# Examples of the use of artificial neural networks in the analysis of geodata

Valery B. Taranchuk Belarusian State University Minsk, Republic of Belarus taranchuk@bsu.by

Abstract—The article discusses the problems of development, tool filling, and usages of the integrated program complex of the composer of digital geoecological models. Possibilities of interactive graphics visualization, and comparison of results are marked. The results of application of artificial neural networks in the analysis and interpretation of geospatial data are presented and discussed.

*Keywords*—digital geoecological model, approximating digital field, computer algebra system Mathematica, geographic information system Surfer, interactive graphics visualization, artificial neural network

# I. INTRODUCTION

The construction of the digital geoecological, geological models at present is a compulsory part of expertise to many areas. Geological simulation is an independent stream, which includes the progressing of mathematical methods and algorithms; development of computer programs, which provide the cycle of models' construction, database creation, provisioning and maintenance. The corresponding geo-environmental information products include 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. There are a lot of software tools for this purpose. As a rule, the corresponding software packages are focused on solving a specific class of problems. One of the actively developing new directions in Geology and Geoecology is the improvement of computer systems based on the integration of powerful tools of spatial and temporal data processing offered by GIS with models of artificial intelligence of neural networks [1] - [4]. Different applications are created that allow not only to implement a huge number of algorithms for data processing, analysis and visualization, but also provide opportunities to obtain new results.

On the other hand, it can be noted, that while general issues are being resolved and there is no proper elaboration of details. Current achievements and tools of intellectual (symbolic) calculations are not fully used, a number of performed products are not justified by the theory of numerical methods. Most of the available packages lack tools to assess the accuracy of the results. Any extension of functionality requires the involvement

of highly qualified programmers. Therefore, an important direction is the development of software systems based on the combination and integration into a single environment of computer algebra systems (CAS) and geographic information systems (GIS). Moreover it should be considered that to solve the problem of processing initial data there's no specific GIS to be the full set of space-analytic methods and analysis tools. In many cases it's necessary to combine the tools provided by GIS with programs for static data analysis, tools for mathematically complex computations which include implementations of modern methods and algorithms of analysis and interpretation of spatial data. A number of methodological and technical solutions aimed at overcoming the mentioned difficulties are proposed and implemented in the integrated computer system GGMD - "The generator of the geological model of deposit" [5] – [7].

### II. THE DEVELOPMENT PLATFORM OF GGMD

GGMD is assigned for creation and estimated accuracy of configurable geological model based on the usage of CAS and GIS, "smart" methods of model adaptation while in service, "self-tuning" of models considering additional data from the actual development of processes. Development platform is computer algebra system Mathematica [8], language is Wolfram Language [9], geographical information system is Golden Software Surfer [10]. While programming in Wolfram Language technical solutions, described in [11], were implemented, moreover software system in a particular configuration can be used after it's built and saved in computable document format [12]. Calculations, user work with CDF version of application are possible on every personal computer. When viewing CDF version, hosted on webserver, viewer is automatically loaded in the form of browser plugin. Offline work is possible after the installation of free distributed CDF Player.

Let's make some clarifications for the illustrations of implementations marked above, the results of the usage of program modules which are selected and modified for the problems we're solving. Below we mention the components, which are actually standalone program modules. They can be also considered as parts of automated workstation of specialist, who during the computational experiments works out techniques of adaptation of digital fields. We should specifically note an important technical solution – all the work steps with the complex of modules are provided with the possibility of import and export of obtained result with several configurations of output format [13]. It provides the user with additional possibilities for performing similar calculations in different (including the others) applications, the comparison of results.

It is necessary to understand that creation and maintenance of geoecological model don't expect to have unique solution to a mathematical problem. Subjective opinion, the qualification of an expert – are the factors that always take place in such activity. While working a user has to operate with data of different accuracy, some initial data is even conflicting; data density with measurements differs on different parts. That's why for construction digital models it's important to have tools for interactive data processing, simulation of possible situations of receiving and correction of input data. All the steps of working with data in GGMD include various options of graphic visualization [14], logging and comparison of incoming and placed to archive results. Complex's tools give a user possibility to "play" with initial data and compare the results with prepared etalons, what is more it's allowed to import and export the data and images and to scale them. Extensive data exchange possibilities are important for simultaneous work in several software environments.

### III. THE COMPONENTS OF GGMD

In computer system GGMD the following tools are implemented:

- tools and patterns for preparation of reference model of digital field, which corresponds to the specified properties ("Digital field constructor");
- tools and several options of "distortion" of reference model;
- tools for data capture simulation, which are used in simulation practice ("Generator of profile observer");
- modules for calculation, visualization, comparison of digital fields approximation by several different methods ("Approximation component");
- tools and adaptation modules for digital model being formed ("Adaptation component").

The main idea and purpose of the development of this computer system is to choose the method of processing the original data by comparing the reference digital field and reconstructed by "observations". The reference distribution for a rectangular area is formed using mathematical descriptions. Each expert determines and includes own typical fragments in the model. Then the instruments of the GGMD perform "observations", simulate the measurement of the reference distribution, and the geometry of the measurement points and their accuracy are also determined by the user of the system and should approximately correspond to the initial data of the of the field of interest. The layout of points with measurements should not be regular. As a result of this stage, the user receives a set of data "observations", the main ones being the coordinates of the point and the value in it. The next step is to select the algorithm for processing the resulting set ("reproducing" the digital field) by performing interpolation and extrapolation. Comparison of the results of "playback" and the reference model will prompt the expert the method of processing, the geometry of the observation points.

# A. Digital field constructor (DFC). Base surface formation

Software components from this group provide in interactive mode the construction of the model's surface from standard elements with accompanied visualization of mathematical description (analytic function), model's surface is interpreted as a relief - set of surface shapes. The construction is made in the module which is programmed in system Mathematica and includes the generation of surface equation - function of two arguments x and y which is continuous (or piecewise continuous) and defined in the rectangle. User defines the boundaries of domain xMin and xMax, yMin and yMax and surface height limits zMin and zMax. Let's mark out that all the notations in DFC are given in format of InputForm (string format), that is accepted specially as some users can use application written in Excel, Delphi, C or others, where mathematics notation isn't supported inside the program code. There are mathematical expressions (elements) which allow for the user of the GGMD to reproduce the behavior of the areas, which are typical for relief, in the set (library) of components of the function being formed. User at the first stage of reference model formation sets up a piecewise-defined function zBasic(x). In the terminology of the complex, this is the basic profile - tape of specified width and length, which imitates the types of relief with the elements of plateau, slope, cliff.

Different variants of determining the basic profile of the base surface are given in [5] - [7]. We note that the user of the system obtains an analytic expression. Analytical expressions (1) are used in the preparation of the following examples:

zBasicQ(x) =	
$If[xMin \le x \le xOtkQ1, fPltQ(x,0), 0] +$	
$+If[xOtkQ1 < x \le xOtkQ2, fOtkQ(x, 0), 0] +$	(1)
$+If[xOtkQ2 < x \le xSklQ2, fPrhQ(x,0), 0] +$	
$+If[xSklQ2 \le x \le xMax, fSklQ(x,0), 0],$	

(x - xSklQ2) + zSklQ2, zSklB = fPrhB(xSklB2, 0),zOtkB2 = fOtkB(xOtkB2, 0).

Further 2 surfaces are considered. For the upper instead of the letter Q should substitute T in expression (1), for the lower – B. xOtkQ1, xOtkQ2 – are coordinates of transition points "plateau – slope", "slope – slight slope", ugOtkQ, ugSklQ defines the inclines of slope, slight slope, zPlt=10. The constants in expression (1): xOtkB1 = 40, xOtkB2 = 70, xSklB2 = 75, zPltB = 10, ugOtkB = 1.1, ugSklB = 0.15, xOtkT1 = 30, xOtkT2 = 45, xSklT2 = 66, zPltT = 10, ugOtkT = 1.2, ugSklT = 0.08. (see Fig. 1).

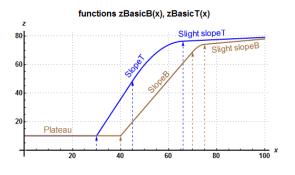


Figure 1. Plot of basic profiles.

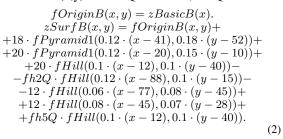
In the given example the base surface model is quasi three dimensional (the level of z doesn't depend on y). Basic surface (tape) is made up of 3 typical sectors: flat horizontal (plateau), flat with fast rising level (slope), flat with slow rising level (slight slope). The connection between the sectors is continuous. Transition "plateau slope" is made at a selected angle, transition "slope slight slope" is smooth.

Then user can add perturbations of different shapes, sizes and orientations to the base surface. The construction with DFC module of basic profile from the fragments is possible with continuous transition "fragment – added fragment", smooth transition, a jump (imitation of split).

### B. Digital field constructor. Reference surface formation

The next step of construction with DFC module is usage of program module's tools to add perturbations, fragments of typical elements of relief to a basic profile. Template (patterns) library includes elements which correspond to perturbations (areas of distortion of basic surface) of different geometrical shape. While connecting the patterns it's possible to set interactively their position and size. Described mathematical elements, which imitate the following shapes of relief: hill, embankment, pit, excavation, trench, canal, quarry, ravine, hollow (vug), are included in basic package ([5], [6]). It should be noted that all the elements listed above are specified by analytic expressions, such as z = fFrgm(x, y). Besides it, those are written for the square [-1,1]×[-1,1], and then in the final function the arguments are scaled. System Mathematica includes a big amount of spatial graphic primitives of which cone, ball, cuboid, cylinder are used in DFC module, also different pyramids are included to the library. Several variants of determining of the reference surface are given in [6], [7].

An example of reference surface model formation, which is obtained from the basic surface, by adding elements of the listed types (2 pyramids, 5 hills) is shown below on Fig. 2. It's important that in the resulting equation (2) the coefficients in the formulas of perturbation elements fHill, fPyramid are chosen by user while visual construction. The constants in expression (2) for zSurfB(x,y) are fh2Q=fh2B=10, fh5Q=fh5B=20.



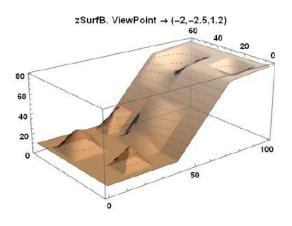


Figure 2. Plot of surface zSurfB.

An example of reference surface model 2 is shown below on Fig. 3. The constants in expression (2) for zSurfT(x,y) are fh2Q=fh2T=8, fh5Q=fh5T=26.

When programming the module DFC, the following functions of the system Mathematica were used: Solve, Piecewise, Cases, Table, Plot, Plot3D, Manipulate. While visual examinations of the plots in DFC module, user have a possibility to define the coefficients of function expression by moving sliders or setting specific values on the panels which are the part of the interface of Manipulate function (more in [14], [11]). At this stage of constructing of the surface, the user of the DFC module also obtains an analytic expression.

# C. Generator of profile observer

More than ten options of 1D, 2D and 3D plots were implemented in GGMD system, including modules for

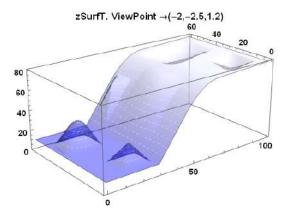


Figure 3. Plot of surface zSurfT.

obtaining and designing maps and inserts on them, plots on profiles, 3D visualizations of possible incisions (vertical, horizontal), clipping (simple and complex). The illustrations of the usage of several visualization tools are given in [5], [7].

Let's consider the examples of obtaining and configuring of graphs of geological profiles with using of the tools of "Generator of profile observer" (GPO) – the illustrations are given on Fig. 4, Fig. 5 for surfaces zSurfT, zSurfB on direction (0,45) - (100,10).

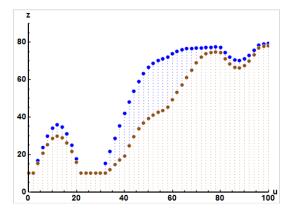


Figure 4. Example of indicating of profiles, graphics of cross-sections.

While solving the problems of the geological simulation, profiles are used to visualize the connection between relief and the structure of Earth's crust. Geological profile is a graphic image in vertical plain of subsoil structure and deposits contained in it. In our case the profile – is a line, which is obtained at the intersection of analyzed surface and vertical surface in a given direction.

We will consider the use of GPO tools to show several methods of simulating observations and obtaining initial data by instruments of the GGMD's module of "distortion" on the example of profiles on reference surfaces.

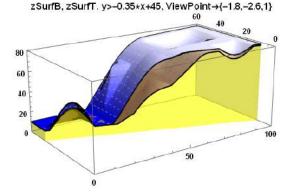


Figure 5. Example of the simple incisions of surfaces zSurfB, zSurfT.

# IV. EXAMPLES OF ANALYSIS OF GEODATA BY TOOLS OF NEURAL NETWORKS

The problems of construction of geological models certainly belong to the class of complex ones. And for the solution of complex problems it is required to provide compatibility and integration of the different models of data representation and knowledge processing algorithms. Approaches may be different, some are considered in [15]. Wolfram Mathematica, as a system of intelligent computing, provides the user with not only the means of mathematical transformations, accurate and approximate calculations, but also the tools of machine learning. They may be used in the interpretation and processing of input data and simulation results. The results and examples of the analysis of geoecological data by means of neural networks are obtained using the corresponding functions, which are available in version 11 of CAS Mathematica [8], [16]. The results below are illustrated by examples of interpretation and processing of data for profiles of zSurfB, zSurfT surfaces.

### A. Preparation of data for numerical experiments

Let's consider the examples of obtaining and configuring of graphs of profiles with using of the tools of "Generator of profile observer". Figure Fig. 6 shows the graphs of the levels of surfaces zSurfT, zSurfB at the profile on direction (0,45) - (100,10). Symbols on the graphs (filled circles and triangles) are marks of surface level values at profile points (regular grid, constant step). These values are calculated using formulas from surface equations. Circles and triangles indicate values that simulate measurements. For the surface zSurfT they are obtained by adding "noise" using generator RandomVariate[NormalDistribution[0,6]], such generator gives a pseudorandom variate from the symbolic distribution dist. NormalDistribution[ $\mu,\sigma$ ] represents a Gaussian distribution with mean  $\mu$  and standard deviation  $\sigma$ ). For the surface zSurfB distortion obtained by adding "noise" using the generator RandomReal[-10,10]

on the entire profile, RandomReal[ $x_{min}, x_{max}$ ] gives a pseudorandom real number in the range  $x_{min}, x_{max}$ .

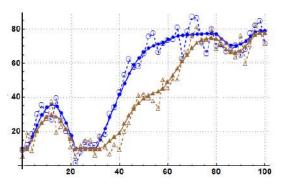


Figure 6. Data for numerical experiments.

## B. Results of application of artificial neural networks

Note that we will process data with "distortions" for two completely different purposes, respectively, different tools of neural networks will be used.

In solving problems of mathematical modeling, the original equations are written in differential form, so the original data (tooling of model) must be continuous, moreover, as a rule, the distribution should be smooth functions (for example, see [17]). In other words, the distribution of the observed parameter along the profile should be transferred to the initial data of the computer model as a smooth function. Note that the original data of figure Fig. 6 the requirement of smoothness is responsible, and data from the "distortions" is unsuitable for numerical models, this baseline data will immediately lead to "bumpy" decisions (computational instability) and the results will be unusable.

Examples of obtaining a "smooth" profile.

The results of obtaining in the module "Adaptation" of GGMD complex of smoothed profiles are presented in figures Fig. 7, Fig. 8.

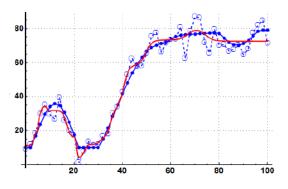


Figure 7. Smoothed profile for zSurfT. Method - RMSProp.

The functions and the options of the neural network for Fig. 7 are next: netA = NetChain[vectLength, Tanh,

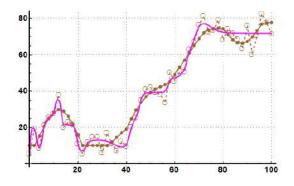


Figure 8. Smoothed profile for zSurfB. Method - ADAM.

vectLength, Tanh, 1, ...]; netA1 = NetTrain[netA, dataProfT1, Method – "RMSProp"], vectLength = 300. Method RMSProp is stochastic gradient descent using an adaptive learning rate derived from exponentially smoothed average of gradient magnitude [18].

For Fig. 8 – vectLength = 25, Method – ADAM, stochastic gradient descent using an adaptive learning rate that is invariant to diagonal rescaling of the gradients.

The results of the neural network application for the tasks to be solved in the interpretation and analysis of geodata.

Figures Fig. 9 and Fig. 10 show the results of using a neural network for a completely different kind of problem – problems that need to be solved in the interpretation and analysis of observational data, measurements of parameters ( (for example, see [19])).

Such tasks can be explained in simple words – it is known that the data to be analyzed contain measurement errors, including system and random ones. System errors are "separated" taking into account the knowledge of their nature and behavior, and random errors are decided by experts.

In this consideration, the task is to train the system to "clean" the data from random noise (extract and discard). Since we know how the data illustrated on the graph were prepared, we can interpret the original set as a standard, but the program (algorithm, executive software module) does not "know" the original.

The illustrations show two different approaches when setting up a neural network. In the variant of Fig. 9 the following condition is accepted – the processed data set contains 2 parts, one of which is the main distribution, and the other is random noise. Information about the nature of the noise in the rules of training is not given. The results of the calculations in figure Fig. 9 are obtained with the following parameters: dataAT2=RandomSample[dataProfT2]; vectLength=300; netAT2 = NetChain[vectLength, Tanh, vectLength, Tanh,1,..]; {trainAT2,testAT2} = Take-Drop[dataAT2,36]; netAT2 = NetTrain[netAT2, trainAT2,

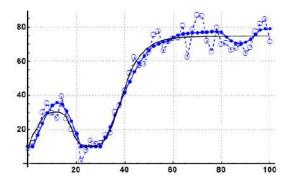


Figure 9. Example 1, zSurfT - "clean" the data from random noise.

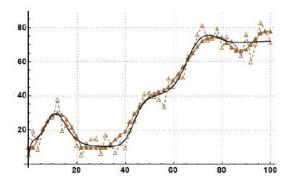


Figure 10. Example 2, zSurfB - "clean" the data from random noise.

# ValidationSet $\rightarrow$ testAT2] ( [20], [21]).

Figure Fig. 10 shows the results of data processing for profile on zSurfB. The figure illustrates the comparison of the source and the result when the program "cleared" the data from random noise. Other functions of neural network configuration are used: NetTrain[netAB1, dataProfB2, Method $\rightarrow$ "ADAM", "L2Regularization" $\rightarrow$ 0.1].

### V. CONCLUSION

The article discusses the problems of development, tool filling, and usages of the integrated program complex of the composer of digital geoecological models. Possibilities of interactive graphics visualization are marked. The presented results, examples of processing and visualization of spatial data, the noted methods of tuning of artificial neural networks tools are confirmation of the wide possibilities of the considered technology. On the other hand, the presented methodological and technical solutions for the generation and filling of a specialized software complex indicate the need to create intelligent medium of automatic connection of computer tools, the formation of a knowledge base with sets of reference and standard examples.

### REFERENCES

[1] Bryan C Pijanowski, Daniel G Brown, Bradley A Shellito, Gaurav A Manik, "Using neural networks and GIS to forecast land use changes: a Land Transformation Model", Computers, Environment and Urban Systems, vol. 26, Issue 6, November 2002, pp. 553–575.

- [2] Vincenzo Barrile, Giuseppe M.Meduri, Giuliana Bilotta, Ugo Monardi Trungadi, "GPS- GIS and Neural Networks for Monitoring Control, Cataloging the Prediction and Prevention in Tectonically Active Areas", Procedia - Social and Behavioral Sciences, vol. 223, 10 June 2016, pp. 909–914.
- [3] SONG Lirong, ZHAO Shiwei, LIAO Weilin, WANG Zhaoli, "Neural Network Application Based on GIS and Matlab to Evaluation of Flood Risk," International Conference on Remote Sensing, Environment and Transportation Engineering (RSETE 2013), pp. 296–299.
- [4] Alper Şen, M. Ümit Gümüşay, Aktül Kavas, and Umut Bulucu, "Programming an Artificial Neural Network Tool for Spatial Interpolation in GIS - A Case Study for Indoor Radio Wave Propagation of WLAN," Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3705544/ (accessed 2018, Dec).
- [5] V. Taranchuk, "The integrated program complex of the composer of geological models. The concept, solutions", Computer Algebra Systems in Teaching and Research, vol. VI, 2017, pp. 186–194.
- [6] V. Taranchuk, Viktoryia Taranchuk, "The integrated program complex of the composer of geological models. Examples of results", Computer Algebra Systems in Teaching and Research, vol. VI, 2017, pp. 195–203.
- [7] V. Taranchuk, "The integrated computer complex of an estimation and adapting of digital geological models", Studia i Materiały, N 2 (14), 2017, pp. 73–86.
- [8] WOLFRAM MATHEMATICA. Available at: http://www.wolfram.com /mathematica/ (accessed 2018, Dec).
- S. Wolfram, An Elementary Introduction to the Wolfram Language. Available at: http://www.wolfram.com/language/elementary-introduction/2nded/ (accessed 2018, Dec).
- [10] Surfer. Explore the depths of your data. Available at: https: //www.goldensoftware.com/products/surfer/ (accessed 2018, Dec).
- [11] Taranchuk V.B. "Osobennosti funkcional'nogo programmirovaniya interaktivnyh graficheskih prilojenii", Vestnik Samarskogo gosudarstvennogo universiteta. Estestvennonauchnaya seriya, razdel Matematika, N 6 (128), 2015, pp. 178–189, (in Russian).
- [12] CDF. Documents Come Alive with the Power of Computation. Available at: https://www.wolfram.com/cdf/ (accessed 2018, Dec).
- [13] Importing and Exporting Data. Available at: https://reference.wolfram.com/language/tutorial/ImportingAndExportingData.html/ (accessed 2018, Dec).
- [14] Annotating & Combining Graphics. Available at: https://reference.wolfram.com/language/guide/AnnotatingAndCombiningGraphics.html/ (accessed 2018, Dec).
- [15] V. Golenkov, N. Guliakina, I. Davydenko, D. Shunkevich, "Semantic Model of Knowledge Bases Representation and Processing", [Electronic resource] // CEUR Workshop Proceedings. Moscow, 2017. Available at: http://ceur-ws.org/Vol-2022/paper51.pdf (accessed 2018, Dec).
- [16] Neural Networks. Available at: https://reference.wolfram.com/language/ guide/NeuralNetworks.html (accessed 2018, Dec).
- [17] A. Chichurin, H. Shvychkina "Computer simulation of two chemostat models for one nutrient resource", Mathematical Biosciences, 278, 2016, pp. 30–36.
- [18] NetTrain. Available at: https://reference.wolfram.com/language/ref/ Net-Train.html (accessed 2018, Dec).
- [19] V. Orlov, E. Detina, O. Kovalchuk "Mathematical modeling of emergency situations at objects of production and gas transportation", MATEC Web of Conferences IPICSE-2018. V. 251, 04059.
- [20] TakeDrop. Available at: https://reference.wolfram.com/language/ref/ Take-Drop.html (accessed 2018, Dec).
- [21] ValidationSet. Available at: https://reference.wolfram.com/language/ref/ ValidationSet.html (accessed 2018, Dec).

# ПРИМЕРЫ ИСПОЛЬЗОВАНИЯ НЕЙРОННЫХ СЕТЕЙ В АНАЛИЗЕ ГЕОДАННЫХ

### В. Б. Таранчук

В статье обсуждаются вопросы разработки, инструментального наполнения и использования интегрированного программного комплекса тестировщика цифровых геоэкологических моделей. Отмечены разные возможности интерактивной графической визуализации и сопоставления результатов. Представлены и обсуждаются результаты применения искусственных нейронных сетей в анализе и интерпретации геоданных.