

Predictive Models for Agricultural Management

Natalia N. Bakhtadze, Evgeny M. Maximov, Natalia E. Maximova, and Lamara N. Kozlovskaya

Abstract—This paper is devoted to the development of predictive models for decision support systems applied in precision farming. Application of predictive models makes it possible to use resources effectively, which reduces the cost of production and increases the efficiency of agricultural production. In addition, the forecast makes it possible to reach a long-term agronomic and ecological effect due to more careful tillage and reduced use of fertilizers. The algorithms using knowledge base for creating models of grain yield are described and the results of applying these models are presented.

Index Terms—Precision farming, soft sensors, predictive models, knowledge, associative search algorithms.

I. INTRODUCTION

Agro-management today actively uses the capabilities of information technologies - both for the implementation of more effective management of a certain technological process, and for organizing the most profitable farming in general for specific agricultural enterprises, taking into account the specifics of their activities and the current situation.

Digital farming is a high-tech approach to managing the state of fields and the efficiency of their use based on the study of the dynamics of their physical and agrochemical properties using modern mathematical and information technologies.

Management within the concept of digital farming is based on the principle that a field that is heterogeneous in topography, soil cover or agrochemical content is subjected to heterogeneous cultivation. The identification of inhomogeneities is carried out on the basis of an analysis of the operation of global positioning systems, aerial photographs and satellite images, geographic information systems, statistical analysis and expert knowledge.

Based on the analysis of data characterizing features of the site, taking into account the peculiarities of soil types and climatic conditions, are carried out: planning of sowing, the calculation of the amount of fertilizer application, crop yield forecasting and financial planning.

This approach allows more rational use of fertilizers and fuel, which reduces the cost of production and increases the efficiency of agricultural production. In addition, a long-term agronomic and ecological effect can be achieved - due to more gentle soil cultivation and a decrease in the intensity of the use of nitrogen fertilizers. Among the modern methods

and means of agricultural management, one can single out: global positioning technologies (GPS, GLONASS – Russian Global Navigation Satellite System), geographic information systems (GIS), technologies for current yield assessment (Yield Monitor Technologies), Variable Rate Technology, earth remote sensing (ERS), Internet of Things, etc.

In this article, we propose an approach to support effective (in technological and economic aspects) management, based on the use of predictive identification models.

The control of technological processes, in particular, the mode of application of mineral fertilizers, in order to obtain the highest yield of grain crops (one of the typical tasks of digital farming) is investigated.

Currently, in developed countries, integrated systems for the differential application of mineral fertilizers are used quite widely [1], [2]. In Russia, such systems are also created. However, these systems do not provide enough information about the effect of the field pieces properties dynamics and external factors on the productivity, which defines the amount of fertilizers to apply [3].

The approach described in this paper is considerably less costly and time consuming in computation than the approach based on using data from multiple sensors with its subsequent processing using specialized software.

In the paper, so-called *Soft Sensors* are proposed [4] built for some indicators for the field pieces by means of predictive models, which are based on intelligent analysis of current and historical data available for measurement.

Intelligence is interpreted here in the aspect of using inductive knowledge about a dynamic process, i.e., certain statistical patterns extracted from the entire array of historical data using Data Mining algorithms.

Knowledge refers to the laws that are extracted from data analysis and refined as information accumulates [5].

To develop the model, a representation is formed of historical and current measurement data, virtual sensors data and knowledge. The algorithm builds a new model for each point in time, and the parameter estimates are the best in terms of the minimum of the root-mean-square error.

The associative search algorithm is based on data mining. To accelerate the associative search, clustering methods are used for processing the data from the fields with similar characteristics.

At the time studied, a set of inputs (in the general case, multidimensional) close to the current input vector in the sense of a certain criterion is selected from the data archive. This criterion is called the associative impulse and can be both a functional and a logical or fuzzy function.

Further, on the basis of the classical least square method (OLS) the model is built and the output value at the next time point is determined. Under the assumption that input actions satisfy Gaussian-Markov conditions, the estimates obtained by the least square method are consistent, unbiased and

Manuscript received January 12, 2021; revised March 29, 2021.

Natalia N. Bakhtadze, Evgeny M. Maximov, and Natalia E. Maximova are with Identification laboratory, V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia (e-mail: sung7@yandex.ru, maxfone@yandex.ru, natalya_maximova@mail.ru).

L. N. Kozlovskaya is with Moscow Timiryazev Agricultural Academy, Moscow, Russia (e-mail: rgau-botanika@timacad.ru).

statistically effective.

Such search method named the associative search [6] makes it possible to build local linear models for different types of non-linear objects. The algorithms have demonstrated high efficiency for non-linear non-stationary objects in industry and power engineering. The use of such algorithms in agromanagement allows making effective management solutions in precise (coordinate) farming systems.

II. FACTORS USED FOR ANALYSIS AND MODEL BUILDING

The yield forecast for a specific field is based on the analysis of the multiple indicators dynamics for the period of several previous years. Natural features, crop rotation, farming technologies and amounts of the applied fertilizers were analyzed [7], [8].

The following factors were selected for building the model forecasting yield and its dependence on using of fertilizers:

- Soil-climatic zone,
- Climate characteristic for the last 8-10 years: monthly precipitation, average monthly air temperature,
- History of crop rotation for the last 8-10 years,
- Crop yield,
- Amount of the mineral fertilizers applied for the last 8-10 years, including the application frequency and doses of fertilizers.

Further, on the basis of data from the Inductive Knowledge Base [9], by means of statistical analysis and application of associative search algorithms, the analysis of the effect of different factors on the yield of main crops was conducted. The purpose of the analysis was the evaluation of the fertilizers application mode effect on the yield.

III. ASSOCIATIVE SEARCH ALGORITHM

Usually, the processes describing crop farming are non-linear. To analyze these processes on the base of current and historical data, an identification algorithm based on virtual models are being built for every specific point in time is proposed.

The identification algorithm consists of building an approximating hypersurface of the space of input vectors and corresponding outputs for every specific point in time. For building the virtual model for a specific time point, the vectors that are close to the current input one in the sense of a certain criterion are selected from the archive.

The criterion for selecting the vectors is described below. Further, on the base of the classical least square method the coefficients of the model are determined.

At any selected point in time a new model is built (instead of approximating the real process in time). To build the model, the representation of historical and current process data is being formed.

The linear dynamic model is of the form:

$$y_N = \sum_{i=1}^m a_i y_{N-i} + \sum_{j=1}^n \sum_{s=1}^S b_{js} x_{N-j,s}, \quad \forall j \in 1, \dots, N \quad (1)$$

where:

y_N is the prediction of the output of the object at the time instant N . In this case, the yield is regarded as the output;

x_N is the input vector (the components of the vector are the factors influencing the yield),

m is the memory depth in the output,

n is the memory depth in the input,

S is the dimension of the input vectors.

For each fixed time point the vectors that are close to the current input in the sense of a certain criterion are chosen from the archive. Thus, in equation (1) n is the number of vectors from the archive (from the time point 1 to the time point N) selected in accordance to the associative search criterion. At each time segment $\{N-1, N\}$ a certain set of n vectors, $1 \leq n \leq N$, is selected.

The criterion of selecting the input vectors from the archive to develop the virtual model at the given time point over the current state of the object may be as follows.

Let us introduce as a distance (a norm in \mathbb{R}^S) between the points of the S -dimensional space of inputs the value (Manhattan distance):

$$d_{N,N-j} = \sum_{s=1}^S |x_{Ns} - x_{N-j,s}|, \quad \forall j \in 1, \dots, N, \quad (2)$$

where x_{Ns} are the components of the input vector at the current time point N .

By virtue of a property of the norm ("the triangle inequality"), we have:

$$d_{N,N-j} \leq \sum_{s=1}^S |x_{Ns}| + \sum_{s=1}^S |x_{N-j,s}|, \quad \forall j \in 1, \dots, N \quad (3)$$

Let the current input vector

$$x_N: \sum_{s=1}^S |x_{Ns}| = d_N.$$

To derive an approximating surface for x_{N-j} , $j = 1, \dots, N$, let us select from the archive of input data such vectors x_{N-j} , that for a set D_N the condition will hold:

$$d_{N,N-j} \leq d_N + \sum_{s=1}^S |x_{N-j,s}| \leq D_N, \quad \forall j = 1, \dots, n \quad (4)$$

where D_N may be selected, for example, from the condition:

$$D_N \geq 2d_N = \max_j \sum_{s=1}^S |x_{N-j,s}| \quad (5)$$

If the selected area does not provide enough inputs for the least square method implementation, i. e. the corresponding system of linear equations proves unsolvable, the chosen criterion of selecting points from the input space may be loosened by increasing the D_N .

The proposed procedure of the approximating surface developing works faster than the usual exhaustive search because at the training stage the

$$d_{N-k} = \sum_{s=1}^S |x_{(N-k),s}|, \quad k = 1, \dots, N-1, \quad (6)$$

values $N-k$ can be once defined and ranged for every time point previous to N , and when a new input enters, a new term supplements this set.

IV. THE CASE STUDIES

By means of the associative search algorithms described above, the models were formed for predicting the crop yield for two climatic zones.

The algorithms select the real data concerning fields (the procedure for selecting data from the archive is described above) that in the sense of a certain criterion are close to the field for which the model is formed.

Further, as described above, the linear model is developed at the current time point. The results of yield prediction obtained by using the models developed are given below.

A. The Results of Model Developing for the Central Black Earth Region

To develop this model, the data was used from the Central Black Earth Region fields with similar climatic conditions. The following influencing factors were revealed:

- nitrogen use (15%)
- air temperature in April
- air temperature in June
- precipitation from April to August

The model accuracy turned out to be very high. The multiple correlation coefficient value was 0.97.

The results of comparison are represented by Fig. 1.

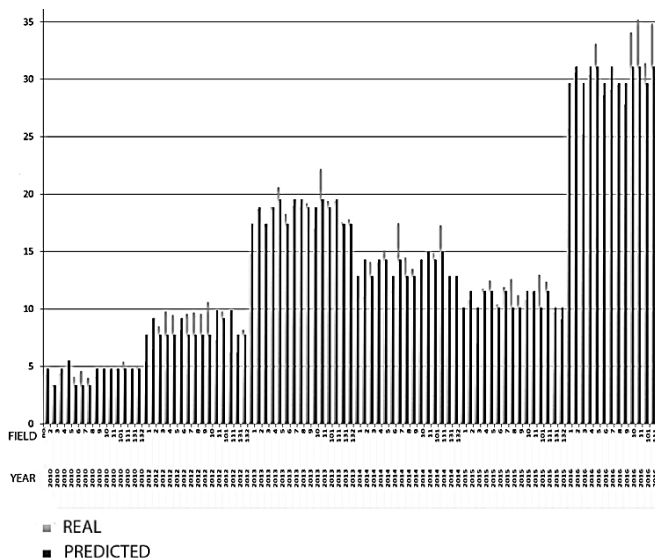


Fig. 1. Wheat yield (t/ha) from the fields of the Central Black Earth Region from 2010 to 2016 and its values predicted by the associative search model.

The yield calculated by using the model was compared with the real yield. The value of the multiple correlation coefficient obtained as a result of comparing the yield calculated by using the model and the real one was 0.95. The results obtained are also illustrated by the yield dynamics and its predicted value (Fig. 2).

To develop this model, the data was used from the Caucasus foothills fields with similar climatic conditions.

B. The Results of Model Developing for the North Caucasus

The following influencing factors were revealed:

- nitrogen use (45%)
- air temperature from April to August
- air temperature in June
- precipitation from April to August.

The model built proved to be rather accurate. The multiple correlation coefficient value was 0.86. The yield calculated by using the model was compared with the real yield.

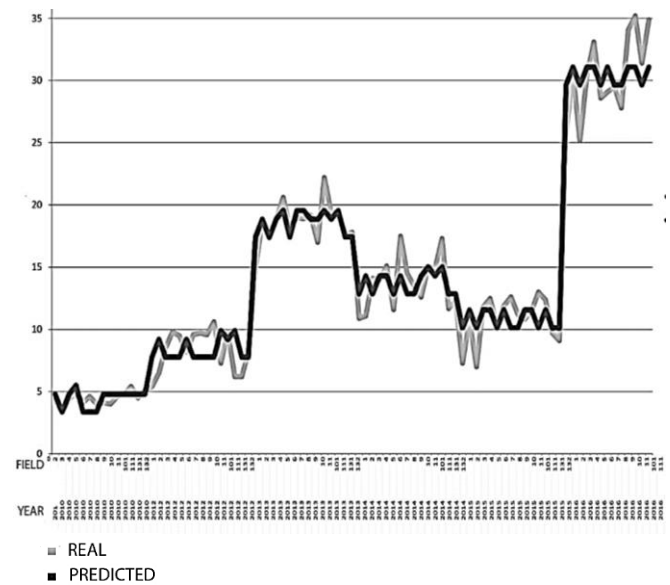


Fig. 2. The wheat yield dynamics (t/ha) from the Central Black Earth Region fields from 2010 to 2016 and its value predicted by the associative search model.

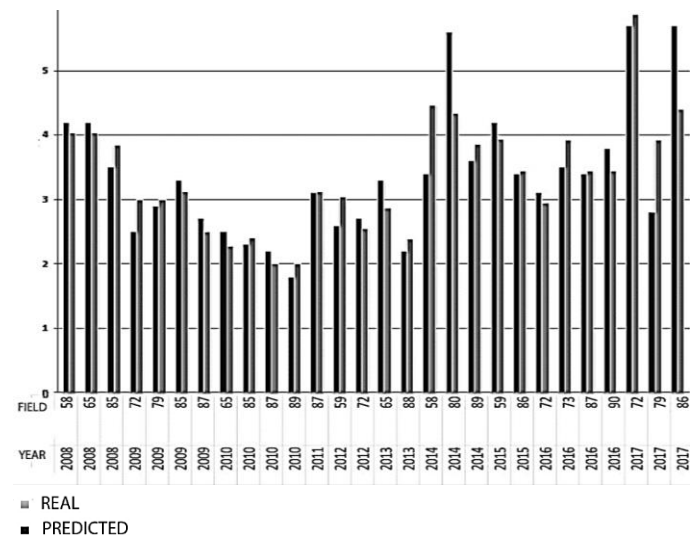


Fig. 3. Wheat yield (t/ha) from the Caucasus foothills fields from 2008 to 2017 and its values by the associative search model.

The value of the multiple correlation coefficient obtained as a result of comparing the yield calculated by using the model and the real one was 0.75.

The results of comparison are represented by Fig. 3.

The results obtained are also illustrated by the dynamics of the yield and its predicted value (Fig. 4).

The analysis of results obtained showed that on the Caucasus foothills fields different schemes of applying mineral fertilizers were used. According to the model, the influence of this factor is quite significant (about 45 %).

However, the model accuracy is lower than in other cases. Most likely, it is a result of the influence of some unaccounted parameters (local relief, insolation etc.) in the foothills conditions.

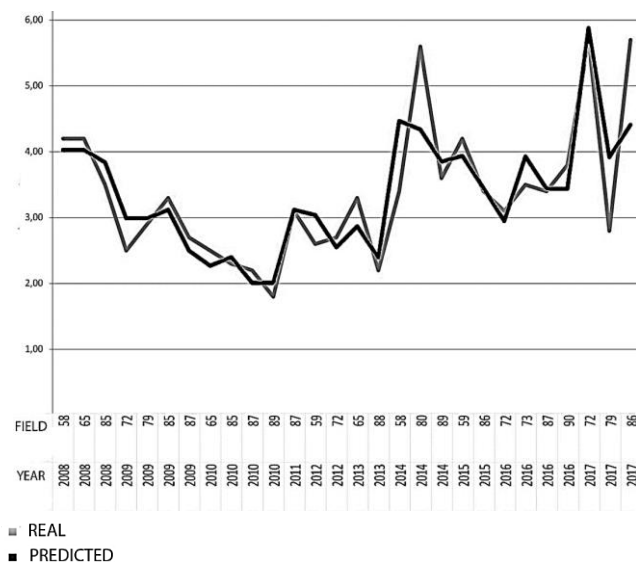


Fig. 4. The dynamics of wheat yield (t/ha) from the Caucasus foothills fields from 2008 to 2017 and its value predicted by the associative search model.

V. PERSPECTIVE RESEARCH

To maximize their profits, agricultural producers need recommendations on what technological operations to conduct and on parameters of these operations. The recommendations are based on yield prediction data.

The results of modeling presented in this paper indicate that the main dynamic factors influencing the yield are climatic conditions and amounts of applied fertilizers. Thus, agricultural producers can control the yield by applying different modes of using fertilizers.

To define the schedule and necessary amounts of fertilizers, the coefficients of crop nutrients consumption are applied. The values of these coefficients are averaged and rather wide-range.

A more precise prediction of crop nutrients consumption can be obtained with use of data on crop yields from the fields located in different soil-climatic zones, data on using fertilizers and climate data.

On the base of this prediction, agricultural producers can obtain recommendations on the schedule and amounts of fertilizers application that would allow reaching the yield they need from a specific field.

The predictive model is given on the Fig. 5.

As the agriculture is an entrepreneurship, agricultural producers pursue maximizing the profits they get from selling the harvest they grew.

That is why besides the yield prediction, they also need profit prediction. Such predictive model is given on the Fig. 6.

Several months before the beginning of the season, so-called technological map is formed for the agricultural producer, which contains the parameters of operations to perform during the season.

The values of the parameters are defined so that the agricultural producer could reach maximum profit by following the recommendations given.

By using predictive models, the NPK levels in the field soil at the beginning of the season are defined, and the schedule of fertilizers application is built.

By using the model of crop nutrients consumption process, the predicted yield value is determined.

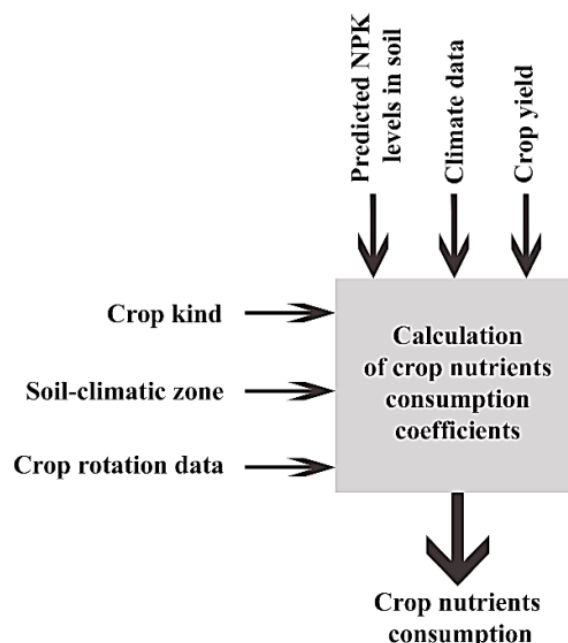


Fig. 5. The predictive model of crop nutrients consumption.

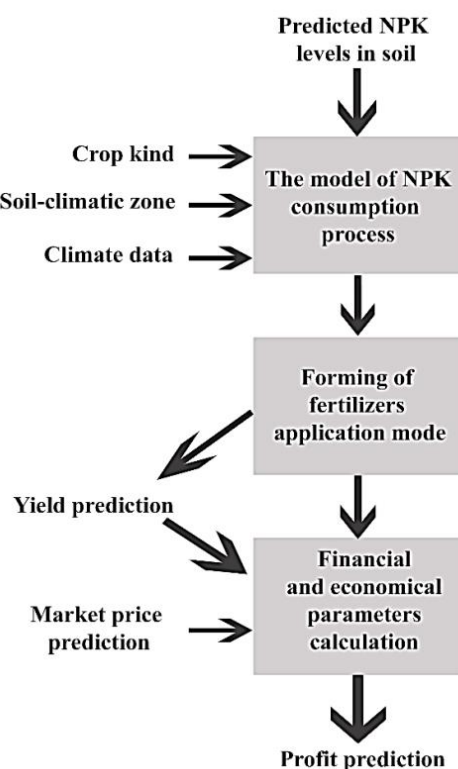


Fig. 6. The predictive model of profit.

By using this value and calculating the technological map operations costs, the predicted profit value is determined.

VI. CONCLUSION

In the paper, the methods of creating highly effective algorithms for decision support systems used in conditions of fertilizers differential applying are represented.

Along with the yield forecast, some agrochemical indicators could be useful. For fields with similar

characteristics (soil type, water-holding properties) from regions with a similar climate, it is possible to predict the concentration of certain chemical elements and compounds in the soil after applying fertilizers.

The input parameters of the forecast model will be statistical data of previous periods, as well as climatic indicators, under the influence of which chemical processes take place in the soil, contributing to (or hindering) the absorption of fertilizers by plants. In addition to mineral fertilizers, other chemicals also get into the soil - means of protection against pests, etc. They may contain compounds that can affect the values of agrochemical parameters. This factor will also be taken into account in the model. For the cultivation of various crops by agricultural enterprises, technological maps are formed in which, taking into account the climatic zone, the recommended scenarios of the work of agricultural producers are given, which make it possible to achieve a given yield.

For the best assimilation of minerals contained in the soil, different cultures need to moisture and temperature of the air were in certain ranges, beyond which the metabolism slows down. If we trace the dependence of the dynamics of the amount of various trace elements in the soil on the amount of precipitation and temperature (i.e., determine the rate at which plants absorb minerals found in the soil or coming in with fertilizers for different values of climatic indicators), we can determine the required amount of fertilizers applied depending on the region's climate (it may be necessary to apply less fertilizer, because the climate does not allow the plants to assimilate all the mineral substances introduced, or on the contrary, plants may need additional feeding during the season).

The algorithms are based on the Data Mining.

Case studies of applying these algorithms and their application in crop farming technological processes are given.

The associative search algorithms application for models developing of crop yield prediction allows compensating of data insufficiency from a specific field with data from another fields with similar characteristics. As a result, high accuracy of the models built can be reached. These models allow more effective decision making on managing agribusiness and obtaining higher income from business as well as reaching positive ecological effect by means of more rational use of mineral fertilizers.

The algorithms proposed have no analogues in the world scientific literature, and results of their practical use show their efficiency.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Natalia N. Bakhtadze developed the concept and forecasting algorithms; Evgeny M. Maximov developed algorithms and implemented software implementation; Natalia E. Maximova proposed a methodology for the management of an agricultural enterprise; Lamara N. Kozlovskaya participated in the development of the concept.

REFERENCES

- [1] B. Aubert, A. Schroeder, and J. Grimaudo, "IT as enabler of sustainable farming: An empirical analysis of farmers' adoption decision of precision agriculture technology," *Decision Support Systems*, vol. 54, pp. 510–520, 2012.
- [2] D. Ehlert, R. Adamek, and H. Horn, "Vehicle based laser range finding in crops," *Sensors*, pp. 3679–3694, 2009.
- [3] N. Pavon-Pulido, J. Lopez-Riquelme, R. Torres, R. Morais, and J. Pastor, "New trends in precision agriculture: A novel cloud-based system for enabling data storage and agricultural task planning and automation," *Precis. Agric.*, vol. 18, pp. 1038–1068, 2017.
- [4] N. Bakhtadze, V. Kulba, V. Lototsky, and E. Maximov, "Identification-based approach to soft sensors design," in *Proc. IFAC Workshop of Intelligent Manufacturing Systems*, Alicante, Spain, 2007, pp. 86–92.
- [5] N. Bakhtadze, V. Lototsky, S. Vlasov, and E. Sakrutina, "Associative Search and Wavelet Analysis Techniques in System Identification," *IFAC Proceedings*, vol. 45, no. 16, 2012, pp. 1227–1232.
- [6] N. Bakhtadze and V. Lototsky, "Knowledge-Based Models of Nonlinear Systems Based on Inductive Learning," *New Frontiers in Information and Production Systems Modelling and Analysis Incentive Mechanisms, Competence Management, Knowledge-Based Production*, Heidelberg: Springer, 2017, ch. 5, pp. 85–104.
- [7] IBM Research Precision Agriculture. (2018). [Online]. Available: www.ibm.com/us-en/marketplace/agri-resrch-innov
- [8] (2018). [Online]. Available: *Biomax. Agriculture IoT System*. www.biomax.com.tw/en/product/Agriculture-IoT-System/agriculture_iot_system.html
- [9] N. Bakhtadze, E. Maximov, D. Donchan, N. Maximova, and L. Kozlovskaya, "Innovation and knowledge management in precision farming," in *Proc. the 12th IADIS International Conference on Information Systems*, Utrecht, Netherlands: International Association for Development of the Information Society, 2019, pp. 265–270.

Copyright © 2021 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).



Natalia N. Bakhtadze was born in Yaroslavl, Russia, in 1956. In 1979, she graduated from the Moscow State Institute of Electronics and Mathematics today - as part of the National Research University Higher School of Economics, majoring in Applied Mathematics. She received her PhD (1993) and doctor degree (2003) in applied mathematics and systems analysis. Her research interests: identification, model predictive control, data mining, wavelet analysis. Some publications:

N. Bakhtadze, et al. Digital Energy Ecosystems. *IFAC PapersOnLine*, Volume 52, Issue 13, (2019) 30–35.

New Frontiers in Information and Production Systems Modelling and Analysis Incentive Mechanisms, Competence Management, Knowledge-based Production. Editors: Różewski, P., Novikov, D., Bakhtadze, N., Zaikin, O. ISBN: 978-3-319-23337-6. Springer, 2016. 268 p. <http://www.springer.com/978-3-319-23337-6>

Bakhtadze, et al. Data Mining-Based Prediction of Manufacturing Situations. *IFAC-PapersOnLine*. 2018. Volume 51, Issue 11. C. 316–321.

Since 1980 to present she works at the V.A. Trapeznikov Institute of Control Sciences (ICS), Moscow, Russia of the Russian Academy of Sciences. Since 2006, she is Head of the Control Systems Identification Laboratory of ICS. Professor of the N.E. Bauman National Technical University, Moscow, Russia.

Prof. Bakhtadze is a vice chair of Technical Committee on Management and Control in Manufacturing and Logistics of International Federation of Automatic Control – IFAC.



Evgeny M. Maximov was born in Sarni, Rivne region, Ukraine (at that time - in the USSR), in 1950. In 1974, he graduated from the Moscow State Institute of Electronics and Mathematics today - as part of the National Research University Higher School of Economics, majoring in Applied Mathematics. He received his PhD in 1987 in applied mathematics and systems analysis. His research interests: identification, adaptive control, data mining, predictive modeling.

Some publications:

Maximov E.M., Yadykin, I.B. Knowledge-Based Models for Smart Grid / *New Frontiers in Information and Production Systems Modeling and Analysis: Incentive Mechanisms, Competence Management, Knowledge-based Production*. Heidelberg, Germany: Springer, 2016. C. 189-210.

Bakhtadze, N.; Yadykin, I.; Pyatetsky, V.; Lototsky, V.; Maximov, E.; Sakrutina, E.; Vershinin, Y. Stability Analysis of Discrete Power Supply Systems in Industry. *International Journal of Production Research*. 1916, TPRS-2016-IJPR-1838

Since the 1974 Dr. Maximov works (at present - as Senior Scientific Researcher) at the V.A. Trapeznikov Institute of Control Sciences (ICS), Moscow, Russia of the Russian Academy of Sciences, in Control Systems Identification Laboratory.



Natalia E. Maximova was born in Moscow, Russia, in 1982. She has a specialist degree in organization management earned at the Moscow State Institute of Electronics and Mathematics today - as part of the National Research University Higher School of Economics, Moscow, Russia, in 2005. Has a PhD degree in Technical Sciences, the specialty is Automation and Control of Technological Processes and Manufactures, earned at V.A. Trapeznikov Institute of Control Sciences, Moscow, Russia, in 2014. Major research interests: control system identification, data mining, associative search, predictive modelling.

She has been working at V.A. Trapeznikov Institute of Control Sciences, Moscow, Russia, since 2007. Current position is a Researcher. Publications: "Innovation and knowledge management in precision farming," in *Proc. 12th IADIS International Conference on Information Systems*, Utrecht, Netherlands, 2019, pp. 265-270;

"Intelligent algorithms of power grid's state identification," *Journal of Information Technologies and Computing Systems*, – no. 3, pp. 45-50, Sept. 2011.

Dr. Maximova is the executive editor-in-chief of "Control Sciences" scientific journal indexed by the Higher Attestation Commission of the Russian Federation for the defense of theses.



Lamara N. Kozlovskaya was born in the Sukhumi city, Georgia (then - in the USSR), 1951. Education: Lomonosov Moscow State University, Faculty of Biology, Department of Molecular Biology, 1975. She got the Ph.D. in biological sciences, 1980, A.N. Bach Institute of Biochemistry, Russian Academy of Sciences, Federal Research Centre "Fundamentals of Biotechnology" of Russian Academy of Sciences. Moscow, Russia.

Her research interests: study of the chemical polymorphism of medicinal essential oil plants.

Since 2000 to the present, she is Associate Professor of the Department Botany, Plant Breeding and Seed Technology, Russian State Agrarian University - Moscow Timiryazev Agricultural Academy, Moscow, Russia. Bakhtadze N., Maximov E., Donchan D., Kozlovskaya L., Maximova N. Innovation and knowledge management in precision farming. *Proceedings of the 12th IADIS International Conference Information Systems*. 2019. P. 265-270; Malankina E.L., Kozlovskaya L.N., Kuzmenko A.N., Evgrafov A.A. Determination of the component composition of essential oil of thyme species by the method of gas chromatography. *Moscow University Chemistry Bulletin*. 2019. T. 74. № 6. P. 310-314.