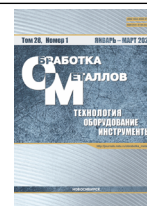




Обработка металлов (технология • оборудование • инструменты)

Сайт журнала: http://journals.nstu.ru/obrabotka_metallov



Определение минимального объема выборки для построения корреляционно-регрессионных моделей параметров шероховатости при фрезеровании

Михаил Гимадеев ^{a, *}, Вадим Стельмаков ^b, Максим Улисков ^c

Тихоокеанский государственный университет, ул. Тихоокеанская, 136, г. Хабаровск, 680035, Россия

^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

ИНФОРМАЦИЯ О СТАТЬЕ

УДК 621:311

История статьи:

Поступила: 25 декабря 2025

Рецензирование: 12 января 2026

Принята к печати: 19 января 2026

Доступно онлайн: 15 марта 2026

Ключевые слова:

Размер выборки

Шероховатость

Корреляция

Регрессионная модель

Фрезерование

Среднеквадратичная ошибка

Финансирование

Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации (проект № FEME–2024–0010 «Экспериментальная лаборатория мехатронных систем ЧПУ»).

АННОТАЦИЯ

Введение. Исследование статистических взаимосвязей между параметрами фрезерования и шероховатости требует корректного выбора объема выборки, поскольку амплитудные параметры и параметры формы по-разному реагируют на ограниченность данных. Надежность корреляционно-регрессионного анализа зависит от соблюдения предпосылок нормальности и устойчивости оценок, что делает задачу определения минимального числа наблюдений ключевой для построения достоверных моделей шероховатости поверхности. **Цель настоящей работы** заключается в разработке подхода к оценке минимального объема выборки, необходимого для построения статистически значимых корреляционно-регрессионных моделей, описывающих взаимосвязь технологических параметров процесса фрезерования и характеристик шероховатости поверхности, а также обеспечивающих статистическую надежность результатов и возможность адекватного прогнозирования качества обработанной поверхности. **Метод и методология.** Нормальность распределений параметров шероховатости после фрезерования оценивалась критериями Шапиро – Уилка, Андерсона – Дарлинга и Пирсона. Мультиколлинеарность технологических факторов контролировалась через VIF, а адекватность регрессионных моделей проверялась по MAE и RMSE. Минимальный объем выборки рассчитывался с учетом мощности теста и преобразования Фишера. **Результаты и обсуждения.** Анализ парных коэффициентов корреляции показал, что амплитудные параметры шероховатости демонстрируют устойчивые взаимосвязи уже при $n = 16$, тогда как параметры формы профиля (Rsk , Rku) характеризуются слабой или отрицательной корреляцией и требуют существенно большего объема выборки. Построенная матрица минимального числа наблюдений подтверждает, что для ряда связей, особенно с участием Rsk , необходимы сотни измерений, что делает оправданным выбор наиболее информативных пар параметров. Увеличение выборки до $n = 128$ снижает смещенность оценок, стабилизирует амплитудные корреляции и выявляет ослабление связей для Rsk и Rku . Ранговый анализ подтверждает монотонную зависимость между Rz и Rt и независимость Rsk от амплитудных характеристик. Проверка нормальности распределений и отсутствие мультиколлинеарности факторов обеспечили корректность построенных регрессионных моделей, отличающихся высокой точностью и согласованностью ошибок.

Для цитирования: Гимадеев М.Р., Стельмаков В.А., Улисков М.В. Определение минимального объема выборки для построения корреляционно-регрессионных моделей параметров шероховатости при фрезеровании // Обработка металлов (технология, оборудование, инструменты). – 2026. – Т. 28, № 1. – С. 64–80. – DOI: 10.17212/1994-6309-2026-28.1-64-80.

Введение

Исследование взаимосвязи технологических параметров фрезерования и шероховатости поверхности имеет ключевое значение для обеспечения требуемого качества обработки и прогно-

зирования эксплуатационных свойств деталей машин [1]. Существенное влияние на трибологические свойства оказывают параметры шероховатости поверхности, включая асимметрию (Rsk) и эксцесс профиля (Rku) [2–4].

Традиционные параметры Ra и Rz недостаточны для полной оценки трибологических свойств поверхностей, поскольку они не отражают особенности распределения микронеровностей. Однако параметры формы профиля Rsk и Rku демонстрируют более устойчивую связь

*Адрес для переписки

Гимадеев Михаил Радикович, к.т.н., доцент
 Тихоокеанский государственный университет,
 ул. Тихоокеанская, 136,
 680035, г. Хабаровск, Россия
 Тел.: +7 924 216-31-39, e-mail: 009063@togudv.ru



с характеристиками трения и износа, что подтверждается рядом исследований [5–7]. Эти параметры рассматриваются как наиболее информативные для рабочих поверхностей, функционирующих в условиях трения скольжения, для базовых, опорных поверхностей и др. Отмечается, что отрицательные значения Rsk , как правило, сочетаются с повышенными значениями Rku , формируя микрорельеф, способствующий снижению коэффициента трения [8] даже при относительно высоких значениях амплитудных параметров шероховатости.

Одним из широко применяемых статистических методов для количественного анализа зависимостей параметров остается корреляционный анализ [6–9]. Однако важно учитывать, что коэффициенты корреляции, особенно коэффициент Пирсона, нередко используются на крайне ограниченном числе наблюдений, что снижает достоверность результатов.

Корректность применения параметрических методов (коэффициент корреляции Пирсона, линейная регрессия и др.) предполагает совместимость распределения анализируемых переменных и резидуалов регрессионных моделей с нормальным законом. Это связано с тем, что статистические критерии значимости, доверительные интервалы и p -значения в параметрических методах выводятся при предположении нормальности. При существенных отклонениях от нормального распределения результаты параметрических методов могут быть некорректными, что требует предварительной проверки распределений или использования альтернативных, непараметрических методов. Для решения этой задачи применяются несколько критериев: Шапиро – Уилка [10], который отличается высокой мощностью при средних объемах выборки и позволяет надежно выявлять отклонения от нормального закона; критерий Андерсона – Дарлинга [11], который обладает повышенной чувствительностью к отклонениям в хвостах распределения, что делает его особенно полезным при анализе данных с редкими, но значимыми выбросами; критерий Пирсона, обеспечивающий интервальную оценку согласия между наблюдаемым и ожидаемым распределением. Такое сочетание критериев позволяет комплексно оценить нормальность распределения параметров, обеспечивает высокую надежность статисти-

ческих выводов и исключает риск ошибочной интерпретации данных.

В современных исследованиях корреляционный анализ дополняется регрессионным моделированием [12], позволяющим не только оценить силу и направление связи между переменными, но и построить уравнения зависимости технологических параметров от показателей шероховатости. При этом анализ резидуалов регрессионных моделей обеспечивает проверку адекватности полученных зависимостей и уточнение значимости отдельных факторов [13]. Дополнительным инструментом повышения надежности регрессионного анализа является индекс инфляции дисперсии (variance inflation factor, VIF) [14], который показывает, насколько дисперсия оценки коэффициента регрессии увеличивается вследствие корреляции независимой переменной (предиктора) с другими. VIF используется как один из ключевых инструментов выявления мультиколлинеарности.

Особое внимание в области обработки металлов резанием уделяется вопросу минимального размера выборки [15–17], необходимого для проведения корреляционно-регрессионного анализа. Недостаточный объем данных может привести к переобучению моделей или искажению выводов, тогда как избыточный – к неоправданным затратам ресурсов. Для технологических процессов фрезерования эта проблема имеет прикладное значение, так как указанные операции часто используются при обработке в том числе сложных поверхностей [18, 19] и требуют контроля параметров качества. В связи с этим актуальной задачей является определение оптимального числа наблюдений, обеспечивающего статистическую достоверность результатов при применении коэффициентов корреляции Пирсона (r) [20] и Спирмена (ρ).

Цель настоящей работы заключается в разработке подхода к оценке минимального объема выборки, необходимого для построения статистически значимых корреляционно-регрессионных моделей, описывающих взаимосвязь технологических параметров процесса фрезерования и характеристик шероховатости поверхности, а также обеспечивающих статистическую надежность результатов и возможность адекватного прогнозирования качества обработанной поверхности.

Для достижения цели необходимо решить следующие **задачи**:

- разработать методический подход к определению минимально необходимого объема выборки, обеспечивающего статистическую надежность получаемых результатов;
- проанализировать влияние числа наблюдений на устойчивость коэффициентов регрессии, точность аппроксимации и величину ошибок предсказания;
- сформулировать практические рекомендации по выбору достаточного объема выборки для решения прикладных задач анализа технологических процессов.

Методика исследований

Методика статистической оценки данных.

В исследовании нормальность проверяется с использованием нескольких взаимодополняющих критериев.

- Критерий Шапиро – Уилка позволяет установить, насколько выборка соответствует нормальному распределению через линейную комбинацию порядковых статистик:

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

где $y_{(i)}$ – упорядоченные значения выборки, y_i – наблюдаемые и $\bar{y}$ – среднее значения; a_i – коэффициенты, зависящие от ковариационной матрицы нормального распределения.

- Критерий Андерсона – Дарлинга, который представляет собой модификацию критерия Колмогорова – Смирнова, придающую больший вес хвостам распределения:

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

где $F(x)$ – функция распределения нормали.

- Критерий согласия Пирсона основан на сравнении наблюдаемых и ожидаемых частот в интервалах $\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$, где O_i – наблюдаемая частота в i -м интервале; E_i – ожида-

емая частота в i -м интервале; k – число интервалов.

Важным фактором, способным исказить результаты анализа, остается мультиколлинеарность предикторов. Для ее выявления и количественной оценки применяется

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

где R^2 – коэффициент детерминации, который показывает, какая доля вариации зависимой переменной объясняется моделью на основе независимых переменных [21].

Если обозначить резидуалы как $e_i = y_i - \hat{y}_i$, где $\hat{y}_i$ – прогнозируемые значения, тогда

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

Чтобы одновременно оценить качество объяснения и точность прогноза, для проверки адекватности модели используют метрики ошибок [22]:

- среднюю абсолютную ошибку (MAE):

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

- квадратный корень из средней квадратичной ошибки (RMSE) [23]:

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

Благодаря тому, что функция RMSE гладкая и дифференцируемая, можно аналитически вывести оптимальные коэффициенты регрессии. Для простой линейной модели $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, где x_i – предиктор; β_0, β_1 – параметры модели; ε_i – случайная ошибка.

Резидуалы определяются как $e_i = y_i - (\beta_0 + \beta_1 x_i)$. Функция RMSE записывается в виде

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

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

Тогда градиент RMSE по параметрам модели через правило цепи имеет вид

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

Для β_0 :

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

Для β_1 :

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

Таким образом, гладкость и дифференцируемость функции RMSE обеспечивают возможность аналитического нахождения оптимальных коэффициентов регрессии и применения градиентных методов для их оценки.

Чтобы рассчитать необходимый объем выборки для исследования, используется формула, основанная на требованиях к точности, уровню доверия и ожидаемой изменчивости данных. Формула для бесконечной генеральной совокупности:

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

где n – объем выборки; Z – квантиль стандартного нормального распределения, соответствующий уровню доверия; p – предполагаемая доля или вероятность; E – допустимая погрешность, которая уменьшается с ростом числа наблюдений:

$E \propto \frac{Z_{1-\alpha/2} \sigma}{\sqrt{n}}$, где $Z_{1-\alpha/2}$ – квантиль уровня

значимости; σ – стандартное отклонение.

Когда размер генеральной совокупности (N) известен, объем выборки уменьшается, учитывая ограниченный масштаб. Тогда применяется уточненная формула для конечной совокупности:

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

При оценке корреляции объем выборки также зависит от ожидаемой величины корреляции (r). Формула может быть адаптирована для использования мощности теста [16, 17] ($1 - \beta$) и уровня значимости (α):

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

где $Z_{1-\beta}$ – квантиль мощности теста.

Следует отметить, что эта формула дает лишь приближенную оценку и не учитывает нелинейность распределения коэффициента корреляции. Более корректный расчет объема выборки при проверке гипотезы о корреляции обычно выполняется на основе преобразования Фишера, позволяющего перевести коэффициент корреляции в нормально распределенную величину и тем самым обеспечить точность статистической оценки:

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

Вместе с этим для конечной генеральной совокупности применяют поправку:

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

Для вычисления квантилей уровня значимости и мощности необходимо рассчитать значение, ниже которого находится определенная часть данных стандартного нормального распределения для заданной вероятности. Это достигается с использованием функции точки вероятности (percent point function, PPF) – обратной функции к кумулятивной функции распределения (cumulative distribution function, CDF):

– для непрерывного распределения

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt,$$

где $F(x)$ – CDF, значение функции распределения для точки x ; $P(X \leq x)$ – вероятность того, что X меньше или равно x ; $f(t)$ – плотность вероятности (probability density function, PDF);

– для дискретного распределения

$$F(x) = P(X \leq x) = \sum_{t \leq x} P(X = t),$$

где $P(X = t)$ – вероятность конкретного значения t .

Для нормального распределения при условии, что среднее равно нулю, а стандартное отклонение равно единице, значение CDF вычисляется как

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

где $e_{\ln}$ – основание натурального логарифма.

Расчеты минимального объема выборки, приведенные выше, основаны на использовании параметрического коэффициента корреляции Пирсона, так как он предполагает линейную зависимость между переменными и требует, чтобы данные были нормально распределены. Однако при практическом применении важно учитывать, что реальное распределение параметров шероховатости может изменяться в зависимости от объема выборки и условий измерений.

Детали эксперимента. Экспериментальные исследования проводились на фрезерном обрабатывающем центре DMC 635V. Были получены объекты исследования – образцы поверхностей для анализа параметров шероховатости. Объекты изготовлены из нержавеющей стали 12X18H10T со следующим химическим составом (масс. %): $C \leq 0,12$; $Si \leq 0,8$; $Mn \leq 2,0$; $P \leq 0,035$; $S \leq 0,02$; $Ni \ 9...11$; $Cr \ 17...19$; $Ti < 0,8$; Fe – остальное. В качестве материала режущего инструмента при резании использовался твердый сплав с мультислойным PVD-покрытием и мелкозернистой основой.

При планировании эксперимента применен полный факторный план 2^3 с двумя центр-точками и двумя повторами, что обеспечило надежность

статистической оценки (табл. 1). В качестве факторов рассматривались диаметр инструмента ($D = 6...12$ мм), угол наклона ($\gamma = 10^\circ...50^\circ$) и подача на зуб ($fz = 0,1...0,5$ мм/зуб). Центр-точки выбраны так, чтобы проверить линейность или нелинейность отклика, при этом учесть применяемые режимы и повысить надежность статистической оценки за счет дополнительных наблюдений в середине диапазона.

Такой дизайн позволяет получить репрезентативные данные о влиянии режимов фрезерования на формирование шероховатости поверхности.

Измерение шероховатости и выбор базовой длины. Измерения шероховатости после фрезерования поверхности проводились при помощи профилометра SURFCOM 1800D [20], фильтр – 50%-й гауссов. В общем случае значение базовой длины выбирается для интервалов значений Ra и Rz согласно ИСО 4288:1996. В настоящей работе по рекомендациям для измерения периодического профиля использовался параметр Rsm со следующими диапазонами: при $0,13 < Rsm \leq 0,4$ базовая длина $lr = 0,8$ мм; при $0,4 < Rsm \leq 1,3$ базовая длина $lr = 2,5$ мм, а длина оценки L составила $5lr$.

Таблица 1

Table 1

Матрица эксперимента $n = 16$

Experiment matrix $n = 16$

№ / No.	Диаметр инструмента D , мм / Tool diameter D , mm	Угол наклона γ , град / Inclination angle γ , grad	Подача на зуб fz , мм/зуб / Feed per tooth fz , mm/tooth
Полный факторный план 2^3 / 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
Центр-точки с двумя повторами / 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
Дополнительные повторы крайних комбинаций / 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

Следует отметить, что использование параметра R_z для оценки периодического профиля поверхности является методологически некорректным. Этот показатель отражает лишь амплитудные характеристики неровностей и не учитывает регулярность структуры, что приводит к искажению интерпретации результатов. Выбранная базовая длина $l_r = 0,8$ мм во всех измерениях могла усилить эффект смещения, снижая достоверность полученных данных. Таким образом, результаты измерений в диапазоне $1,6 < R_z < 12,5$ не могут рассматриваться как полностью корректные и требуют уточнения с применением более адекватных параметров для периодических профилей.

Наличие повторяющихся элементов в профиле шероховатости может быть обусловлено износом инструмента, вибрациями или различными технологическими дефектами [19]. Поэтому перед выбором базовой длины необходимо убедиться, что анализируемый профиль обладает периодичностью. В частности, в настоящей работе при фрезеровании профиль шероховатости поверхности рассматривается как кинематико-геометрическая проекция формы инструмента, полученная на окончательных операциях механической обработки.

Результаты и их обсуждение

В этом разделе представлены результаты статистического анализа параметров шероховатости поверхности, включающие в себя оценку

корреляционных зависимостей, проверку достаточности выборки, анализ распределений параметров формы профиля, а также построение регрессионных моделей. Такой подход позволяет комплексно оценить структуру данных, выявить устойчивые зависимости и определить информативность отдельных параметров.

В табл. 2 представлены парные коэффициенты корреляции Пирсона между параметрами шероховатости поверхности при объеме выборки $n = 16$. Полученные значения демонстрируют высокую степень взаимосвязи между амплитудными параметрами, что подтверждает их линейную зависимость и общность механизмов формирования. При этом параметры, характеризующие форму профиля (R_{sk} и R_{ku}), проявляют отрицательные или умеренные корреляции с амплитудными характеристиками, что указывает на их независимость и различную чувствительность к структуре профиля.

Для оценки достаточности выборки при выявлении статистически значимых коэффициентов корреляции использован расчет минимального числа наблюдений, результаты которого представлены в табл. 3.

Анализ показывает, что минимальный объем выборки существенно зависит от ожидаемой величины коэффициента корреляции, уровня значимости и мощности теста. Так, при слабых корреляциях ($r \leq 0,2$) требуемое число наблюдений достигает сотен, тогда как для умеренных и сильных связей ($r \geq 0,4$) достаточно выборки в несколько десятков измерений.

Таблица 2

Table 2

Парные коэффициенты корреляции Пирсона ($\alpha = 0,01$; $n = 16$)Pearson correlation coefficients ($\alpha = 0.01$; $n = 16$)

	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

Таблица 3

Table 3

Минимальный размер выборки для выявления статистически значимых коэффициентов корреляции r Пирсона

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

Ожидаемый коэффициент корреляции / Anticipated correlation coefficient	Уровень значимости α / Significance level α			
	0,05		0,01	
	Ошибка второго рода $1-\beta$ / Statistical power $1-\beta$			
	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

Матрица минимального объема выборки (табл. 4) позволяет наглядно сопоставить требуемое количество данных для различных пар параметров шероховатости.

Для большинства амплитудных параметров достаточно 16 наблюдений, однако для параметров, связанных с формой профиля, объем выборки значительно возрастает, что подтверждает их высокую чувствительность к статистической устойчивости. Например, для корреляции Rsk с Rp расчетный объем выборки достигает 1034 при $\alpha = 0,01$, что делает проведение столь масштабного эксперимента практически невозможным в производственных условиях. Это обстоятельство подчеркивает необходимость поиска компромиссного подхода, позволяющего снизить требования к объему выборки при сохранении статистической надежности результатов. Целесообразно сосредоточиться на измерениях для наиболее информативных связей – корреляциях Rsk с Rz , Rt и Rsm . При этом требуемый объем выборки для данных пар (при $\alpha = 0,01$ соответственно 39, 81 и 104 наблюдения) остается

трудоемким, но реалистичным. Такой выбор позволяет сохранить баланс между статистической надежностью и ресурсными ограничениями, обеспечивая достаточное количество данных для анализа влияния технологических характеристик процесса фрезерования на параметры шероховатости.

Важно учитывать общие критерии статистической достоверности. Для построения статистически значимых моделей допустимая погрешность принимается в пределах $E \approx 1...3\%$, и при $\alpha = 0,01$ значение $n = 128$. Это значение обеспечивает достаточную точность оценок коэффициентов и предсказаний без чрезмерного увеличения объема выборки.

Вместе с тем одними из наиболее распространенных и нормативно закреплённых показателей в практике оценки шероховатости являются Ra и Rz , которые целесообразно подвергнуть отдельной проверке, что позволит подтвердить корректность полученных результатов. Параметры Ra и Rz чувствительны к числу наблюдений и при малом объеме выборки их значения могут

Матрица минимального объема выборки по коэффициентам корреляции параметров шероховатости

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

	<i>Ra</i>	<i>Rq</i>	<i>Rz</i>	<i>Rp</i>	<i>Rt</i>	<i>Rc</i>	<i>Rv</i>	<i>Rsm</i>	<i>Rsk</i>	<i>Rku</i>
<i>Ra</i>	*	11	11	11	11	11	11	12	218	27
<i>Rq</i>	16	*	11	11	11	11	11	12	199	27
<i>Rz</i>	16	16	*	11	11	11	11	12	169	28
<i>Rp</i>	16	16	16	*	12	11	11	13	730	33
<i>Rt</i>	16	16	16	17	*	11	11	12	57	25
<i>Rc</i>	16	16	16	16	16	*	11	12	169	27
<i>Rv</i>	16	16	16	16	16	16	*	12	169	28
<i>Rsm</i>	17	17	17	18	17	17	17	*	73	17
<i>Rsk</i>	308	282	239	1034	81	239	239	104	*	33
<i>Rku</i>	38	38	39	46	35	38	39	23	46	*

* Над диагональю $\alpha = 0,05$; под диагональю $\alpha = 0,01$.
 * Values above the diagonal are for significance level $\alpha = 0.05$; values below the diagonal are for $\alpha = 0.01$.

быть завышены из-за влияния отдельных точек (рис. 1).

По мере увеличения количества данных наблюдается тенденция к снижению математического ожидания (M) параметров Ra и Rz . Это связано с тем, что добавление новых измерений сглаживает влияние выбросов и формирует более репрезентативное представление о структуре поверхности. Однако оценка устойчивости отдельных параметров по мере роста выборки требует не только анализа их средних значений,

но и проверки взаимосвязей между ними при статистически достаточном объеме данных.

С этой целью была рассчитана матрица парных коэффициентов корреляции Пирсона (табл. 5), что соответствует ранее установленному минимальному количеству наблюдений для формирования достоверных оценок. Полученные значения позволяют оценить степень линейной зависимости между параметрами шероховатости и подтвердить устойчивость выявленных связей при необходимом объеме выборки.

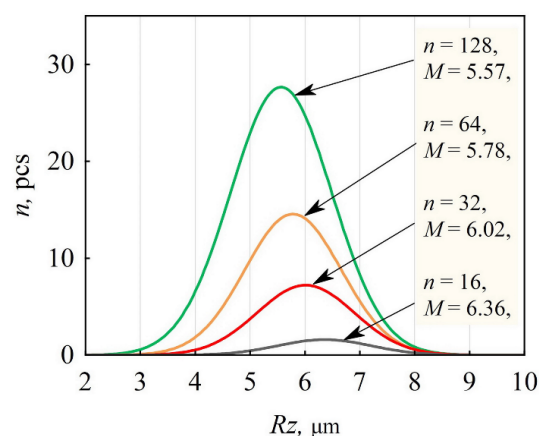
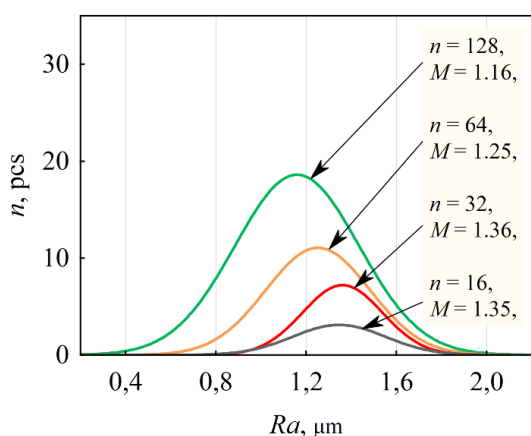


Рис. 1. Распределение параметров шероховатости Ra и Rz

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

Таблица 5

Table 5

Парные коэффициенты корреляции Пирсона ($\alpha = 0,01; n = 128$)

Pearson's correlation coefficients ($\alpha = 0.01; n = 128$)

	<i>Ra</i>	<i>Rq</i>	<i>Rz</i>	<i>Rp</i>	<i>Rt</i>	<i>Rc</i>	<i>Rv</i>	<i>Rsm</i>	<i>Rsk</i>	<i>Rku</i>
<i>Ra</i>	1	0,99	0,99	0,98	0,98	0,99	0,97	0,91	−0,11	−0,32
<i>Rq</i>	0,99	1	0,99	0,99	0,99	0,99	0,98	0,91	−0,12	−0,30
<i>Rz</i>	0,99	0,99	1	0,99	0,99	0,99	0,99	0,89	−0,15	−0,25
<i>Rp</i>	0,98	0,99	0,99	1	0,99	0,99	0,97	0,88	−0,07	−0,26
<i>Rt</i>	0,98	0,99	0,99	0,99	1	0,99	0,99	0,89	−0,16	−0,24
<i>Rc</i>	0,99	0,99	0,99	0,99	0,99	1	0,98	0,91	−0,12	−0,27
<i>Rv</i>	0,97	0,98	0,99	0,97	0,99	0,98	1	0,88	−0,25	−0,22
<i>Rsm</i>	0,91	0,91	0,89	0,88	0,89	0,91	0,88	1	−0,13	−0,32
<i>Rsk</i>	−0,11	−0,12	−0,15	−0,07	−0,16	−0,12	−0,25	−0,13	1	−0,33
<i>Rku</i>	−0,32	−0,30	−0,25	−0,26	−0,24	−0,27	−0,22	−0,32	−0,33	1

Выборка $n = 16$ показывает завышенные оценки корреляции между параметрами шероховатости. Многие корреляции выглядят экстремально сильными (Rsm и Rku равны $-0,81$), тогда как при увеличении объема выборки до $n = 128$ большинство амплитудных связей остаются сильными, но ряд пар существенно ослабевает или меняет знак (особенно пары с участием Rsk и Rku). Это указывает на нестабильность оценок при меньшей выборке, обусловленную влиянием отдельных выбросов и гетерогенностью подвыборок, что статистически ожидаемо, поскольку рост n снижает дисперсию оценок корреляции.

Для наиболее сильных взаимодействий при $n = 128$ получены следующие корреляционные уравнения, выраженные через параметры Ra и Rz :

$$Ra = -0,0763 + 0,2412Rz;$$

$$Rz = 0,5593 + 4,0352Ra;$$

$$Rq = -0,0803 + 0,2820Rz;$$

$$Rq = 0,0274 + 1,1609Ra;$$

$$Rp = 0,0991 + 0,5868Rz;$$

$$Rp = 0,4055 + 2,3779Ra;$$

$$Rt = 0,4797 + 1,0220Rz;$$

$$Rt = 1,0576 + 4,1214Ra;$$

$$Rc = -0,0514 + 0,8691Rz;$$

$$Rc = 0,3608 + 3,5409Ra;$$

$$Rv = -0,0991 + 0,4132Rz;$$

$$Rv = 0,1538 + 1,6573Ra;$$

$$Rsm = 120,76 + 52,03Rz;$$

$$Rsm = 133,61 + 217,39Ra.$$

Параметры Rsk и Rku анализируются отдельно, включая проверку распределения на нормальность и выявление возможных выбросов, а для оценки силы и направления предполагаемого нелинейной связи между переменными применяется ранговая корреляция Спирмена (табл. 6). Корреляция ρ эквивалентна обычной корреляции r , если сначала заменить исходные значения их рангами $r_{x,i} = R(x_i)$, $r_{y,i} = R(y_i)$, а затем вычислить r между этими рангами:

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

При наличии выбросов или нелинейных монотонных связей ρ дает более робастную оценку взаимосвязей, поэтому его результаты предпочтительны для данных Rsk и Rku .

Коэффициент $\rho = 0,98$ указывает на практически монотонную зависимость между Rz и Rt , что ожидаемо, поскольку оба параметра характеризуют амплитудные параметры профиля и формируются сходными механизмами. Параметр Rsm демонстрирует умеренную положительную корреляцию с Rz и Rt ($\rho = 0,64$) и указывает на то, что увеличение высоты неровностей

Таблица 6

Table 6

Матрица коэффициентов корреляции Спирмена ($\alpha = 0,01$; $n = 128$)Spearman's rank correlation coefficient matrix ($\alpha = 0.01$; $n = 128$)

	Rz	Rt	Rsm	Rsk	Rku
Rz	1	0,98	0,64	-0,20	-0,43
Rt	0,98	1	0,64	-0,20	-0,40
Rsm	0,64	0,64	1	-0,01	-0,47
Rsk	-0,20	-0,20	-0,01	1	0,19
Rku	-0,43	-0,40	-0,47	0,19	1

сопровождается ростом среднего расстояния между ними. Параметр асимметрии Rsk практически не коррелирует с амплитудными параметрами (ρ от $-0,20$ до $-0,01$). Это свидетельствует о том, что асимметрия распределения высот является независимой характеристикой, не связанной напрямую с абсолютными значениями неровностей или Rsm . Параметр эксцесса Rku демонстрирует умеренные отрицательные связи с Rz , Rt и Rsm (ρ от $-0,40$ до $-0,47$). Профили с более высокими амплитудными значениями и увеличенным шагом неровностей характеризуются менее «остроконечным» распределением высот.

Полученные результаты указывают на относительную независимость Rsk и Rku от других параметров, в связи с этим дальнейший анализ требует обращения не только к корреляционной структуре, но и к характеру распределения самих параметров.

Поскольку параметры Rsk и Rku описывают форму распределения высот, а не геометрию профиля, то их корректная интерпретация требует проверки соответствия эмпирического распределения нормальному закону. Параметр асимметричности Rsk характеризует различие топографии выступов и впадин. Он обладает большей дисперсией по сравнению с параметром Rq и вычисляется по выражениям

$$Rsk = \frac{1}{Rq^3} \left[\frac{1}{lr} \int_0^l z^3(x) dx \right]$$

или

$$Rsk = \frac{1}{NRq^3} \sum_{i=1}^N z^3.$$

Здесь Rq – типовой статистический параметр распределения, характеризующий ширину гистограммы распределения ординат профиля $z(x)$:

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

Параметр Rku характеризует отличие такого распределения от нормального:

$$Rku = \frac{1}{Rq^4} \left[\frac{1}{lr} \int_0^l z^4(x) dx \right]$$

или

$$Rku = \frac{1}{NRq^4} \sum_{i=1}^N z^4.$$

Для корректной интерпретации этих параметров необходимо проверить, насколько распределение высот профиля соответствует закону нормального распределения. Отклонения от нормальности напрямую отражаются в величинах Rsk и Rku , что делает анализ распределения ключевым этапом исследования.

При условии $f > 100$, если нулевую гипотезу о соответствии распределения параметров Rsk и Rku нормальному закону по критерию согласия Пирсона отклонить нельзя ($\alpha > 0,01$), распределение считается нормальным, так как графики, значения асимметрии и эксцесса не свидетельствуют о противоположном (рис. 2).

Определено, что значения ряда отклоняются от среднего $0,52$ мкм в среднем на $0,17$ мкм для Rsk , а для Rku – от среднего $2,09$ мкм на $0,18$ мкм. Проверка по критерию Шапиро – Уилка показала следующие результаты: $W = 0,98$; $p = 0,76 > 0,05$ для Rsk и $W = 0,99$; $p = 0,97 > 0,05$ для Rku , что свидетельствует об отсутствии осно-

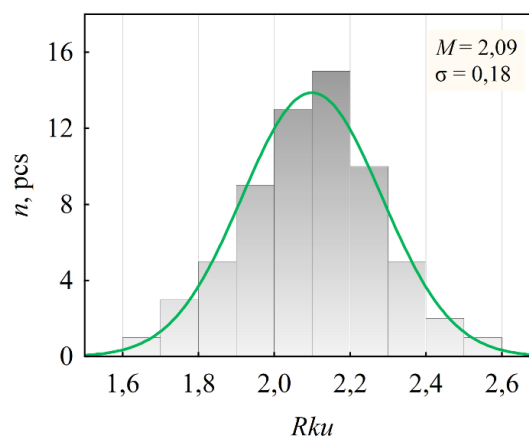
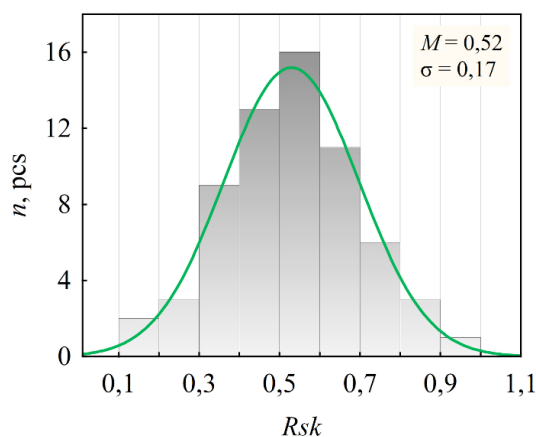


Рис. 2. Распределение параметров Rsk и Rku с гистограммами числа наблюдений

Fig. 2. Distribution of the Rsk and Rku parameters with frequency histograms

ваний для отклонения предположения о нормальном законе распределения рассматриваемых параметров. Проверка нормальности распределения с использованием критерия Андерсона – Дарлинга показала, что статистика для обеих выборок оказалась ниже критического значения: $A^2(Rsk) = 0,27 < 0,73$ и $A^2(Rku) = 0,16 < 0,73$. Это означает, что нет оснований отвергать нулевую гипотезу о соответствии распределения нормальному закону.

Результаты проверки гипотез для параметров Rsk и Rku свидетельствуют об отсутствии оснований для отклонения предположения о нормальном законе распределения, что обеспечивает методическую валидность последующего анализа.

Установленное соответствие распределения высот нормальной модели позволяет корректно использовать параметры формы в составе регрессионных зависимостей. На следующем этапе исследования была проведена оценка мультиколлинеарности факторов, включающих в себя технологические параметры подачи, угла и диаметра, с целью исключения линейной зависимости между предикторами и обеспечения устойчивости оценок коэффициентов. Если значение VIF превышает 10, то фиксируется высокая мультиколлинеарность. Усредненные значения VIF для fz , γ и D приблизительно равны единице, что свидетельствует об их независимости и отсутствии критической взаимосвязи. Таким образом, указанные факторы могут быть одновременно включены в регрессионные модели без риска искажения коэффициентов. При этом целесообразно использовать все три предиктора

для моделей, а выбор откликов (Rz , Rku , Rsk) будет определять их специфику.

После алгебраического развертывания в исходных переменных получим следующие модели:

$$Rz = -3,0975 + 37,1750fz + 0,0228\gamma + \\ + 0,1800D - 0,2775fz\gamma - 2,1667fzD - \\ - 0,0013\gamma D + 0,0167fz\gamma D;$$

$$Rsk = 1,1671 - 1,4562fz - 0,0346\gamma - 0,0283D + \\ + 0,0849fz\gamma + 0,0968fzD + \\ + 0,0027\gamma D - 0,0076fz\gamma D;$$

$$Rku = 4,2348 - 3,5884fz - 0,0392\gamma - \\ - 0,1822D + 0,0515fz\gamma + 0,2972fzD + \\ + 0,0035\gamma D - 0,0041fz\gamma D.$$

Для модели Rku : MAE = 0,1159 и RMSE = 0,1300. Упростим модель, исключив слабые взаимодействия. Тогда

$$Rku = 3,4087 - 2,8967fz - 0,0055\gamma - 0,0942D + \\ + 0,0067fz\gamma + 0,2422fzD.$$

Здесь значения абсолютной ошибки между предсказанными и фактическими значениями MAE = 0,1159 и RMSE = 0,1303. Упрощенная регрессионная модель демонстрирует сопоставимые значения метрик качества по сравнению с полной, при этом обладает меньшей структурной сложностью и более высокой интерпретируемостью коэффициентов.

Для оценки согласованности трех регрессионных моделей и анализа взаимосвязи ошибок

была рассчитана ковариационная матрица остатков. При оценке ковариационной матрицы ошибок использовался фрагмент экспериментальных данных объемом $n = 20$ наблюдений, соответствующих реализованным сочетаниям уровней факторов fz , γ и D . Каждое регрессионное уравнение содержит $k = 8$ параметров (свободный член, три линейных эффекта и четыре члена взаимодействия), что определяет число степеней свободы для ошибки как $n - k = 20 - 12 = 8$. Оценка ковариационной матрицы ошибок выполнялась с нормированием на $n - k$, что обеспечивает несмещенность оценок дисперсий и ковариаций остатков. На основе выборки из 20 наблюдений для каждого уравнения были вычислены прогнозные значения Rz' , Rsk' и Rku' , после чего сформирована матрица остатков E размерности 20×3 . Оценка ковариационной матрицы ошибок выполнялась по матричной формуле $\Sigma' = \frac{1}{n - k} E^T E$. Полученная ковариационная

матрица имеет вид

$$\Sigma' = \begin{bmatrix} 0,0000252 & -0,0000165 & 0,0000230 \\ -0,0000165 & 0,0000110 & -0,0000153 \\ 0,0000230 & -0,0000153 & 0,0000214 \end{bmatrix}.$$

Диагональные элементы матрицы Σ' отражают дисперсии ошибок отдельных моделей и показывают, что все три уравнения обладают высокой точностью аппроксимации. Дисперсии ошибок моделей соответствуют малым отклонениям между фактическими и рассчитанными значениями: $\varepsilon(Rz) = 2,52 \cdot 10^{-5}$; $\varepsilon(Rsk) = 1,10 \cdot 10^{-5}$; $\varepsilon(Rku) = 2,14 \cdot 10^{-5}$. Внедиагональные элементы матрицы отражают ковариации ошибок между моделями, демонстрируя слабую отрицательную взаимосвязь между парами Rz - Rsk и Rsk - Rku , а также умеренную положительную связь между ошибками Rz и Rku . Такая структура ковариаций свидетельствует о более согласованной реакции параметров Rz и Rku на изменения технологических факторов, тогда как параметр Rsk проявляет относительную независимость.

Полученная ковариационная структура подтверждает корректность построенных моделей и отсутствие систематических смещений в остатках. Низкие значения дисперсий ошибок и отсутствие выраженной мультиколлинеарности свидетельствуют о статистической устойчиво-

сти регрессионных уравнений и возможности их дальнейшего использования для прогнозирования параметров шероховатости в зависимости от технологических условий фрезерования.

Таким образом, проведенный анализ показал, что амплитудные параметры шероховатости обладают высокой взаимосвязанностью и устойчивостью оценок при увеличении объема выборки, тогда как параметры формы распределения (Rsk и Rku) требуют отдельного рассмотрения из-за их чувствительности к структуре данных. Проверка нормальности распределения подтвердила корректность их использования в регрессионных моделях, что позволило построить статистически значимые зависимости между технологическими параметрами и параметрами шероховатости поверхности.

Заключение

В совокупности полученные результаты подтверждают, что предложенная методика обеспечивает статистически надежное построение корреляционно-регрессионных моделей, адекватно описывающих влияние технологических параметров фрезерования на характеристики шероховатости поверхности. Полученные данные позволяют выделить следующие наиболее существенные выводы.

1. Разработан методический подход к определению минимально необходимого объема выборки, основанный на сочетании анализа парных коэффициентов корреляции Пирсона и Спирмена, расчета требуемого числа наблюдений для заданных уровней значимости и мощности теста, а также построения матрицы минимального объема выборки для различных пар параметров шероховатости. Показано, что амплитудные параметры (Ra , Rq , Rz , Rt , Rc , Rv) демонстрируют высокую устойчивость оценок и могут быть надежно охарактеризованы уже при относительно малых выборках, тогда как параметры формы распределения (Rsk и Rku) требуют существенно большего объема данных, что подчеркивает их специфическую информативность и необходимость отдельного анализа. Тем самым разработанный подход позволяет не только формально оценить минимальное число наблюдений, но и дифференцированно подходить к выбору объема выборки в зависимости от природы анализируемых параметров.

2. Сопоставлены результаты, полученные при различных объемах выборки ($n = 16$ и $n = 128$), что позволило проанализировать влияние числа наблюдений на устойчивость коэффициентов корреляции, а также на точность аппроксимации и величину ошибок предсказания. Установлено, что при малом объеме выборки корреляционные зависимости, особенно с участием параметров Rsk и Rku , имеют тенденцию к завышению и неустойчивости, тогда как при увеличении числа наблюдений корреляционная структура выравнивается, а оценки становятся более робастными по отношению к отдельным выбросам и гетерогенности подвыборок. Дополнительно в данном исследовании определено, что при $n \approx 128$ достигается баланс между статистической точностью оценок и практической реализуемостью эксперимента.

3. Показано, что для задач анализа и моделирования, связанных преимущественно с амплитудными параметрами шероховатости, может быть использован сравнительно небольшой объем выборки, тогда как при включении в модели параметров формы распределения (Rsk и Rku) требуется существенное увеличение числа наблюдений и проверка нормальности распределения. Разработанный подход позволяет целенаправленно планировать эксперимент, выбирая объем выборки с учетом как статистической мощности и требуемой точности оценок, так и ресурсных ограничений производства.

Список литературы

1. Система автоматизированного технологического управления износостойкостью деталей машин при обработке резанием / А.Г. Суслов, Д.М. Медведев, Д.И. Петрешин, О.Н. Федонин // Научные технологии в машиностроении. – 2018. – № 5 (83). – С. 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. A correlational study of areal surface texture parameters on some typical machined surfaces / Q. Qi, T. Li,

P.J. Scott, X. Jiang // Procedia CIRP. – 2025. – Vol. 27. – P. 149–154. – DOI: 10.1016/j.procir.2015.04.058.

5. Study on contact performance of ultrasonic-assisted grinding surface / Y. Wen, J. Tang, W. Zhou, C. Zhu // Ultrasonics. – 2019. – Vol. 91. – P. 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. – P. 102–112. – DOI: 10.1016/j.triboint.2011.11.008.

7. Формирование параметров шероховатости на основе корреляционных связей при чистовом фрезеровании пространственно-сложных поверхностей / М.Р. Гимадеев, В.М. Давыдов, А.В. Никитенко, А.В. Сарыгин // Упрочняющие технологии и покрытия. – 2019. – Т. 15, № 6 (174). – С. 243–248.

8. Correlation between surface roughness parameters and contact stress of gear / D. Yang, J. Tang, W. Zhou, Y. Wen // Proceedings of the institution of mechanical engineers, Part J: Journal of Engineering Tribology. – 2020. – Vol. 235 (3). – P. 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). – IEEE, 2018. – P. 927–936. – DOI: 10.1109/IAEAC.2018.8577706.

10. A surface quality prediction model considering the machine-tool-material interactions / M. Guo, W. Xia, C. Wu, C. Luo, Z. Lin // The International Journal of Advanced Manufacturing Technology. – 2024. – Vol. 131 (7–8). – P. 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). – P. 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). – P. 800–806. – DOI: 10.3103/S1068798X24701259.

13. 3D curved surface milling modeling for the topography simulation and surface roughness prediction / C. Chen, C. Wu, T. Zhang, S.Y. Liang // Journal of Manufacturing Processes. – 2025. – Vol. 137. – P. 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). – P. 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). – P. 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. – P. 37–46. – DOI: 10.22158/wjssr.v3n1p37.

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

18. *Пономарев Б.Б., Нгуен Ш.Х.* Оценка шероховатости при пятикоординатном чистовом фрезеровании поверхностей сфероцилиндрической фрезой // *Известия высших учебных заведений. Машиностроение*. – 2020. – № 5 (722). – С. 21–31. – DOI: 10.18698/0536-1044-2020-5-21-31.

19. Экспериментальное исследование динамики процесса механообработки концевыми сфероци-

линдрическими фрезами / М.Р. Гимадеев, А.А. Ли, В.О. Беркун, В.А. Стельмаков // *Обработка металлов (технология, оборудование, инструменты)*. – 2023. – Т. 25, № 1. – С. 44–56. – DOI: 10.17212/1994-6309-2023-25.1-44-56.

20. *Гимадеев М.Р., Ли А.А.* Анализ систем автоматизированного обеспечения параметров шероховатости поверхности на основе динамического мониторинга // *Advanced Engineering Research*. – 2022. – Т. 22 (2). – С. 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). – P. 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). – P. 1247–1250. – DOI: 10.5194/gmd-7-1247-2014.

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

© 2026 Авторы. Издательство Новосибирского государственного технического университета. Эта статья доступна по лицензии Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная (<https://creativecommons.org/licenses/by/4.0>).



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 Tihokeanskaya 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).

References

1. Suslov A.G., Medvedev D.M., Petreshin D.I., Fedonin O.N. Sistema avtomatizirovannogo tekhnologicheskogo upravleniya iznosostoykost'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.

* Corresponding author

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

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., Erzincanlı 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. Mashinostroyeniye = 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.

© 2026 The Authors. Published by Novosibirsk State Technical University. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0>).