Obrabotka metallov (tekhnologiya, oborudovanie, instrumenty) = Metal Working and Material Science. 2026 vol. 28 no. 1 pp. 64–80
ISSN: 1994-6309 (print) / 2541-819X (online)
DOI: 10.17212/1994-6309-2026-28.1-64-80

NSTU NETI

Obrabotka metallov -
Metal Working and Material Science

Journal homepage: http://journals.nstu.ru/obrabotka_metallov

Minimum sample size requirements for reliable correlation-regression modeling of surface roughness in milling

Mikhail Gimadeev [a, *], *Vadim Stelmakov* [b], *Maksim Uliskov* [c]

Pacific National University, 136 Tihookeanskaya St., Khabarovsk, 680035, Russian Federation

[a] https://orcid.org/0000-0001-6685-519X, 009063@togudv.ru; [b] https://orcid.org/0000-0003-2763-1956, 009062@togudv.ru;
[c] https://orcid.org/0009-0001-9858-423X, 2016104779@togudv.ru

ARTICLE INFO

Article history:
Received: 25 December 2025
Revised: 12 January 2026
Accepted: 19 January 2026
Available online: 15 March 2026

Keywords:
Sample size
Surface roughness
Correlation
Regression model
Milling
Root mean square error

Funding
This work has funded by the Ministry of science and higher education of Russian Federation (project № FEME–2024–0010).

ABSTRACT

Introduction. Investigating the statistical relationships between milling parameters and surface roughness requires a correct selection of the sample size, as amplitude parameters and form parameters respond differently to data limitations. The reliability of correlation-regression analysis depends on meeting the assumptions of normality and the stability of the estimates, making the determination of the minimum number of observations essential for constructing reliable surface roughness models. ***The purpose of this work*** is to develop a methodology for estimating the minimum sample size required to build statistically significant correlation-regression models that describe the relationship between the technological parameters of the milling process and surface roughness characteristics, ensuring statistical reliability of the results and enabling accurate prediction of machined surface quality. **Methodology.** The normality of the surface roughness parameter distributions after milling was assessed using the *Shapiro–Wilk*, *Anderson–Darling*, and *Pearson's* chi-squared tests. Multicollinearity among the technological factors was analyzed using the variance inflation factor (*VIF*), while the adequacy of the regression models was verified using the mean absolute error (*MAE*) and root mean square error (*RMSE*) metrics. The minimum required sample size was determined by considering statistical test power and *Fisher*'s *z*-transformation. **Results and discussions.** The analysis of pairwise correlation coefficients revealed that amplitude roughness parameters exhibit stable relationships even at $n = 16$, while profile shape parameters (***Rsk***, ***Rku***) are characterized by weak or negative correlations and require a substantially larger sample size. The constructed matrix of the minimum number of observations confirms that for a number of relationships, especially those involving ***Rsk***, hundreds of measurements are necessary, which justifies the selection of the most informative parameter combinations. Increasing the sample size to $n = 128$ reduces estimation bias, stabilizes the correlations for amplitude parameters, and reveals a weakening of the relationships for ***Rsk*** and ***Rku***. Rank correlation analysis confirmed a monotonic dependence between ***Rz*** and ***Rt*** and the independence of ***Rsk*** from amplitude characteristics. The verification of the normality of the distributions and the absence of multicollinearity among the factors ensured the validity of the constructed regression models, which demonstrated high accuracy and consistent error behavior.

For citation: Gimadeev M.R., Stelmakov V.A., Uliskov M.V. Minimum sample size requirements for reliable correlation-regression modeling of surface roughness in milling. *Obrabotka metallov (tekhnologiya, oborudovanie, instrumenty) = Metal Working and Material Science*, 2026, vol. 28, no. 1, pp. 64–80. DOI: 10.17212/1994-6309-2026-28.1-64-80. (In Russian).

Introduction

Investigation of the relationship between milling process parameters and surface roughness is key to ensuring the required machining quality and predicting the performance of machine parts [1]. Surface roughness parameters, including asymmetry (skewness) (***Rsk***) and profile excess (kurtosis) (***Rku***), have a significant impact on tribological properties [2–4].

Traditional parameters ***Ra*** and ***Rz*** are insufficient for a complete assessment of the tribological properties of surfaces, since they do not reflect the features of the distribution of microroughnesses. In contrast, the

*** Corresponding author**
Gimadeev Mikhail R., Ph.D. (Engineering), Associate Professor
Pacific National University,
136 Tihookeanskaya st.,
680035, Khabarovsk, Russian Federation
Tel.: +7 924 216-31-39, **e-mail:** 009063@togudv.ru

profile shape parameters ***Rsk*** and ***Rku*** show a more stable relationship with friction and wear characteristics, which is confirmed by several studies [5–7]. These parameters are considered to be the most informative for working surfaces operating under sliding friction conditions, for base and supporting surfaces, etc. It is noted that negative values of ***Rsk***, as a rule, are combined with increased values of ***Rku***, forming a microrelief that helps to reduce the friction coefficient [8], even with relatively high values of the amplitude parameters of roughness.

Statistical methods are used for quantitative analysis of parameter dependencies, among which correlation analysis remains one of the most widely used [6–9]. However, it is important to take into account that correlation coefficients, especially the *Pearson* coefficient, are often applied to an extremely limited number of observations, which reduces the reliability of the results.

The correctness of the application of parametric methods (*Pearson's* correlation coefficient, linear regression, etc.) presupposes the conformity of the distribution of analyzed variables and residuals of the regression models to a normal law. This is because statistical tests of significance, confidence intervals, and ***p***-values in parametric methods are derived under the assumption of normality. In the case of significant deviations from the normal distribution, the results of parametric methods may be incorrect, which requires preliminary testing of distributions or the use of alternative non-parametric ones. Several criteria are used to solve this problem: *Shapiro-Wilk's* [10], which is characterized by high power for medium sample sizes and allows reliable detection of deviations from the normal law; *Anderson-Darling's* [11], which has increased sensitivity to variation in distribution tails, making it particularly useful in analyzing data with rare but significant outliers; and *Pearson's*, providing an interval estimate of agreement between observed and expected distributions. This combination of criteria allows a comprehensive assessment of the normality of the distribution of parameters, ensures high reliability of statistical conclusions**,** and eliminates the risk of erroneous interpretation of data.

In modern studies, correlation analysis is supplemented by regression modeling [12], which allows one not only to evaluate the strength and direction of the relationship between variables, but also to construct equations for the dependence of process parameters on roughness indicators. At the same time, the analysis of residuals of regression models ensures verification of the adequacy of the obtained dependencies and clarification of the significance of individual factors [13]. An additional tool for increasing the reliability of regression analysis is the variance inflation factor (*VIF*) [14], which shows how much the dispersion of the regression coefficient estimate increases due to the correlation of the independent variable (predictor) with others, and is used as one of the key tools for identifying multicollinearity.

Particular attention in the field of metal cutting is paid to the issue of minimum sample size [15–17] required for correlation and regression analysis. Insufficient data can lead to model overfitting or distortion of conclusions, while excessive data can lead to unjustified resource costs. For technological milling processes, this problem has practical significance, since these operations are often used in machining, including of complex surfaces [18, 19], and require control of quality parameters. In this regard, an urgent task is to determine the optimal number of observations that ensures statistical reliability of the results when using *Pearson's* (***r***) [20] and *Spearman's* (**ρ**) correlation coefficients.

The purpose of this work is to develop an approach to determining the minimum sample size required to construct statistically significant correlation-regression models describing the relationship between the technological parameters of the milling process and the characteristics of surface roughness, ensuring the statistical reliability of the results and the ability to adequately predict the quality of the machined surface.

To achieve this purpose, the following ***tasks*** must be solved:

− to develop a methodological approach to determining the minimum required sample size necessary to ensure the statistical reliability of the results obtained;

− to analyze the influence of the number of observations on the stability of regression coefficients, the accuracy of approximation**,** and the magnitude of prediction errors;

− to formulate practical recommendations for selecting a sufficient sample size to solve applied problems of technological process analysis.

Research methodology

Methodology for statistical evaluation of data. In the study, normality is tested using several complementary criteria:

− The *Shapiro–Wilk* test allows one to establish the degree to which the sample conforms to the normal distribution through a linear combination of order statistics:

$$W = \frac{\left(\sum_{i=1}^{n} a_i y_{(i)}\right)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2},$$

where $\boldsymbol{y_{(i)}}$ are the ordered sample values; $\boldsymbol{y_i}$ is the observable value; $\boldsymbol{\bar{y}}$ is the average value; $\boldsymbol{a_i}$ are the coefficients depending on the covariance matrix of the normal distribution;

− The *Anderson–Darling* test is a modification of the *Kolmogorov–Smirnov* test that gives more weight to the tails of the distribution:

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n} (2i-1)\left[lnF\left(x_{(i)}\right) + ln\left(1 - F\left(x_{(n+1-i)}\right)\right)\right],$$

where $\boldsymbol{F(x)}$ is the normal cumulative distribution function;

− *Pearson's* chi-squared test is based on a comparison of observed and expected frequencies across intervals:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i},$$

where $\boldsymbol{O_i}$ is the observed frequency in the ***i**-th* interval; $\boldsymbol{E_i}$ is the expected frequency in the ***i**-th* interval; $\boldsymbol{k}$ is the number of intervals.

Multicollinearity of predictors remains a significant factor that can distort the results of the analysis. To identify and quantify it, the variance inflation factor (*VIF*) is used:

$$VIF_i = \frac{1}{1 - R_i^2},$$

where $\boldsymbol{R^2}$ is the coefficient of determination shows what proportion of the variation in the dependent variable is explained by the model based on the independent variables [21].

If we denote residuals as $e_i = y_i - \hat{y}_i$, where $\hat{y}_i$ are the predicted values, then:

$$R^2 = \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2}.$$

To simultaneously assess the quality of the explanation and the accuracy of the prediction, error metrics are used to verify the adequacy of the model [22]:

− Mean absolute error (*MAE*):

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |e_i|.$$

− Square root of mean square error (*RMSE*) [23]:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} |e_i|}.$$

Because the *RMSE* function is smooth and differentiable, optimal regression coefficients can be analytically derived. For a simple linear model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

where $\boldsymbol{x_i}$ is the predictor; $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$ are the model parameters; $\boldsymbol{\varepsilon_i}$ is the random error.

Residuals are defined as $e_i = y_i - (\beta_0 + \beta_1 x_i)$. The *RMSE* function is written as:

$$RMSE(\beta_0, \beta_1) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|e_i|}.$$

We denote:

$$S(\beta_0, \beta_1) = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \Rightarrow RMSE\sqrt{S}.$$

Then the *RMSE* gradient according to the model parameters obtained through the chain rule has the form:

$$\frac{\partial RMSE}{\partial \beta_j} = \frac{1}{2\sqrt{S}}\frac{\partial S}{\partial \beta_j}, \quad j \in \{0, 1\}.$$

$$\text{For } \beta_0\text{: } \frac{\partial S}{\partial \beta_0} = -\frac{2}{n}\sum_{i=1}^{n} e_i \Rightarrow \frac{\partial RMSE}{\partial \beta_0} = \frac{1}{n \cdot \mathrm{RMSE}}\sum_{i=1}^{n} e_i .$$

$$\text{For } \beta_1\text{: } \frac{\partial S}{\partial \beta_1} = -\frac{2}{n}\sum_{i=1}^{n} x_i\, e_i \Rightarrow \frac{\partial RMSE}{\partial \beta_1} = \frac{1}{n \cdot \mathrm{RMSE}}\sum_{i=1}^{n} x_i\, e_i .$$

Thus, the smoothness and differentiability of the *RMSE* function provide the possibility of analytically finding optimal regression coefficients and using gradient methods for their estimation.

To calculate the required sample size for a study, a formula is used based on the requirements for accuracy, confidence level, and expected variability of the data. The formula for an infinite population is:

$$n = \frac{Z^2 p(1-p)}{E^2},$$

where $\boldsymbol{n}$ is the sample size; $\boldsymbol{Z}$ is the quantile of the standard normal distribution corresponding to the confidence level; $\boldsymbol{p}$ is the expected proportion or probability.

$\boldsymbol{E}$ is the permissible error, which decreases with the number of observations:

$$E \propto \frac{Z_{1-\alpha/2} \cdot \sigma}{\sqrt{n}},$$

where $\boldsymbol{Z_{1-\alpha/2}}$ is the quantile corresponding to the significance level $\boldsymbol{\alpha}$; $\boldsymbol{\sigma}$ is the standard deviation.

When the population size ($\boldsymbol{N}$) is known, the sample size is reduced to account for the limited scale. Then a more precise formula for a finite population is used:

$$n = \frac{NZ^2 p(1-p)}{E^2(N-1) + Z^2 p(1-p)}.$$

When estimating correlation, the sample size also depends on the expected magnitude of the correlation ($\boldsymbol{r}$). The formula can be adapted to use the power of the test [16, 17] ($\boldsymbol{1-\beta}$) and level of significance ($\boldsymbol{\alpha}$):

$$n = \frac{\left(Z_{1-\alpha/2} - Z_{1-\beta}\right)^2}{r^2},$$

where $\boldsymbol{Z_{1-\beta}}$ is the quantile corresponding to the test power.

It should be noted that this formula gives only an approximate estimate and does not take into account the nonlinearity of the correlation coefficient distribution. A more accurate calculation of the sample size when testing a correlation hypothesis is usually performed on the basis of the *Fisher* transformation, which allows the correlation coefficient to be converted into a normally distributed value and thereby ensure the accuracy of the statistical estimate:

$$n = \left(\frac{Z_{1-\alpha/2} - Z_{1-\beta}}{\frac{1}{2} ln\left(\frac{1+r}{1-r}\right)} \right)^2 + 3.$$

At the same time, for a finite population, the following correction is applied:

$$n = \frac{n}{1 + \frac{n-1}{N}}.$$

In order to calculate the significance level and power quantiles, it is necessary to calculate a value such that a given proportion of the standard normal distribution lies below it. This is achieved using the percent point function (*PPF*), which is the inverse of the cumulative distribution function (*CDF*):

– for a continuous distribution: $F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$. Here $\boldsymbol{F(x)}$ is the *CDF*, the value of the distribution function for a point $\boldsymbol{x}$, $P\ (X \le x)$ is the the probability that $\boldsymbol{X}$ is less than or equal to x, $f(t)$ is the probability density function (*PDF*).

– for a discrete distribution: $F(x) = P(X \le x) = \sum_{t \le x} P(X = t)$. Here $P\ (X = t)$ is the probability of a specific value t.

For a standard normal distribution, with mean = 0 and the standard deviation = 1, the *CDF* value is calculated as:

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e_{\ln}^{-\frac{t^2}{2}}\,dt,$$

where $\boldsymbol{e_{ln}}$ is the base of natural logarithm.

The minimum sample size calculations above are based on *Pearson's* parametric correlation coefficient, since it assumes a linear relationship between variables and requires that the data be normally distributed. However, in practical applications, it is important to consider that the actual distribution of roughness parameters may vary depending on the sample size and measurement conditions.

Details of the experiment. Experimental studies were carried out on a *DMC 635V* machining center, where study objects – surface samples for roughness parameter analysis – were obtained. The objects were made of *AISI 321* stainless steel with a chemical composition in wt.%: $C \le 0.12$; $Si \le 0.8$; $Mn \le 2.0$; $P \le 0.035$; $S \le 0.02$; *Ni* 9–11; *Cr* 17–19; $Ti < 0.8$; *Fe* — balance. The cutting tool material used was carbide with a multilayer *PVD* coating and a fine-grained substrate.

When planning the experiment, a 2^3 full factorial design with two center points and two replicates was used, which ensured the reliability of the statistical assessment (Table 1). The following factors were considered: tool diameter ($\boldsymbol{D}$ = 6–12 mm), edge inclination angle (γ = 10–50°), and feed per tooth ($\boldsymbol{fz}$ = 0.1–0.5 mm/tooth). The center points were chosen to test the linearity or nonlinearity of the response, taking into account the applied parameters and increasing the reliability of the statistical estimate due to additional observations in the middle of the range.

This design allows obtaining representative data on the influence of milling parameters on the formation of surface roughness.

Table 1

Experiment matrix *n* = 16

No.	Tool diameter ***D***, mm	Inclination angle γ, °	Feed per tooth f_z, mm/tooth
Full factorial design 2^3			
1	6	10	0.1
2	6	10	0.5
3	6	50	0.1
4	6	50	0.5
5	12	10	0.1
6	12	10	0.5
7	12	50	0.1
8	12	50	0.5
Center points with two replicates			
9	8	30	0.3
10	10	30	0.3
11	8	30	0.3
12	10	30	0.3
Additional replicates of extreme combinations			
13	6	10	0.1
14	12	50	0.5
15	6	50	0.5
16	12	10	0.1

Roughness measurement and selection of base length. Roughness measurements after milling were carried out using a *SURFCOM 1800D* profilometer [20] with a 50% *Gaussian* filter. In general, the base length is selected based on ***Ra*** and ***Rz*** value ranges according to *ISO 4288:1996*. In this work, following the recommendations for measuring periodic profiles, the ***Rsm*** parameter was used with the following ranges: for 0.13 < ***Rsm*** ≤ 0.4 mm, the base length ***lr*** = 0.8 mm; for 0.4 < ***Rsm*** ≤ 1.3 mm, the base length ***lr*** = 2.5 mm, and the evaluation length ***L*** was **5*lr***.

It should be noted that the use of the ***Rz*** parameter to estimate the periodic surface profile is methodologically incorrect. This indicator reflects only the amplitude characteristics of the irregularities and does not take into account the regularity of the structure, which leads to distortion in the interpretation of the results. The chosen base length ***lr*** = 0.8 mm for all measurements could have increased the bias effect, reducing the reliability of the obtained data. Thus, the measurement results in the range 1.6 < ***Rz*** < 12.5 μm cannot be considered fully correct and require clarification using more adequate parameters for periodic profiles.

The presence of repeating elements in the roughness profile may be due to tool wear, vibrations, or various technological defects [19]. Therefore, before choosing a base length, it is necessary to ensure that the analyzed profile is periodic. In particular, in this work, in milling, the surface roughness profile is considered as a kinematic-geometric projection of the tool shape obtained in finishing machining operations.

Results and discussion

This section presents the results of a statistical analysis of surface roughness parameters, including an assessment of correlations, a sample size adequacy check, an analysis of profile shape parameter distributions, and the construction of regression models. This approach allows for a comprehensive assessment of the data structure, the identification of consistent relationships, and the determination of the informative value of individual parameters.

Table 2 presents the paired *Pearson* correlation coefficients between the surface roughness parameters for a sample size of n = 16. The obtained values demonstrate a high degree of interrelation between the amplitude parameters, which confirms their linear dependence and the commonality of the formation mechanisms. The parameters characterizing the shape of the profile (***Rsk*** and ***Rku***) exhibit negative or moderate correlations with amplitude characteristics, which indicates their independence and different sensitivity to the profile structure.

Table 2

Pearson correlation coefficients (α = 0.01, *n* = 16)

	R_a	R_q	R_z	R_p	R_t	R_c	R_v	R_{sm}	R_{sk}	R_{ku}
R_a	1	0.99	0.99	0.99	0.97	0.99	0.99	0.94	−0.22	−0.63
R_q	0.99	1	0.99	0.99	0.97	0.99	0.99	0.94	−0.23	−0.63
R_z	0.99	0.99	1	0.98	0.98	0.99	0.99	0.94	−0.25	−0.62
R_p	0.99	0.99	0.98	1	0.94	0.99	0.99	0.91	−0.12	−0.57
R_t	0.97	0.97	0.98	0.94	1	0.98	0.98	0.95	−0.43	−0.66
R_c	0.99	0.99	0.99	0.99	0.98	1	0.99	0.95	−0.25	−0.63
R_v	0.99	0.99	0.99	0.99	0.98	0.99	1	0.94	−0.25	−0.62
R_{sm}	0.94	0.94	0.94	0.91	0.95	0.95	0.94	1	−0.38	−0.81
R_{sk}	−0.22	−0.23	−0.25	−0.12	−0.43	−0.25	−0.25	−0.38	1	0.57
R_{ku}	−0.63	−0.62	−0.62	−0.57	−0.66	−0.63	−0.62	−0.81	0.57	1

To assess the sufficiency of the sample size for identifying statistically significant correlation coefficients, the minimum number of observations was calculated; the results are presented in Table 3.

Table 3

Minimum sample size for statistically significant *Pearson's r* correlation coefficients

Anticipated correlation coefficient	Significance level **α**			
	0.05		0.01	
	Statistical power 1−**β**			
	0.80	0.90	0.8	0.9
0.10	785	1051	1168	1488
0.15	349	467	520	662
0.20	197	263	292	372
0.25	126	169	187	239
0.30	88	117	130	166
0.35	65	86	96	122
0.40	50	66	73	93
0.45	39	52	58	74
0.50	32	43	47	60
0.55	26	35	39	50
0.60	22	30	33	42
0.65	19	25	28	36
0.70	17	22	24	31
0.75	14	19	21	27
0.80	13	17	19	24
0.85	11	15	17	21
0.90	10	13	15	19
0.95	9	12	13	17

The analysis shows that the minimum sample size depends significantly on the expected value of the correlation coefficient, the significance level and the power of the test. Thus, for weak correlations ($r \leq 0.2$), the required number of observations reaches the hundreds, whereas for moderate and strong relationships ($r \geq 0.4$), a sample of several dozen measurements is sufficient.

The minimum sample size matrix (Table 4) allows for a visual comparison of the required amount of data for different pairs of roughness parameters.

Table 4

Matrix of minimum required sample sizes calculated from *Pearson's* correlation coefficients for roughness parameters

	R_a	R_q	R_z	R_p	R_t	R_c	R_v	R_{sm}	R_{sk}	R_{ku}
R_a	*	11	11	11	11	11	11	12	218	27
R_q	16	*	11	11	11	11	11	12	199	27
R_z	16	16	*	11	11	11	11	12	169	28
R_p	16	16	16	*	12	11	11	13	730	33
R_t	16	16	16	17	*	11	11	12	57	25
R_c	16	16	16	16	16	*	11	12	169	27
R_v	16	16	16	16	16	16	*	12	169	28
R_{sm}	17	17	17	18	17	17	17	*	73	17
R_{sk}	308	282	239	1034	81	239	239	104	*	33
R_{ku}	38	38	39	46	35	38	39	23	46	*

* Values above the diagonal are for significance level **α** = 0.05; values below the diagonal are for **α** = 0.01.*

For most amplitude parameters, 16 observations are sufficient, but for parameters related to the shape of the profile, the sample size required increases significantly, which confirms their high sensitivity to statistical stability. For example, for the correlation of ***Rsk*** with ***Rp***, the calculated sample size reaches 1,034 at **α** = 0.01, which makes conducting such a large-scale experiment practically impossible under production conditions. This highlights the need to find a compromise approach to reduce sample size requirements while maintaining statistical robustness. In this regard, it is advisable to focus on measurements for the most informative relationships – specifically, the correlations of ***Rsk*** with ***Rz***, ***Rt***, and ***Rsm***. For these pairs, the required sample size (at **α** = 0.01: 39, 81, and 104 observations, respectively) remains labor-intensive but realistic. This choice allows for a balance between statistical reliability and resource constraints, providing sufficient data to analyze the effect of milling process parameters on roughness parameters.

It is important to consider established criteria for statistical validity. To build statistically significant models, the permissible error is taken as $E \approx 1$–3%, and at **α** = 0.01, the required sample size is $n = 128$. This value ensures sufficient accuracy of coefficient estimates and predictions without excessively increasing the sample size.

At the same time, two of the most common and standardized indicators in roughness assessment practice are ***Ra*** and ***Rz***, which should be subjected to separate testing to confirm the correctness of the results obtained. The parameters ***Ra*** and ***Rz*** are sensitive to the number of observations; with a small sample size, their values may be overestimated due to the influence of individual data points (Fig. 1).

As the amount of data increases, there is a tendency for the expected values (***M***) of the parameters ***Ra*** and ***Rz*** to decrease. This is because adding new measurements reduces the impact of outliers and forms a more representative picture of the surface structure. However, assessing the robustness of individual parameters as the sample size increases requires not only analyzing their average values, but also checking the relationships between them with a statistically sufficient amount of data.

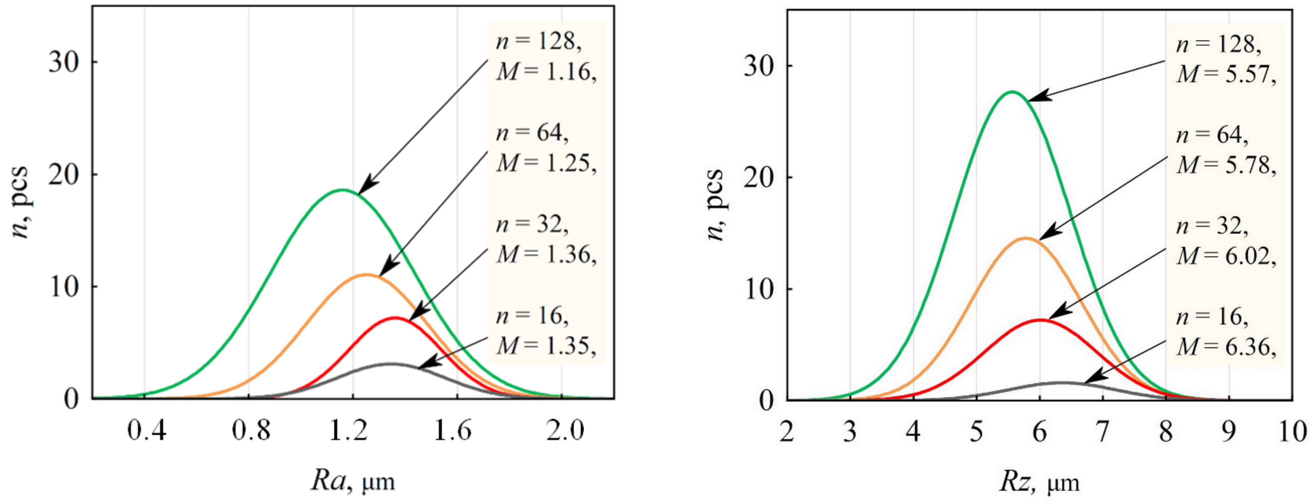

Fig. 1. Distribution of the surface roughness parameters ***Ra*** and ***Rz***

For this purpose, a matrix of paired *Pearson* correlation coefficients was calculated (Table 5), which corresponds to the previously established minimum number of observations required to obtain reliable estimates. The obtained values make it possible to assess the degree of linear dependence between the roughness parameters and confirm the stability of the identified relationships with the required sample size.

Table 5

***Pearson's* correlation coefficients (α = 0.01, *n* = 128)**

	R_a	R_q	R_z	R_p	R_t	R_c	R_v	R_{sm}	R_{sk}	R_{ku}
R_a	1	0.99	0.99	0.98	0.98	0.99	0.97	0.91	−0.11	−0.32
R_q	0.99	1	0.99	0.99	0.99	0.99	0.98	0.91	−0.12	−0.30
R_z	0.99	0.99	1	0.99	0.99	0.99	0.99	0.89	−0.15	−0.25
R_p	0.98	0.99	0.99	1	0.99	0.99	0.97	0.88	−0.07	−0.26
R_t	0.98	0.99	0.99	0.99	1	0.99	0.99	0.89	−0.16	−0.24
R_c	0.99	0.99	0.99	0.99	0.99	1	0.98	0.91	−0.12	−0.27
R_v	0.97	0.98	0.99	0.97	0.99	0.98	1	0.88	−0.25	−0.22
R_{sm}	0.91	0.91	0.89	0.88	0.89	0.91	0.88	1	−0.13	−0.32
R_{sk}	−0.11	−0.12	−0.15	−0.07	−0.16	−0.12	−0.25	−0.13	1	−0.33
R_{ku}	−0.32	−0.30	−0.25	−0.26	−0.24	−0.27	−0.22	−0.32	−0.33	1

For a sample size of n = 16, the correlation estimates between roughness parameters appear inflated. Many correlations appear extremely strong (***Rsm*** and ***Rku*** = −0.81), whereas when the sample size is increased to ***n*** = 128, most amplitude relationships remain strong, but several pairs weaken significantly or change sign (especially pairs involving ***Rsk*** and ***Rku***). This indicates the instability of the estimates obtained from a smaller sample, due to the influence of individual outliers and heterogeneity of subsamples – a statistically expected outcome, since increasing ***n*** reduces the variance of the correlation estimates.

For the strongest interactions at ***n*** = 128, the following correlation equations were obtained, expressed in terms of the parameters ***Ra*** and ***Rz***:

$$Ra = -0.0763 + 0.2412Rz; \quad Rz = 0.5593 + 4.0352Ra;$$

$$Rq = -0.0803 + 0.2820Rz; \quad Rq = 0.0274 + 1.1609Ra;$$

$$Rp = 0.0991 + 0.5868Rz; \quad Rp = 0.4055 + 2.3779Ra;$$

$$Rt = 0.4797 + 1.0220Rz; \quad Rt = 1.0576 + 4.1214Ra;$$

$$Rc = -0.0514 + 0.8691Rz; \quad Rc = 0.3608 + 3.5409Ra;$$

$$Rv = -0.0991 + 0.4132Rz; \quad Rv = 0.1538 + 1.6573Ra;$$

$$Rsm = 120.76 + 52.03Rz; \quad Rsm = 133.61 + 217.39Ra.$$

The parameters ***Rsk*** and ***Rku*** are analyzed separately, including testing the distribution for normality and identifying possible outliers**;** *Spearman's* rank correlation coefficient **(ρ)** is used to assess the strength and direction of potentially nonlinear relationships between the variables (Table 6). The correlation **ρ** is equivalent to usual correlation ***r*** if we first replace the original values with their ranks $r_{x,i} = R(x_i)$, $r_{y,i} = R(y_i)$, and then calculate ***r*** between these ranks:

$$\rho = \frac{\sum_{i=1}^{n}\left(r_{x,i}-\bar{r}_x\right)\left(r_{y,i}-\bar{r}_y\right)}{\sqrt{\sum_{i=1}^{n}\left(r_{x,i}-\bar{r}_x\right)^2}\sqrt{\sum_{i=1}^{n}\left(r_{y,i}-\bar{r}_y\right)^2}}.$$

Table 6

***Spearman's* rank correlation coefficient matrix (α = 0.01; *n* = 128)**

	R_z	R_t	R_{sm}	R_{sk}	R_{ku}
R_z	1	0.98	0.64	−0.20	−0.43
R_t	0.98	1	0.64	−0.20	−0.40
R_{sm}	0.64	0.64	1	−0.01	−0.47
R_{sk}	−0.20	−0.20	−0.01	1	0.19
R_{ku}	−0.43	−0.40	−0.47	0.19	1

In the presence of outliers or nonlinear monotonic relationships, **ρ** provides a more robust assessment of the associations; therefore, its results are preferred for the ***Rsk*** and ***Rku*** data.

The coefficient **ρ** = 0.98 indicates a practically monotonic relationship between ***Rz*** and ***Rt***, which is expected, since both parameters are amplitude parameters of the profile and are formed by similar mechanisms. The parameter ***Rsm*** shows a moderate positive correlation with ***Rz*** and ***Rt*** (**ρ** = 0.64)**,** indicating that an increase in the height of irregularities is accompanied by an increase in the mean distance between them. The asymmetry (skewness) parameter ***Rsk*** shows practically no correlation with the amplitude parameters (**ρ** from –0.20 to –0.01). This indicates that the asymmetry of the height distribution is an independent characteristic, not directly related to the absolute values of roughness or ***Rsm***. The profile excess (kurtosis) parameter ***Rku*** shows moderate negative relationships with ***Rz***, ***Rt***, and ***Rsm*** (**ρ** from –0.40 to –0.47). Profiles with higher amplitude values and increased spacing of irregularities are characterized by a less **"**peaked**"** distribution of heights.

The obtained results demonstrate the relative independence of ***Rsk*** and ***Rku*** from other parameters; therefore, further analysis requires consideration not only of the correlation structure, but also of the nature of the distribution of the parameters themselves.

Since the parameters ***Rsk*** and ***Rku*** describe the shape of the height distribution, and not the profile geometry, their correct interpretation requires checking the conformity of the empirical distribution to the normal law. The asymmetry parameter (skewness) ***Rsk*** characterizes the difference between the topography of protrusions and depressions. It has a larger variance compared to the parameter ***Rq*** and is calculated using the expressions:

$$Rsk = \frac{1}{Rq^3}\left[\frac{1}{lr}\int_0^l z^3(x)\,dx\right] \text{ or } Rsk = \frac{1}{NRq^3}\sum_{i=1}^{N} z^3 .$$

Here ***Rq*** is a standard statistical parameter of the distribution, characterizing the width of the histogram of the distribution of the profile ordinates ***z***(***x***):

$$Rq = \sqrt{\frac{1}{lr}\int_0^l z^2(x)\,dx} \text{ or } Rq = \sqrt{\frac{1}{N}\sum_{i=1}^{N} z^2} .$$

The parameter ***Rku*** characterizes the deviation of the distribution from normality:

$$Rku = \frac{1}{Rq^4}\left[\frac{1}{lr}\int_0^{lr} z^4(x)\,dx\right] \text{ or } Rku = \frac{1}{NRq^4}\sum_{i=1}^{N} z^4 .$$

For the correct interpretation of these parameters, it is necessary to check the extent to which the distribution of profile heights corresponds to the normal distribution. Deviations from normality are directly reflected in the ***Rsk*** and ***Rku*** values, making distribution analysis a key step in the study.

With $f > 100$, if the null hypothesis concerning the conformity of the distribution of ***Rsk*** and ***Rku*** parameters to the normal law according to *Pearson's* chi-squared test cannot be rejected (**α** > 0.01), the distribution is considered normal, since the graphs, skewness, and kurtosis values do not indicate the opposite (Fig. 2).

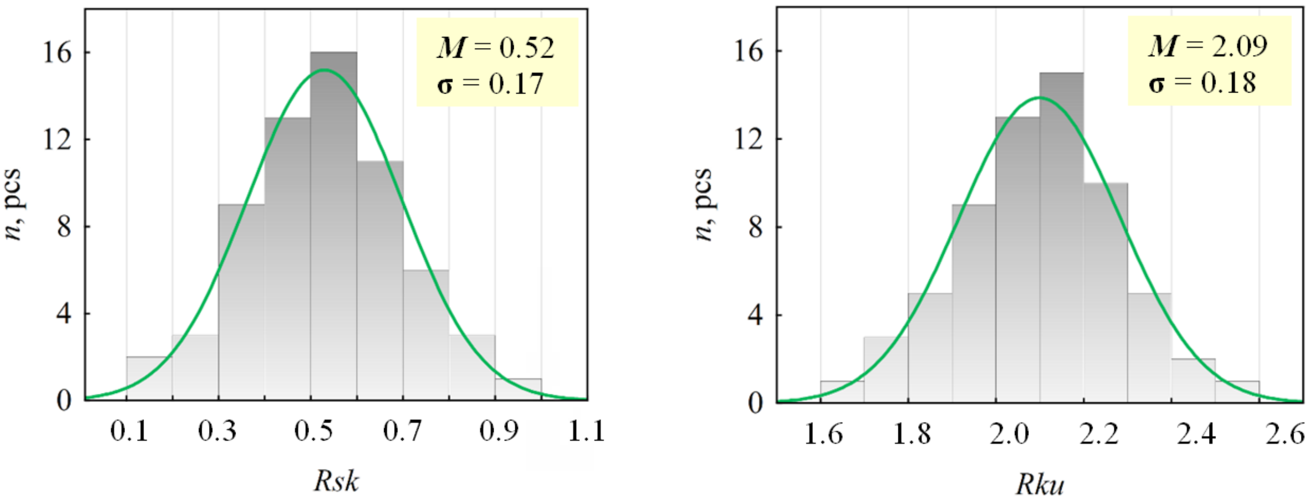

Fig. 2. Distribution of the R_{sk} and R_{ku} parameters with frequency histograms

It was determined that each value in the series deviates from the mean of 0.52 µm by an average of 0.17 µm for ***Rsk***, and from the mean of 2.09 µm by 0.18 µm for ***Rku***.

The *Shapiro–Wilk* test showed the following results: $W = 0.98$; $p = 0.76 > 0.05$ for ***Rsk***, and $W = 0.99$; $p = 0.97 > 0.05$ for ***Rku***, which indicates that there is **no** basis for rejecting the assumption of a normal distribution for the parameters under consideration. Testing the normality of distribution using the *Anderson–Darling* test showed that the statistics for both samples were below the critical value: $A^2(Rsk) = 0.27 < 0.73$ and $A^2(Rku) = 0.16 < 0.73$. This means that there is no reason to reject the null hypothesis that the distribution is normal.

The results of hypothesis testing for the parameters ***Rsk*** and ***Rku*** indicate that there is no basis for rejecting the assumption of a normal distribution law, which ensures the methodological validity of the subsequent analysis.

The established correspondence of the height distribution to the normal model allows correct use of the shape parameters in regression dependencies. At the next stage of the study, an assessment of the multicollinearity of factors was carried out, including the technological parameters of feed rate, inclination angle, and diameter, in order to exclude any linear relationship between the predictors and ensure the stability of the coefficient estimates. If the *VIF* value exceeds 10, high multicollinearity is indicated. The average *VIF* values for ***fz***, **γ**, and ***D*** are approximately equal to 1, indicating their independence and the absence of a critical relationship. Thus, these factors can be included simultaneously in regression models without the risk of distorting the coefficients. In this case, it is advisable to use all three predictors in the models, and the choice of responses (***Rz***, ***Rku***, ***Rsk***) will determine their specific form.

After algebraic expansion in the original variables, we obtain the following models:

$$Rz = -3.0975 + 37.1750fz + 0.0228\gamma + 0.1800D - 0.2775fz\gamma - 2.1667fzD -$$

$$- 0.0013\gamma D + 0.0167fz\gamma D;$$

$$Rsk = 1.1671 - 1.4562fz - 0.0346\gamma - 0.0283D + 0.0849fz\gamma + 0.0968fzD +$$

$$+ 0.0027\gamma D - 0.0076fz\gamma D;$$

$$Rku = 4.2348 - 3.5884fz - 0.0392\gamma - 0.1822D + 0.0515fz\gamma + 0.2972fzD +$$

$$+ 0.0035\gamma D - 0.0041fz\gamma D.$$

For the ***Rku*** model: *MAE* = 0.1159 and *RMSE* = 0.1300. We simplify the model by excluding weak interactions, obtaining:

$$Rku = 3.4087 - 2.8967fz - 0.0055\gamma - 0.0942D + 0.0067fz\gamma + 0.2422fzD.$$

For this model, the absolute error values between the predicted and actual values are *MAE* = 0.1159 and *RMSE* = 0.1303. The simplified regression model demonstrates comparable values of quality metrics compared to the full model, while having lower structural complexity and higher interpretability of coefficients.

To assess the fit of the three regression models and analyze the relationship between errors, the covariance matrix of the residuals was calculated. When estimating the covariance matrix of errors, a fragment of experimental data with a sample size of ***n*** = 20 observations corresponding to the realized combinations of levels of the factors levels ***fz*, γ, *D*** was used. Each regression equation contains ***k*** = 8 parameters (intercept, three linear effects and four interaction terms), which determines the number of degrees of freedom for the error as ***n*** – ***k*** = 20 – 12 = 8. The estimation of the error covariance matrix was performed with normalization by ***n*** – ***k***, which ensures unbiased estimates of the variances and covariances of the residuals. Based on a sample of 20 observations, predicted values of ***Rz'***, ***Rsk'***, ***Rku'*** were calculated for each equation, after which a residual matrix ***E*** of dimension 20 × 3 was formed.

The estimation of the error covariance matrix was performed using the matrix formula: $\Sigma' = \dfrac{1}{n-k} E^T E$

. The resulting covariance matrix has the form:

$$\Sigma' = \begin{bmatrix} 0.0000252 & -0.0000165 & 0.0000230 \\ -0.0000165 & 0.0000110 & -0.0000153 \\ 0.0000230 & -0.0000153 & 0.0000214 \end{bmatrix}.$$

The diagonal elements of the matrix Σ' reflect the error variances of individual models and show that all three equations have high approximation accuracy. The model error variances correspond to small deviations between actual and calculated values: **ε(*Rz*)** = 2.52×10^{-5}; **ε(*Rsk*)** = 1.10×10^{-5}; **ε(*Rku*)** = = 2.14×10^{-5}. Off-diagonal matrix elements reflect error covariances between models, demonstrating a weak negative relationship between **the** pairs ***Rz–Rsk*** and ***Rsk–Rku***, as well as a modest positive association between ***Rz*** and ***Rku*** errors. Such a structure of covariances indicates a more consistent response of the parameters ***Rz*** and ***Rku*** to changes in technological factors, while the parameter ***Rsk*** exhibits relative independence. The obtained covariance structure confirms the correctness of the constructed models and the absence of systematic shifts in the residuals. Low values of error variances and the absence of pronounced multicollinearity indicate the statistical stability of **the** regression equations and the possibility of their further use for predicting roughness parameters depending on the technological conditions of milling.

Thus, the analysis showed that the amplitude roughness parameters exhibit high correlation and stability of estimates with increasing sample size, while the distribution shape parameters (***Rsk*** and ***Rku***) require separate consideration due to their sensitivity to the data structure. A normality check confirmed the validity of their use in regression models, enabling the construction of statistically significant relationships between process parameters and surface roughness parameters.

Conclusion

Collectively, the results obtained confirm that the proposed method provides statistically reliable construction of correlation-regression models that adequately describe the effect of milling process parameters on surface roughness characteristics. The obtained data allow for identification of the following most significant conclusions:

1. A methodological approach has been developed to determine the minimum required sample size, based on a combination of paired *Pearson* and *Spearman* correlation coefficient analysis, calculation of the required number of observations for given levels of significance and test power, as well as the construction of a minimum sample size matrix for various pairs of roughness parameters. It is shown that the amplitude parameters (***Ra***, ***Rq***, ***Rz***, ***Rt***, ***Rc***, ***Rv***) demonstrate high stability of estimates and can be reliably characterized even with relatively small samples, while the distribution shape parameters (***Rsk*** and ***Rku***) require a significantly larger amount of data, which emphasizes their specific informational content and the need for separate analysis. Thus, the developed approach allows not only formal estimation of the minimum number of observations, but also a differentiated approach to sample size selection depending on the nature of the analyzed parameters and their effectiveness, and the need for separate analysis.

2. The results obtained with different sample sizes (n = 16 and n = 128) were compared, which made it possible to analyze the influence of the number of observations on the stability of the correlation coefficients, as well as on the accuracy of approximation and the magnitude of prediction errors. It was found that with a small sample size, correlation relationships, especially involving the ***Rsk*** and ***Rku*** parameters, tend to be overestimated and unstable, while with an increase in the number of observations, the correlation structure stabilizes, and estimates become more robust with respect to individual outliers and subsample heterogeneity. Furthermore, it was determined that at $n \approx 128$, a balance is achieved between the statistical accuracy of the estimates and the practical feasibility of the experiment**.**

3. It has been demonstrated that for analysis and modeling tasks dealing primarily with amplitude roughness parameters, a relatively small sample size can be used, whereas including distribution shape parameters (***Rsk*** and ***Rku***) in the models requires a significant increase in the number of observations and verification of distributional normality. The developed approach enables targeted experimental design, selecting the sample size based on both statistical power and the required accuracy of estimates, as well as on production resource constraints.

References

1. Suslov A.G., Medvedev D.M., Petreshin D.I., Fedonin O.N. Sistema avtomatizirovannogo tekhnologicheskogo upravleniya iznosostoikost'yu detalei mashin pri obrabotke rezaniem [System for automated wear–resistance technological control of machinery at cutting]. *Naukoemkie tekhnologii v mashinostroenii = Science Intensive Technologies in Mechanical Engineering*, 2018, no. 5 (83), pp. 40–44. DOI: 10.30987/article_5ad8d291cddcd8.06334386.

2. Wen Y., Zhou W., Tang J. Research on the correlation between roughness parameters and contact stress on tooth surfaces and its dominant characteristics. *Measurement*, 2024, vol. 238, p. 115399. DOI: 10.1016/j.measurement.2024.115399.

3. Podulka P. Selection of methods of surface texture characterisation for reduction of the frequency-based errors in the measurement and data analysis processes. *Sensors*, 2022, vol. 22 (3), p. 791. DOI: 10.3390/s22030791.

4. Qi Q., Li T., Scott P.J., Jiang X. A correlational study of areal surface texture parameters on some typical machined surfaces. *Procedia CIRP*, 2025, vol. 27, pp. 149–154. DOI: 10.1016/j.procir.2015.04.058.

5. Wen Y., Tang J., Zhou W., Zhu C. Study on contact performance of ultrasonic-assisted grinding surface. *Ultrasonics*, 2019, vol. 91, pp. 193–200. DOI: 10.1016/j.ultras.2018.08.009.

6. Sedlaček M., Podgornik B., Vižintin J. Correlation between standard roughness parameters skewness and kurtosis and tribological behaviour of contact surfaces. *Tribology International*, 2012, vol. 48, pp. 102–112. DOI: 10.1016/j.triboint.2011.11.008.

7. Gimadeev M.R., Davydov V.M., Nikitenko A.V., Sarygin A.V. Formirovanie parametrov sherokhovatosti na osnove korrelyatsionnykh svyazei pri chistovom frezerovanii prostranstvenno-slozhnykh poverkhnostei [Formation of roughness parameters based on correlation relations during finishing milling of spatially complex surfaces]. *Uprochnyayushchie tekhnologii i pokrytiya = Strengthening Technologies and Coatings*, 2019, vol. 15, no. 6 (174), pp. 243–248.

8. Yang D., Tang J., Zhou W., Wen Y. Correlation between surface roughness parameters and contact stress of gear. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, 2020, vol. 235 (3), pp. 551–563. DOI: 10.1177/1350650120928661.

9. Li T., Huang X., Luo M. Analysis on the correlation between plunge milling parameters and plunge milling force and force coefficient. *2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 2018, pp. 927–936. DOI: 10.1109/IAEAC.2018.8577706.

10. Guo M., Xia W., Wu C., Luo C., Lin Z. A surface quality prediction model considering the machine-tool-material interactions. *The International Journal of Advanced Manufacturing Technology*, 2024, vol. 131 (7–8), pp. 1–19. DOI: 10.1007/s00170-024-13072-2.

11. Paturi U.M.R., Devarasetti H., Narala S.K.R. Application of regression and artificial neural network analysis in modelling of surface roughness in hard turning of AISI 52100 steel. *Materials Today: Proceedings*, 2018, vol. 5 (2), pp. 4766–4777. DOI: 10.1016/j.matpr.2017.12.050.

12. Zakharova O.V., Suleimanova F.D. Linear regression equations for determining the roughness of machined surfaces. Part 1. Turning, face milling, surface grinding, and polishing. *Russian Engineering Research*, 2024, vol. 44 (6), pp. 800–806. DOI: 10.3103/S1068798X24701259.

13. Chen C., Wu C., Zhang T., Liang S.Y. 3D curved surface milling modeling for the topography simulation and surface roughness prediction. *Journal of Manufacturing Processes*, 2025, vol. 137, pp. 150–165. DOI: 10.1016/j.jmapro.2025.02.003.

14. Wibowo A., Desa M.I. Kernel based regression and genetic algorithms for estimating cutting conditions of surface roughness in end milling machining process. *Expert Systems with Applications*, 2012, vol. 39 (14), pp. 11634–11641. DOI: 10.1016/j.eswa.2012.04.004.

15. Oktem H., Erzurumlu T., Erzincanli F. Prediction of Minimum Surface Roughness in End Milling Mold Parts Using Neural Network and Genetic Algorithm. *Materials & Design*, 2006, vol. 27 (9), pp. 735–744. DOI: 10.1016/j.matdes.2005.01.010.

16. Bujang M.A., Baharum B. Sample size guideline for correlation analysis. *World Journal of Social Science Research*, 2016, vol. 3, pp. 37–46. DOI: 10.22158/wjssr.v3n1p37.

17. Faul F., Erdfelder E., Buchner A., Lang A.-G. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 2009, vol. 41, pp. 1149–1160. DOI: 10.3758/BRM.41.4.1149.

18. Ponomarev B.B., Nguen S.H. Otsenka sherokhovatosti pri pyatikoordinatnom chistovom frezerovanii poverkhnostei sferotsilindricheskoi frezoi [the influence of tool orientation on cutting forces during end milling]. *Izvestiya vysshikh uchebnykh zavedenii. Mashinostroenie = BMSTU Journal of Mechanical Engineering*, 2020, no. 5 (722), pp. 21–31. DOI: 10.18698/0536-1044-2020-5-21-31.

19. Gimadeev M.R., Li A.A., Berkun V.O., Stelmakov V.A. Experimental study of the dynamics of the machining process by ball-end mills. *Obrabotka metallov (tekhnologiya, oborudovanie, instrumenty) = Metal Working and Material Science*, 2023, vol. 25, no. 1, pp. 44–56. DOI: 10.17212/1994-6309-2023-25.1-44-56.

20. Gimadeev M.R., Li A.A. Analysis of automated surface roughness parameter support systems based on dynamic monitoring. *Advanced Engineering Research (Rostov-on-Don)*, 2022, vol. 22 (2), pp. 116–129. DOI: 10.23947/2687-1653-2022-22-2-116-129.

21. Chen C.H., Jeng S.Y., Lin C.J. Prediction and analysis of the Surface roughness in CNC end milling using neural networks. *Application Science*, 2022, vol. 12 (1), p. 393. DOI: 10.3390/app12010393.

22. Manjunath K., Tewary S., Khatri N. Surface roughness prediction in milling using long-short term memory modelling. *Materials Today: Proceedings*, 2022, vol. 64 (3), pp. 1300–1304. DOI: 10.1016/j.matpr.2022.04.126.

23. Chai T., Draxler R.R. Root mean square error (RMSE) or mean absolute error (MAE) arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 2014, vol. 7 (3), pp. 1247–1250. DOI: 10.5194/gmd-7-1247-2014.

Conflicts of Interest

The authors declare no conflict of interest.