SYMBOLIC REGRESSION FOR LEARNING SCALE TRANSITION EQUATIONS IN SYNTHETIC FRACTAL SURFACE ROUGHNESS

Aneesh Chatrathi
Rock Ridge High School

Zayan Hasan
Academies of Loudoun

ABSTRACT

Modeling surface roughness in materials science is a challenging multiscale problem, as surface textures often exhibit hierarchical (fractal-like) structure across multiple scales. In this work, we present a synthetic data-driven approach to studying scale transitions in surface roughness using fractal data generation and symbolic regression. We construct coarse-grained representations of synthetic fractal surfaces and apply symbolic regression to derive interpretable mathematical expressions that map fine-scale features to coarse-scale behavior. On controlled synthetic data, our approach achieves high predictive accuracy (R^2 near 1, low MSE), serving as a baseline validation. While the data is idealized, these results suggest that symbolic regression can capture scale-transition relationships in hierarchical surface structures and may also be able to support future efforts in data-driven multiscale modeling. This work highlights the potential of symbolic learning in accelerating modeling workflows for complex physical systems.

1 INTRODUCTION

Surface roughness plays a crucial role in determining mechanical, tribological, and adhesive properties, affecting applications ranging from machining to contact mechanics (Greenwood & Williamson, 1966). The statistical nature of surface asperities influences real contact area, friction, and wear. The fractal characteristics of fracture surfaces in metals were demonstrated by Mandelbrot et al. (1984), revealing self-affine properties that contribute to the understanding of roughness at multiple scales. This fractal perspective was later applied to contact mechanics by Majumdar & Bhushan (1991), who developed a fractal-based elastic-plastic contact model, establishing a power-law relationship between contact area and load.

Machine learning (ML) facilitates data-driven roughness modeling, complementing physics-based approaches that require extensive domain knowledge and can be computationally expensive (Motta et al., 2022; Benardos & Vosniakos, 2002). Symbolic regression (SR) has emerged as a powerful method for discovering analytical equations governing physical systems (Schmidt & Lipson, 2009), though its direct application to hierarchical roughness transitions remains largely unexplored.

We propose a synthetic fractal roughness generator and use SR to derive interpretable mathematical expressions for scale transitions. Our coarse-graining approach extracts multiscale features, and SR identifies relationships characterizing roughness behavior across scales. The results demonstrate high predictive accuracy while preserving interpretability, highlighting SR's potential for multiscale modeling in materials science.

2 RELATED WORKS

ML-based surface roughness modeling has primarily emphasized predictive accuracy, with Motta et al. (2022) and Benardos & Vosniakos (2002) employing machine learning models to predict roughness from machining parameters. These approaches, however, rely on black-box models that lack interpretability. In contrast, SR offers explicit mathematical expressions. Zhao & Zhao (2025) utilized SR to formulate an empirical equation for the joint roughness coefficient (JRC), while Torres-Treviño et al. (2013) and Asadollahi-Yazdi et al. (2021) applied SR to model roughness in machining and 3D-printed surfaces, respectively. Nevertheless, these studies do not explore hierarchical scale transitions in surface roughness modeling.

Multiscale modeling techniques, such as coarse-graining, have been extensively employed to link fine-scale and macro-scale properties in materials science (Steinhauser, 2017). These methods provide computational efficiency but typically depend on predefined physics-based assumptions rather than data-driven discovery.

Our approach integrates fractal roughness generation, coarse-graining, and SR into a unified framework for learning scale transition equations. Unlike prior ML-based models, our method explicitly derives interpretable mathematical relationships, offering new insights into hierarchical roughness modeling.

3 METHODOLOGY

3.1 SYNTHETIC FRACTAL SURFACE ROUGHNESS GENERATION

To model hierarchical surface roughness, we generate synthetic fractal-like data using a multiscale perturbation approach. We generate synthetic fractal roughness by refining a base function $f(x)$, typically chosen as $sin(x)$, with progressively smaller perturbations:

$$y(x) = f(x) + \sum_{d=1}^{D} A_d \sin(2\pi S^d x + \phi_d) \tag{1}$$

where:

- D is the number of refinement levels (depth),
- S is the scale factor controlling frequency increase per level,
- $A_d = A_0 \cdot \lambda^d$ is the amplitude decay per level, with $0 < \lambda \leq 1$,
- $\phi_d \sim U(-\pi, \pi)$ is a random phase shift per level.

The generated roughness is normalized for numerical stability:

$$y_{\text{norm}} = \frac{y - \mu_y}{\sigma_y} \tag{2}$$

where μ_y and σ_y are the mean and standard deviation of y, respectively.

3.2 SCALE TRANSITION VIA COARSE-GRAINING

To learn scale transition rules, we apply a coarse-graining operation where fine-scale roughness data $y(x)$ is aggregated into larger-scale descriptors. Coarse-graining reduces the dimensionality of roughness data by aggregating local values within a window W:

$$Y_i = g(y_{iW}, \ldots, y_{(i+1)W}) \tag{3}$$

where $g(\cdot)$ is an aggregation function:

- Mean smoothing: $g = \frac{1}{W} \sum y$ (reduces noise while preserving structure),
- Median filtering: $g = \text{median}(y)$ (robust to outliers),
- Trimmed mean: $g = \text{trim}(y, p)$ (ignores extreme values).

If the dataset length is not evenly divisible by W, the remaining points are aggregated as a smaller window. To retain the most informative features, we apply variance-based feature selection to the fine-scale data windows after aggregation, retaining the top k features with the highest variance.

3.3 Symbolic Regression for Scale Transitions

We employ symbolic regression via PySR to discover an interpretable mathematical equation that governs the relationship between fine-scale and coarse-grained roughness representations. Symbolic regression searches for an equation mapping fine-scale to coarse-scale roughness. We use a constrained operator set with complexity penalization to favor interpretable expressions.

We apply early stopping and an 80/20 train-test split to ensure generalization.

3.4 Validation, Residual Analysis, and Uncertainty Estimation

Model performance is evaluated on an 80/20 train-test split using MSE, MAE, RMSE, and R^2. We also analyze the residuals $(Y_{\text{test}} - Y_{\text{pred}})$ to assess model fit. Uncertainty is quantified by training multiple bootstrapped models (each with reduced iterations, $max\_iters = 50$) on resampled training data and computing the standard deviation of their predictions:

$$\sigma_{\text{uncertainty},i} = \text{std}(\hat{Y}_{1,i}, \hat{Y}_{2,i}, \ldots, \hat{Y}_{B,i}) \tag{4}$$

where B is the number of bootstrap iterations and $\hat{Y}_{b,i}$ is the prediction from the b-th model for test sample i.

4 Results and Discussion

4.1 Synthetic Surface Roughness Generation

The synthetic roughness generator produced fractal-like structures by iteratively perturbing a base function. Figure 1 shows the generated profile, capturing both large-scale curvature and fine-scale variations. Normalization ensured numerical stability.

4.2 Scale Transition Mapping

Coarse-graining was applied to analyze scale transitions. Figure 2 shows fine-to-coarse representations, with red markers indicating aggregated values. The process preserves key roughness features while capturing large-scale trends.

4.3 Symbolic Regression Modeling

Symbolic regression was employed to derive an explicit mathematical relationship governing the transition between fine and coarse scales. Figure 3 compares the true coarse-scale values with predictions made by the symbolic regression model. The alignment between predicted and actual values indicates that the discovered equation generalizes well to unseen data. Error bars derived from bootstrap resampling provide an estimate of uncertainty, suggesting that the model remains robust across different test samples.

4.4 Residual Analysis and Model Evaluation

To assess model accuracy, standard error metrics were computed. The symbolic regression model achieved a mean squared error (MSE) of 3×10^{-6}, root mean squared error (RMSE) of 0.00185, and coefficient of determination (R^2) of 0.999997, indicating a high degree of accuracy. The residual distribution (Figure 4) exhibits a visually symmetric error pattern, suggesting little systematic bias in the predictions.

Additionally, bootstrap uncertainty quantification revealed a mean uncertainty of 0.0024 with a standard deviation of 0.00056, confirming that the model produces consistent results across resampled datasets.

4.5 Implications for Surface Roughness Modeling

These results suggest that symbolic regression can capture interpretable scale-transition equations in synthetic hierarchical roughness data. Unlike traditional black-box models, it yields explicit functional forms that could inform physics-based roughness modeling or improve coarse-graining in multiscale simulations. On clean synthetic data with known structure, the model achieves high accuracy, serving as a baseline validation. Future work will address more complex and real-world scenarios to assess generalizability.

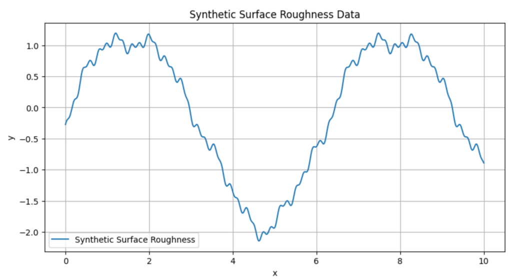

Figure 1: Synthetic surface roughness data.

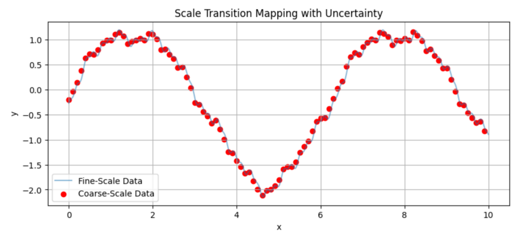

Figure 2: Scale transition mapping.

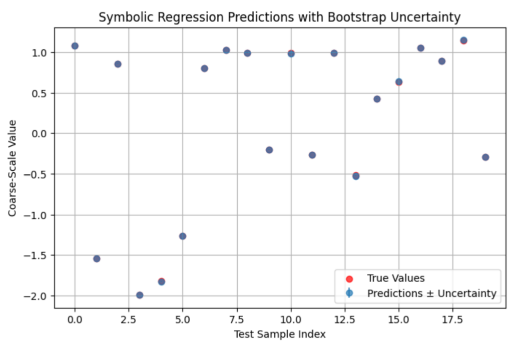

Figure 3: Symbolic regression predictions.

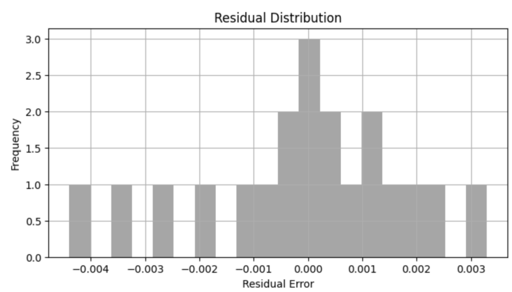

Figure 4: Residual distribution of the symbolic regression model.

Acknowledgments

This manuscript was prepared with the assistance of generative AI tools, specifically OpenAI's ChatGPT-4o, throughout its development. AI was used to refine language clarity, grammar, sentence structure, and overall readability. The AI's role was strictly limited to non-substantive tasks, such as improving linguistic coherence, without altering or contributing to the technical content, analyses, methodologies, or results.

Specifically, AI tools were used to:

Enhance the readability and flow of the Abstract, Introduction, Related Works, Methodology, Results, and Discussion sections. Verify grammar and structure in all technical descriptions and explanations, ensuring accessibility without modifying the underlying scientific meaning. All conceptual, analytical, and experimental work, derivation of equations, design of experiments, implementation of algorithms, and interpretation of results, was conducted solely by the authors.

The authors carefully reviewed and validated all AI-edited content to ensure its alignment with the intended meaning and scientific integrity of the work. No AI was involved in generating technical content, equations, analyses, results, or figures.

REFERENCES

E. Asadollahi-Yazdi, J. Gardan, and P. Lafon. Generic roughness meta-model in 3d printing by fused deposition modeling. *Progress in Additive Manufacturing*, 2021. doi: 10.1007/s40964-021-00237-8.

P. G. Benardos and G.-C. Vosniakos. Predicting surface roughness in machining: A review. *International Journal of Machine Tools and Manufacture*, 2002. doi: 10.1016/S0890-6955(03)00059-2.

J. A. Greenwood and J. B. P. Williamson. Contact of nominally flat surfaces. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 1966. doi: 10.1098/rspa.1966.0242.

A. Majumdar and B. Bhushan. Fractal model of elastic-plastic contact between rough surfaces. *Journal of Tribology*, 1991. doi: 10.1115/1.2920588.

B. B. Mandelbrot, D. E. Passoja, and A. J. Paullay. Fractal character of fracture surfaces of metals. *Nature*, 1984. doi: 10.1038/308721a0.

M. P. Motta, C. Pelaingre, A. Delamézière, L. Ben Ayed, and C. Barlier. Machine learning models for surface roughness monitoring in machining operations. *Procedia CIRP*, 2022. doi: 10.1016/j.procir.2022.03.110.

M. Schmidt and H. Lipson. Distilling free-form natural laws from experimental data. *Science*, 2009. doi: 10.1126/science.1165893.

M. Steinhauser. Multiscale modeling, coarse-graining, and shock wave computer simulations in materials science. *AIMS Materials Science*, 2017. doi: 10.3934/matersci.2017.6.1319.

L. M. Torres-Treviño, I. Escamilla, B. Gonzalez, R. Praga-Alejo, and P. Pérez-Villanueva. Modeling cutting machining process using symbolic regression alpha-beta. *The International Journal of Advanced Manufacturing Technology*, 2013. doi: 10.1007/s00170-012-4655-5.

Y. Zhao and H. Zhao. Symbolic regression for the determination of joint roughness coefficient. *Mathematical and Computational Applications*, 2025. doi: 10.3390/mca30010017.