Journal homepage: https://mining-journal.com.ua/en

MINING JOURNAL
of Kryvyi Rih National University

Vol. 59, No. 1, 2025

Received: 15.02.2025; Revised: 27.05.2025; Accepted: 30.06.2025

UDC 004.4:622.142:622.271

DOI: 10.31721/2306-5435-2025-1-10-21

Oleksandr Zelensky*

Doctor of Technical Sciences, Professor
State University of Economics and Technology
50005, 16 Medychna Str., Kryvyi Rih, Ukraine
https://orcid.org/0000-0001-8780-587X

Oleksii Donchenko

Master, Assistant
State University of Economics and Technology
50005, 16 Medychna Str., Kryvyi Rih, Ukraine
https://orcid.org/0009-0002-4671-2823

Research and application of polynomial models in the development of an automated geological and mine surveying support system for open-pit mining

Abstract. The relevance of the research was determined by the necessity to improve the accuracy and efficiency of geological and mine surveying operations under conditions of complex geological structures and limited resources for field investigations. The development and implementation of automated software enabled the prompt analysis of large data volumes, optimisation of the testing network, and reduction of exploration costs. The aim of the research was the development of software for polynomial models of curves and surfaces with a proposed simplified algorithm for implementing the least squares method. The desktop application software was developed in Microsoft Visual Studio 2019 using the Microsoft Foundation Classes library. The mobile application was developed in Android Studio. For 3D graphics, the OpenGL library with shaders was employed, ensuring high operational performance. Parameters for evaluating the adequacy of the model were defined, as standard standalone packages provided insufficient objectivity. It was demonstrated in automated mode that the Lagrange polynomial represents a particular case of the developed polynomial model in curve construction. Furthermore, the interpolation method for surface construction was improved by incorporating spatial variability of indicators through the use of autocorrelation functions. During the construction of autocorrelation functions, spatial variability was assessed based on autocorrelation coefficients between values at adjacent control points, as well as the critical correlation radius, which made it possible to evaluate the predictive potential beyond the polynomial construction zone. The application of critical correlation radius in the automated geological and mine surveying system for analysing the testing network of all boreholes (exploratory and blasting) was described. It was noted that the network parameters should reflect the natural variability of the studied indicators. In cases where the deposit depth exceeded the bench height, the feasibility of predicting qualitative characteristics based on the results of blasting boreholes located above the mineral deposit was considered, allowing for a significant reduction in the scope of costly exploratory works. It was established that the multi-module automated geological and mine surveying system had been developed and implemented over many years in open-pit mines, particularly

Suggested Citation:
Zelensky, O., & Donchenko, O. (2025). Research and application of polynomial models in the development of an automated geological and mine surveying support system for open-pit mining. *Mining Journal of Kryvyi Rih National University*, 59(1), 10-21. doi: 10.31721/2306-5435-2025-1-10-21.

*Corresponding author

at the Erdenet enterprise (Mongolia), which develops a copper-molybdenum deposit. The results of the conducted research were utilised in the educational process and in the automated geological and mine surveying system for planning and managing mining operations in open-pit environments

Keywords: regression model; autocorrelation function; least squares method; Lagrange method; frame model; critical correlation radius; indicatrix

Introduction

Under conditions of increasingly complex subsurface geological structures, limited resources for fieldwork, and the necessity for prompt engineering decision-making, emerged an urgent need for the implementation of automated tools for analysing geological and mine surveying information. The enhancement of the efficiency of such operations was directly associated with the use of modern mathematical methods and software, which enabled a reduction in exploration costs and an improvement in the reliability of forecasts. Particular attention was paid to the advancement of spatial interpolation models adapted to real-world environments, where the indicators exhibit complex variability, and traditional statistical methods proved insufficiently effective.

Regression analysis, as highlighted by D. Mountgomery *et al*. (2021), represented a powerful statistical instrument widely applied in the natural, technical, and socio-economic sciences. I. Pardoe (2021) emphasised the role of the coefficient of determination R^2 as a key metric for assessing a model's capacity to explain the variation of the target variable. Meanwhile, the studies of I. David *et al*. (2020) demonstrated that traditional adequacy criteria could be complemented by new ones, such as median squared error prediction and robust model efficiency, which better accounted for the presence of extreme or leading data.

According to the findings of S. Lapach (2025), the least squares method (LSM) remained one of the most effective techniques for parameter estimation in regression analysis, particularly in the presence of large data scatter. It was also emphasised that under conditions of heteroscedasticity, various methods produced significantly different results, and the choice of model ought to be based on the characteristics of the dataset. This was supported by the research of V.V. Khlivnyi & C.V. Bazilo (2023), which substantiated the appropriateness of using the Fisher criterion and the multiple correlation coefficient R^2 for verifying the informativeness of metamodels.

In the field of geoinformation modelling, an essential aspect involved the consideration of spatial autocorrelation. As demonstrated in the works of P. Bidiuk *et al*. (2020), effective modelling of temporal and spatial processes was achieved through the application of the Box-Jenkins methodology as well as intelligent approaches that allowed for the refinement of regression model structures. These approaches ensured high modelling accuracy through the analysis of autocorrelation and partial autocorrelation functions.

Special attention was paid to the issue of interpolation model accuracy. In the study by O. Prokaza & O. Kuznetsova (2022), it was justified that high model significance could be attained through the inclusion of a greater number of relevant factors. At the same time, the insufficient significance of individual parameters was attributed to the influence of random disturbances and unaccounted variables. In the context of improving model accuracy, the approach of J. Bell (2020) appeared appropriate, where in unique orthogonal polynomial estimators were proposed to strike a balance between minimising the mean squared error and ensuring result stability under dynamic conditions.

Despite existing achievements, the automation of assessing the spatial variability of geological indicators using critical correlation radius, and their integration with interpolation models, remained insufficiently explored. Likewise, limited attention had been given to the possibilities of predicting the qualitative characteristics of mineral resources based on limited or indirect data, particularly from blasting boreholes. The research focused on the development of automated software for constructing polynomial models of curves and surfaces, incorporating spatial autocorrelation and evaluating the boundaries of reliable forecasts beyond the zone of direct measurements. The aim of the study was to construct polynomials $y = f(x)$ and $z = f(x, y)$, which most accurately described the dependency of the resultant indicator on the selected factor.

Materials and Methods

A simplified algorithm for implementing the LSM in the construction of polynomial models $y = f(x)$ and $z = f(x, y)$ was proposed. This approach eliminated the need for calculating partial derivatives, as presented in the textbook by E. Manoukian (2021), where the procedure was described in detail. The task was reduced to the elementary formulation of a system of linear equations based on the proposed algorithm. As outlined in the textbook by V.M. Horbachuk & O.I. Kushlyk-Dyvulska (2023), this system was solved using the Gaussian elimination method. The adequacy of the polynomial model was subsequently evaluated. This study introduced enhancements to polynomial models and provided an assessment of their accuracy, along with interpolation methods that accounted for indicator variability in constructing surfaces using polynomial and frame-based models.

Mathematical modelling is typically conducted in various standalone software packages. These tools are often difficult to integrate as components within larger systems, particularly in the context of automated generation of input data and transmission of results within a system. Furthermore, such packages usually offer insufficient assessment of model adequacy. Within the geological and mine surveying software (GMSS) module, the following methods for curve construction were implemented: 1) Polynomial method (based on the least squares method); 2) Lagrange polynomial; 3) Cubic spline interpolation; 4) Hermite spline interpolation; 5) Bézier curves; 6) B-splines. For surface modelling, the following methods were developed: 7) Polynomial method (based on the least squares method); 8) Interpolation method (grid-based and frame-based); 9) B-splines (NURBS surfaces); 10) Nearest neighbourhoods (polygons).

In methods 2 to 4, the curves strictly passed through the control points. Methods 5 and 6 ensured that the curves passed exactly through the first and last points. In many cases, these constraints were unnecessary. However, for solving a range of problems, the involvement of all control points in curve and surface construction was essential. These conditions were satisfied in polynomial models. Nevertheless, these models did not account for values located between adjacent points, which significantly affected the evaluation of model adequacy. This factor had to be considered when assessing the spatial variability of indicators during borehole testing for the selection of test networks, reserve estimation, and ore body delineation along quarry horizons. An example of generating a curve polynomial $y = f(x)$ could be found in Microsoft Excel, although such tools typically failed to provide any evaluation of variability between control points.

The article proposed methods for evaluating model adequacy, ensuring a more accurate selection of a polynomial model. Similar to time series, polynomial construction was considered under the condition $x_{i+1} > x_i$. Particular attention was given to the B-spline method (methods 6 and 9), where a dynamic controllable model was formed through the adjustment of weights at the control points. This method was implemented in the construction of the open-pit surface based on the coordinates of bench edges. The GMSS had been continuously expanded and improved, both in terms of software and calculation accuracy. This fully applied to the curve and surface modelling module, which was employed in tasks such as assessing spatial variability of indicators, resource estimation (both operational and complete), ore extraction accounting, forming the information base for mining planning and management, and modelling of the deposit and open-pit contours.

In its initial implementation, O. Zelensky & V. Lysenko (2022) introduced the GMSS module at the Erdenet enterprise (Mongolia), which develops a copper-molybdenum deposit. The new version was developed for the conditions of an iron ore deposit, using the example of the Southern Mining and Processing Plant (Southern MPP). In this context, significant improvements in software and mathematical support were made by O. Zelensky (2023). To construct the polynomial model, an array of control points was utilised, generated either automatically from shuffled or sequential data. A sinusoidal function was selected for forming sequential data due to its continuous, periodic, and smooth nature, without phase shift as observed in the cosine function. Based on the sine function, a clear relationship between data values was identified, resulting in a high level of model adequacy. This visibly demonstrated the relevance of the developed indicator system.

To determine the acceptable level of model adequacy, dozens of test scenarios were conducted, allowing the identification of deviation levels from standard indicators. The number of points *N* did not directly influence the performance of the system, but it was taken into account when constructing the model and determining the maximum degree of the polynomial. The desktop application software for constructing polynomial models was developed in C++ using the Microsoft Foundation Classes (MFC) library. Model visualisation was performed with the OpenGL library. User interaction was facilitated through the graphical interface of the MFC application. Peripheral devices such as a keyboard and mouse were employed for parameter control, as further detailed in the Results section. Users were provided with the ability to adjust the polynomial degree and the number of control points. All computational algorithms were implemented manually by the authors, without reliance on additional libraries. To support comprehensive research execution, a mobile application for the Android operating system was also developed (Zelensky, 2023), offering functionality similar to the desktop version. The article provided an in-depth investigation of the polynomial method and the Lagrange polynomial, while other methods were described in the works of O. Zelensky (2023) and O. Zelensky & V. Lysenko (2022).

Results

When constructing curves, it is necessary to determine such a function $y = f(x)$ that most accurately characterises the dependence of the resulting indicator on the chosen factor. To obtain a polynomial curve, the LSM was used, which ensures the minimisation of the sum of squared deviations of the given indicators from the resulting curve. The following formula was used (1):

$$\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 \rightarrow min, \quad (1)$$

where y_i – the actual value of the resulting indicator; $\hat{y}_i$ – the computed value of the indicator; n – the number of observations. In general, the curve formula has the following form (2):

$$f(x) = A_0 + A_1x + A_2x^2 + \ldots + A_M^M, \text{ or, } f(x) = \textstyle\sum_{i=0}^{M} A_i xi, \quad (2)$$

where M – the degree of the polynomial. Obtaining A_i by the least squares method is reduced to solving a system of linear equations consisting of $M+1$ linear algebraic equations (3):

$$\begin{cases} a_{00}A_0 + a_{01}A_1 + *** + a_{0M}A_M = b_0 \\ a_{10}A_0 + a_{11}A_1 + *** + a_{1M}A_M = b_1 \\ ****************************** \\ a_M A_0 + a_{M+1}A_1 *** + a_{2*M}A_M = b_M. \end{cases} \quad (3)$$

When determining the coefficients a_{ij} and b_{ij}, the literature often provides a rather complex solution algorithm. O. Zelensky & V. Lysenko (2022) proposed a simple method for obtaining these coefficients using the formula (4):

$$a_{ij} = \textstyle\sum_{k=1}^{n} x_k^{i+j}; \; b_i = \sum_{k=1}^{n} x_k^{i} \cdot y_k, \quad (4)$$

where x_k, y_k – the coordinates of the input control points; n – the number of control points.

Further, solving the system of linear algebraic equations using classical methods is not difficult, that is, obtaining the values of the coefficients A_i. In the program developed by the authors, the Gauss method was used. For the study of the polynomial curve, the sine wave $y = sin(x)$ in the range from -3π to 3π was used as the source data. The input data is the number of control points N. In this case, the value of the control points (y) starts from the point $x = -3\pi$, $y = 0$. Then y is determined by increasing the argument x by 6π / $(N-1)$. Naturally, the last point will have the value $y = 0$, $x = 3\pi$. The obtained control points are used in two variants: sequential data and shuffled data.

In the first case, the sequence of control point formation was described above, in the second – their order is randomly changed. Figure 1 shows the dialog window for generating input data to build two polynomials. The number of control points for the first calculation is 9, and for the second – 101. In this case, the data is not "shuffled". The location of the control points is shown in Figure 2.

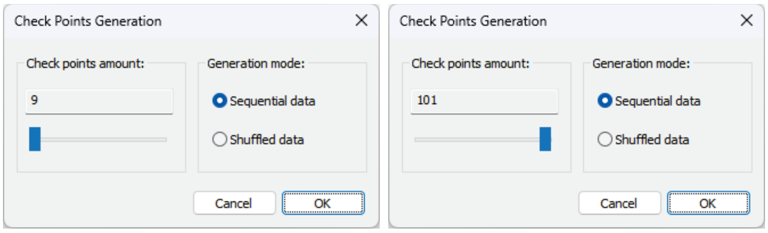


Figure 1. *Message box for generating input data*

Source: *software developed by the authors*

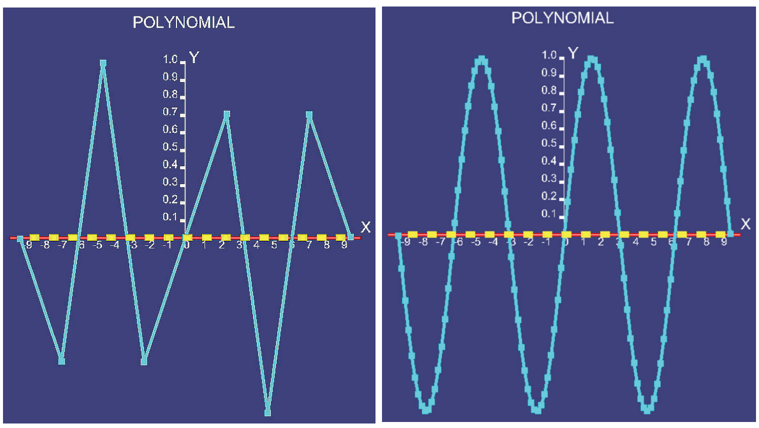


Figure 2. *Location at 9 and 101 point*

Source: *software developed by the authors*

In Microsoft Excel, the degree of correspondence between the model and the real data is determined by the coefficient of determination, R^2. In other words, this is a measure of the model's accuracy. R^2 is calculated using formula (5):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}; \; \bar{y} = \frac{1}{n}\textstyle\sum_{i=1}^{n} y_i, \quad (5)$$

where n – the number of input data points; y_i and $\hat{y}_i$ – the actual and predicted values of the indicator.

If $R^2 = 1$, this corresponds to a perfect model in which all observation points lie exactly on the regression line – that is, the sum of squared deviations is zero. If $R^2 = 0$, there is no relationship between the variables in the polynomial model, and instead, the mean of the observed values can be used to estimate the output variable. A model with R^2 in the range of 0.5 to 0.8 is considered satisfactory. If $R^2 > 0.8$, the model is regarded as highly reliable. Values below 0.5 indicate that the model is ineffective. A polynomial passes through all control points when its degree is maximal – that is, one less than the number of input data points. Figure 3 presents an 8th-degree polynomial for 9 control points.

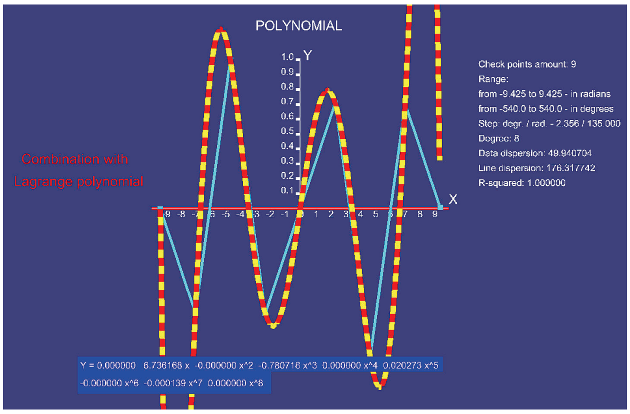


Figure 3. *Polynomial of degree 8*
Notes: *below is the equation of polynomial of degree 8*
Source: *software developed by the authors*

As shown in Figure 3, the polynomial passes through all control points, and in this case $R^2 = 1$. However, the use of such a polynomial is impractical due to sharp fluctuations between the control points. In this case, it is proposed to compare the variances of the data and the curve using formula (6):

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2. \qquad (6)$$

These values are presented in Figure 3. Here, the variance of the input data is 49.9, and the estimated variance of the curve (based on the function values at increasing arguments with a fixed step) is 176.3. In this case, even though $R^2 = 1$, adopting such a model would be incorrect. Therefore, in addition to R^2, the variance ratio should be considered. It is proposed that the acceptable difference in variances should not exceed 20%, although this may vary depending on the specifics of the domain under study. Figures 4 and 5 show 7th- and 10th-degree polynomials constructed for 40 input points.

It was visually evident that the 10th-degree polynomial model proved to be more effective. This was also confirmed by quantitative evaluation. As observed from the data presented in Figures 4-5, the variance of the input data equalled 50. The variances of the 7th- and 10th-degree polynomials amounted to 32.2 and 48.4, respectively. The deviations of the variances of the 7th-degree and 10th-degree polynomials from the variance of the input data (k_7 and k_{10}) were determined as follows:

For the 7th-degree polynomial:
$k_7 = (50 - 32.2)/50 \cdot 100 = 35.6\%$

For the 10th-degree polynomial:
$k_{10} = (50 - 48.4)/50 \cdot 100 = 3.2\%$

The value of k_{10} was significantly below 20% in comparison to k_7.

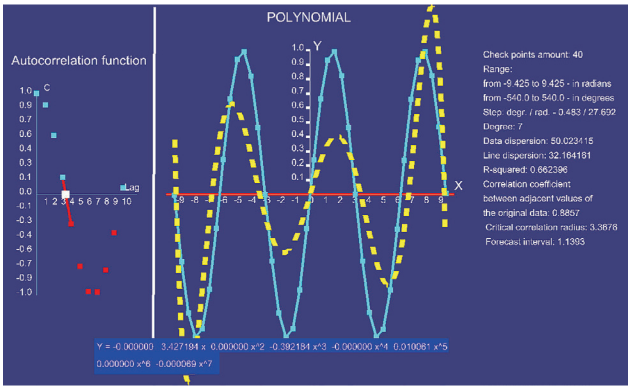


Figure 4. *Polynomial of degree 7*
Source: *software developed by the authors*

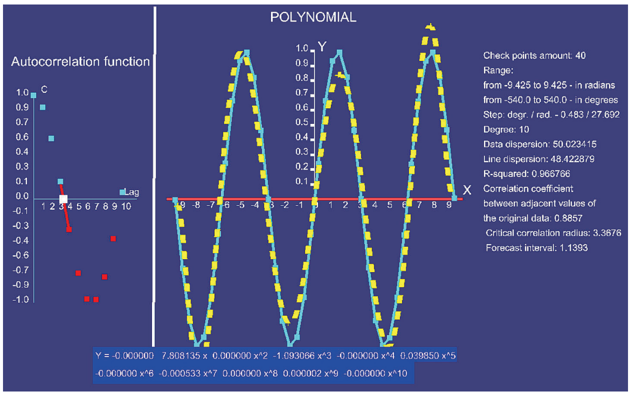


Figure 5. *Polynomial of degree 10*
Source: *software developed by the authors*

As shown in Figure 5, the same input data were used ($N = 40$, degree = 10), but they were arranged in a random order. These were generated by selecting the "Shuffled data" option in the dialogue window (Fig. 1). In this case, the polynomial variance ratio equalled (Fig. 6): $k_{10} = (50 - 16.2)/50 \cdot 100 = 67.6\%$. Such a model could not be considered acceptable.

As previously indicated, a polynomial passed through all N control points if its degree equalled N – 1. This was demonstrated in Figure 3, which presented an 8th-degree polynomial for 9 control points. The curve was represented by a line comprising alternating red and yellow segments. In this case, the constructed polynomial fully coincided with the Lagrange polynomial.

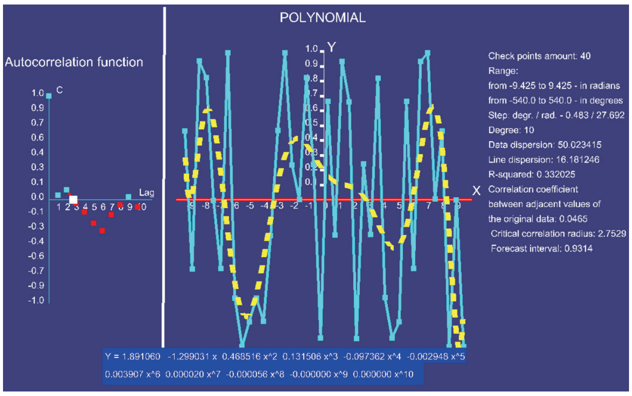


Figure 6. *Polynomial of degree 10 ("shuffled data")*
Source: *software developed by the authors*

This polynomial strictly passed through all control points. This result indicated that the constructed polynomial was sufficiently universal for any given degree and, in particular, coincided completely with the Lagrange polynomial when the degree equalled $N - 1$. In the Lagrange polynomial, for $N + 1$ numbers (x_0, y_0), (x_1, y_1), ..., (x_n, y_n), there existed a unique polynomial of degree N such that (7):

$$L(x) = \sum_{i=0}^{n} y_i \, l_i(x), \tag{7}$$

where $l_i(x)$ was defined by formula (8):

$$l_i(x) = \\ = \prod_{j=0, j\neq i}^{n} \frac{x - x_j}{x_i - x_j} = \frac{x - x_0}{x_i - x_0} \cdots \frac{x - x_{i-1}}{x_i - x_{i-1}} \cdot \frac{x - x_{i+1}}{x_i - x_{i+1}} \cdots \frac{x - x_n}{x_i - x_n}. \tag{8}$$

For any $i = 0, \ldots, n$, the polynomial l_i had a degree of n and satisfied the condition (9):

$$l_i(x_j) = \begin{cases} 0, & j \neq i, \\ 1 & j = i. \end{cases} \tag{9}$$

Hence, $L(x)$ was a linear combination of the polynomials $l_i(x)$ and had a degree of no more than n, while also satisfying $L(x_i) = y_i$. During the study of polynomial models, an important indicator for evaluating spatial variability was examined. For this purpose, the autocorrelation function (ACF) was utilised. Quantitatively, the ACF could be measured using autocorrelation coefficients. The ACF was a function $r_m = f(m)$, where r_m represented the autocorrelation coefficient between the initial sequence of control points and the sequence shifted relative to the initial one by a lag of m. For example, the initial sequence included n numbers. In this case, $n = 8$ with the following values: 5, 8, 1, 9, 4, 6, 9, 10. For a shift $m = 1$, the autocorrelation coefficient (degree of association) was determined between $n - m$ pairs of numbers: 5, 8, 1, 9, 4, 6, 9 – 8, 1, 9, 4, 6, 9, 10. For $m = 2$, the autocorrelation coefficient was determined between $n - m$ pairs: 5, 8, 1, 9, 4, 6 – 1, 9, 4, 6, 9, 10, and so forth. Thus, in constructing the ACF, the values of r_m (autocorrelation coefficients for increasing m) were used. The value of r_m was determined by formula (10):

$$r_m = \frac{\sum_{i=0}^{n-m}(x_i - \bar{x}_1)(x_{i+m} - \bar{x}_2)}{\sqrt{\sum_{i=0}^{n-m}(x_i - \bar{x}_1)^2 (x_{i+m} - \bar{x}_2)^2}}, \tag{10}$$

where $\bar{x}_1$ – the arithmetic mean of the initial sequence (from the first to the $n - m$ elements), and $\bar{x}_2$ – the arithmetic mean of the "shifted" sequence (from the m-th to the n-th elements).

It was noted that with each increment of lag m by one, the number of value pairs used to calculate the ACF coefficient decreased by one. Therefore, the maximum lag (m) was usually recommended to be equal to $n/4$. The ACF reached its maximum value of 1 at $m = 0$ (indicating full self-correlation of the series), and equalled 0 when the initial sequence and its shifted counterpart were uncorrelated. The faster the autocorrelation function decreased with increasing m, the weaker the autocorrelation was, and vice versa. The ACF gradually decreased and intersected the abscissa axis at a distance r_k from the origin. The parameter r_k was referred to as the critical correlation radius. At the initial shift ($m = 1$), r_1 was defined as the autocorrelation coefficient between values at adjacent control points. When $r_1 > 0.7$, a consistent trend in the variation of the studied parameter was observed. When $r_1 \leqslant 0.7$, the dataset was considered a set of random, independent variables. Figures 4-6 illustrated the autocorrelation functions and their corresponding critical correlation radius. The number of control points equalled 40. Figures 4 and 5 employed identical input data; therefore, the results of the autocorrelation analysis were also identical: the autocorrelation coefficient between adjacent values in the input data was $r_1 = 0.88$; the critical correlation radius equalled $r_k = 3.36$. In Figure 6, using the same 40 input values but arranged in random order, the results were as follows: $r_1 = 0.039$, $r_k = 1.78$.

Three factors were identified as determining the adequacy level of the polynomial model: 1) the coefficient of determination, which depended on the degree α; 2) the ratio of the variances between the curve and the control points; 3) the autocorrelation between adjacent values in the input data (r_1). Among these, the second and third factors were of greatest significance. Moreover, these factors were interrelated. The model was considered effective when $r_1 > 0.7$ and the ratio of the variances between the curve and input data was less than 20%. While r_1 depended solely on the input data, the variance ratio depended on the selected polynomial degree α.

When evaluating the spatial variability of indicators, the critical correlation radius r_k, as described by O. Zelensky & V. Lysenko (2022), was regarded as an important parameter. This value allowed for the estimation of forecast capability beyond the polynomial construction range. The application of r_k was further examined within the GMSS system. For many years, the GMSS system, as part of an automated quarry management system, had been implemented in non-ferrous and iron ore deposits. In this context, the input data mainly comprised the coordinates from borehole collar surveys, bench edges, tracks, and other objects, as well as the results of blast hole testing and both detailed and operational exploration.

Implementation experience demonstrated that at non-ferrous deposits, blast hole testing was carried out almost comprehensively, i.e. for all boreholes. The critical correlation radius r_k, determined in the study of variability of total copper content (Cu_{tot}) in the porphyry copper ores of the Erdenet GMSS, were presented in Table 1. In this case, the network of blast and exploration boreholes revealed the natural characteristics of the studied parameters along the quarry horizons, as the correlation radius exceeded the testing intervals. In terms of deposit depth, the critical correlation radius was significantly less than the bench height (15 m), indicating the appropriateness of forecasting quality parameters based on blast hole test data from the overlying horizon for the purpose of operational resource estimation under conditions of limited exploration data availability. This approach allowed for significant cost reduction in conducting expensive exploration operations.

Table 1. *Critical correlation radius for Cutotal (in meters)*

Test Network	By the area of the deposit		By depth
	along the strike	across the strike	
By blast borehole, 8×8	24 ± 7	20 ± 12	120 ± 20
By exploration borehole, 30×60	130 ± 40	100 ± 40	

Source: *developed by the authors on the basis of their own research*

Anisotropy is present, and to evaluate it, an indicatrix in the form of an ellipse is used, which illustrates the variability of parameters in different directions. The indicatrix is constructed based on the critical correlation radius in various directions. Figure 7(a) shows the autocorrelation function and the corresponding indicatrix. The use of indicatrices to assess the variability of a key parameter in the mining face-based on blast hole testing data from the digital model of the deposit-allows for a reduction in fluctuations of the parameter in the ore flow by selecting the mining direction along the major axis of the indicatrix. In Figure 7(b), the arrow indicates the direction of the least variability, which is the proposed mining direction.

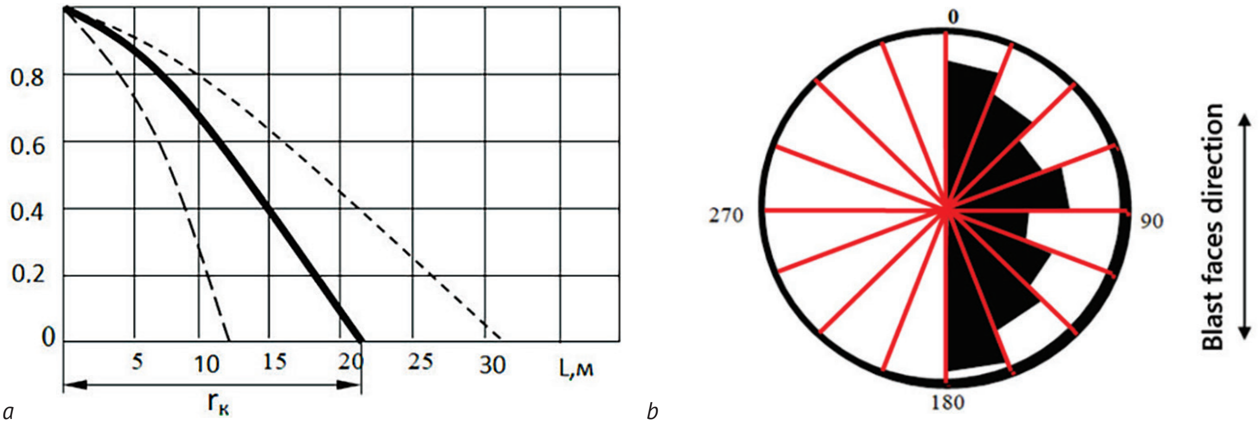
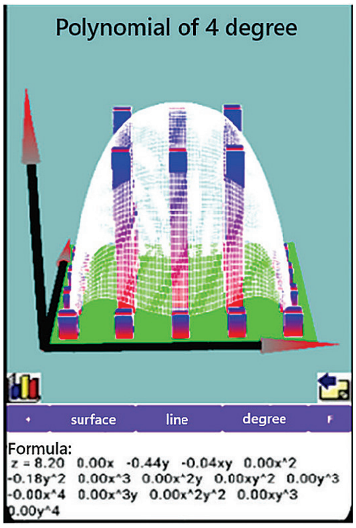


Figure 7. *Indicators of variability of indicators*

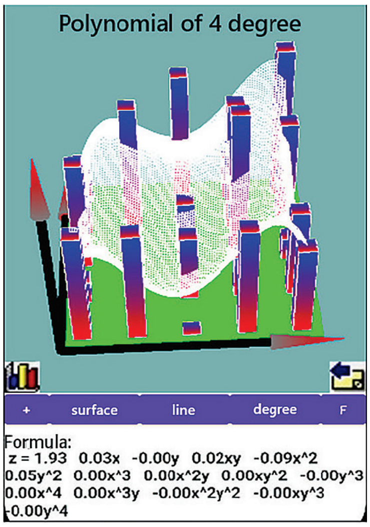
Notes: *a – autocorrelation function; b – indicatrix*

Source: *software developed by the authors*

The construction of a surface based on a polynomial model is presented in the works of O. Zelensky (2023). These works also provided a simplified algorithm for implementing the least squares method. The core of the algorithm is minimising the sum of squared deviations of the function $z = f(x, y)$ from the input data. Figure 8 shows two degree-4 polynomials (surfaces) based on 25 input data points. In Figure 8(a), model adequacy can be assessed using the coefficient of determination. However, this is a specific case where the indicator values change smoothly. In cases of more complex variability, a second indicator must be used: the deviation between the variance of the input data and the variance of the fitted surface. This is evident in the example shown in Figure 8(b). A detailed selection of surface construction methods requires further research.



a



b

Figure 8. *Building a surface with a polynomial of degree 4*
Notes: *a – sequential data; b – shuffled data*
Source: *software developed by the authors*

The construction of a surface within a specified contour by pit horizons is carried out as follows. The user defines a contour that is covered with a regular grid of a given step size. At each grid node Z_A, the indicator value is interpolated based on test data from the nearest boreholes. The algorithm for finding these nearest boreholes is described in the work of O. Zelensky & V. Lysenko (2022). The indicator value at a node is determined using the inverse distance method (11):

$$Z_A = \frac{\sum_{i=1}^{N} Z_i \cdot d_i^{-\alpha}}{\sum_{i=1}^{N} d_i^{-\alpha}}, \quad (11)$$

where α – the interpolation power (usually $\alpha = 2$); d_i – the distance between the node and the i-th closest borehole; Z_i – the indicator value at the i-th borehole; N – the number of nearest boreholes to the node.

This well-known formula does not account for the natural component of indicator variability. This can be addressed by incorporating the indicatrix, which reflects the presence of critical correlation radius in all directions. In this case, the inverse distance method is refined as follows (12, 13):

$$Z_A = \frac{\sum_{i=1}^{N} Z_i \cdot d_i^{-\alpha ot_i}}{\sum_{i=1}^{N} d_i^{-\alpha ot_i}}, \quad (12)$$

$$ot_i = \frac{r_{ki}}{r_{kmax}}, \quad (13)$$

where r_{ki} – the critical correlation radius in the i-th direction; r_{kmax} – the maximum critical correlation radius.

One of the modules of the developed automated GMSS for the conditions of the Southern MPP includes an interpolated 3d model, shown in Figure 9. The user defines a contour, which is covered with a regular grid with a defined step size *sh*. For each grid node, rectangular prisms are constructed, with heights equal to the interpolated Fe_{mag} values, and the base being a square with side length *sh*. The coordinates *x*, *y* of the square's centre correspond to the *x*, *y* coordinates of the grid node. An alternative to the interpolation model is a wireframe model (Fig. 10), where interpolated values are simply connected by straight lines.

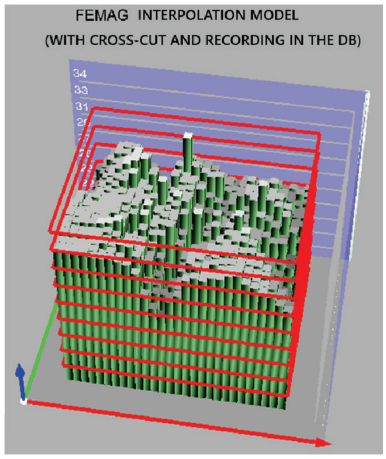


Figure 9. Interpolation model of Fe_{mag} variability
Source: *software developed by the authors*

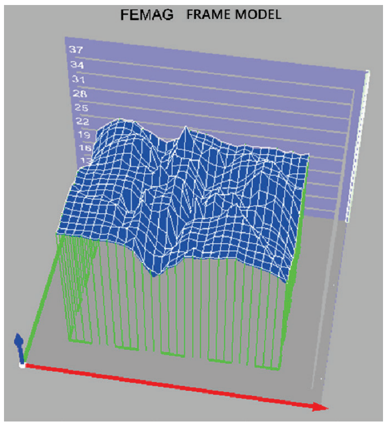


Figure 10. *Wireframe model of* Fe_{mag} *variability*
Source: *software developed by the authors*

Particular attention within the conducted study was devoted to the assessment of the adequacy of a polynomial model reflecting the relationship between input and output data in forecasting tasks. The results obtained during the experiments confirmed the necessity of a more thorough analysis of both the accuracy of the models themselves and the interpretability of the resulting coefficients, taking into account the spatial and statistical properties of the data.

One of the fundamental tools for assessing model quality was the coefficient of determination (R^2), which is used, in particular, within the Microsoft Excel software environment. However, its application was not always sufficient for adequate evaluation, especially in cases involving spatial or temporal variability of data between control points. In this study, the problem was partially addressed through additional analysis of spatial variability by means of the autocorrelation function and variance ratio, which enabled the assessment of model smoothness and stability in the intervals between known values. This approach shared common features with the method of statistical modelling of random fields proposed by Z. Vizhva *et al*. (2023), where emphasis was likewise placed on accounting for structural heterogeneity and constructing adequate realisations to supplement incomplete data. The distinction lay in the application area (geophysics versus general-purpose regression modelling); however, the approach to handling the spatial components of the data was conceptually similar.

At the same time, as noted by B. Sencer (2024), Excel possesses significant limitations in regression analysis, particularly concerning the assessment of non-linear models. The author pointed to the systematic overestimation of the coefficients of determination in the case of non-linear functions approximated as linear. In this study, such issues were addressed through the use of R^2, root mean square error, and mean absolute error calculations within an external environment, which allowed the avoidance of misleading conclusions typical for Excel's built-in functions. Hence, the obtained results confirmed the need for a critical evaluation of standard statistical analysis tools, especially in the context of complex models. An alternative perspective was presented by U. Srilakshmi *et al*. (2024), who, working in the field of predictive modelling in healthcare, proposed an innovative approach to improving the efficiency of linear regression models by reducing the dimensionality of variables. Unlike traditional methods, this approach involved a preliminary transformation of input data, which significantly reduced the sum of squared errors. The research described in this article did not involve prior dimensionality reduction; however, the influence of the random component on model accuracy was taken into account. Moreover, both studies indicated the advantage of using non-parametric methods in cases where classical assumptions were violated.

To support this, reference may be made to the conclusions of W. Laverty & I. Kelly (2021), who compared the least squares method with the non-parametric Kendall and Theil approach. In cases where data exhibited heteroscedasticity or asymmetry, the latter method provided high estimation efficiency. In the cases considered in this article, some datasets displayed signs of non-normality, making the application of adaptive or non-parametric modelling approaches particularly relevant. This aligned with the findings of O. Tkachenko (2021), who proposed an approach for adaptive regression parameter estimation using higher-order statistics. Such an approach enabled the adjustment of models to real data violating Gaussian assumptions, which were not accounted for in classical regression analysis within Excel. These considerations were further supported by the conclusions of D. Tadashi *et al*. (2024), who focused on the development of local polynomial models of software reliability. Although the applied domain differed, their approach to polynomial degree control and adaptive forecasting was highly relevant in terms of modelling complex dependencies in data. The idea of using semi-parametric models could also be adapted within general regression tasks to achieve greater flexibility in model construction.

In the study by O. Zorin & V. Palahin (2023), a new approach to the statistical processing of RZ-signals was proposed, based on the use of moment-cumulant models for describing the fine structure of random processes. Although the subject area differed slightly, the proposed method involved constructing polynomial stochastic solvable rules with coefficient optimisation based on a moment criterion. The practical implementation and testing of such algorithms became possible through MATLAB tools, which were detailed in the work of O. Romanyuk *et al*. (2024). The MATLAB environment provided the necessary functionality for implementing numerical methods, system modelling, and visualisation of results, which was critically important for evaluating the effectiveness of new reception methods under complex noise conditions. This once again highlighted the importance of polynomial models and their feasibility of implementation across various programming environments or mathematical software packages.

Thus, the comparative analysis of the results of the study presented in this article with existing approaches in the literature enabled several important generalisations to be drawn. Standard analytical tools, such as Microsoft Excel, remained limited in the context of evaluating complex models, particularly in cases of non-linear or adaptive regression. Spatial or statistical data analysis between control points allowed for improved modelling accuracy, as confirmed both in this study and in the works of other authors. The application of adaptive and non-parametric approaches, such as those proposed by W. Laverty & I. Kelly (2021), O. Tkachenko (2021), and U. Srilakshmi *et al*. (2024), opened new possibilities for improving model accuracy, especially when classical statistical assumptions were violated. The development of more sophisticated models, including semi-parametric or local polynomial types, enabled effective modelling of complex systems, with wide-ranging applications not only in engineering or economics, but also in healthcare, software reliability, and geophysics.

In summary, it may be stated that the effective application of methods for assessing the adequacy of polynomial models under modern conditions required a comprehensive approach, combining classical techniques with adaptive methods, the use of new spatial data processing tools, and non-parametric statistical approaches. The obtained results confirmed the potential for further research into the improvement of model quality assessment, taking into account the spatio-temporal heterogeneity of data, which had become particularly relevant in interdisciplinary areas of mathematical modelling application.

Conclusions

The article presented a study of polynomial functions $y = f(x)$ and $z = f(x, y)$ of a given degree. The implementation of the polynomial model was based on the method of least squares. Due to the complexity of the task, a simplified algorithm was proposed, which reduced the problem to solving a system of linear equations. The adequacy of the polynomial model was assessed using integrated packages of the coefficient of determination R^2, as implemented in Microsoft Excel. However, this assessment remained limited, as it did not take into account the nature of the curve's behaviour between control points. This significant drawback was addressed by two additional indicators. The first indicator involved the ratio of the variances of the input data and the curve points. The second indicator comprised the assessment of spatial variability of indicators using an autocorrelation function. All studies described in the article were carried out in an automated mode, which considerably simplified the process and enabled the simulation of a substantial amount of data for obtaining results. An example included the construction of a polynomial curve of a selected degree α, equal to one less than the number of control points (N - 1). In this case, the curve passed precisely through the control points. An identical curve was obtained using the Lagrange method. When constructing the autocorrelation function, the assessment of spatial variability of the indicator was determined by the autocorrelation coefficient between values at neighbouring control points, as well as by the critical correlation radius r_k, which enabled the evaluation of the forecasting capability beyond the constructed polynomial. In the automated HMSS system, critical correlation radius was used for assessing the test network across all boreholes (exploratory and blasting). The parameters of the network had to reflect the natural characteristics of the studied indicators.

In the case of non-ferrous metals, tests were generally conducted on blasting boreholes. If the deposit depth exceeded r_k, i.e. the height of the bench, it was considered appropriate to forecast the qualitative indicators based on test data from blasting boreholes of the overlying horizon. This significantly reduced the need for costly exploratory operations. Based on r_k values in different planar directions, an indicatrix was constructed. Its use enabled the implementation of interpolation and framework models that accounted for the spatial variability of indicators. The construction of curves and surfaces was performed for both desktop and mobile applications and was successfully implemented at non-ferrous and iron ore deposits as part of an automated geological and mine surveying system. Future research prospects involved improving spatial forecasting algorithms using surfaces as an example, and integrating the software with artificial intelligence for adapting results to various geological and technical conditions.

Acknowledgements

None.

Funding

None.

Conflict of Interest

None.

References

[1] Bell, J.W. (2020). Polynomial least squares multiple-model estimation: Simple, optimal, adaptive, and practical. *SN Applied Sciences*, 2, article number 1964. doi: 10.1007/s42452-020-03439-x.

[2] Bidiuk, P.I., Demkivskyi, Y.O., & Demkivska, T.I. (2020). Method of development of heteroskedastic processes models. *Bulletin of the Kyiv National University of Technology and Design. Technical Science Series*, 142(1), 19-26. doi: 10.30857/1813-6796.2020.1.2.

[3] David, I.J., Adubisi, O.D., Ogbaji, O.E., Eghwerido, J.T., & Umar, Z.A. (2020). Resistant measures in assessing the adequacy of regression models. *Scientific African*, 8, article number e00437. doi: 10.1016/j.sciaf.2020.e00437.

[4] Horbachuk, V.M., & Kushlyk-Dyvulska, O.I. (2023). *Probability theory and mathematical statistics*. Kyiv: Igor Sikorsky Kyiv Polytechnic Institute.

[5] Khlivnyi, V.V., & Bazilo, C.V. (2023). Influence of operating parameters of hydroabrasive cutting of various materials on the accuracy of observing the cutting width. *Bulletin of Cherkasy State Technological University*, 28(2), 38-49. doi: 10.24025/2306-4412.2.2023.280719.

[6] Lapach, S.M. (2025). Conflict of user accuracy requirements and regression model indicators. *Mathematical Machines and Systems*, 2025(1), 91-102. doi: 10.34121/1028-9763-2025-1-91-102.

[7] Laverty, W.H., & Kelly, I.W. (2021). Exploring the effects of assumption violations on simple linear regression and correlation using Excel. *American Journal of Theoretical and Applied Statistics*, 10(4), 194-201. doi: 10.11648/j.ajtas.20211004.13.

[8] Manoukian, E.B. (2021). *Mathematical nonparametric statistics*. Hoboken: Taylor & Francis Group.

[9] Mountgomery, D.C., Peck, E.A., & Vining, G.G. (2021). *Introduction to linear regression analysis* (6th ed). Hoboken: John Wiley & Sons, Inc.

[10] Pardoe, I. (2021). *Applied regression modeling* (3d ed.). Hoboken: John Wiley & Sons, Inc.

[11] Prokaza, O.I., & Kuznetsova, O.V. (2022). Construction of mathematical models of evaporation plant by experimental-statistical method. *Visnik of the Volodymyr Dahl East Ukrainian National University*, 271(1), 36-40. doi: 10.33216/1998-7927-2022-271-1-36-40.

[12] Romanyuk, O.N., Romanyuk, O.V., & Tsikhanovskaya, O.M. (2024). Information support for economic disciplines. In *Proceedings of the XVII international scientific and practical conference of the information technologies and automation – 2024* (pp. 393-396). Odesa: Odesa National University of Technology.

[13] Sencer, B. (2024). On the pros and cons of using Excel for regression analysis. *Turkish Journal of Agriculture – Food Science and Technology*, 12(s2), 2234-2241. doi: 10.24925/turjaf.v12is2.2234-2241.6931.

[14] Srilakshmi, U., Manikandan, J., Thanmayee, V., Abhinav, G., Tharun, K., & Saideep, D. (2024). A new approach to computationally-successful linear and polynomial regression analytics of large data in medicine. *Journal of Computer Allied Intelligence*, 2(2), 35-48. doi: 10.69996/jcai.2024009.

[15] Tadashi, D., Siqiao, L., & Okamura, H. (2024). Local polynomial software reliability models and their application. *Information and Software Technology*, 166, article number 107366. doi: 10.1016/j.infsof.2023.107366.

[16] Tkachenko, O.M. (2021). *Polynomial methods and tools for estimating regression parameters using Non-Gaussian error models*. (Doctoral dissertation, Cherkasy State Technological University, Cherkasy, Ukraine).

[17] Vizhva, Z., Demydov, V., & Vizhva, A. (2023). Statistical modeling of random fields with Gaussian correlation function for magnetometry data analysis. *Bulletin of Taras Shevchenko National University of Kyiv: Geology*, 3(102), 81-87. doi: 10.17721/1728-2713.102.

[18] Zelensky, O. (2023). Construction surface using desktop and mobile applications. *Scientific and Practical Journal "Economics and Technical Engineering"*, 1(1), 87-102. doi: 10.62911/ete.2023.01.01.07.

[19] Zelensky, O.S., & Lysenko, V.S. (2022). Improvement of automated contouring of ore grades along the horizons of ore quarries. *Journal of Kryvyi Rih National University*, 55, 15-24. doi: 10.31721/2306-5451-2022-1-55-15-24.

[20] Zorin, O., & Palahin, V. (2023). Adaptive system for receiving data by bipolar discrete signals in information and measurement systems in asymmetric non-Gaussian noise. In *International scientific conference of the information technologies and computer modelling* (pp. 181-182). Ivano-Frankivsk: Vasyl Stefanyk Precarpathian National University.

Олександр Зеленський
Доктор технічних наук, професор
Державний університет економіки і технологій
50005, вул. Медична, 16, м. Кривий Ріг, Україна
https://orcid.org/0000-0001-8780-587X

Олексій Донченко
Магістр, асистент
Державний університет економіки і технологій
50005, вул. Медична, 16, м. Кривий Ріг, Україна
https://orcid.org/0009-0002-4671-2823

Дослідження та використання поліноміальних моделей при розробці автоматизованого геолого-маркшейдерського забезпечення гірничих робіт у кар'єрі

Анотація. Актуальність дослідження зумовлена потребою в підвищенні точності та ефективності геолого-маркшейдерських робіт в умовах складної геологічної будови та обмежених ресурсів для польових досліджень. Розробка та впровадження автоматизованого програмного забезпечення надали змогу оперативно аналізувати великі обсяги даних, оптимізувати мережу випробувань і зменшити витрати на розвідувальні роботи. Метою дослідження була розробка програмного забезпечення поліноміальних моделей кривих і поверхонь із запропонованим спрощеним алгоритмом реалізації методу найменших квадратів. Програмне забезпечення для десктопного додатку розроблено в середовищі Microsoft Visual Studio 2019 на базі бібліотеки Microsoft Foundation Classes. Мобільний додаток розроблено в середовищі Android Studio. Для роботи з 3D-графікою застосовується бібліотека OpenGL з використанням шейдерів, що забезпечує високу продуктивність експлуатації. Було визначено параметри для оцінки адекватності моделі, яка в стандартних автономних пакетах недостатньо об'єктивна. В автоматизованому режимі показано, що поліном Лагранжа є окремим випадком розробленої поліноміальної моделі при побудові кривих. Крім того, вдосконалено інтерполяційний метод при побудові поверхні. При цьому враховується просторова мінливість показників з використанням автокореляційних функцій. При побудові автокореляційних функцій було здійснено оцінку просторової мінливості показника на основі коефіцієнтів автокореляції між значеннями у сусідніх контрольних точках, а також критичного радіуса кореляції, що дало змогу оцінити потенціал прогнозування за межами зони побудови полінома. Було описано застосування критичних радіусів кореляції в автоматизованій системі геолого-маркшейдерського забезпечення для аналізу мережі випробувань усіх свердловин (вибухових і розвідувальних). Було зазначено, що параметри мережі мають відображати природну варіативність досліджуваних показників. У випадках, коли глибина покладів перевищує висоту уступу, було розглянуто доцільність прогнозування якісних характеристик за результатами випробувань вибухових свердловин, розташованих вище залягання корисної копалини, що дозволяє суттєво скоротити обсяги дорогих розвідувальних робіт. Було встановлено, що багатомодульна автоматизована система геолого-маркшейдерського забезпечення протягом багатьох років удосконалюється та впроваджується на кар'єрах, зокрема на підприємстві «Ерденет» (Монголія), яке розробляє мідно-молібденове родовище. Результати проведених досліджень використано у навчальному процесі та в автоматизованій системі геолого-маркшейдерського забезпечення при плануванні та керуванні гірничих робіт у кар'єрі

Ключові слова: регресійна модель; автокореляційна функція; метод найменших квадратів; метод Лагранжа; каркасна модель; критичний радіус кореляції; індикатриса