Data processing for mathematical model building with taking into account heteroscedasticity

Valeriyi Kuzmin[1,†], Maksym Zaliskyi[1,*,†], Onyedikachi Chioma Okoro[2,†] and Oleksandr Bondarev[1,†]

[1] State University "Kyiv Aviation Institute", Liubomyra Huzara Ave., 1, Kyiv, 03058, Ukraine
[2] Vamooose Technologies, Calgary, Canada

Abstract

The paper concentrates on a comparative analysis of the use of various mathematical models, as well as a justification for choosing the best of them. To limit the types of possible models, the unacceptability of utilizing a linear approximating function was proved. Therefore, functions based on parabolas of the second degree and an exponential function were chosen as approximating functions, namely: general parabola, conditional parabola, conditional parabola with heteroscedasticity, conditional parabola built through cluster centers, exponential functions, and two-segmented linear-quadratic regression. To obtain specific equation, ordinary and weighted least squares methods were used. To account for heteroskedasticity, a new approach is considered for determining both a quantitative measure of heteroscedasticity in the analyzed data and calculating heteroscedasticity weighting coefficients for each empirical point. For two-segmented linear-quadratic regression, the abscissa of the switching point was optimized, which made it possible to obtain the best approximating function. A single analytical expression for the two-segmented linear-quadratic regression was obtained using the Heaviside function. Two-segmented linear-quadratic regression was substantiated as the best mathematical model.

Keywords

mathematical model, ordinary least squares, weighted least squares, heteroscedasticity, two-segmented regression, switching point

1. Introduction

The mathematical models building is an important branch of scientific activity [1]. It allows to establish a functional dependence between two or more variables [2, 3], explains natural phenomena [4, 5], the functioning of technical devices [6, 7], social and economic processes [8, 9], and others.

To build mathematical models, scientists use theoretical and experimental research [10]. In the first case, the basis is the laws of physics, mathematics, and other fundamental sciences. In the second case, it is necessary to rely on the means of planning and conducting experiments, as well as methods of computer science and mathematical statistics [11, 12].

When mathematical models building, scientists try to find a compromise between the accuracy and simplicity of describing the phenomenon under study [13]. In general, the factors that should be taken into account when building mathematical models are shown in Figure 1.

The mathematical models building for data obtained through experimentation is widely studied today in various technical fields, in particular, in econometrics [14, 15], telecommunications [16, 17], radio engineering [18, 19], control theory [20, 21], cybersecurity [22, 23], maintenance theory [24, 25], artificial intelligence application [26, 27], and other.

WDA'25: International Workshop on Data Analytics, January 6, 2025, Kyiv, Ukraine
* Corresponding author.
† These authors contributed equally.
valeriyikuzmin@gmail.com (V. Kuzmin); maximus2812@ukr.net (M. Zaliskyi); okorokachi7@gmail.com (O.C. Okoro); alexbondarev1990@gmail.com (O. Bondarev)
0000-0003-4461-9297 (V. Kuzmin); 0000-0002-1535-4384 (M. Zaliskyi); 0000-0001-5968-0424 (O.C. Okoro); 0009-0006-5391-8829 (O. Bondarev)

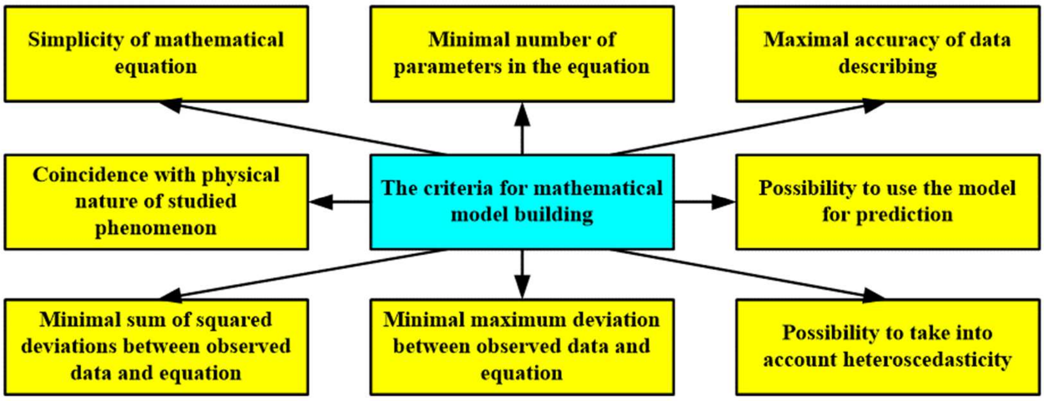

Figure 1: The criteria for building mathematical models.

The classical means of mathematical models building is the method of ordinary least squares (OLS). However, this method has a number of limitations, in particular, the assumption of a Gaussian distribution of errors and the constancy of the standard deviation [28]. Therefore, alternative methods of regression analysis currently exist [29]. These methods include:

- least absolute deviation regression;
- lasso regression;
- ridge regression;
- weighted least squares (WLS);
- elastic regression, and others [30].

The literature also presents a large number of heuristic engineering approaches [31, 32]. One such approach is the median center method. To implement this method, the initial data are divided into groups (clusters), in each of which there are medians. Then a smooth curve is drawn through the obtained median centers by the OLS method. This method allows to take into account bias and to reduce the impact of outliers. At the same time, there are other methods of approximation, for example, minimization of the maximum absolute deviation, minimization of Mahalanobis distance, minimization of the range for cumulative curve of deviations, and others [33, 34]. The use of such methods and approaches is the basis for building adequate mathematical models. Choosing the best mathematical model is a very complex scientific and technical task, and it will become more complicated and sophisticated because new approaches and methods of approximation of empirical data are developed [35]. Different scientific publications generally consider various approaches to approximate empirical data but heteroscedasticity was not taken into account during the study.

Recently, the applied use of the theory of approximation during the mathematical models building is developing at an active pace, namely methods of accounting for heteroscedasticity and for the bias that arises as a result of not fulfilling the condition of constant variance of empirical data within a data band. The theoretical base for detecting heteroscedasticity is sufficiently described in [36, 37] but practical methods are still absent.

It should be noted that heteroscedasticity taking into account often improves the accuracy of approximation, so heteroscedasticity analysis should be considered a necessary tool for building reliable mathematical models. However, the detection and substantiation of heteroscedasticity is a complex scientific and practical task. The analysis of the literature and the practice of mathematical models building shows that at the present time insufficient attention is paid to the issues of heteroscedasticity calculation.

Therefore, this paper describes several approaches to solving the recent scientific and technical problem of data approximation in the conditions of heteroscedasticity. The aim of the paper is to present step-by-step methodology of mathematical models building in case of heteroscedasticity for specific numerical example.

2. Materials and methods

When processing data on the strength of various materials, the tasks of mathematical models building and choosing the best ones arise. The analysis shows that these empirical data in most cases have heterogeneous variance, which leads to the need to account for heteroscedasticity. The use of the OLS as the approximation toll of empirical data leads to the significant errors of mathematical models since it does not take into account the real physical and natural conditions of data change.

Initial data on the dependence of the hardness y of various types of wood on its average density x (kg/m^3) [38] are given in Table 1. The sample size of data is $n = 36$.

Table 1
Initial dataset

x	y	x	y	x	y	x	y
24.7	484	38.5	914	42.9	1270	57.3	2020
24.8	427	38.5	1070	45.8	1180	57.6	1980
27.3	413	39.3	1020	46.9	1400	59.2	2310
28.4	517	39.4	1210	48.2	1760	59.8	1940
28.4	549	39.9	989	51.5	1710	66	3260
29	648	40.3	1160	51.5	2000	67.4	2700
30.3	587	40.6	1010	53.4	1880	68.8	2890
32.7	704	40.7	1100	56	1980	69.1	2740
35.6	979	40.7	1130	56.5	1820	69.1	3140

During the preprocessing we need to check the possibility of approximating the initial data using a linear function. For this, we use the linearity test. Calculation of unknown coefficients of approximation was carried based on OLS. As a result, the equation of the following form was obtained

$$y(x) = -1159 + 57.478x. \quad (1)$$

The initial data and their approximation using linear function and OLS are shown in Figure 2. Visual analysis of obtained dependence shows that the linear function drops significantly below zero, which does not correspond to the physical nature of the empirical data under study.

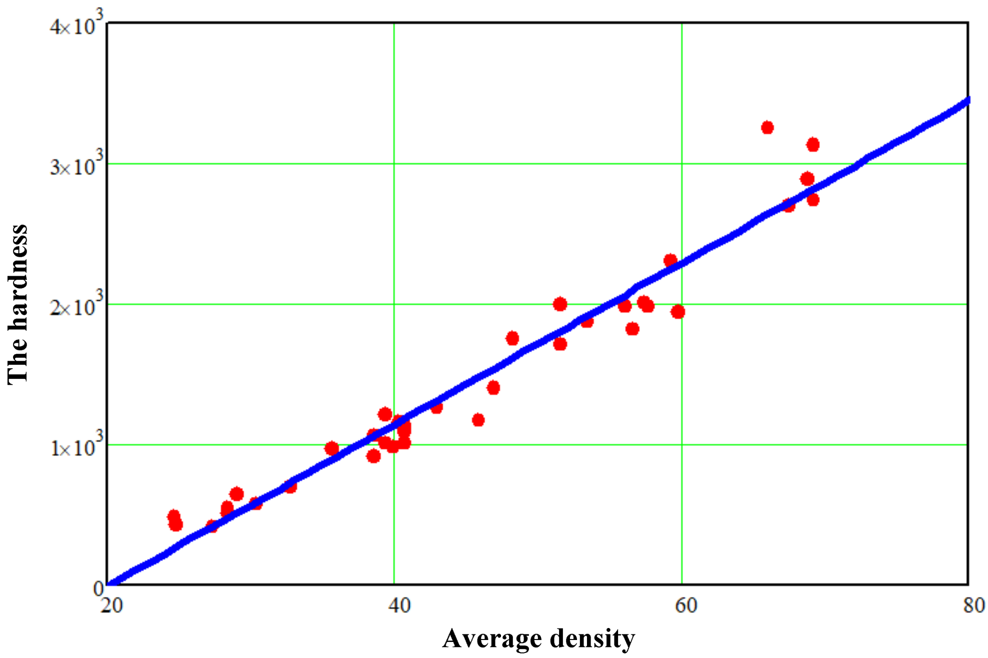

Figure 2: Data approximation using linear function.

For a justified statistical conclusion, a cumulative curve of residuals was calculated, the graphic representation of which is presented in Figure 3. The range of the cumulative residual curve is 1685 with a standard deviation of 182.75. The ratio of the range of the cumulative residual curve to the standard deviation is 9.22, which with a confidence probability of 0.99 indicates that the data under study cannot be described by the linear function [39].

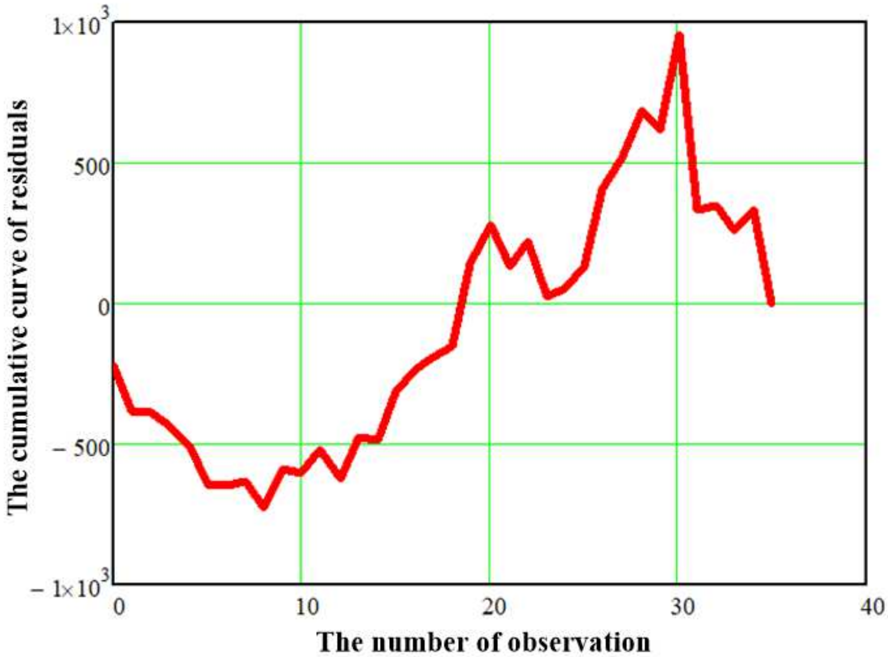

Figure 3: The cumulative curve of residuals for model (1).

Let's approximate the data by a general parabola of second degree using the OLS method. We get the equation of the form

$$y(x) = -115.888 + 9.373x + 0.5094x^2. \quad (2)$$

Approximation of the data by a general parabola of the second degree is presented in Figure 4.

For the resulting equation (2), the sum of the squares of the deviations is 161.379. To refine the linearity test, let's recalculate the ratio of the range of the cumulative residual curve to the standard deviation, which in this case is 19.6. This shows that with a confidence probability of 0.999 the data under study are significantly nonlinear.

As we can see from the graph in Figure 4, the parabola passes almost through the origin of the coordinates, so it could be replaced by a conventional one that can be described using equation $y(x) = bx + cx^2$. In this case, we get the equation

$$y(x) = 4.275x + 0.561x^2. \quad (3)$$

For the resulting equation, the sum of the squares of the deviations is 159.277.

Let's consider three variants of taking into account the heteroscedasticity.

1. The method of sliding regressions.

Let us perform a sliding approximation of a group of data with a size of 12 points using a parabola of the second degree by the OLS method. The width of the sliding window was chosen from the following considerations:

- visual analysis of the initial data made it possible to distinguish clusters with a number of points from 6 to 9;
- it is desirable to use data from two adjacent clusters so that there is no sudden change in the structure of the approximating function.

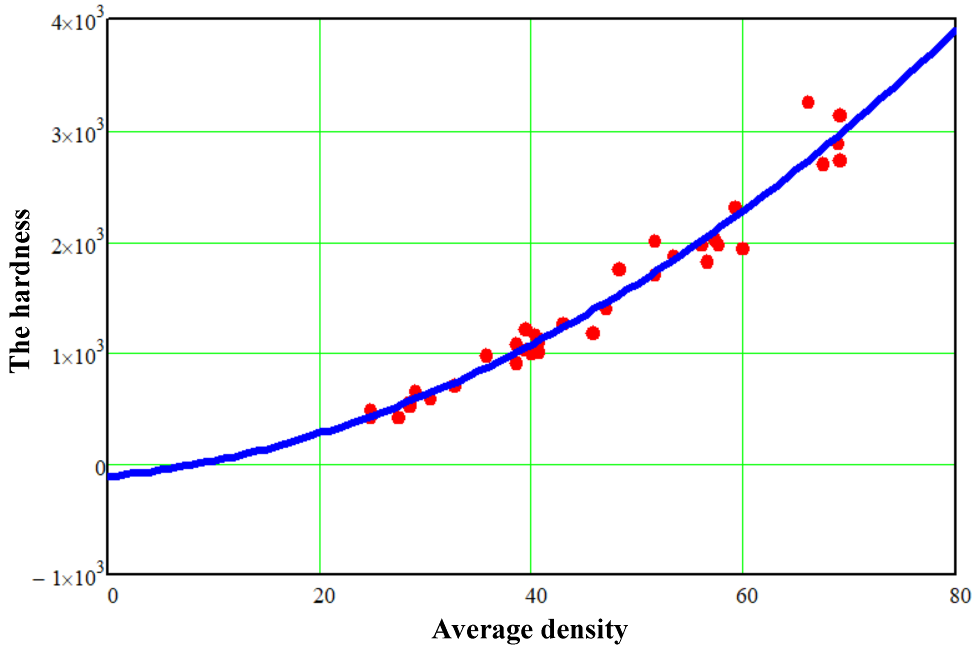

Figure 4: Data approximation using parabola.

Standard deviations σ_i and average values $\bar{y}_i$ along the ordinate were found for each approximation variant. Two statistics of 25 points each were obtained. To construct the heteroscedasticity equation, this data was approximated by a linear function using the OLS method. As a result, the equation was obtained

$$\sigma_i\,(\bar{y}_i) = -0.09864 + 0.1052\bar{y}_i. \tag{4}$$

The resulting equation can be used to calculate a system of weighting coefficients when approximating the initial data by a function of any type.

We can find the current weighting coefficients for each empirical point using the formula

$$W_i = \left(\frac{\bar{\sigma}}{\sigma_i\,(\bar{y}_i(x_i))}\right)^2, \tag{5}$$

where $\bar{\sigma}$ is the average value of standard deviations.

As a result, the equation of the general parabola was obtained, taking into account heteroscedasticity coefficients

$$y(x) = -102.01 + 8.471x + 0.522x^2. \tag{6}$$

As we can see from the obtained equation (6), the free coefficient decreased in absolute value compared to the corresponding coefficient of equation (2). For the resulting equation, the weighted sum of the deviations is 160.011. The result of the approximation is shown in Figure 5.

2. A new general approach to approximation taking into account heteroscedasticity.

To construct the best general parabola of the second degree, we will use the following scheme for determining the weighting coefficients for each empirical point

$$W_i = \left(\frac{\bar{y}}{y(x_i)}\right)^{2\alpha}, \tag{7}$$

where α is the heteroscedasticity parameter, $\bar{y}$ is the average strength value for the entire sample, $y(x_i)$ is the current value calculated according to equation (2).

To obtain the best weighted parabola, it is necessary to find the optimal value of the heteroscedasticity parameter, which characterizes the structure of the initial empirical data. This will be done by calculating the weighted sum Δ of squares deviations for seven variants of the parameter α values. Such approach of optimization is suitable for numerical tasks [40]. The results of the calculations are shown in Table 2.

Table 2
Data on the dependence of the weighted sum of squares of deviations on the parameter α

α	Δ
0	161.379
0.25	159.96
0.5	158.979
0.75	158.969
1	158.955
1.25	159.187
1.5	159.707

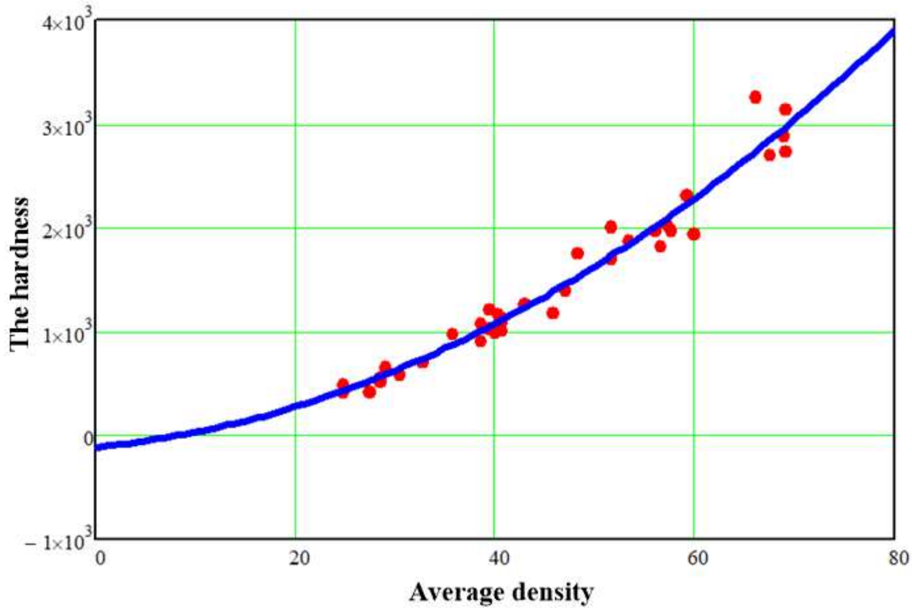

Figure 5: Data approximation using parabola and taking into account heteroscedasticity.

To find the optimal value of parameter α_{opt} based on these data, we will construct a parabola of the second degree using the OLS. As a result, we get the parabola of the form

$$\Delta(\alpha) = 161.379 - 0.814\alpha + 0.408\alpha^2. \tag{8}$$

For this data set and this type of approximating function, we obtain

$$\alpha_{\text{опт}} = -\frac{-0.814}{2 \cdot 0.408} = 0.998. \tag{9}$$

As a result, the optimal equation of the general parabola was obtained taking into account the optimal indicator of heteroscedasticity

$$y(x) = -1.273 + 0.978x + 0.507x^2. \tag{10}$$

For this equation (10), the weighted sum of the deviations is 160.951.

The optimal equation for the conditional parabola will have the following form:

$$y(x) = 3.765x + 0.571x^2. \tag{11}$$

For the resulting equation, the weighted sum of the deviations is 159.358.

3. The cluster method for obtaining the heteroscedasticity equation.

This algorithm can be described by the following sequence:

1. Determining the conditional parabola over the entire set of points. Equations of the conditional parabola using the OLS method for the studied set of data have the form (3). For the resulting equation, the weighted sum of the deviations is 159.277.

2. Division of the entire set of data into m compact clusters. In this case, there will be four clusters, and their grouping is shown in Figure 6.
3. For each cluster, we find the cluster center.
4. The centers of the clusters are approximated by the conditional parabola of the second degree.

In this case, the following equation is obtained

$$y(x) = 2.623x + 0.596x^2. \tag{12}$$

Let's perform an analysis of the logarithms of the initial data, which is shown in Figure 7.

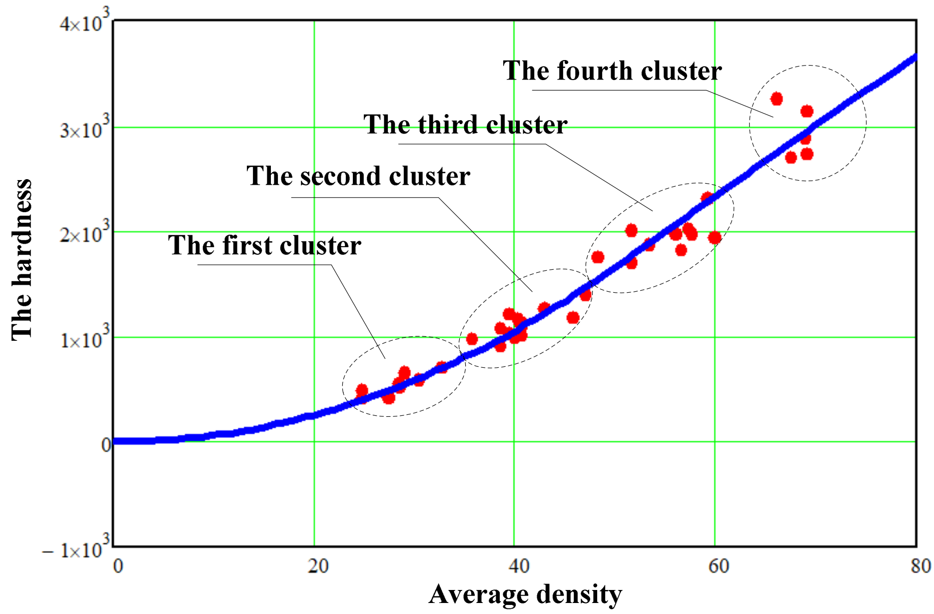

Figure 6: Clusters formation and data approximation using conditional parabola and taking into account heteroscedasticity.

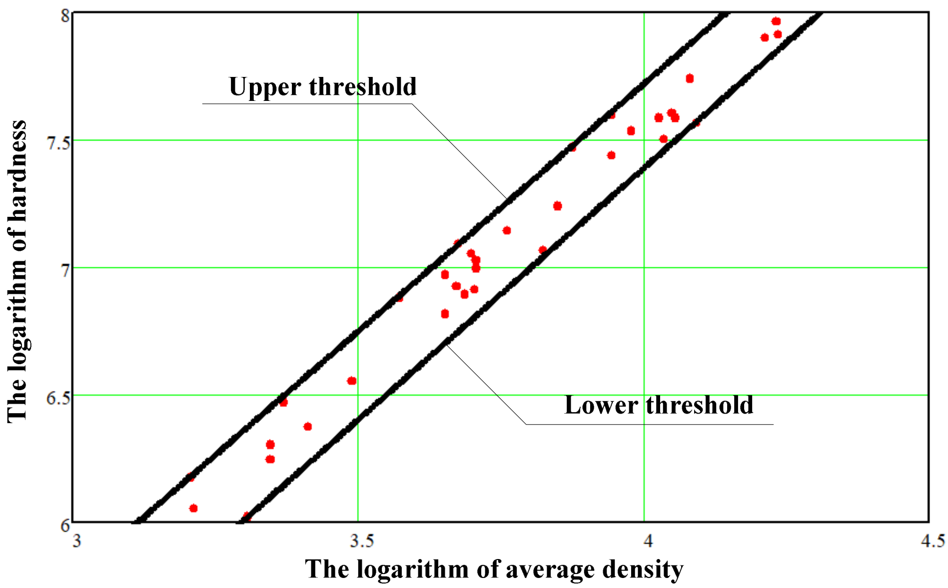

Figure 7: Confidence intervals for logarithms of initial data.

Figure 7 shows the confidence intervals that have been calculated. Visual analysis of the logarithmic values of the initial data allows to make a conclusion about the absence of heteroscedasticity in the logarithms of the values.

For the case of approximation using power function and the OLS method, the following equation was obtained

$$y(x) = 1.0195x^{1.8838}. \tag{13}$$

For the equation (13), the weighted sum of the deviations is 159.307.

We also performed the approximation of the initial data using the tool of segmented regression analysis. To build a more correct mathematical model, we used two segments: quadratic and linear. The equation of the approximating function in its general form has the form:

$$y(x) = ax + bx^2 - b(x - x_{sw})^2 h(x - x_{sw}), \tag{14}$$

where a, b are the regression coefficients, x_{sw} is the abscissa of switching point, $h(x)$ is the Heaviside function.

In the equation (14), the third coefficient is equal to the second coefficient, but it is opposite in sign. This condition ensures the linearity of the second segment.

In this case, the problem of finding the optimal abscissa of the switching point arises. This problem is solved by calculating standard deviations for several variants of values [41]. The obtained values x_{sw} are approximated by the parabola of the second degree with use of the OLS method. The minimum of this parabola determines the optimal abscissa of the switching point, which is 52.706.

As a result, the resulting final equation has the form:

$$y(x) = -1.927x + 0.692x^2 - 0.692(x - 52.706)^2 h(x - 52.706). \tag{15}$$

For the resulting equation (15), the weighted sum of the deviations is 157.981.

3. Comparative analysis of mathematical models

We will perform a comparative analysis of the obtained mathematical models. The results of the comparison are shown in Figure 8.

A comparative analysis of the obtained equations and standard deviations is shown in Table 3.

Table 3

Comparative analysis of the obtained equations and standard deviations

Number	Mathematical model	Equation	Standard deviation
1	Parabola	$y(x) = -115.888 + 9.373x + 0.5094x^2$	161.379
2	Conditional parabola	$y(x) = 4.275x + 0.561x^2$	159.277
3	Parabola with taking into account heteroscedasticity	$y(x) = 3.765x + 0.571x^2$	159.358
4	Cluster approach	$y(x) = 2.623x + 0.596x^2$	160.9
5	Power function	$y(x) = 0.78x^{1.884}$	159.307
6	Linear-quadratic	$y(x) = -1.927x + 0.692x^2 - -0.692(x - 52.706)^2 h(x - 52.706)$	157.981

The analysis shows that the cluster method is the worst both from the point of view of predictive properties and from the point of view of the value of the standard deviation. The general parabola has the negative free coefficient, which does not correspond to the physical nature of the phenomenon under study. In addition, this model does not take into account heteroscedasticity.

The lack of the general parabola in the form of the negative free coefficient is eliminated by using the conditional parabola. However, this parabola also does not take into account heteroscedasticity.

Of all the mathematical models with the use of parabolas, the best one is the conditional with taking into account heteroskedasticity. However, it has unsatisfactory predictive properties.

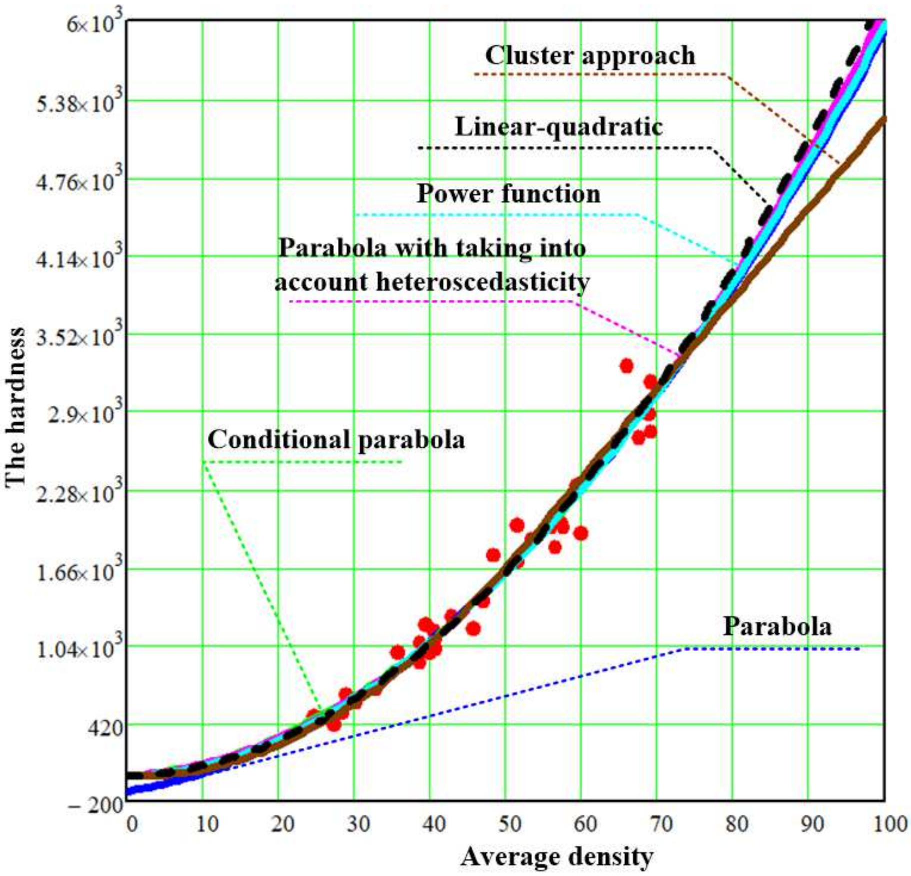

Figure 8: The results of approximation using different functions and approaches.

Approximation using the exponential function has an acceptable standard deviation but unsatisfactory predictive properties (sharp growth). The use of the exponential model according to formula (13) assumes that within the band of data the law of probability distribution is logarithmically normal, however, reliable proof of this statement is impossible to make for the researched dataset. The use of an exponential function always has underestimated predictive values compared to the conditional parabola.

The best mathematical model is the optimal two-segmented model consisting of segments with the conditional parabola and linear function. This is explained both by the minimum value of the standard deviation and by the best predictive properties.

4. Conclusions

The paper considers the issues of mathematical models building and choosing the best of them. Six variants of approximation were studied with a preliminary analysis of the data for linearity. As approximating functions, the following were used: general parabola, conditional parabola, conditional parabola with heteroscedasticity, conditional parabola built through cluster centers, exponential function, and two-segmented linear-quadratic regression.

Alternative variants of taking into account the heteroscedasticity are discussed in the paper. At the same time, a new approximation approach is proposed, which involves the calculation of the heteroscedasticity parameter for correct definition of the approximating function.

For two-segmented linear-quadratic regression, optimization of the abscissa of the switching point was performed to find the best option from the point of view of minimizing the standard deviation. The comparative analysis showed that the two-segmented linear-quadratic regression is the best among the considered mathematical models. This is explained by the minimal standard deviation and the best predictive properties.

Acknowledgements

This research is partially supported by the Ministry of Education and Science of Ukraine under the projects "Methods of building protected multilayer cellular networks 5G / 6G based on the use of artificial intelligence algorithms for monitoring critical infrastructure objects of country" (# 0124U000197) and is partially supported by EURIZON project # 871072 (Project EU #3035 EURIZON "Research and development of Ukrainian ground network of navigational aids for increasing the safety of civil aviation").

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] J. Stark, Product Lifecycle Management, Volume 1: 21st Century Paradigm for Product Realisation, Springer, London, 2019. doi: 10.1007/978-3-030-98578-3.

[2] M. Ezekiel, K. A. Fox, Method of Correlation and Regression Analysis. Linear and Curvilinear, Wiley, New York, 1959.

[3] N. Kuzmenko, et al., Airplane flight phase identification using maximum posterior probability method, in: Proceedings of IEEE 3rd International Conference on System Analysis & Intelligent Computing (SAIC), IEEE, Kyiv, Ukraine, 2022, pp. 1–5. doi: 10.1109/SAIC57818.2022.9922913.

[4] J. Andre, Industry 4.0: Paradoxes and Conflicts, Wiley, New York, 2019.

[5] A. Zaporozhets, V. Babak, V. Isaienko, K. Babikova, Analysis of the air pollution monitoring system in Ukraine. In: V. Babak, V. Isaienko, A. Zaporozhets, (Eds.), Systems, Decision and Control in Energy I. Studies in Systems, Decision and Control, vol 298, Springer, Cham, 2020, pp. 85–110. doi: 10.1007/978-3-030-48583-2_6.

[6] V. Larin, et al., Prediction of the final discharge of the UAV battery based on fuzzy logic estimation of information and influencing parameters, in: Proceedings of IEEE 3rd KhPI Week on Advanced Technology (KhPIWeek), IEEE, Kharkiv, Ukraine, 2022, pp. 1–6. doi: 10.1109/KhPIWeek57572.2022.9916490.

[7] K. Dergachov, et al., GPS usage analysis for angular orientation practical tasks solving, in: Proceedings of IEEE 9th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T), IEEE, Kharkiv, Ukraine, 2022, pp. 187–192. doi: 10.1109/PICST57299.2022.10238629.

[8] M. Zaliskyi, et al., Methodology for substantiating the infrastructure of aviation radio equipment repair centers, CEUR Workshop Proceedings 3732 (2024) 136–148. URL: https://ceur-ws.org/Vol-3732/paper11.pdf.

[9] S. Smerichevskyi, Z. Poberezhna, O. Mykhalchenko, Y. Petrova, V. Smilyanets, Ensuring the sustainable development of the aviation enterprise in the context of forming innovative potential using digital technologies, CEUR Workshop Proceedings 3732 (2024) 174–185. URL: https://ceur-ws.org/Vol-3732/paper14.pdf.

[10] Th. P. Ryan, Modern Regression Methods, Second Edition, Wiley, New York, 1997.

[11] F. C. Mills, Statistical Methods, Pitman Publishing, New York, 1965.

[12] D. M. Himmelblau, Process Analysis by Statistical Methods, Wiley, New York, 1970.

[13] M. Zaliskyi, et al., Model building for diagnostic variables during aviation equipment maintenance, in: Proceedings of IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT), IEEE, Lviv, Ukraine, 2022, pp. 160–164. doi: 10.1109/CSIT56902.2022.10000556.

[14] D. Gujarati, Econometrics by Example, Red Globe Press, London, 2014.

[15] Z. Poberezhna. Formation of a system for optimizing business processes of aviation enterprises based on their automation, in: I. Ostroumov, M. Zaliskyi (Eds.), Proceedings of the 2nd International Workshop on Advances in Civil Aviation Systems Development. ACASD 2024,

volume 992 of Lecture Notes in Networks and Systems, Springer, Cham, 2024, pp. 325–338. doi: 10.1007/978-3-031-60196-5_24.
[16] R. S. Odarchenko, S. O. Gnatyuk, T. O. Zhmurko, O. P. Tkalich, Improved method of routing in UAV network, in: Proceedings of International Conference Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD), IEEE, Kyiv, Ukraine, 2015, pp. 294–297. doi: 10.1109/APUAVD.2015.7346624.
[17] O. Lavrynenko, R. Odarchenko, G. Konakhovych, A. Taranenko, D. Bakhtiiarov, T. Dyka, Method of semantic coding of speech signals based on empirical wavelet transform, in: Proceedings of IEEE 4th International Conference on Advanced Information and Communication Technologies (AICT), IEEE, Lviv, Ukraine, 2021, pp. 18–22. doi: 10.1109/AICT52120.2021.9628985.
[18] I. Zhukov, B. Dolintse, S. Balakin, Enhancing data processing methods to improve UAV positioning accuracy, International Journal of Image, Graphics and Signal Processing 16 (3) (2024) 100–110. doi: 10.5815/ijigsp.2024.03.08.
[19] O. Solomentsev, et al., Data processing through the lifecycle of aviation radio equipment, in: Proceedings of IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT), IEEE, Lviv, Ukraine, 2022, pp. 146–151. doi: 10.1109/CSIT56902.2022.10000844.
[20] O. Sushchenko, et al., Airborne sensor for measuring components of terrestrial magnetic field, in: Proceedings of IEEE 41st International Conference on Electronics and Nanotechnology (ELNANO), IEEE, Kyiv, Ukraine, 2022, pp. 687–691. doi: 10.1109/ELNANO54667.2022.9926760.
[21] O. Sushchenko, et al., Integration of MEMS inertial and magnetic field sensors for tracking power lines, in: Proceedings of IEEE XVIII International Conference on the Perspective Technologies and Methods in MEMS Design (MEMSTECH), IEEE, Polyana, Ukraine, 2022, pp. 33–36. doi: 10.1109/MEMSTECH55132.2022.10002907.
[22] M. Zaliskyi, R. Odarchenko, S. Gnatyuk, Y. Petrova, A. Chaplits, Method of traffic monitoring for DDoS attacks detection in e-health systems and networks, CEUR Workshop Proceedings 2255 (2018) 193–204. URL: https://ceur-ws.org/Vol-2255/paper18.pdf.
[23] R. Odarchenko, O. Lavrynenko, D. Bakhtiiarov, S. Dorozhynskyi, V. A. O. Zharova, Empirical wavelet transform in speech signal compression problems, in: Proceedings of IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T), IEEE, Kharkiv, Ukraine, 2021, pp. 599–602. doi: 10.1109/PICST54195.2021.9772156.
[24] I. Ostroumov, et al., Relative navigation for vehicle formation movement, in: Proceedings of IEEE 3rd KhPI Week on Advanced Technology (KhPIWeek), IEEE, Kharkiv, Ukraine, 2022, pp. 1–4. doi: 10.1109/KhPIWeek57572.2022.9916414.
[25] O. Solomentsev, et al., Efficiency analysis of current repair procedures for aviation radio equipment, in: I. Ostroumov, M. Zaliskyi (Eds.), Proceedings of the 2nd International Workshop on Advances in Civil Aviation Systems Development. ACASD 2024, volume 992 of Lecture Notes in Networks and Systems, Springer, Cham, 2024, pp. 281–295. doi: 10.1007/978-3-031-60196-5_21.
[26] O. Holubnychyi, et al., Self-organization technique with a norm transformation based filtering for sustainable infocommunications within CNS/ATM systems, in: I. Ostroumov, M. Zaliskyi (Eds.), Proceedings of the 2nd International Workshop on Advances in Civil Aviation Systems Development. ACASD 2024, volume 992 of Lecture Notes in Networks and Systems, Springer, Cham, 2024, pp. 262–278. doi: 10.1007/978-3-031-60196-5_20.
[27] O. Tachinina, O. Lysenko, I. Alekseeva, V. Novikov, Mathematical modeling of motion of iron bird target node of security data management system sensors, CEUR Workshop Proceedings 2711 (2020) 482–491. URL: https://ceur-ws.org/Vol-2711/paper37.pdf.
[28] S. Weisberg, Applied Linear Regression, Wiley, New York, 2005.
[29] D. Birkes, Y. Dodge, Alternative Methods of Regression, Wiley, New York, 1993.
[30] M. Gopal, Applied Machine Learning, McGraw Hill Education, India, 2018.
[31] G. Snedecor, W. Cochran, Statistical Methods, Iowa State University Press, Iowa, 1989.

[32] D. C. Montgomery, G. C. Runger, Applied Statistics and Probability for Engineers, Fours Edition, Wiley, New York, 2007.
[33] G. A. F. Seber, C. J. Wild, Nonlinear Regression, Wiley, New York, 2003.
[34] S. Huet, A. Bouvier, M.-A. Poursat, E. Jolivet, Statistical Tools for Nonlinear Regression. A Practical Guide With S-PLUS and R Examples, Springer-Verlag, New York, 2004.
[35] Z. Poberezhna, Y. Petrova, K. Slimani, Information technologies in logistics processes of enterprises in the aviation industry, CEUR Workshop Proceedings 3732 (2024) 90–102. URL: https://ceur-ws.org/Vol-3732/paper07.pdf.
[36] R. L. Kaufman, Heteroskedasticity in Regression, SAGE Publications, 2013.
[37] M. Zaliskyi, et al., Heteroskedasticity analysis during operational data processing of radio electronic systems, in: S. Shukla, A. Unal, J. Varghese Kureethara, D.K. Mishra, D.S. Han (Eds.), Data science and security, volume 290 of Lecture Notes in Networks and Systems, Springer, Singapore, 2021, pp. 168–175. doi: 10.1007/978-981-16-4486-3_18.
[38] A. Dlin, Factor Analysis in Production, Statistics, 1975.
[39] V. N. Kuzmin, New statistical method for identification of nonlinearity of empirical data, in: Proceedings of Fifth International Conference on Computer Data Analysis and Modeling, Volume 1: A-M, 1998, pp. 159–164.
[40] G. V. Reklaitis, A. Ravindran, K. M. Ragsdell, Engineering Optimization. Methods and Applications, Wiley, New York, 1983.
[41] J. Al-Azzeh, A. Mesleh, M. Zaliskyi, R. Odarchenko, V. Kuzmin, A method of accuracy increment using segmented regression, Algorithms 15(10) (2022) 1–24. doi: 10.3390/a15100378.