CHEMICAL ENGINEERING TRANSACTIONS
VOL. 119, 2025
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.cetjournal.it
Guest Editors: David Bogle, Flavio Manenti, Piero Salatino

ISBN 979-12-81206-20-5; **ISSN** 2283-9216
DOI: 10.3303/CET25119086

Machine Learning-Assisted Optimization of Ultrasonic Dispersion Parameters for Graphene Nanoribbons in Paint Solvents: a Statistical Analysis

Manh Huan Nguyen [a,*], Huynh Hung My Nguyen [a], Minh Thuan Huynh[a], Binh Phan Minh Quoc [a], Ngoc Kien Pham [a], Thanh Sang Nguyen [a]

[a]Vietnam Petroleum Institute, Vietnam Petroleum Institute Building, Lot E2b-5, D1 Street, High-Tech Park, TanPhu Ward, ThuDuc City, HoChiMinh City, Vietnam
huannm.pvpro@vpi.pvn.vn

This study investigates the optimization of graphene nanoribbons (GNRs) dispersion in coating solvents using high-power ultrasonication, through an integrated framework that combines experimental design with machine learning techniques. Dispersion stability was assessed via polydispersity index (PI) measurements, with a stability threshold defined at PI < 0.5. A full factorial Design of Experiments (DOE) was implemented to examine the influence of three key parameters: sonication time (5–60 minutes), temperature (40–70 °C), and GNRs concentration (50–1000 ppm). Statistical analysis using ANOVA confirmed the significant impact of all three factors ($p < 0.05$), with sonication time and concentration exerting the strongest effects on dispersion stability ($F = 69.156$ and 67.561, respectively). Machine learning models were developed to predict optimal dispersion conditions. Among the models evaluated, Random Forest exhibited the best predictive performance (MSE = 0.251, MAE = 0.373, $R^2 = 0.770$), outperforming both Linear Regression ($R^2 = 0.672$) and Decision Tree ($R^2 = 0.110$). This suggests that Random Forest effectively captures both linear and mildly nonlinear relationships between process variables and dispersion outcomes. A PI of 0.20 ±0.02 was found during experimental validation under the best conditions indicated by the model. This showed that the model is robust and reliable. This research provides a scalable approach to improve the dispersion of nanomaterials in solvents used in coating systems, doing this using traditional statistical tools and advanced machine learning methods that can work together to provide more accurate process control and predictions.

1. Introduction

The effective dispersion of graphene nanoribbons (GNRs) in coating solvents poses a significant challenge in the development of high-performance protective coatings. Nielsen's barrier theory (Nielsen, L.E., 1967) illustrates that the incorporation of well-dispersed nanomaterials can markedly improve the protective characteristics of coatings by establishing tortuous pathways that impede corrosive agents. Achieving stable dispersions of GNRs is challenging due to the inherent aggregation tendencies induced by strong van der Waals forces in nanomaterial systems. Dispersion stability is crucial for assessing the performance of coatings that contain GNR. Min et al. (2016) conducted studies on the dispersion and stabilization of graphene nanoribbons in organic solvents, revealing the critical impact of solvent choice and processing parameters on dispersion stability. The parameters include sonication time, power, temperature, and GNR concentration, all of which demonstrate intricate interrelationships that influence the final dispersion quality. Efficient dispersion of graphene nanoribbons (GNRs) in solvents, the main components of coating systems, poses a significant challenge in the development of high-performance protective coatings. Nielsen's barrier theory (Nielsen, L.E., 1967) suggests that the incorporation of well-dispersed nanomaterials can significantly improve the protective properties of coatings by establishing tortuous pathways that prevent corrosive agents from contacting the protective surface. The stable dispersion of GNRs in solvent-based paint systems is a challenge due to the inherent tendency of nanoparticles to agglomerate due to van der Waals forces between them and between

Paper Received: 8 June 2025; Revised: 17 June 2025; Accepted: 29 September 2025
Please cite this article as: Nguyen M.H., Nguyen H.H.M., Huynh M.T., Phan Minh Quoc B., Pham N.K., Nguyen T.S., 2025, Machine Learning-Assisted Optimization of Ultrasonic Dispersion Parameters for Graphene Nanoribbons in Paint Solvents: a Statistical Analysis, Chemical Engineering Transactions, 119, 511-516 DOI:10.3303/CET25119086

them and the solvent, the stability of the dispersion is important to evaluate the performance of GNRs-containing coatings. Min et al. (2016) studied the dispersion and stability of graphene nanoribbons in organic solvents, showing that sonication time, power, temperature and GNRs concentration affected the final dispersion quality of the nanosystem.

The intricacy of nanomaterial dispersion has resulted in more advanced optimization strategies. Wan et al. (2021) employed Response Surface Methodology (RSM) to statistically optimize the process parameters for graphene dispersion, demonstrating non-linear relationships among various processing variables. Turki and Kiran (2024) advanced modern computational methods by developing deep learning approaches for the characterization and optimization of graphene-based nanomaterials, illustrating the potential of artificial intelligence in dispersion optimization. The assessment of dispersion stability typically depends on the polydispersity index (PI) and zeta potential measurements as primary indicators. Sabih et al. (2024) developed a detailed experimental design framework aimed at optimizing ultrasonic dispersion parameters, identifying essential process variables and their interactions. Siti et al. (2018) established that PI values under 1 signify favourable dispersion stability in nanomaterial systems, with decreasing values associated with enhanced dispersion quality.

Koh et al. (2023) contributed to this field by creating machine learning models that predict nanoparticle dispersion stability in various solvent systems, demonstrating significant prediction accuracy. Their research enabled the application of machine learning to enhance dispersion conditions for targeted applications, thereby reducing the time and resources needed for experimental optimization. Despite these advancements, there is a scarcity of systematic studies that integrate experimental design methodologies with machine learning techniques for the optimization of GNRs dispersion. This research integrates Design of Experiments (DOE) with advanced machine learning algorithms to optimize the dispersion of GNRs in coating solvents, thereby addressing an existing gap in the literature. This study aims to: (1) create a comprehensive database detailing the effects of ultrasonic parameters on the dispersion stability of GNRs, (2) develop and train machine learning models to accurately predict dispersion quality based on input parameters, and (3) enhance the dispersion process by integrating experimental data with model predictions. This method seeks to reduce development time and enhance resource efficiency while maintaining uniform dispersion quality in industrial coating applications. This research marks a crucial advancement in the systematic application of learning-assisted optimization in nanomaterial processing, potentially establishing a novel framework for process optimization in advanced materials manufacturing.

2. Material and methods

2.1. Materials

Multilayer graphene (10 - 20 layers) with outer diameter (OD) of 7 - 12 nm and length > 1 μm was used in this study. Laboratory grade xylene was employed as the dispersion medium without any pre-treatment.

2.2. Experimental Design

The impacts of three critical process factors on graphene dispersion were studied using a full factorial design. These parameters were sonication duration (t) ranging from 5 to 60 minutes, temperature (T) from 40 to 70 °C, and graphene concentration (C) from 50 to 1000 ppm. The experimental matrix consisted of 27 trials incorporating diverse combinations of these characteristics. Dispersion stability was assessed using the Polydispersity Index (PI), aiming for a target value of PI < 0.5. PI quantifies sample heterogeneity according to size distribution, indicating possible agglomeration or aggregation during isolation and analysis. The selected parameter ranges were informed by prior research (Min et al., 2016) and initial trials to guarantee successful dispersion while preventing material degradation.

2.3. Experimental Procedure

The experimental investigation was carried out following a methodical sequence of procedures. Initially, dispersions of graphene-xylene were formulated at different concentrations following the established experimental design parameters. These dispersions were subsequently subjected to ultrasonic treatment using a high-power sonicator (2,000 W, 20 kHz) under carefully controlled temperature and temporal conditions. The temperature during processing was maintained using a thermostatic water bath connected to a jacketed beaker. The probe used was made of titanium alloy, minimizing the risk of contamination. The Polydispersity Index (PI) measurements were then performed utilizing Dynamic Light Scattering (DLS) technique, employing a HORIBA SZ-100 particle size analyzer. The resultant data were collected and subjected to comprehensive analysis to evaluate the dispersion characteristics.

2.4. Data Analysis Methods

The experimental data were analyzed through a comprehensive analytical framework encompassing statistical analysis and machine learning approaches. Analysis of Variance (ANOVA) was employed to evaluate the statistical significance of process parameters and their interactions, while multiple regression analysis was utilized to establish mathematical relationships between variables. All statistical analyses were performed using Python's scientific computing libraries (scipy.stats and statsmodels, version 1.7.0).

For predictive modeling, three machine learning algorithms were implemented using the scikit-learn library (version 0.24.2): Linear Regression (LR) for baseline linear relationships, Decision Tree (DT) for capturing non-linear parameter interactions, and Random Forest (RF) for ensemble-based predictions. Model performance evaluation was conducted using two complementary metrics: Mean Squared Error (MSE) and Mean Absolute Error (MAE). These metrics were selected to provide comprehensive assessment of prediction accuracy, with MSE being sensitive to large deviations through squared terms and MAE offering direct interpretation of average prediction errors. All computations and analyses were executed using Python 3.8 with NumPy (version 1.21.0) for numerical operations and data manipulation.

3. Result and discussion

At fixed conditions (T = 70 °C, C = 50 ppm), sonication time significantly influenced dispersion stability, with PI decreasing from 1.5 to 0.6 (5 to 32.5 minutes) and further to 0.2 (60 minutes). This reduction indicates effective disruption of GNRs agglomerates through cavitation effects and mechanical shearing forces, with 60 minutes representing optimal sonication time.

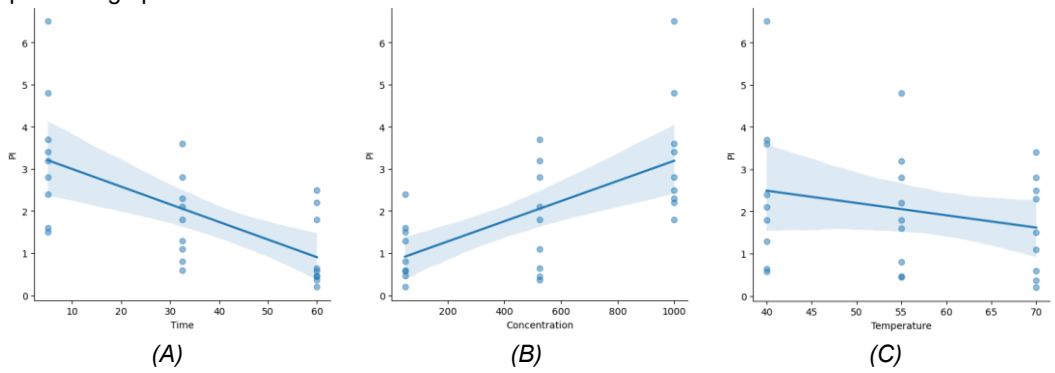

Figure 1. Relationships between PI and Variables of time (A), concentration (B), and temperature (C)

Temperature effects were evaluated at 50 ppm concentration. At 40 °C, PI values decreased with time: 2.4 (5 minutes), 1.3 (32.5 minutes), and 0.58 (60 minutes). Higher temperature (70 °C) showed improved stability: 1.5 (5 minutes), 0.6 (32.5 minutes), and 0.2 (60 minutes). Enhanced dispersion at elevated temperatures results from reduced solvent viscosity, increased molecular kinetic energy, and intensified cavitation effects, aligning with Sabih (2024).

Table 1. Full Factorial Design of Experiments and PI Values

No	Time (minutes)	Temperature (°C)	Concentration (ppm)	PI
1	5.0	40	50	2.40
2	5.0	40	525	3.70
3	5.0	40	1000	6.50
4	5.0	55	50	1.60
5	5.0	55	525	3.20
6	5.0	55	1000	4.80
7	5.0	70	50	1.50
8	5.0	70	525	2.80
9	5.0	70	1000	3.40
10	32.5	40	50	1.30
11	32.5	40	525	2.10
12	32.5	40	1000	3.60

13	32.5	55	50	0.80
14	32.5	55	525	1.80
15	32.5	55	1000	2.80
16	32.5	70	50	0.60
17	32.5	70	525	1.10
18	32.5	70	1000	2.30
19	60.0	40	50	0.58
20	60.0	40	525	0.65
21	60.0	40	1000	1.80
22	60.0	55	50	0.46
23	60.0	55	525	0.45
24	60.0	55	1000	2.20
25	60.0	70	50	0.20
26	60.0	70	525	0.36
27	60.0	70	1000	2.50

The observed improvement in dispersion stability at higher temperatures can be attributed to: (1) reduced solvent viscosity facilitating GNRs movement, (2) increased molecular kinetic energy promoting deagglomeration, and (3) enhanced cavitation effects from ultrasonication. GNRs concentration studies at optimal conditions (70 °C, 60 minutes) showed PI values increasing from 0.2 to 2.5 as concentration rose from 50 to 1000 ppm. This stability deterioration at higher concentrations stems from increased particle interactions and collision-induced aggregation. The optimal concentration was identified as 50 ppm, yielding PI = 0.2, well below Siti (2018) stability threshold of 0.5. While optimal dispersion was observed at the extremes of the studied ranges (e.g., maximum temperature and sonication time, minimum concentration), these choices were made based on practical constraints and material limitations. Extending temperature or time beyond current values risks degradation of the graphene nanoribbons and solvent evaporation, thus limiting process reproducibility.

The statistical parameters of the regression model are presented in Table 2. The model demonstrated robust statistical significance, with the intercept coefficient (3.760 ± 0.1825) and all process variables showing p-values substantially below the 0.05 significance threshold. Specifically, time (-0.040 ± 0.0028, $p = 1.28\times10^{-5}$), temperature (-0.030 ± 0.0031, $p = 6.50\times10^{-7}$), and concentration (0.024 ± 0.0018, $p = 8.11\times10^{-6}$) exhibited strong correlations with PI values. The negative coefficients for time and temperature indicate their inverse relationship with PI, while the positive coefficient for concentration suggests a direct correlation with the dependent variable.

Table 2. The statistical parameters of the regression model

Predictors	Coefficients	Std. Error	t value	Pr(>|t|)
Intercept	+ 3.760	0.566	6.650	0.000
Time	- 0.040	0.005	-0.316	0.000
Temperature	- 0.030	0.009	-3.162	0.004
Concentrate	+ 0.024	0.000	8.200	0.000

The Analysis of Variance (ANOVA) results (Table 3) further validated the model's reliability and significance. The time parameter demonstrated the strongest effect with a sum of squares of 23.805 ($F = 69.156$, $p = 2.19\times10^{-8}$), followed closely by concentration with a sum of squares of 23.256 ($F = 67.561$, $p = 2.67\times10^{-8}$). Temperature showed a comparatively smaller but still significant effect (sum of squares = 3.440, $F = 9.996$, $p = 4.36\times10^{-3}$). The resulting multiple linear regression equation was established as:

$PI = 3.760 - 0.040t - 0.030T + 0.024C$

where t represents sonication time (minutes), T represents temperature (°C), and C represents concentration (ppm).

Table 3. Analysis of Variance (ANOVA) for the regression model

	df	Sum_sq	Mean_sq	F	Pr(>|t|)
Time	1.0	23.805	23.805	69.156	$2.19.10^{-8}$
Temperature	1.0	3.440	3.440	9.996	$4.36.10^{-3}$
Concentration	1.0	23.256	23.256	67.561	$2.67.10^{-8}$
Residual	23.0	7.917	0.344	-	-

The model exhibited exceptional predictive capability with a coefficient of determination (R^2) of 0.864, indicating that 86.4% of the variance in PI values can be explained by the selected process parameters. This high R^2 value, combined with the extremely low p-values across all variables ($p < 0.001$ for time and concentration, $p <$

0.005 for temperature), confirms the model's robust predictive power for GNRs dispersion optimization. These findings align with similar statistical analyses reported in recent literature for nanoparticle dispersion systems (Koh, 2021).

The F-values from the ANOVA analysis indicate that time and concentration exert similar effects on the PI, whereas temperature has a lesser yet still significant influence. The residual mean square error of 0.344 suggests a strong correlation between predicted and observed values, thereby reinforcing the model's reliability for practical applications in GNRs dispersion optimization.

Three machine learning models—Linear Regression, Decision Tree, and Random Forest—were evaluated for their predictive efficacy in modeling the dispersion stability of GNRs, using the metrics of MSE, MAE, and R^2. The Random Forest model exhibited superior performance, with a mean squared error of 0.251, a mean absolute error of 0.373, and a coefficient of determination of 0.770. The results demonstrate that the model efficiently captures both linear and somewhat nonlinear connections. The elevated R^2 indicates that almost 77% of the variance in the dispersion index is elucidated by the model, signifying robust performance in practical experimental contexts. This validates the Random Forest model as a reliable and precise option, particularly in scenarios with minor nonlinearity.

The Linear Regression model followed, with MSE = 0.358, MAE = 0.465, and R^2 = 0.672. Although less precise than Random Forest, this model indicates that the correlation between ultrasonic processing parameters and dispersion stability is primarily linear. An R^2 value of 0.672 is typically seen acceptable for experimental data, indicating that Linear Regression is a viable and interpretable choice, especially in contexts where simplicity and transparency are stressed. The Decision Tree model exhibited subpar performance, yielding an MSE of 0.973, an MAE of 0.823, and a notably low R^2 of 0.110. This reflects severe overfitting and limited generalization, indicating that the model is unsuitable for this application, particularly with a small dataset.

In regression analysis, R^2 values above 0.75 are typically considered strong, 0.60–0.75 are moderate and acceptable in many applied research settings, while values below 0.40 suggest poor model performance. Therefore, the R^2 values of 0.770 (Random Forest) and 0.672 (Linear Regression) indicate that both models are sufficiently effective for predicting GNRs dispersion, especially when the dataset is limited in size and contains experimental variability. This further supports the use of these models in practical optimization scenarios. Overall, the findings demonstrate that Random Forest offers the highest predictive power, while Linear Regression balances accuracy with interpretability. The poor performance of the Decision Tree model reinforces the importance of model selection, particularly when working with small or noisy datasets.

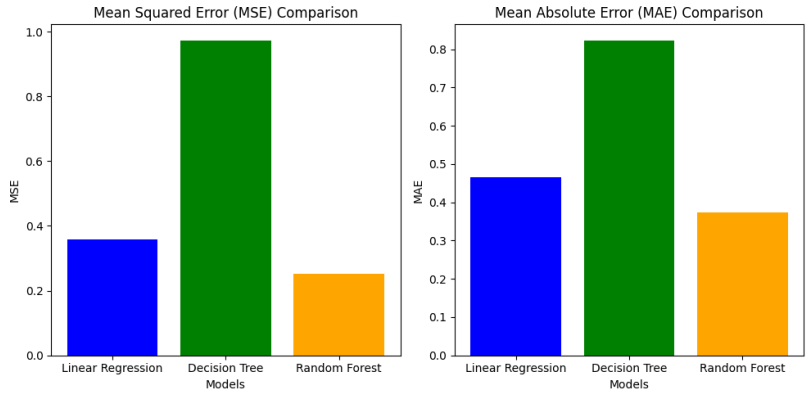

Figure 2. Comparison of Machine Learning Models through MSE and MAE

Validation of the optimal processing conditions (60 minutes sonication, 70 °C, 50 ppm GNRs concentration) through repeated trials yielded consistent PI values of 0.2 ± 0.02, closely matching the model's prediction of 0.187. The methodology shows promise for broader applications, including adaptation to various solvent, applying the model to butyl acetate and toluene solvents allowed for a significant reduction in the number of experiments, from 27 in the full DOE to only 7–10 to achieve comparable results. This efficiency offers clear advantages in terms of cost and R&D time. Furthermore, although 50 ppm provided the best dispersion results, extremely low concentrations may not be feasible in applications where surface coverage or conductivity are critical. A future multi-objective optimization method, considering both dispersion stability and application performance, will further enhance the industrial applicability of this work.

4. Conclusions

This study proposes an integrated methodology combining Design of Experiments, statistical modeling, and machine learning to optimize the dispersion of graphene nanoribbons in xylene solvent. Optimal dispersion was achieved at 60 minutes of sonication, 70 °C, and 50 ppm GNRs concentration, resulting in a polydispersity index (PI) of 0.20 ± 0.02—well below the threshold of 0.5 for stable nanodispersions. Statistical analysis using ANOVA and multiple linear regression revealed strong correlations between process parameters and dispersion stability. The regression model (PI = 3.760 – 0.040t – 0.030T + 0.024C) demonstrated high predictive reliability with R^2 = 0.864. Sonication time and GNRs concentration were identified as the most influential factors (F = 69.156 and 67.561, respectively; $p < 0.001$), while temperature also exhibited a statistically significant but less dominant effect (F = 9.996; $p < 0.005$).In the evaluation of machine learning models, Random Forest outperformed others with the highest R^2 (0.770) and the lowest prediction errors (MSE = 0.251, MAE = 0.373), suggesting that it effectively captures both linear and subtle nonlinear patterns in the data. Linear Regression also showed acceptable performance (R^2 = 0.672), supporting the presence of predominantly linear relationships. In contrast, Decision Tree displayed limited generalization capability (R^2 = 0.110), likely due to overfitting on the small dataset.These results validate that longer sonication times and elevated temperatures promote GNRs exfoliation through intensified cavitation and reduced solvent viscosity, while moderate concentrations prevent re-agglomeration. The integrated optimization framework presented here, anchored by statistically grounded modeling and effective machine learning tools, provides a practical and scalable approach for the industrial deployment of GNRs-based dispersions. Moreover, it offers a replicable methodology for optimizing other nanomaterial dispersion systems in future research and development efforts.

Acknowledgments

This research was funded by Vietnam Oil and Gas Group, grant numbers 3155/HD-DKVN, 10th June 2022, Code 01/HCBDK(VPI)/2022/KHCN.

Nomenclature

C - concentration (ppm)
DLS - Dynamic Light Scattering
DOE - Design of Experiments
GNRs - Graphene Nanoribbons
MAE - Mean Absolute Error
MSE - Mean Squared Error
OD - Outer Diameter (nm)
PI - Polydispersity Index
R^2 - coefficient of determination
t - sonication time (minutes)
T - temperature (°C)

References

Koh S., Yusuke S., Yasuhiko O, 2023, Prediction of Nanoparticle Dispersion by Machine Learning Using Various Molecular Descriptors, Industrial & Engineering Chemistry Research, 62 (46), 19683 – 19689. DOI: https://doi.org/10.1021/acs.iecr.3c02581

Nielsen, L.E., 1967, Models for the Permeability of Filled Polymer Systems. Journal of Macromolecular Science Part A: Chemistry, 5, 929 – 942. DOI: https://doi.org/10.1080/10601326708053745

Min Y. S., Young S. J., Na R. K., Hyoung J. J., 2016, Dispersion stability of chemically reduced graphene oxide nanoribbons in organic solvents, RSC advances, 23, 19389 – 19393. DOI: https://doi.org/10.1039/C5RA23801C

Sabih Q., Naveed R., Waquas A., 2024, Graphene dispersion, functionalization techniques and applications: A review, Synthetic Metals, 307, 117697. DOI: https://doi.org/10.1016/j.synthmet.2024.117697

Siti S. S., Ahmad A., Mohd N. M. Z., Shaifulazuar R., Mohd Z. Z., Mohd F. M. S., Mohammadreza S., 2018, Investigation of the thermophysical properties and stability performance of non-covalently functionalized graphene nanoplatelets with Pluronic P-123 in different solvents, Materials Chemistry and Physics, 206, 94 – 102. DOI: https://doi.org/10.1016/j.matchemphys.2017.12.008

Turki S. A., Kiran A., 2024, Machine learning approaches to predict the strength of graphene nanoplatelets concrete: Optimization and hyper tuning with graphical user interface, Materials Today Communications, 40(1), 109946. DOI: http://dx.doi.org/10.1016/j.mtcomm.2024.109946

Xin L., Bing Y., Jie C., Dongxia H., Jun L., Ding N., 2024, Optimized Functionalization of Graphene Oxide for Enhanced Mechanical Properties in Epoxy Resin Composites, Coattings, 14 (5), 609. DOI: https://doi.org/10.3390/coatings14050609

Wan N. W. B., Ruey S. C., Dalila S., Muhamamad J., Mohd Y., Mohamad J. S., Sahrim A., 2021, Statistical Optimization Using Response Surface Methodology for Enhanced Tensile Strength of Polyethylene/Graphene Nanocomposites, International Journal of Integrated Engineering, 13 (6), 109 – 117. DOI: https://doi.org/10.30880/ijie.2021.13.06.010