materials proceedings

Proceeding Paper

Solubility-Driven Prediction of Electrospun Nanofibers' Diameters via Generalized Linear Models

Marco Antonio Pérez-Castillo, Rubén Caro-Briones, Mariangely López-González, Gabriela Martínez-Mejía, Mónica Corea and Lazaro Ruiz-Virgen

MDPI

https://doi.org/10.3390/materproc2025025008

materials proceedings

MDPI

Proceeding Paper

Solubility-Driven Prediction of Electrospun Nanofibers' Diameters via Generalized Linear Models †

Marco Antonio Pérez-Castillo [1], Rubén Caro-Briones [1,2,*], Mariangely López-González [3,*], Gabriela Martínez-Mejía [1], Mónica Corea [1] and Lazaro Ruiz-Virgen [1]

[1] Escuela Superior de Ingeniería Química e Industrias Extractivas, Instituto Politécnico Nacional, Av. Luis Enrique Erro S/N, Unidad Profesional Adolfo López Mateos, Zacatenco, Alcaldía Gustavo A. Madero, Mexico City 07738, Mexico; mperezc1703@alumno.ipn.mx (M.A.P.-C.); gamartinezm@ipn.mx (G.M.-M.); mcorea@ipn.mx (M.C.); lruizv1900@alumno.ipn.mx (L.R.-V.)

[2] Escuela Superior de Ingeniería Mecánica y Eléctrica, Instituto Politécnico Nacional, Av. Luis Enrique Erro S/N, Unidad Profesional Adolfo López Mateos, Zacatenco, Alcaldía Gustavo A. Madero, Mexico City 07738, Mexico

[3] Departamento de Matemáticas, Facultad de Ciencias, Ciudad Universitaria, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico

* Correspondence: rcaro@ipn.mx (R.C.-B.); mariangelylg@ciencias.unam.mx (M.L.-G.)

† Presented at the 5th International Online Conference on Nanomaterials, 22–24 September 2025; Available online: https://sciforum.net/event/IOCN2025.

Abstract

Electrospinning is a versatile technique for producing polymer nanofibers whose morphology strongly influences their properties. This work developed predictive and inferential models for fiber diameter based on solution and process parameters. Polymer–solvent compatibility was described through cohesive energy-based solubility parameters such as Hansen and Flory–Huggins (χ). Twenty Generalized Linear Models (GLMs) were trained using both the raw response (Y) and its natural logarithm ($\ln Y$) under Gaussian, Gamma, and Inverse Gaussian distributions with different link functions. Models using $\ln Y$ showed better goodness-of-fit, with the Gamma distribution and identity link performing best. The final model, optimized via AIC-forward selection, achieved RMSE = 0.5862, $Corr^2$ = 0.7803, and MAPE = 0.0775. The Flory–Huggins parameter and solution concentration were identified as the most influential predictors, providing a reliable framework for controlling nanofiber diameter in electrospinning processes.

Keywords: electrospinning; nanofiber diameter prediction; generalized linear models; solubility interaction

Academic Editor: Sotirios Baskoutas

Published: 25 November 2025

Citation: Pérez-Castillo, M.A.; Caro-Briones, R.; López-González, M.; Martínez-Mejía, G.; Corea, M.; Ruiz-Virgen, L. Solubility-Driven Prediction of Electrospun Nanofibers' Diameters via Generalized Linear Models. *Mater. Proc.* **2025**, *25*, 8. https://doi.org/10.3390/materproc2025025008

1. Introduction

Electrospinning is an electrostatic nanofiber fabrication technique. Fibers can be made from many substrates, such as molten salts, composite materials, and mainly polymer solutions. Polymer solutions are made of polymers dissolved in a proper solvent. Many of the parameters of polymer solutions, such as viscosity and superficial tension, impact the morphology of the fiber [1], and these are strongly related to polymer-solvent system affinity [2].

The molecular compatibility of the polymer solution system can be determined via solubility parameters, e.g., Hansen ($\delta_d, \delta_p, \delta_h$), Hildebrand ($\delta$), and Flory–Huggins ($\chi$), which describe the cohesion energy likeliness as well as the molecular ordering and interaction strength between polymer and solvent through the Relative Energy Difference (RED) of the

species [3]. Figure 1 shows, in yellow and blue, the operational and solution parameters, respectively, for the electrospinning process that affects the nanofiber diameter.

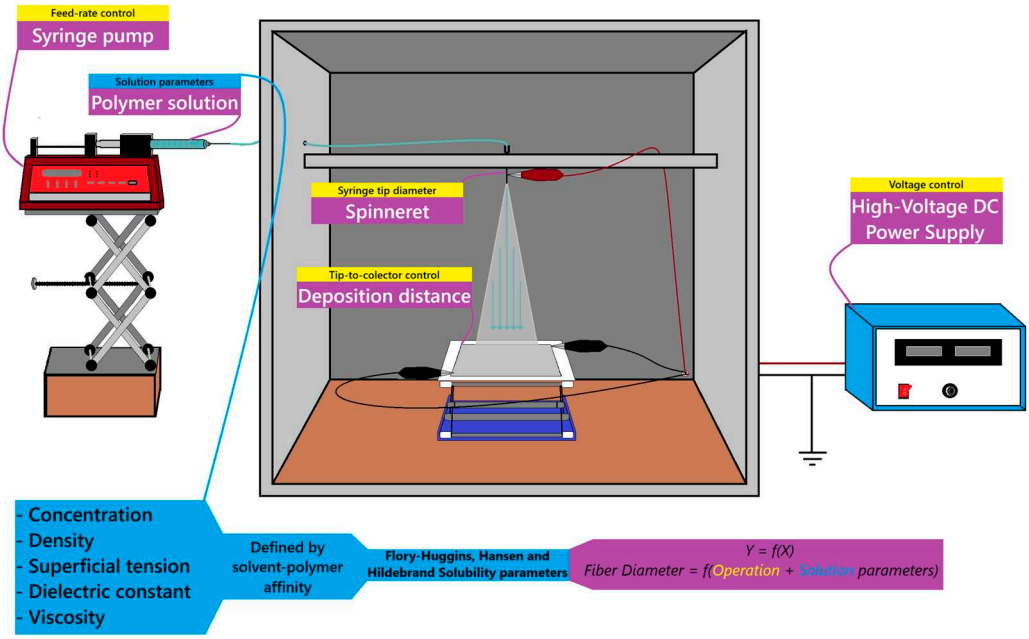


Figure 1. Experimental setup and parameters governing the fiber diameter in the electrospinning process.

Hildebrand first introduced the concept of solubility parameters in 1950 to describe miscibility between species based on their cohesive energy. He proposed that substances with similar solubility parameter values (δ) are likely to mix without phase separation [4]. Later, Hansen, from 1995 to 2000, refined this concept by identifying that solubility depends on different types of molecular interactions. He divided the total cohesive energy into three components—dispersive (δ_d), polar (δ_p), and hydrogen bonding (δ_h)—and represented them in a three-dimensional coordinate system to assess compatibility between species [3]. The relationship between the Hansen and Hildebrand parameters is given by

$$\delta = \sqrt{\delta_d^2 + \delta_p^2 + \delta_h^2} \quad (1)$$

The three-dimensional distance (Ra) and the RED can quantify the affinity of a solvent–polymer system:

$$Ra = \sqrt{4\left(\delta_{d2} - \delta_{d1}\right)^2 + \left(\delta_{p2} - \delta_{p1}\right)^2 + \left(\delta_{h2} - \delta_{h1}\right)^2} \quad (2)$$

Then

$$\mathrm{RED} = \frac{Ra}{R_0} \quad (3)$$

where

Ra = three-dimensional distance between solvent–polymer coordinates ($\mathrm{MPa}^{1/2}$).
δ_{d1}, δ_{p1}, δ_{h1} = Hansen solubility parameters of the polymer ($\mathrm{MPa}^{1/2}$).
δ_{d2}, δ_{p2}, δ_{h2} = Hansen solubility parameters of the solvent ($\mathrm{MPa}^{1/2}$).
R_0 = radius of interaction of the polymer ($\mathrm{MPa}^{1/2}$).
RED = Relative Energy Difference between the species. Systems with RED ≤ 1 are considered miscible, while RED > 1 are not.

Shrutidhara Sarma employed voltage, collector distance, and feed rate, together with the Hansen solubility parameters of the polymer solution, to evaluate polymer–solvent affinity through the Relative Energy Difference (RED) and predict the diameter of polyvinylidene fluoride (PVDF) fibers using artificial intelligence. The models were of the black-box type, meaning the relationship between inputs and outputs was not explicitly defined [5]. Model performance was assessed through the determination coefficient (R^2), representing the percentage of explained variance, and the Root Mean Squared Error (RMSE), which measures the average prediction error [6]. The best model, a Gradient Boosting Regressor (GBR), achieved an R^2 of 0.98 and an RMSE of 76.37 nm. SHapley Additive exPlanations (SHAP) analysis revealed that feed rate was the most influential process variable, followed by polymer concentration and the Flory–Huggins interaction parameter [7].

Figure 2 shows the lignin sphere of Hansen compared to proper (blue dots) and improper (red dots) solvents. The units for three axes are $MPa^{1/2}$.

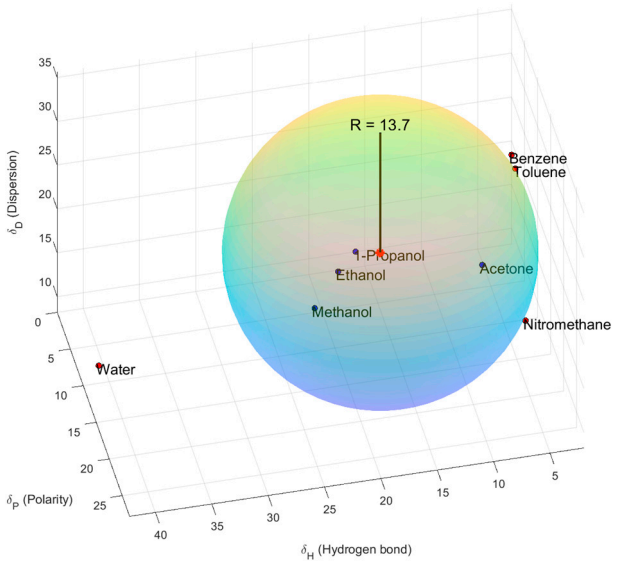


Figure 2. Three-dimensional sphere for lignin and various solvents.

Mohammad Golbabaei designed classification (qualitative) and prediction (quantitative) black-box models to determine the fiber diameter and electrical conductivity. For the quantitative diameter prediction, the best performing model was an Artificial Neural Network (ANN), which reached an amazingly high R^2 value of 0.8. The dataset used consisted of chitosan, polyvinyl alcohol (PVA), polycaprolactone (PCL), and gelatin, among other elements [8].

While there has been work that has successfully designed models to determine nanofiber diameter, there is yet to be a white-box model that explains a direct relationship, $f(X_1,X_2,\ldots) = Y$, between solution and operation parameters and the nanofiber diameter of electrospun fibers.

Ordinary Least Squares (OLS) classic linear regression models assume that data follow a Gaussian distribution when trained, where the mean, median, and mode coincide and observations are symmetrically distributed (hence the non-dependence of mean on variance) [9]. However, many phenomena—such as electrospinning—are not likely to exhibit this behavior due to their high degrees of freedom. Generalized Linear Models (GLMs)

are powerful tools for analyzing non-Gaussian data, as they extend classical regression by incorporating distributions from the exponential family, expressed as

$$\mathrm{E}[Y] = g(\mu) = \eta = X\beta \tag{4}$$

where

$\mathrm{E}[Y]$ = expected value of Y;
$g(\mu)$ = link function;
$\eta = X\beta$ = linear predictor.

The link functions transform the linear predictor matrix–input to fit the response variable–output, where β is calculated via Maximum Likelihood Estimation (MLE). This approach identifies the coefficients that maximize the likelihood of observing the given data based on the assumed probability density function. As a result, the estimated coefficients describe the relationship between predictors and the response within the chosen link–distribution framework [10]. Table 1 summarizes the link functions associated with each exponential-family distribution. The canonical link, indicated in the table footnotes, represents the function that best fits each distribution; however, alternative link–distribution combinations commonly used in practice are also included [10].

Table 1. Link function relations for exponential probability distribution family.

Link Function	$g(\mu) = \eta = X\beta$	$g^{-1}(\eta) = \mu$	Response Variable Y	Commonly Used Distributions
Identity	μ	η	Continuous	Gaussian [1], Poisson, Gamma, Inverse Gaussian
Log	$\ln \mu$	e^{η}	Continuous	Gaussian, Poisson [2], Gamma, Inverse Gaussian
Inverse	μ^{-1}	η^{-1}	Continuous	Gaussian, Poisson, Gamma [3], Inverse Gaussian
Inverse squared	μ^{-2}	η^{-2}	Continuous	Inverse Gaussian [4]
Square root	$\mu^{-\frac{1}{2}}$	$\eta^{-\frac{1}{2}}$	Discrete	Poisson
Logit	$\ln\left(\frac{\mu}{1-\mu}\right)$	$\frac{e^{\eta}}{1+e^{\eta}}$	Discrete	Bernoulli [5], Binomial [6]
Probit	$\phi^{-1}(\mu)$	$\phi(\eta)$	Discrete	Bernoulli, Binomial
Log-Log	$-\ln(-\ln \mu)$	$\exp(-\exp(-\eta))$	Discrete	Bernoulli, Binomial
C-Log-Log	$\ln(-\ln(1-\mu))$	$1-\exp(-\exp(\eta))$	Discrete	Bernoulli, Binomial

[1] Identity is the canonical link for Gaussian distribution. Default framework used in OLS. [2] Log is the canonical link for Poisson distribution. [3] Inverse is the canonical link for Gamma distribution. [4] Inverse squared is the canonical link for Inverse Gaussian distribution. [5,6] Logit is the canonical link for Bernoulli and Binomial distributions.

2. Materials and Methodology

2.1. Data Collection

A meta-analysis of electrospun nanofibers was conducted, searching for common synthetic and non-synthetic polymers used in electrospinning. The polymer Hansen parameters of solvent and polymer were used to calculate the RED of the system and the Flory–Huggins parameter, following Shrutidhara's process [7]:

$$\chi = \frac{M_v Ra^2}{16.736RT} \tag{5}$$

where

χ = Flory–Huggins parameter;
M_v = molar volume ($\mathrm{cm}^3/\mathrm{mol}$);

Ra = three-dimensional distance between solvent–polymer coordinates ($MPa^{1/2}$);
R = 1.987 cal/mol K;
T = temperature (K).

Figure 3 shows the total records of the polymers, while Table 2 shows the range of every *X* in the database.

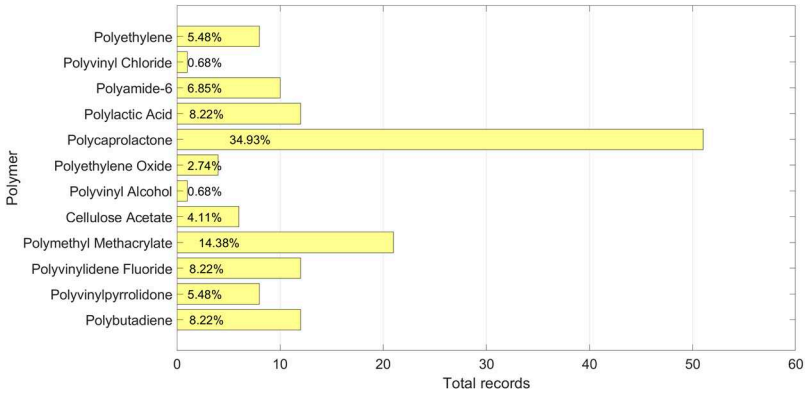


Figure 3. Number of records and percentages by polymer in the training database.

Table 2. Value ranges across *Y*'s and *X*'s in database.

Variable	Parameter	Values Range
Response variable (*Y*)	Fiber Diameter (nm)	45.0000–7670.0000
X_1	RED	0.2416–1.2962
X_2	Flory–Huggins Interaction Parameter (F_H)	0.0074–0.7752
X_3	Jet speed-Velocity (cm/h)	6.0172–9417.4520 [1]
X_4	Temperature (K)	293.1000–393.1000
X_5	Syringe diameter (cm)	0.0260–0.4600
X_6	Feed Rate (cm^3/h)	0.1000–12.000
X_7	Distance (cm)	6.0000–20.0000
X_8	Voltage (kV)	5.5000–45.0000
X_9	Polymer Concentration (wt%)	0.1000–22.0000
X_{10}	Type of electrospinning	0, 1, 2 [2]

[1] Jet speed was calculated by continuity fluid law using Syringe Diameter (X_5) and Feed Rate (X_6). [2] Categorical *X*: 1 and 2 are inner and outer polymers in coaxial electrospinning, respectively, while 0 means non-coaxial electrospinning.

2.2. Materials, Resources, and Process

Modeling was conducted in RStudio software v.2025.05.1+513 (Allaire, NJ, USA) [11]. Table 3 displays the main libraries used for data analysis and modeling.

Table 3. Libraries' functions used throughout data processing.

Library	Function/Syntaxis	Usage
stats	glm(formula, data, family("link function")) [1]	Model training
car	residualPlots(model)	Check linearity across *X*'s
glmtoolbox	BoxTidwell(model, transf~variable)	Fix linearity across *X*'s [2]
DHARMa [3]	plot(simulateResiduals(model, seed))	Check normality of GLM
MASS	stepAIC(model, direction = c("both"), k) [4]	AIC/BIC variable reduction [5]

[1] Default value for "formula" is "*Y* ~ ." which represents *Y*'s dependence on all X's found in Table 1, while "family" can be either "gaussian", "Gamma", or "inverse.gaussian" for this work. [2] Function output is better λ value to which X^{λ} more effectively fixes residual linearity in the model. [3] Diagnostics for Hierarchical Regression Models. [4] k-Parameter is a natural logarithm of database sample size (ln *n*) to simulate BIC variable selection. [5] Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC).

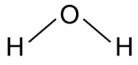
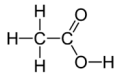
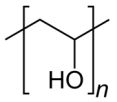
The electrospinning parameters and synthesis materials for experimental verification are presented in Tables 4 and 5, respectively. Some variables are not shown since they were fixed to constant values to analyze the voltage variation on fiber diameter, such as syringe diameter—0.072 cm; velocity—491.219 cm/h; and temperature—297.150 K.

Table 4. Electrospinning parameters of PVA nanofibers.

Sample [1]	PVA Concentration (wt%)	HAc [2] Concentration (wt%)	Feed Rate (mL/h)	Voltage (kV)	Distance (cm)
A	10.00	5.00	2.00	30.00	18.00
B	10.00	5.00	2.00	25.00	18.00
C	10.00	5.00	2.00	20.00	18.00

[1] All experimental samples were non-coaxial electrospinning (Type = 0). [2] Acetic acid (HAc).

Table 5. Materials used for PVA polymeric solution synthesis.

Material	Description	Chemical Formula
Deionized water	Barnstead MicroPure ST Thermo Scientific-processed (Waltham, Massachusetts, USA)	
Acetic Acid (HAc)	Reagent Grade (R.G.), Merck (Darmstadt, Germany)	
Polyvinyl Alcohol	R.G., Merck (Darmstadt, Germany) Molecular Weight (M.W.) 89,000–98,000 99+% hydrolyzed	

2.3. Inference Modeling

Twenty GLMs were trained using Gaussian, Gamma, and Inverse Gaussian families to evaluate the goodness-of-fit across different continuous response frameworks. The response variable was modeled as the natural logarithm of the fiber diameter (ln Y) to improve normality and variance stability; this assumption was later verified through the Box–Cox transformation test.

Cook's distance, leverage, and box plots were used to identify influential records in the Y and X variables. Model assumptions were assessed through QQ-plots (Quantile-Quantile-plots), residual histograms, and DHARMa residual analyses to verify normality and homoscedasticity.

Box–Tidwell transformation tests were performed for each parameter to examine predictor linearity and determine whether the linearity in the model could be fixed or not by transforming X.

Variable selection was guided by the Akaike (AIC) and Bayesian (BIC) Information Criteria using forward, backward, and bidirectional procedures, with assumptions verified at each step.

2.4. Prediction Modeling

The best inference framework was applied to train models, including main, quadratic, and interaction effects. Data were split following an 80:20 ratio using the repeated hold-out sampling method, and AIC/BIC-based reduction was performed to obtain the most parsimonious and best-fitting models.

Model performance was evaluated through the Mean Squared Error (MSE) and its root (RMSE), Mean Absolute Error (MAE), and its Percentage (MAPE). The squared correlation coefficient ($Corr^2$) between predicted and observed responses was also calculated to assess predictive accuracy.

2.5. Experimental Testing

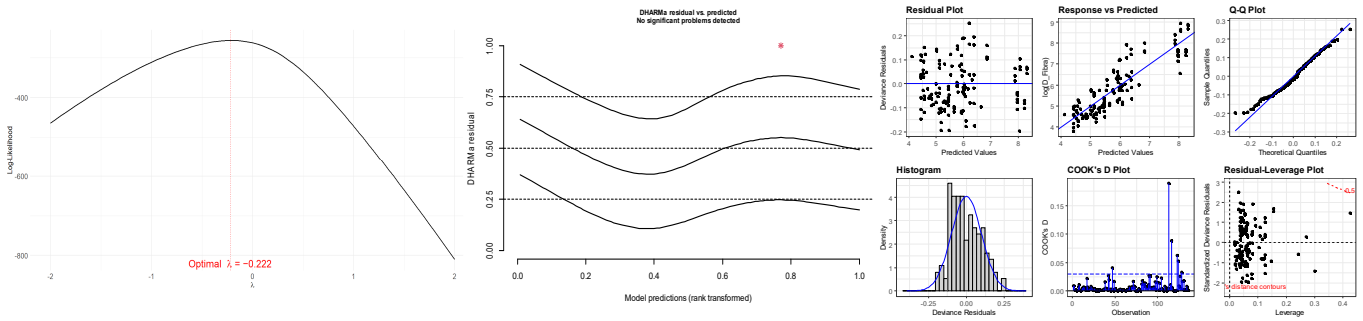
The samples were prepared by dissolving 1 g of PVA in 8.5 g of deionized water previously acidified with 0.5 g of HAc and heated at 60 °C for 5 min. This mixture was then mechanically stirred for 15 h until complete dissolution. Electrospinning was conducted on the same day to ensure solution stability.

The morphology of the electrospun PVA fibers was examined using Scanning Electron Microscopy (SEM) (JSM-7800F, JEOL, Tokyo, Japan). Prior to imaging, the samples were coated with a thin gold layer (~10 nm) by sputtering at 20 mA for approximately 60 s. Micrographs were obtained at an accelerating voltage of 10 kV, with magnifications between 1000× and 100,000×.

3. Results and Discussion

3.1. Inference Modeling

Figure 4a shows the best transformation for the Box–Cox test, which determined that the best transformation for Y was $\lambda = -0.222$ (p-value = 2.2×16^{-16}). Logarithmic transformation ln(Y) was performed, giving λ closeness to 0 and to simplify interpretation, as suggested by Remi Sakia [12].



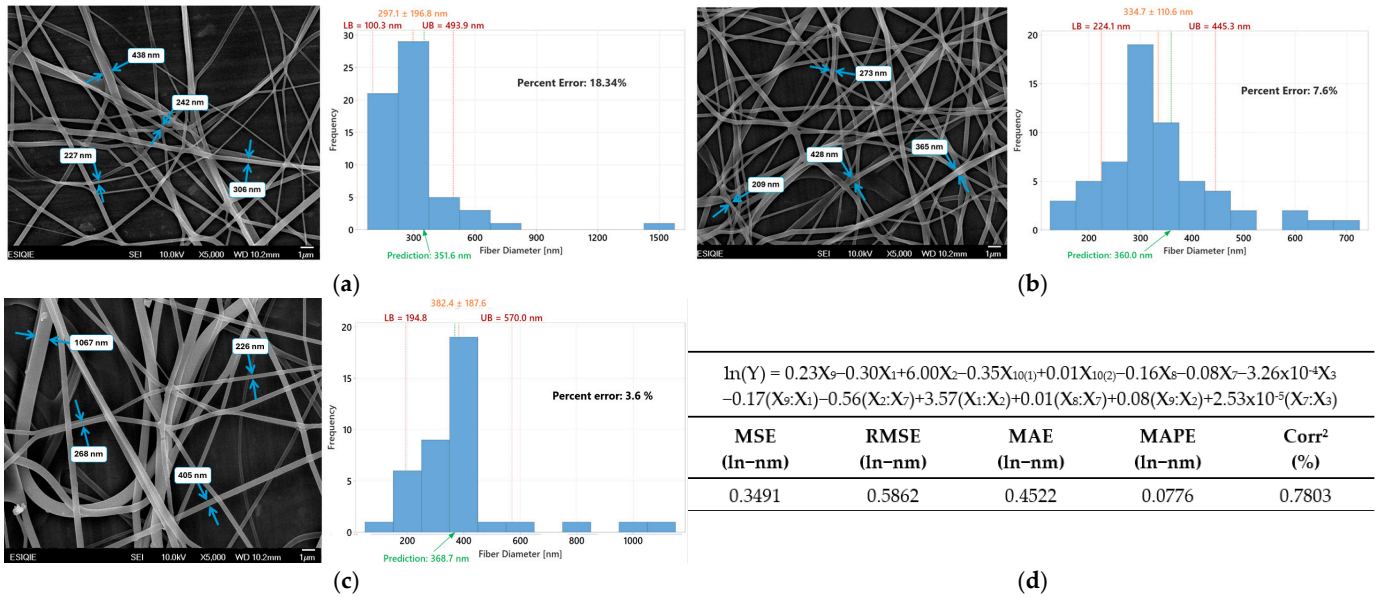
(**a**) (**b**) (**c**)

Figure 4. (**a**) Log-likelihood maximization for λ-transformation values, (**b**) DHARMa residuals, and (**c**) residual panel for Gamma(μ) model: response ln(Y).

The best goodness-of-fit framework was identified as the Gamma distribution with an identity link using the ln(Y)-transformed response. Figure 4b presents the DHARMa and Figure 4c shows the leverage and Cook's distance plots, indicating no influential observations, except for record 113. This observation corresponds to the highest jet speed (X_3) in the dataset (noted as a red asterisk in the DHARMa plot); however, it remains within acceptable limits for both Cook's distance and leverage.

3.2. Prediction Modeling and Experimental Verification

The best performing model, obtained by AIC forward reduction in Gamma(μ), and the transformed response ln(Y) of main effects + interactions was used to predict nanofiber diameter on Samples A, B, and C, shown in Table 4. Figure 5 displays SEM nanofibers as well as the distribution of their measured dimensions compared to their predicted dimensions. In orange the mean plus minus the deviation is presented ($\mu \pm \sigma$), while red dotted lines display lower (LB) and upper (UB) bounds of it.



$$\ln(Y) = 0.23X_9 - 0.30X_1 + 6.00X_2 - 0.35X_{10(1)} + 0.01X_{10(2)} - 0.16X_8 - 0.08X_7 - 3.26\times10^{-4}X_3 - 0.17(X_9{:}X_1) - 0.56(X_2{:}X_7) + 3.57(X_1{:}X_2) + 0.01(X_8{:}X_7) + 0.08(X_9{:}X_2) + 2.53\times10^{-5}(X_7{:}X_3)$$

(d)

MSE (ln−nm)	RMSE (ln−nm)	MAE (ln−nm)	MAPE (ln−nm)	Corr² (%)
0.3491	0.5862	0.4522	0.0776	0.7803

Figure 5. Real diameter represented by SEM analysis and fiber measured histogram vs. predicted for (**a**) sample A, (**b**) sample B, (**c**) sample C, and in (**d**) the simplified formula, as well as the metrics of the prediction model.

4. Conclusions

The developed GLM-based Gamma(μ) successfully predicted electrospun log-nanofiber diameters from solubility and process parameters, achieving strong correlation and low error. These results demonstrate the potential of interpretable statistical models to guide polymer–solvent selection and process optimization in electrospinning.

Author Contributions: R.C.-B. and M.A.P.-C. synthesized copolymers, prepared polymeric composite solutions, electrospun yarns, and collected and interpreted data. M.A.P.-C. and M.L.-G. collected and analyzed data for modeling. M.C. and G.M.-M. performed experiments, collected and interpreted data. The manuscript was written by M.A.P.-C. and revised by R.C.-B. and L.R.-V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Instituto Politécnico Nacional (SIP projects 20242529 and 20240655).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors would like to acknowledge Centro de Nanociencias y Micro-Nanotecnologías (CNMN) from Instituto Politécnico Nacional for the SEM assays.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GLM	Generalized Linear Models
AIC	Akaike Information Criteria
BIC	Bayesian Information Criteria
MSE	Mean Squared Error
RMSE	Root Mean Squared Error

MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
RED	Relative Energy Difference
PVDF	Polyvinylidene Fluoride
GBR	Gradient Boosting Regressor
SHAP	SHapley Additive exPlanation
ANN	Artificial Neural Network
PVA	Polyvinyl alcohol
PCL	Polycaprolactone
MLE	Maximum Likelihood Estimation
OLS	Ordinary Least Squares

References

1. Bhardwaj, N.; Kundu, S.C. Electrospinning: A fascinating fiber fabrication technique. *Biotechnol. Adv.* **2010**, *28*, 325–347. [CrossRef] [PubMed]
2. Francis, L.F. *Materials Processing: A Unified Approach to Processing of Metals, Ceramics and Polymers*, 1st ed.; Elsevier: Amsterdam, The Netherlands, 2016; Volume 6.
3. Hansen, C.M. *Hansen Solubility Parameters: A User's Handbook*, 1st ed.; CRC Press LLC: New York, NY, USA, 2000.
4. Hildebrand, J.H.; Scott, R.L. *The Solubility of Nonelectrolytes*, 3rd ed.; Science: New York, NY, USA, 1950; p. 488.
5. Loyola-González, O. Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view. *IEEE Access* **2019**, *7*, 154096–154113. [CrossRef]
6. Starbuck, C. Linear Regression. In *The Fundamentals of People Analytics*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 181–206.
7. Sarma, S.; Verma, A.K.; Phadkule, S.S.; Saharia, M. Towards an interpretable machine learning model for electrospun polyvinylidene fluoride (PVDF) fiber properties. *Comput. Mater. Sci.* **2022**, *213*, 111661. [CrossRef]
8. Golbabaei, M.H.; Varnoosfaderani, M.S.; Hemmati, F.; Barati, M.R.; Pishbin, F.; Ebrahimi, S.A.S. Machine learning-guided morphological property prediction of 2D electrospun scaffolds: The effect of polymer chemical composition and processing parameters. *RSC Adv.* **2024**, *14*, 15178–15199. [CrossRef] [PubMed]
9. Socuéllamos, J.M. *Modelos Lineales Generalizados*, 1st ed.; Universidad Miguel Hernández: Elche, Spain, 2001; p. 247.
10. Nelder, J.; Wedderburn, W. Generalized Linear Models. *J. R. Stat. Soc.* **1972**, *135*, 370–384. [CrossRef]
11. Allaire, J.J. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2025.
12. Sakia, R. The Box-Cox Transformation Technique: A Review. *J. R. Stat. Soc.* **1992**, *41*, 169–178. [CrossRef]