

# Examples of intelligent adaptation of digital fields by means of the system GeoBazaDannych

Valery B. Taranchuk

*Department of Computer Applications and Systems*

*Belarusian State University*

Minsk, Republic of Belarus

taranchuk@bsu.by

**Abstract**—Examples of interactive formation of digital models of geological objects in computational experiments that meet the intuitive requirements of the expert are given and discussed. Special tools of the system GeoBazaDannych are noted, and the results of comparison with standard solutions in the complex "The Generator of the geological model of the deposit" are presented. Examples of approximation and reconstruction of the digital field and its interactive adaptation by means of the system GeoBazaDannych are given and discussed; the obtained solutions and their accuracy are illustrated by maps of isolines.

**Keywords**—digital geological model, system GeoBazaDannych, interactive graphical visualization, intelligent adaptation of digital fields, "smart" methods of computer model adaptation

## I. INTRODUCTION

Geological modeling is an independent direction that includes the improvement of mathematical methods and algorithms; development of computer programs that provide a cycle of model construction, forming, filling and maintenance of databases [1]. The corresponding software includes the loading from different sources and data preprocessing, correlation, creation of digital cubes of reservoir properties, interactive data analysis, visualization with the help of any type graphics, mapping. The task of developing and implementing various computer-based geological models with self-tuning tools is one of the priorities. Herewith, an important component is the task of evaluating the adequacy and accuracy of the proposed digital models. The key issues are automation, adaptation of models taking into account continuously incoming additional data, as well as a revision of the results of processing the initial information using new interpretation methods [2].

The data used in geological and geoecological models are a representative part of the geodata, which classify, summarize information about processes and phenomena on the earth's surface. This information becomes really useful when integrated into a single system. Geodata, as a generalization of accumulated information, include information not only from the field of Earth sciences, but also from others, such as transport, economics, ecology, management, education, analysis, artificial intelligence.

Technological, system, and information features of geodata are noted in [3].

Geodata volumes are growing at a very high rate. Accordingly, it is natural to use "big data" technologies (the specifics for geodata are described in [4]), including automated data mining. One of the main aims of data mining is to find previously unknown, non-trivial, practically useful and understandable interpretations of knowledge in "raw" (primary) data sets [5], [6]. At the same time, following [5], "data mining does not exclude human participation in processing and analysis, but significantly simplifies the process of finding the necessary data from raw data, making it available to a wide range of analysts who are not specialists in statistics, mathematics or programming. Human participation is expressed in the cognitive aspects of participation and the application of informational cognitive models".

Geodata mining tools are the same as for usual data; the basis is the theory, methods, and algorithms of applied statistics, databases, artificial intelligence, and image recognition. There are many different active and applied software tools for data mining, for example, 8 classes of data mining systems are identified in [5]. The variety of proposed methods and software tools make it necessary to assess the quality of geodata and determine their main characteristics. Criteria for evaluating the quality of geodata are discussed in [7].

A number of issues related to the analysis and evaluation of spatial data quality can be solved using the computer system GeoBazaDannych [8] – [10]. Possible options, methodological solutions, and software tools that allow you to confirm the validity of interpretations, visualize and obtain numerical values of errors calculated by different methods of intellectual data processing results included and used in computer geological models are discussed below. For illustrations, the key task of forming and processing digital fields used in computer models is selected. In particular, we discuss the proposed methods that have been tested for solving different applied problems, as well as implemented in the interactive computer system GeoBazaDannych specialized algorithms for calculating approximating digital fields.

The interactive computer system GeoBazaDannych is the complex of intelligent computer subsystems, mathematical, algorithmic and software for filling, maintaining and visualizing databases, input data for simulation and mathematical models, tools for conducting computational experiments, algorithmic tools and software for creating continuously updated computer models. By means of the system GeoBazaDannych, it is possible to generate and visualize digital descriptions of spatial distributions of data on sources of contamination, on the geological structure of the studied objects; graphically illustrate solutions to problems describing the dynamic processes of multiphase filtration, fluid migration, heat transfer, moisture, and mineral water-soluble compounds in rock strata; design and implement interactive scenarios for visualization and processing the results of computational experiments. GeoBazaDannych's subsystems allow you to calculate and perform expert assessments of local and integral characteristics of ecosystems in different approximations, calculate distributions of concentrations and mass balances of pollutants; create permanent models of oil production facilities; generate and display thematic maps on hard copies. The main components of the system GeoBazaDannych [8] – [12]:

- the data generator Gen\_DATv;
- the generator and editor of thematic maps and digital fields Gen\_MAPw;
- modules for organizing the operation of geographic information systems in interactive or batch modes;
- the software package Geo\_mdl – mathematical, algorithmic and software tools for building geological models of soil layers, multi-layer reservoirs; modules for three-dimensional visualization of dynamic processes of distribution of water-soluble pollutants in active soil layers;
- software and algorithmic support for the formation and maintenance of permanent hydrodynamic models of multiphase filtration in porous, fractured media;
- the integrated software complex of the composer of digital geological and geocological models (GGMD).

## II. EXAMPLES OF INTERACTIVE ADAPTATION OF DIGITAL FIELDS

Integration of the capabilities of various geographic information systems (GIS) and GeoBazaDannych is provided by a wide range of tools of the system for importing and exporting data, images, and functions. Several non-standard solutions that are recognized as difficult for all geodata processing packages are discussed below. The examples of approximation and reconstruction of the digital field, its interactive adaptation by means of the system GeoBazaDannych and evaluation of the accuracy of results using the tools of the GGMD complex illustrate

the unique capabilities of the developed methods and software. The task of reconstruction of the grid function involves calculating the values of the approximating function at regular grid points from the values of randomly located experimental data points (observations), i.e. creating a regular array of Z values of node points from an irregular array of (X,Y,Z) values. The term “irregular array of values” means that the X, Y coordinates of data points are distributed irregularly across the function definition area. The procedure for constructing a regular network of level values and restoring the grid function is an interpolation or extrapolation of values from a collection of scattered sets of source points and values of surface levels in them to uniformly distributed nodes in the study area.

Methods for restoring grid functions and the corresponding algorithms are implemented in several specialized computer graphics and GIS packages. They can be divided into two classes: exact and smoothing interpolators [13], [14]. In fact, the method falls into a particular class depending on the user-defined settings when performing value calculations. Most methods for restoring the function and constructing a digital field are based on calculating weighted coefficients that are used to weigh the values of the measurement data at the nearest points. This means that, all other things being equal, the closer the data point is to a network node, the more weight it has in determining the value of the function being restored at that node.

It should be understood that restoring a grid function does not imply finding a single solution to a certain mathematical problem. Subjective opinion and expert qualifications are factors that are always present in such activities [15]. Therefore, for building digital models, it is important to have tools for interactive data processing, simulation of possible situations for obtaining and correcting input information, and modules for mathematical processing and statistical analysis [13], [16]. The multi-format data exchange tools mentioned above is required for simultaneous work in multiple software environments. It is important for the user to have tools that allow them to “play” with the source data and compare the results with the prepared reference models. How this is implemented in the system GeoBazaDannych is described below, the corresponding preparatory actions in the GGMD complex are described in detail in [11], [12], [17].

The reference surfaces in the following examples are objects described by the expressions (1) and (2). The functions *fOrigin*, *fHill*, *fPyramid*, *fTrench*, and *zBasic* included in these expressions are described in [11], [17].

$$\begin{aligned}
zSurfB(x, y) = & fOriginB(x, y) + \\
& +45 \cdot fHillZ(0.004 \cdot (x - 400), 0.002 \cdot (y - 1900)) + \\
& +37 \cdot fPyramidA(0.004 \cdot (x - 350), 0.002 \cdot (y - 600)) + \\
& +15 \cdot fTrenchB(0.003 \cdot (x - 1100), 0.0015 \cdot (y - 400)) + \\
& +25 \cdot fHillyX(0.003 \cdot (x - 1100), 0.0014 \cdot (y - 1700)) + \\
& +15 \cdot fHill(0.005 \cdot (x - 2350), 0.003 \cdot (y - 950)) - \\
& -15 \cdot fHill(0.006 \cdot (x - 1900), 0.003 \cdot (y - 1300)) - \\
& -32 \cdot fHill(0.0012 \cdot (x - 3000), 0.0025 \cdot (y - 2000)) + \\
& +35 \cdot fPyramidB(0.002 \cdot (x - 3300), 0.002 \cdot (y - 650)), \\
fOriginB(x, y) = & zBasicB(x).
\end{aligned}
\tag{1}$$

$$\begin{aligned}
zSurfT(x, y) = & fOriginT(x, y) + \\
& +42 \cdot fHillZ(0.004 \cdot (x - 400), 0.002 \cdot (y - 1900)) + \\
& +37 \cdot fPyramidA(0.004 \cdot (x - 350), 0.002 \cdot (y - 600)) + \\
& +10 \cdot fTrenchB(0.004 \cdot (x - 900), 0.0015 \cdot (y - 500)) + \\
& +15 \cdot fHill(0.004 \cdot (x - 1100), 0.0014 \cdot (y - 1700)) + \\
& +15 \cdot fHill(0.005 \cdot (x - 2150), 0.003 \cdot (y - 1050)) - \\
& -15 \cdot fHill(0.006 \cdot (x - 1900), 0.003 \cdot (y - 1300)) - \\
& -13 \cdot fHill(0.0012 \cdot (x - 3000), 0.0025 \cdot (y - 2000)) + \\
& +25 \cdot fPyramidB(0.002 \cdot (x - 3300), 0.002 \cdot (y - 650)), \\
fOriginT(x, y) = & zBasicT(x).
\end{aligned}
\tag{2}$$

An example of creating a model of the reference surface, which is obtained from the base surface by adding elements of the listed types (2 pyramids, 5 hills, 1 trench), is shown below in Fig. 1, Fig. 2. It's important that in the resulting equations (1), (2) the coefficients in the formulas of perturbation elements  $fHill$ ,  $fPyramid$ ,  $fTrench$  are chosen by user while visual construction. Shown in figures images of surfaces  $zSurfB(x, y)$ ,  $zSurfT(x, y)$  give an aggregate vision of these surfaces. Individual details can be studied on the maps of isolines Fig. 3, Fig. 4. We emphasize attention on fragments of the  $zSurfB$  surface, namely, perturbations  $fHillZ(400, 1900)$ ,  $fHillyZ(1100, 1700)$ . The element  $fHillZ(400, 1900)$  simulates the presence of a plateau (clipping with a horizontal plane),  $fHillyZ(1100, 1700)$  simulates a split (in geology – shift-overshoot), and is mathematically described by clipping a smooth hill-type shape with a vertical plane. Similar perturbations are specified for the  $zSurfT$  surface, but instead of  $fHillyZ(1100, 1700)$ , the expression  $fHill(1100, 1700)$  is accepted – without clipping (no split). Such fragments of  $fHillZ$  and  $fHillyZ$  cannot be reproduced using standard methods for restoring digital fields. Also note that in the following illustrations, the isolines corresponding to the plateau levels (for  $zSurfB$  – level 9) are not displayed. In the  $zSurfB$  illustration, level 10 is shown to further emphasize the boundaries of disturbances placed on the plateau section.

Illustrations Fig. 5 – Fig. 10 show examples of possible solutions using the methods of “adjustment” distributions in the system GeoBazaDanych, and for

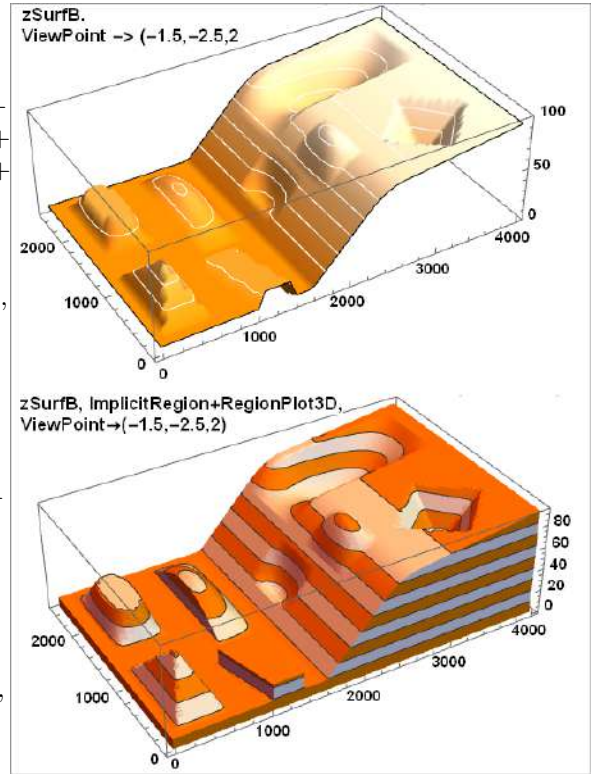


Figure 1. ZsurfB surface and volume graphs.

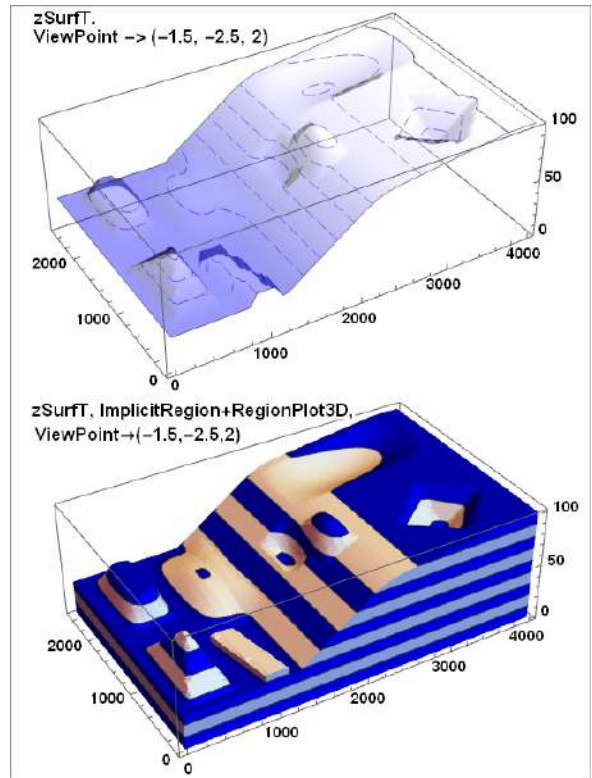


Figure 2. ZsurfT surface and volume graphs.

this purpose, the steps of simulating observations and restoring the distribution over a scattered set of points with measurements are performed. The corresponding results obtained in the [11] are shown on the isolines map in Fig. 4. The measurement points (163 dots) are placed on the profiles, they are given for different observation profiles by different primitives, and the centers of the perturbation forms are marked with circles (red circles, 10 dots). On maps, the isolines of the reference field are given as solid lines; to the restored field, they are dot-dashed green lines. There was no goal to choose a good system of profiles, they are formed sketchily, about how often designers do this. The reconstructed field is obtained by the 2nd-order interpolation method in Wolfram Mathematica. The implementation features are described in [12], note that in Mathematica, when working with irregularly placed data points, interpolation can only be performed using the 1st order method. Using the mentioned initial data, in Gen\_MAPw obtained the results of digital field reconstruction by spline approximation of GeoBazaDannych's method, the results are shown in Fig. 6.

How to understand the results shown in Fig. 7 – Fig. 10?

In this case, the task is to correct the digital fields obtained by the approximation algorithm via using such GeoBazaDannych's elements as subdomain, selected, split, and boundary conditions [10]. In fact, it is an interactive intellectual adaptation.

To get results in the system GeoBazaDannych, the data of measurements on profiles and control points in GGMD are exported to an Xls file that is imported into Gen\_MAPw. The corresponding illustration is given in Fig. 5, where the profiles, points of observation are shown; the values of measurements are displayed near their primitives. In the database, values are stored with machine precision, for brevity, the output of the map is made in the format of a single significant number after a decimal point. You should also note that the “overclocking numbers” command is applied to the image in Gen\_MAPw – where there was an overlap, the program shifted the digits of the values, and turned some of them.

The possibilities of adapting digital fields in Gen\_MAPw are illustrated in Fig. 7 – Fig. 10. The boundary of the area, the calculated grid is not changed, the boundaries of subdomains, inclusions, and the contour-split are entered (Fig. 7). “Allocated” sections (they're 4, different boundary conditions) provide autonomous calculation in nodes within these subdomains. Points that belong to these regions with measurements outside the “Selected” sections are not taken into account. Two sections marked with horizontal hatching have attributes that identify the areas bounded by them as inclusions. In inclusions, an autonomous calculation of the digital field is also performed, and points with

measurements that get inside “give” their values to the border, and they are taken into account in the external part.

The isolines of the distribution calculated in this way and the isolines of the reference field (Fig. 8 – Fig. 10) are very close. It should be noted that the initial capabilities of the system GeoBazaDannych have been significantly expanded in recent years. This was achieved by integrating and addition the system with the tools of the GGMD complex and the functions of the computer algebra system. In the current state, GeoBazaDannych provides users not only with the means to solve specific industrial tasks, but also with the possibility of scientific research on new methods of analysis and processing of initial data and used computer models. In particular, for the above problem, error estimates are obtained using the method described in [12]. We do not give specific digits here, because the errors of the so-calculated adapted field are comparable to the accuracy of obtaining grid functions, and these errors are an order of magnitude lower than in the unadapted field.

### III. CONCLUSION

The article discusses the questions of instrumental filling and use of the interactive computer system GeoBazaDannych. The results of intellectual processing of data included and used in computer geological models are presented and discussed.

### REFERENCES

- [1] P. Bedrikovetsky, *Mathematical Theory of Oil and Gas Recovery: With Applications to ex-USSR Oil and Gas*. Springer Science & Business Media, 2013.
- [2] 13th International Symposium on Software Engineering for Adaptive and Self-Managing Systems. Gothenburg, Sweden, May 28-29, 2018. Available at: <https://conf.researchr.org/track/seams-2018/seams-2018-papers> (accessed 2019, Dec).
- [3] V. P. Savinyh, V. Ya. Cvetkov, “Geodannye kak sistemnyi informatsionnyi resurs”, *Vestnik Rossiiskoi akademii nauk*, vol. 84, no. 9, pp. 826–829, 2014, (in Russian).
- [4] V. Ya. Tsvetkov, A. A. Lobanov, “Big Data as Information Barrier”, *European Researcher*, vol. 78, no. 7–1, pp. 1237–1242, 2014.
- [5] S. V. Shaitura, “Intellectual'nyi analiz geodannyh”, *Perspektivy nauki i obrazovaniya*, no. 6 (18), pp. 24–30, 2015, (in Russian).
- [6] V. V. Golenkov, “Ontology-based design of intelligent systems”, *Otkrytye semanticheskie tehnologii proektirovaniya intellektual'nyh sistem [Open semantic technologies for intelligent systems]*, pp. 37–56, 2017.
- [7] S. G. Dyshlenko, “Analiz i razrabotka harakteristik kachestva geodannyh”, *Perspektivy nauki i obrazovaniya*, no. 2 (20), pp. 23–27, 2016, (in Russian).
- [8] V. B. Taranchuk, “Chislennoe modelirovanie processov dvuhfaznoi mnogokomponentnoi fil'tracii v neodnorodnoi poristoi srede”, *Doktorskaya dissertatsiya*, Kazan' State University, 1992, (in Russian).
- [9] S. A. Barvenov, M. F. Kibash, V. B. Taranchuk, “Metodika, instrumentarii adaptatsii matematicheskikh modelei processov podzemnoi gidrodinamiki”, *Vybrannyya navukovyia pracy BDU*, t. “Matematika”, Minsk, pp. 34–65, 2001, (in Russian).
- [10] V. B. Taranchuk, “Postroenie cifrovyyh geologo-ekologicheskikh modelei v sisteme GBD-e”, *Geoekologiya Belarusi*. Minsk : Pravo i ekonomika, Nauchnoe izdanie, pp. 72–96, 2006, (in Russian).

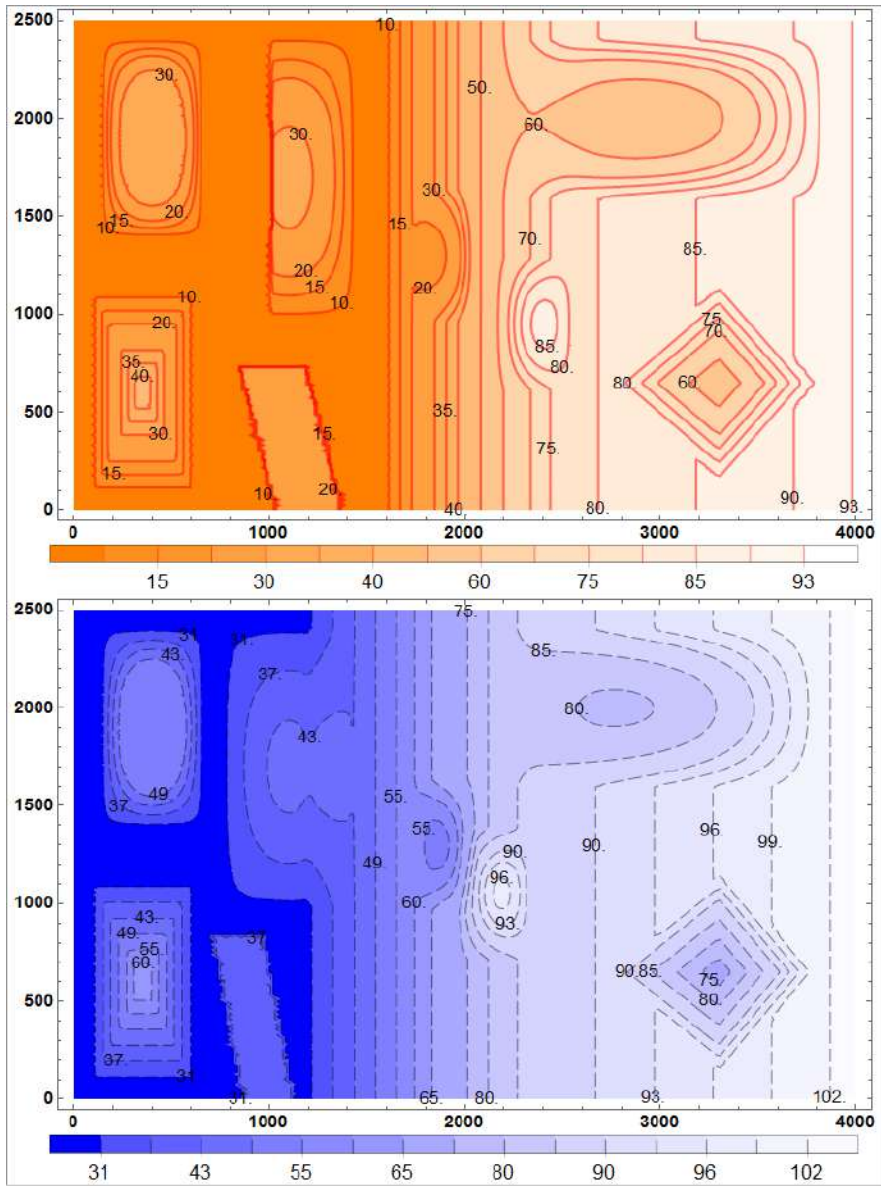


Figure 3. Isolines and density maps (color filling of intervals) of digital fields of  $z_{SurfB}(x,y)$ ,  $z_{SurfT}(x,y)$  functions. Also, color intervals are explained by legends and value labels, which are the same as the contour levels.

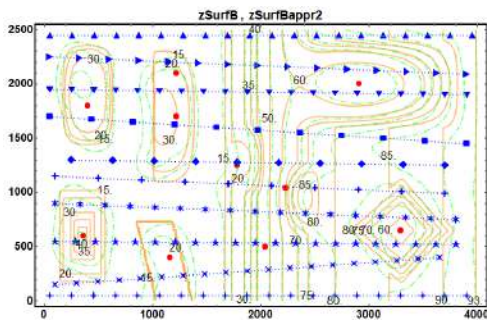


Figure 4. Reference centers of perturbations of the base surface (tape) with measurements and measurement points on the profiles. Isolines of the digital fields of the reference and restored distributions in GGMD.

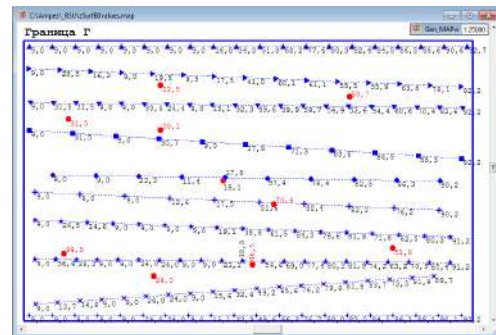


Figure 5. Results of exporting points with measurements from GGMD to Gen\_MAPw.



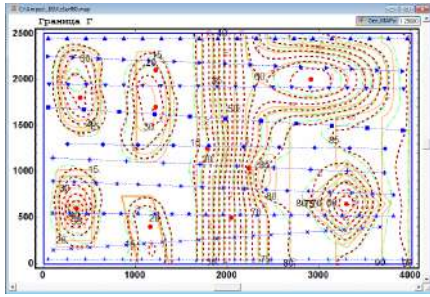


Figure 6. Isolines of reference and restored fields in GGMD and in Gen\_MAPw (spline approximation algorithm).

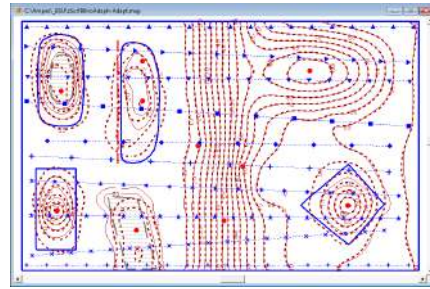


Figure 9. Result of Gen\_MAPw-no\_adapted and Gen\_MAPw-adapted distributions.

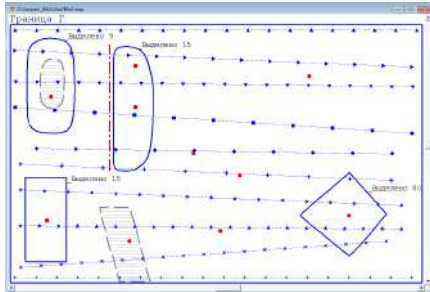


Figure 7. Input map for Gen\_MAPw: subdomains, inclusions, splits, boundary conditions. Version with expert additions for calculating the adapted digital distribution.

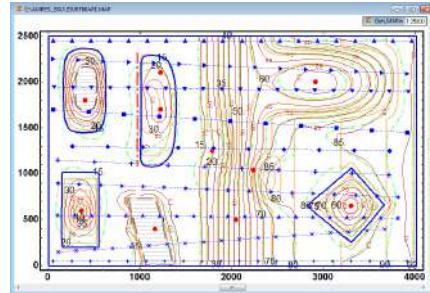


Figure 10. Result of Gen\_MAPw-adapted distribution and fields in GGMD.

- [11] V. Taranchuk, "The integrated computer complex of an estimation and adapting of digital geological models", *Studia i Materialy*, no. 2 (14), pp. 73–86, 2017.
- [12] V. B. Taranchuk, "Metody i instrumentarii ochenki tochnosti komp'yuternykh geologicheskikh modelei", *Vestnik BrGTU*, no. 2, pp. 64–69, 2018, (in Russian).
- [13] J. C. Davis, *Statistics and data analysis in geology*. New York: John Wiley & Sons Inc., 1986.
- [14] J. Sacks, W. J. Welch, T. J. Mitchell, H. P. Wynn, "Design and Analysis of Computer Experiments", *Statistical Science*, no. 4, pp. 409–435, 1989.
- [15] E. M. Halimov, "Detal'nye geologicheskie modeli i trehmernoe modelirovanie", *Neftegazovaya geologiya. Teoriya i praktika*, no. 3(7), pp. 1–10, 2012, (in Russian).
- [16] WOLFRAM MATHEMATICA. Available at: <http://www.wolfram.com/mathematica/> (accessed 2019, Dec).
- [17] V. B. Taranchuk, "Examples of the use of artificial neural networks in the analysis of geodata", *Open semantic technologies for intelligent systems*, vol. 3, pp. 225–230, 2019.

## Примеры интеллектуальной адаптации цифровых полей средствами системы ГеоБаза Данных

Таранчук В. Б.

В статье обсуждаются вопросы инструментального наполнения и использования интерактивной компьютерной системы ГеоБаза Данных. Представлены и обсуждаются результаты интеллектуальной обработки данных, включаемых и используемых в компьютерных геологических моделях.

Received 28.01.2020

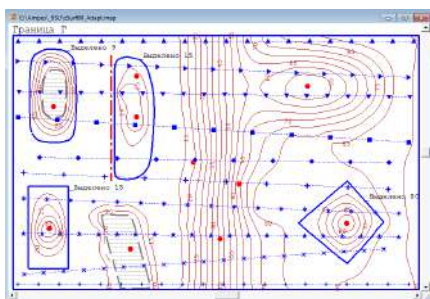


Figure 8. Result of Gen\_MAPw-adapted distribution.