Interactive Adaptation of Digital Fields in the System GeoBazaDannych

Valery B. Taranchuk

Belarusian State University, 4, Nezavisimosti Avenue,
220030 Minsk, Republic of Belarus
taranchuk@bsu.by

Abstract. The interactive computer system GeoBazaDannych is a complex of intelligent computer subsystems, mathematical, algorithmic and software tools for filling, maintaining and visualizing input and output data, creating continuously updated computer models. The purpose and functionality of the main components of the system GeoBazaDannych are briefly described. Examples of interactive formation of digital models of geological objects in computational experiments that meet the intuitive requirements of the expert are discussed. Methodological and algorithmic solutions, corresponding special tools of the system GeoBazaDannych for the formation of digital distributions that meet the requirements set by the expert are noted. The results of comparison with standard solutions in the complex "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

1 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 [1,2]. 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

Supported by BSU Research Institute for Applied Mathematics and Informatics.

V. Golenkov et al. (Eds.): OSTIS 2020, CCIS 1282, pp. 1–12, 2020.
https://doi.org/10.1007/978-3-030-60447-9_14

taking into account continuously incoming additional data, as well as a revision of the results of processing the initial information using new interpretation methods [2,3].

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 [4].

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 [5]), 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 [6,7]. At the same time, following [6], "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 [6]. 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 [8].

A number of issues related to the analysis and evaluation of spatial data quality can be solved using the computer system GeoBazaDannych [9–11]. 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 [9–13]:

- 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 geoecological models (GGMD).

2 Examples of Interactive Adaptation of Digital Fields

Integration of the capabilities of various geographic information systems (GIS) and the GeoBazaDannych is provided by a wide range of tools of the system for importing and exporting data, images, and functions. This article discusses several non-standard solutions that are recognized as difficult for all geodata processing packages. 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 [14,15]. 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 [16]. To create computer models, you need to have tools for interactive data processing and implementation of possible situations of receiving and correcting input information, modules for mathematical processing, interpretation and statistical analysis [14,17]. For the compiler of any computer model, it is important to be able to perform data analysis using different methods and algorithms in several software environments, to have tools that allow him to "play" with the source data and compare the results with the prepared reference models. This assumes that there are tools for exchanging data between modules that use different formats. How these requirements are implemented in the system GeoBazaDannych is described 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 GGMD complex illustrate the unique capabilities of the developed methods and software. Using the tools of the GGMD complex [12,13,18], estimates of the accuracy of digital field reconstruction are obtained and illustrated with graphics.

The capabilities of the algorithms implemented in the system GeoBazaDannych for interactive preparation and adaptation of digital parameter distributions in relation to the tasks of composing computer models of geological objects are illustrated by examples using the results of [13,18]. At the same time, we will not consider typical graphic objects (pyramids, parallelepipeds, cylinders, cones). The reason is that three-dimensional geometric patterns can be processed separately in the system GeoBazaDannych. A special module of the system can be used to "restore surfaces when the digital field is not calculated using the approximation method on the selected sections, but is filled in with the calculated values of functions from the template library (by the generator of defining functions)" [11]. Also, to demonstrate the special capabilities of the system GeoBazaDannych, in addition to the perturbations of the type (2) considered in [13], perturbations simulating a plateau and a split are introduced. They are set by expressions (1)–(3):

$$fH(x,y) = cos(\pi x/2) + cos(\pi y/2),$$

$$fHill3(x,y) = \begin{cases} fH(x,y), -1 \le x \le 1 \cap -1 \le y \le 1 \cap fH(x,y) \le 1/4, \\ 1/4, -1 \le x \le 1 \cap -1 \le y \le 1 \cap fH(x,y) > 1/4, \end{cases} \quad (1)$$

$$fHill5(x,y) = \begin{cases} fH(x,y), -1 \le x \le 1 \cap -1 \le y \le 0, \\ 0, \end{cases} \quad (2)$$

$$fHill6(x,y) = \begin{cases} fH(x,y), -1 \le x \le 1 \cap -1 \le y \le 0 \cap -1 \le x+y \cap x+y > 0, \\ 0. \end{cases} \quad (3)$$

The reference surface in the following examples is objects described by the expression (4)

$$\begin{gathered} fOriginF(x,y) = zBasicF(x), \\ zSurfF(x,y) = fOriginF(x,y) + \\ +400 \cdot fHill6(0.005 \cdot (x-250), 0.007 \cdot (y-400)) + \\ +500 \cdot fHill3(0.01 \cdot (x-200), 0.01 \cdot (y-150)) - \\ -150 \cdot fHill(0.01 \cdot (x-880), 0.015 \cdot (y-100)) - \\ -150 \cdot fHill(0.02 \cdot (x-920), 0.004 \cdot (y-500)) + \\ +200 \cdot fHill5(0.008 \cdot (x-550), 0.01 \cdot (y-150)). \end{gathered} \quad (4)$$

Its important that in the resulting expression (4) coefficients in the formulas of perturbation elements are chosen by the user while visual construction. An example of creating a model of the reference surface, which is obtained from the base surface by adding elements of the listed types, is shown in Fig. 1, Fig. 2.

We emphasize attention on fragments of the zSurfF surface, namely, perturbations fHill3, fHill5, fHill6. The element *fHill3*(200,150) simulates the presence of a plateau (clipping with a horizontal plane), *fHill5*(550,150) and *fHill6*(250,400) simulate splits (in geology shift-overshoot), and is mathematically described by clipping a smooth hill-type shape with a vertical plane. Such fragments of fHill5 and fHill6 cannot be reproduced using standard methods for restoring digital fields. Also note that in the following illustrations, the isolines corresponding to the plateau level (for zSurfF level 225) are not displayed. In the zSurfF illustration, level 210 is shown to further emphasize the boundaries of disturbances placed on the plateau section.

The images zSurfF in Fig. 1 give an aggregate vision of the reference surface. Individual, details can be studied on the contour map Fig. 2. It displays the isolines and density distribution (color filling of intervals) of the digital field of the function *zSurfF*(x,y). In addition, color intervals are explained by the legend and value labels that match the levels of the isolines.

The illustrations below (Fig. 3–Fig. 7) show examples of reconstruction of the grid function using standard algorithms and application of possible solutions and methods for "adjustment" distributions in the system GeoBazaDannych. Typical steps for simulating observations and preparing initial data for the subsequent reconstruction of the digital field from a scattered set of points with measurements were performed beforehand. It is assumed that the measurement points

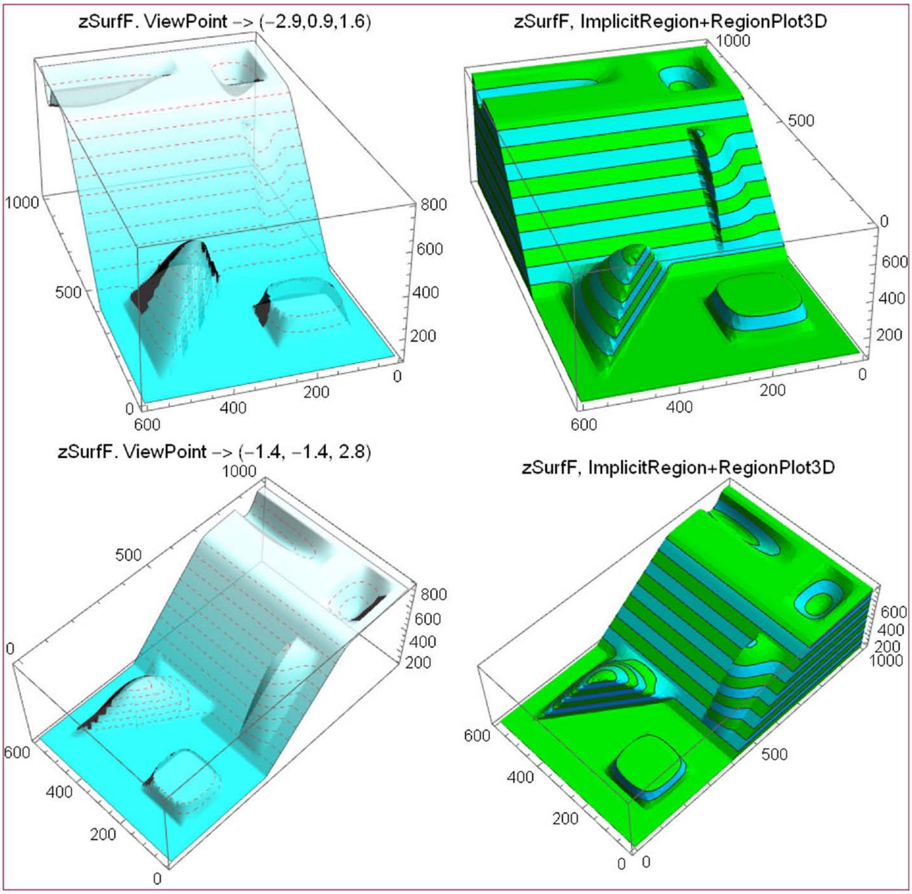

Fig. 1. ZsurfF surface and volume graphs.

are placed on profiles; they are shown for different observation profiles by different primitives (112 dots). The centers of form perturbations are marked with red circles, which are 6 points (Fig. 3). Two pairs of red diamond markers show split segments. There was no goal to choose a good system of profiles, they are formed sketchily, and about how often designers do this.

Figure 3 shows the results obtained in the complex GMMD of digital field reconstruction, when the simplest variant is implemented using the standard interpolation function of the 1st order in Wolfram Mathematica. Note that when working with irregularly placed data points in the system Mathematica, interpolation can only be performed using this method. But this is enough, since in other specialized packages, the most common, this problem is solved by the method of triangulation, which also corresponds to linear interpolation. The corresponding results obtained in the GMMD are shown on the contour map in Fig. 3, where the isolines of the reference field are given as solid purple lines, and in the restored field – as dashed black lines.

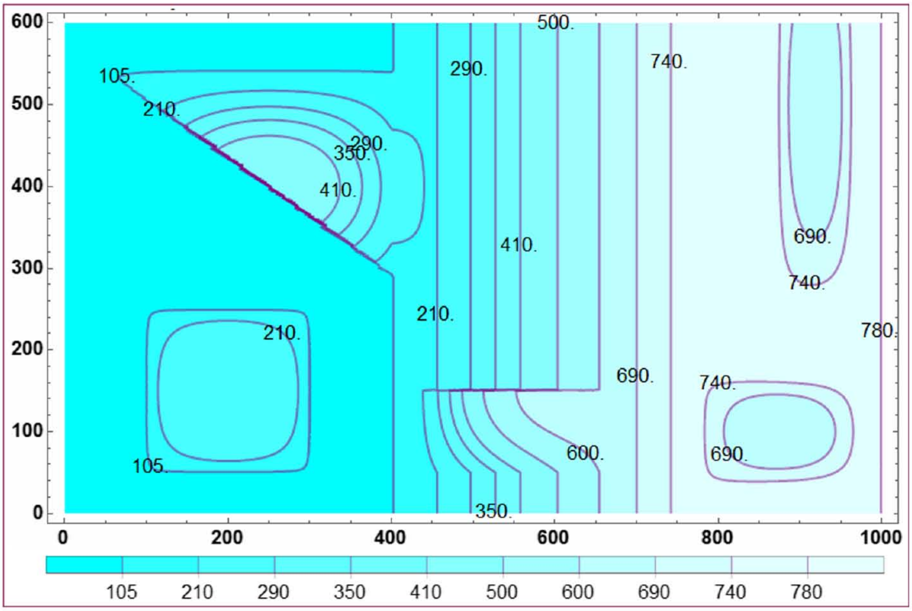

Fig. 2. Isolines and density maps of *zSurfF*(x,y) function.

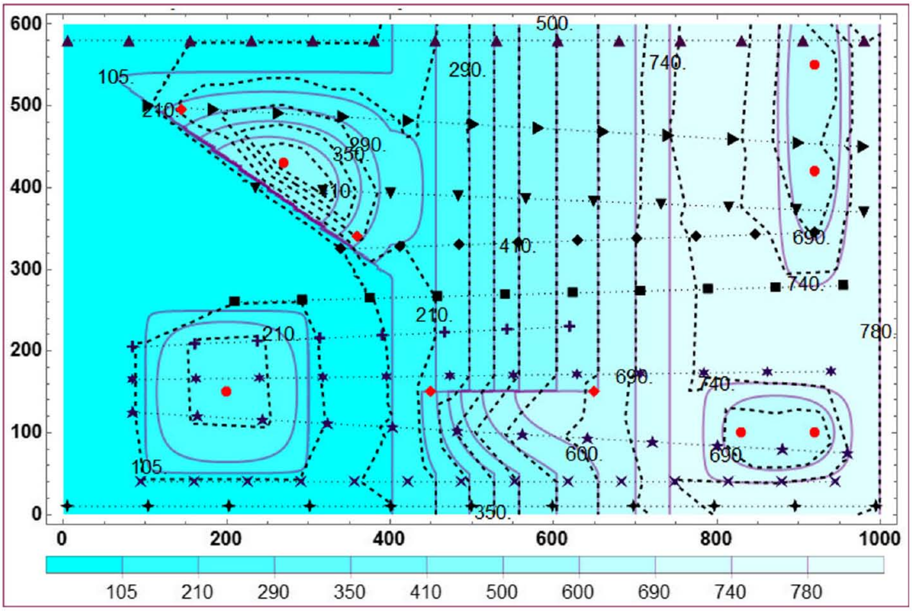

Fig. 3. Isolines of the digital fields of the reference (purple lines) and restored in the GMMD (dashed black lines) distributions. (Color figure online)

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. 4, 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 the integer numbers.

Figure 5–Fig. 7 show the results of restoring the digital field obtained in the system GeoBazaDannych. At the initial stage (to compare the results of restoring the digital field), calculations of the field reconstruction using the GeoBazaDannych's spline approximation method were performed in Gen_MAPw without any additions. The calculated grid was formed in the Gen_MAPw automatically taking into account the coverage of the entire area edged by the Granica G border, its step was set to 10. In this case, the GeoBazaDannych "subdomain" element (Granica G) "works". The results are shown on the contour map Fig. 5, the corresponding isolines are displayed as dash-dotted green polygons. The dashed isolines of the graphic layer from Fig. 4 is also added to the illustration. There is an understandable difference, because the results are calculated using different interpolation methods. Basically, both methods provide visualization of fHill perturbation forms, but with significant reproduction errors for fHill5 and fHill6.

How to understand the results shown in Fig. 6, Fig. 7? In these cases, the task is to correct the digital fields obtained by the approximation algorithm via using such GeoBazaDannych's elements as "selected", "split", and "boundary conditions" [11]. In fact, it is an interactive intellectual adaptation.

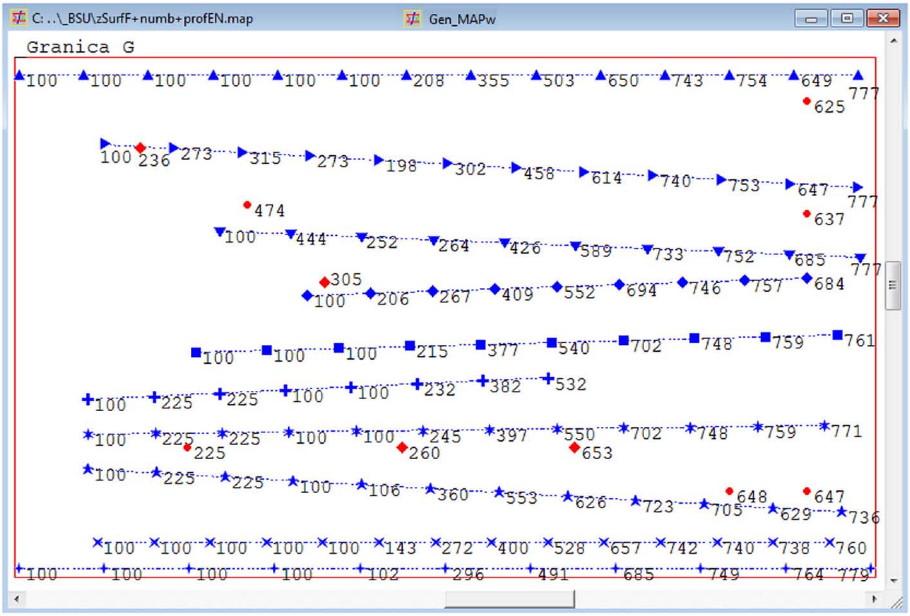

Fig. 4. Results of exporting points with measurements from GGMD to Gen_MAPw.

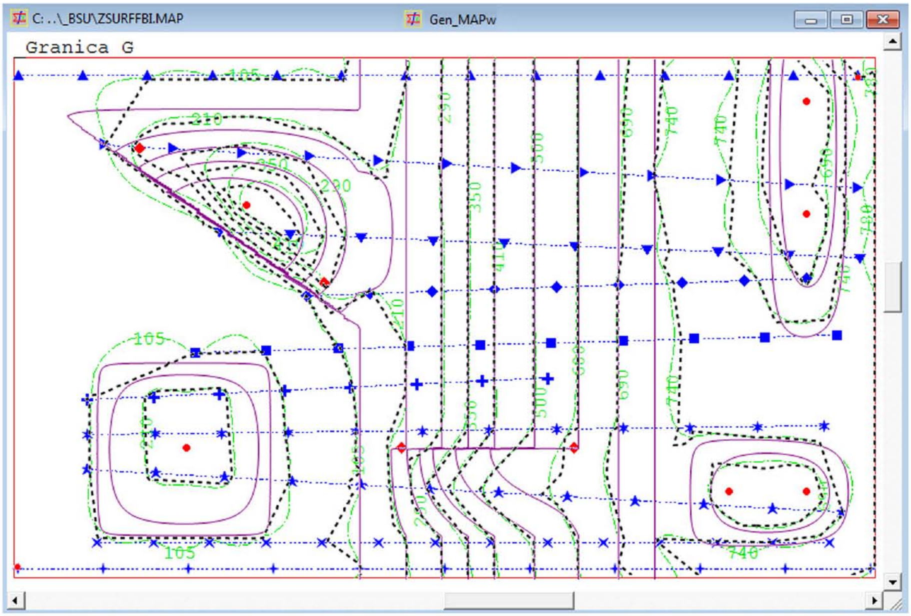

Fig. 5. Isolines of reference and restored fields in GGMD (dashed black lines) and in Gen_MAPw (spline approximation algorithm, dash-dotted green lines). (Color figure online)

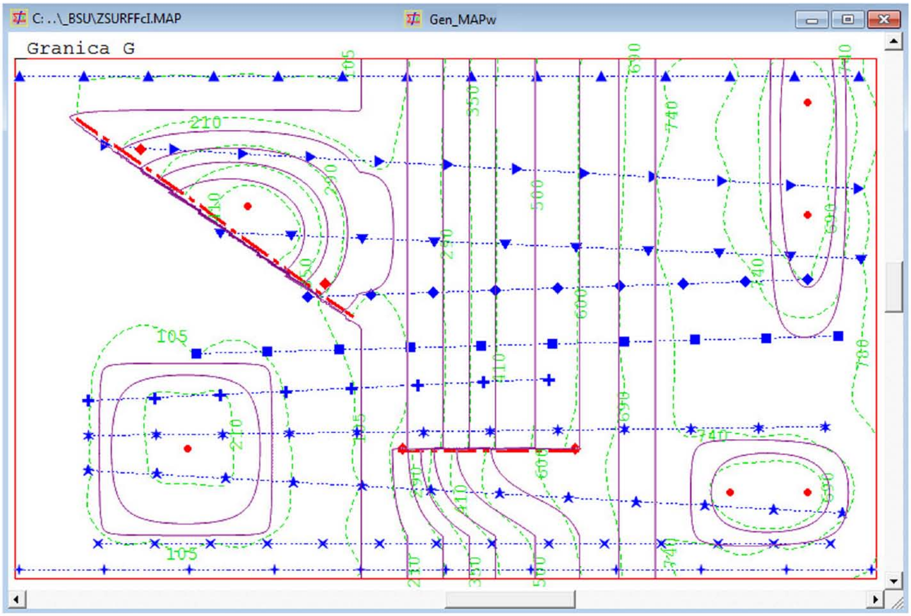

Fig. 6. Effects of correction of the reconstructed digital field using the "split" method. (Color figure online)

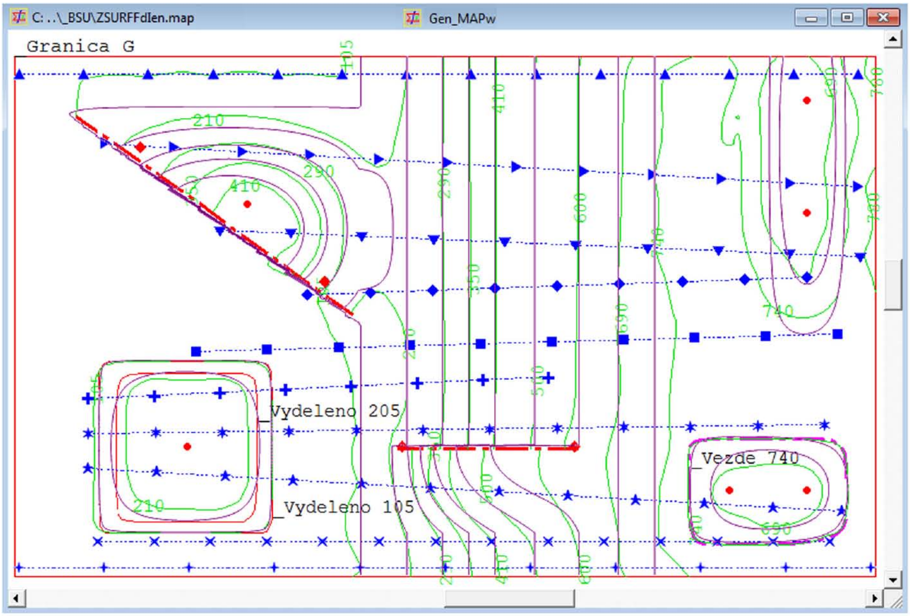

Fig. 7. Effects of correction of the reconstructed digital field using the "selected", and "boundary conditions" methods. (Color figure online)

An example of adapting digital fields in Gen_MAPw using the "split" element of the GeoBazaDannych is illustrated in Fig. 6. When executing the approximation (forming a digital field on the grid), in addition to the Granica G, 2 contours were added to the input map – bold dash-dotted red lines. In [11] it is explained that the system GeoBazaDannych "recognizes" special contours (in particular, subdomains, inclusions, boundary conditions, splits) by line attributes, which include: color, thickness, line type, closed or not. The isolines in this version (with the addition of "split" category contours to the input data) are shown as dotted green lines. The illustration Fig. 6 expressively demonstrates the effect of the of "split" element in the lower central part, in particular, the levels 290, 350, 410, 500, 600.

Figure 7 illustrates the possibilities of additional refinements of the digital field that should correspond to the user's a priori opinion using the "Selected" and "Boundary conditions" GeoBazaDannych's tools. The "Selected" method allows you to perform a separate (autonomous) calculation of grid values within a user-defined subdomain, taking into account only the points with measurements that are included in the subdomain. When approximating outside the subdomain, measurements that fall within the subdomain are not taken into account. Such a subdomain is enough to sketch a closed non-self-intersecting polyline on the area of the function definition, and set the appropriate attribute for this contour. In addition, in the Gen_MAPw, the user can "assign" a numeric value to the polyline nodes, which will be taken into account when calculating the grid function. In fact, in this case, the original set of values is supplemented with measurements. The grid function (digital field) adopted by the approximation algorithm will be gradually independently calculated in the subdomain and outside it separately, in each only by "their" reference points with values. The "Boundary conditions" method should be used in the part where it is necessary to "correct" the digital field obtained during approximation. The user draws a polyline with the corresponding attributes in the definition area and "assigns" values, one or more (the interpolation algorithm for setting multiple values is described in [11]). The reference nodes of this contour "go" as an addition to the set of values with measurements.

In the calculation variant illustrated in Fig. 7, the "Selected" and "Boundary conditions" methods are used. In the illustration in this version, the isolines are shown as green solid lines, unlike the dot dashed and dashed lines in the previous versions. In the lower-left part of the grid function definition area, the conditions "Vydeleno_105" and "Vydeleno_205" are used. The corresponding contour borders are drawn on the input map with red solid thin lines. Comparing the isolines of Fig. 6 and Fig. 7, the effect of the applied method of digital field adaptation is clearly positive. You can also clearly see the positive effect of applying the boundary condition in the lower-right part of the definition area. The effect is provided by the introduction of a contour border with a value of 740 (a dash-dotted bold line of magenta color).

It should be noted that the 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, the 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 [13]. We do not give particular error digits here, because in the model problem under consideration, the eliminated errors on the adapted field are comparable to the accuracy of obtaining grid functions, although these errors are noticeably lower than in the unadapted field.

3 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. Bedrikovetsky, P.: Mathematical Theory of Oil and Gas Recovery: With Applications to ex-USSR Oil and Gas. Springer, Heidelberg (2013)
2. Dmitrievskii, A.N., Eremin, N.A.: Bolshie geodannye v cifrovoi neftegazovoi ekosisteme. Energeticheskaya politika **2**, 31–39 (2018). (in Russian)
3. 13th International Symposium on Software Engineering for Adaptive and Self-Managing Systems. Gothenburg, Sweden, 28–29 May 2018. https://conf.researchr.org/track/seams-2018/seams-2018-papers. Accessed 29 Jun 2020
4. Savinyh, V.P., Cvetkov, V.Ya.: Geodannye kak sistemnyi informacionnyi resurs. Vestnik Rossiiskoi akademii nauk **84**(9), 826–829 (2014). (in Russian)
5. Tsvetkov, V.Ya., Lobanov, A.A.: Big Data as Information Barrier. Eur. Res. **78**(7–1), 1237–1242 (2014)
6. Shaitura, S.V.: Intellektual'nyi analiz geodannyh. Perspektivy nauki i obrazovaniya **18**(6), 24–30 (2015). (in Russian)
7. Golenkov, V.V.: Ontology-based design of intelligent system. In: Golenkov, V.V. (ed.) Conference 2017, OSTIIS, pp. 37–56. BSUIR, Minsk (2017)
8. Dyshlenko, S.G.: Analiz i razrabotka harakteristik kachestva geodannyh. Perspektivy nauki i (2016)
9. Taranchuk, V.B.: Chislennoe modelirovanie processov dvuhfaznoi mnogokomponentnoi fil'tracii v neodnorodnoi poristoi srede. Doktorskaya dissertaciya, Kazan' State University (1992). (in Russian)
10. Barvenov, S.A., Kibash, M.F., Taranchuk, V.B.: Metodika, instrumentarii adaptacii matematicheskih modelei processov podzemnoi gidrodinamiki. Vybranyya navukovyya pracy BDU, Matematika, 34–65 (2001). (in Russian)
11. Taranchuk, V.B.: Postroenie cifrovyh geologo-ekologicheskih modelei v sisteme GBD-e. Geoekologiya Belarusi, Pravo i ekonomika, 72–96 (2006). (in Russian)
12. Taranchuk, V.: The integrated computer complex of an estimation and adapting of digital geological models. Studia i Materialy **14**(2), 73–86 (2017)
13. Taranchuk, V.B.: Metody i instrumentarii ocenki tochnosti komp'yuternyh geologicheskih modelei Vestnik BrGTU (2), 64–69 (2018). (in Russian)
14. Davis, J.C.: Statistics and Data Analysis in Geology. Wiley, New York (1986)
15. Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P.: Design and analysis of computer experiments. Stat. Sci. **4**, 409–435 (1989)
16. Halimov, E.M.: Detal'nye geologicheskie modeli i trehmernoe modelirovanie. Neftegazovaya geologiya. Teoriya i praktika **7**(3), 1–10 (2012)
17. Wolfram Mathematica. http://www.wolfram.com/mathematica/. Accessed 29 Jun 2020
18. Taranchuk, V.B.: New computer technologies, analysis and interpretation of geodata. In: MATEC Web of Conferences IPICSE-2018, vol. 251, pp. 1–8 (2018). https://doi.org/10.1051/matecconf/201825104059

Metadata of the chapter that will be visualized in SpringerLink

Book Title	Open Semantic Technologies for Intelligent System	
Series Title		
Chapter Title	Interactive Adaptation of Digital Fields in the System GeoBazaDannych	
Copyright Year	2020	
Copyright HolderName	Springer Nature Switzerland AG	

Corresponding Author	Family Name	**Taranchuk**
	Particle	
	Given Name	**Valery B.**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	Belarusian State University
	Address	4, Nezavisimosti Avenue, 220030, Minsk, Republic of Belarus
	Email	taranchuk@bsu.by
	ORCID	http://orcid.org/0000-0003-2210-652X

Abstract The interactive computer system GeoBazaDannych is a complex of intelligent computer subsystems, mathematical, algorithmic and software tools for filling, maintaining and visualizing input and output data, creating continuously updated computer models. The purpose and functionality of the main components of the system GeoBazaDannych are briefly described. Examples of interactive formation of digital models of geological objects in computational experiments that meet the intuitive requirements of the expert are discussed. Methodological and algorithmic solutions, corresponding special tools of the system GeoBazaDannych for the formation of digital distributions that meet the requirements set by the expert are noted. The results of comparison with standard solutions in the complex “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