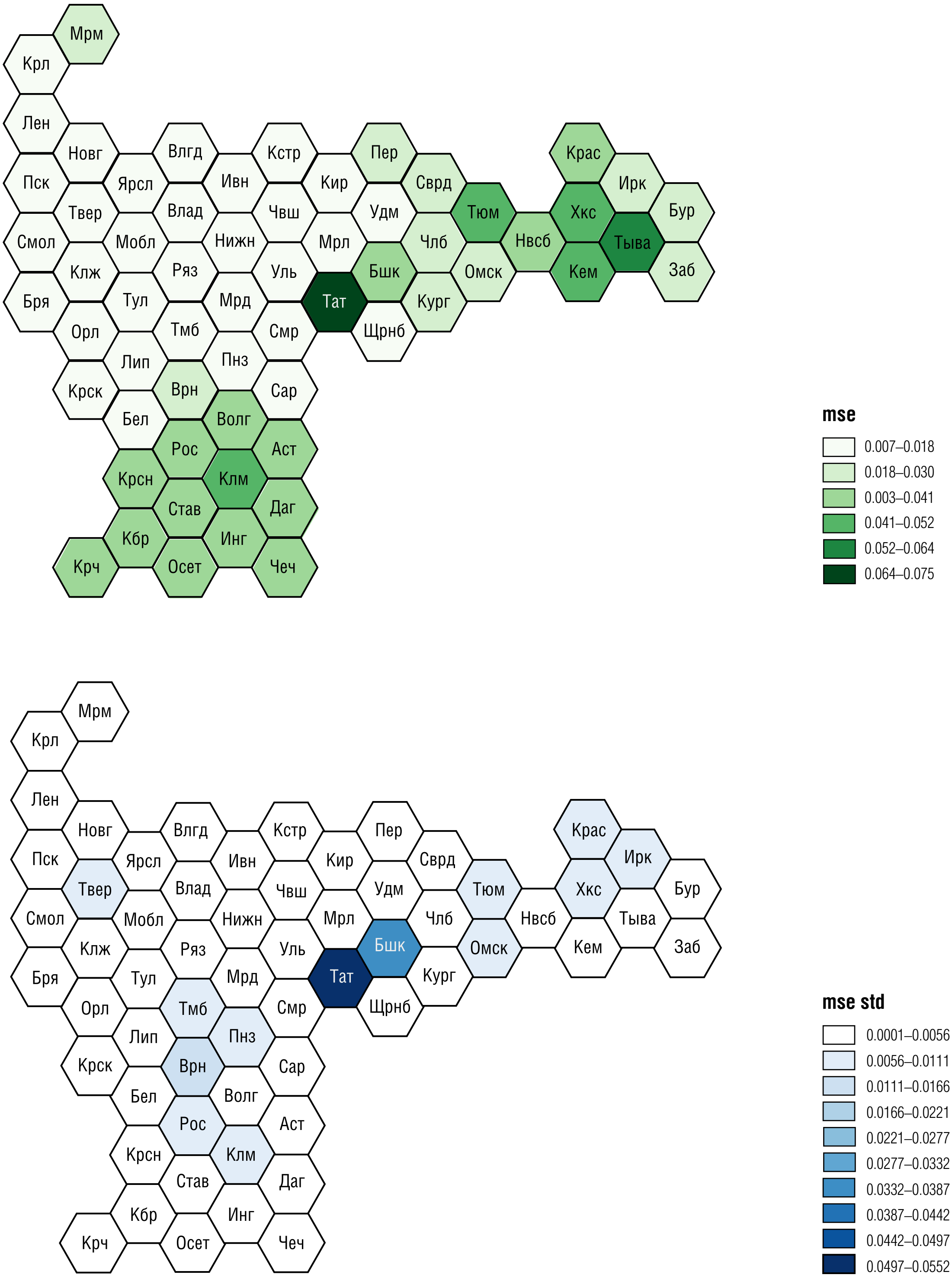


Fig. 7. The average error of the generative-adversarial model in the test sample (graph below) and its variance (right graph) for five random nodes in the region.

The hypothesis of the group equality of the Levene variances for the root-mean-square errors obtained at randomly selected nodes of the power system within a single region. The null hypothesis of the test is that "all subsamples have equal variances". When outliers are excluded (12 regions with the largest standard deviation of error), the null hypothesis of the test is not rejected for the remaining 54 regions. Thus, for most regions, the proposed architecture of the generative model is universal and is able to provide a relatively low error on the test sample for various nodes of the country's power system.

Conclusion

The results of the study demonstrate that the proposed generative-adversarial network can be used to predict prices in DAM nodes for the day ahead without significant deviations in accuracy for 54 of the 66 regions of Russia under consideration. The network architecture is universal (it does not change during the transition from region to region of the Russian Federation) and uses a limited number of explanatory factors. The network needs to be retrained on data specific to the target node. The neural network model included the following set of variables: historical price values in the target and geographically close to it nodes of the power system (node-by-node prices in the electricity market are correlated both spatially and in time), ambient temperature and seasonal factors.

The proposed generative-adversarial model reduced the mean square error by 22% on a test sample of a static generator model based on a recurrent neural network, and also surpassed the quality of alternative benchmark models: convolutional neural network and autoregressive linear model (ARIMA). ■

References

1. Akopov A.S., Beklarjan G.L. (2005) Analysis of the effectiveness of the regulatory policy of the state using the regional CGE model of the behavior of natural monopolies (on the example of the electric power industry). *Jekonomicheskaja nauka sovremennoj Rossii*, no. 4(31), pp. 130–139 (in Russian).

2. Akopov A.S., Beklarjan A.L., Hachatrjan N.K., Fomin A.V. (2017) Oil production dynamics forecasting system using simulation modeling. *Informacionnye tehnologii*, vol. 23, no. 6, pp. 431–436 (in Russian).

3. Makarov V.L., Bahtizin A.R., Beklarjan G.L., Akopov A.S. (2021) Digital plant: methods of discrete-event modeling and optimization of production characteristics. *Business Informatics*, vol. 15, no. 2, pp. 7–20. https://doi.org/10.17323/2587-814X.2021.2.7.20

4. Hachatrjan N.K., Kravchenko T.K., Akopov A.S., Uvarova O.M. (2016) Forecasting of the main indicators of the Russian stock market by autoregressive models with distributed lags. *Audit and financial analysis*, vol. 3, pp. 128–133 (in Russian).

5. Gonzalez J.P., San Roque A.M., Perez E.A. (2017) Forecasting functional time series with a new Hilbertian ARMAX model: Application to electricity price forecasting. *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 545–556. https://doi.org/10.1109/TPWRS.2017

6. Zhao Z., Wang C., Nokleby M., Miller C.J. (2017) Improving short-term electricity price forecasting using day-ahead LMP with ARIMA models. *2017 IEEE Power & Energy Society General Meeting*, pp. 1–5. https://doi.org/10.1109/PESGM.2017.8274124

7. Vehviläinen I., Pyykkönen T. (2005) Stochastic factor model for electricity spot price – the case of the Nordic market. *Energy Economics*, no. 2, pp. 351–367.

8. Howison S., Coulon M. (2009) Stochastic behaviour of the electricity bid stack: from fundamental drivers to power prices. *The Journal of Energy Markets*, pp. 29–69.

9. Weron R. (2014) Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, vol. 30, no. 4, pp. 1030–1081. https://doi.org/10.1016/j.ijforecast.2014.08.008

10. Garcia R.C., Contreras J., Akkeren M., Garcia J.B. (2005) A GARCH forecasting model to predict day-ahead electricity prices. *IEEE Transactions on Power Systems,* vol. 20, no. 2, pp. 867–874. https://doi.org/10.1109/TPWRS.2005.846044

11. Lynch C., O'Leary C., Sandareshan P.G., Akin Y. (2021) Experimental analysis of GBM to expand the time horizon of Irish electricity price forecasts. *Energies*, vol. 14, no. 22, 7587. https://doi.org/10.3390/en14227587

12. Deng Z., Liu C., Zhu Z. (2021) Inter-hours rolling scheduling of behind-the-meter storage operating systems using electricity price forecasting based on deep convolutional neural network. *International Journal of Electrical Power & Energy Systems*, vol. 125, 106499. https://doi.org/10.1016/j.ijepes.2020.106499

13. Zhang J., Tan Z., Yang S. (2012) Day-ahead electricity price forecasting by a new hybrid method. *Computers and Industrial Engineering*, vol. 63, pp. 695–701. https://doi.org/10.1016/j.cie.2012.03.016

14. Kim T.Y., Cho S.B. (2019) Predicting residential energy consumption using CNN-LSTM neural networks. Energy, vol. 182, pp. 72–81. https://doi.org/10.1016/j.energy.2019.05.230

15. Mar'yasin O.Yu., Lukashov A.I. (2020) Prediction of free electricity prices using neural networks. *Neiroinformatika-2020. Sbornik nauchnykh trudov. XXII mezhdunarodnaya nauchno-tekhnicheskaya konferentsiya*, pp. 294–303 (in Russian).

16. Zolotova I.Yu., Dvorkin V.V. (2017) Short-term forecasting of prices in the Russian wholesale electricity market based on neural networks. *Problemy prognozirovaniya*, no. 6, pp. 47–57 (in Russian).

17. Guo X., Zhao Q., Zheng D., Ning Y., Gao, Y. (2020) A short-term load forecasting model of multi-scale CNN-LSTM hybrid neural network considering the real-time electricity price. *Energy Reports*, vol. 6, pp. 1046–1053. https://doi.org/10.1016/j.egyr.2020.11.078

18. Wu W., Huang F., Kao Y., Chen Z., Wu Q. (2021) Prediction method of multiple related time series based on generative adversarial networks. *Information*, vol. 12, no. 2, 55. https://doi.org/10.3390/info12020055

19. Nikolenko S., Kadurin A., Arkhangel'skaya E. (2019) *Glubokoe obuchenie*. Sankt-Peterburg, Piter (in Russian).

20. Goodfellow I., Pouget-Abadie I., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. (2014) Generative adversarial networks. *arXiv*:1406.2661. https://doi.org/10.48550/arXiv.1406.2661

21. Zhang H., Zhang H., Yu Koh J., Baldridge J., Lee H., Yang Y. (2021) Cross-modal contrastive learning for text-to-image generation. Proceedings of the *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 833–842.

22. Kadurin A., Nikolenko S., Khrabrov K., Aliper A., Zhavoronkov A. (2017) DruGAN: An advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico. *Molecular Pharmaceutics*, vol. 14, no. 9, pp. 3098–3104. https://doi.org/10.1021/acs.molpharmaceut.7b00346

23. Yoon J., Jarrett D., van der Schaar M. (2019) Time-series generative adversarial networks. *Neural Information Processing Systems*.

24. Koochali A., Dengel A., Ahmed S. (2021) If you like it, GAN it — Probabilistic multivariate times series forecast with GAN. *Engineering Proceedings*, vol. 5, no. 1, 40. https://doi.org/10.3390/engproc2021005040

25. Zhang Z., Wu M. (2020) Real-time locational marginal price forecasting using generative adversarial network. *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids*, pp. 1–6. https://doi.org/10.1109/SmartGridComm47815.2020.9302938

26. Kabir H.M., Khosravi A., Nahavadini S., Kavousi-Fard A. (2019) Partial adversarial training for neural network-based uncertainty quantification. *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 5, no. 4, pp. 595–606. https://doi.org/10.1109/TETCI.2019.2936546

27. *Otchet o ravnovesnykh tsenakh v naibolee krupnykh uzlakh raschetnoi modeli*. Administrator Torgovoi Sistemy. Available at: https://www.atsenergo.ru/nreport?rname=big_nodes_prices_pub&region=eur (accessed 02.03.2023) (in Russian).

28. Porokhova N.V., Rudakov E.N., Saakyan Yu.Z. (2009) Alternative methods for monitoring the dynamics of industrial production. *Problemy Prognozirovaniya*, no. 4, pp. 36–53 (in Russian).

29. Zhang K., Zhong G., Dong J., Wang S., Wang Y. (2019) Stock market prediction based on generative adversarial network. *Procedia Computer Science*, vol. 147, pp. 400–406. https://doi.org/10.1016/j.procs.2019.01.256

30. Mathieu M., Couprie C., LeCun Y. (2015) Deep multi-scale video prediction beyond mean square error. *arXiv*:1511.05440. https://doi.org/10.48550/arXiv.1511.05440

31. Radford A., Metz L., Chintala S. (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv*:1511.06434. https://doi.org/10.48550/arXiv.1511.06434

32. Staffini A. (2022) Stock price forecasting by a deep convolutional generative adversarial network. *Frontiers in Artificial Intelligence*, vol. 5, 837596. https://doi.org/10.3389/frai.2022.837596

33. Liu S., Ji H., Wang M.C. (2020) Nonpooling convolutional neural network forecasting for seasonal time series with trends. *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 8, pp. 2879–2888. https://doi.org/10.1109/tnnls.2019.2934110

34. Izmailov P., Maddox J., Kirichenko P., Garipov T., Vetrov D., Gordon A. (2020) Subspace inference for Bayesian deep learning. Proceedings of the *35th Uncertainty in Artificial Intelligence Conference. Proceedings of Machine Learning Research*, vol. 115, pp. 1169–1179.

About the authors

Andrej S. Kaukin

Cand. Sci. (Econ.);

Head of the Laboratory of Industry Markets and Infrastructure, Gaidar Institute for Economic Policy, 3–5, Gazetnyy Lane, Moscow 125993, Russia;

Head of the Laboratory of System Analysis of Industrial Markets, Institute of Applied Economic Research, Russian Presidential Academy of National Economy and Public Administration, 82, Vernadskogo Prospect, Moscow 119571, Russia;

E-mail: kaukin@iep.ru

ORCID: 0000-0003-2892-5278

Pavel N. Pavlov

Cand. Sci. (Econ.);

Senior researcher, Laboratory of System Analysis of Industrial Markets, Institute of Applied Economic Research, Russian Presidential Academy of National Economy and Public Administration, 82, Vernadskogo Prospect, Moscow 119571, Russia;

E-mail: pavlov@ranepa.ru

ORCID: 0000-0001-6200-6481

Vladimir S. Kosarev

Research Fellow, Laboratory of System Analysis of Industrial Markets, Institute of Applied Economic Research, Russian Presidential Academy of National Economy and Public Administration, 82, Vernadskogo Prospect, Moscow 119571, Russia;

E-mail: kosarev-vs@ranepa.ru

ORCID: 0000-0001-5435-9076

DOI: 10.17323/2587-814X.2023.3.24.37

Influence of algorithmization and interface for the preparation of management decisions*

Rimma D. Gutgarts
E-mail: gutgarc@gmail.com

Irkutsk National Research Technical University
Address: 83, Lermontov St., Irkutsk 664074, Russia

Abstract

In modern conditions, managerial decision-making is carried out using automated systems under the general name "Decision Support Systems" (DSS). When creating them, it is important to consider two key points. The first is the algorithmic component, which reflects the logic of the system as a whole and its individual parts. The second is the application interface through which the user interacts with it. The interface is a graphical interpretation of the algorithms that are implemented within the system. Therefore, it is very important to design and create such a relationship between the algorithm and the interface so that the user is as comfortable as possible using the DSS to solve current tasks (information input, its processing, presentation and analysis for decision making). Thus, there is a directly proportional relationship between the interface and the algorithm. Moreover, despite the fact that there are many studies on these aspects, both theoretical and practical, there are

* The article is published with the support of the HSE University Partnership Programme

still questions to which one should pay attention to in terms of application. The purpose of this study is to formulate practical recommendations to prevent the entry of incorrect information into the DSS database and to present the results in a form convenient for its analysis. The main tasks of the work are to show by means of examples which errors can contribute to the entry of unreliable information into the database, as well as how best to present information on the monitor screen in accordance with the psychophysiological characteristics of a person in order to reduce the time for its analysis and decision-making.

Keywords: decision support systems, features of algorithmization, interrelation of the interface and algorithms, presentation of information, error handling when entering information

Citation: Gutgarts R.D. (2023) Influence of algorithmization and interface for the preparation of management decisions. *Business Informatics*, vol. 17, no. 3, pp. 24–37. DOI: 10.17323/2587-814X.2023.3.24.37

Introduction

Key indicators in making managerial decisions are that they are reasonable and timely. This can only be achieved by using relevant information that meets all of its classical requirements: reliability, consistency, completeness, relevance, value, understandability, accessibility, etc.

Therefore, on the one hand, it is very important to algorithmically track and eliminate all potential errors at the stage of initial input of information into the database, since subsequent processing of incorrect information will inevitably generate incorrect results. Their analysis, in turn, will lead to wrong conclusions in decision making.

On the other hand, the aspect of presenting information for analysis and decision-making is no less important. And this depends on the features of the interface, through which the visual interpretation of the results of processing information intended for decision making is carried out.

Algorithmization and interface are very multifaceted areas. In the article, attention is paid only to certain points related to some applied issues, the knowledge of which will allow decision makers (DM) to take them into account when formulating tasks for the preparation of analytical information, as well as to reduce the time for its analysis in the future.

The presentation of materials on the subject of managerial decision-making can be considered in two interrelated aspects. The first aspect is related to issues of an organizational and methodological nature (in particular, the mathematical apparatus used, the features of the application of system analysis, organizational interaction between interested levels of management). The second aspect is specialized software, i.e. the decision support system (DSS) focused on the processing of information for analysis and decision making. Currently, they are based on applications in the form of expert systems and artificial intelligence [1]. In addition, such canonical forms of analytical content as tables (formed according to certain rules) and graphs remain in demand. They may reflect: the results of information processing within the framework of expert systems or artificial intelligence systems, business intelligence systems or any other calculations performed in the conditions of any specialized systems.

1. About the role of information in the decision-making process

Kravchenko and Isaev give 13 stages in the decision-making process [2, p. 21–26]. The first stage consists in the awareness of the decision maker of the need to make a decision (non-formalizable part). The second and third stages ("Obtaining information" and "Analysis of information", respectively), from which, in fact, the process itself begins, are associated with the information aspect. The subsequent stages define the classical actions in the course of system analysis, allowing one to come to the choice of a solution.

Speaking about the use of a computer in solving applied problems in [3, p. 17], it is noted that among the relevant software (software) for the subject area "Economic activity" there are accounting programs, programs for calculating estimates, evaluating the effectiveness of investment projects, real estate valuation, etc. In addition, the spreadsheet MS Excel can be used quite effectively.

In [4] it is indicated that the term "Management Decision" (DM) "... is used in two main meanings – as a process and as a phenomenon. As an SD process, there are the following basic procedures: information preparation, development of options, coordination of options, choice of one option, approval, implementation, control over the implementation of SD and informing the initiator of the decision. As a phenomenon of SD, there is a set of measures aimed at resolving the economic problem under consideration in the form of a resolution, order or instruction given orally or in writing" [4, p. 22].

In [5], there is a special section devoted to information support for decision-making, where information for the designated purposes is presented as a key factor in the context of its canonical features, including its modern interpretations in the Internet format.

Tabekin notes that one of the factors determining the quality of management decisions is "... the volume and value of the information available; for the successful adoption of a managerial decision, the main thing is not so much the amount of information as its value (relevance) and timeliness, combined with the level of professionalism, experience, intuition of personnel making and implementing a managerial decision ..." [6, p. 36].

Krushanov draws attention to the fact that "... early cybernetics was characterized by two diverse controversial tasks ...", the first of which was that generalized control processes and the phenomenon of information were singled out as new subjects of scientific knowledge. This is designated as the management-informational aspect of cybernetics. The author understands management as "... purposeful information impact, carried out according to the feedback scheme" [7].

From the above definitions, we can conclude that information is the "main character" in solving problems related to analysis and management decisions.

The decision-making process itself is based on the following key points:

1. Information provided in a timely manner, presented in a correct and easy-to-understand form.
2. Mathematical tools that allow you to process the original information in order to interpret it for decision making.
3. Software for automated processing of initial information according to the applied algorithms.
4. Non-formalizable aspects that depend, for example, on the intuition of decision makers, on their competence in interpreting special information depending on the final or intermediate goals of the analysis.

2. Impact of errors when entering information into the database to its subsequent processing for decision-making

Using the example of *Table 1*, we will comment on some types of algorithmic errors that may occur during the initial formation of the Database (DB) and subsequently provoke the appearance of incorrect results in order to process information for analysis and decision making.

Table 1.

Types and examples of errors (based on [8, pp. 38–39])

View errors	Error example	Comment
1. Errors in data recording	1. Recorded 5 instead of 4. 2. Instead of code 36, denoting enterprise "A", code 35 was introduced.	These are random "manual" errors. It is very difficult to identify them. It is possible only by chance or in the context of subsequent decisions, in which incorrectly entered information will affect the results and attract the attention of the analyst. The second error is possible even if the code (or company name) is selected from the drop-down list.
2. Not installed data	A payment was received from the buyer under No. 153, but the buyer with this number was not found in the list of accounts payable.	This type of error can occur for two reasons. First: since the payment is generated "on the side" and there is no way to check the correctness of its filing, then it will not be possible to enter information from it into the database. If the document is critically important and its contents cannot be ignored, then there is only one way out: through communication channels, contact the source of information and clarify the issue "manually." Second: the list of accounts payable was formed incorrectly. Solutions: 1) reshape the list; 2) check the list generation algorithm; 3) check the database (perhaps the mistake in the account number was "manual").
3. Distortion of the regulated limit.	For certain types of deliveries for buyers, there are restrictions on the quantity of the corresponding product within one purchase (for example, 100 products). But if a manager has issued two orders to one buyer (for example, 75 items each), the total volume of which exceeds the established norm, then this buyer will be able to receive a quantity of goods that exceeds the established limit value.	Incorrect algorithm. When trying to issue a second order for more than 100 items, the system should either prompt the clerk to issue an order for 25 items, or block his actions when trying to issue a second order for 75 items and inform the user about this.
4. Missed entries	The entry in the database was deleted for some reason.	Such an error may be the cause of error #2. Actions related to the removal of certain information should be a lgorithmically strictly regulated. For example, apply the question "Are you sure you want to delete?". Or deleted information may be in the "recycle bin" for some time with the possibility of restoring it if necessary.
5. Errors when generating reports	The decision maker, when analyzing the report, assumes that it includes information on sold and paid products. However, in fact, in addition to paid goods, the document contains goods sold on credit.	Similar errors can occur quite often due to an incorrect description of the task for programming. For the formation of reporting (analytical, statistical) documents, it is advisable to always offer their graphic structure with the correct names of the columns, and, if possible, it is necessary to determine the details of the database, which are the basis for compiling the document.

View errors	Error example	Comment
6. Data Entry Errors	For some reason, the increase in wages for an employee with personnel number 174 is recorded in the database for an employee with personnel number 175.	On the one hand, such an error can be "manual". On the other hand, deliberate. In any case, if the employee knows about the increase in salary, then the first time after the increase, the error will be revealed. If the employee is not notified of the salary increase, then the situation relates to the moral and ethical sphere.
7. Repeated mistake	An incorrect price for a product has been entered into the database, therefore products of this type are billed for payment at this incorrect price.	The error could be "manual" or intentional. Since the source for entering this kind of information is usually a paper document, its scanned copy can always be saved for some time in special folders or files. Periodically, you can perform "manual" checks. Knowing this, the employee is unlikely to take risks.
8. Incorrect differen-tiation by periods	Information about the order was entered into the database on the last day of the month. But the actual shipment of goods under this order was carried out only a few days later in the next month. However, information about the sale is reflected in the database for the previous month.	The reason for the error is an incorrect algorithm. This is a typical example of ignoring exception handling. The decision should be based on compliance with the conditions specified by the potential user.
9. Data falsification	An employee of the organization, who has access to the financial information of other employees and has the rights to change information, deleted records of unpaid invoices of his colleague, which provided the latter with the opportunity not to pay for purchased goods and (or) services.	The reason for the error is the incorrect differentiation of access rights for an employee when working with information. Removing information is always a critical aspect. Therefore, all actions related to removal are subject to strict regulation and control. This can be done algorithmically through a procedure for logging (journaling) jobs.
10. Incorrect accounting	Money that was given to the buyer's representative for one purpose was spent on completely different purposes (for example, instead of paying for tools for work, a business dinner at a restaurant was paid for).	These errors can be detected either by chance, or by the results of the analysis of reporting documents, or during the audit. Since the source for entering this kind of information is probably a paper document, it is always possible to determine a logical relationship between its details. And when filling out the corresponding form on the screen, algorithmically reveal such a connection. If a recognition procedure is used to enter information from a document, then a special algorithm can also be applied that determines the correlation between logically related attributes.

The examples of errors given in the table refer to the subject area "Accounting". However, it is not difficult to project them to any other subject areas.

3. Algorithmic features of the interface

More than 20 years ago, Nielsen proposed 10 basic principles that must be considered when designing user interaction with a system [9]. These principles are universal in nature. They can be considered classic, so they have not lost their significance at the present time. Ignoring these principles can provoke various kinds of problems when using various types of automated information systems, including decision support systems.

Let us briefly comment on the principles that should be reflected in the algorithmic context and that a potential user can take into account when formulating functional requirements for decision support systems.

Principle 1: Visibility of the state of the system. The duration of the solution of essentially different tasks may be different. There are tasks, the solution of which is carried out so quickly that, at the current speed of information processing on modern computers, the result is issued almost instantly after pressing the Enter key. However, there are other tasks (for example, drawing up production plans, calculating some indicators for a large number of employees, processing significant amounts of statistical information to make forecast estimates, etc.), the solutions of which can be quite lengthy. In this case, it is useful for the user responsible for completing the task to provide visual information about its chronology. This information can be either discrete or continuous. You can use a horizontal indicator, circular indicators (hours or sectoral filling), or some other.

Principle 2: Correspondence between the system and the real world. It should always be remembered that the DSS is intended for use in a specific subject area in which specific terminology is necessarily present. Therefore, all information presented to the user on the monitor screen must fully comply with those semantic meanings (concepts, definitions, designations, names of objects, indicators, coefficients, processes, situations, phenomena, dependencies, etc.) that are accepted in this area. The interface should not contain terminology typical for the IT field, with the exception of procedures and actions that have become "de facto" standards (copy, print, cut, save, etc.), provided that their terminological replacement is inappropriate. In all other cases, it is always necessary to find an analogy between the terminology of the IT sphere and the subject area. For example, the interface should not contain the word "Database," since in reality it may correspond to the "Chart of Accounts," "Personal Files of Employees," "Production Equipment," "Payment Schedule," etc.

Principle 3: User control and freedom. The user should not be afraid to make a mistake. The system should always "insure" him. It could be:

- a warning message (for example, "Are you sure you want to delete the entry about the employee with personnel number ...?," "After changing the indicator "Ω," it will be impossible to restore its initial value," "Check the time synchronization with the Internet," "You did not specify the calculated period," "The accuracy of calculations should not exceed 2 decimal places");
- context prompt when entering information (for example, "Attribute value must not exceed 10 characters," "Enter only numeric data");
- the ability to select a value from a drop-down list when filling in a field (using the necessary directories, dictionaries, classifiers and other ordered sets of certain objects), which significantly reduces the likelihood of entering an incorrect value (in the DSS, this can be used to make a request for obtaining the necessary information);
- the ability to return to previous stages of processing.

Principle 4: Consistency and standards. Within the framework of one system, all used names of objects (processes, phenomena, situations, etc.) must be unified. For these purposes, in particular, classifiers, reference books, dictionaries are created, which are usually referred to as reference information. Ordered sets of corresponding values make it possible to save time when entering information, minimizing errors, optimizing the amount of computer memory when storing the entered data, and unifying the required information in all documents reflecting any kind of analytics (*Fig. 1*).

Technologically, the use of values from classifiers or their analogues is carried out through drop-down lists.

Principle 5: Prevention of errors. A well-designed interface should either prevent the user from making mistakes altogether, or minimize their occurrence. When entering information, this can be done, for example, by context prompts, selecting values from the proposed lists, using masks or templates. For data of the type "year" (when entering a value manually), it is advisable to specify an additional range of change, since an incorrectly entered value may be "outside" the allowable historical period.

To indicate dates and (or) time periods, as a rule, modern tools are used that allow you to almost completely eliminate the error when choosing dates. The date in the "DD.MM.YYYY" format can be selected either from the built-in calendar, or "assembled" from parts (day, month, year) – from sequences of logically reasonable proposed numbers. The user only needs to specify the desired values. The error in this case may be accidental due to inattention.

It is enough to simply detect an error in the boolean type indicator (i.e., it can take one of two values – "0" or "1").

In any case, "manual" input, especially for text and (or) character information, should be kept to a minimum.

Principle 6: Recognition, not recollection. The user should feel comfortable and safe when working with the interface. Therefore, he needs to create conditions that will not require him to remember the exact sequence of any of his actions when performing certain tasks. He must always be sure that the system will either prompt him with a simple logical way out when he makes a mistake, or provide a link to the appropriate fragment of the instructions for working with the application.

Principle 7: Flexibility and efficiency of use. In any decision support system, there may always be settings that the user can customize to his individual preferences if he uses the relevant information to solve only his current tasks. For example, in the 1C:Enterprise system, there are 76 (!) interface objects for which the user can independently change their color scheme according to their requirements [10]. If the information (in particular, the generated report) is transferred to the external part of the system in which settlements are carried out, or to a system located outside the perimeter of the relevant organization, then it must comply with accepted corporate standards.

The remaining three principles ("Aesthetic and minimalistic design," "Helping users to recognize, diagnose and correct errors" and "Help and documen-

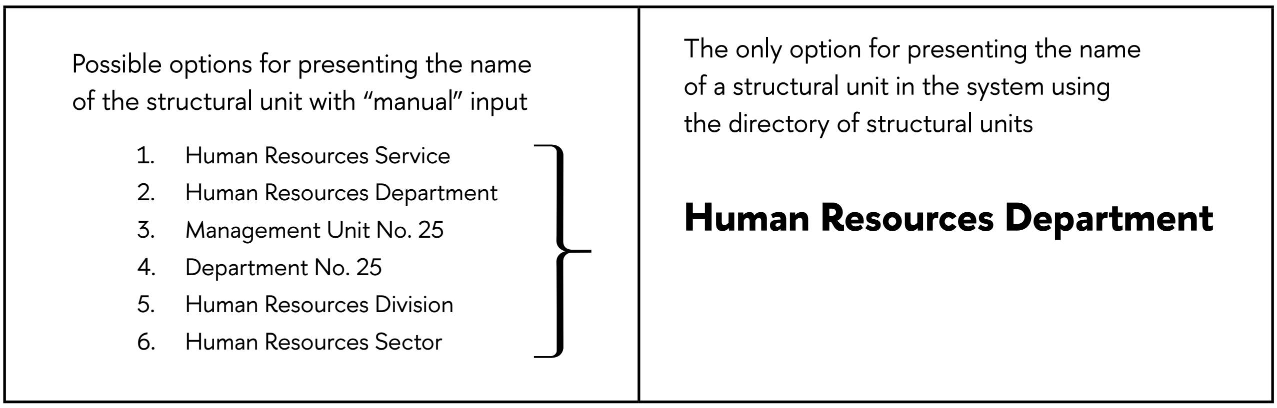


Fig. 1. Illustration for unification of the name of the management unit in the enterprise (in the organization).

tation") are not directly related to algorithmization and therefore do not affect the formulation of user functional requirements. Professional application developers should implement these principles "by definition."

4. Features of the use of text and symbolic information in the interface

Information intended for analysis must meet certain requirements. This is explained by the fact that working in a virtual environment should not create discomfort due to the psychophysiological characteristics of a person.

The interface, being a virtual environment from the user's point of view, should not contradict his actions in the real world. This is due to the general principles of human mental activity (in the broad sense), as well as to the peculiarities of information perception through the senses. Weinshank reflects such aspects in [11].

As an example, consider the impact of different fonts and markers that can be used to represent insights.

There is a very wide variety of fonts. More than 80 types are presented on the resource [12].

Here are some recommendations for using some fonts [13]:

1. Serif fonts (the most popular – Times New Roman) have a universal purpose and can be used for a wide variety of text. But it is especially important that they represent the text well even when using a small font size and are therefore suitable for displaying a large amount of text.
2. Sans-serif or sans-serif fonts (examples: Arial, Tahoma, Verdana) and monospaced fonts (example: Courier New) are recommended for headings or selections of text, but not for text as a whole.
3. Fonts stylized as handwriting make it difficult to read large texts. Therefore, they are recommended as headings or highlighting fragments of text.
4. Decorative (for example, Gothic or Old Slavonic) should be used with extreme caution, as they do not correspond to the present and therefore can be perceived with difficulty.
5. An outline font, a font with a shadow, as well as underlined text are not recommended for large text fragments, as this makes it difficult to read (the clarity of the text is lost, which complicates its perception and understanding).
6. There is such a thing as "reverse," when, for example, white letters are located on a black background. Using this technique for long text presented in small print is also not very comfortable and will make it difficult to read. However, this technique can be used for headers.

There are special free resources on the Internet for converting text to any font, for example, [14, 15].

The provision of public information, especially within the framework of functional software (SW), must comply with the general rules for its perception. As an example, we present the text located at the link [16] (*Fig. 2*). On *Fig. 3* the same text is transformed into a form convenient for perception and understanding. This was done by structuring information, i.e. splitting the text into fragments and highlighting the same type of semantic meanings in the form of lists. The difference in the presentation of the same information is obvious.

As markers for lists, it is recommended to use symbols filled inside (for example: ■ ● ♦ ◆ –), i.e. contrasting in relation to the text, since their style is fundamentally different from letters, and the essence of the marker is precisely to highlight the elements of the list. In contrast to this, it is not recommended to use symbols that are "empty" inside for the same purposes, since in their outline they resemble letters (for example: ❖ ☐ ❑ ❒ ☒ ☐ ○), although they are not. However, the clarity of the lists in this case will be "blurred."

Equally important is the use of icons and pictures. You can either draw your original ones, or choose only those that are presented on special free resources. You cannot copy ready-made drawings, pictures, icons, logos or their fragments if they are copyrighted. Otherwise, it will be contrary to the Law of the Russian Federation "On Trademarks, Service Marks and Appellations of Origin."

Program description:
"Orakl–Kadry" («ORAKUL – Kadry Frames»)

"HR document management and personnel management system" is the most powerful personnel program under Microsoft Windows 95/98/NT4/2000/ME/XP, covering all aspects of personnel records and document management. The program was created with the help of leading HR managers in St. Petersburg and has been popular for over 3.5 years. The program will take over everything that you previously had to do manually, sorting through the "mass of paper." "Orakl-K" is distinguished by simplicity and ease of management; it is universal and easy to use for any computer user, both experienced and novice. The "standard version" includes: a password at the entrance, 4 databases (regular, personnel reserve, dismissed, archive) with employee registration cards (more than 130 topics & windows!), maintaining all records (timesheets, vacations, business trips, sick leave, loans, mat .help, compensation, etc.), a huge library of personnel template documents, automatic creation of orders, customizable staffing (of any complexity!), Your personal notes, all military records, experience, analysis block, more than 33 thousand (!!!) sample-reports (i.e. finding any information).

Fig. 2. Example of "unreadable" text (fragment) [16].

Description of the program: "Orakl–Kadry" ("ORAKUL–Kadry")

"HR document management and personnel management system" is the most powerful personnel program under Microsoft Windows 95/98/NT4/2000/ME/XP, covering all aspects of personnel records and document management.

The program was created with the help of leading HR managers in St. Petersburg and has been popular for over 3.5 years. The program will take over everything that you previously had to do manually, sorting through the "mass of paper." "Orakl-K" is distinguished by simplicity and ease of management; it is universal and easy to use for any computer user, both experienced and novice.

The "standard version" includes:

- password at the entrance, 4 databases (regular, personnel reserve, dismissed, archive) with employee registration cards (more than 130 topics & windows!);
- maintaining all accounting (time sheets, vacations, business trips, sick leave, loans, financial assistance, compensation, etc.);
- a huge library of HR document templates;
- automatic creation of orders;
- customizable staffing (of any complexity!);
- Your personal notes;
- all military records;
- seniority;
- block of analysis;
- more than 33 thousand (!!!) sample-reports (i.e., finding any information);
- blocks of testing and certification and much, much more.

Both standard delivery and complete set under your specific order are possible. The program takes into account the requirements of the Decree of the Government of the Russian Federation and the corresponding Decree of the State Statistics Committee. For non-standard solutions, an additional 35 standard expansion modules or custom-made functions are available.

Fig. 3. Converted text example (fragment).

5. Features of presenting information for analysis and decision-making

Since the quality of decision-making is directly dependent on the information used, it is very important to visualize it in a certain form. On the one hand, it should really contain only those indicators (dependencies, calculation results, statistics, etc.) that are necessary for analysis and management decision-making in specific situations. On the other hand, the presentation of information should be so clear that the time spent on its consideration is minimal, no matter how quickly the decision needs to be made.

The scenario of the dialogue within the framework of the DSS may include the indication of some operational conditions for the selection of information and its processing. These include, for example:

- period of time (from ... to ...);
- structural subdivision (production or management of any level);
- a group of goods or services;
- sales market (district, city, region, country);
- price or value characteristics;
- method of delivery of goods;
- ordering in ascending or descending order of indicators;
- summing up intermediate and (or) final results;
- marking results that go beyond the established limits (up or down) or have deviations that exceed the regulated limits.
- etc.

The dialogue scenario may also provide for the need and (or) variability of a subsequent decision, depending on the intermediate results obtained either at a certain stage of the decision, or before a regulated point in time.

As it is known, any resulting information for analysis can be obtained in two ways. First: document structures are known and therefore the algorithm for their formation is pre-programmed. In this case, it is enough for the decision maker to select the desired document name (in the broad sense, since in addition to the traditional table, the document can be presented, for example, in the form of a graph, "flat" text or an animated "picture") from the proposed list. Second: the structure of the document form is set by the user himself immediately before the formation of the desired document. To do this, the user must have the skills to compose a request for the selection of information (i.e., to know the special query language provided for in a particular DSS), which allows one, without special programming, to quickly generate any form of documents at the request of the user himself.

In tabular documents intended for analysis, there should be no "information noise" which does not carry any semantic load but on a subconscious level distracts the user's attention and slows down his work. An example is shown in *Fig. 4*.

The option of getting rid of this kind of "information noise," which makes it difficult to perceive information, is quite simple. For the given table, you need to add in the name of the corresponding column, separated by a "comma," the value (in the example, this is the unit of measurement), which is repetitive (*Fig. 5*).

Computer technologies make it possible to single out separate fragments of analytical documents in a special way which helps to attract the attention of analysts. For these purposes, you can use, for example, changing the font (by style, color, size), "painting" or framing certain cells (rows, columns) of the table, as well as other features. It is advisable to use such a "design" in order to show values that are significantly different from others, a change in some indicators by a critically permissible value, an indicator value approaching a regulated limit, etc.

In some cases, color coding may be used. This will be useful, for example, in the case when one analytical document simultaneously needs to reflect the different statuses of objects and (or) their characteristics. Thus, within the framework of one document, in particular, goods can be noted: awaiting shipment, shipped, awaiting payment, paid.

When designing tables, one feature must be taken into account. If it turns out that in one of the columns

Year Result: 2022 Smartphone shipments						
Rank	OEM	Y2022		Y2021		YoY
		Shipment	M/S	Shipment	M/S	
1	Samsung	259	21%	272	20%	-4.8%
2	Apple	231	19%	235	18%	-1.3%
3	Xiaomi	152	13%	190	14%	-20.0%
4	Oppo Group	107	9%	144	11%	-25.6%
5	vivo	98	8%	134	10%	-27.1%
6	Transsion	68	6%	75	6%	-8.7%

Fig. 4. An example of an analytical document (excerpt) containing “information noise” [17].

Year Result: 2022 Smartphone shipments						
Rank	OEM	Y2022		Y2021		YoY, %
		Shipment	M/S, %	Shipment	M/S, %	
1	Samsung	259	21	272	20	-4.8
2	Apple	231	19	235	18	-1.3
3	Xiaomi	152	13	190	14	-20.0
4	Oppo Group	107	9	144	11	-25.6
5	vivo	98	8	134	10	-27.1
6	Transsion	68	6	75	6	-8.7

Fig. 5. Sample policy paper (detail) without “information noise.”

in all lines there is the same value (textual, digital, logical, “date” or some other), then this contradicts one of the properties of information – informativeness. In such cases, it is advisable to reflect the name of such a column in the heading of the table and exclude this column from the structure of the table. The table will become more compact and easy to analyze.

It is known that in some cases, for analytical information, a graphical form will be more preferable and visual. It can exist as the only option for presenting relevant information or as an addition to the classic table. A good example of analytics and statistics are the static yearbooks published by the Higher School of Economics. On *Fig. 6* and *7* such examples are shown.

Conclusion

The subject, reflecting theoretical and practical aspects in the field of DSS, is of interest to different groups of specialists. The main part of the publications is devoted to general issues of a canonical nature in this subject area, for example [20], the use of spe-

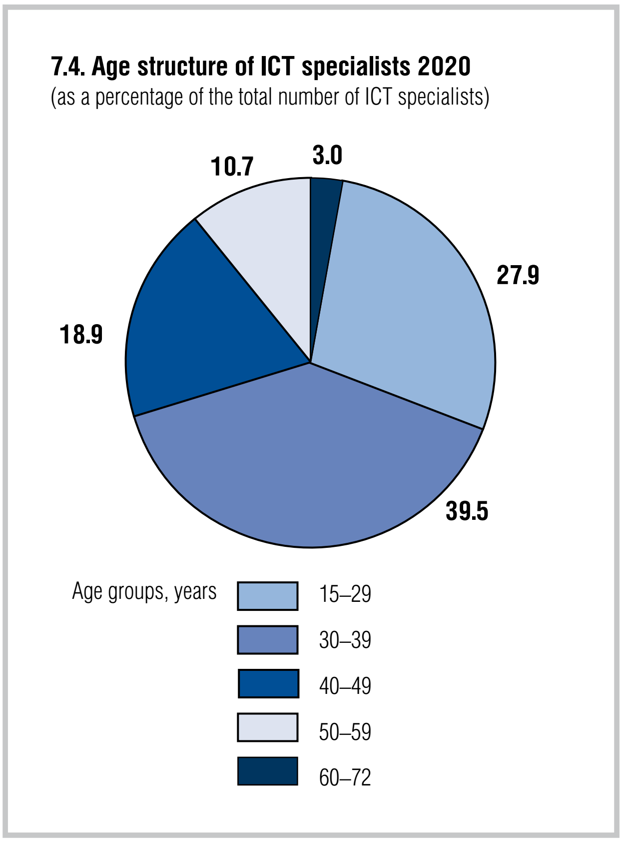


Fig. 6. Example 1 of graphical representation of analytical information [18, p. 80].

cific methods and models, a description of the application and related features of DSS in various functional areas (medicine [21–23], specialized universities [24], defense technology [25], coal preparation enterprises [26], forestry [27–28], transport [29], financial issues [30–31], etc.), as well as a description of the DSS used in the form of specific software.

Nevertheless, insufficient attention is paid to applied issues that must be taken into account both when formulating functional requirements and evaluating DSS by potential users, and when designing and developing systems of this kind by developers. This article deals with certain issues in these aspects.

Decision support algorithms and the accompanying interface are dynamic and evolving areas that include many nuances that are often overlooked by users and developers. This is what can create problems in the application of various information systems, including DSS. Therefore, the issues discussed in the article concerning the algorithmization and interface can be considered with varying degrees of depth and differentiation, as well as a critical analysis of DSS present on the market, which may be the subject of further research.

The materials of this article may be useful to specialists who act as decision makers. They can use them when formulating functional requirements for the development of DSS on order, as well as when evaluating the interface of already existing systems acquired by an enterprise or organization to solve their tasks of supporting managerial decision-making.

The information presented in the article can also be applied in the applied aspect by students, undergraduates and graduate students who are engaged in research in the field of managerial decision-making and (or) the creation of a DSS.

Certain points outlined in the article may be of interest to DSS developers. ■

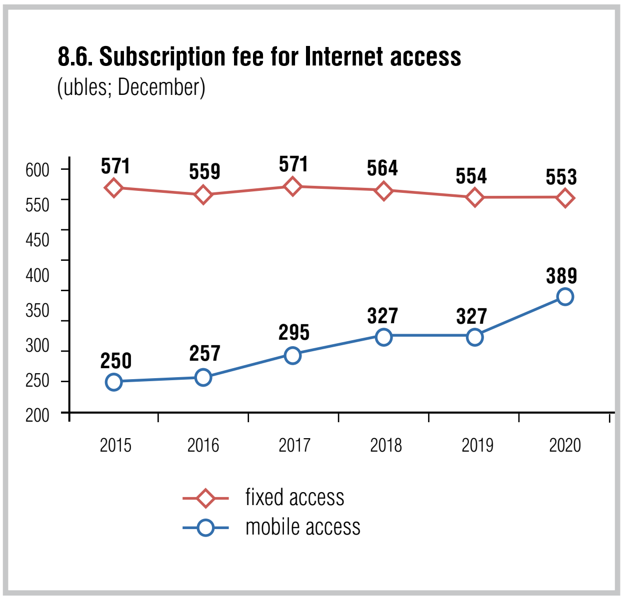


Fig. 7. Example 2 of graphical representation of analytical information [18, p. 93].

References

1. Demidovskij A.V., Babkin E.A. (2021) Integrated neurosymbolic decision support systems: problems and opportunities. *Business Informatics*, vol. 15, no 3, pp. 7–23. https://doi.org/10.17323/2587-814X.2021.3.7.2

2. Kravchenko T.K., Isaev D.V. (2022) *Decision support systems: textbook and workshop for universities*. Moscow: Urait (in Russian).

3. Nabatova D.S. (2023) *Mathematical and instrumental methods of decision support: textbook and workshop for universities*. Moscow: Urait (in Russian).

4. Khalin V.G. et al. (2023) *Theory of decision making. In two volumes. Vol. 1: textbook and workshop for universities*. Moscow: Urait (in Russian).

5. Khalin V.G. et al. (2023) *Theory of decision making. In two volumes. Vol. 2: textbook and workshop for universities*. Moscow: Urait (in Russian).

6. Tabekin A.V. (2023) *Methods of making managerial decisions: a textbook for universities*. Moscow: Urait (in Russian).

7. Krushanov A.A. (2017) The concept of «management» in the cybernetic context. *Vox. Filosofskiy zhurnal* (electronic journal), vol. 23 (in Russian). https://doi.org/10.24411/2077-6608-2017-00020

8. Perry W. (1986) *Computer and the organization of accounting*. Moscow: Finance and statistics (in Russian).

9. Nielsen J. (1994) *10 Usability Heuristics for User Interface Design*. Nielsen Norman Group. Available at: https://www.nngroup.com/articles/ten-usability-heuristics/ (accessed 24 July 2023).

10. *1C:Enterprise. Demo applications*. Available at: https://v8.1c.ru/podderzhka-i-obuchenie/demonstratsionnye-konfiguratsii/ (accessed 24 July 2023).

11. Weinshank S. (2015) *100 main principles of design: how to keep attention*. St. Petersburg: Piter (in Russian).

12. *Free fonts*. Available at: https://ffont.ru/en/fonts (accessed 24 July 2023).

13. Massel G.G. (2007) *Psychological aspects of the development of computer systems*. Melentiev Energy Systems Institute Siberian Branch of the Russian Academy of Sciences. Irkutsk: ESI SB RAS (in Russian).

14. *Beautiful fonts online*. Available at: https://beautifulfonts.net/en/ (accessed 24 July 2023).

15. *Beautiful fonts online*. Available at: https://prettyfont.net/ (accessed 24 July 2023).

16. *Kazus.ru*. Available at: http://kazus.ru/programs/viewdownloaddetails/kz_0/lid_4350.html (accessed 07/24/2023).

17. *Smartphones (Global Market)*. TAdviser. Available at: https://tadviser.com/a/e.php?id=47257 (accessed 24 July 2023).

18. Gokhberg L., Kuzminov Y.I., et al. (eds.) (2022) *Digital economy, 2022: Pocket data book*. Moscow: HSE.

19. Zaslavskaya V.L. (2022) Decision support systems and their role in information management systems. *Economics and management: problems, solutions*, vol. 2, no. 12, pp. 144–153. https://doi.org/10.36871/ek.up.pr2022.12.02.017

20. Ahmed A., Spiess B., Kortsmit J., van den Ham R., Erdoes G., Klein A. (2023) A narrative review of clinical decision support systems for perioperative bleeding management in cardiac surgery. *Journal of Cardiothoracic and Vascular Anesthesia*, vol. 37, no. 9, pp. 1804–1812. https://doi.org/10.1053/j.jvca.2023.05.008

21. Toffaha K.M, Simsekler M.C.E., Omar M.A. (2023) Leveraging artificial intelligence and decision support systems in hospital-acquired pressure injuries prediction: A comprehensive review. *Artificial Intelligence in Medicine*, vol. 141. 102560. https://doi.org/10.1016/j.artmed.2023.102560

22. Malykh V.L. (2019) Decision support systems in medicine. *Program systems: Theory and applications*, vol. 10, no. 2(41), pp. 155–184. https://doi.org/10.25209/2079-3316-2019-10-2-155-184

23. Vecherskaya S.E. (2021) Selection of criteria for a decision support system for an art university. *Informatics and Education*, no. 3, pp. 56–62. https://doi.org/10.32517/0234-0453-2021-36-3-56-62

24. Orkin V.V., Nesterenko O.E., Platonov S.A. (2021) A model of a situational control system in an automated decision support system. *Questions of defense technology. Series 16: Technical means of countering terrorism*, no. 1–2(151–152), pp. 40–45.

25. Lepilo N.N., Katan K.S. (2019) Improving the efficiency of a coal preparation enterprise based on the DSS. *Economic Bulletin of the Donbass State Technical University*, no. 2, pp. 52–59.

26. Punit B., Reddy A.S., Dave T.N. (2023) A spreadsheet-based decision support system for selection of optimal soil liquefaction mitigation technique. *Decision Analytics Journal*, vol. 6, 100154. https://doi.org/10.1016/j.dajour.2022.100154

27. Aldea J., Bianchi S., Nilsson U., Hynynen J., Lee D., Holmström E., Huuskonen S. (2023) Evaluation of growth models for mixed forests used in Swedish and Finnish decision support systems. *Forest Ecology and Management*, vol. 529, 120721. https://doi.org/10.1016/j.foreco.2022.120721

28. Kaltsidis A., Ketikidis K., Basbas S., Aifadopoulou G., Grau J.M.S. (2023) A decision support system for taxi drivers. *Transportation Research Procedia*, vol. 69, pp. 123–130.

29. Siconolfi S.F. (2022) An analytics-based decision support system for evaluating the fiscal health of academic programs. *Decision Analytics Journal*, vol. 4, 100091. https://doi.org/10.1016/j.dajour.2022.100091

30. Borrero-Domínguez C., Escobar-Rodríguez T. (2023) Decision support systems in crowdfunding: A fuzzy cognitive maps (FCM) approach. *Decision Support Systems*, 114000. https://doi.org/10.1016/j.dss.2023.114000

About the author

Rimma D. Gutgarts

Dr. Sci. (Econ.);

Professor, Institute of Information Technologies and Data Analysis, Irkutsk National Research Technical University, 83, Lermontov St., Irkutsk 664074, Russia;

E-mail: gutgarc@gmail.com

ORCID: 0000-0001-9881-1976

DOI: 10.17323/2587-814X.2023.3.38.52

Application of measures of heavy-tailedness in problems for analysis of financial time series

Lilia A. Rodionova
E-mail: lrodionova@hse.ru

Elena D. Kopnova
E-mail: ekopnova@hse.ru

HSE University
Address: 11, Pokrovsky Bulvar, Moscow 109028, Russia

Abstract

An important feature when working with financial data is the fact that the residuals of GARCH-models often have fatter tails than the tails of a normal distribution due to the large number of "outliers" in the data. This requires more detailed study. Kurtosis and quantile-based measure of heavy-tailedness were analyzed and compared in the article in relation to the problem of choosing the GARCH(1,1)-model specification. The data of indices of the Moscow Exchange were considered for the period from April 01, 2019 to February 22, 2022. Kurtosis values ranged from 3 to 52. Empirical data showed that kurtosis was very sensitive to "outliers" in the data, which made it difficult to make assumptions about the distribution of model residuals. The approach considered in this paper based on the heavy-tailedness measure made it possible to justify the choice of degrees of freedom of the t-distribution for the model residuals to explain the fat tails in financial data. It was found that GARCH(1,1)-models with $t(5)$-distribution in the residuals are common.

Keywords: GARCH, kurtosis, quantile-based measure of heavy-tailedness, t-distribution of residuals, degrees of freedom, fat tails

Citation: Rodionova L.A., Kopnova E.D. (2023) Application of measures of heavy-tailedness in problems for analysis of financial time series. *Business Informatics*, vol. 17, no. 3, pp. 38–52. DOI: 10.17323/2587-814X.2023.3.38.52

Introduction

As is known, many financial time series are characterized by certain regularities: asset returns are weakly stationary, volatility clustering is observed, distribution normality is rejected in favor of a distribution with thick tails, etc. [1]. To describe and forecast the processes with such properties, wide use is made of the class of models with conditional heteroscedasticity (ARCH, GARCH models) proposed by Angle [2] and Bollerslev [3] and their modifications. An important feature when working with financial data which we would like to consider in detail in this article is the fact that the residuals of ARCH/GARCH models have fatter tails than the tails of a normal distribution due to the large number of "outliers" in the data, and this fact requires more detailed study. To account for fat tails in econometric practice, several alternative distributions had been proposed: Student's t-distribution [3, 4], generalized error distribution (GED) [5, 6], Student's skew t-distribution [7], etc. Note that the possibility of choosing the Student's t-distribution and GED-distribution for estimating the GARCH model is implemented in econometric packages (for example, Stata16). This is of practical interest in substantiating the choice of the appropriate distribution in modeling and forecasting. The proposed distributions differ in properties; therefore, these distributions will not equally well characterize the "thickness" of the tails of the distribution. Thus, the *research problem* is how to choose the type of distribution that best characterizes the heavy-tailed distribution. The correct specification of the GARCH model, taking into account heavy tails, allows us to get more accurate forecasts of returns and the maximum profit for investors. This fact determines the *relevance* of the study.

The main *goal* of this paper is to analyze the behavior of the quantile-based measure of tailedness in relation to the choice of degrees of freedom of the Student's t-distribution in the residuals of the GARCH model. Note that the measures of heavy-tailedness are widely discussed in foreign literature and are an alternative approach to choosing the number of degrees of freedom of the t-distribution. Let us check how applicable these measures are in econometric practice for analyzing financial data and compare them with the classical approach of choosing the degree of freedom of the t-distribution based on a comparison of maximum likelihood estimates.

1. Measurement of heavy-tailedness

In this section, we will analyze what approaches are used for measuring the "thickness" of the tail of the distribution. The heavy-tailedness of a distribution for a random variable (r.v.) X is usually understood as

$$P(|X| > x) \sim \frac{C}{x^{\xi}}, \tag{1}$$

where $C, \xi > 0$ are the constants, $f(x) \sim g(x)$ means:

$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = 1.$$

The parameter ξ is commonly called the tail index of the distribution X. It characterizes the decay rate of the tails of the power-law distribution (1) and the probability of observing extreme values of the r.v. As the probability mass in the tails increases, the tail index parameters decrease, and vice versa [8].

Note that in the literature, heavy-tailed distributions are divided into three subclasses: fat-tailed distributions, long-tailed distributions, and subexponential distributions [9, 10]. A fat tail distribution exhibits more skewness or kurtosis than a normal distribution. The terms “fat tail” and “heavy tail” are often used as synonyms in financial analysis papers. In our work, we will use the term “heavy tail”, and we will consider the “fat tails” of the distribution as a special case of “heavy tails.” In practice, we have the question of how to measure the heavy-tailedness, and how to assess the degree of “heaviness” of the tail of the distribution. There are parametric and nonparametric approaches to estimating the tail index [11]. In our article, we will analyze the “heavy-tailedness” in the context of time series modeling based on GARCH models. Kurtosis is one of the measures used to detect “outliers” in time series. In 1905 Pearson introduced the concept of kurtosis through 4th order moments:

$$K = \frac{\mu_4}{\sigma^4}, \tag{2}$$

where μ_4 is the central moment of the 4th order;

σ^4 is the square of the variance.

All distributions were classified as plateaucurtic, mesocurtic or leptokurtic with respect to normal [12]. For a normal distribution $K = 3$, in connection with which an excess kurtosis (modified kurtosis indicator) is often used:

$$E_k = \frac{\mu_4}{\sigma^4} - 3.$$

Kurtosis K in the form (2) will be used further in the article. Fat tails (as a special case of heavy tails) are characterized by excessive kurtosis $K > 3$, and the distribution is called leptokurtic [13].

In this paper, we will compare and explore the residuals of GARCH(1,1) models as often used in econometric practice [14]. Recall the definition. The process ε_t follows the generalized autoregressive conditional heteroscedasticity or GARCH(1,1)-model if $\varepsilon_t = \sigma_t z_t$, $t = 1, 2, \ldots$, where $z_t \sim N(0,1)$ is independent normally distributed random variables, and the conditional variance of the process has the form:

$$\sigma_t^2 = \omega + \gamma\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2. \tag{3}$$

Note that in practice, model (3) only partly explains the fat tails, and it is necessary to refine the specification of the distribution of residuals. Student’s t-distribution is often used as an alternative to the normal distribution [3, 4]. The standardized Student’s t-distribution with zero mean and unit variance has a density:

$$f(z_t, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{1}{2}\nu\right)\sqrt{\pi(\nu-1)}}\left[1 + \frac{z_t^2}{\nu-2}\right]^{-\frac{\nu+1}{2}},$$

where $\Gamma(.)$ is Euler gamma function;

$\nu > 2$ is number of degrees of freedom.

The kurtosis of the distribution z_t is

$$K = 3\frac{\nu-2}{\nu-4}, \ \nu > 4.$$

The kurtosis of the errors ε_t is

$$K_u = 3\frac{\nu-2}{\nu-4}\frac{E(\sigma_t^4)}{\left[E(\sigma_t^2)\right]^2}. \tag{4}$$

The factor containing ν in expression (4) makes it possible to take into account the excess kurtosis [1]. It can be seen from formula (4) that the kurtosis of the t-distribution depends on the degree of freedom and the degree of volatility of the process. Thus, by varying the degrees of freedom of the t-distribution, we can obtain different values of kurtosis and different degrees of heavy-tailedness. However, kurtosis based on 4th-order moments is very sensitive to outliers, and therefore, in the presence of outliers, can lead to false conclusions about the nature of the distribution of the residuals.

The question of whether kurtosis measures heavy-tailedness and how to determine which distribution has a heavier tail based on kurtosis is debatable. If, for example, we look at kurtosis as an average of outliers, then large kurtosis indicates large heavy tails [15]. Some authors describe kurtosis as a measure of both “pointiness” and “tail thickness” [16]. In general, there are three approaches to comparing the heaviness of the tails of distributions: the usual kurtosis K, the measures of "heaviness" of the tails, and the ratio of interquantile intervals. Consider one of the approaches – a measure of heavy-tailedness based on quantiles, in the form:

$$K_\alpha\left(y_t\right)=\frac{Q_{1-\alpha}(y_t)-Q_\alpha(y_t)}{Q_{1-\tau}(y_t)-Q_\tau(y_t)}, \quad (5)$$

where $Q_\theta(y_t)$ is θ-quantile y_t, $0<\alpha<\tau<0.5$ [16].

Following article [17], we choose $\tau = 0.25$ and $\alpha = 0.01$, $\alpha = 0.05$. It is believed that the quantile measure (5) is free from assumptions about distributions and from kurtosis values. Therefore it is resistant to misclassifications of distributions and can be used to compare distributions. We calculate and study the behavior of kurtosis (2) and measure of heavy-tailedness (5) for the normal distribution and Student's t-distribution, as a commonly used distribution in econometric practice to account for fat tails, and compare it with the distributions of empirical returns data, which will allow us to make an assumption about the number degrees of freedom t-distribution for empirical data when specifying GARCH(1,1)-models. We will compare this with the distributions of empirical data of returns, which will allow us to make an assumption about the degrees of freedom t-distribution for empirical data when specifying GARCH(1,1)-models.

2. Analysis of heavy tails of distribution in Russian studies

In this section, we will analyze how heavy tails of distributions are taken into account in practice when forecasting financial series of returns in recent studies. To account for fat tails in econometric practice, Student's t-distribution and its variations are often used. Shvedov noted the importance of using t-distribution for MLE (maximum likelihood estimation) estimates in the case of outliers in the data. The author compared the EM-algorithm (expectation-maximization algorithm) and LSM (least-squares method) for linear regression model estimates on generated data with different distributions of errors [18]. Balaev considered and compared the two-dimensional t-distribution with a vector and a scalar of degrees of freedom, the generalized error distribution and the Gram-Charlier distribution according to the daily closing prices of stock indices of various countries: S&P 500, FTSE 100, CAC 40, DAX, Hang Seng, Nikkei during the period November 26, 1990 – November 18, 2012. The author noted that the distributions of all considered returns have heavy tails. The kurtosis coefficient varied from 5.21 to 9.45. It was found that the model based on t-distribution with a vector of degrees of freedom was more preferable [19]. Works of Fantazzini were of a survey nature and were devoted to modeling multivariate distributions based on copula functions. The paper introduced the concepts of upper and lower tail dependences for random variables with a certain probability of outliers and considers Student's copula functions [20]. Balazs in his work analyzed the influence of external sources of information (news and trading volumes) on the volatility of securities using GARCH(1,1)-models. The author noted that the hypothesis of normality was rejected for most of the securities under consideration (returns on shares of 19 companies from the FTSE100 list for the period July 01, 2005 – July 01, 2008) [21]. Some authors used a class of special models Go-GARCH, GJR-GARC, which allow one to estimate the degrees of freedom of the t-distribution along with other model parameters [22, 23]. Lakshina simulated returns with further calculation of the dynamic hedge ratio for eight shares of Russian companies traded on the RTS for the period January 01, 2007 – October 01, 2014. Kurtosis ranged from 18 to 42. Based on the GO-GARCH model, it was calculated that the residuals were Student's distribution with 2 degrees of freedom.

Note that the authors of the papers under consideration used the kurtosis as an "indicator" of heavy tails, and the question of choosing the degrees of freedom t-distribution usually remained outside the scope of such studies. It is also known that the inclusion of the degrees of freedom t-distribution in the arguments of the likelihood function in the MLE is not always correct [24]. Thus, some alternative characteristics of the "heavy-tailedness" are needed, which are easily implemented in practice. In our work, we decided to make an attempt to fill this gap.

3. Analysis of the measure of heavy-tailedness and kurtosis for Student's distribution on generated data

Consider the behavior of the measure of heavy-tailedness in the form (5) and kurtosis (2) for Student's $t(v)$-distribution for different degrees of freedom v. Using the Monte Carlo method, we will generate 5000 repetitions over $N = 200$, 750 and 1000 observations, calculate and compare the kurtosis K and measures of heavy-tailedness K_{01}, K_{05} for quantiles $\alpha = 0.01$, $\alpha = 0.05$, respectively: the interval of variation from minimum to maximum (min-max) and average (mean). The simulation results are shown in *Table 1.* Random variables were generated in the Stata16 package. The generation of pseudo-random numbers was implemented on the basis of the algorithm proposed in the work [25]. Due to the properties of t-distribution, we considered degrees of freedom from 3 to 10. Note that theoretical kurtosis exists for $v > 4$.

The values of the measures from *Table 1* will be further used for comparison with the measures of heavy-tailedness of the analyzed returns of the Moscow Exchange indices for further specification of the GARCH(1,1)-model in choosing the assumption about the distribution of residuals.

Table 1.

Measures of heavy-tailedness (5) and kurtosis (2) for $t(v)$-Student's distribution and normal distribution

	$K_{01min-max}$	K_{01mean}	$K_{05min-max}$	K_{05mean}	$K_{min-max}$	K_{mean}
$N = 1000$						
$t(3)$	4.8–7.7	5.96	2.71–3.62	3.08	5.7–464.07	29.883
$t(4)$	3.9–7.08	5.085	2.51–3.37	2.875	4.09–304.55	12.653
$t(5)$	3.76–5.61	4.645	2.41–3.17	2.772	3.88–115.6	7.550
$t(6)$	3.71–5.35	4.388	2.35–3.13	2.708	3.5–127.91	6.166
$t(7)$	3.57–5.02	4.224	2.29–3.11	2.662	3.35–26.54	4.679
$t(8)$	3.45–4.97	4.103	2.26–3.13	2.633	3.30–16.28	4.486
$t(9)$	3.37–4.76	4.016	2.28–3.13	2.609	3.04–13.85	4.186
$t(10)$	3.38–4.68	3.947	2.29–3.11	2.589	3.01–13.83	4.013
$N(0,1)$	3.01–4.07	3.467	2.18–2.74	2.442	2.57–3.84	3.004
$N = 750$						
$t(3)$	4.37–8.94	5.989	2.57–3.75	3.084	4.28–572.55	27.596
$t(4)$	3.81–7.52	5.080	2.39–3.49	2.881	3.34–614.12	12.881
$t(5)$	3.45–6.49	4.649	2.31–3.37	2.771	3.31–541.35	7.391
$t(6)$	3.39–5.82	4.397	2.27–3.25	2.707	3.26–147.19	5.672
$t(7)$	3.34–5.31	4.216	2.27–3.21	2.663	3.06–48.49	4.909
$t(8)$	3.21–5.42	4.105	2.23–3.25	2.631	3.02–55.67	4.411
$t(9)$	3.21–5.21	4.013	2.15–3.07	2.606	2.89–23.87	4.152
$t(10)$	3.12–5.11	3.951	2.18–3.20	2.591	2.89–24.09	3.975
$N(0,1)$	2.82–4.37	3.465	2.13–2.84	2.442	2.49–4.36	2.995

	$K_{01min-max}$	K_{01mean}	$K_{05min-max}$	K_{05mean}	$K_{min-max}$	K_{mean}
$N = 200$						
$t(3)$	3.87–12.45	6.259	2.25–4.31	3.108	3.28–137.78	13.783
$t(4)$	3.42–9.54	5.237	2.21–3.74	2.897	3.12–141.47	8.775
$t(5)$	3.17–7.75	4.756	2.14–3.73	2.790	2.78–62.72	6.123
$t(6)$	2.81–6.46	4.459	1.97–3.64	2.714	2.71–65.19	5.200
$t(7)$	2.97–7.72	4.307	1.99–3.75	2.670	2.52–21.59	4.648
$t(8)$	2.96–6.31	4.155	2.01–3.40	2.633	2.61–26.43	4.266
$t(9)$	2.96–5.56	4.058	2.05–3.62	2.611	2.51–26.48	4.086
$t(10)$	2.73–6.24	3.987	2.00–3.51	2.588	2.31–30.24	3.943
$N(0,1)$	2.72–4.92	3.504	2.01–3.23	2.446	2.32–4.28	3.007

Behavior of measures of heavy-tailedness – key findings:

- Measures of heavy-tailedness $K_{01,}$ K_{05} (5) based on quantile estimates are more robust to outliers than kurtosis calculated from distribution moment estimates. Measures lie in fixed ranges depending on the degree of freedom, while the kurtosis varies greatly for all considered degrees of freedom v from 3 to 10. It is rather difficult to make assumptions about the degrees of freedom of the t-distribution over the kurtosis values.
- Measures of heavy-tailedness K_{01} for quantile 0.01 are more informative, as K_{05} measures have many overlapping intervals. In what follows, to justify the choice of the degrees of freedom v of the t-distribution, we will use the K_{01} measure, and the K_{05} measure will be used to control the range of the measure using empirical data.
- The resulting intervals of variation of measures of heavy-tailedness for theoretical distributions will be used in further work to compare with measures of heavy-tailedness based on empirical data and substantiate the assumption regarding the degrees of freedom of the t-distribution of residuals in GARCH(1,1)-model.
- It can be assumed that the measures of heavy-tailedness comparison approach works well on large samples (N = 700 and more). Note that in [17], measures of heavy-tailedness were also calculated based on interfractile range (α = 0.125), which requires a sample of at least 1000 observations. This can be attributed to the disadvantages of this approach, since economic data are not always expressed in long time series.

4. Measures of heavy-tailedness and kurtosises for Moscow Exchange indices

In this paper, we examined the returns (logarithmic differences) of the MOEX Index: major and sectoral indices for the period from April 01, 2019 to February 23, 2022 (732 trading days) [26]. The opening and closing prices (the price of the first and last transaction on the trading day) were studied. Graph of returns of the oil and gas index (closing price) is shown in *Fig. 1*. The series of returns has a constant zero mean value, and clustering volatility is observed. The period t < 200 (April 01, 2019 – February 20, 2020) is marked as a period of low volatility. The returns dynamics of other indices behave in a similar way. Periods of high volatility, as a rule, are characterized by abnormally high values (in absolute value) of returns, which leads to high values of kurtosis and the appearance of “fat” distribution tails. Note that indices with different kurtosis values were taken for further analysis.

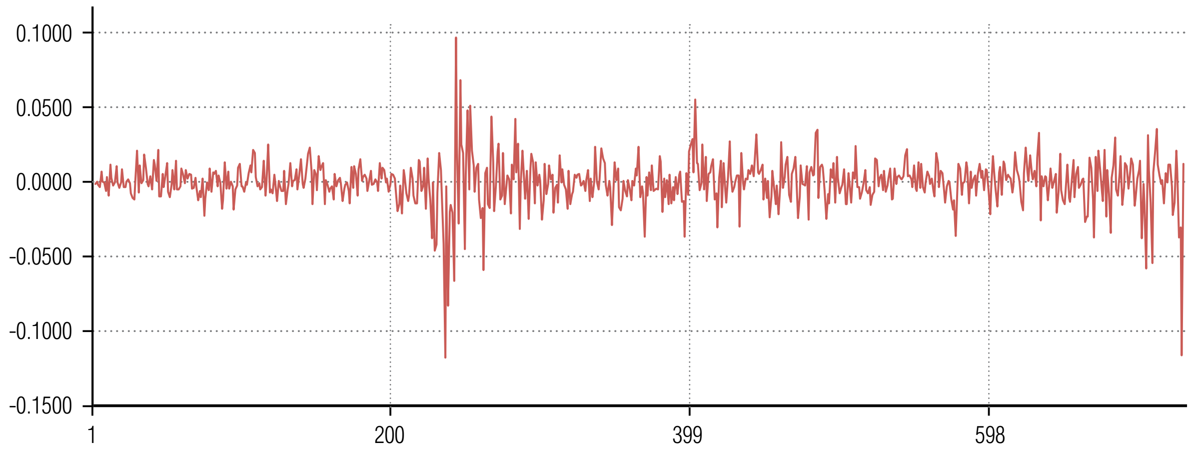


Fig. 1. Logarithmic difference of Oil and Gas Index (Closing Price) (*y*-axis) for the period from April 01, 2019 to February 23, 2022 (*x*-axis).

Let us introduce the notation in *Tables 2, 3*:

- Period 1: April 01, 2019 – February 23, 2022 ($N = 732$), period 2: April 01, 2019 – February 20, 2020 ($N = 200$);
- $K1$, $K2$ are kurtoses (2) of the returns of the Moscow Exchange indices for period 1 and 2, respectively;
- $K1_{01}$, $K1_{05}$ and $K2_{01}$, $K2_{05}$ are measures of heavy-tailedness (5) for periods 1 and 2 for 0.01 and 0.05 quantiles, respectively.

Tables 2, 3 contain the values of kurtoses $K1$ and $K2$ of index returns for two periods: period 1 has a pronounced clustering volatility, period 2 – the period of low volatility. Note that period 1 is characterized by a significant difference in the value of the kurtosis between the opening and closing prices by almost two times, and the kurtosis varies from 6.77 to 51.87. The opening price (variables with index 2 in *Tables 2, 3*) and the closing price (variables with index 1 in *Tables 2, 3*) as different variables were used further for modeling. It is obvious that the value of kurtoses of indicators for period 1 indicates that the assumption of normality of residuals in GARCH(1,1) cannot be used. In period 2, the differences in the value of kurtoses for the opening and closing prices are insignificant, and for some indices they coincide with the kurtosis of the normal distribution ($K = 3$) (*Tables 2, 3*). Thus, further analysis and comparison of estimates of GARCH(1,1)-models in the work was carried out for period 1.

It should be noted that the measures of heavy-tailedness $K2_{01}$ (5) of most of the analyzed indices (85%) calculated for period 2 fall within the intervals of varying measures for a normal distribution ($N = 200$): 2.72–4.92 (*Table 1*). The measures of heavy-tailedness $K2_{01}$ (5) of some variables calculated for period 2 are given in *Table 2*. Thus, it can be assumed that GARCH(1,1)-models, assuming normal residuals, will be the best model for a period with low volatility.

Assumptions about the degrees of freedom of the $t(v)$-distribution of residuals for further specification of GARCH(1,1)-models by comparing the calculated measures of heavy-tailedness with the value of measures of theoretical distributions (*Table 1*) based on the calculated measures were made for period 1. For example, index blue1 has measure of heavy-tailedness $K1_{01}$ = 6,009 (*Table 2*). This measure corresponds to the variation interval $K01_{\min-\max}$ ($N = 750$) for $t(3)$: 4.37–8.94; $t(4)$: 3.81–7.52; $t(5)$: 3.45–6.49 (*Table 1*). Therefore, $t(3)$–$t(5)$ will be the assumed distributions of the residuals when estimating the GARCH(1,1)-models. We will consider how well the considered measure of heavy-tailedness allows us to correctly specify GARCH(1,1). From *Tables 2, 3* it can be seen that mea-

Table 2.

Values of kurtoses (2) and measures of heavy-tailedness (5) of index returns (logarithmic differences), *main equity* indices

MOEX Index		Period 1				Period 2		
		$K1$	$K1_{01}$	$K1_{05}$	Estimated $t(v)$ for residuals	$K2$	$K2_{01}$	$K2_{05}$
blue1	MOEX Blue Chip Index	15.194	6.009	2.946	$t(3)-t(5)$	3.597	3.991	2.287
blue2		33.434	6.225	2.966	$t(3)-t(5)$	3.579	4.541	2.509
imoex1	MOEX Russia Index	15.596	6.347	3.019	$t(3)-t(5)$	3.406	3.792	2.347
imoex2		32.844	6.763	2.993	$t(3)-t(4)$	3.392	4.041	2.319
rts1	RTS Index	14.432	6.716	3.179	$t(3)-t(4)$	4.982	4.547	2.411
rts2		29.391	6.508	3.087	$t(3)-t(5)$	4.177	5.333	2.789

Note: $var1$ is closing price, $var2$ is opening price. Period 1: April 01, 2019 – February 23, 2022 (N = 732), period 2: April 01, 2019 – February 20, 2020 (N = 200).

Table 3.

Values of kurtoses (2) and measures of heavy-tailedness (5) of index returns (logarithmic differences), *sectoral* indices

MOEX Index		Period 1				Period 2
		$K1$	$K1_{01}$	$K1_{05}$	Estimated $t(v)$ for residuals	$K2$
gaz1	Oil and gas	14.669	5.602	3.007	$t(3)-t(6)$	3.233
gaz2		27.428	6.221	3.388	$t(3)-t(5)$	3.529
chem1	Chemicals	6.772	5.785	3.528	$t(3)-t(6)$	5.527
chem2		10.558	6.034	3.270	$t(3)-t(5)$	7.789
electro1	Electric Utilities	21.194	6.604	3.107	$t(3)-t(4)$	3.519
electro2		36.633	6.775	2.937	$t(3)-t(4)$	4.044
telecom1	Telecoms	13.667	6.373	3.109	$t(3)-t(5)$	7.295
telecom2		51.869	7.142	3.461	$t(3)-t(4)$	5.977

<table>
<tr><th colspan="2" rowspan="2">MOEX Index</th><th colspan="4">Period 1</th><th>Period 2</th></tr>
<tr><th>$K1$</th><th>$K1_{01}$</th><th>$K1_{05}$</th><th>Estimated $t(v)$ for residuals</th><th>$K2$</th></tr>
<tr><td>metal1</td><td rowspan="2">Metals & Mining</td><td>12.399</td><td>5.621</td><td>2.842</td><td>$t(3)–t(6)$</td><td>3.295</td></tr>
<tr><td>metal2</td><td>25.064</td><td>5.975</td><td>3.049</td><td>$t(3)–t(5)$</td><td>3.472</td></tr>
<tr><td>finan1</td><td rowspan="2">Financials</td><td>13.063</td><td>5.973</td><td>3.309</td><td>$t(3)–t(5)$</td><td>4.097</td></tr>
<tr><td>finan2</td><td>37.475</td><td>6.557</td><td>3.327</td><td>$t(3)–t(5)$</td><td>4.421</td></tr>
<tr><td>potreb1</td><td rowspan="2">Consumer goods and Services</td><td>12.399</td><td>5.621</td><td>2.842</td><td>$t(3)–t(6)$</td><td>3.295</td></tr>
<tr><td>potreb2</td><td>25.064</td><td>5.975</td><td>3.049</td><td>$t(3)–t(5)$</td><td>3.472</td></tr>
<tr><td>trans1</td><td rowspan="2">Transport</td><td>16.832</td><td>7.081</td><td>3.457</td><td>$t(3)–t(4)$</td><td>5.569</td></tr>
<tr><td>trans2</td><td>18.288</td><td>7.712</td><td>3.368</td><td>$t(3)$</td><td>7.355</td></tr>
</table>

Note: *var*1 is closing price, *var*2 is opening price. Period 1: 01.04.2019–23.02.2022 (N = 732), period 2: 01.04.2019–20.02.2020 (N = 200).

sures of heavy-tailedness did not give a clear answer, but they allowed us to narrow down the number of models that need to be further evaluated.

5. Results

In this section, we compare the GARCH(1,1) specifications for the MOEX indices assuming different types of distribution in the residuals: normal and t-distribution with degrees of freedom from 3 to 9. The results of estimation and comparison of models for the indicator gaz1 are given as an example in *Table 3*. The likelihood ratio test [27, p.171] and the comparison of information criteria by Akaike and Schwartz were used to compare models. The results showed the same result, so only the value of the maximum of the likelihood function is given below in the text. Estimates of the parameters γ and β of the GARCH(1,1)-model in the form (3) are given in *Table 4*. LLF values are the value of the maximum likelihood function for the current model. Note that the full log-likelihood function with the inclusion of terms without optimization parameters is calculated in Stata. The form of the log-likelihood function under the assumption of normal and t-distribution is given in [28]. All model coefficients were statistically significant at the 1% significance level.

The results show (*Table 4*) that for various specifications, the coefficient $\beta \approx 0.84–0.89$, which indicates the persistence of volatility over time, $(\gamma + \beta)$ exceeds 0.9, which indicates the presence of a pronounced GARCH effect. The coefficients γ and β for various specifications behave quite steadily. $(\gamma + \beta) > 1$ is for the case $t(3)$, which violates the condition of positivity of the conditional variance of the model. This model is also not adequate in terms of modeling heavy tails, since the kurtosis of the t-distribution is defined and greater than 3 for the degrees of freedom $v > 4$. The model with $t(5)$-distribution in the residuals is the best model in terms of the minimum values of information criteria, for which AIC = –4300.17 and BIC = –4281.79. For this model, there is also a maximum of LLF = 2154.09. GARCH(1,1) assuming a $t(5)$ distribution in residuals would be the most preferred model for predicting vola-

Table 4.

Estimates of parameters of the GARCH(1,1) model for the variable gaz1 assuming different degrees of freedom of the *t*-distribution for residuals

Distribution of residuals	γ	β	$\gamma + \beta$	LLF
N	0.155	0.836	0.991	2130.876
$t(3)$	0.136	0.892	1.028	2149.819
$t(4)$	0.110	0.884	0.993	2153.393
$t(5)$	**0.104**	**0.879**	**0.982**	**2154.087**
$t(6)$	0.102	0.875	0.978	2153.702
$t(7)$	0.103	0.871	0.974	2151.963
$t(8)$	0.103	0.871	0.974	2151.963
$t(9)$	0.104	0.869	0.973	2150.997

tility. Note that this model is in line with the assumptions about model according the measures of heavy-tailedness: $t(3)$–$t(6)$ (*Table 3*).

Note that the indicator gaz1 has a kurtosis $K1 = 14.67$, which indicates outliers in the data and does not allow using the assumption of normality of the residuals in the GARCH(1,1)-model. GARCH(1,1) with normality in residuals has the highest AIC, BIC and the lowest LLF. Measures of heavy-tailedness $K1_{01} = 5.60$ and $K1_{05} = 3.01$ fall within the variation intervals for the measures of heavy-tailedness for distributions $t(3)$–$t(6)$ (*Table 1*), which in this case coincided with the results of estimating the GARCH(1,1)-model by the enumeration method.

GARCH(1,1)-models were evaluated in the same way for all other indicators. The best models with maximum LLF are given in *Table 5*. Comparison of various specifications of GARCH(1,1) models for each indicator is given in the Appendix.

Models with $t(5)$-distribution assumptions in residuals are the most common and best models as shown by the analysis of GARCH(1,1)-model parameter estimates for MOEX Indices. Such specifications of the model amounted to 60%, while the kurtosis of the logarithmic returns of indicators varied from 6 to 51 (*Tables 2, 3*). The ability to evaluate GARCH(1,1)-models under the assumption of a $t(v)$ distribution in residuals is available in Stata16 with $v \to \infty$ (for example, you can choose $v = 1000$). As the analysis of empirical data shows, the degrees of freedom v of t-distributions for the considered indicators vary from 4 to 7 (*Table 5*), and in terms of measures of heavy-tailedness at $v = 10$, the t-distribution approaches normal. Degrees of freedom assumptions based on measures of heavy-tailedness agreed with the results of empirical analysis by model enumeration in 68% of cases. A few indexes, for example trans2, became an exception. Models with an assumption of a $t(3)$-distribution in the residuals, for which there is no theoretical kurtosis, were not among the best. You can also notice that there is not a single model with the assumption of normal residuals for period 1. Such models had, as a rule, the worst LLF characteristics (application).

Table 5.

Estimates of the parameters of the GARCH(1,1)-model for MOEX Indices assuming different degrees of freedom of the *t*-distribution of residuals

MOEX Index	Estimated distribution of residues according to measures of heavy-tailedness (5)	Distribution of residuals of the best model by LLF	γ	β	$\gamma+\beta$	LLF
Main Equity Indice						
blue1	$t(3)–t(5)$	$t(6)$	0.115	0.869	0.984	2233.683
blue2	$t(3)–t(5)$	$t(5)$	0.123	0.861	0.984	2219.847
imoex1	$t(3)–t(5)$	$t(6)$	0.116	0.871	0.987	2295.943
imoex2	$t(3)–t(4)$	$t(7)$	0.112	0.863	0.975	2294.071
rts1	$t(3)–t(4)$	$t(5)$	0.097	0.901	0.998	2052.176
rts2	$t(3)–t(5)$	$t(5)$	0.121	0.871	0.992	2046.424
Sectoral Indices						
gaz1	$t(3)–t(6)$	$t(5)$	0.104	0.879	0.982	2154.087
gaz2	$t(3)–t(5)$	$t(5)$	0.145	0.84	0.985	2109.548
chem1	$t(3)–t(6)$	$t(5)$	0.05	0.955	1.005	2345.464
chem2	$t(3)–t(5)$	$t(5)$	0.149	0.829	0.978	2313.152
electro1	$t(3)–t(4)$	$t(5)$	0.136	0.842	0.978	2350.902
electro2	$t(3)–t(4)$	$t(5)$	0.137	0.843	0.98	2350.902
telecom1	$t(3)–t(5)$	$t(4)$	0.119	0.864	0.983	2437.535
telecom2	$t(3)–t(4)$	$t(4)$	0.152	0.857	1.009	2418.431
metal1	$t(3)–t(6)$	$t(5)$	0.096	0.879	0.975	2275.479
metal2	$t(3)–t(5)$	$t(5)$	0.142	0.806	0.948	2262.375
finan1	$t(3)–t(5)$	$t(5)$	0.102	0.898	1.000	2167.337
finan2	$t(3)–t(5)$	$t(4)$	0.124	0.883	1.007	2129.232
potreb1	$t(3)–t(6)$	$t(4)$	0.104	0.881	0.985	2274.596
potreb2	$t(3)–t(5)$	$t(5)$	0.142	0.806	0.948	2262.375
trans1	$t(3)–t(4)$	$t(4)$	0.212	0.732	0.944	2251.718
trans2	$t(3)$	$t(6)$	0.301	0.668	0.969	2217.500

Note that in this work we did not consider alternative approaches to estimating the heavy-tailedness distribution, for example, other types of measures, and this is also of scientific interest for further research. The GED-distribution is another possible distribution in the specification of GARCH-models, which was not considered in this study. Often the distribution of financial indicators has asymmetry, which also needs to be taken into account when choosing the specification of GARCH models, but we have not considered it. As noted above, measures of heavy-tailedness based on interfractile ranges are also used in works which require a sample size of $N \geqslant 1000$ [17]. These measures may also be the subject of further research.

Conclusion

We considered generalized autoregressive models of conditional heteroscedasticity for 22 MOEX Indices (Main Equity Indices and Sectoral Indices) with different kurtosis values from 3 to 52 in order to study the heavy-tailedness of distributions and the influence of kurtosis on the choice of the distribution type assumption in the residuals of the model to explain the fat tails. As the analysis showed, kurtosis is only partly an "indicator" of fat tails: on its basis, it is difficult to make an assumption about the form of the distribution of residuals, since it is sensitive to outliers. So, for example, the kurtoses for chem1 and blue2 were 6.77 and 33.43, but for these indicators the best model turned out to be the same model specification – GARCH(1,1) with $t(5)$-distribution in the residuals. It was shown in the work that the considered measures of heavy-tailedness are sufficiently robust to outliers and allow us to *partially* justify the choice of the degree of freedom for the t-distribution when evaluating GARCH(1,1)-models. It should be noted that the use of the model comparison approach based on maximum likelihood estimates gives similar results in 68% of cases (*Table 5*). Perhaps the classical approach is more preferable in econometric practice for the analysis of financial time series on samples of size $N < 1000$. However, the analysis of measures of heavy-tailedness is of great practical importance for modeling time series with heavy tails and substantiating the choice of degrees of freedom of the t-distribution, since kurtosis is not a good quantitative measure of the "heaviness" of distribution tails. In the opinion of the authors, measures of heavy-tailedness and their properties can be useful to a wide range of researchers working with financial time series in order to obtain more accurate profitability forecasts. This article is a small contribution to the further development of time series analysis tools. ■

Appendix.

Comparison of GARCH(1,1) models for MOEX Index by LLF

	MOEX Index										
Distribution of residuals	blue1	blue2	imoex1	imoex2	rts1	rts2	gaz2	chem1	chem2	electro1	electro2
N	2217.00	2189.29	2279.85	2277.39	2021.94	2017.52	2067.12	2321.80	2285.55	2321.49	2266.05
$t(3)$	2225.95	2213.30	2288.15	2284.26	2048.65	2041.62	2104.99	2342.08	2309.40	2347.75	2322.18
$t(4)$)	2231.36	2218.40	2293.59	2290.61	2051.96	2045.63	2108.85	2345.10	2312.76	2350.74	2325.95
$t(5)$	2233.23	**2219.85**	2295.48	2293.09	**2052.18**	**2046.42**	**2109.55**	**2345.46**	**2313.15**	**2350.90**	**2326.57**
$t(6)$	**2233.68**	2219.83	**2295.94**	2293.95	2051.30	2046.02	2109.00	2344.89	2312.49	2350.07	2325.93
$t(7)$	2233.51	2219.18	2295.78	**2294.07**	2050.04	2045.15	2107.95	2343.97	2311.44	2348.89	2324.75

	MOEX Index										
Distribution of residuals	blue1	blue2	imoex1	imoex2	rts1	rts2	gaz2	chem1	chem2	electro1	electro2
$t(8)$	2233.06	2218.27	2295.33	2293.83	2048.67	2044.10	2106.71	2342.94	2310.28	2347.63	2323.37
$t(9)$	2232.49	2217.26	2294.77	2293.41	2047.32	2043.02	2105.41	2341.91	2309.12	2346.38	2321.91

	MOEX Index									
Distribution of residuals	telecom1	telecom2	metal1	metal2	finan1	finan2	potreb1	potreb2	trans1	trans2
N	2383.40	2355.57	2244.45	2216.95	2142.97	2077.17	2244.45	2259.31	2198.72	2192.53
$t(3)$	2435.90	2416.07	2270.33	2259.31	2166.45	2126.57	2270.33	2259.31	2250.95	2210.66
$t(4)$	**2437.54**	**2418.43**	2274.60	2262.29	2166.45	**2129.23**	**2274.60**	2262.29	**2251.72**	2215.79
$t(5)$	2436.19	2417.67	**2275.48**	**2262.38**	**2167.34**	2129.05	2275.48	**2262.38**	2249.93	2217.34
$t(6)$	2433.95	2415.87	2275.07	2261.43	2167.05	2127.84	2275.07	2261.43	2247.41	**2217.50**
$t(7)$	2431.52	2413.74	2274.14	2260.10	2166.30	2126.24	2274.14	2260.10	2244.78	2217.06
$t(8)$	2429.13	2411.56	2273.01	2258.66	2165.37	2124.53	2273.01	2258.66	2242.26	2216.37
$t(9)$	2426.88	2409.43	2271.83	2257.22	2164.41	2122.84	2271.83	2257.22	2239.91	2215.57

Note. For gaz1, a comparison of the models is given in *Table 4*.

References

1. Rossi E. (2010) One-dimensional GARCH models: A review. *Quantile*, vol. 8, pp. 1–67 (in Russian).
2. Engle R. (1982) Autoregressive conditional Heteroscedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica*, vol. 50, pp. 987–1007. https://doi.org/10.2307/1912773
3. Bollerslev T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, vol. 31, pp. 307–327. https://doi.org/10.1016/0304-4076(86)90063-1
4. Bollerslev T., Baillie R. (1989) The message in daily exchange rates: A conditional-variance tale. *Journal of Business & Economic Statistics,* vol. 7(3), pp. 297–305. https://doi.org/10.2307/1391527
5. Hsieh D.A. (1989) Modeling heteroscedasticity in daily foreign-exchange rates. *Journal of Business & Economic Statistics,* vol. 7(3), pp. 307–317. https://doi.org/10.2307/1391528

6. Nelson D. (1991) Conditional heteroskedasticity in asset returns: A new approach. *Econometrica,* vol. 59, no. 2, pp. 347–370. https://doi.org/10.2307/2938260

7. Hansen B. (1994) Autoregressive conditional density estimation. *International Economic Review,* vol. 35, no. 3, pp. 705–730. https://doi.org/10.2307/2527081

8. Ibragimov M., Ibragimov R., Walden J. (2015) *Heavy-tailed distributions and robustness in economics and finance*. Springer International Publishing.

9. Asmussen S.R. (2003) Steady-state properties of GI/G/1. *Applied Probability and Queues. Stochastic Modelling and Applied Probability*, vol. 51, pp. 266–301. https://doi.org/10.1007/0-387-21525-5_10

10. Foss S., Korshunov D., Zachary S. (2013) *An introduction to heavy-tailed and subexponential distributions*. Springer Science & Business Media.

11. Novak S.Y. (2011) *Extreme value methods with applications to finance*. London: CRC.

12. Fiori A.M., Zenga M. (2009) Karl Pearson and the origin of kurtosis. *International Statistical Review*, vol. 77(1), pp. 40–50.

13. Tsyplakov A. (2008) Mini-dictionary of English econometric terms. *Quantile*, vol. 5, pp. 45–48 (in Russian).

14. Engle R. (2001) GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of economic perspectives*, vol. 15, no. 4, pp. 157–168. https://doi.org/10.1257/jep.15.4.157

15. Hogg R.V. (1974) Adaptive robust procedures: A partial review and some suggestions for future applications and theory. *Journal of the American Statistical Association*, vol. 69, pp. 909–927.

16. Ruppert D. (1987) What is kurtosis?: An influence function approach. *The American Statistician*, vol. 41(1), pp. 1–5. https://doi.org/10.2307/2684309

17. Xiaochun L. (2019) On tail fatness of macroeconomic dynamics. *Journal of Macroeconomics (electronic journal)*, vol. 62(C). https://doi.org/10.1016/j.jmacro.2019.103154

18. Shvedov A.S. (2011) Robust regression using *t*-distribution and EM-algorithm. *The HSE Economic Journal*, vol. 1, pp. 68–87 (in Russian).

19. Balaev A.I. (2013) Multivariate time series analysis of financial returns: Comparison of different heavy tail modeling approaches. *The HSE Economic Journal*, vol. 2, pp. 239–263 (in Russian).

20. Fantazzini D. (2011) Modeling multivariate distributions using copula functions. II. *Applied Econometrics*, vol. 3, pp. 98–132 (in Russian).

21. Balash V.A., Sidorov S.P. (2013) Using news analytics data in GARCH models. *Applied Econometrics*, vol. 1, pp. 82–96 (in Russian).

22. Lakshina V.V. (2016) Dynamic hedging with risk aversion. *The HSE Economic Journal*, vol. 1, pp. 156–174 (in Russian).

23. Asaturov K.G., Teplova T.V. (2014) Construction of hedging ratios for highly liquid shares of the Russian market based on GARCH class models. *Economics and mathematical methods*, vol. 1, pp. 37–54 (in Russian).

24. Lucas A. (1997) Robustness of the Student-t based M-estimator. *Communications in Statistics – Theory and Methods*, vol. 26, no. 5, pp. 1165–1182. https://doi.org/10.1080/03610929708831974

25. Matsumoto M., Nishimura T. (1998) Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator // *ACM Transactions on Modeling and Computer Simulation*, vol. 8, pp. 3–30.

26. *Moscow Exchange Indices*. Available at: https://www.moex.com/ru/indices (accessed 10 December 2022).

27. Ayvazyan S.A. (2010) *Methods of econometrics*. Moscow: Magistr (in Russian).

28. *State documentation*. Available at: https://www.stata.com/manuals/tsarch.pdf (accessed 10 December 2022).

About the authors

Lilia A. Rodionova

Cand. Sci. (Econ.);

Associate Professor, Department of Statistics and Data Analysis, HSE University, 11, Pokrovsky Bulvar, Moscow 109028, Russia;

E-mail: lrodionova@hse.ru

ORCID: 0000-0002-0310-6359

Elena D. Kopnova

Cand. Sci. (Tech.);

Associate Professor, Department of Statistics and Data Analysis, HSE University, 11, Pokrovsky Bulvar, Moscow 109028, Russia;

E-mail: ekopnova@hse.ru

ORCID: 0000-0002-8429-141X

DOI: 10.17323/2587-814X.2023.3.53.69

Formation of the causal field of indicators for an organization's intellectual capital development: A concept and a fuzzy economic and mathematical model

Georgij S. Zavalin
E-mail: georgiy.zavalin08@vvsu.ru

Olga V. Nedoluzhko
E-mail: olga.nedoluzhko25@gmail.com

Konstantin S. Solodukhin
E-mail: k.solodukhin@mail.ru

Vladivostok State University
Address: 41, Gogolya Street, Vladivostok 690014, Russia

Abstract

The development of intellectual capital theory through the introduction of the concept of implicitness involves considering intellectual capital as an implicit factor, so that the process of its formation is largely determined by the impact of specific hidden factors whose impact is expressed implicitly and is difficult to formalize. Currently, the process of selecting explicit and implicit factors affecting intellectual capital is not formalized in domestic and foreign studies, and therein is the relevance of this work. The purpose of this study was to develop a scheme for selecting explicit and implicit factors in the development of the organization's intellectual capital in conjunction with its strategy based on a modified Balanced Scorecard, taking into account the distribution of indicators by types of cognitive activity. The implementation of this scheme was carried out by developing a fuzzy economic and mathematical model suitable for practical use. The main feature of the model is the possibility of fuzzy setting of "cut-off boundaries" for explicit and implicit factors. We present the results of testing the model on the example of a large regional university. Sets of explicit and implicit factors of the university's intellectual capital are given for various "cut-off boundaries" using various defuzzification methods.

Keywords: cognitive activity, fuzzy model, implicit factor, intellectual capital, strategic management

Citation: Zavalin G.S., Nedoluzhko O.V., Solodukhin K.S. (2023) Formation of the causal field of indicators for an organization's intellectual capital development: A concept and a fuzzy economic and mathematical model. *Business Informatics*, vol. 17, no. 3, pp. 52–69. DOI: 10.17323/2587-814X.2023.3.53.69

Introduction

In the context of the formation of the knowledge-based economy, the main sources of competitive advantages for organizations are intangible factors of production, including the organization's intellectual capital (hereinafter referred to as IC). IC is the instrumental core of the knowledge-based economy. The nature of IC development is largely determined by the impact of specific hidden factors whose impact on the development process is implicit and difficult to formalize. This circumstance necessitates the identification of such factors [1–3].

The development of the IC theory through the introduction of the concept of implicitnes made it possible to provide an explanation and interpretation of the business processes of economic systems at a fundamentally new level of generalization. Within the framework of the theory, the IC itself is an implicit factor, the process of formation of which is largely determined by the impact of a number of implicit factors affecting it [4]. The concept of implicitness was originally formulated in cognitive psychology [5] developed in linguistics [6, 7], and at the present stage it has found practical application in economic sciences [8, 9].

Implicit factors are non-obvious factors that have a significant impact on the business processes of an economic entity which are based on hidden information [10]. In the context of the knowledge-based economy, when the impact of information as the most important resource becomes most significant, the impact of implicit factors in the management system of an organization increases [11, 12]. Accordingly, it seems appropriate to single out, in addition to factors that clearly affect the IC development (explicit), factors of hidden, indirect impact (implicit). Taken together, the selected groups of key indicators of IC development, explicit, or obvious factors that have a direct impact on the IC development, as well as implicit factors, make up the causal field of indicators of the IC development in an organization [13].

The IC development as a source of competitive advantages is carried out within the framework of strategic management of an organization. The most important tool for structuring and implementing the strategy is the Balanced Scorecard (BSC). This system management method proposed by Kaplan and Norton allows us to translate formulated strategic goals and objectives of the organization, considering all aspects of its further development, into specific actions [14–16]. Over its thirty-year history, the BSC concept has undergone significant evolution, not only retaining but also strengthening its popularity. It is used by almost all well-known consulting companies, while all major developers of enterprise information systems offer BSC tool support [17–19].

The possibility of using the BSC in relation to the assessment of organizations' IC is due, firstly, to the emphasis on intangible indicators, and secondly, the relationship between the traditionally distinguished main structural IC components (human capital, organizational capital, relational capital) and the prospects of the BSC [15, 20–23].

Despite its recognized advantages, the BSC is not devoid of shortcomings and has been criticized throughout its evolution [13, 24–27]. Let us single out two shortcomings that are critical from the point of view of the objectives of this work.

Firstly, the traditional BSC model does not consider the indirect impact of implicit factors on the organizations' key performance indicators. Cause-and-effect relationships in strategic maps reflect factors of direct impact (obvious dependencies).

Secondly, in the classical version of the BSC, the inequality of the organization's stakeholders from the point of view of taking into account their interests was initially laid down. However, the most important structural component of the IC is relational capital, which is determined by the nature of the organization's relationships with external entities [28–32].

The solution to the first problem is offered by Nazarov [13], who has developed a model for the reflexive selection of implicit factors for an organization's management activities and its application to the development of a modified BSC. In turn, in works [33, 34], a modification of the BSC is proposed for the so-called stakeholder-company. Within its framework, among other things, they propose a method of constructing a strategic objectives map which eliminates the initial inequality of stakeholders' interests inherent in the classical BSC. According to the objectives of this work, it seems promising to combine the described modifications of the BSC – namely, to apply the model of implicit factors reflexive selection within the framework of the "stakeholder" modification of the BSC.

It is important to note that the formation of an organization's IC is carried out by identifying its basic characteristic – cognitive activity. The cognitive activity reflects the main condition for the emergence of various IC types and is carried out through various mental processes and states [35–37]. The identification of the possible types of cognitive activity in an organization (education, involvement, production rationalization, self-improvement, customer-oriented rationalization, innovation) enables us to implement specific managerial interventions for them at various levels. The types of cognitive activity can be correlated with the structural components of the IC as follows: education and self-improvement contribute to development of human capital; involvement and production rationalization develop organizational capital; customer-oriented rationalization and innovation provide an increase in relational capital.

The hiddenness of implicit factors and the mediation of their impact on IC development (which, in turn, is an implicit factor) leads to the need to use fuzzy tools in their identification. A significant advantage of using fuzzy models and methods is the possibility of formalizing various kinds of uncertainties (primarily linguistic uncertainty). The use of fuzzy tools in relation to a wide variety of objects and areas of knowledge has proven itself well in conditions of incomplete information and various uncertainties. Unfortunately, in the scientific literature, we were unable to find fuzzy models for identifying implicit factors of the IC. At the same time, there are works that offer fuzzy tools in relation to a wide variety of implicit factors of socio-economic systems [38–41].

The work [4] proposes a fuzzy model for identifying implicit factors in an organization's BSC. Identification of indirect impacts within the framework of the model is based on the technology for evaluating fuzzy binary relations on a certain set. At the same time, the elements of the matrices of fuzzy binary relations are single-point fuzzy sets, which to a certain extent narrows the possibilities of using the model. It seems promising to develop this model in relation to the IC within the framework of a new modification of the BSC in relation to the main structural components of

IC, considering the distribution of indicators by types of cognitive activity with a change in the technology for assessing fuzzy binary relations.

Thus, the purpose of this study is to develop a method for selecting explicit and implicit factors in the development of an organization's IC in conjunction with its strategy based on the modified BSC, which considers the distribution of indicators by types of cognitive activity in a fuzzy setting.

1. The method of forming the causal field of IC development indicators

The formation of the causal field of an organization's IC development indicators involves the allocation of three groups of indicators:

1. the key IC development indicators;
2. he explicit IC factors (having an obvious direct impact on IC development);
3. the implicit IC factors (having an indirect impact on the IC development).

The formation of the causal field of an organization's IC development indicators is proposed to be carried out within the framework of the modified BSC. It is proposed to group the organization's strategic objectives that are significantly related to the development of its IC into three groups corresponding to the main structural components of IC. At the same time, integral indicators corresponding to the main structural components of the IC can be considered as the IC key indicators.

Since, as noted above, each structural component of the IC can be correlated with two types of cognitive activity, in fact, there is a grouping of strategic objectives into six groups.

The preliminary selection of indicators applying for inclusion in the groups "explicit IC factors" and "implicit IC factors" takes place among the indicators of strategic objectives from six groups. The BSC concept assumes that each strategic objective corresponds to a set of the lagging indicators, the values of which make it possible to judge the degree of the objective's achievement. Strategic objectives that contribute, to some extent, to the development of the organization's IC, may be directed towards the development of other key aspects of the organization's activities. Therefore, not all the lagging indicators of these objectives will be indicators of IC development.

The formed set of indicators should be divided into three subgroups: explicit IC factors; implicit IC factors; indicators whose impact on the development of the IC can be neglected (for a specific organization within its strategy at this stage of its development).

To do this, at the first stage, it is necessary to assess the impact of all selected indicators on the key IC indicators. Under the indicator's impact on the IC, we will understand the integral degree of impact of this indicator on the key IC indicators. All indicators, the degree of impact of which on the IC exceeds a certain boundary, will be referred to as explicit factors of the IC.

At the second stage, it is necessary to evaluate the impact of all the remaining indicators on the already selected explicit IC factors. Here, following Nazarov [2], we accept the hypothesis that implicit factors affect the key performance indicators of an organization indirectly. Moreover, explicit factors act as indirect indicators. Accordingly, the impact of the remaining indicators on the IC development can be assessed as a superposition of the impact of these indicators on the explicit IC factors and the explicit IC factors on key IC indicators. All indicators, the degree of the final (indirect) impact of which on the IC exceeds a certain boundary, will be referred to as implicit factors of the IC. Note that in the general case, the "cut-off boundaries" in the selection of explicit and implicit factors may not coincide. We will assume that the impact on the IC of indicators remaining after the selection of explicit and implicit factors can be neglected.

In general, the basic scheme for forming the causal field of IC development indicators is presented in *Fig. 1*.

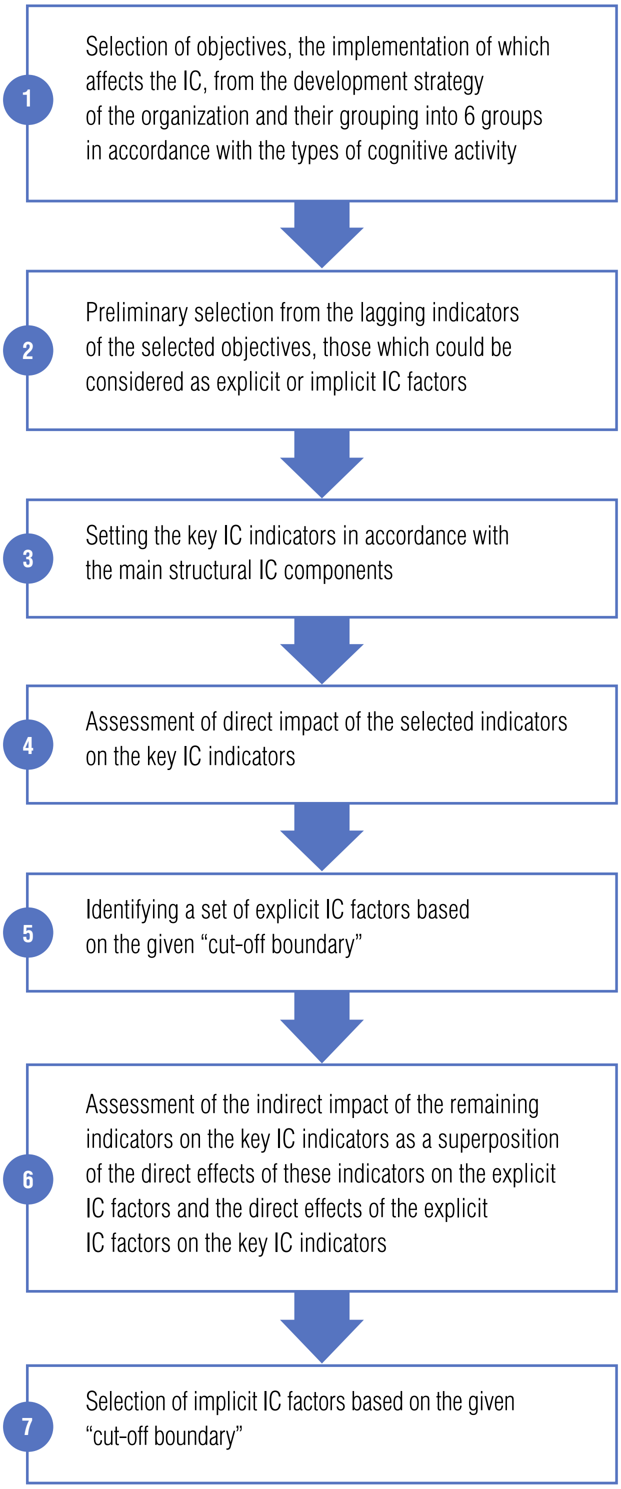


Fig. 1. The basic scheme for forming the causal field of IC development indicators.

2. Fuzzy model

Let $C = \{c_1, c_2, \ldots, c_k\}$ be the set of key indicators of the IC development;

$E = \{e_1, e_2, \ldots, e_t\}$ – the set of strategic objectives indicators that affect the IC development;

$B = \{b_1, b_2, \ldots, b_m\}$ – the set of explicit IC factors;

$A = \{a_1, a_2, \ldots, a_n\}$ – the set of implicit IC factors;

$D = \{d_1, d_2, \ldots, d_s\}$ – the set of factors whose impact on the IC development can be neglected.

Thus, $E = B \cup A \cup D$, and $B \cap A \cap D = \varnothing$, that is $t = m + n + s$.

The degree of impact of the set E indicators on the set C indicators are determined by experts in a given linguistic scale. *Table 1* shows a possible linguistic scale and the membership functions of fuzzy sets corresponding to linguistic variables.

Table 1.

Term set of the linguistic variable "the impact of the indicator e_i on the indicator c_j"

Value of the linguistic variable	Trapezoidal membership function
Very weak	<0; 0; 0.5; 1.5>
Weak	<0.25; 1.0; 1.5; 2.75>
Average	<1.0; 2.0; 3.0; 4.0>
Strong	<2.25; 3.5; 4.0; 4.75>
Very strong	<3.5; 4.5; 5.0; 5.0>

The experts' responses should be verified for consistency and averaged. Each expert may be assigned with a crisp or fuzzy weighting coefficient reflecting their level of competence.

As a result, we have a matrix M_{EC} of dimension $t \times k$, the elements of which are fuzzy numbers. Note that the elements of this matrix and subsequent fuzzy matrices can be fuzzy numbers of an arbitrary type (not necessarily singleton fuzzy numbers).

Let us associate a column vector M^*_{EC} of length t with the matrix M_{EC} as follows:

$$(M^*_{EC})_i = \sum\nolimits_{j=1}^{k} w_j (M_{EC})_{ij}, \qquad (1)$$

where w_j are weight coefficients of key indicators of IC development. Note that in the general case the coefficients w_j can be fuzzy (in this particular case, we can consider $w_1 = w_2 = w_3 = 1/3$). The elements of the column vector M^*_{EC} determine the impact of the set E indicators on the IC.

Then we will consider the explicit factors of the IC to be the indicators e_i, for which $(M^*_{EC})_i$ exceed the exogenously set "cut-off boundary". The "cut-off boundary" of explicit factors in general case can be defined fuzzily. In this case, it is necessary to use one of existing methods for comparing fuzzy sets [42]. If the "cut-off boundary" is a crisp number, then the fuzzy elements of the column vector M^*_{EC} can be defuzzified; after that the resulting crisp numbers can be compared with a crisp "cut-off boundary."

Note that traditionally the "cut-off boundary" for explicit factors is set verbally. For example, explicit factors in a key performance indicator of an organization are usually understood as indicators whose impact is "strong" or "very strong". Sometimes (rarely) indicators with an "average" impact are also added to them. In this case, a fuzzy "cut-off boundary" should be understood as a fuzzy set with a membership function corresponding to a given verbal assessment.

Let $F = \{f_1, f_2, ..., f_{n+s}\}$ be the set of strategic objective indicators that are not explicit factors. That is, $F = E \backslash B = A \cup D$.

Let us determine expertly in a given linguistic scale the degree of impact of the set F indicators on the set B indicators. As a result, we have a matrix M_{FB} of dimension $(t - m) \times m$, the elements of which are fuzzy numbers.

Consider a matrix M_{BC} of dimension $m \times k$, obtained from the M_{EC} matrix by deleting rows corresponding to the indicators of the set F. The elements of the M_{BC} matrix reflect the degree of impact of explicit factors on the key indicators of the IC development.

Let the M'_{FC} matrix obtained as a result of the product of the matrices M_{FB} and M_{BC}:

$$(M'_{FC})_{ij} = \sum\nolimits_{k} (M_{FB})_{ik} \cdot (M_{BC})_{kj}. \qquad (2)$$

The product and addition of matrix elements in this case is carried out according to the given rules for the product and addition of fuzzy numbers.

There are two main approaches to the implementation of fuzzy arithmetic operations: the α-cut approach using interval arithmetic, and the extension principle approach using different t-norms. For trapezoidal fuzzy numbers, within the framework of the first approach, one can use the well-known addition and product formulas [43].

There are more sophisticated ways to implement fuzzy arithmetic using computational methods that eliminate the shortcomings of the two main approaches (overestimation of the uncertainty in the resulting fuzzy numbers in the first approach and high sensitivity to changes in the input fuzzy numbers in the second approach). However, in some cases, the complexity of performing computational operations within the framework of these methods can be unacceptably high. In this regard, there are simplifications of the procedure for arithmetic operations on fuzzy numbers of certain types, including trapezoidal ones [44]. The paper [45] proposes a unified system of rules for performing arithmetic operations on $(L\text{-}R)$-type fuzzy numbers.

Note that when using basic formulas for addition and multiplication of trapezoidal fuzzy numbers, the weighted expert assessments will also be trapezoidal fuzzy numbers. However, when applying the mentioned arithmetic operations system to ((L-R)-type fuzzy numbers, the weighted expert assessments may have exponential (Gaussian) membership functions (more precisely, the membership functions of the obtained fuzzy sets are well approximated by Gaussians).

If necessary, we normalize the elements of the matrix M'_{FC} in such a way that the universal set of the resulting fuzzy numbers coincides with the original universal set (in our case [0; 5]). The resulting matrix will be denoted by M_{FC}.

The elements of the matrix M_{FC} reflect the degree of impact of the set F indicators on the key indicators of the IC development.

Let us associate a column vector M^*_{FC} of length t with the matrix M_{FC} as follows:

$$(M^*_{FC})_i = \sum_{j=1}^{k} w_j (M_{FC})_{ij}. \qquad (3)$$

The elements of the column vector M^*_{FC} determine the impact of the set F indicators on the IC.

Then the implicit factors of the IC will be considered the indicators f_i, for which $(M^*_{FC})_i$ exceed the exogenously set "cut-off boundary". The "cut-off boundary" for implicit factors can also be defined fuzzily and, in the general case, does not coincide with the "cut-off boundary" for explicit factors.

3. Approbation of the model

The model was tested on the example of a large regional university (Vladivostok State University, VVSU). VVSU has developed a strategy for the university's development formalized as strategic maps in accordance with the "stakeholder" modification of the BSC. Strategic objectives that are significantly related to the development of the university's IC have been grouped into six categories according to types of cognitive activity (*Table 2*).

Table 2.

University strategic objectives in the field of the IC development (fragment)

Stakeholder group	BSC perspective	Objective	Indicator	Cognitive activity	Structural component of the IC
Employees	Resource	Implementation of procedures and criteria for evaluating the quality and effectiveness of e-learning courses used	Use of e-learning (E_1)	Education	Human capital
Employees	Resource	Implementation of procedures and criteria for evaluating the quality and effectiveness of e-learning courses used	Effectiveness of using distance education technologies (E_2)	Education	Human capital
Employees	Resource	Establishment of a university-business interaction center	Internship activity (E_3)	Education	Human capital
…					
Employees	Resource	Modernization of the university's material and technical infrastructure	Infrastructure provision (E_7)	Involvement	Organizational capital
Employees	Stakeholder	Formation of a unique corporate environment promoting the development and maintenance of corporate culture	Socio-psychological satisfaction (E_8)	Involvement	Organizational capital
Business community	Process	Forming a portfolio of projects and research topics, demanded by the business	Level of scientific and scientific-production cooperation with partners (E_9)	Production rationalization	Organizational capital
…					

Stakeholder group	BSC perspective	Objective	Indicator	Cognitive activity	Structural component of the IC
Clients	Process	Creation of a system for evaluating the effectiveness of the use of e-learning courses in the educational process	Digitalization of the educational process (E_{11})	Production rationalization	Organizational capital
…					
Clients	Process	Inclusion of Russian and foreign internships into higher education and secondary vocational education programs	Efficiency of networking with partners (E_{14})	Production rationalization	Organizational capital
…					
State; Society	Stakeholder	Formation of scientific schools	Publication activity (E_{21})	Self-improvement	Human capital
…					
Employees	Stakeholder	Creation of a system of staff motivation to achieve high performance and career growth	Personal growth of teaching staff (E_{27})	Self-improvement	Human capital
…					
Clients	Stakeholder	University brand development	Student satisfaction with the quality of education (E_{30})	Customer-oriented rationalization	Relational capital
Clients; Business community; Society	Stakeholder	University brand development	Brand management effectiveness (E_{31})	Customer-oriented rationalization	Relational capital
Business community; Society; State	Stakeholder	Creation of a comfortable envi-ronment and modern developed infrastructure necessary for hosting major significant events	Efficiency of public and business initiatives (E_{32})	Customer-oriented rationalization	Relational capital
…					
Employees; Clients; Business community; State	Stakeholder	Development of interdisciplinary scientific research projects	Interdisciplinary scientific projects (E_{40})	Innovation	Relational capital
Business community; State	Stakeholder	Creation of an R&D system potentially demanded by the real sector of the economy	R&D income (E_{41})	Innovation	Relational capital
…					
Business community; State; Society	Stakeholder	Ability to execute scientific projects and, in particular, to lead student teams in carrying out scientific projects, fostering STEMskills	Patent activity (E_{44})	Innovation	Relational capital

Among the lagging indicators of the selected objectives, indicators were selected whose values make it possible to judge the degrees of achieving the objectives in the aspect of developing the IC:

1. Use of e-learning (E_1).
2. Effectiveness of using distance education technologies (E_2).
3. Internship activity (E_3).
4. Efficiency of internship activity (E_4).
5. Degree of staff's qualifications matching the tasks being solved (E_5).
6. Staff retention (E_6).
7. Infrastructure provision (E_7).
8. Socio-psychological satisfaction (E_8).
9. Level of scientific and scientific-production cooperation with partners (E_9).
10. Degree of correspondence of the staff's motivation system to the tasks being solved (E_{10}).
11. Digitalization of the educational process (E_{11}).
12. Infrastructure efficiency (E_{12}).
13. Degree of individualization of educational trajectories (E_{13}).
14. Efficiency of networking with partners (E_{14}).
15. Level of advanced technologies adaptation (E_{15}).
16. Level of automation of management processes (E_{16}).
17. Level of accessibility of digital educational resources (E_{17}).
18. Level of use of open educational platforms (E_{18}).
19. Level of expert support according to WorldSkills standards (E_{19}).
20. Mastery level of WorldSkills standards (E_{20}).
21. Publication activity (E_{21}).
22. Grant activity (E_{22}).
23. Dissertation defenses (E_{23}).
24. Organizational culture formation (E_{24}).
25. International science degree (E_{25}).
26. International academic mobility (E_{26}).
27. Personal growth of teaching staff (E_{27}).
28. Innovative and entrepreneurial activity of teaching staff (E_{28}).
29. Student employment (E_{29}).
30. Student satisfaction with the quality of education (E_{30}).
31. Brand management effectiveness (E_{31}).
32. Efficiency of public and business initiatives (E_{32}).
33. Level of support for student entrepreneurship activity (E_{33}).
34. Uniqueness of a university's educational program portfolio (E_{34}).
35. Level of digital marketing use in interacting with applicants (E_{35}).
36. Internal demand for additional educational university programs (E_{36}).
37. External demand for additional educational university programs (E_{37}).
38. International educational activity (E_{38}).
39. Implemented scientific projects (E_{39}).
40. Interdisciplinary scientific projects (E_{40}).
41. R&D income (E_{41}).
42. Qualification of staff in the field of R&D (E_{42}).
43. Efficiency of the innovation business incubator's activities (E_{43}).
44. Patent activity (E_{44}).

At the next stage, an expert survey was conducted which included representatives of the academic and administrative staff of the university, as well as specially invited external experts. The experts, within the given linguistic scale, assessed the degree of impact of the selected indicators on the key indicators of the IC development corresponding to the main structural components of the IC. Experts' answers were checked for consistency and averaged considering exogenously given expert competence levels. Note that each individual expert assessed the impact of not all 44 indicators on key IC indicators, but only those in respect of which he had the appropriate expert knowledge (competencies). The results of this stage of the expert survey are weighted average expert assessments represented as Gaussian-type fuzzy numbers. *Table 3* shows the parameters of the corresponding approximating Gaussians.

Since the "cut-off boundaries" of explicit and implicit factors were not known in advance, experts also needed to assess the mutual impact of all 44 indicators on each other. In this case, each expert also answered only questions related to their area of expertise. Thus, each expert needed to answer a reasonable number of questions within an acceptable time frame. This approach allows the decision maker to have a wide range of options for varying the "cut-off boundaries" without requiring additional expert questions. The results of the second stage of the expert survey (in the form of parameters of the Gaussians, approximating the weighted average fuzzy expert assessments) are partially shown in *Table 4.*

To conduct an expert survey, process expert answers and perform the necessary calculations based on the fuzzy model, a software package was developed. Among other things, it allows us to form sets of explicit and implicit IC factors for given "cut-off boundaries" and selected defuzzification methods (if "cut-off boundaries" are defined as crisp numbers).

Table 5 shows the sets of explicit and implicit factors of the university's IC for various "cut-off boundaries" obtained using three defuzzification methods (Center of Gravity / Maximum of Maximums / Median).

The decision-maker is able to set the first ("explicit") "cut-off boundary" based on the requirements for the strength of the direct impact of the selected factors on the lagging IC indicators. As a result, a set of explicit IC factors will be formed. Then, based on the requirements for the strength of the indirect impact of the selected factors on the lagging indicators, the second ("implicit") "cut-off boundary" is selected. Thus, a set of implicit IC factors is formed.

Table 3.

Fuzzy assessments of the impact of the set E indicators on the IC development key indicators (fragment)

Indicator	Human capital (C_1)		Organizational capital (C_2)		Relational capital (C_3)		Intellectual capital	
	μ	σ	μ	σ	μ	σ	μ	σ
E_1	2.4472	0.2675	1.2536	0.1401	3.7685	0.2108	2.4999	0.3247
E_2	3.8603	0.3763	1.2991	0.2477	0.1345	0.2980	1.8461	0.2185
E_3	2.3443	0.3999	2.4939	0.3249	3.6572	0.1670	2.8657	0.2251
…								
E_{42}	2.4143	0.3371	2.3756	0.4054	4.8342	0.2865	3.1524	0.1353
E_{43}	3.9457	0.1746	3.8787	0.3795	4.7375	0.1445	3.8849	0.1578
E_{44}	3.5705	0.1556	3.7692	0.1379	3.6417	0.2386	3.5971	0.3190

Table 4.

Fuzzy assessments of the impact of the set E indicators on the IC development key indicators (fragment)

<table>
<tr><th rowspan="2">Indicator</th><th colspan="2">E_1</th><th colspan="2">E_2</th><th colspan="2">E_3</th><th rowspan="2">…</th><th colspan="2">E_{42}</th><th colspan="2">E_{43}</th><th colspan="2">E_{44}</th></tr>
<tr><th>μ</th><th>б</th><th>μ</th><th>б</th><th>μ</th><th>б</th><th>μ</th><th>б</th><th>μ</th><th>б</th><th>μ</th><th>б</th></tr>
<tr><td>E_1</td><td colspan="2">*</td><td>0.26</td><td>0.14</td><td>4.72</td><td>0.14</td><td></td><td>4.93</td><td>0.14</td><td>4.85</td><td>0.28</td><td>3.55</td><td>0.34</td></tr>
<tr><td>E_2</td><td>1.26</td><td>0.27</td><td colspan="2">*</td><td>1.07</td><td>0.35</td><td></td><td>2.55</td><td>0.29</td><td>1.23</td><td>0.26</td><td>2.31</td><td>0.33</td></tr>
<tr><td>E_3</td><td>3.83</td><td>0.33</td><td>4.57</td><td>0.20</td><td colspan="2">*</td><td></td><td>4.64</td><td>0.20</td><td>3.74</td><td>0.35</td><td>3.83</td><td>0.14</td></tr>
<tr><td>…</td><td colspan="13"></td></tr>
<tr><td>E_{42}</td><td>3.61</td><td>0.17</td><td>3.56</td><td>0.19</td><td>4.81</td><td>0.13</td><td></td><td colspan="2">*</td><td>0.42</td><td>0.26</td><td>4.73</td><td>0.21</td></tr>
<tr><td>E_{43}</td><td>4.66</td><td>0.39</td><td>4.66</td><td>0.18</td><td>2.46</td><td>0.24</td><td></td><td>2.62</td><td>0.25</td><td colspan="2">*</td><td>0.35</td><td>0.23</td></tr>
<tr><td>E_{44}</td><td>4.89</td><td>0.18</td><td>4.73</td><td>0.39</td><td>3.63</td><td>0.28</td><td></td><td>1.28</td><td>0.38</td><td>0.19</td><td>0.33</td><td colspan="2">*</td></tr>
</table>

Table 5.

Sets of explicit and implicit IC factors

First cut-off boundary	Numbers of IC indicators taken as explicit	Second cut-off boundary	Numbers of IC indicators taken as implicit
2	1, 2, 4, 5, 6, 9, 11, 14, 21, 22, 23, 24, 31, 39, 40, 41, 42 / 1, 2, 4, 5, 6, 9, 11, 14, 21, 22, 23, 24, 31, 39, 40, 41, 42 / 1, 2, 4, 5, 6, 9, 11, 14, 21, 22, 23, 24, 31, 39, 40, 41, 42	1.5	3, 7, 8, 10, 12, 15, 26, 27, 28, 32, 33 / 3, 7, 8, 10, 12, 15, 26, 27, 28, 32, 33, 43, 44 / 3, 7, 8, 10, 12, 15, 26, 27, 28, 32, 33, 43, 44
		1.75	3, 7, 8, 12, 27, 33 / 3, 7, 8, 12, 27, 33, 43 / 3, 7, 8, 12, 27, 33, 43
		2	7, 8, 12 / 7, 8, 12 / 7, 8, 12
		2.25	None / None / None
2.5	1, 4, 6, 9, 11, 14, 21, 22, 23, 24, 31, 39, 40, 41, 42 / 1, 4, 6, 9, 11, 14, 21, 22, 23, 24, 31, 39, 40, 41, 42 / 1, 4, 6, 9, 11, 14, 21, 22, 23, 24, 31, 39, 40, 41, 42	1.5	2, 3, 5, 7, 8, 10, 12, 15, 26, 27, 28, 32, 33, 43, 44 / 2, 3, 5, 7, 8, 10, 12, 15, 26, 27, 28, 32, 33, 43, 44 / 2, 3, 5, 7, 8, 10, 12, 15, 26, 27, 28, 32, 33, 43, 44
		1.75	2, 3, 7, 8, 12, 27, 28, 33, 43 / 2, 3, 7, 8, 12, 27, 28, 33, 43 / 2, 3, 7, 8, 12, 27, 28, 33, 43
		2	2, 7, 8, 12 / 2, 7, 8, 12 / 2, 7, 8, 12
		2.25	2 / 2 / 2

First cut-off boundary	Numbers of IC indicators taken as explicit	Second cut-off boundary	Numbers of IC indicators taken as implicit
3	1, 6, 9, 11, 14, 21, 22, 23, 39, 40, 41, 42 / 1, 6, 9, 11, 14, 21, 22, 23, 39, 40, 41, 42 / 1, 6, 9, 11, 14, 21, 22, 23, 39, 40, 41, 42	1.5	2, 3, 4, 5, 7, 8, 10, 12, 15, 24, 26, 27, 28, 32, 33, 43, 44 / 2, 3, 4, 5, 7, 8, 10, 12, 15, 24, 26, 27, 28, 32, 33, 43, 44 / 2, 3, 4, 5, 7, 8, 10, 12, 15, 24, 26, 27, 28, 32, 33, 43, 44
		1.75	2, 3, 4, 7, 8, 12, 24, 27, 28, 32, 33, 44 / 2, 3, 4, 7, 8, 12, 24, 27, 28, 32, 33, 44 / 2, 3, 4, 7, 8, 12, 24, 27, 28, 32, 33, 44
		2	2, 3, 4, 7, 8, 12, 24, 28 / 2, 3, 4, 7, 8, 12, 24 / 2, 3, 4, 7, 8, 12, 24, 28
		2.25	2, 4, 7, 12, 24 / 2, 4, 7, 12, 24 / 2, 4, 7, 12, 24
3.5	1, 6, 9, 11, 14, 22, 41, 42 / 1, 6, 9, 11, 14, 22, 41, 42 / 1, 6, 9, 11, 14, 22, 41, 42	1.5	2, 3, 4, 5, 7, 8, 10, 12, 13, 15, 17, 21, 23, 24, 28, 31, 32, 33, 39, 40, 43, 44 / 2, 3, 4, 5, 7, 8, 10, 12, 13, 15, 17, 21, 23, 24, 28, 31, 32, 33, 39, 40, 43, 44 / 2, 3, 4, 5, 7, 8, 10, 12, 13, 15, 17, 21, 23, 24, 28, 31, 32, 33, 39, 40, 43, 44
		1.75	2, 3, 4, 5, 7, 8, 12, 21, 23, 24, 28, 32, 33, 39, 40, 43, 44 / 2, 3, 4, 5, 7, 8, 12, 21, 23, 24, 28, 32, 33, 39, 40, 43, 44, / 2, 3, 4, 5, 7, 8, 12, 21, 23, 24, 28, 32, 33, 39, 40, 43, 44,
		2	2, 3, 4, 7, 8, 12, 24, 32, 33, 39, 40 / 2, 3, 4, 7, 8, 12, 24, 32, 33, 39, 40 / 2, 3, 4, 7, 8, 12, 24, 32, 33, 39, 40
		2.25	2, 4, 7, 12, 24, 39, 40 / 2, 4, 7, 12, 24, 39, 40 / 2, 4, 7, 12, 24, 39, 40

4. Discussion

The analysis of the obtained results shows the following.

A change in the first ("explicit") "cut-off boundary" leads to a change in the sets of the IC factors taken as explicit. At the same time, the larger the "cut-off boundary" (which means stricter requirements for the strength of the direct impact of the selected factors on the lagging indicators), the smaller the number of explicit factors, and vice versa. Interestingly, with different defuzzification methods, the sets of explicit factors do not change for a fixed "cut-off boundary". This is due to the fact that the crisp estimates of the strength of the direct impact of factors obtained using different defuzzification methods differ insufficiently to change the composition of explicit factors. This, in turn, is most likely due to the trapezoidal type of the chosen membership functions.

2. A change in the second (“implicit”) “cut-off boundary” also leads to a change in the sets of the IC factors taken as implicit. Moreover, the higher the “cut-off boundary”, the fewer implicit factors are included. When choosing implicit factors, the choice of the defuzzification method begins to play a role, but only for small values of both “cut-off boundaries.”

3. Changes in the sets of implicit factors when changing the second “cut-off boundary” largely depend on the selected first “cut-off boundary”, regardless of the defuzzification method.

4. Some factors can be defined as explicit (for some “cut-off boundaries”) and implicit (for other “cut-off boundaries”). This is related, firstly, to the requirements for the strength of the direct or indirect impact of the factor on the lagging indicators to assign it to a particular group, and secondly, to the linguistic uncertainty in formulating such requirements and expert evaluation of the strength of the impact. That is why it became necessary to develop a fuzzy model.

5. The proposed method of forming the causal field of IC indicators is generic in the sense that it is applicable to various types of organizations of different industry affiliations. The key IC indicators corresponding to its main structural components (human capital, organizational capital, relational capital), types of cognitive activity (education, involvement, production rationalization, self-improvement, customer-oriented rationalization, innovation), and the correspondence between types of cognitive activity and IC structural components are universal. All stages of the basic method scheme are universal as well.

6. At the same time, the sets of explicit and implicit IC factors for different organizations may differ significantly for the following reasons. Firstly, the set and composition of stakeholders in organizations and their requests to organizations can vary considerably. Consequently, the strategic maps of an organization’s objectives will differ significantly, including the objectives related to IC development and their lagging indicators (i.e., the initial set of IC development indicators from which explicit and implicit factors are selected). Even if the initial sets of indicators are relatively similar in composition, the degrees of impact of these indicators on key IC indicators and on each other can vary significantly. Finally, decision-makers may choose different “cut-off boundaries” and defuzzification methods.

Conclusion

A conceptual scheme for the formation of the causal field of the IC indicators in conjunction with the organization’s strategy and types of cognitive activity is proposed. The implementation of this scheme was carried out by developing a fuzzy economic-mathematical model that makes it possible to identify explicit and implicit factors of IC. The proposed scheme and model have the following distinctive features. The set of the IC indicators is formed based on the lagging indicators of strategic objectives selected from the objective map of the modified BSC grouped by six types of cognitive activity. The key IC indicators are the main structural components of the IC (human capital, organizational capital, relational capital). The explicit IC factors are selected based on the results of assessing the direct impact on the key IC indicators by setting a “cut-off boundary”. The implicit IC factors are selected based on the results of assessing the indirect impact on the key IC indicators through explicit factors by setting another “cut-off boundary”. Estimates of direct impact are carried out expertly in a given linguistic scale with the corresponding membership functions of fuzzy sets. Estimates of indirect impact are calculated based on operations with matrices whose elements are fuzzy numbers. The results of testing the model on the example of a university are presented. It is shown that the sets of explicit and implicit factors of the university’s IC vary depending on the given “cut-off boundaries” and the chosen defuzzification method. ■

Acknowledgments

The study was sponsored by the Russian Science Foundation (RSF) as part of research project No. 23-28-01091 (https://rscf.ru/project/23-28-01091/).

References

1. Kucharska W. (2022) Tacit knowledge impact on intellectual capital and innovativeness in the healthcare sector: A cross-country study of Poland and the US. J*ournal of Business Research*, vol. 149, pp. 869–883. https://doi.org/10.1016/j.jbusres.2022.05.059

2. Nazarov D.M. (2016) *Methodology of fuzzy set evaluation of implicit factors in organizational activities*. Ekaterinburg: Ural State Economic University Press (in Russian).

3. Ding Y., Li G. (2010) Study on the management of intellectual capital. *International Journal of Business and Management*, vol. 5, no. 2, pp. 213–216. https://doi.org/10.5539/ijbm.v5n2p213

4 Nazarov D.M. (2016) The evaluation model of implicit factors on the basis of fuzzy-set descriptions. *Far Eastern Federal University News*, no. 4 (80), pp. 3–17 (in Russian). https://doi.org/10.5281/zenodo.220793

5. Luria A.R. (1979) Language and consciousness. Moscow: Moscow University Press (in Russian).

6. Prikhodko A.I. (2012) Implicity as a means of latent evaluation. *Current Issues in Philology and Pedagogical Linguistics*, no. 14, pp. 48–54 (in Russian).

7. Miloserdova E.V. (2011) Implicitness as a linguo-pragmatic category and its role in different types of discourse. *Bulletin of Tomsk State University*, vol. 104, no. 12–2, pp. 727–732 (in Russian).

8. del-Río-Ortega A., Resinas Arias de Reyna M., Durán Toro A., Ruiz-Cortés A. (2012) Defining process performance indicators by using templates and patterns. *Business Process Management*, vol. 7481, pp. 223–228. https://doi.org/10.1007/978-3-642-32885-5_18

9. Tsiutsiura S.V., Kryvoruchko O.V., Tsiutsiura M.I. (2012) Theoretical basis and essence of managerial decisions. Models for accepting managerial decisions. *Development Management of Folding Systems*, no. 9, pp. 50–58 (in Ukrainian).

10. Nazarov D.M. (2015) The fundamental basis of implied in the system of economic development organizations. *Bulletin of Saint Petersburg State University of Economics*, no. 3(93), pp. 7–14 (in Russian).

11. Markić M., Požega Ž., Crnković B. (2022) The impact of knowledge management on the economic indicators of the companies. *South East European Journal of Economics and Business*, vol. 17, no. 2, pp. 34–48. https://doi.org/10.2478/jeb-2022-0013

12. Van Looy A., Shafagatova A. (2016) Business process performance measurement: A structured literature review of indicators, measures and metrics. *SpringerPlus*, vol. 5, pp. 1–24. https://doi.org/10.1186/s40064-016-3498-1

13. Nazarov D.M. (2017) Model of reflexive selection of implicit indicators of management activities of an organization. *Bulletin of Saint-Petersburg State University*, vol. 33, no. 3, pp. 498–518 (in Russian). https://doi.org/10.21638/11701/spbu05.2017.308

14. Lawrie G.J.G., Cobbold I. (2004) Third-generation balanced scorecard: Evolution of an effective strategic control tool. *International Journal of Productivity and Performance Management*, vol. 53, no. 7, pp. 611–623. https://doi.org/10.1108/17410400410561231

15. Olve N.-G., Roy J., Wetter M. (1999) *Performance drivers: A practical guide to using the balanced scorecard*. Chichester: Wiley.

16. Kaplan R.S., Norton D.P. (1992) The balanced scorecard — measures that drive performance. *Harvard Business Review*, vol. 70, no. 1, pp. 71–79.

17. Tawse A., Tabesh P. (2022) Thirty years with the balanced scorecard: What we have learned. *Business Horizons*, vol. 66, no. 1, pp. 123–132. https://doi.org/10.1016/j.bushor.2022.03.005

18. Pierce E. (2022) A balanced scorecard for maximizing data performance. *Frontiers in Big Data*, vol. 5, 821103. https://doi.org/10.3389/fdata.2022.821103

19. Kopecka N. (2015) The balanced scorecard implementation, integrated approach and the quality of its measurement. *Procedia Economics and Finance*, vol. 25, pp. 59–69. https://doi.org/10.1016/S2212-5671(15)00713-3

20. Mouritsen J., Larsen H., Bukh P. (2005) Dealing with the knowledge economy: intellectual capital versus balanced scorecard. *Journal of Intellectual Capital*, vol. 6, no. 1, pp. 8–27. https://doi.org/10.1108/14691930510574636

21. Johanson U., Martensson M., Skoog M. (2001) Mobilising change through the management control of intangibles. *Accounting, Organizations and Society*, vol. 26, no. 7/8, pp. 715–733. https://doi.org/10.1016/S0361-3682(01)00024-1

22. Petty R., Guthrie J. (2000) Intellectual capital literature review: measurement, reporting and management. *Journal of Intellectual Capital*, vol. 1, no. 2/3, pp. 155–176. https://doi.org/10.1108/14691930010348731

23. Bontis N., Dragonetti N.C., Jacobsen K., Roos G. (1999) The knowledge toolbox: A review of the tools available to measure and manage intangible resources. *European Management Journal*, vol. 17, no. 4, pp. 391–402.

24. Nedosekin A.O., Shkatov M.Yu., Abdulayeva Z.I. (2013) Developing a balanced scorecard system for a mixed marine oil and gas company using fuzzy sets. *Audit and Financial Analysis*, no. 4, pp. 126–134 (in Russian).

25. Horváth & Partners (eds.) (2007) *Balanced Scorecard umsetzen (Balanced Scorecard implementation)*, 4th ed. Stuttgart, Schäffer-Poeschel.

26. Kovalev D.V. (2006) The balanced scorecard (BSC) — a big right mistake? *Controlling*, vol. 2, no. 18, pp. 26–29 (in Russian).

27. Nørreklit H. (2003) The balanced scorecard: What is the score? A rhetorical analysis of the balanced scorecard. *Accounting, Organizations and Society*, vol. 28, no. 6, pp. 591–619. https://doi.org/10.1016/S0361-3682(02)00097-1

28. Sharabati A.A.A., Naji Jawad S., Bontis N. (2010) Intellectual capital and business performance in the pharmaceutical sector of Jordan. *Management Decision*, vol. 48, no. 1, pp. 105–131. https://doi.org/10.1108/00251741011014481

29. Lervik E. (2006). *Relational capital: A study on its importance, quantification and its impact on business sectors and markets*. The European Inter-University Association on Society, Science and Technology (ESST).

30. Montequin V.R., Fernandez F.O., Cabal V.A., Gutierrez N.R. (2006) An integrated framework for intellectual capital measurement and knowledge management implementation in small and medium-sized enterprises. *Journal of Information Science*, vol. 32, pp. 525–538. https://doi.org/10.1177/0165551506067127

31. Roos G., Pike S., Fernstrom L. (2006) *Intellectual capital: Practice of management*. Abingdon, Routledge.

32. Bontis N., Keow W.C.C., Richardson S. (2000) Intellectual capital and business performance in Malaysian industry. *Journal of Intellectual Capital*, vol. 1, no. 1, pp. 85–100.

33. Solodukhin K.S., Dzina G.A. (2009) Applying a balanced scorecard system at a university based on the stakeholder theory. *Controlling*, no. 1(29), pp. 12–23 (in Russian).

34. Solodukhin K.S. (2009) The implementation of a balanced scorecard system in a stakeholder company. *Controlling*, no. 2(30), pp. 64–69 (in Russian).

35. Korobova E., Kardovich I., Konysheva M., Mironova D. (2018) Cognitive activity: Philosophical analysis, psychological and pedagogical characteristics. *SHS Web of Conferences*, vol. 50, 01083. https://doi.org/10.1051/shsconf/20185001083

36. Korotaeva E.V., Nefedova A.N. (2012) Development of cognitive activity in senior preschoolers. *Pedagogical Education in Russia*, no. 3, pp. 176–180 (in Russian).

37. Nedoluzhko O.V. (2016) A new approach to forming the conceptual space of the intellectual capital phenomenon of an organization. *Bulletin of Udmurt University*, no. 2(26), pp. 42–49 (in Russian).

38. Bazil G.D., Adilova S.K., Abzhanova L.K., Sugurova L.A., Yerzhanova M.E. (2021) Fuzzy simulation of organizational adjustment processes management based on heat supply balanced scorecard. *Innovative Infrastructure Solutions*, vol. 6, pp. 77. https://doi.org/10.1007/s41062-020-00435-3

39. Biktimerov V.A. (2019) Methodology for analyzing digital content to identify patterns. Proceedings of the *BI-technologies and corporate information systems in optimizing business processes of the digital economy: VI International scientific-practical full-time and correspondence conference*. Yekaterinburg, pp. 30–31 (in Russian).

40. Hushko S., Temchenko O., Kryshtopa I., Temchenko H., Maksymova I., Huk O. (2018) Modelling of management activity of the organization considering the impact of implicit factors in business processes. *Eastern-European Journal of Enterprise Technologies*, vol. 3, no. 91, pp. 13–21. https://doi.org/10.15587/1729-4061.2018.121647

41. Pakseresht M., Seyyedi M.A., Zade M.M., Gardesh H. (2009) Business process measurement model based on the fuzzy multi-agent systems. *AIKED Proceedings of WSEAS*, pp. 501–506.

42. Chen S., Hwang C. (1992) *Fuzzy multiple attribute decision making: Methods and applications*. New York: Springer-Verlag.

43. Anshin V.M., Dyomkin I.V., Tsarkov I.V., Nikonov I.M. (2008) On application of fuzzy set theory to the problem of project portfolio selection. *Issues of Risk Analysis*, vol. 5, no. 3, pp. 8–21 (in Russian).

44. Vahidi J., Rezvani S. (2013) Arithmetic operations on trapezoidal fuzzy numbers. *Journal of Nonlinear Analysis and Application*, vol. 2013, pp. 1–8. https://doi.org/10.5899/2013/jnaa-00111

45. Raskin L., Sira O. (2020) Performing arithmetic operations over the (L-R)-type fuzzy numbers. *Eastern-European Journal of Enterprise Technologies*, vol. 3, no. 4 (105), pp. 6–11. https://doi.org/10.15587/1729-4061.2020.203590

About the authors

Georgij S. Zavalin

Head of Data Mining Department, Intern Researcher at Strategic Planning Laboratory, Vladivostok State University, 41, Gogolya Str., Vladivostok, 690014, Russia;

E-mail: georgiy.zavalin08@vvsu.ru

ORCID: 0000-0003-4519-0242

Olga V. Nedoluzhko

Cand. Sci. (Econ.), Associate Professor;

Associate Professor, Economics and Management Department, Associate Professor, Mathematics and Modeling Department, Researcher, Strategic Planning Laboratory, Vladivostok State University, 41, Gogolya Str., Vladivostok, 690014, Russia;

E-mail: olga.nedoluzhko25@gmail.com

ORCID: 0000-0002-3622-7501

Konstantin S. Solodukhin

Dr. Sci. (Econ.), Professor;

Professor, Mathematics and Modeling Department, Head of Strategic Planning Laboratory, Vladivostok State University, 41, Gogolya Str., Vladivostok, 690014, Russia;

E-mail: k.solodukhin@mail.ru

ORCID: 0000-0003-3619-1219

DOI: 10.17323/2587-814X.2023.3.70.86

Simulation model of an intelligent transportation system for the “smart city” with adaptive control of traffic lights based on fuzzy clustering

Armen L. Beklaryan [a, b]
E-mail: abeklaryan@hse.ru

Levon A. Beklaryan [b]
E-mail: beklar@cemi.rssi.ru

Andranik S. Akopov [b]
E-mail: akopovas@umail.ru

[a] HSE University
Address: 20, Myasnitskaya Street, Moscow 101000, Russia

[b] Central Economics and Mathematics Institute, Russian Academy of Sciences
Address: 47, Nakhimovsky Prospect, Moscow 117418, Russia

Abstract

This article presents a new simulation model of an intelligent transportation system (ITS) for the “smart city” with adaptive traffic light control. The proposed transportation model, implemented in the AnyLogic, allows us to study the behavior of interacting agents: vehicles (V) and pedestrians (P) within the framework of a multi-agent ITS of the "Manhattan Lattice" type. The spatial dynamics of agents in such an ITS is described using the systems of finite-difference equations with the variable structure, considering the controlling impact of the "smart traffic lights." Various methods of traffic light control aimed at maximizing the total traffic of the ITS output flow have been studied, in particular, by forming the required duration phases with the use of a genetic optimization algorithm, with a local ("weakly adaptive") switching control and based on the proposed fuzzy clustering algorithm. The possibilities of optimizing the characteristics of systems for individual control of the behavior of traffic lights under various scenarios, in particular, for the ITS with spatially homogeneous and periodic characteristics, are investigated. To determine the best values of individual parameters of traffic light control systems, such as the phases' durations, the radius of observation of traffic and pedestrian flows, threshold coefficients, the number of clusters, etc., the previously proposed parallel real-coded genetic optimization algorithm (RCGA type) is used. The proposed method of adaptive control of traffic lights based on fuzzy clustering demonstrates greater efficiency in comparison with the known methods of collective impact and local ("weakly adaptive") control. The results of the work can be considered a component of the decision-making system in the management of urban services.

Keywords: intelligent transportation system, "smart city", "smart traffic lights", agent-based modeling, adaptive control, fuzzy clustering, AnyLogic

Citation: Beklaryan A.L., Beklaryan L.A., Akopov A.S. (2023) Simulation model of an intelligent transportation system for the "smart city" with adaptive control of traffic lights based on fuzzy clustering. *Business Informatics*, vol. 17, no. 3, pp. 70–86. DOI: 10.17323/2587-814X.2023.3.70.86

Introduction

Currently, there is an increase in the need for the design and implementation of intelligent transport systems (ITS) for a "smart city" due to the ever-increasing traffic, causing the formation of multiple traffic jams. At the same time, one of the most promising directions of the ITS evolutionary development is the use of "smart traffic lights" that analyze the dynamics and structure of traffic and pedestrian flows [1].

Various approaches to rational traffic light management are known, in particular, based on information exchange [2] using machine learning methods with reinforcement [3] based on mixed integer programming [4, 5], using genetic and swarm optimization algorithms [6–8], as well as artificial neural networks (ANS), fuzzy logic, clustering and adaptive control for the ITS [11–13].

To study the behavior and optimize the characteristics of the ITS, various combined approaches are used, for example, agent-based and discrete-event modelling methods supported in AnyLogic [14, 15], joint control of traffic lights and vehicle trajectories [16], adaptive control based on a predictive model and reinforcement learning [17]. At the same time, most of these approaches are used for ITS with a simplified configuration, for example, for two consecutive intersections [15], one intersection consisting of two roads, etc. [17]. Various scenarios that determine the periodic dynamics of interacting transport and pedestrian routes are not considered.

As a rule, significant difficulties arise when managing the characteristics of the ITS with a more complex geometry of the "Manhattan Lattice" type [18, 19]. In such an ITS, inconsistent control of the states of at least one traffic light, as a rule, leads to a change in vehicle speed and traffic density on all connected routes. At the same time, in order to maximize the output traffic flow, it is necessary to effectively manage traffic lights, ensuring periodic prioritization between vehicles and pedestrians. So, for example, when a significant number of people gather at a regulated pedestrian crossing, the inclusion of a traffic light permitting signal is justified (a similar approach, in particular, has already been successfully applied in the street road network of some cities in Austria). At the same time the main purpose of "smart traffic lights" is to monitor traffic flows and select the optimal time points for switching control signals. The greatest difficulties in managing traffic flows are caused by the effect of "wave speed reduction" [20], when, as a result of a vehicle braking at a traffic light, all subsequent drivers inadvertently seek to increase the safe distance, contributing to the formation of traffic congestion. Therefore, it is necessary to study the heterogeneous spatial dynamics of agents and use data on the structure of traffic and pedestrian flows for adaptive traffic light control.

In this article, we propose a new simulation model of heterogeneous traffic flows in a "smart city" with adaptive traffic light behavior control based on fuzzy clustering. Within the framework of such a model, individual decisions on switching traffic light control signals are based on a fuzzy assessment of the traffic situation, including the evolutionary dynamics of both traffic and pedestrian flows (i.e. with equal priority in relation to cars and pedestrians). At the same time, an important task is solved to maximize the total traffic of the output stream under various scenarios, in particular, for the ITS with spatially homogeneous and periodic flow characteristics.

The scenarios presented in the paper, the corresponding optimal controls, as well as, in general, the proposed universal simulation model with the possibility of further modification of the studied geometry of intersections, as the authors see, can be considered to be an element of an integrated decision-making system in the management of urban services.

1. Description of the model

A key fragment of a multi-agent transport system of the Manhattan Lattice type is considered, consisting of four interconnected nodes-intersections that allow arbitrary change of vehicle directions, i.e. movement in a straight line, turns to the left and right, as well as a U-turn and movement in the opposite direction (*Fig. 1*).

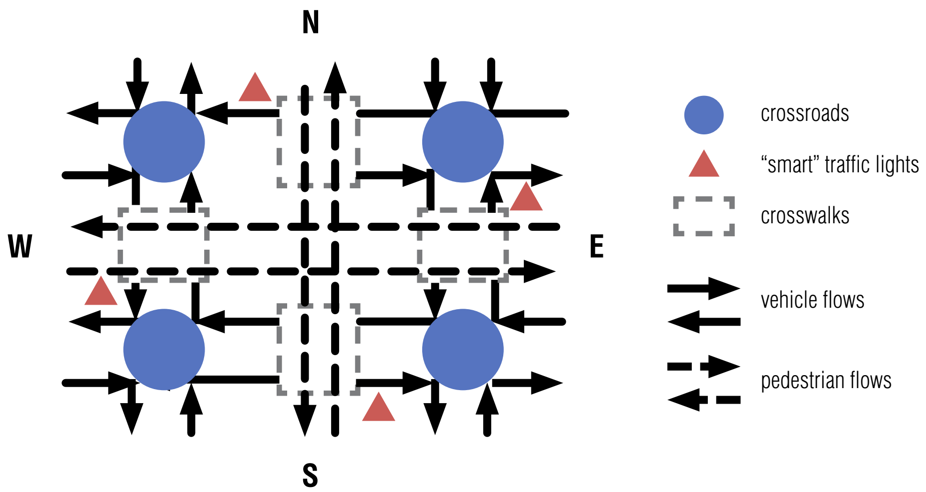


Fig. 1. General scheme of a multi-agent transport system of the Manhattan Lattice type with controlled traffic at pedestrian crossings.

Earlier, in [18, 19], the dynamics of traffic flows in Manhattan Lattice class systems was examined and various ways of their optimization were proposed, mainly based on the management of vehicle routes, i.e. the search and assignment of optimal routes for each agent-vehicle [18], including using genetic algorithms [19]. At the same time, one of the important ways to reduce the load of the ITS is to improve the manoeuvrability of the vehicle, including by choosing the least loaded traffic lanes, the determination of which is implemented using the fuzzy clustering algorithm [21, 22]. The existing methods of improving traffic flow are implemented mainly for unmanned vehicles (UVs), which can be "assigned" the optimal route depending on the current situation. The spatial dynamics of conventional vehicles (CVs) is most influenced by "smart traffic lights" that regulates the movement of traffic and pedestrian flows. In particular, they make it possible to effectively redistribute vehicle flows at intersections and pedestrian crossings, preventing the formation of traffic congestion.

Such traffic jams are formed mainly as a result of the "wave speed reduction." first studied in [20] and illustrated in *Fig. 2.* When braking a vehicle, for example, at the stop line of an adjustable pedestrian crossing or in front of the nearest obstacle in the form of another vehicle (*Fig. 2*), the car following it, as a rule, will brake harder in order to maintain a safe distance by increasing the radius of his personal space, due to the psychological characteristics of the driver's reaction. Further, the effect of an increase in the "safe" distance ("expansion" of personal space) spreads along the chain, reducing the flow rate as you move away from the original source of congestion (traffic lights), up to a complete stop.

To model the spatial dynamics of agents within the ITS (vehicles and pedestrians), systems of finite-difference equations with a variable structure can be used [21, 23]. This allows one to consider various scenarios of interaction of vehicles with each other and with the external environment (such as V2V, V2P, V2I, etc.) and the influence of the radius of each agent's personal space.

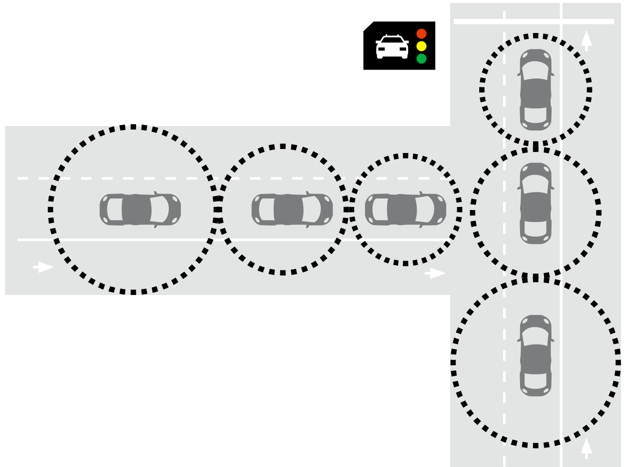


Fig. 2. Illustration of the effect of "wave speed reduction" in a road network with an adjustable pedestrian crossing.

Here is a brief formal description of the developed simulation model of vehicle movement, considering the influence of "smart traffic lights" regulating traffic and pedestrian traffic within the ITS.

Here,

$T = \{t_0, t_1, ..., |T|\}$ is the set of time moments (in minutes), $|T|$ is the total number of time moments; $t_0 \in T$, $t_{|T|} \in T$ are initial and final moments of time;

$L = \{l_0, l_1, ..., l_{|L|}\}$ is the set of indices of "smart traffic lights", where$|L|$ is the total number of "smart traffic lights";

$s_l(t_{k-1}) \in \{1, 2, 3\}$, $l \in L$ are the states of the phase of the l-th "smart traffic light" at the moment t_{k-1} ($t_{k-1} \in T$): $s_l(t_{k-1}) = 1$ is the prohibiting (for agents-vehicle) traffic light signal ("red"), $s_l(t_{k-1}) = 2$ is the warning signal of the traffic light ("yellow"), $s_l(t_{k-1}) = 3$ is the permitting (for agents-vehicle) traffic light signal ("green");

$\{\tau_{l1}, \tau_{l2}, \tau_{l3}\} \in T$, $l \in L$ is the duration of the phases of the l-th "smart traffic light" (in seconds) (*control parameter of the model*);

$\tilde{\tau}_l$ is the minimum required (to ensure safe traffic) duration of the main phase ("red" or "green") for "smart traffic lights" (in seconds) (control parameter of the model);

$\{P_l(t_{k-1}),\ V_l(t_{k-1})\}$ is the total number of pedestrians and vehicles, respectively, located in the monitoring zones of the l-th "smart traffic light" at the moment $t_{k-1}\,(t_{k-1} \in T)$;

$\{\tilde{V}_l(t_{k-1}),\ \tilde{D}_l(t_{k-1})\}$ is the total number of vehicles in clusters and the average inter-cluster distance calculated using the fuzzy clustering algorithm for agents located in the monitoring zones of the l-th smart traffic light at the moment$t_{k-1}\,(t_{k-1} \in T)$;

$\delta_l,\ l \in L$ is the threshold ratio between the number of pedestrians at the crossing regulated by the l-th smart traffic light and the total number of agent-vehicles planning to move this crossing (in any direction), at which it is necessary to turn on the traffic light permitting signal (control parameter of the model);

$\alpha_l,\ l \in L$ is the coefficient of significance of the average inter-cluster distance (for vehicles), estimated using the fuzzy clustering algorithm, when controlling traffic flows regulated by the l-th smart traffic light (control parameter of the model);

The phase status of the l-th "smart traffic light" $(l \in L)$ at the moment t_k $(t_k \in T)$ is set up according to the following rules:

$$s_l(t_k) = \begin{cases} 1, \text{ if I, II, or III is true,} \\ 2, \text{ if IV is true,} \\ 3, \text{ if V, VI, or VII is true,} \end{cases} \tag{1}$$

under conditions:

I. $(t_k \leqslant t_{k-1} + \tau_{l1}$ and $s_l(t_{k-1}) = 1)$ or $(t_k > t_{k-1} + \tau_{l2} + + \tau_{l3}$ and $s_l(t_{k-1}) = 3$ with the **first method** of controlling the duration of phases, based on the **collective impact on traffic lights**;

II. $((t_k \leqslant t_{k-1} + \tau_{l1}$ and $s_l(t_{k-1}) = 1)$ or $(t_k > t_{k-1} + \tau_{l2} + + \tau_{l3}$ and $s_l(t_{k-1}) = 3))$ or

$$\left(\frac{P_l(t_{k-1})}{V_l(t_{k-1})} > \delta_l \text{ and } s_l(t_{k-1}) = 3 \text{ and } t_k > t_{k-1} + \tilde{\tau}_l\right)$$

with the **second method** of controlling the duration of phases based on local ("weakly adaptive") control of traffic light switching, considering the **prioritization of pedestrian traffic**;

III. $((t_k \leqslant t_{k-1} + \tau_{l1}$ and $s_l(t_{k-1}) = 1)$ or

$(t_k > t_{k-1} + \tau_{l2} + \tau_{l3}$ and $s_l(t_{k-1}) = 3))$or

$$\left(\frac{P_l(t_{k-1})}{\tilde{V}_l(t_{k-1})\left(\tilde{D}_l(t_{k-1})\right)^{-\alpha_l}} > \delta_l \text{ and } s_l(t_{k-1}) = 3 \text{ and } t_k > t_{k-1} + \tilde{\tau}_l\right)$$

with the **third (adaptive) method** of controlling the duration of phases based on the fuzzy clustering algorithm, considering the **prioritization of pedestrian traffic**;

IV. $((t_k > t_{k-1} + \tau_{l1}$ and $s_l(t_{k-1}) = 1)$ or $(t_k > t_{k-1} + \tau_{l3}$ and $s_l(t_{k-1}) = 3))$ or $(t_k \leqslant t_{k-1} + \tau_{l2}$ and $s_l(t_{k-1}) = 2)$, which means that one of the main traffic lights ("red" or "green") has expired or continues to operate, the previously included warning ("yellow") signal;

V. $(t_k \leqslant t_{k-1} + \tau_{l3}$ and $s_l(t_{k-1}) = 3)$ or $(t_k > t_{k-1} + \tau_{l2} + \tau_{l1}$ and $s_l(t_{k-1}) = 1)$ with the **first method** of controlling the duration of phases, based on the **collective impact on traffic lights**;

VI. $((t_k \leqslant t_{k-1} + \tau_{l3}$ and $s_l(t_{k-1}) = 3)$ or

$(t_k > t_{k-1} + \tau_{l2} + \tau_{l1}$ and $s_l(t_{k-1}) = 1))$ or

$$\left(\frac{P_l(t_{k-1})}{V_l(t_{k-1})} < \delta_l \text{ and } s_l(t_{k-1}) = 1 \text{ and } t_k > t_{k-1} + \tilde{\tau}_l\right)$$

with the **second method** of controlling the duration of phases based on local ("weakly adaptive") control of traffic light switching, considering the **prioritization of traffic flow traffic**;

VII. $((t_k \leqslant t_{k-1} + \tau_{l1}$ and $s_l(t_{k-1}) = 3)$ or

$(t_k > t_{k-1} + \tau_{l2} + \tau_{l1}$ and $s_l(t_{k-1}) = 1))$ or

$$\left(\frac{P_l(t_{k-1})}{\tilde{V}_l(t_{k-1})\left(\tilde{D}_l(t_{k-1})\right)^{-\alpha_l}} < \delta_l \text{ and } s_l(t_{k-1}) = 1 \text{ and } t_k > t_{k-1} + \tilde{\tau}_l\right)$$

with the **third (adaptive) method** of controlling the duration of phases, based on the **fuzzy clustering algorithm**, considering the **prioritization of the transport stream traffic**.

The total number of vehicles and pedestrians located in the monitoring zones of the l-th "smart traffic light" $(l \in L)$ at the moment t_k $(t_k \in T)$ calculated with the second method of local ("weakly adaptive") switching control is equal to

$$V_l(t_k)=\sum_{i=1}^{|I|} v_{il}(t_k),\ P_l(t_k)=\sum_{\tilde{i}=1}^{|\tilde{I}|} p_{\tilde{i}l}(t_k), \qquad (2)$$

where

$$v_{il}(t_k)=\begin{cases}1, \text{ if } d_{il}(t_k)\le R_{l1},\\ 0, \text{ if } d_{il}(t_k)> R_{l1},\end{cases}$$
$$p_{\tilde{i}l}(t_k)=\begin{cases}1, \text{ if } \tilde{d}_{\tilde{i}l}(t_k)\le R_{l1},\\ 0, \text{ if } \tilde{d}_{\tilde{i}l}(t_k)> R_{l2},\end{cases} \qquad (3)$$

where

$I=\{i_1, i_2, ..., i_{|I|}\}$ is the set of indices of agent-vehicles, where $|I|$ is the total number of vehicles;

$\tilde{I}=\{\tilde{i}_1, \tilde{i}_2, ..., \tilde{i}_{|\tilde{I}|}\}$ is the set of indices of agent-pedestrians, where $|\tilde{I}|$ is the total number of pedestrians;

$\{R_{l1}, R_{l2}\}$, $l\in L$ are the radiuses of traffic monitoring zones for road and pedestrian traffic, respectively, for the l-th "smart traffic light" (*control parameter of the model*);

$\{d_{il}(t_k), \tilde{d}_{\tilde{i}l}(t_k)\}$, $i\in I, \tilde{i}\in\tilde{I}, l\in L$ is the distance from the i-th agent-vehicle and the $\tilde{i}$-th agent-pedestrian to the l-th "smart traffic light" at the momen t_k $(t_k\in T)$.

The total number of vehicles in clusters and the average inter-cluster distance for traffic flows located in the monitoring zones of the l-th "smart traffic light" $(l\in L)$ at the moment t_k $(t_k\in T)$ calculated with the third method of adaptive switching control using the fuzzy clustering algorithm are equal

$$\tilde{V}_l(t_k)=\sum_{i=1}^{|I|}\sum_{c_l=1}^{|C_l|}\tilde{v}_{ic_l}(t_k),\ \tilde{D}_l(t_k)=\frac{1}{|C_l|}\sum_{c_l=1}^{|C_l|}\sum_{\tilde{c}_l=1}^{|C_l|}\hat{d}_{c_l\tilde{c}_l}(t_k), \qquad (4)$$

where

$C_l=\{c_{l1}, c_{l2}, ..., c_{|C_l|}\}$, $l\in L$ is the set of cluster indices determined for the analysis of the traffic situation in the location area of the l-th "smart traffic light" using a fuzzy clustering algorithm, where $|C_l|$ is the total number of clusters (*control parameter of the model*);

$\tilde{v}_{ic_l}(t_k)$, $i\in I, \tilde{i}\in\tilde{I}, l\in L$ is the total number of vehicles belonging to the c_l-th cluster at the moment t_k $(t_k\in T)$;

$\hat{d}_{c_l\tilde{c}_l}(t_k)$, $c_l, \tilde{c}_l\in C_l$, $c_l\ne\tilde{c}_l$, $l\in L$ are pairwise distances between the centers of clusters belonging to the l-th "smart traffic light" at the moment t_k $(t_k\in T)$.

The spatial dynamics of vehicle agents and pedestrians can be modelled using systems of finite-difference equations with the variable structure, considering the regulatory impact of "smart traffic lights."

Here,

$\{x_{il}(t_k), y_{il}(t_k)\}$, $\{\tilde{x}_{\tilde{i}l}(t_k), \tilde{y}_{\tilde{i}l}(t_k)\}$, $i\in I$, $\tilde{i}\in\tilde{I}$, $l\in L$ are coordinates of the i-th agent-vehicle and the $\tilde{i}$-th agent-pedestrian located in the monitoring zone of the l-th "smart traffic light" at the moment t_k $(t_k\in T)$;

$\{v_i(t_{k-1}), \tilde{v}_{\tilde{i}}(t_{k-1})\}$, $i\in I, \tilde{i}\in\tilde{I}$ is the preferred speed of the $\tilde{i}$-th agent-vehicle and the $\tilde{i}$-th agent-pedestrian at the moment t_{k-1} $(t_{k-1}\in T)$;

$\{r_i(t_{k-1}), \tilde{r}_{\tilde{i}}(t_{k-1})\}$, $i\in I$, $\tilde{i}\in\tilde{I}$, $l\in L$ are the radius of personal spaces of the i-th agent-vehicle and the $\tilde{i}$-th agent-pedestrian, the values of which depend on the density of the transport (pedestrian) flow consisting of agents that reduce their speed and are located in the direction of travel (see *Fig. 2*) at the moment t_{k-1} $(t_{k-1}\in T)$;

$\{m_{ib}(t_k), \tilde{m}_{\tilde{i}b}(t_k)\}$, $i\in I$, $\tilde{i}\in\tilde{I}$, $b\in I\cup\tilde{I}$ is the distance from the i-th agent-vehicle and the $\tilde{i}$-th agent-pedestrian to the nearest b-th agent-obstacle at the moment t_{k-1} $(t_{k-1}\in T)$;

$\{w_i(t_{k-1}), \tilde{w}_{\tilde{i}}(t_{k-1})\}$, $\{q_i(t_{k-1}), \tilde{q}_{\tilde{i}}(t_{k-1})\}\in\{-1, 0, 1\}$, $i\in I$, are parameters that determine the direction of movement of the i-th agent-vehicle and the $\tilde{i}$-th agent-pedestrian at the moment t_{k-1} $(t_{k-1}\in T)$:

$w_i(t_{k-1})$, $\tilde{w}_{\tilde{i}}(t_{k-1})=-1$ when moving in the direction of the **E-W** (see *Fig. 1*),

$w_i(t_{k-1})$, $\tilde{w}_{\tilde{i}}(t_{k-1})=0$ when moving in the direction of the **N-S** or **S-N**,

$w_i(t_{k-1})$, $\tilde{w}_{\tilde{i}}(t_{k-1})=1$ when moving in the direction of the **W-E**,

$q_i(t_{k-1})$, $\tilde{q}_{\tilde{i}}(t_{k-1})=-1$ when moving in the direction of the **S-N**,

$q_i(t_{k-1}), \tilde{q}_{\tilde{i}}(t_{k-1}) = 0$ when moving in the direction of the **W-E** or **E-W**,

$q_i(t_{k-1}), \tilde{q}_{\tilde{i}}(t_{k-1}) = 1$ when moving in the direction of the **S-N**;

λ is the coefficient that specifies the ratio of the scales of real and model time.

The spatial dynamics of the i-th agent-vehicle ($i \in I$) and the i-th agent-pedestrian ($i \in I$), located in the monitoring zone of the l-th smart traffic light ($l \in L$) at the moment t_k ($t_k \in T$) without taking into account internal manoeuvring (associated with overtaking, lane changes, etc.) is given by the following system of finite difference equations with a variable structure:

$$x_{il}(t_k) = \begin{cases} x_{il}(t_{k-1}) + w_i(t_{k-1})\lambda v_i(t_{k-1}), \\ \text{if VIII is true,} \\ x_{il}(t_{k-1}), \text{ if IX is true,} \end{cases} \tag{5}$$

$$y_{il}(t_k) = \begin{cases} y_{il}(t_{k-1}) + q_i(t_{k-1})\lambda v_i(t_{k-1}), \\ \text{if VIII is true,} \\ y_{il}(t_{k-1}), \text{ if IX is true,} \end{cases} \tag{6}$$

$$\tilde{x}_{\tilde{i}l}(t_k) = \begin{cases} \tilde{x}_{\tilde{i}l}(t_{k-1}), \text{ if X is true,} \\ \tilde{x}_{\tilde{i}l}(t_{k-1}) + \tilde{w}_{\tilde{i}}(t_{k-1})\lambda \tilde{v}_{\tilde{i}}(t_{k-1}), \\ \text{if XI is true,} \end{cases} \tag{7}$$

$$\tilde{y}_{\tilde{i}l}(t_k) = \begin{cases} \tilde{y}_{\tilde{i}l}(t_{k-1}), \text{ if X is true,} \\ \tilde{y}_{\tilde{i}l}(t_{k-1}) + \tilde{q}_{\tilde{i}}(t_{k-1})\lambda \tilde{v}_{\tilde{i}}(t_{k-1}), \\ \text{if XI is true,} \end{cases} \tag{8}$$

$i \in I, \tilde{i} \in \tilde{I}, b \in I \cup \tilde{I}, l \in L,$

where

VIII. $s_l(t_{k-1}) = 3$ and $m_{ib}(t_{k-1}) > (r_i(t_{k-1}) + r_b(t_{k-1}))$ for the nearest agent ($b \in I \cup \tilde{I}$), which means that the permissive (for agent-vehicles) traffic light signal ("green") is in effect and there are no obstacles in the form of other vehicles or pedestrians on the way of the i-th agent-vehicle ($i \in I$);

IX. $s_l(t_{k-1}) = 1$ and $m_{ib}(t_{k-1}) \leqslant (r_i(t_{k-1}) + r_b(t_{k-1}))$ for the nearest agent ($b \in I \cup \tilde{I}$), which means that a prohibitor (for agent-vehicles) traffic light signal ("red") is in effect, or there is an obstacle in the form of another vehicle or a pedestrian on the way of the i-th agent-vehicle ($i \in I$);

X. $s_l(t_{k-1}) = 1$ and $\tilde{m}_{\tilde{i}b}(t_{k-1}) > (\tilde{r}_{\tilde{i}}(t_{k-1}) + r_b(t_{k-1}))$ for the nearest agent ($b \in I \cup \tilde{I}$), which means that a prohibitor (for agent-vehicles) traffic light signal ("red") is in effect and there are no obstacles in the form of other pedestrians or vehicles on the way of the $\tilde{i}$-th agent-pedestrian ($\tilde{i} \in \tilde{I}$);

XI. $s_l(t_{k-1}) = 1$ or $\tilde{m}_{\tilde{i}b}(t_{k-1}) \leqslant (\tilde{r}_{\tilde{i}}(t_{k-1}) + r_b(t_{k-1}))$ for the nearest agent ($b \in I \cup \tilde{I}$), which means that the permissive (for agent-vehicles) traffic light signal ("green") is in effect, or there is an obstacle in the form of another pedestrian or vehicle on the way of the $\tilde{i}$-th agent-pedestrian ($\tilde{i} \in \tilde{I}$).

The total traffic of the output stream that should be maximized is equal to

$$N = \sum_{t_{k=0}}^{|T|} \left(\sum_{i=1}^{|I|} n_i + \sum_{\tilde{i}=1}^{|\tilde{I}|} \tilde{n}_{\tilde{i}} \right), \tag{9}$$

where

$$n_i(t_k) = \begin{cases} 1, \text{ if } \{x_i(t_{k-1}), y_i(t_{k-1})\} \notin \{X, Y\}, \\ 0, \text{ if } \{x_i(t_{k-1}), y_i(t_{k-1})\} \in \{X, Y\}, \end{cases} \tag{10}$$

$$\tilde{n}_{\tilde{i}}(t_k) = \begin{cases} 1, \text{ if } \{\tilde{x}_{\tilde{i}}(t_{k-1}), \tilde{y}_{\tilde{i}}(t_{k-1})\} \notin \{X, Y\}, \\ 0, \text{ if } \{\tilde{x}_{\tilde{i}}(t_{k-1}), \tilde{y}_{\tilde{i}}(t_{k-1})\} \in \{X, Y\}, \end{cases} \tag{11}$$

where

$\{x_i(t_{k-1}), y_i(t_{k-1})\}, \{\tilde{x}_{\tilde{i}}(t_{k-1}), \tilde{y}_{\tilde{i}}(t_{k-1})\}, i \in I, \tilde{i} \in \tilde{I}$ are coordinates of the i-th agent-vehicle and the $\tilde{i}$-th agent-pedestrian within the ITS at the moment t_{k-1} ($t_{k-1} \in T$);

$\{X, Y\}$ is the set of all coordinates of the ITS digital road network.

Then, it is possible to formulate the following optimization problem to be solved considering the chosen method of controlling "smart traffic lights."

Problem A. The need to maximize the total traffic of the output flow by the set of control parameters $\{\tau_{l1}, \tau_{l2}, \tau_{l3}, \tilde{\tau}_l, \delta_l, R_{l1}, R_{l2}, |C_l|, \alpha_l\}$:

$$\max_{\{\tau_{l1}, \tau_{l2}, \tau_{l3}, \tilde{\tau}, \delta_l, R_{l1}, R_{l2}, |C_l|, \alpha_l\}} N \tag{12}$$

s.t.

$\tau_{l1}, \tau_{l2}, \tau_{l3}, \tilde{\tau}_l \in [\underline{\tau}, \overline{\tau}], \delta_l \in [\underline{\delta}, \overline{\delta}], R_{1l} \in [\underline{R}_1, \overline{R}_1], R_{2l} \in [\underline{R}_2, \overline{R}_2],$
$|C_l| \in [\underline{C}, \overline{C}], \alpha_l \in [\underline{\alpha}, \overline{\alpha}],$

where $\{\underline{\tau}, \underline{\delta}, \underline{R}_1, \underline{R}_2, \underline{C}, \underline{\alpha}\}$, $\{\overline{\tau}, \overline{\delta}, \overline{R}_1, \overline{R}_2, \overline{C}, \overline{\alpha}\}$ are the lower and upper boundary values of the control parameters of the model.

To solve **Problem A**, the previously proposed genetic optimization algorithm of real coding (RCGA class) is used [20, 25], aggregated by target functionality with the developed simulation model of the transport system implemented in AnyLogic.

2. Fuzzy clustering algorithm

To assess the structure of traffic flow and adaptive control of "smart traffic lights", it is proposed to use the fuzzy clustering algorithm (Fuzzy C-means) [21, 22, 26, 27]. The choice of this algorithm is primarily due to the possibility of considering various characteristics of moving vehicles in the formation of clusters, in particular, density, speed, distance from the traffic light regulating traffic at the transition, etc. The inclusion of such characteristics in cluster analysis makes it possible to achieve maximum "likelihood" when assessing the structure of the traffic flow. Unlike classical algorithms, Fuzzy C-means does not assign an object unambiguously to any cluster, but compares each cluster with the probability of assigning observed objects to it, forming a so-called membership matrix.

The enlarged scheme of the proposed fuzzy clustering algorithm is shown in *Fig. 3.* An important difference between the developed algorithm and those previously known is that its key characteristics (for example, the number of clusters, the radius of the traffic monitoring zone, etc.) are calculated using a genetic optimization algorithm (RCGA class) as part of solving the main problem of maximizing output stream traffic. As a result, the results of fuzzy clustering directly affect the possibilities of finding optimal solutions for the ITS being studied.

Figure 3 uses the following notation:

- $z \in [0, 1]$ is the measure of fuzziness;
- $M(k)$ is the membership matrix at the kth step of the algorithm, $k = 1, 2, ..., |K|$ where $|K|$ is the maximum number of iterations;
- ε is a small parameter that is a criterion for the algorithm stopping.

Thus, the proposed fuzzy clustering algorithm is aggregated by the target functional (the total traffic of the ITS output stream), with the real-coded genetic algorithm (*Fig. 3*). RCGA uses heuristic crossing-over and mutation operators (for example, LX, SBX, SNUM, see [21, 24, 25]) to form new potential solutions with the best characteristics. The Fuzzy C-means algorithm was built into the ITS simulation model implemented in AnyLogic and is executed at each step of the model time, providing an assessment of the structure of the traffic flow located in the monitoring area of each "smart traffic light."

3. Software implementation of the model

The key fragment of the software implementation of the proposed ITS simulation model performed in the AnyLogic environment is shown in *Fig. 4.*

An important feature of the software implementation of the model (*Fig. 4*) is the combined use of discrete-event and agent methods, including those supported in the AnyLogic traffic library [28, 29]. In particular, elements of the **carSource** and **pedSource** types provide generation of new agents and their addition to the corresponding populations of vehicles and pedestrians, elements of the **SelectOutput** type (s1, s2 in *Fig. 4b*) are used to distribute traffic flow along possible routes when the vehicle reaches intersections; **CarMoveTo** and **pedGoTo** elements move vehicle agents and pedestrians to a given goal, according to their predefined characteristics (preferred speed, intensity of arrival, etc.); **carDispose** and **pedSink** ensure the removal of agents from the corresponding populations and the calculation of the output traffic.

4. Results of optimization experiments

Optimization experiments were carried out for the ITS with spatially homogeneous and periodic flow characteristics with three methods of controlling "smart traffic lights":

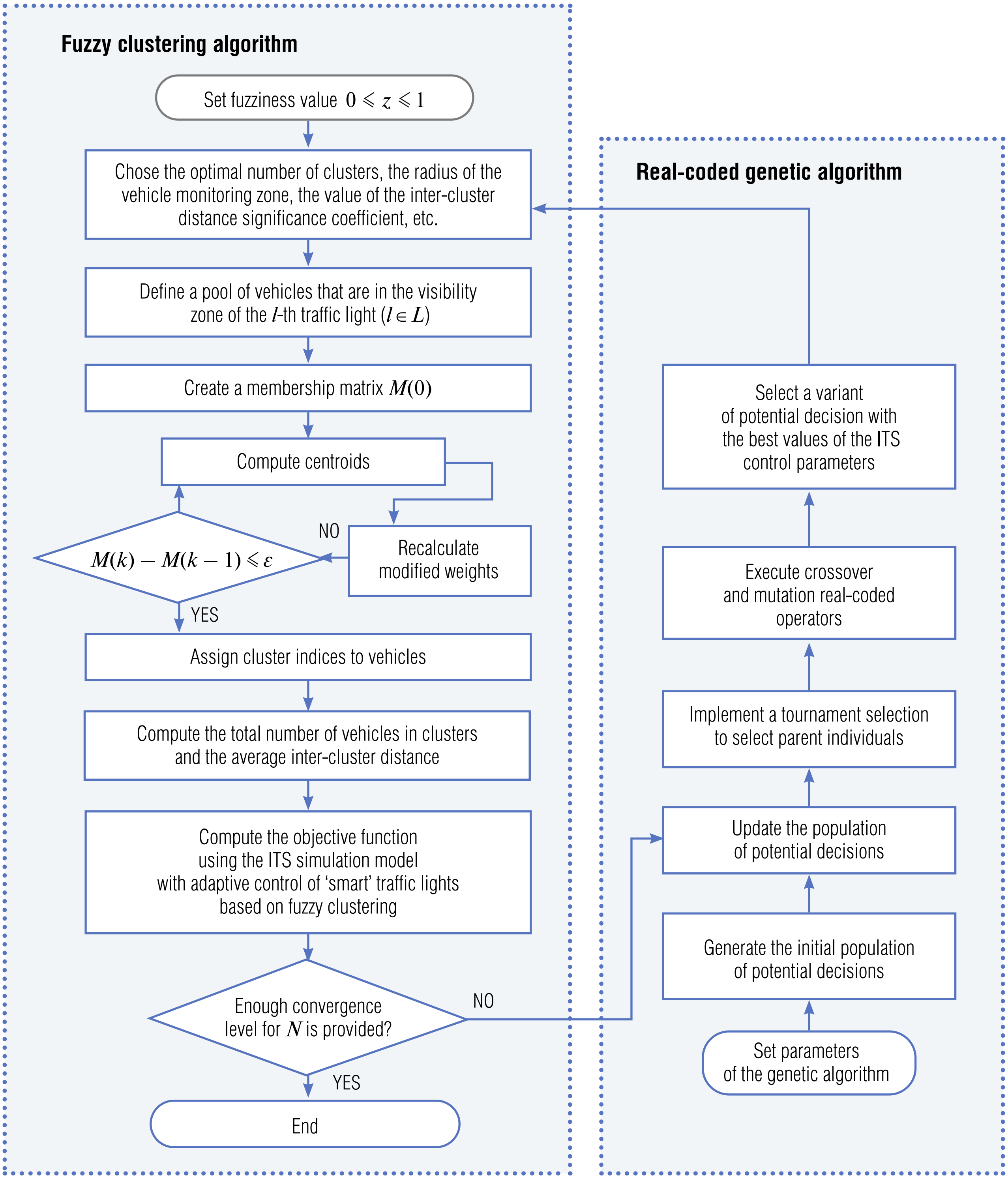


Fig. 3. Fuzzy clustering algorithm for adaptive traffic light control.

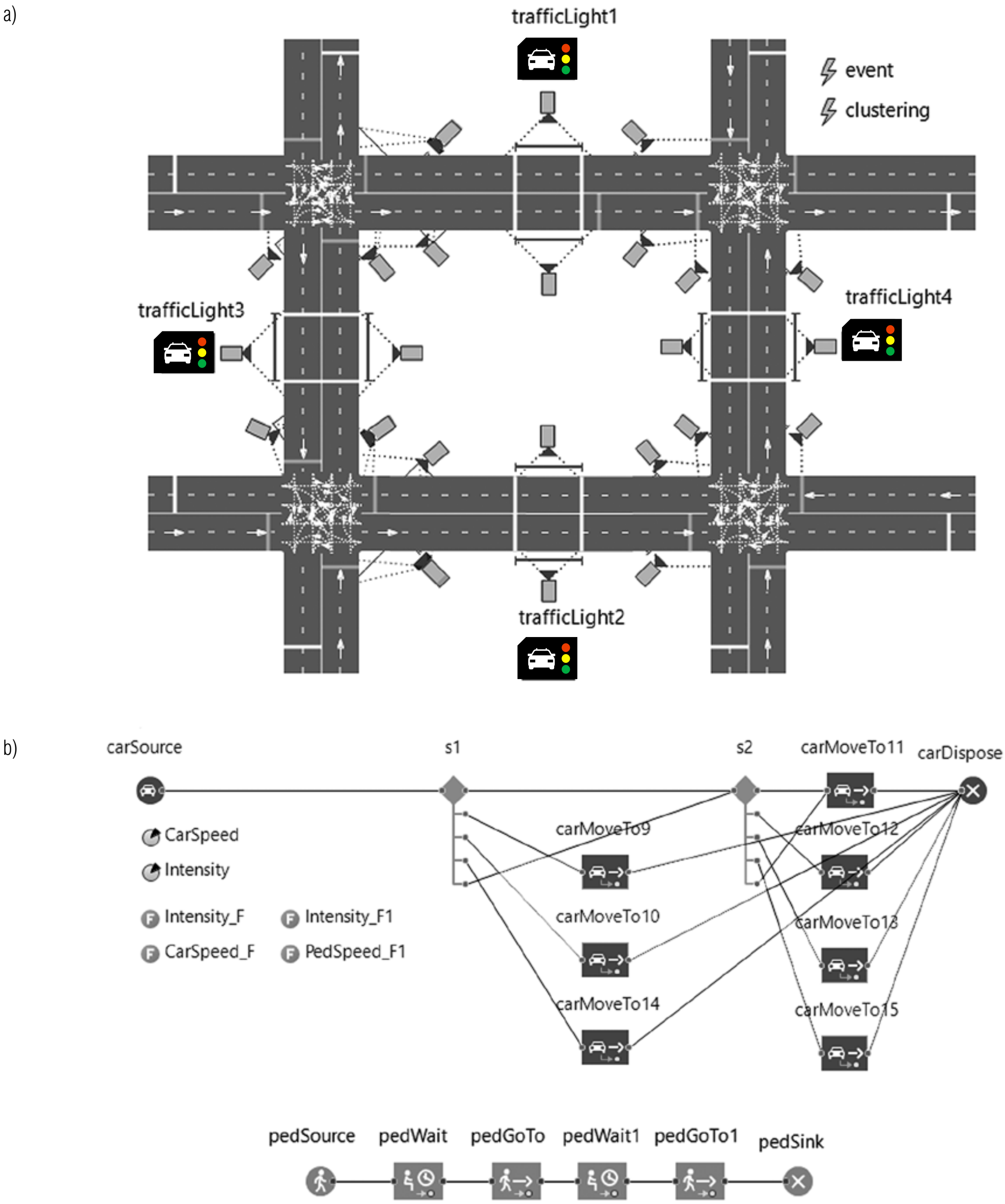


Fig. 4. Software implementation of the ITS simulation model in AnyLogic:
a) a diagram of a digital road network with "smart traffic lights"
b) a fragment of a discrete-event model of the movement of agents-vehicles and pedestrians along specified routes.

- by forming the required duration of phases using a genetic optimization algorithm;
- using local (“weakly adaptive”) switching control.
- based on the proposed fuzzy clustering algorithm.

For a system with spatially homogeneous flow characteristics, the intensity of arrival of agents and their preferred speeds are constants.

The intensity of the arrival of agents the ITS with periodic characteristics simulating the presence and absence of peak loads is calculated at each moment of time t_k ($t_k \in T$):

$$\theta(t_k) = \begin{cases} \text{truncnormal}(\hat{\theta},\ \sigma,\ \theta,\ \theta),\ \text{if}\ \ t_k \le \frac{1}{4}|T|, \\ \tilde{\theta},\ \text{if}\ \ \frac{1}{4}|T| < t_k < \frac{3}{4}|T|, \\ \text{truncnormal}(\hat{\theta},\ \sigma,\ \theta,\ \theta),\ \text{if}\ \ t_k \ge \frac{3}{4}|T|, \end{cases} \qquad (13)$$

where

$\text{truncnormal}(\hat{\theta}, \sigma, \underline{\theta}, \overline{\theta})$ is the random value of the intensity of the arrival of agents, set using a truncated normal distribution with the mean $\hat{\theta}$, standard deviation σ, lower and upper boundary values $\underline{\theta}$, $\overline{\theta}$, corresponding to the conditions of extreme traffic;

$\tilde{\theta}$ is the intensity of arrival corresponding to the conditions of normal traffic;

In a similar way, the average speeds of the agents are set. The main model assumptions (initial data) are presented in *Table 1*.

At the first stage, using the Monte Carlo type method [30], numerical experiments were carried out to assess the sensitivity of the target functional (the total traffic of the output stream) with respect to the values of the ITS control parameters with spatially homogeneous and periodic flow characteristics (*Fig. 5*).

It follows from *Fig. 5* that the total traffic of the output stream is sensitive with respect to the values of the ITS control parameters, both with spatially homogeneous and periodic flow characteristics. At the same time, the most likely ranges of values of the total traffic of the output stream are 1800–1900 agents (vehicles and pedestrians).

Table 1.

Initial data of the simulation model

No.	Model parameters		Values	
1	Length and width of roads, m.		155	
2	Number of intersections		4	
3	Distance between adjacent intersections, m.		65	
4	Number of traffic lanes for each road		2	
5	The width of the dividing strip, m.		2	
6	The number of pedestrian crossings regulated by “smart traffic lights”		4	
7	Simulation period, min.		20	
			vehicles	pedestrians
8	The intensity of the arrival of vehicles and pedestrians to the ITS with spatially homogeneous characteristics at each entrance of the road network (agents per hour)		500	1000
9	Preferred speed of vehicles and pedestrians within the ITS with spatially homogeneous characteristics (km/h for vehicles and m/s for pedestrians)		100	0.75
10	Parameters for calculating the intensity of arrival of vehicles and pedestrians to the ITS with periodic flow characteristics (agents per hour)	$\hat{\theta}$	500	1000
		σ	100	100
		$\underline{\theta}$	100	500
		$\overline{\theta}$	1500	1500
		$\tilde{\theta}$	100	500
11	Parameters for calculating the preferred speed of vehicles and pedestrians within the ITS with periodic flow characteristics (km/h for vehicles and m/s for pedestrians)	$\hat{\theta}$	45	1.3
		σ	10	0.5
		$\underline{\theta}$	20	1
		$\overline{\theta}$	60	1.5
		$\tilde{\theta}$	100	1

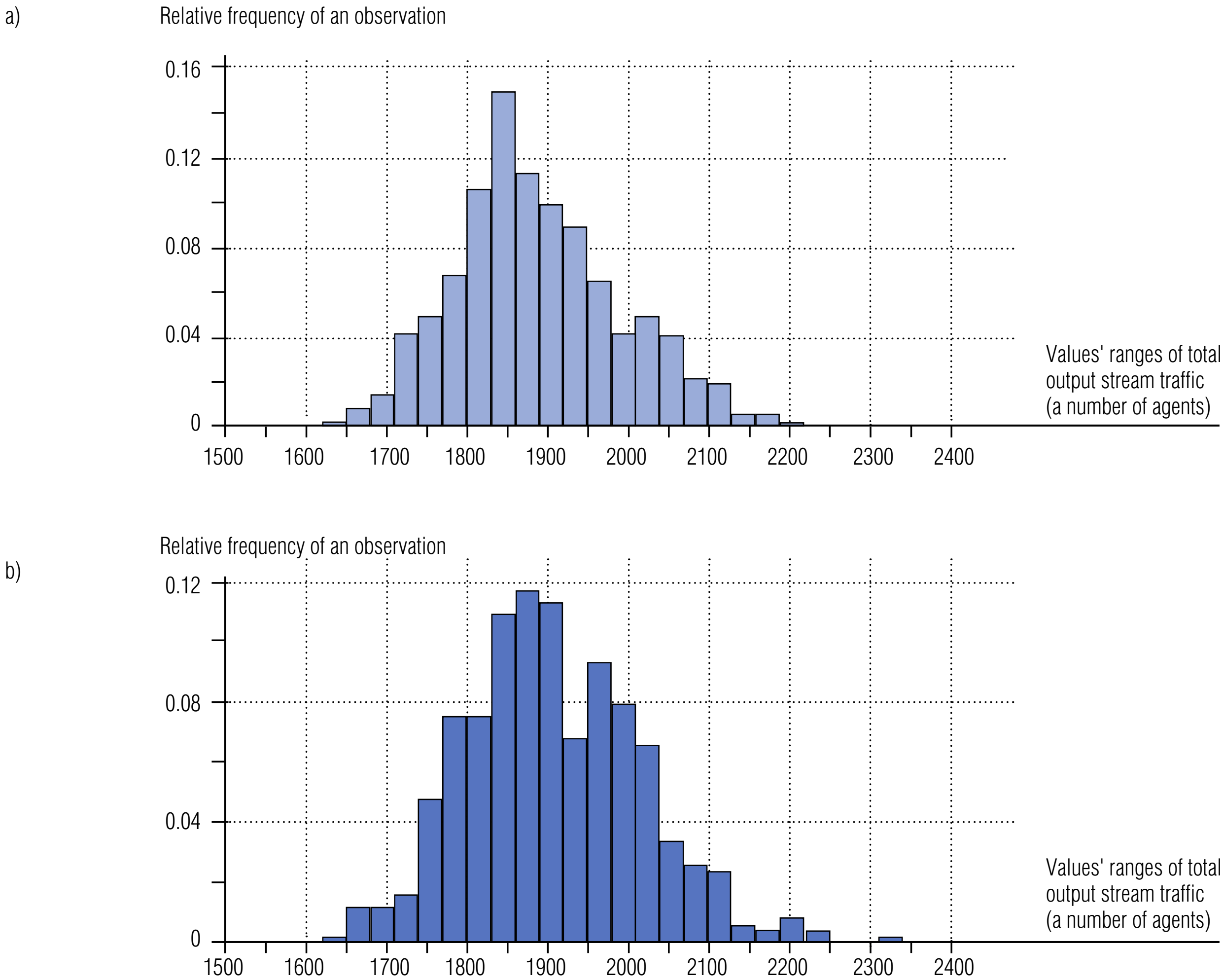


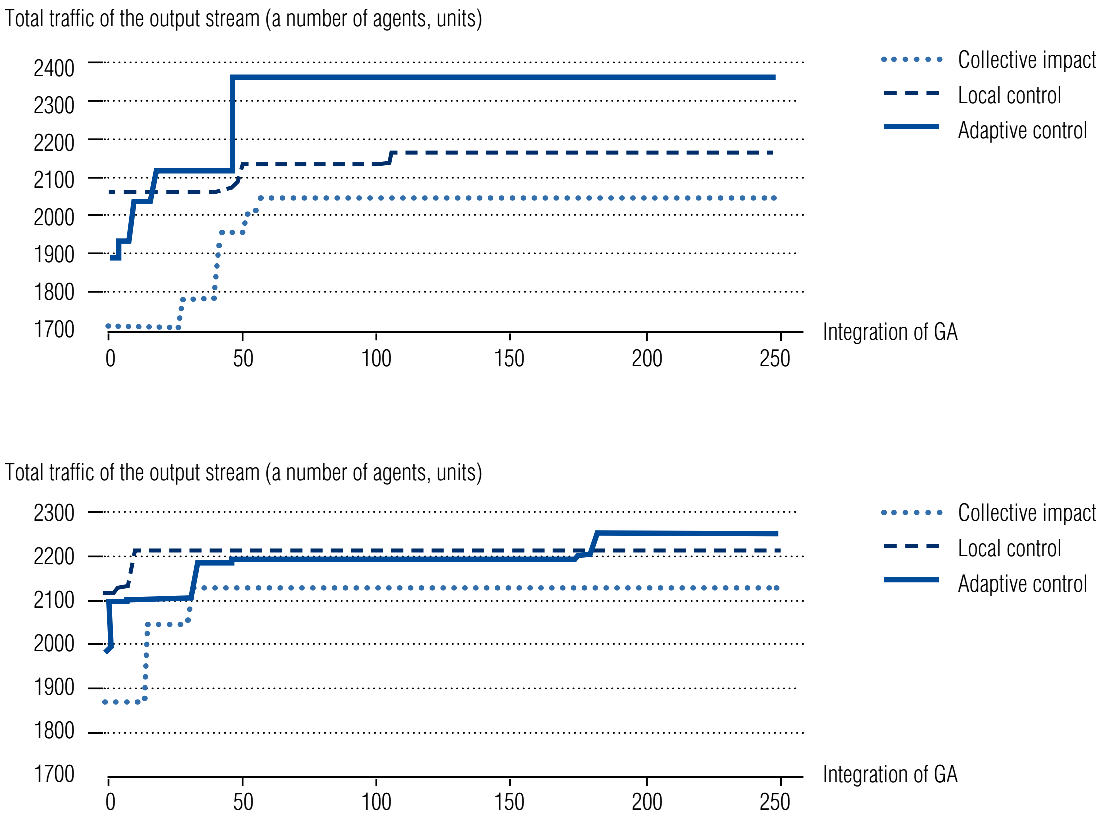
Fig. 5. Estimation of sensitivity (probability density) of the total output stream traffic for the ITS: a) with spatially homogeneous and b) periodic flow characteristics.

Figure 6 shows the dynamics of convergence of the objective function obtained using the developed simulation model aggregated by the target functional with a genetic algorithm (GA).

The maximum possible values of the total output stream for the ITS with periodic flow characteristics are, on average, less than for the ITS with spatially homogeneous characteristics (*Fig. 5*). The obtained suboptimal values of the control parameters of the model corresponding to the scenarios of the ITS implementation in an enlarged form discussed above are presented in *Table 2*.

It follows from *Fig. 6* and *Table 2* that the most promising way to control "smart traffic lights" is adaptive switching control based on fuzzy clustering. The proposed approach demonstrates its effectiveness even for the ITS with periodic flow characteristics, providing the best final value of the objective function.

a)



b)

Fig. 6. Dynamics of convergence of the objective function (total traffic of the output stream) for the ITS: a) with spatially homogeneous and b) periodic flow characteristics.

Conclusion

This article presents a new simulation model of the intelligent transport system (ITS) of a "smart city" with adaptive traffic light control. We propose a model of the movement of a vehicle ensemble using systems of finite-difference equations with a variable structure, considering the regulatory effect of "smart traffic lights." To assess the structure of traffic flow and adaptive control of "smart traffic lights." a fuzzy clustering algorithm is proposed, the key characteristics of which are calculated using a genetic optimization algorithm (RCGA class) as part of solving the main task of maximizing output traffic. With the help of the developed simulation model, the possibilities of rational management of "smart traffic lights" are investigated, in particular, for ITS with spatially homogeneous and periodic characteristics. As a result, a model example demonstrates the greater efficiency of adaptive switching control based on fuzzy clustering.

Further research will be aimed at designing the large-scale agent-based model of the ITS "smart city" using the FLAME GPU. ■

Acknowledgments

The research was carried out based on a grant from the Russian Science Foundation (project No. 23-11-00080).

Table 2.

The obtained suboptimal values of the control parameters of the model

			A system with spatially homogeneous flow characteristics	A system with periodic flow characteristics
Collective impact				
Total traffic, agents			**2392**	**2125**
Duration of traffic light phases, sec.	first and second traffic lights	red	11.327	22.675
		green	25.816	18.061
	third and fourth traffic lights	red	10.136	18.768
		green	11.933	103.071
	all traffic lights	yellow	1.098	1.874
Local management				
Total traffic, agents			**2166**	**2209**
Minimum required duration of the main phase, min.			1.839	4.918
Radius of the traffic monitoring zone, m.			31.59	12.25
Radius of the pedestrian traffic monitoring zone, m.			19.13	2.56
Threshold ratio between the number of pedestrians at the crossing and the total number of vehicles			88.5	241.7
Adaptive management based on fuzzy clustering				
Total traffic, agents			**2357**	**2246**
Minimum required duration of the main phase, min.			1.514	3.555
Radius of the traffic monitoring zone, m.			26.11	17.89
Radius of the pedestrian traffic monitoring zone, m.			20.21	25.01
Threshold ratio between the number of pedestrians at the crossing and the total number of vehicles (adjusted for inter-cluster distance)			162.5	255.1
Coefficient of significance of the average inter-cluster distance			0.602	1
Number of clusters			3	3

References

1. Li L., Wen D., Yao D. (2014) A survey of traffic control with vehicular communications. *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 1, pp. 425–432. https://doi.org/10.1109/TITS.2013.2277737

2. Tang C., Wei X., Hao M., Zhu C., Wang R., Chen W. (2018) Traffic signal phase scheduling based on device-to-device communication. *IEEE Access,* vol. 6, pp. 47636–47645. https://doi.org/10.1109/ACCESS.2018.2867553

3. Tong W., Pan Z., Liu K., Yali Y., Xiumin W., Huawei H., Wu D.O. (2020) Multi-agent deep reinforcement learning for urban traffic light control in vehicular networks. *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 8243–8256. https://doi.org/10.1109/TVT.2020.2997896

4. Lin W-H., Wang C. (2004) An enhanced 0-1 mixed-integer LP formulation for traffic signal control. *IEEE Transactions on Intelligent Transportation Systems,* vol. 5, no. 4, pp. 238–245. https://doi.org/10.1109/TITS.2004.838217

5. Gartner N.H., John D.C., Gabbay L.H. (1975) Optimization of traffic signal settings by mixed-integer linear programming. *Transportation Science*, vol. 9, no. 4, pp. 321–343. https://doi.org/10.1287/trsc.9.4.321

6. Davydov I., Tolstykh D., Kononova P., Legkih I. (2019) Genetic based approach for Novosibirsk traffic light scheduling. Proceedings of the *15th International Asian School-Seminar Optimization Problems of Complex Systems (OPCS), Novosibirsk, Russia*, pp. 31–36. https://doi.org/10.1109/OPCS.2019.8880158

7. Turky A.M., Ahmad M.S., Yusoff M.Z.M., Sabar N.R. (2009) Genetic algorithm application for traffic light control. *Information Systems: Modeling, Development, and Integration. UNISCON 2009. Lecture Notes in Business Information Processing*, vol. 20. (eds. J. Yang, A. Ginige, H.C. Mayr, R.D. Kutsche). Berlin: Springer, Heidelberg, pp. 115–120. https://doi.org/10.1007/978-3-642-01112-2_12

8. García-Nieto J., Alba E., Carolina Olivera A. (2012) Swarm intelligence for traffic light scheduling: Application to real urban areas. *Engineering Applications of Artificial Intelligence*, vol. 25, no. 2, pp. 274–283. https://doi.org/10.1016/j.engappai.2011.04.011

9. Royani T., Haddadnia J., Alipoor M. (2013) Control of traffic light in isolated intersections using fuzzy neural network and genetic algorithm. *International Journal of Computer and Electrical Engineering*, vol. 5, no. 1, pp. 142–146. https://doi.org/10.7763/IJCEE.2013.V5.682

10. Agramelal F., Sadik M., Hannani A.E., Moubarak Y. (2022) A traffic-aware street lighting system based on fuzzy logic controller. Proceedings of the *2022 IEEE 18th International Colloquium on Signal Processing & Applications (CSPA), Selangor, Malaysia*, pp. 132–137. https://doi.org/10.1109/CSPA55076.2022.9782045

11. Rashid H., Ashrafi M.J.F., Azizi M., Heydarinezhad M.R. (2015) Intelligent traffic light control based on clustering using vehicular ad-hoc networks. Proceedings of the *2015 7th Conference on Information and Knowledge Technology (IKT), Urmia, Iran*, pp. 1–6. https://doi.org/10.1109/IKT.2015.7288801

12. Sachan A., Kumar N. (2023) SDN control-enabled and time-quantum-driven max-pressure approach for intersection management in smart city. *IEEE Systems Journal,* vol. 17, no. 1, pp. 1694–1702. https://doi.org/10.1109/JSYST.2022.3211933

13. Marino F., Leccese F., Pizzuti S. (2017) Adaptive street lighting predictive control. *Energy Procedia,* vol. 111, pp. 790–799. https://doi.org/10.1016/j.egypro.2017.03.241

14. Makarov V.L., Bakhtizin A.R., Epstein J.M. (2022) *Agent-based modeling for a complex world*. Moscow: Scientific publications department, GAUGN. https://doi.org/10.18254/978-5-604-5843-5-4

15. Veremchuk N.S. (2022) Elements of simulation modeling in issues of traffic optimization. *Vestnik kibernetiki*, vol. 4, no. 48, pp. 23–28 (in Russian). https://doi.org/10.34822/1999-7604-2022-4-23-28

16. Agafonov A., Yumaganov A., Myasnikov V. (2023) Cooperative control for signalized intersections in intelligent connected vehicle environments. *Mathematics,* vol. 11, no. 6, 1540. https://doi.org/10.3390/math11061540

17. Myasnikov V.V., Agafonov A.A., Yumaganov A.S. A deterministic predictive traffic signal control model in intelligent transportation and geoinformation systems. *Computer Optics*, vol. 45, no. 6, pp. 917–925 (in Russian). https://doi.org/10.18287/2412-6179-CO-1031

18. Sarma S.S., Sinha K., Sub-r-pa C., Chakraborty G., Sinha B.P. (2021) Optimal distribution of traffic in Manhattan road networks for minimizing routing-time. *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 11, pp. 6799–6820. https://doi.org/10.1109/TITS.2020.2994836

19. Akopov A.S., Beklaryan L.A., Beklaryan A.L. Simulation-based optimisation for autonomous transportation systems using a parallel real-coded genetic algorithm with scalable nonuniform mutation. *Cybernetics and Information Technologies*, 2021, vol. 21, no. 3, pp. 127–144. https://doi.org/10.2478/cait-2021-0034

20. Sugiyama Y., Fukui M., Kikuchi M., Hasebe K., Nakayama A., Nishinari K., Tadaki S., Yukawa S. (2008) Traffic jams without bottlenecks – experimental evidence for the physical mechanism of the formation of a jam. *New Journal of Physics*, vol. 10, no. 3, 033001. https://doi.org/10.1088/1367-2630/10/3/033001

21. Akopov A.S., Beklaryan L.A., Thakur M. (2022) Improvement of maneuverability within a multiagent fuzzy transportation system with the use of parallel biobjective real-coded genetic algorithm. *IEEE Transactions on Intelligent Transportation Systems*. vol. 23, no. 8. pp. 12648–12664. https://doi.org/10.1109/TITS.2021.3115827

22. Akopov A.S., Khachatryan N.K., Beklaryan L.A., Beklaryan A.L. (2020) Unmanned vehicle control system based on fuzzy clustering. Part 2. Fuzzy clustering and software implementation. *Vestnik komp'iuternykh i informatsionnykh tekhnologii,* vol. 17, no. 10(196), pp. 21–29. https://doi.org/10.14489/vkit.2020.10.pp.021-029

23. Akopov A.S., Beklaryan L.A. (2022) Simulation of rates of traffic accidents involving unmanned ground vehicles within a transportation system for the 'smart city'. *Business Informatics*, vol. 16, no. 4, pp. 19–35. https://doi.org/10.17323/2587-814X.2022.4.19.35

24. Herrera F., Lozano M., Verdegay J.L. (1998) Tackling real-coded genetic algorithms: Operators and tools for behavioural analysis. *Artificial Intelligence Review*, vol. 12, no. 4, pp. 265–319. https://doi.org/10.1023/A:1006504901164

25. Akopov A.S., Beklaryan L.A., Thakur M., Verma B.D. (2019) Parallel multi-agent real-coded genetic algorithm for large-scale black-box single-objective optimisation. *Knowledge-Based Systems*, vol. 174, pp. 103–122. https://doi.org/10.1016/j.knosys.2019.03.003

26. Bezdek J.C. (1974) Cluster validity with fuzzy sets. *Journal of Cybernetics*, vol. 3, no. 3, pp. 58–73. https://doi.org/10.1080/01969727308546047

27. Beklaryan A.L., Akopov A.S. (2016) Simulation of agent-rescuer behaviour in emergency based on modified fuzzy clustering. Proceedings of the *2016 International Conference on Autonomous Agents & Multiagent Systems (AAMAS'16). International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC*, pp. 1275–1276.

28. Kuklová J. (2021) Highway modeling in anylogic for multi-agent approach to smart city management. Proceedings of the *2021 Smart City Symposium Prague (SCSP), Prague, Czech Republic*, pp. 1–6. https://doi.org/10.1109/SCSP52043.2021.9447402

29. Oo T.M., Lupin S., Khaing M.T., Thu A. (2021) Using hybrid modeling for estimation of the efficiency of service systems. Proceedings of the *IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), St. Petersburg, Moscow, Russia,* pp. 2277–2280. https://doi.org/10.1109/ElConRus51938.2021.9396596

30. Metropolis N. (1987) The beginning of the Monte Carlo method. *Los Alamos Science Special Issue*, vol. 15, pp. 125–130.

About the authors

Armen L. Beklaryan

Cand. Sci. (Tech.);

Associate Professor, Department of Business Informatics, HSE University, 20, Myasnitskaya Street, Moscow 101000, Russia;

Senior Researcher, Laboratory of Social Modeling, Central Economics and Mathematics Institute, Russian Academy of Sciences, 47, Nakhimovsky Prospect, Moscow 117418, Russia;

Email: abeklaryan@hse.ru

ORCID: 0000-0001-5141-5140

Levon A. Beklaryan

Dr. Sci. (Phys.-Math.), Professor;

Chief Researcher, Laboratory of Dynamic Models of Economy and Optimization, Central Economics and Mathematics Institute, Russian Academy of Sciences, 47, Nakhimovsky Prospect, Moscow 117418, Russia;

E-mail: beklar@cemi.rssi.ru

ORCID: 0000-0001-5695-4450

Andranik S. Akopov

Dr. Sci. (Tech.), Professor; Professor of the Russian Academy of Sciences;

Chief Researcher, Laboratory of Dynamic Models of Economy and Optimization, Central Economics and Mathematics Institute, Russian Academy of Sciences, 47, Nakhimovsky Prospect, Moscow 117418, Russia;

E-mail: akopovas@umail.ru

ORCID: 0000-0003-0627-3037

DOI: 10.17323/2587-814X.2023.3.87.100

Forecasting financial time series using singular spectrum analysis*

Anna V. Zinenko
E-mail: anna-z@mail.ru

Siberian Federal University
Address: 79, Svobodny Prospect, Krasnoyarsk 660041, Russia

Abstract

Financial time series are big arrays of information on quotes and trading volumes of shares, currencies and other exchange and over-the-counter instruments. The analysis and forecasting of such series has always been of particular interest for both research analysts and practicing investors. However, financial time series have their own features, which do not allow one to choose the only correct and well-functioning forecasting method. Currently, machine-learning algorithms allow one to analyze large amounts of data and test the resulting models. Modern technologies enable testing and applying complex forecasting methods that require volumetric calculations. They make it possible to develop the mathematical basis of forecasting, to combine different approaches into a single method. An example of such a modern approach is the Singular Spectrum Analysis (SSA), which combines the decomposition of a time series into a sum of time series, principal component analysis and recurrent forecasting. The purpose of this work is to analyze the possibility of applying SSA to financial time series. The SSA method was considered in comparison with other common methods for forecasting financial time series: ARIMA, Fourier transform and recurrent neural network. To implement the methods, a software algorithm in the Python language was developed. The method was also tested on the time series of quotes of Russian and American stocks, currencies and cryptocurrencies.

* The article is published with the support of the HSE University Partnership Programme

Keywords: non-stationary time series, forecasting, singular spectrum analysis, error metrics

Citation: Zinenko A.V. (2023) Forecasting financial time series using singular spectrum analysis. *Business Informatics*, vol. 17, no. 3, pp. 87–100. DOI: 10.17323/2587-814X.2023.3.87.100

Introduction

Datasets on the market prices of various financial instruments, such as stocks, currencies, derivatives and precious metals can be referred to as financial time series. Time series of this type have certain peculiarities which have to be taken into account in the application of various forecasting methods. The background investment theories, such as the Markowitz and SHARP model or the Black-Scholes option pricing model [1], implied that market time series were random and followed the law of normal distribution. However, simultaneous investigations carried out by Mandelbrot [1, 2] and by Peters [3, 4] proved price time series to be non-random and, consequently, non-stationary; thus, for their forecast, the application of the methods based on the assumption of random processes would not give any adequate result.

The following peculiarities of financial time series can be specified.

1. As is indicated above, time series of market prices are non-stationary. This means that the average value and variance are unstable within the interval under study. Therefore, before applying to these time series the methods which function well with random processes, they have to be transformed to the stationary form. This can be done, for example, by differencing, as is suggested by the ARIMA model.
2. Financial time series are persistent. This implies that subsequent parameters strongly depend on the previous ones. Close to this situation is the sensitivity to the initial conditions, which is typical for chaos. The Hurst exponent [3] allows one to determine whether the time series is persistent. If this exponent fluctuates within the limits from 0.5 to 1, then this time series is persistent. Hurst referred to the processes of this type as "long memory processes." The calculation of the Hurst exponent for numerous time series confirms their persistent character [4, 5].
3. There is a large amount of open access information on market time series, which makes it possible to apply methods of machine learning and data analysis to these time series (including deep learning which requires a big volume of training and test sets). Analysts and traders successfully use trading algorithms based on artificial neural networks and other machine learning methods. Financial data are available for various time intervals, from minutes to weeks. On the one hand, dealing with minute data for several years one can obtain a huge amount for training a neural network. However, on the other hand, such sets can be rather noisy [6].
4. Financial time series are non-differentiable, though continuous. If we consider a graph of stock market quotes, it is possible to see that it is not smooth. This means that it is impossible to draw a tangent to it, and thus, to calculate a derivative in any point. An example of a non-differentiable continuous function is the Weierstrass function [1], which resembles a no-trend stock exchange series. This peculiarity has to be taken into account in the application of certain methods, for example, the Fourier transform, which decomposes a time series into a sum of trigonometric functions. In the case of applying the Fourier transform to non-differentiable series, the original series has to be smoothed, for example, by a moving average.

All the above mentioned features should be considered when choosing methods to forecast financial time series. For example, long memory can be taken into account by certain types of recurrent neural networks, such as LSTM neural networks. In the process of learning, they are capable of choosing from the past data those which have the most significant influence on the values being forecast.

To test the forecast adequacy and to compare the forecasting methods, the time series is to be divided into the training and test components. By using the training component, the algorithm learns (for example, ARIMA parameters, Fourier coefficients, regression coefficients are chosen etc.). Then, the model obtained for the training time series is applied for forecasting and the predictive time series is compared with the test series using metrics [7].

All the internal metrics are based on estimating the variance between actual and predictive values. The mean absolute error (MAE) shows the average variance between actual and predictive values; the mean squared error (MSE) is the average of the squared differences; and the mean absolute percentage error (MAPE) is the average of the relative differences. MAPE is calculated by the following formula:

$$\mathrm{MAPE} = \frac{1}{n}\sum\nolimits_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\cdot 100\right|, \qquad (1)$$

where n is the number of observations in the test set;

y_i is the actual value of the parameter in the test set;

$\hat{y}_i$ is the predictive value of the parameter.

In testing the suggested method, use was made of the MAPE metric since it allows us to estimate the deviation of the forecast data from the actual ones in relative terms, i.e. it shows the deviation of the forecast from the fact in percentage.

1. Materials and methods

In this study, a forecast of a financial time series was made using singular spectrum analysis (SSA), i.e. the "Caterpillar" method. As is the case with the Fourier transform, SSA decomposes the original time series into a sum of components. However, in the Fourier analysis, the time series (the summands) are periodic functions of different frequency and amplitude, while SSA decomposes the original series into trending, periodic and noise elements [8, 9]. Thus, SSA takes into account such peculiarities of the financial time series as non-stationarity and non-differentiability.

The caterpillar method is a variety of the SSA analysis independently developed in the USSR at the end of the 80s. At present, the study by Golyandina "Caterpillar Method – SSA Analysis of Time Series" [10] gives the most comprehensive description of the method. It is worth noting that foreign authors also refer to the study by Golyandina as the original source [11, 12]. Other studies devoted to SSA analysis also belong to the Russian authors Leontyeva [13] and Danilov [14]. The authors of research works and guidebooks specify two problems in implementing the method: the choice of the caterpillar length and the choice of the main components. The second problem can be solved by considering the contribution of each component to the total variance; however, recommendations concerning the caterpillar length are more likely to have a heuristic character [15].

The essence of the method of singular spectrum analysis is that the original time series is transformed into a matrix, and then, the matrix is divided into components by singular value decomposition (the main components method is used here). The next step depends on the aims of the analysis: either the components are decomposed into the trending, periodic and noise elements and they are used in the analysis and in forecasting (with the noise elements being removed), or the main components are chosen and they are used to continue the series to the length of the step which was set upon the formation of the initial matrix. In this study, the second approach is used. The SSA algorithm is the following.

The time series being analyzed has the length n. The caterpillar length L is chosen, L, $2 \leqslant L \leqslant n/2$, and the matrix X is constructed, as obtained by the shift of each following column by one value of the dimension $L \times (n - L + 1)$:

$$\begin{matrix} x_1 & x_2 & \cdots & \cdots & x_{n-L+1} \\ x_2 & x_3 & \cdots & \cdots & x_{n-L+2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_L & x_{L+1} & \cdots & \cdots & x_n \end{matrix} \qquad (2)$$

Then, the singular value decomposition of the matrix X was performed:

$$X = U \sigma V^T, \tag{3}$$

where U is the left singular vector;

σ is the diagonal matrix of the values;

V^T is the right singular vector.

The obtained elements of the vectors are sorted in descending order of the eigenvalues. The initial matrix X is decomposed into L elementary matrices:

$$X_j = \sum_{j=1}^{d} \sigma_j U_j V_j^T, \ j = \overline{1L}, \tag{4}$$

where σ_j is the j-th eigenvalue;

$U_j V_j^T$ are the left and right singular vectors corresponding to the eigenvalue;

d is the rank of the matrix X.

The matrices X_j have the dimension $L \times (n - L + 1)$. In order to transform the given matrices to the one-dimensional form, diagonal averaging is used. Each row of the matrix is shifted by the value $i = 1 \ldots n - L + 1$, and then, the mean values are calculated in the column which represents the calculated values of the j-th component of the original time series. The procedure of diagonal averaging is shown by formula (5):

$$\begin{array}{llllllllllll} x_{11} & x_{12} & \ldots & \ldots & x_{1(n-L+1)} & & & & & & & \\ & x_{21} & x_{22} & \ldots & \ldots & x_{2(n-L+1)} & & & & & & \\ & & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \\ & & & & & & x_{L1} & x_{L2} & \ldots & \ldots & x_{L(n-L+1)} & \\ \bar{x}_1 & \bar{x}_2 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \bar{x}_n \end{array} \tag{5}$$

Thus, the original time series is decomposed to the sum of L series. Further, the obtained series are analyzed for trending, periodicity or noise, or the main components are selected, namely the components of the original time series which most significantly influence its dynamics. It is possible to analyze the character of the components, for example, with the help of heat maps; however, the goal of the present study is forecasting rather than analysis, and thus, we used only the method of the main components, as is indicated above. The contribution of each obtained component to the total variance is calculated by the formula:

$$\frac{\sigma_j^2}{\sum_{j=1}^{d} \sigma_j^2}, \tag{6}$$

where σ_j^2 is the square of the j-th eigenvalue.

Figure 1 shows the variances of 10 components. As is clear from the figure, the first eigenvalue makes the most significant contribution to the total variance. Let us recall that the eigenvalues and their corresponding left and right singular vectors are sorted in descending order of the eigenvalues. Therefore, as is indicated in *Fig. 1*, only the series X_1 can be chosen for the forecast.

The final step of the analysis is the actual forecast using the selected main components. Note that if there are more than one component which have a significant impact on the variance, they are to be summed up. In the present study, use was made of the method of recurrent forecasting. For this purpose, the last $2L + 1$ elements of the obtained time series were used and the following system of linear equations was constructed:

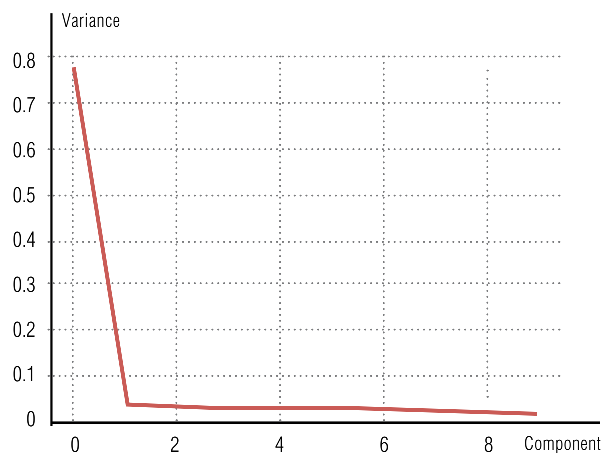


Fig. 1. Contribution of the component variances.

$$\begin{aligned} &a_1X_{n-2L+1} + a_2X_{n-2L+2} + ... + a_LX_{n-L-1} = X_{n-L} \\ &a_1X_{n-2L+2} + a_2X_{n-2L+3} + ... + a_LX_{n-L} = X_{n-L+1} \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &a_1X_{n-L+1} + a_2X_{n-L+2} + ... + a_LX_{n-1} = X_n. \end{aligned} \tag{7}$$

Solving this system in terms of the coefficients a, one obtains a matrix of coefficients which is then substituted into the matrix constructed in a manner analogous to formula (7) from the final L time series values and thus, obtains predictive L values:

$$\begin{aligned} &a_1X_{n-L+1} + a_2X_{n-L+2} + ... + a_LX_n = X_{n+1} \\ &a_1X_{n-L+2} + a_2X_{n-L+3} + ... + a_LX_{n+1} = X_{n+2} \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &a_1X_{L+1} + a_2X_{L+2} + ... + a_LX_{n+L-1} = X_{n+L}. \end{aligned} \tag{8}$$

Also, for comparison of the results obtained by SSA, forecasting by other methods was performed in this study, namely using ARIMA (autoregressive moving average model), the Fourier transform and the recurrent neural network. The ARIMA model was chosen, being the most popular for forecasting financial time series. The Fourier transform was chosen since it applies a principle somewhat similar to SSA: the original time series is also decomposed into a sum of several time series (only in the case of Fourier analysis all these series are periodic). As concerns the recurrent neural network, this method was chosen due its perspectives and due to the fast development of machine learning methods, in particular, deep learning.

The ARIMA model has three parameters: the autoregression order p, the order of differencing d and the moving average order q. The order of differencing is determined by the Dickey-Fuller test. The autoregression order is determined by the autocorrelation plot of the series levels, in which the time lags are shown on the X-axis, while the values of the correlation coefficient between the levels corresponding to the lag are given on the Y-axis. The autoregression order is chosen to be equal to such a time lag at which the correlation coefficient has the last maximum value which is different from zero. The moving average order is chosen in the same way; though, instead of the autocorrelation coefficients, partial autocorrelation coefficients are calculated. Partial autocorrelation differs from autocorrelation by not taking into account the impact of the levels which are located between the current level and the level which has a single time lag. It is obvious that in the case of a single lag, autocorrelation and partial autocorrelation coincide.

To do ARIMA forecasting in Python, it is sufficient to determine the parameters p, d, q of the model. Formally, the ARIMA model is described as follows:

$$\Delta^d y_t = \sum\nolimits_{i=1}^{p} \alpha_i \,\Delta^d y_{t-i} - \sum\nolimits_{j=1}^{q} \beta_j \,\varepsilon_{t-j}, \tag{9}$$

where Δ^d is the difference of d-order required to achieve stationarity;

α_i are the p-order autoregression coefficients;

β_j are the q-order moving average coefficients;

ε_{t-j} are the moving average forecasting errors.

The coefficients α_i and β are estimated and substituted into the forecast.

The Fourier transform decomposes the periodic function into a sum of sinusoids and cosine curves with known frequencies, amplitudes and phases. In general terms, the Fourier transform is represented by formula (10):

$$y_t = \overline{y_t} + \sum\nolimits_{i=1}^{N/2} (\alpha_i \cos \omega_i t + b_i \sin \omega_i t), \tag{10}$$

where y_t is the transformed value of the time series;

$\overline{y_t}$ is the average value of the original time series;

ω_i is the frequency of the i-th harmonic (the first frequency corresponds to the period of the function, with the other ones being multiples of it);

α_i, β_i are the coefficients to be estimated.

The coefficients of the Fourier series are calculated using the following formulas:

$$a_1 = \frac{2}{N} \sum\nolimits_{t=0}^{2\pi(N-1)/N} y_t \cos t, \;\; b_1 = \frac{2}{N} \sum\nolimits_{t=0}^{2\pi(N-1)/N} y_t \sin t \tag{11}$$

for the first harmonic,

$$a_2 = \frac{2}{N}\sum_{t=0}^{2\pi(N-1)/N} y_t \cos 2t, b_2 = \frac{2}{N}\sum_{t=0}^{2\pi(N-1)/N} y_t \sin 2t \quad (12)$$

for the second harmonic and so on. Usually, the first two or three harmonics are sufficient for forecasting. The Fourier coefficients obtained by formulas (11) and (12) are substituted into the equation (10), and forecasting is performed using this equation. In this work, for forecasting by the Fourier transform method, use was made of the code in the Python programming language developed by the author.

The time series forecasting algorithm using a **recurrent neural network** can be represented as follows.

1. The original time series of the length n is transformed into a matrix in which the rows correspond to the step L for building a forecast. An array of dimension $(L \times (n - L + 1))$ is fed to the input of the recurrent layer. Also, m is set, which is a step describing how many periods ahead the forecast is made.
2. Inside the recurrent layer, the given array is processed by the activation function, and the output is an array of the dimension specified by the user. If this layer is the last one, then the dimension of the output array is $(m \times 1)$.
3. The final solution is compared with the actual data. The loss function is set (the difference between the actual data and those predicted by the neural network), and using the optimization function, the error backpropagation algorithm is implemented. The error backpropagation algorithm changes the weights given randomly at the stage of the forward operation of the neural network in such a way as to minimize the loss function. Large data arrays for "fitting" the weights are divided into packages (batches), i.e., the optimizer changes the weights after the package is sent rather than after each signal sent. The package size is set upon constructing the neural network and it is usually a power of two. The number of epochs determines the number of "runs" of the neural network for successful learning. The training quality is determined by the loss function on the training data and by the error metric on the test data.

2. Results and discussion

For the practical application of SSA, an algorithm was developed in Python using the numpy, pandas, matplotlib, sclearn libraries. Data on foreign financial time series were taken from the Yahoo! Finance website using the Python yfinance library, and data on domestic stock quotes were obtained using the Python apimoex library.

For the analysis, daily stock quotes of the top Russian and American companies were chosen for the time period from June 2022 to March 2023. A total of 30 companies were analyzed: 15 of them Russian and 15 American. Data for 110 days were divided into training and test sets. The training set included 100 values, and the test set had 10 values. Since the forecast was made for the period L – the caterpillar length; then, accordingly, the parameter L was chosen equal to 10. Thus, the original matrix X had the dimension 10×91, and as a result of the singular value decomposition, it was decomposed into 10 matrices of the same dimension. These matrices were transformed into one-dimensional arrays of dimension 100, and those were selected from them whose matrix eigenvalues make the greatest contribution to the variance. Then, the obtained predictive values were compared with the test set using the MAPE metric. *Table 1* presents the MAPE values of the SSA-forecast for the analyzed stocks.

The average forecast error was 5.54%, including 4.62% for domestic stocks and 6.46% for US stocks. Note that there is a strong outlier: for the company JPMorgan Chase & Co., the error turned out to be significant (26%). For the Russian companies, there is a rather large error only for RUSAL, 11% (which is quite an acceptable result for the forecast). Regarding the stocks of US companies, there is also a significant error for Mastercard (17%). The errors are also quite large for United Health (9.5%) and Advanced Micro Devices (11%). The best forecasts were obtained for Norilsk Nickel (the error being 0.5%), MTS (with the error of 1.7%) and McDonald's (1.2%). *Fig. 2* shows several plots illustrating the SSA method, with both "bad" and "good" forecasts visualized.

Table 1.

MAPE of the SSA method for stocks

Russia	MAPE	USA	MAPE
Polymetal	7.00%	Amazon	5.00%
Polyus Gold	5.30%	Apple	4.70%
Mechel	7.80%	American Express Company	6.20%
MMC Norilsk Nickel	0.50%	Tesla	2.60%
Yandex	5.20%	Advanced Micro Devices	11.00%
Aeroflot - Russian Airlines	2.90%	Pfizer	2.10%
VTB Bank	4.60%	Netflix	2.12%
Magnit	2.90%	Microsoft	2.00%
Alrosa	5.20%	Mastercard	17.00%
Tinkoff Group	6.10%	Visa	2.10%
RUSAL	11.00%	Starbucks	2.30%
Novatek	2.40%	JPMorgan Chase & Co.	26.00%
Surgutneftegas	3.20%	McDonald's	1.20%
MTS	1.70%	Boeing	3.10%
Severstal	3.50%	United Health	9.50%

To compare the SSA analysis with ARIMA, the Fourier transform and recurrent neural network, a forecast was made for the same stocks over the same time period. The *p* and *d* parameters of the ARIMA model for all the stocks were 2 and 1, respectively, while the *q* parameter varied, depending on the autocorrelation plot. The Fourier transform was performed in three harmonics. The recurrent neural network was constructed using the Keras and Tensor Flow Python libraries. Since the analyzed interval was very small for the neural network, one RNN layer was chosen to avoid overtraining. The activation function, loss function and optimizer function were chosen based on the parameters recommended in [6] for predictive recurrent neural networks. The minimum package size was chosen, i.e. 2, and the number of epochs was chosen such that the loss function and the error (mean absolute error and mean absolute error in percent, respectively) stopped changing. Most experiments used 20 epochs. The forecast was made for 10 days ahead similar to the SSA method.

As is seen from the *Table 2*, the ARIMA method significantly outperformed other methods, including SSA, in terms of the forecast accuracy. However, the SSA analysis showed more accurate results than the Fourier transform and recurrent neural network, and regarding the shares of Norilsk Nickel, MTS and McDonald's it is rather comparable to ARIMA. In general, the Fourier transform also showed good results, with error outliers being present only for the shares of Mechel, Advanced Micro Devices and Pfizer. As concerns the recurrent neural network, it does not work for the stocks for a short period, which was confirmed by the unacceptable size of errors.

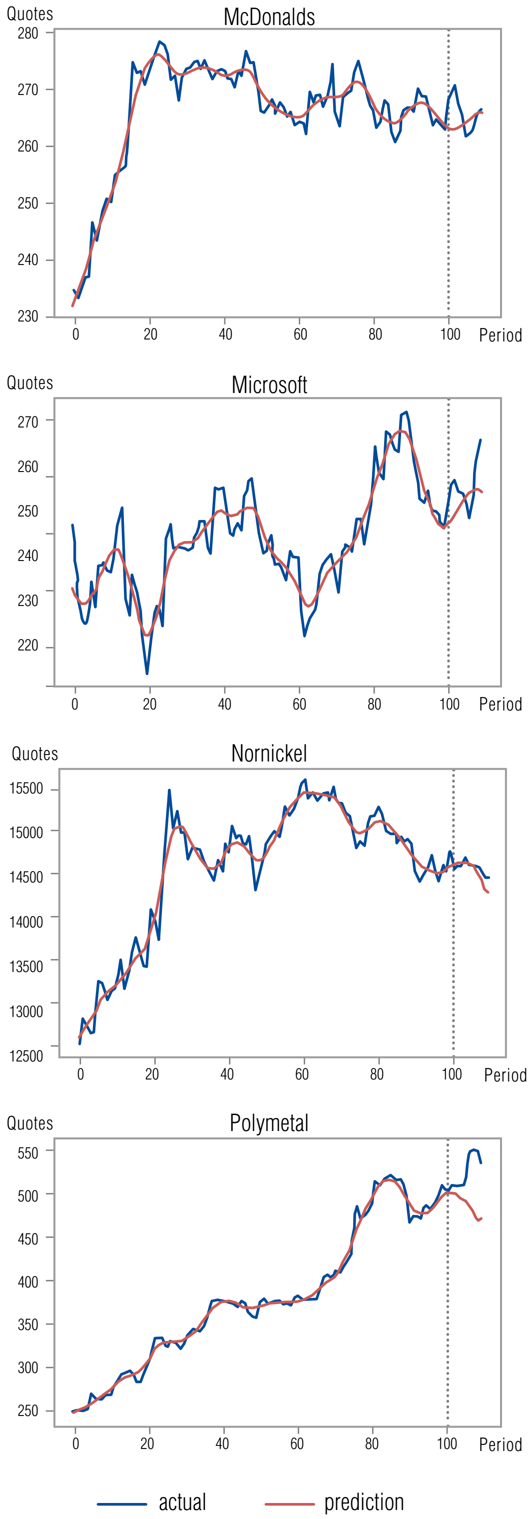


Fig. 2. Visualization of the SSA forecast.

Next, singular spectrum analysis as compared with other methods was applied to the foreign exchange market and cryptocurrency market, which are more dynamic than the stock market. The time period was the same – from June 2022 to March 2023. Ten currency pairs and ten cryptocurrency quotes against the US dollar were taken. The average absolute error in percent for the compared methods is given in *Table 3*.

The SSA method did not show very good results both for fiat and crypto currencies. While for the cryptocurrencies, there were outliers only for BNB and Tron currencies; for the fiat currency pairs, the method showed a large error in four cases out of ten. The ARIMA method worked just as well on the currencies as it did on the stocks, and the Fourier transform and recurrent neural network significantly improved the forecast accuracy. But, the recurrent neural network forecast still remained unsatisfactory. Note that all the considered methods had worse results on the cryptocurrencies than on the fiat ones. This can be due to the fact that there are still loopholes for arbitrageurs since the crypto market is rather young. *Figure 3* shows the forecast plots for some fiat and cryptocurrencies, as obtained by the considered methods.

To solve the problem of the poor SSA forecast for both fiat and cryptocurrencies, we doubled the time interval. New training and test sets for the same currencies were 200 and 20 (the length of the caterpillar, accordingly, was also chosen to be 20). The average absolute error in percent under the new conditions is shown in *Table 4*.

With the increase in the time interval, the results of the SSA forecast improved significantly, while the accuracy of the ARIMA forecast remained almost unchanged, and that of the Fourier forecast was even worse. The recurrent neural network significantly improved the forecast accuracy, though significant outliers remained for two fiat currencies and three cryptocurrencies. The best accuracy was still obtained with the ARIMA model. *Figure 4* shows the forecasts for the currencies over a long time period.

Table 2.

MAPE of the methods ARIMA, Fourier transform, RNN

Russia	MAPE, ARIMA	MAPE, Fourier	MAPE, RNN
Polymetal	1.70%	27.60%	95.77%
Polyus Gold	1.50%	17.00%	99.79%
Mechel	1.80%	22.20%	19.50%
MMC Norilsk Nickel	0.50%	1.50%	54.83%
Yandex	1.20%	3.20%	95.51%
Aeroflot	0.70%	8.80%	5.86%
VTB	0.80%	1.50%	3.60%
Magnit	0.40%	3.30%	99.05%
Alrosa	0.70%	2.30%	29.19%
Tinkoff Group	1.00%	3.90%	98.34%
RUSAL	0.70%	5.00%	2.65%
Novatek	0.80%	2.60%	95.85%
Surgutneftegas	0.80%	3.50%	2.19%
MTS	1.00%	8.40%	81.83%
Severstal	1.10%	18.40%	95.12%
Average for Russian stocks	**0.98%**	**8.61%**	**58.61%**
Amazon	1.50%	3.10%	56.39%
Apple	1.60%	6.40%	69.55%
American Express Company	2.30%	8.50%	70.46%
Tesla	3.10%	5.90%	76.90%
Advanced Micro Devices	2.00%	17.00%	34.24%
Pfizer	1.00%	17.10%	8.81%
Netflix	1.70%	3.00%	83.86%
Microsoft	1.30%	5.90%	81.80%
Mastercard	1.40%	3.80%	86.55%
Visa	1.30%	4.80%	78.09%
Starbucks	1.10%	3.50%	53.21%
JPMorgan Chase & Co.	2.70%	4.10%	64.73%
McDonald’s	0.80%	2.00%	83.24%
Boeing	2.10%	10.80%	72.31%
United Health	0.70%	9.40%	91.70%
Average for US stocks	**1.64%**	**7.18%**	**67.46%**
Total average	**1.31%**	**7.90%**	**63.03%**

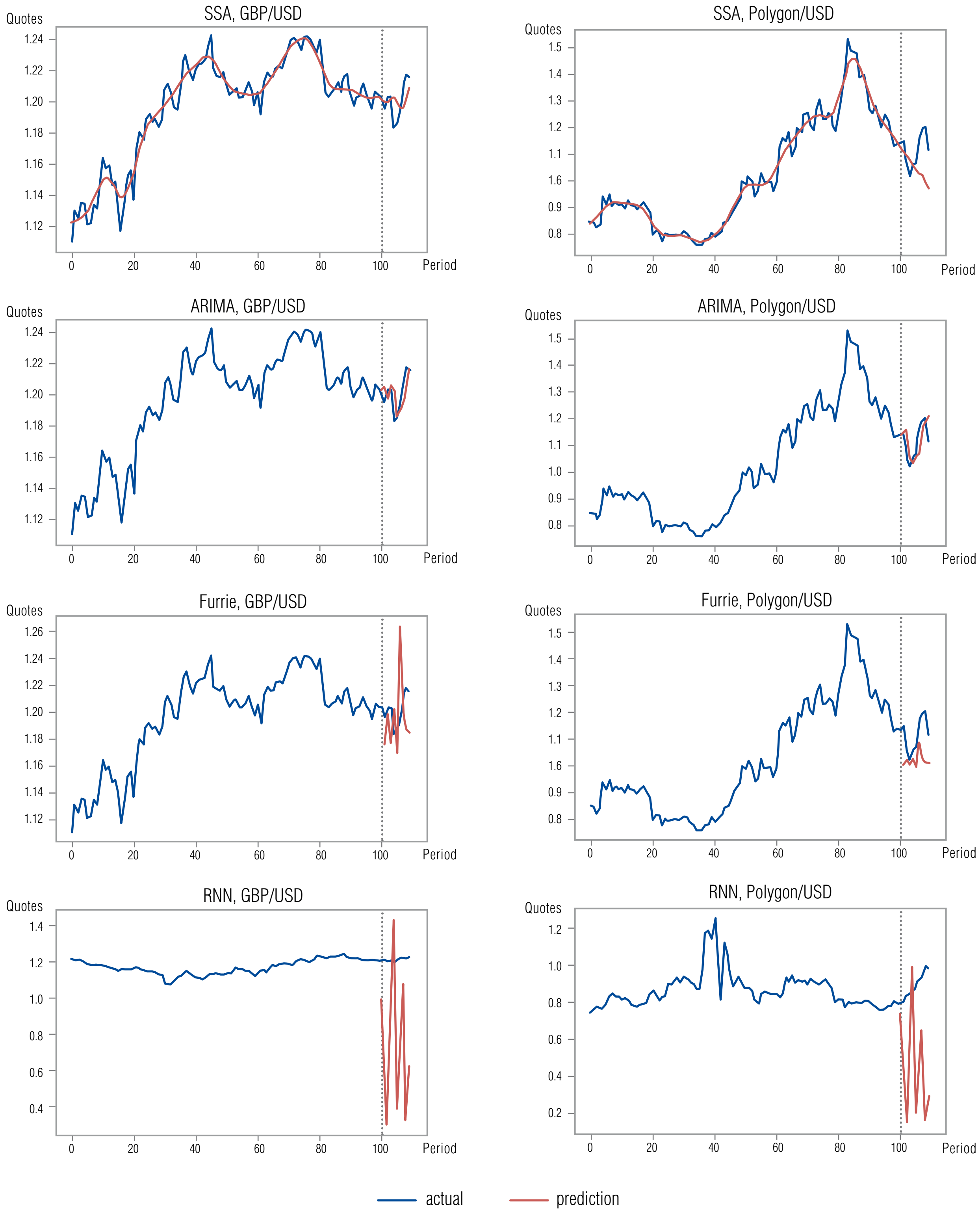


Fig. 3. Visualization of the currency forecasts by SSA, ARIMA, Fourier transform and RNN.

Table 3.

MAPE of SSA, ARIMA, Fourier transform and RNN for currencies

Currency	MAPE, SSA	MAPE, ARIMA	MAPE, Fourier	MAPE, RNN
Euro / US Dollar	59.00%	0.50%	2.70%	1.89%
Pound / US Dollar	0.70%	0.60%	2.10%	2.22%
US Dollar / Yuan	3.70%	0.50%	1.70%	1.47%
US Dollar / Rouble	37.90%	0.30%	12.20%	30.46%
US Dollar / Yen	40.20%	0.60%	1.80%	69.06%
US Dollar / Hong Kong Dollar	43.60%	0.00%	1.80%	0.33
US Dollar / South African Rand	8.30%	0.80%	5.00%	2.27%
Australian Dollar / US Dollar	0.90%	0.60%	1.90%	0.33%
US Dollar / Mexican Peso	1.70%	1.10%	5.20%	1.36%
New Zealand Dollar / US Dollar	2.00%	0.70%	1.90%	2.42%
Average for fiat currencies	**16.38%**	**0.57%**	**3.63%**	**15.90%**
Bitcoin	6.80%	3.10%	9.80%	99.90%
Ethereum	8.20%	3.10%	8.30%	98.75%
Binance Coin	47.3%	1.80%	4.00%	6.57%
Polygon	8.10%	4.00%	8.00%	6.25%
Lightcoin	12.30%	5.70%	8.40%	74.24%
Ripple	1.90%	1.90%	2.50%	6.29%
Polkadot	6.40%	3.50%	4.10%	5.70%
Chainlink	7.40%	3.30%	4.60%	9.44%
Avalanche	17.40%	4.70%	4.8%	11.13%
Tron	45.50%	4.00%	6.70%	4.52%
Average for cryptocurrencies	**16.13%**	**3.51%**	**6.08%**	**34.82%**
Total average	**16.26%**	**2.04%**	**4.85%**	**25.36%**

Conclusion

In this paper, a method of forecasting time series, namely the SSA analysis (caterpillar method) was considered. The method was implemented by developing an algorithm in the Python language, and then, was tested on 30 time series of Russian and US stock quotes, as well as on 20 fiat and cryptocurrencies against the US dollar. For comparison, three forecasting methods were taken: ARIMA, the Fourier transform and recurrent neural network. In all the cases, the SSA analysis, except for the currencies with a short time period, showed the second most accurate result after the ARIMA method. For some securities and shares, the SSA error is comparable to that of ARIMA. At the same time, an increase in the time interval significantly improved the SAA results, while the ARIMA results remained unchanged.

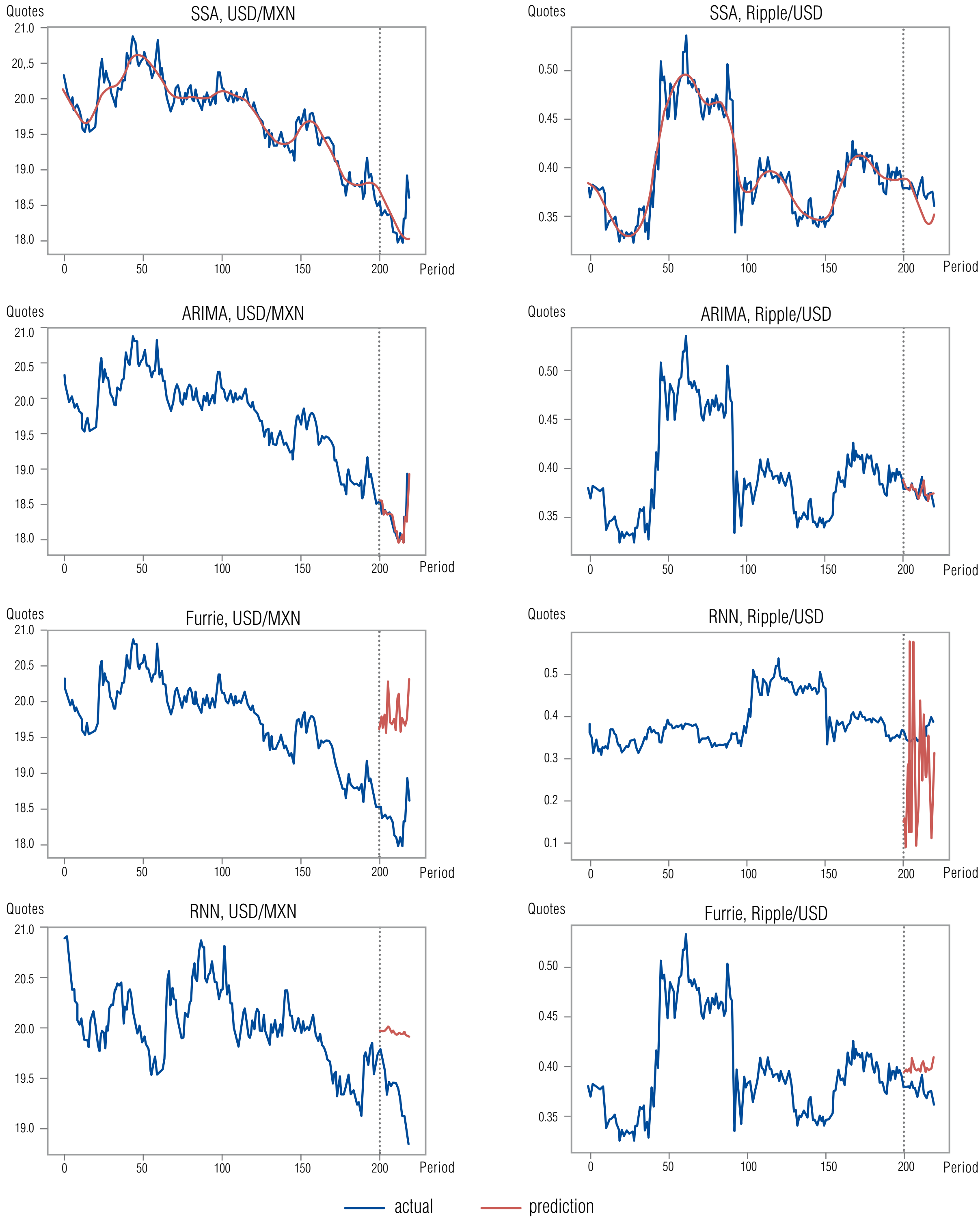


Fig. 4. Visualization of the forecasts for the currencies obtained by SSA, ARIMA, Fourier transform and RNN (long time period).

Table 4.

MAPE of the SSA method and recurrent neural network for currencies (long period)

Currency	MAPE, SSA	MAPE, ARIMA	MAPE, Fourier	MAPE, RNN
Euro / US dollar	2.60%	0.40%	3.00%	2.44%
Pound / US dollar	8.20%	0.50%	1.20%	2.62%
US dollar / Yuan	0.90%	0.40%	1.10%	2.40%
US dollar / Rouble	1.50%	1.10%	16.30%	50.53%
US dollar / Yen	6.50%	0.50%	1.50%	78.31%
US dollar / Hong Kong dollar	1.00%	0.00%	0.90%	0.22%
US Dollar / South African Rand	1.80%	0.60%	6.60%	2.94%
Australian Dollar / US Dollar	3.10%	0.50%	1.70%	0.38%
US Dollar / Mexican Peso	1.10%	0.70%	8.1%	1.53%
New Zealand Dollar / US Dollar	3.20%	0.50%	1.10%	2.99%
Average for fiat currencies	**3.16%**	**0.52%**	**4.48%**	**14.44%**
Bitcoin	9.90%	2.30%	13.30%	99.83%
Ethereum	48.80%	2.40%	9.90%	97.76%
Binance Coin	3.20%	1.50%	3.80%	7.22%
Polygon	16.50%	3.70%	20.50%	7.23%
Lightcoin	12.40%	3.90%	19.40%	50.07%
Ripple	4.40%	1.50%	6.10%	8.88%
Polkadot	19.70%	3.20%	5.80%	6.32%
Chainlink	10.80%	2.90%	5.20%	8.55%
Avalanche	7.40%	4.00%	6.2%	18.19%
Tron	7.00%	2.70%	9.80%	4.73%
Average for cryptocurrencies	**14.01%**	**2.81%**	**10.02%**	**30.88%**
Total average	**8.58%**	**1.67%**	**7.25%**	**22.66%**

It can be concluded that although the SSA analysis shows a lower forecast accuracy than ARIMA generally accepted in the analysis of financial time series, it can be applied both to stocks and to other financial instruments, even to cryptocurrencies which are volatile. It can be used to confirm the ARIMA results, as well as separately as a forecasting method. Note that for the analysis of a large set of stocks, the SSA method is more convenient than ARIMA, since ARIMA requires recalculation at least of the order of the moving average for each time series, while a single main component is sufficient for the SSA forecast. In addition, when considering other components of the singular value decomposition, we can draw conclusions about the ratio of trending, periodicity and noise in the analyzed time series, which cannot be done using any other considered methods. ■

References

1. Mandelbroit B., Fisher A., Calvet L. (1997) A multifractal model of asset returns. *Yale Cowles Foundation for Research in Economics*, Discussion Paper No. 1164.
2. Mandelbrot B. (2004) *Fractals, case and finance*. Moscow, Izhevsk: Research Center “Regular and Chaotic Dynamics” (in Russian).
3. Peters E. (2004) *Fractal analysis of financial markets. Application of chaos in investment and economics*. Moscow: Internet Trading (in Russian).
4. Peters E. (2000) *Chaos and order in capital markets. A new analytical perspective on cycles, prices and market volatility*. Moscow: Mir (in Russian).
5. Zinenko A.V. (2012) R/S analysis in the stock market. *Business Informatics,* no. 3(21), pp. 21–27 (in Russian).
6. Cholet F. (2018) *Deep learning with Python*. Saint Petersburg: Peter (in Russian).
7. Kong Q., Han J., Jin X., Li C., Wang T., Bai Q., Chen Y. (2023) Polar motion prediction using the combination of SSA and ARMA. *Geodesy and Geodynamics*, vol. 14, no. 4, pp. 368–376. https://doi.org/10.1016/j.geog.2022.12.004
8. Li K., Zhang Z., Guo H., Li W., Yan Y. (2023) Prediction method of pipe joint opening-closing deformation of immersed tunnel based on singular spectrum analysis and SSA-SVR. *Applied Ocean Research*, vol. 135, 103526. https://doi.org/10.1016/j.apor.2023.103526
9. Montalvo C., Pantera L., Lipcsei S., Torres L.A. (2022) Signal processing applied in cortex project: From noise analysis to OMA and SSA methods. *Annals of Nuclear Energy*, vol. 175, 109193. https://doi.org/10.1016/j.anucene.2022.109193
10. Golyandina N.E. (2004) *Method “Caterpillar”-SSA: analysis of time series*. Saint Petersburg: St. Petersburg State University (in Russian).
11. Coussin M. (2022) Singular spectrum analysis for real-time financial cycles measurement. *Journal of International Money and Finance*, vol. 120, 102532. https://doi.org/10.1016/j.jimonfin.2021.102532
12. Lahmiri S., Bekiros S., Bezzina F. (2022) Evidence of the fractal market hypothesis in European industry sectors with the use of bootstrapped wavelet leaders singularity spectrum analysis. *Chaos, Solitons & Fractals*, vol. 165, part 1, 112813. https://doi.org/10.1016/j.chaos.2022.112813
13. Leontyeva L.N. (2011) Multidimensional caterpillar, choice of length and number of components. *Machine Learning and Data Analysis*, no. 1, pp. 5–15 (in Russian).
14. Solntsev V.N., Danilov D.L., Zhiglyavsky A.A. (1997) *Principal components of time series: Method “Caterpillar”*. Saint Petersburg: St. Petersburg State University (in Russian).
15. Baharanchi S.A., Vali M., Modares M. (2022) Noise reduction of lung sounds based on singular spectrum analysis combined with discrete cosine transform. *Applied Acoustics*, vol. 199, 109005. https://doi.org/10.1016/j.apacoust.2022.109005

About the author

Anna V. Zinenko

Cand. Sci. (Tech.), Associate Professor;

Associate Professor, Department of Economic and Financial Security, Siberian Federal University, 79, Svobodny Prospect, Krasnoyarsk 660041, Russia;

E-mail: anna-z@mail.ru