UDC 004.832;338.27 **doi: 10.26906/SUNZ.2021.3.108**

E. Skakalina

National University "Yuri Kondratyuk Poltava Polytechnic", Poltava, Ukraine

SYNTHESIS TRENDS OF FORECASTING USING INDUCTIVE MODELING METHODS

Abstract.Modern development of computer technology and the possibility of implementing calculations in parallel allow to solve increasingly large-scale problems of numerical modeling. The development of multiprocessor computing and parallel computing makes it important to solve problems of optimization analysis. The optimization analysis is based on the mass solution of inverse problems when the defining parameters of the considered class of problems change in certain ranges. Thus, calculations of not only direct problems where it is necessary to model the phenomenon at the known initial data, but also calculations of inverse problems where it is necessary to define on what defining parameters there is this or that phenomenon become more and more demanded. This formulation requires multiple solutions of direct problems and solving the problem of optimization analysis and construction of predictive trends. Sets of multidimensional parametric data in the paper are considered as numerical solutions of the optimization problem. The construction of predictive trends is implemented on the basis of the group method of data handling as a direction of induction modeling. The methodology of visualization of results of calculation of parametric functions is realized. The scheme of Data Mining with application of methods of visualization by means of the Matlab software environment is described.

Keywords: multiparameter functions, inductive modeling, group method of data handling, visualization, trends, optimization of results, Matlab.

Introduction

Statement of the problem of synthesis of forecast trends requires multiple solutions of direct problems and solution of the problem of optimization analysis.

The final solution of such problems are multidimensional arrays of discrete quantities that express the dependence of the desired function (control parameter) on the defining parameters of the problem. The thus obtained multidimensional numerical results require processing and analysis.

Initially, the methods of data analysis were focused in the course of historical development on the processing of the results of physical and engineering experiments, as well as on the processing of the results of statistical observations. Real physical experiments were conducted in three-dimensional space. Therefore, the methods designed to work with one-dimensional, two-dimensional and three-dimensional results are well-established and well-known. With the development of computing power and algorithms, it has become possible to process huge arrays of observational data and experiments in various fields. Solutions of inverse and optimization problems became possible, which became multidimensional arrays that express the dependences of control parameters in the multidimensional space of defining parameters.

We needed tools to analyze data implemented not in three-dimensional but in multidimensional space.

Over the past decade, the analysis of multidimensional data has become one of the main areas of applied mathematics, actively developing and applied in almost all areas of research.

Multidimensional Data Analysis (MDA) is one of the most popular and sought-after interdisciplinary fields of knowledge and an active tool for the synthesis of various disciplines.

At first, the most intensive tools of AMD were developed in applied analytical chemistry. That is why there is another common name for AMD methods - Chemometrics [1]. Today, the study of multidimensional data is a key section of modern mathematical statistics, analytical chemistry, environmental and geographical research.

MDA methods are used in econometrics in the analysis of financial and economic indicators, in psychometry in the analysis of the results of psychological surveys, in biology and medicine in the processing of observations [2]. When creating databases and developing DBMS methods and algorithms MDA are used to create matrix operators designed to process data placed in a multidimensional form [3].

One of the most well-known methods of analyzing multidimensional data is the principal components method and its generalization for nonlinear cases. Methods of analysis of multidimensional data are implemented in close relationship and interaction with the methods of factor and cluster analysis. In problems of mechanics of continuous media and computational physics, which involve optimization analysis of phenomena, the analysis of multidimensional data allows to find the optimal conditions for the occurrence of a physical phenomenon in the multidimensional space of the defining parameters of the problem.

There are many different definitions of AMD, dictated by the specific area and objectives of research, the algorithms used. In our case, where the source of multiparametric data functions are solutions to optimization problems and visualization of results, the following definition is used: This definition allows you to formulate the problem in a more general form and avoid the restrictions imposed by the use of specific methods solved by the problem and other similar factors.

The results of such calculations are multidimensional arrays, the dimension of which corresponds to the number of defining parameters. These arrays require processing and visual representation for data analysis. This requires a general scheme for working with such data.

1. Analysis of recent research and publications

Inductive modeling is a self-organizing process of evolutionary transition from primary data to explicit

mathematical models, reflecting the patterns of functioning of simulated objects and systems, which are implicitly contained in the available experimental, experimental, statistical data.

Inductive modeling (IM) allows you to select a model of optimal complexity from a given class of models to describe a limited set of experimental data. IM is effective when there is no or partial a priori information about a possible object model. This approach was proposed in the 80s by Acad. A.G. Ivakhnenko [4] and now IM is an integral part of Machine Learning and Data Mining technologies. The theory of inductive modeling based on the group method of data handling (GMDH) using the method of critical variances [5], allowed to explain the nature of the efficiency of GMDH as a method of constructing noise-tolerant models with minimal prediction error variance, and to solve the problem of optimizing]. The two-criteria method of redefining the choice of the model using the criterion of non-displacement of errors [7] allows to eliminate the ambiguity of the choice of the optimal model.

The principles of designing and implementing high-performance search algorithms of GMDH on the basis of recurrent calculations [8], procedures of parallelization of operations [9] and sequential selection of informative variables [10] allow to increase the dimensionality of the solved problems. The principles of constructing hybrid architectures of iterative algorithms of GMDH as a generalization of algorithmic structures of multirow, relaxation and combinatorial types allowed to develop a generalized iterative algorithm of GIA GMDH [11] as a neural network with active neurons in the form of COMBI algorithm for automatic detection. The Department of Information Technologies of Inductive Modeling of the International Research and Training Center for Information Technologies and Systems of the National Academy of Sciences of Ukraine has developed theoretical and practical foundations and algorithms:

- generalized relaxation iterative algorithm GRIA GMDH, based on the use of high-speed recurrent calculations and matrices of normal equations, which allows to solve problems of inductive modeling of ultra-large dimension [12];

- theoretical bases of a fundamentally new class of enumerative and iterative algorithms of GMDH with the use of recurrent-parallel calculations on cluster systems [13] as a basis of highly efficient intelligent technologies of inductive modeling;

- principles of designing technologies of intelligent modeling of complex systems based on the use of knowledge bases, inductive data analysis tools and intelligent user interface [14]. Such technologies should have three main tool levels: autonomous modeling from an existing database; built-in modeling as part of a real-time control system; complex modeling of a complex system to identify the optimal modes of its operation and critical scenarios;

- theoretical principles and tools for forecasting interconnected socio-economic processes based on statistical data in the class of discrete dynamic models of vector autoregression [15];

- principles of hybridization of GMDH search algorithms and genetic algorithms, on the basis of which the COMBI-GA search algorithm was developed [16]

2. Statement of the research problem

It is necessary to synthesize a multi-parameter function that will display the production data of the main trends over the years of the agricultural holding, use the algorithm of the process of optimization and visualization of decision-making results. Agricultural production was chosen as the subject area.

To solve the optimization problem, you must first choose a mathematical method that would lead to the final results with the lowest computational costs, or to obtain the largest amount of solution search information, which widely uses the method of minimizing and maximizing linear and nonlinear functions.

In the search for a solution, the problems of processing and analysis of multidimensional volumes of numerical information given in the form of multidimensional arrays are considered. Multiparametric data are considered in the work as numerical solutions of the optimization problem and visualization of the results of parametric functions.

The function of the synthesized mathematical algorithm should be optimized, then display the visualization of trends, for this purpose the mathematical software environment MATLAB was used.

Matlab Numerical Analysis Software Package, which has an Optimization Toolbox, designed to solve complex mathematical problems, contains programs of well-known methods for minimizing and maximizing linear and nonlinear optimization functions for reliability, quality for various applications and tasks for implementing various construction methods. graphs [17].

The practical use of the mathematical method is largely determined by the formulation of the optimization problem, as well as the used mathematical model of the object of optimization, the characteristics of a particular subject area.

3. Basic material and results

The choice of a method of synthesis of forecast trends is largely determined by the formulation of a specific optimization problem, as well as a mathematical model of the object of optimization [18]. The general scheme of analysis of numerical data is given in Fig. 1.

The calculation must be implemented through a multiparameter function that will display the production data of the main trends of the agroholding over the years, the algorithm of the process of visualization and optimization of the program, which widely uses the method of minimizing and maximizing linear and nonlinear functions Optimization Toolbox, which is designed to solve complex mathematical problems of optimizing the cost of reliability, quality for various applications and the implementation of various methods of graph visualization.

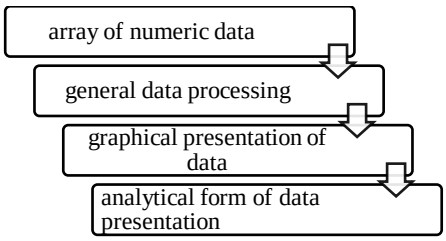

Fig. 1. General scheme of numerical data analysis

MATLAB is a high-level language of technical calculations, an interactive algorithm development environment and a modern data analysis tool. MATLAB in comparison with traditional programming languages (C / C ++, Java, Pascal, FORTRAN) allows to reduce time of the decision of typical problems and considerably simplifies development of new algorithms [18]. MATLAB is the basis of the entire family of MathWorks products and is the main tool for solving a wide range of scientific and applied problems in such areas as: object modeling and control systems development, communication systems design, signal and image processing, signal measurement and testing, financial modeling, computational biology, etc.

The MATLAB core allows you to easily work with matrices of real, complex and analytical data types. Contains built-in functions of linear algebra (LAPACK, BLAS), fast Fourier transform (FFTW), functions for working with polynomials, functions of basic statistics and numerical solution of differential equations. All built-in MATLAB kernel functions are designed and optimized by specialists and work faster or in the same way as their equivalent in C / C ++, which is a programming language for engineering and mathematical calculations. Large library of functions simplifies work (including graphical display of data). Key features:

•platform independence, high-level programming language focused on matrix computing and algorithm development

• Interactive environment for code development, file and data management

• Functions of linear algebra, statistics, Fourier analysis, solutions of differential equations, etc. Quantitative results of forecasting the main production and financial performance of agricultural holdings are shown in table 1.

Table 1 − **Generalized results of forecast trends**

Characteristic	**The range of deviation of the predicted value from the control (%)**
The cost of goods sold	0,23 – 5,08
Gross profit	0,64 – 3,77
Net profit	3,30 – 9,96
Costs	0,74 – 5,06

• Rich visualization tools, 2-D and 3-D graphics

• Built-in UI development tools for creating complete applications on MATLAB

• Tools for integration with C / C ++, code inheritance, ActiveX technology

• Access to .NET features.

MATLAB has more than a million production users and scientists. The price of the basic commercial version without tools is about 2000 dollars. US dollars and only 100 dollars. USA for educational institutions with a minimum set of tools.

MATLAB provides the user with a large number of data analysis functions that cover almost all areas of mathematics, including:

♦ Matrices and linear algebra - matrix algebra, linear equations, eigenvalues and vectors, singularities, matrix factorization and more.

♦ Polynomials and interpolation - roots of polynomials, operations on polynomials and their differentiation, interpolation and extrapolation of curves

♦ Mathematical statistics and data analysis - statistical functions, statistical regression, digital filtering, fast Fourier transform and others.

♦ Data processing - a set of special functions, including graphing, optimization, zero search, numerical integration and more.

♦ Differential equations - solution of differential and differential-algebraic equations, differential equations with delay, equations with constraints, equations in partial derivatives and more.

♦ Sparse matrices - a special data class of the MATLAB package used in specialized applications.

♦ Integer arithmetic - performing integer arithmetic operations in the MATLAB environment.

MATLAB got its name from MATrix LABoratory, which was founded in the late 1970s by Clive Moler, who later became head of the Department of Computer Science at the University of New Mexico. He designed it to give his students access to LINPACK and EISPACK without having to master Fortran. MATLAB soon became popular at other universities and attracted the attention of applied mathematicians. Engineer John Little fell in love with this product when he visited Moller at Stanford University in 1983. Predicting MATLAB's commercial success, he joined Moller and Steve Bangert. They rewrote MATLAB in C and founded.

The MathWorks in 1984. The rewritten libraries became known as JACKPAC. Apart from the recognition of teachers of linear algebra and numerical analysis, MATLAB has been recognized de facto by specialists in working with digital images (eg tomography).

MATLAB supports the creation of applications with graphical user interface properties. MATLAB includes a GUIDE (GUI development environment) for graphical design of graphical user interfaces. Visualization of imported data in the MATLAB environment is shown in Fig. 2. Visualization of forecast trends for the years 2019 - 2020 is shown in Fig. 3,. 4. All modifications of GMDH for autoregressive and distributive lag models demonstrate high accuracy of the forecast.

The best accuracy is not clear GMDH, and GMDH with fuzzy inputs. In addition, their advantage over clear GMDH is that they do not use MNCs and are not sensitive to the poor conditionality of the matrix of input factors and autocorrelation of random variables, which is so important for autoregressive models.

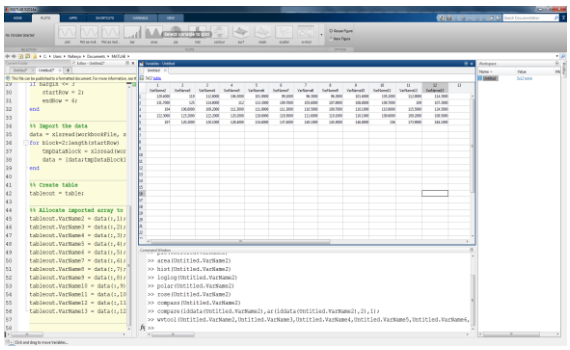

Fig. 2. Imported data

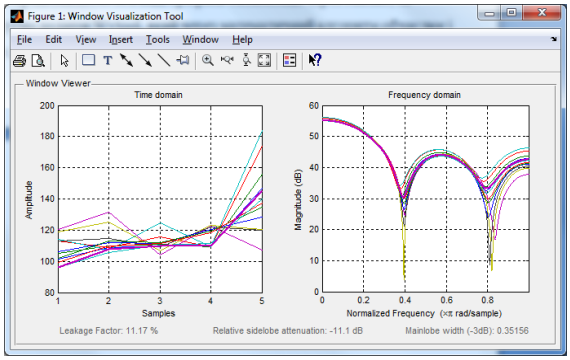

Fig. 3. Graphs of trends in the optimization of multiparameter functions for 2019

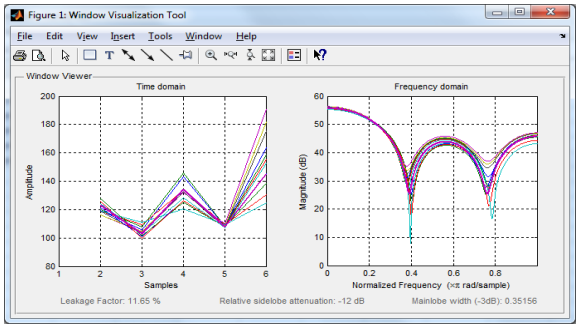

Fig. 4. Graphs of trends in the optimization of multiparameter functions for 2020

The basic two-stage methodology for the synthesis of forecast trends using the method of group consideration of arguments is shown in Fig. 5. At the first stage the synthesis of forecast trends with the use of inductive modeling is realized [19,20]. The second stage is the visualization of numerical trends in the MATLAB software environment.

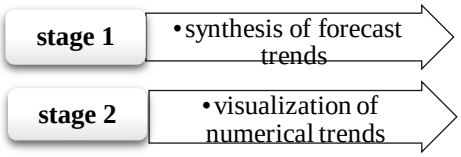

Fig. 5. Methodology of synthesis of forecast trends using the Group method of data handling

Conclusions

Taking into account the formulated input requirements, a multiparameter function was implemented and optimized to solve the problem of optimizing the results of the main gross trends of the agricultural holding by years. Any complex system consists of separate simpler subsystems (elements). Therefore, naturally, solving the problem of multiparametric optimization for the system as a whole, to set and solve the problem of multiparametric optimization for its individual subsystems. At the same time coordination of optimality of subsystems according to their purpose and communications existing between subsystems should be carried out.

The set of system quality indicators can be considered as a vector, so multiparameter optimization is also called vector.

To solve the problems of multiparameter optimization, certain conditions must be provided. In particular, it should be possible to change, within certain limits, independent variables that affect quality criteria.

Any independent variable that can be changed within certain limits and that has some effect on all quality criteria or only some of them, is called a controlled change.

This terminology is in a sense consonant with the terminology of management theory. The set of all controlled variables can be considered as a control vector.

It corresponds to the point of n-dimensional control space. The dependence of quality criteria on managed variables is some reflection of the management space on the goal space.

An effective compromise is a set of all target points that cannot be further and evenly improved within the available management capabilities. Thus, this includes all points that are incomparable with each other in the sense of improving or deteriorating the control effect.

When implementing the derivation of optimization data in the price indices of industrial producers and sales of agricultural products of the agricultural holding.

The optimal search for the best solutions was carried out, which significantly improves the value of the system efficiency criterion. Thesis examines the theoretical issues of optimization of multiparameter functions, production data by years of the main trends of the agroholding to optimize results, statistical decision rules with procedures for solving the multiparameter function for calculating statistics of agroholding trends and effective management decisions.

The results of the study can be used for further research on the synthesis of forecast trends and optimization of modeling results by year of production of agricultural enterprises and decision-making. A numerical experiment was performed to process the actual input data for multiparameter functions on the price indices of industrial producers and the realization of agricultural prices of gross trends by years. Visualization of forecast trends with the help of MATLAB software environment is performed. Predicting trends are synthesized using the method of group consideration of arguments, which is the direction of inductive modeling, which shows the full adequacy of the proposed methodology for the task.

REFERENCES

1. Чарльз Генрі Едвардс, Девід Е. Пенні. Диференціальні рівняння і проблема власних значень: моделювання та обчислення за допомогою Mathematica, Maple і MATLAB Differential Equations and Boundary Value Problems: Computing and Modeling. - 3-е изд. - М.: «Вільямс», 2007. - 397 с, ISBN 978-5-8459-1166-7.
2. С.П. Іглін. Математичні розрахунки на базі Matlab. Видавництво "BHV-Санкт-Петербург" 2005р. 640 с.
3. Є.Р. Алексєєв, О.В. Чеснокова. Рішення задач обчислювальної математики в пакетах Mathcad 12, MATLAB 7, Maple 9. Серія: Самовчитель. Видавництво: НТ Пресс, 2006р. 496 с. ISBN 5-477-00208-5.
4. Ивахненко А.Г. Индуктивный метод самоорганизации моделей сложных систем. – Киев: Наук. думка, 1982. – 296 с.
5. Samoilenko O., Stepashko V. A method of Successive Elimination of Spurious Arguments for Effective Solution of the Search-Based Modelling Tasks // Proc. of the II Int. Conf. on Inductive Modelling, Sept. 15-19, 2008, Kyiv, Ukraine. – IRTC ITS NASU, Kyiv, 2008. – P. 36-39.
6. Stepashko V., Bulgakova O. Generalized Iterative Algorithm GIA GMDH // Proc. of the 4th Int. Conf. on Inductive Modelling ICIM-2013, Sept. 16-20, 2013, Kyiv, Ukraine. – Kyiv: IRTC ITS NASU, 2013. – P. 119-123.
7. Yefimenko S., Stepashko V. Intelligent Recurrent-and-Parallel Computing for Solving Inductive Modeling Problems // Proc. of 16th Int. Conf. on Computational Problems of Electrical Engineering CPEE'2015, Lviv, Ukraine, September 2-5, 2015. – Lviv: LNPU, 2015. – P. 236-238.
8. Степашко В.С. Концептуальные основы интеллектуального моделирования // УСиМ. – 2016. – № 4. – С. 3-15.
9. Ефименко С.Н. Построение систем прогнозных моделей многомерных взаимосвязанных процессов // УСиМ. – 2016. – № 4. – С. 80-86.
10. Stepashko V., Moroz O. Hybrid Searching GMDH-GA Algorithm for Solving Inductive Modeling Tasks // Proc. of the 1st IEEE International Conference on Data Stream Mining & Processing, 23-27 August 2016, Lviv, Ukraine. – P. 350-355.
11. Self-organizing methods in modeling: GMDH type algorithms / Ed. S.J. Farlow. – New York, Basel: Marcel Decker Inc., 1984. – 350 p.
12. Bodyanskiy Ye., Zaychenko Yu., Pavlikovskaya Ye. The Neo-Fuzzy Neural Network Structure Optimization Using the GMDH for the Solving Forecasting and Classification Problems / Proc. of the 3rd Int. Workshop on Inductive Modeling IWIM–2009, 14–19 Sept. 2009, Krynica, Poland. – Prague: Czech Techn. Univ., 2009. – P. 100–107.
13. Булгакова О.С., Зосімов В.В., Степашко В.С. Програмний комплекс моделювання складних систем на основі ітераційних алгоритмів МГУА з можливістю мережевого доступу // Системні дослідження та інформаційні технології. – 2014. – № 1. – С. 43-55.
14. Павлов А.В. Проектирование системы автоматизированной структурно-параметрической идентификации // Індуктивне моделювання складних систем. – Вип. 7. – К.: МННЦ ІТС НАНУ, 2015. – С. 202-219.
15. [Stepashko V., Samoilenko O., Voloschuk R. Informational Support of Managerial Decisions as a New Kind of Business Intelligence Systems. – Computational Models for Business and Engineering Domains / G.Setlak, K.Markov (Eds.). – Rzeszow, Poland; Sofia, Bulgaria: ITHEA, 2014. – P. 269-279.
16. Zosimov V., Stepashko V., Bulgakova O. Inductive building of search results ranking models to enhance the relevance of the text information retrieval // Proc. of the 26th Intern. Workshop "Database and Expert Systems Applications", 1-4 Sept., Valencia, Spain / Ed. by Markus Spies at al. – Los Alamitos: IEEE Computer Society, 2015. – P. 291-295.
17. Оленєв М.М., Печонкін Р.В., Ченців А.М. Паралельне програмування в MATLAB і його додатки. М .: ВЦ РАН. 2007. 120 с. - ISBN 5-201-09865-7
18. E. Skakalina. INFORMATION TECHNOLOGY FOR BUILDING A MODEL OF FINANCIAL FORECASTING / E.Skakalina // Proceedings of the 16th INTERNATIONAL SCIENTIFIC CONFERENCE INFORMATION TECHNOLOGIES AND MANAGEMENT 2018. April 26-27, 2018. ISMA University, Riga, Latvia, P.195-196.
19. Optimization Toolbox - Оптимізація [Електронний ресурс]. - 2021. - Режим доступу: http://matlab.exponenta.ru/optimiz/index.php
20. Skakalina E. Investigation of intelligent technologies for formation forecasting models / Elena Skakalina // International Journal of Engineering & Technology.- 2018.- 7(3.2). – P.413-418. DOI: 10.14419/ijet.v7i3.2.14563

Received (Надійшла) 15.06.2021
Accepted for publication (Прийнята до друку) 11.08.2021

Синтез прогнозних трендів із застосуванням методів індукційного моделювання

О. В. Скакаліна

Анотація.Сучасний розвиток обчислювальної техніки і можливість реалізації обчислень в паралельному режимі дозволяють вирішувати все більш масштабні завдання чисельного моделювання. Розвиток багатопроцесорний обчислювальної техніки та паралельних обчислень робить актуальним вирішення задач оптимізаційного аналізу. Оптимізаційний аналіз заснований на масовому рішенні зворотних задач при змінюються в певних діапазонах визначальних параметрах розглянутого класу задач. Так, все більш затребуваними стають розрахунки не тільки прямих завдань, де потрібна моделювати явище при відомих вихідних даних, але і розрахунки зворотних задач, де необхідно визначити за яких визначальних параметрах виникає те чи інше явище. Така постановка вимагає багаторазового розв'язання прямих задач і рішення задачі оптимізаційного аналізу та побудови прогнозуючих трендів.. В роботі розглядаються проблеми Data Mining багатовимірних обсягів числової інформації, заданих у вигляді багатовимірних масивів. Множини багатовимірних параметричних даних в роботі розглядаються, як чисельні рішення задачі оптимізації. Побудова прогнозуючих трендів реалізується на базі методу групового урахування аргументів як напрямку індукційного моделювання. Реалізована методологія візуалізації результатів обчислення параметричних функцій. Описана схема Data Mining із застосуванням методів візуалізації засобами програмного середовища Matlab.

Ключові слова: багатопараметричні функції, індуктивне моделювання, метод групового урахування аргументів, візуалізація, тренди, оптимізація результатів, Matlab.