INEMATD SCIENCE

Ingeniería, Energía, Materiales, Tecnología y Desarrollo

ISSN XXXX-XXXX

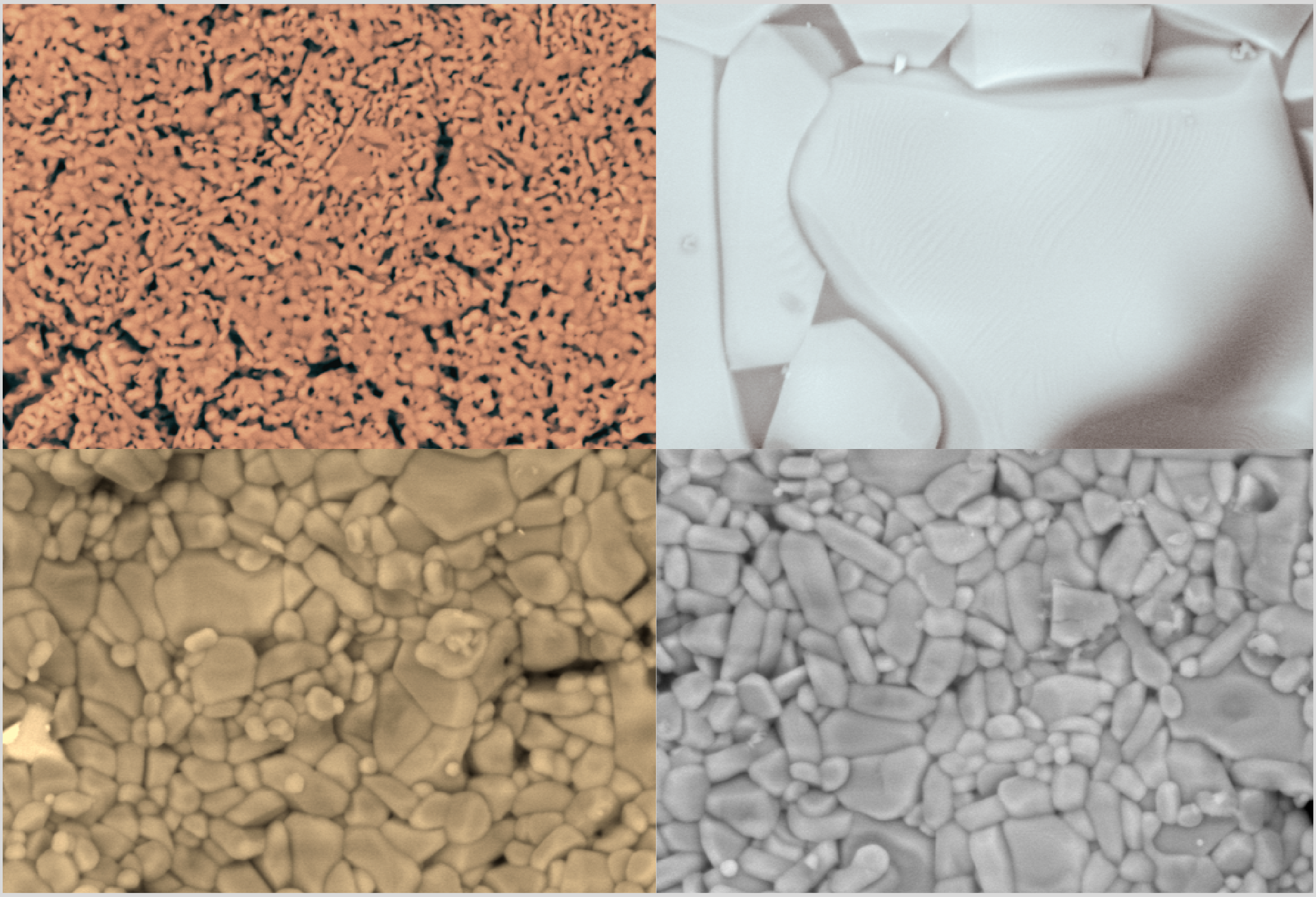


Foto de portada por Cristian Rodríguez-Gómez

VOLUMEN 2, NÚMERO 1 | 2026
PUBLICACIÓN SEMESTRAL

INEMATD SCIENCE:
Ingeniería, Energía, Materiales, Tecnología y Desarrollo

INEMATD SCIENCE: Ingeniería, Energía, Materiales, Tecnología y Desarrollo Número 1, enero de 2026 es una publicación semestral editada y distribuida por Cristian Gómez Rodríguez, Calle Monte Aconcagua 117, C.P. 66612, Apodaca, Nuevo León Teléfono: +528116235104. Editor responsable: Cristian Gómez Rodríguez. Reservas de Derecho al Uso Exclusivo No.04-2025-112610233900-102, ISSN: (en trámite), ambos otorgados por el Instituto Nacional de Derecho de Autor, Licitud de Título. INEMATD SCIENCE: Ingeniería, Energía, Materiales, Tecnología y Desarrollo es una publicación 100% digital. El editor responsable: Cristian Gómez Rodríguez.

DIRECTORIO

CONTENIDO

PERFORMANCE ESTIMATION IN PHOTOVOLTAIC-THERMAL SYSTEMS USING POLYNOMIAL REGRESSION

D. Domínguez-Bautista [1], L. V. García–Quiñonez [2] and C. Gómez-Rodríguez [3,*]

[1] Maestría en Ciencias en Tecnología Energética, Centro de Investigación en Recursos Energéticos y Sustentables, Campus Coatzacoalcos, Av. Universidad km. 7.5, Santa Isabel, 96538, Coatzacoalcos, Veracruz, México; zS24019638@estudiantes.uv.mx
[2] Centro de Investigación en Recursos Energéticos y Sustentables, Campus Coatzacoalcos, Av. Universidad km. 7.5, Santa Isabel, 96538, Coatzacoalcos, Veracruz, México; lingarcia@uv.mx
[3] Departamento de Mecánica, Facultad de Ingeniería, Campus Coatzacoalcos, Universidad Veracruzana, Av. Universidad Km 7.5 Col. Santa Isabel, Coatzacoalcos, 6535,Veracruz, México; crisgomez@uv.mx
* Correspondence: crisgomez@uv.mx.

Abstract

Hybrid solar panels, or photovoltaic-thermal (PVT) systems, simultaneously convert solar radiation into electricity and heat within the same unit. This study analyses a dataset obtained from a hybrid solar panel with an experimental setup. A polynomial regression model based on three variables: solar irradiance, fluid flow rate, and thermal output energy. Analysis produced a polynomial model correlating these variables with thermal efficiency (ηt), achieving a determination coefficient of $R^2 = 0.8870$. Contour and surface plots of the performance coefficient as functions of input variables are provided, along with a residual analysis to assess the model's accuracy.

Keywords: Hybrid solar panel; Photovoltaic-thermal system; Residual analysis.

1. Introduction

Hybrid solar panels, or photovoltaic-thermal (PVT) systems, this dual-use design improves energy utilization by capturing and managing heat generated by photovoltaic cells, improving their efficiency and regulating operating temperatures [1].

A PVT system integrates a photovoltaic module with a thermal collector, offering higher energy generation density compared to separate systems. This unified approach contributes to the development of sustainable solutions for both electricity and heat production [2]. Prior studies have shown that transparent covers minimize thermal losses and energy dissipation, thereby improving system performance [3].

Since the early 2000s, global collaboration on PVT technologies has intensified. However, deployment challenges persist, requiring further research. Accurate modelling of input-output relationships is crucial for optimizing performance. Researchers have employed experimental mathematical models, regression techniques, artificial neural networks, and time-series-based statistical methods to predict solar system performance [4][5].

Among regression techniques, polynomial fitting offers a practical and accurate method for modelling nonlinear relationships. Escobedo et al. [6] compared artificial neural networks (ANNs) with polynomial models, concluding that although ANNs may outperform in accuracy, polynomial models are advantageous due to their simplicity and ease of derivation.

This study proposes an efficient methodology to model PVT system performance using polynomial regression, reducing reliance on complex equations while achieving accurate coefficient estimation.

2. Materials and Methods

Figure 1 presents a schematic diagram of the experimental setup installed on a hybrid solar panel. System consists of a variable-speed circulator, Rilsan tubing, a thermostatic water bath, and a flow sensor used to measure the flow rate, as described by Matteo et al. [7].

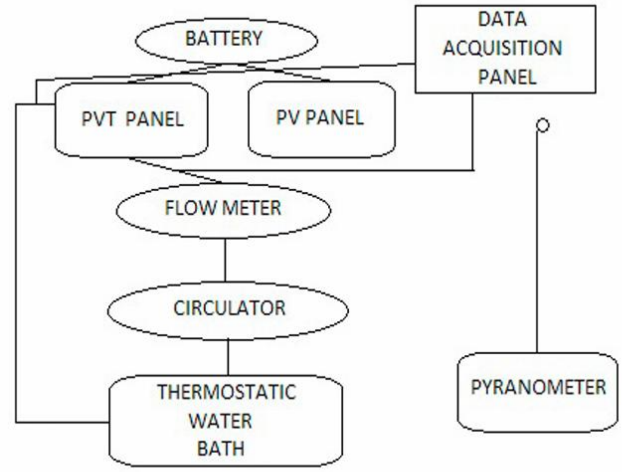


Figure 1. The experimental setup. Matteo et al [7].

After exiting the aluminum block, working fluid is cooled in a thermostatic water bath before being recirculated by a pump to the cold inlet side, maintaining a constant temperature. To measure thermal and flow parameters, type K thermocouples were installed and connected to a DAP interface, which also records the frequency signal from a flow meter to calculate the system's mass flow rate. Additionally, a pyranometer was used to measure the solar radiation flux density (W/m²) in the experimental environment. Data acquisition and processing were optimized using an integrated LabVIEW interface, enabling efficient management of the experimental data [7]. Table 1 presents the operating range of the measured pyranometers.

Table 1. Experimental operating conditions range of PVT system.

Variable Name	Operating Range	Units	Label
Irradiance	897 – 1087	W/m²	V1
Fluid flow rate	1.9 – 23	L/min	V2
Thermal energy	63.9 – 100.9	W	V3

A dataset was generated based on the specific operating conditions of the system, considering various parameters including fluid flow velocity, inlet and outlet sink temperatures, solar irradiance, ambient temperature and cell's open-circuit voltage.

The instantaneous thermal efficiency of the PVT system (ηt) was determined as the ratio of the thermal power generated to the incident solar power, as expressed in Equation (1):

$$\eta_t = Q_t / Q_i, \quad (1)$$

The input power (Q_i) is calculated as the product of the solar irradiance (G) and the collector's aperture area (A), as shown in Equation (2):

$$Q_i = G.A, \tag{2}$$

Thermal power output (Qt) corresponds to the amount of heat extracted by working fluid. It is proportional to difference between outlet and inlet fluid temperatures (ΔT), mass flow rate ($\dot{m}$), and specific heat capacity of the fluid (Cp), as described in Equation (3):

$$Q_t = m.Cp.\Delta T, \tag{3}$$

Methodology

This section presents a methodology for developing a polynomial model that defines relationship between performance coefficient and system's operational variables. This approach enables accurate system characterization and enhances predictive analysis. Construction of polynomial model is based on previously established input variables. To generally represent relationship between these variables and performance coefficient, Equation (4) is introduced.

$$n_t = p\,(V1, V2, V3) + \epsilon, \tag{4}$$

Let p be an indeterminate polynomial function and ϵ a random error term. It is assumed that this error term follows a standard normal distribution.
Figure 2 shows correlation matrix between input variables, which is used to assess the degree of correlation among independent variables.

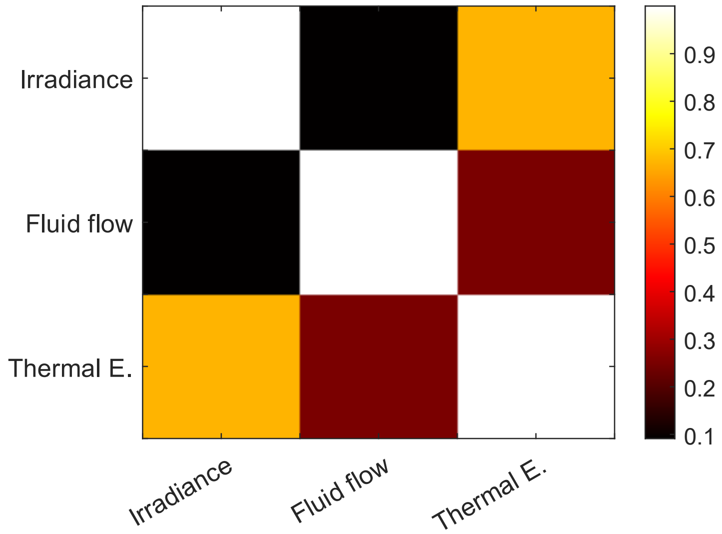


Figure 2. Input variables correlation matrix.

3. Results

With purpose of determining a polynomial model that offers best balance between model fit and parsimony, multiple linear regression is used to analyze the relationship between variables (V1, V2, V3) and the experimental performance coefficient (ηt).

Table 2, presents performance coefficients corresponding to a linear combination of monomial terms included in model, excluding only constant term, which is equal to 50.1626.

Table 2. Performance coefficient of each term in polynomial model.

Factor	Performance coefficient
$V1^2$	0.3715
V2	- 46.4592
$V2^2$	- 227.5152
$V2^3$	124.9192
$V3^2$	- 0.0032
(V1)(V2)	55.2697
$(V1)(V2^2)$	242.5250
(V2)(V3)	- 121.0381
(V2)(V3)	17.4685
$(V2^2)(V3)$	- 577.8241
$(V1)(V2^2)(V3)$	65.4707
$(V1)(V2)(V3^2)$	- 4.3044

In pursuit of a more streamlined polynomial model—with fewest variables and highest possible coefficient of determination (R^2)— Several term combinations were evaluated. Finally, Equation (5) was identified as optimal model, achieving an R^2 value of 0.8870.

$$\begin{aligned} n_t = {} & 50.1626 + 0.3715(V1^2) - 46.4592(V2) - 227.5152(V2^2) + 124.9192(V2^3) \ldots \\ & - 0.0032(V3^2) + 55.2697(V1)(V2) + 242.5250(V1)(V2^2) - 121.0381(V2)(V3) \ldots \\ & + 17.4685(V2)(V3^2) - 577.8241(V2^2)(V3) + 65.4707(V1)(V2^2)(V3) - \\ & 4.3044(V1*V2)(V3^2) \end{aligned} \tag{5}$$

Polynomial in Equation (5) stands out for having the highest R^2 compared to other combinations and, furthermore, for its compact structure with a smaller number of related variables. As R^2 increases, complexity of the model also tends to increase

In Figure 3, it can be observed that relationship is approximately linear between experimental performance coefficients and predicted values, to demonstrate validity of this polynomial regression model.

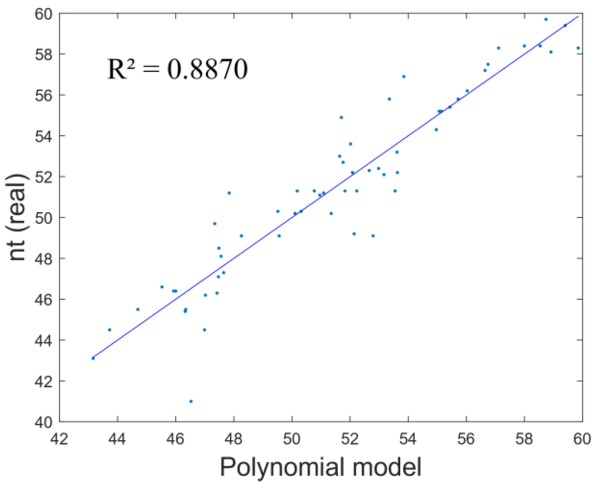


Figure 3. Experimental performance coefficient versus predicted values.

Residual Analysis

Given that variables in polynomial Equation (5) include irradiance, flow rate, and thermal energy, a graphical representation of the polynomial predictor's behavior is presented. In this plot, experimentally obtained performance coefficients are compared with the predicted values (Figure 3).

Since polynomial regression model incorporates three distinct variables, contour plots are generated in Figure 4 and surface plots in Figure 5, illustrate correlation between predicted values versus experimental data points.

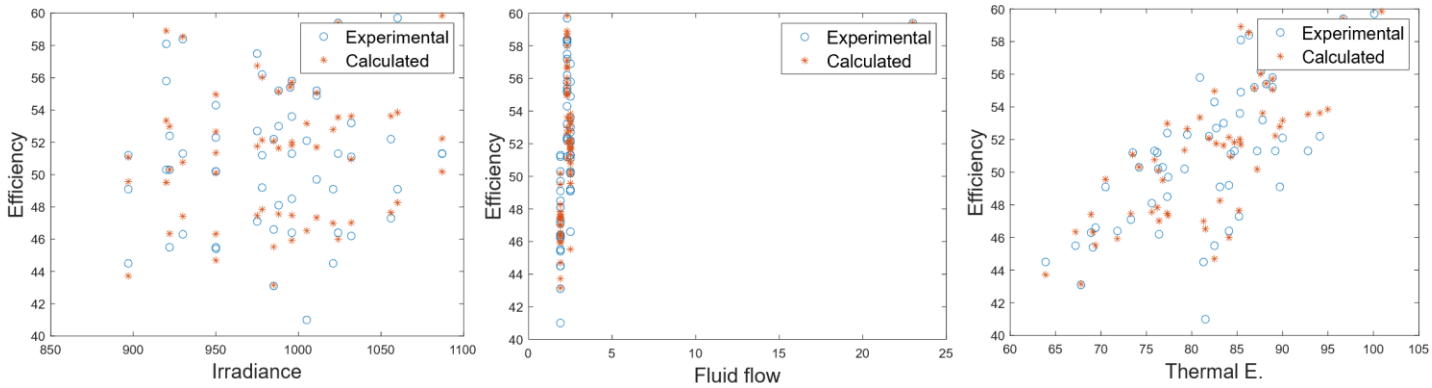


Figure 4. Contour plots. Comparison of experimental performance coefficient and predicted values for each significant variable in polynomial model.

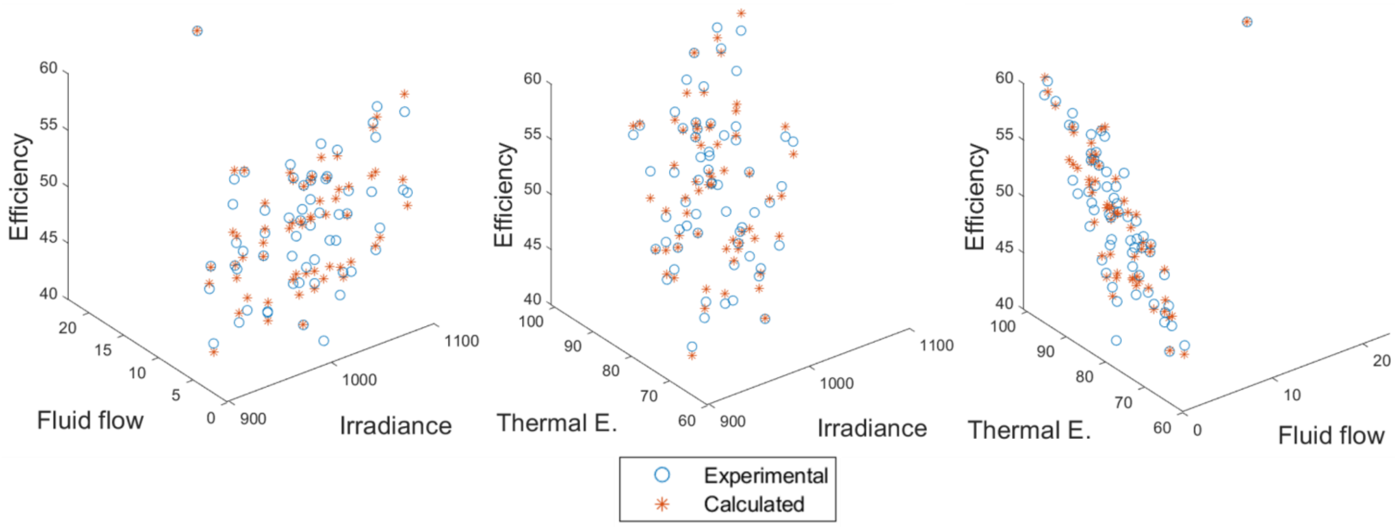


Figure 5. Surface plots. Comparison of experimental performance coefficient and predicted values for each pair of significant variables in polynomial model.

Figure 6 presents residuals histogram, which exhibits a bell-shaped distribution. On the right end of the diagram, slight deviations indicate an imperfect symmetry. This characteristic is further examined in Figure 7, where the relationship between the predicted performance coefficient and the standardized residuals is shown. Two horizontal reference lines at values of 3 and -3 are included. Outside these boundaries, five outliers are identified, which may influence polynomial model.

Improve the accuracy of the fit, these outliers were excluded from the dataset, resulting in an improved coefficient of determination of $R^2 = 0.9466$.

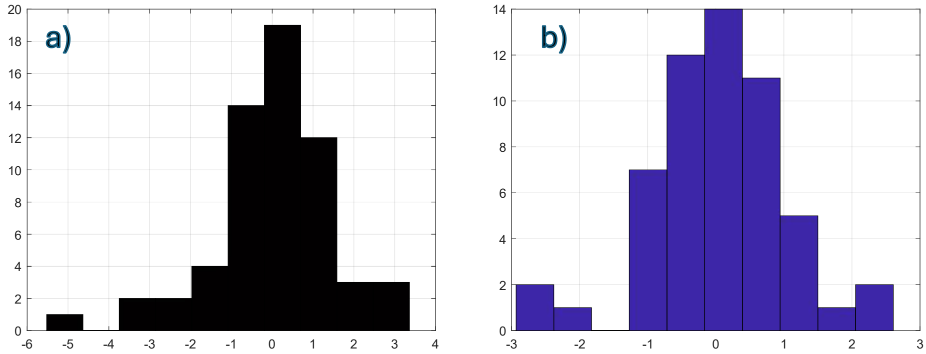


Figure 6. Histogram of standardized residuals a) with outliers, b) without outliers.

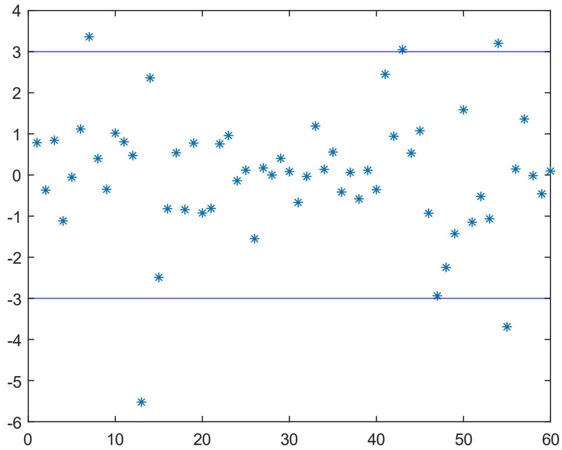


Figure 7. Standardized residuals.

Finally, to assess independence of residuals, Figures 7 and 8 are presented. Figure 8 specifically illustrates confidence intervals for each residual (green lines), with central point’s indicating residual values and red lines identifying outliers. These visualizations confirm that presence of a substantial number of outliers negatively impacts model's stability, resulting in a decreased correlation coefficient.

Outliers introduce significant variability into datasets, reducing the model's ability to accurately capture relationship between variables. Notable difference compared to initial R^2 value of 0.8870 demonstrates that these anomalous data points indeed interfere with the model's performance. Furthermore, standardized residuals deviate from a Gaussian distribution cantered at zero and exhibit non-constant variance, violating key assumptions of polynomial regression.

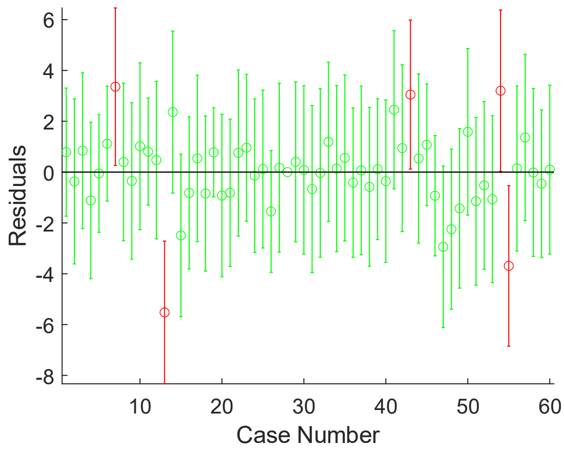


Figure 8. Predicted performance coefficient versus standardized residuals.

Primary purpose of ANOVA analysis is to evaluate experimental errors and perform a significance test to determine correlation between factors versus experimental performance coefficient (η_t). Table 3 presents correlation degree of factors based on variables V1, V2, and V3, including their three-way interaction, according to analysis of variance.

F-values obtained for each factor are remarkably high (F = 7433.32 for individual variables and F = 6797.58 for the interaction), accompanied by extremely low p-values (all $\ll 0.05$), clearly indicating that the contributions of V1, V2, and V3, both individually and jointly are statistically significant.

Table 3. ANOVA analysis of independent variables.

Factor	SS	DF	MS	F	Prob > F
V1	79480	1	79479.97	7433.32	2.0358 e-108
V2	79480	1	79479.97	7433.32	2.0358 e-108
V3	79480	1	79479.97	7433.32	2.0358 e-108
(V1)(V2)(V3)	1379.7	3	39738.99	6797.58	9.61865 e-229
Error	1261.7	118	10.65		

The polynomial regression model yields:

$$F=30.74 \quad \text{and} \quad p<0.001$$

These values indicate a strong statistical significance. Since:

$$p<\alpha=0.05$$

The null hypothesis H_0: $\beta 1 = \beta 2 \cdots \beta k = 0$ is rejected, confirming that the model terms, taken together, significantly explain the variance in the dependent variable (ηt). Results confirm model validity and usefulness for predictive applications.

Table 4 presents a comparative summary of modelling approaches used to estimate efficiency in photovoltaic-thermal (PVT) systems, focusing on predictive accuracy based on R^2. Although complex models tend to yield higher R^2 values, they aren't always the most practical choice. In many cases, a simple polynomial model offers advantages, lower complexity, easier interpretation, and reduced overfitting risk, especially when working with limited or extrapolated data. Slightly lower R^2 may be acceptable if it improves robustness and generalization, making model applicable under real-world conditions where data often deviate from ideal patterns.

Table 4. Evaluation of Accuracy and Deviation in Efficiency Prediction Models for PVT Systems with Extrapolated Data.

Author	Method	R^2
Payman et al. [8]	Evolutionary Polynomial Regression	0.8836
García [9]	Linear	0.8865
	Polynomial	0.8686
	Persistent	0.8114
Hossein et al. [10]	Random Forest	0.9862
	Multilayer Perceptron	0.8775
	Support Vector Regression	0.7639
Bautista et al.	Polynomial	0.8870

5. Conclusions

A polynomial model was successfully developed using three operational variables to predict performance coefficient of a photovoltaic-thermal (PVT) system. Results show that increasing coefficient of determination (R^2) leads to greater polynomial complexity. Model achieved an $R^2 = 0.8870$, outperforming results reported by Payman et al. [8], García [9], and Hossein et al. [10], whose approaches involved higher error terms. Despite its simplicity, proposed method proves advantageous by delivering lower prediction errors.

Findings indicate that, even with a simplified structure, proposed model surpasses conventional approaches in accuracy. In particular, the estimation of PVT system's independent variables shows high agreement with actual values, highlighting the model's reliability and effectiveness in characterizing system behavior.

References

1. Matteo Greppi, Giampietro Fabbri, Experimental characterization of a hybrid industrial solar tile, Energy Procedia, Volume 126, 2017, https://doi.org/10.1016/j.egypro.2017.08.243.
2. M. Arif Hasan, K. Sumathy, Photovoltaic thermal module concepts and their performance analysis: A review, Renewable and Sustainable Energy Reviews, Volume 14, Issue 7 2010, Pages 1845-1859, https://doi.org/10.1016/j.rser.2010.03.011.
3. Ahmad Fudholi, Kamaruzzaman Sopian, Mohammad H. Yazdi, Mohd Hafidz Ruslan, Adnan Ibrahim, Hussein A. Kazem, Performance analysis of photovoltaic thermal (PVT) water collectors, Energy Conversion and Management, Volume 78, 2014, https://doi.org/10.1016/j.enconman.2013.11.017.
4. Kamran Moradi, Fereshteh Jafari, Fariba Moghaddam, Qobad Shafiee, Dynamic modeling of photovoltaic–thermal systems using polynomial regression, Renewable Energy, Volume 250, 2025, https://doi.org/10.1016/j.renene.2025.123242.
5. Ali H.A. Al-Waeli, K. Sopian, Jabar H. Yousif, Hussein A. Kazem, John Boland, Miqdam T. Chaichan (2019). Artificial neural network modeling and analysis of photovoltaic/thermal system based on the experimental study, Energy Conversion and Management, Volume 186, Pages 368-379, https://doi.org/10.1016/j.enconman.2019.02.066.
6. Escobedo-Trujillo, B.A., Alaffita-Hernández, F.A., Colorado, D., & Siqueiros, J.. (2014). Coefficent of performance prediction by a polynomial model of absorption heat transformer. Revista mexicana de ingeniería química, 13(3), 907-917.
7. Matteo Greppi, Giampietro Fabbri, Experimental dataset of seasonal behaviour of a hybrid solar tile, Data in Brief, Volume 34, 2021, https://doi.org/10.1016/j.dib.2020.106649.
8. Salami, P., Ajabshirchi, Y., Abdollahpoor, S., & Behfar, H. The relation between different indices and temperature by Evolutionary Polynomial Regression for a PV/T system. 2016.
9. García, Juan Carlos Sosa. Predicción de producción de energía fotovoltaica a corto plazo. 2017.
10. Hossein Gharaee, Mohammad Erfanimatin, Ammar M. Bahman, Machine learning development to predict the electrical efficiency of photovoltaic-thermal (PVT) collector systems, Energy Conversion and Management, Volume 315, 2024. https://doi.org/10.1016/j.enconman.2024.118808.

ENERGY EFFICIENCY COEFFICIENT PREDICTION ON INDUCTION MOTORS BY A POLYNOMIAL REGRESSION MODEL

R. Martínez-Ronquillo [1*], L.V. García-Quiñonez [1], F.A. Alaffita-Hernández[1], C. Gómez-Rodríguez[2], M.I. Mendivil-Palma [3]

[1] Centro de Investigación en Recursos Energéticos y Sustentables, Universidad Veracruzana, Av. Universidad Km 7.5 Col. Santa Isabel, Coatzacoalcos Veracruz, C.P. 36538, México.
[2] Universidad Veracruzana Campus Coatzacoalcos Veracruz, Facultad de Ingeniería, Departamento de Mecánica, C.P. 96535, Coatzacoalcos, Veracruz, México.
3 Centro de Investigación en Materales Avanzados (CIMAV), Unidad Monterrey, Alianza Norte 202, Parque de Investigación Tecnológica (PIIT), Apodaca, Nuevo León 66600, México.
* Correspondence: zs24019640@estudiantes.uv.mx

Abstract

Several methods have been developed to estimate energy efficiency of induction motors, and the accuracy of these methods varies with load factor, voltage and harmonics. In this work, a polynomial model was developed to predict energy efficiency coefficient in a induction motor (De Lorenzo DL 1021) using electrical and mechanical parameters, through a dependent variable and independent variables, obtaining an R^2 value of 0.9926.

Keywords: Energy efficiency, Induction motor, Polynomial model, Polynomial, Balanced voltage

1. Introduction

Induction motors (IM) represent approximately 70% of energy consumed by industry. Currently, only motors with a greater power than 500 hp are generally monitored due to they cost. Nonetheless, motors less than 500 hp represent 97% of motors in operations and consume 71% of energy used. On average, these motors do not operate more than 60% of their rated load due to oversized installations of low-load conditions, resulting in poor efficiency and wasted energy [1].

Regulatory frameworks have been established in different countries to reduce energy consumption. These regulations aim to improve average efficiency of IM available on the market. Requiring companies to comply with efficiency standards [2]. Efficiency and load factor monitoring in real-time are essential to evaluate the energy efficiency of inductions motors. However, assessing these parameters requires highly intrusive and costly methods.

As measurements in-situ alternatives, several methods have been developed to estimate efficiency and load factor, known as energy efficiency estimation methods. Some of these methods include the slip method, current method, equivalent circuit method, torque method, among others [3], which are mainly influenced by load factor [4].

One of the most effective approaches to analyzing IM, is characterize the parameters of its equivalent circuit. These parameters allows for determinant of a mathematical model representing electrical and magnetic effects occurring within induction motor [5].

In this study, an empirical prediction model was developed and validated based on polynomial construction through multiple linear regression analysis to behavior predict of energy efficiency coefficient using electrical and mechanical parameters. Unlike existing studies that focus on achieving the highest prediction accuracy, our objective is construct a simple and easy-to-use model for industrial applications.

2. Experimental Data

The data used to adjust a model were obtained from experiment conducted on a test bench with a three-phase induction motor of 1.1 kW (De Lorenzo DL 1021), a brake control unit (De Lorenzo DL 1054TT) and a magnetic powder brake (De Lorenzo DL 1019P) [6]. Figure 1 shows configuration diagram of experimental test used for balanced sinusoidal voltage.

Database used to get a polynomial was obtained varying torque from 0.50 to 3 Nm in intervals of 0.25 Nm, while simultaneously measuring electrical and mechanical parameters shown in Table 1. To ensure the reliability and consistency of results, each test was repeated 10 times. A stable temperature was maintained during test; for this purpose, in each interval, measurements were taken after temperature had stabilized.

Mechanical output power, Equation 1, was calculated using torque and rotor speed measurements [7]:

$$P_{out} = \frac{T_{shaft} \cdot n_m}{9.549} \tag{1}$$

Efficiency, Equation 2, was calculated using input and output power values [8]:

$$\eta = \frac{P_{out}}{P_{in}} \cdot 100\% \tag{2}$$

Table 1. Electrical and mechanical parameters.

Electrical Parameters	Symbol	Mechanical Parameters	Symbol
Line voltage	V_{ab}, V_{bc}, V_{ca}	Output power	P_{out}
Voltage angles	$\theta_{ab}, \theta_{bc}, \theta_{ca}$	Shaft torque	T_{shaft}
Line current	I_a, I_c, I_c	Rotor speed	n_m
Current angles	ϕ_a, ϕ_b, ϕ_c	Load factor	L_c

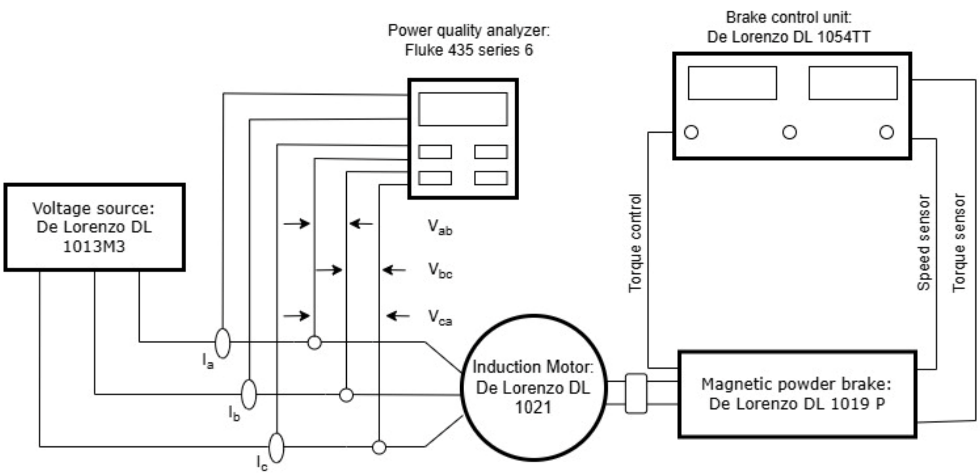


Figure 1. Schematic diagram of experimental test setup for balanced sinusoidal voltage.

3. Methodology

Generally, relationship between variables of interest, Energy Efficiency coefficient (EEC), and associated variables (inserter variables), are unknown but can be approximated using a polynomial model as follows [9]:

$$ECC = p(T_{shaft}, P_{in}, n_m, V_{ab}, V_{bc}, V_{ca}, \theta_{ab}, \theta_{bc}, \theta_{ca}, I_a, I_c, I_c, \phi_a, \phi_b, \phi_c, L_c) + \varepsilon \quad (3)$$

Where p is an unknown polynomial function and ε is random error. Is assumed that ε is a random variable with a standard normal distribution. Equation 3 defines the following set of predictor variables (Equation 4)

$$A := \{V_1, \dots, V_{12} \mid i = 1, \dots, 12\} \quad (4)$$

In Table 2, measured data assigned to a variable for polynomial construction can be observed. To obtain a simple polynomial, variable combinations used of first degree. Cardinality of set A is equal 12, due to are 12 terms in form V. Therefore, there are $2^{12} - 1$ subsets of A, excluding the empy set.

A linear regression was performed between subsets of A and EEC to find a simple polynomial that fits data. As mentioned in the previous paragraph, there are too many subsets of A; therefore, depend variables were discarded, resulting in variables shown in Table 3.

Thus, number of combinations required for test is: $2^7 - 1 = 127$.

Then, coefficient of determination (R^2) was calculated for each combination of variables and Energy Efficiency coefficient (EEC), and combination with highest coefficient (R^2) was selected.

Table 2. Measured variables

V_i	Variable	Unit
V_1	T_{shaft}	Nm
V_2	P_{in}	W
V_3	n_m	Rpm
V_4	V_{ab}	V
V_5	V_{bc}	V
V_6	V_{ca}	V
V_7	θ_{ab}	θ
V_8	θ_{bc}	θ
V_9	θ_{ca}	θ
V_{10}	I_a	A
V_{11}	I_b	A
V_{12}	I_c	A
V_{13}	ϕ_a	θ
V_{14}	ϕ_b	θ
V_{15}	ϕ_c	θ
V_{16}	L_c	p.u.

Table 3. Selected variables.

V_i	Variable	Unit
V_2	P_{in}	W
V_4	V_{ab}	V
V_6	V_{ca}	V
V_8	θ_{bc}	θ
V_{10}	I_a	A
V_{12}	I_c	A
V_{14}	ϕ_b	θ

4. Results

Considering a polynomial that fits experimental EEC, highest determination coefficient is sought. Due to simplicity of polynomial obtained, Equation 5 was chosen, with a determination coefficient R^2 of 0.9926, indicating a polynomial model accounts for 99.26% of variability in energy efficiency coefficient.

$$
\begin{aligned}
EEC = 229.9383 + 0.2726(V_2) + 3.1024(V_4) - 2.3638(V_6) \\
- 1.0961(V_8) - 36.3166(V_{10}) - 14.8443(V_{12}) \\
- 54.1092(V_{14})
\end{aligned} \tag{5}
$$

In Figure 2, a positive relationship between predicted EEC and experimental EEc is shown, suggesting, in general as predicted efficiency coefficient increases, actual also tends to increase. This positive correlation indicates that model can capture general trend of EEC's behavior.

Considering variables involved in polynomial, graphs were generated to improve a geometric perspective on behavior of predicted polynomial, experimental data and efficiency coefficients.

To obtain a geometric representation of predicted polynomial's behavior, given that our polynomial consists of seven variables, level curves and surface plots were generated for each of them, as shown in Figures 3 and 4, respectively. These figures illustrate the behavior of the variables in relation to our polynomial versus the experimental data.

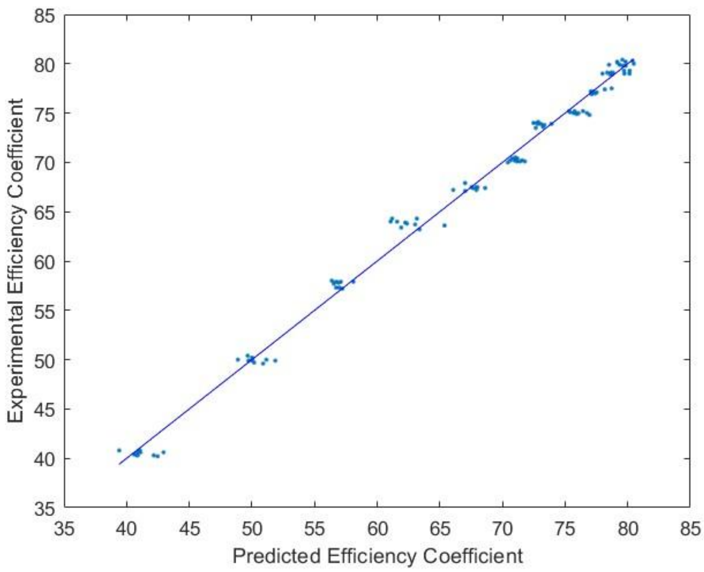


Figure 2. Experimental vs Predicted Efficiency Coefficient for the database.

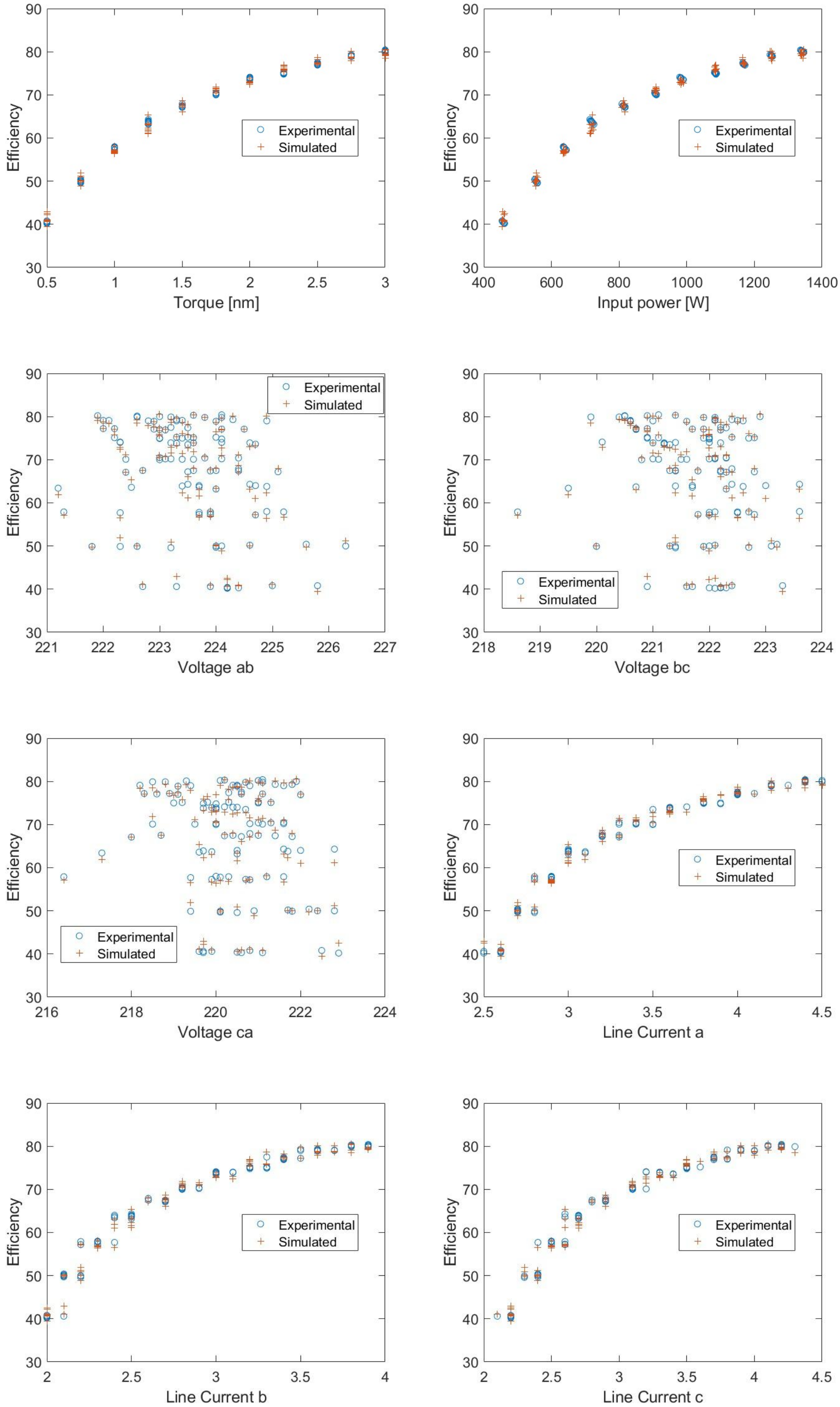


Figure 3. Level curves comparing experimental and predicted efficiency coefficients for each significant variable in polynomial model.

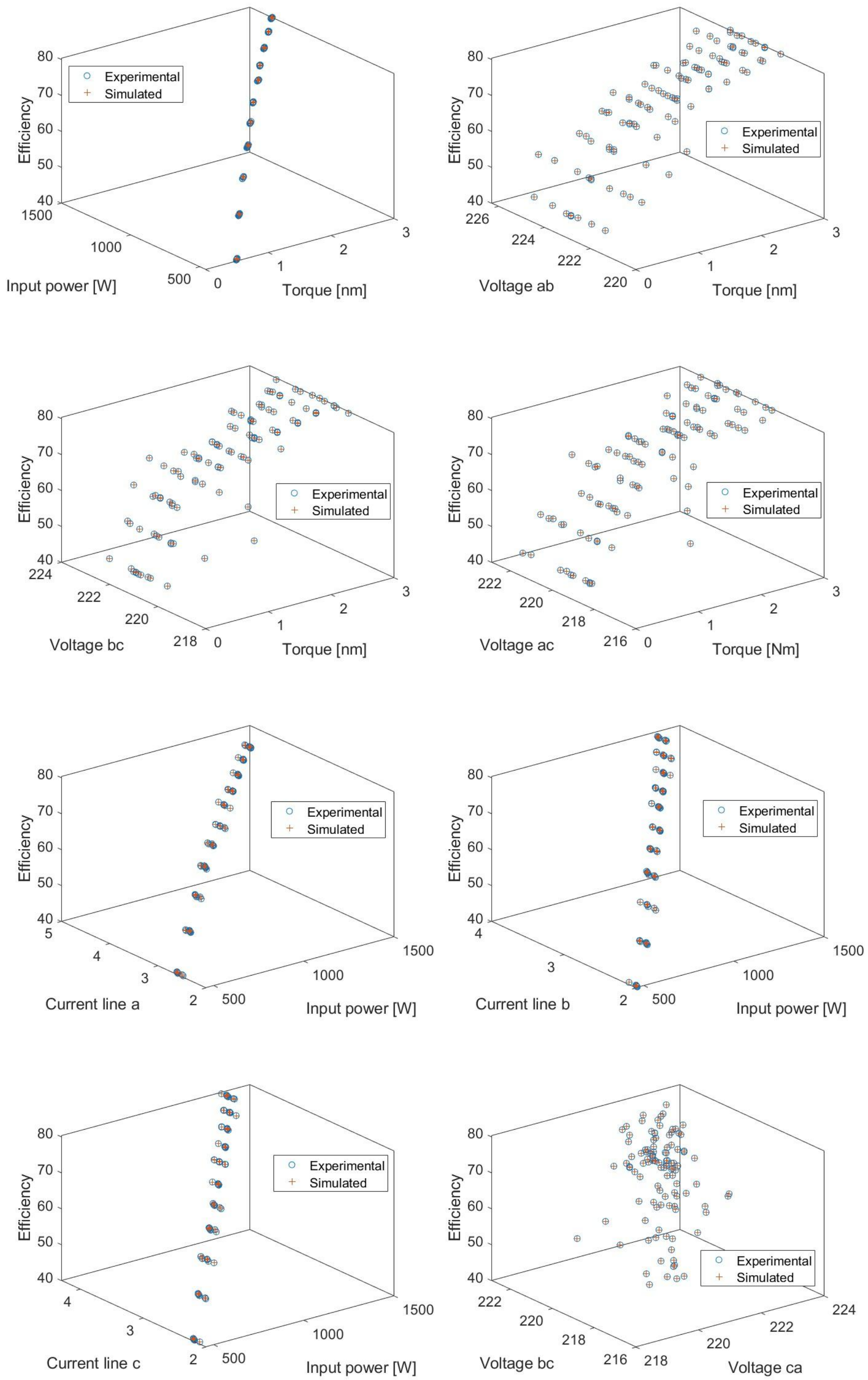


Figure 4. Level surfaces comparing experimental and predicted efficiency coefficients for each significant variable in polynomial model.

4.1. Residual analysis

In this section, a residual analysis will be performed to verify polynomial fit of Equation 3. According to hypothesis, Equation 3 follows a Gaussian distribution, "under the assumption that it has a zero mean", expected value (E) can be taken on both sides of equation as follows:

$$E(ECC) = E\big(p(T_{shaft}, P_{in}, n_m, V_{ab}, V_{bc}, V_{ca}, \theta_{ab}, \theta_{bc}, \theta_{ca}, I_a, I_c, I_c, \phi_a, \phi_b, \phi_c, L_c) + \varepsilon\big) = E\left(p(T_{shaft}, P_{in}, n_m, V_{ab}, V_{bc}, V_{ca}, \theta_{ab}, \theta_{bc}, \theta_{ca}, I_a, I_c, I_c, \phi_a, \phi_b, \phi_c, L_c)\right) + E(\varepsilon)$$

$$ECC = E\left(p(T_{shaft}, P_{in}, n_m, V_{ab}, V_{bc}, V_{ca}, \theta_{ab}, \theta_{bc}, \theta_{ca}, I_a, I_c, I_c, \phi_a, \phi_b, \phi_c, L_c)\right)$$

Considering last equation, experimental performance coefficient and polynomial p would be equal on average. To determine whether e follows a normal distribution with zero mean and constant variance, a histogram of residuals was used (Figure 5). As observed, histogram has a bell-shaped form, suggesting that most residuals are concentrated near 0. This is done to ensure that residuals have zero mean and variance 1.

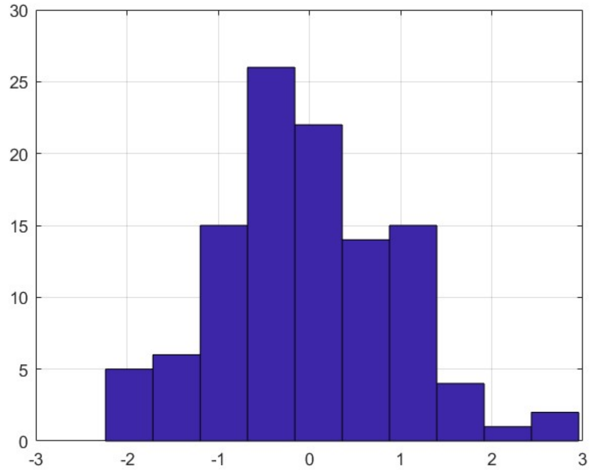


Figure 5. Histogram of residuals.

Residuals were plotted in Figure 6, along with two horizontal lines at 3 and -3. In this experiment, no points were found outside the interval; only one value approached these lines, so it considering that all residuals comply with the 3-sigma rule. This means there are no significant outliers, helping to confirm that residuals follow the expected distribution.

In Figure 7, outliers are observed. To verify whether these values affect polynomial model, five outliers were removed, and a new linear regression was performed without them. A new R2 value of 0.9942 was obtained. The difference compared to R2 = 0.9926 with outliers is not significant, meaning outliers don't interfere with our sample. Thus, it can be concluded that residuals follow a Gaussian distribution with a mean of zero and a variance 1.

To verify that the data obtained from experiment are reliable for our polynomial, statistical F-test and p-value were applied [10]. To calculate the F-statistic, Equation 6 is used.

$$F_{value} = \frac{\frac{SSR}{df_{Model}}}{\frac{SSE}{df_{Error}}} \quad (6)$$

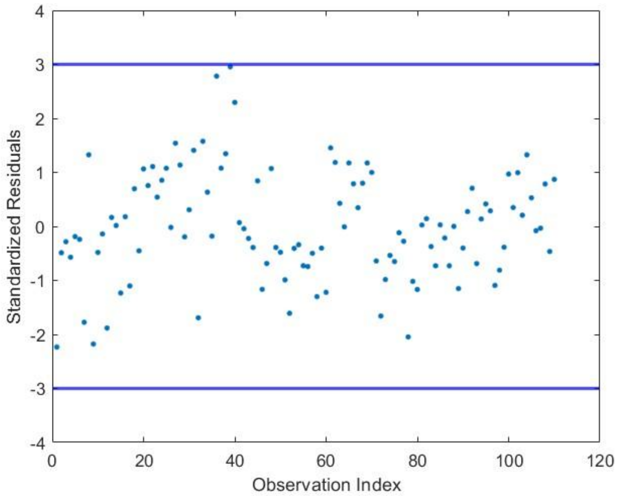


Figure 6. Standardized residuals.

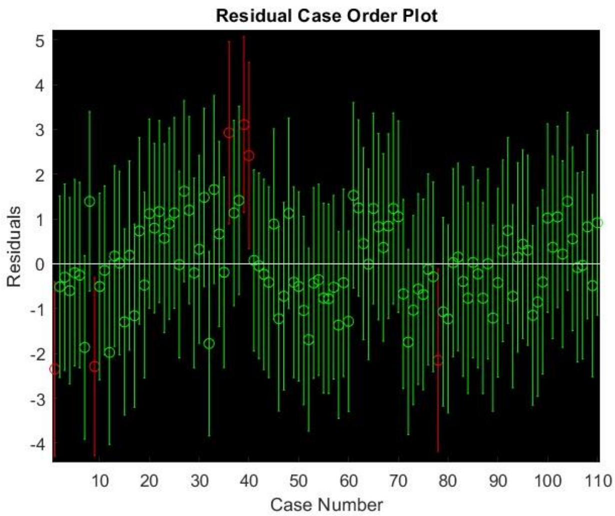


Figure 7. Predicted EEC vs residuals.

Where SSR is sum of squared differences between model data and mean of experimental data, SSE is sum of squared residuals, df_{Model} represents degrees of freedom in polynomial, df_{Error} corresponds to degrees of freedom of error as expressed in Equation 7, and n is number of rows in the database:

$$df_{Error} = n - df_{Model} - 1 \quad (6)$$

These data were used for hypothesis, which consisted of two hypotheses: a null hypothesis H_0 and an alternative hypothesis H_1. For null hypothesis, assumption was made that there is no relationship between independent variables and energy efficiency coefficient, while alternative hypothesis represents opposite-that there is a relationship between variables and EEC. Aim was to validate only one of these hypotheses.

P-value obtained is extremely small 1.1311e-^{105}, indicating that probability of results occurring by chance is practically zero, meaning that null hypothesis is highly improbable. Regarding $F_{value} > F_{critical}$. In other words, data suggest that effect of independent variable is significant, validating that our polynomial has a real impact on variability of data. Thus, we accept the alternative hypothesis.

Table 4. ANOVA table results.

Source	SS	df	MS	F	$Prob > F$
Columns	7.31469×10^7	8	Nm	1047.27	0
Error	8.56477×10^6	981	W		
Total	8.17117×10^7	989	Rpm		

To verify that variables used in our polynomial are independent, an analysis of variance (ANOVA) was performed between variables used for polynomial and experimental EEC, with results shown in Table 4.

Since p-value is 0, we reject null hypothesis, which in this case states that there are no significant differences between means of analyzed groups. This means that there are indeed significant differences among the groups (columns). In simpler terms, at least one of the group means is significantly different from others, suggesting that analyzed factor has a real effect on measurements.

5. Conclusions

Losses in a three-phase induction motor can use many adverse effects, such as excessive heating, reduced life span increased energy consumption. These factors contribute to higher operating costs and greater environmental damage resulting from inefficient energy use. Therefore, real-time monitoring of motor energy efficiency during operation in essential for any induction motor (IM). This requires a sensor in-situ capable of providing such information without the need for additional equipment. Resulting polynomial can calculate EEC, thus reducing costs by relying solely on motor's existing electrical parameters.

In this study, a polynomial was used to predict energy efficiency and induction motor. Favorable results were obtained and a polynomial model with a good determination coefficient (R^2) was achieved, capable of predicting energy efficiency in situ. This high precision suggests that polynomial model captures relationships between operating variables and energy efficiency coefficient.

References

1. N. None, "United states industrial electric motor systems market opportunities assessment," EERE Publication and Product Library, Washington, DC (United States), techreport, 2002.
2. R. Saidur, "A review on electrical motors energy use and energy savings," Renewable and sustainable energy reviews, jourvol 14, number 3, pages 877–898, 2010.
3. J. S. Hsu, J. D. Kueck, M. Olszewski, D. A. Casada, P. J. Otaduy and L. M. Tolbert, "Comparison of induction motor field effciency evaluation methods," inIAS'96. Conference Record of the 1996 IEEE Industry Applications Conference Thirty-First IAS Annual Meeting IEEE, volume 1, 1996, pages 703–712.
4. V. Sousa Santos, J. J. Cabello Eras, A. Sagastume Gutierrez and M. J. Cabello Ulloa, "Assessment of the energy efficiency estimation methods on induction motors considering real-time monitoring," Measurement, jourvol 136, pages 237–247, 2019, issn: 0263-2241.
5. R. J. Rodríguez Lozoya and R. Vásquez León, Análisis del desempeño de motores de inducción trifásicos (Héctor Merino Rodríguez (Guías y manuales de enseñanza de ingeniería)). Universidad Juárez Autónoma de Tabasco, 2025.
6. V. Sousa Santos, J. J. Cabello Eras, A. Sagastume Gutiérrez and M. J. Cabello Ulloa, "Data to support the assessment of the energy efficiency estimation methods on induction motors considering real-time monitoring.,"Data in Brief, jourvol 30, page 105 512, 2020, issn: 2352-3409.
7. B. Lu, T. G. Habetler and R. G. Harley, "A nonintrusive and in-service motor-efficiency estimation method using air-gap torque with considerations of condition monitoring," IEEE Transactions on Industry Applications, jourvol 44, number 6, pages 1666–1674, 2008.

8. A. G. Siraki and P. Pillay, “An in situ efficiency estimation technique for induction machines working with unbalanced supplies,” IEEE Transactions on Energy Conversion, jourvol 27, number 1, pages 85–95, 2012.
9. B. Escobedo-Trujillo, F. Alaffita-Hernández, D. Colorado and J. Siqueiros, “Coefficent of performance prediction by a polynomial model of absorption heat transformer,” Revista mexicana de ingeniería qíımica, jourvol 13, number 3, pages 907–917, 2014.
10. W. Mendenhall, R. J. Beaver, B. M. Beaver and F. S. Fragoso, Introducción a la probabilidad y estadística. Wadsworth Internacional/IberoAmérica, 1982.

COEFFICIENT OF PERFORMANCE PREDICTION BY A POLYNOMIAL MODEL OF MOBILE WICK SOLAR STILL

R. Prieto-Ramírez[1*], L. V. García-Quiñonez[1], F.A. Alaffita-Hernández[1],C. Gómez-Rodríguez[2], M. I. Mendivil-Palma[3]

[1] Centro de Investigación en Recursos Energéticos y Sustentables, Universidad Veracruzana, Av. Universidad Km 7.5 Col. Santa Isabel, Coatzacoalcos Veracruz, CP 96538, México
[2] Universidad Veracruzana Campus Coatzacoalcos, Facultad de Ingeniería, Departamento de Mecánica, C.P, 96535 Coatzacoalcos, Veracruz, México.
[3] Centro de Investigación en Materiales Avanzados (CIMAV), Unidad Monterrey, Alianza Norte 202, Parque de Investigación e Innovación Tecnológica (PIIT), Apodaca, Nuevo León 66600, México.
* Correspondence: zS24019637@estudiantes.uv.mx

Abstract

Prototype used is a mobile wick solar still with passive external condenser, powered by solar energy, developed by the company IPFH2O. Performance modeling involved analysis of global irradiation and temperature parameters. Two polynomial models are proposed to estimate coefficient of performance energy (COP) and coefficient of performance exergy (COP_{exe}), with coefficients of determination of 0.9855 and 0.9985, respectively. Study provides a fast method to estimate energy efficiency and exergy efficiency without requiring complex measurements.

Keywords: Solar still, Global irradiation, Performance coefficient.

1. Introduction

Water purification and distillation using solar radiation is an evolving technique that offers economic advantages, such as electricity savings and water availability in remote areas, avoiding transportation costs [1]. However, this technology suffers from drawback of low efficiency, which may explain its limited adoption [2]. Water production in passive solar stills [3] is influenced by key parameters such as solar irradiation and temperature gradients [4]. Furthermore, these systems can be employed for water desalination via distillation, recognized as one most fundamental and cost-effective purification techniques [5]. Prototype features a 45° tilt angle and a double-glazing configuration to enhance thermal performance and overall system efficiency [6].

2. Materials and Methods

Tests were conducted over four days since October until December 2021 under climate conditions of Rennes, France. Manual solar monitoring of system was conducted at one hour intervals during the test days. [2]. Figure 1 shows prototype used.

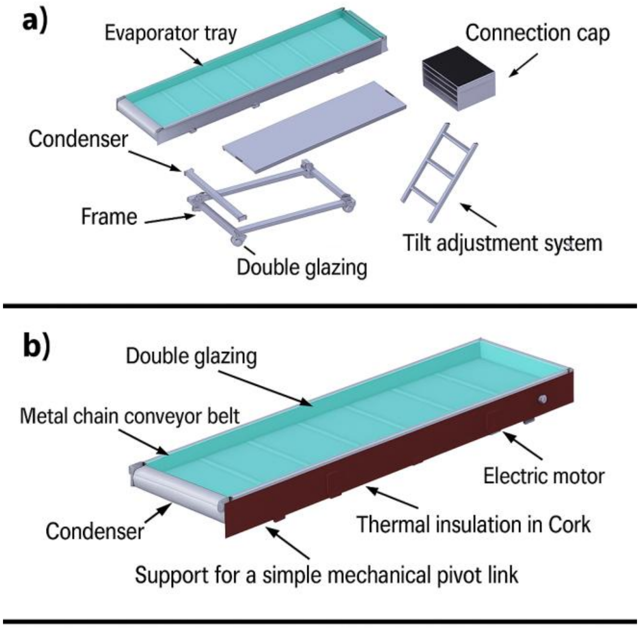


Figure 1. a) Components of the mobile solar still with Passive Condenser, b) Components of the evaporator tray of the solar still [2]

In this section the objective is to describe a model for each coefficient of performance (COP and COP_{exe}). Polynomial fitting was performed using experimental energy efficiency data as a function of global irradiation (I_r) and mobile wick temperature (T_w). Hourly energy efficiency was determined according to Equation 1.

$$n_H = (m_H L)/(A_w I_t) \quad (1)$$

$$L = 3.1615 \times 10^6 - (7.616 \times 10^{-4} \cdot T_w) \quad (2)$$

Where:
n_H: Hourly energy efficiency.
m_H: Mass of water produced per hour in kg.
L: Latent heat of vaporization of water in J/kg.
A_w: Area of the wick surface exposed to the sun in m^2.
T_w: Wick temperature in °C.
I_t: Global irradiation in W/m^2.

The fitting is similar to Equation 3 with two input variables, where p is a polynomial and ϵ represents random noise assumed to follow a normal distribution [7].

$$COP = p(T_w, I_r) + \epsilon \tag{3}$$

Model has a coefficient of determination R^2 of 0.9855, indicating that it explains most of the variability. Similarly, experimental exergy data were used to fit a polynomial model based on collected variables: Wick temperature (T_w), Global Irradiation (I_r), Ambient Temperature (T_a), and Interior Glass Temperature (T_{gi}). Exergy was calculated using Equation 4.

$$n_{exer} = Ex_{ou}t/Ex_{inp} \tag{4}$$

Where:
n_{exer}: Exergy
Ex_{out}: Exergy related to evaporation process in the space between wick and inner glass.
Ex_{inp}: Exergy of the solar radiation absorbed by the wick.

Fitted model for exergy is presented in Equation 5, with four input variables.

$$COP_{exe} = p(T_w, I_r, T_{gi}, T_a) + \epsilon \tag{5}$$

It achieved a coefficient of determination R^2 of 0.9985, indicating that 99.85% of the variability in exergy coefficient of performance can be explained by this polynomial model.

System variables are shown in Table 1.

Table 1. System variables.

	Variable	Unidad
x1	T_w	°C
x2	I_r	W/m^2
x3	T_a	°C
x4	T_{gi}	°C

Input 1 (x1): Temperature of the mobile wick, 2 °C lower than temperature inside the evaporator tray. Due to wick is absorption of water and heat, its temperature can be slightly lower [8].

Input 2 (x2): Global irradiation, corresponding to values obtained with an analog pyranometer during the experiment.
Input 3 (x3): Ambient temperature.
Input 4 (x4): Inner glass temperature.

Polynomial fitting method is an effective strategy for predicting the behavior of a dependent variable within a system, using measurements from collected data [9]. Two polynomial models were developed to predict behavior of two indicators: the energy performance coefficient (COP) and the exergy performance coefficient (COP_{exe}).

Both polynomials were obtained through multivariable polynomial regression, seeking the relationship between the independent variables and the performance coefficients. Models include both linear and interaction terms.

3. Results

Energy Efficiency Model (Eq. 6): This model is a third degree polynomial with two independent variables, T_w and I_r, composed of nine terms. The selection of the independent variables was based on Eq. 1, where water mass produced per hour (m_H) and surface area of the wick exposed to the sun (A_w) were considered constant.

$$COP = 110.4348 + 4.4378(x1) - 1.3355(x2) - 0.2966(x1)^2 + 0.0045(x2)^2 - 3.9751x10^{-4}(x1)(x2) + 0.0063(x1)^3 - 4.7245x10^{-6}(x2)^3 \quad (6)$$

Exergy Model (Eq. 7): A quadratic polynomial was constructed, consisting of 15 terms. Four independent variables were considered for its development: T_w, I_r, T_a, an T_{gi}, which are used in the experimental exergy.

$$COP_{exe} = -238.7276 - 1.7037(x1)^2 + 0.0006(x2)^2 - 3.3545(x3)^2 - 3.7648(x4)^2 - 0.0310(x1)(x2) + 5.7139(x1)(x3) + 5.2319(x1)(x4) + 43.000(x2)(x3) + 0.0271(x2)(x4) - 5.5808(x3)(x4) - 74.8790(x1) + 0.3003(x2) + 47.8111(x3) + 82.5614(x4) \quad (7)$$

Figure 2b shows low correlation between independent variables, which benefits obtained model for Energy Efficiency. On the left, plot of Experimental vs predicted Efficiency confirms a satisfactory data fit, showing how points align with trend line.

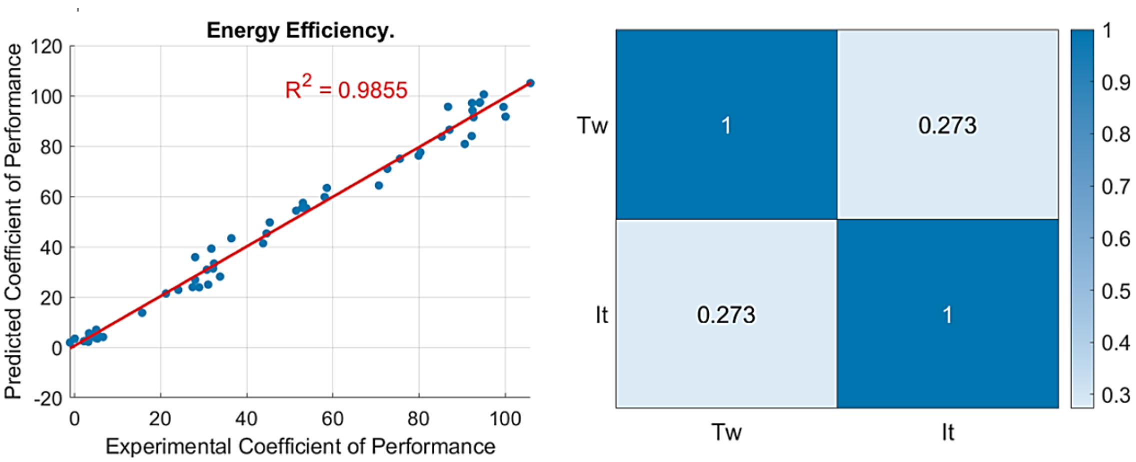


Figure 2. a)Experimental versus predicted coefficient of performance for database, b) correlation matrix between input variables. Energy Efficiency.

Exergy indicates how much energy content can be utilized from system. In a solar still, energy losses are lower compared to other equipment such as boilers [10], which facilitates model application. This is reflected in graph of experimental vs calculated exergy values (Figure 3), where proximity of points to trend line indicates a good model fit.

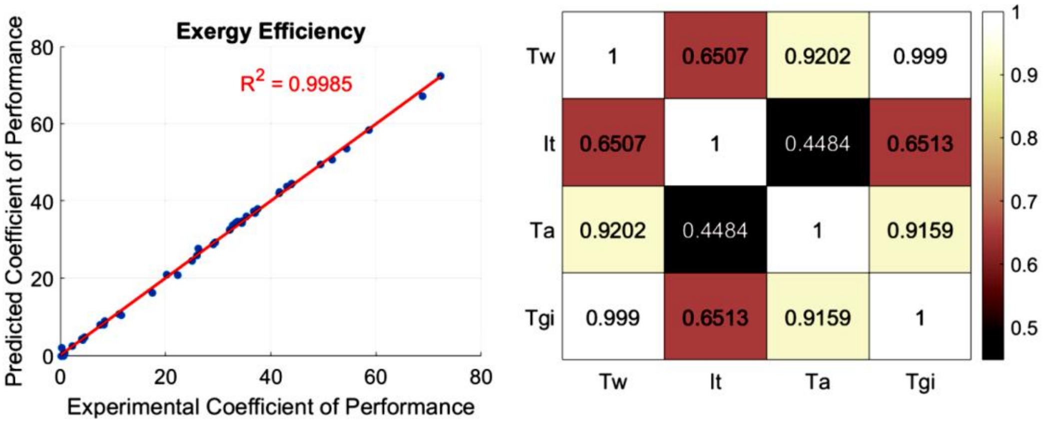


Figure 3. a) Experimental versus predicted coefficient of performance for database, b) correlation matrix between input variables. Exergy.

3.1 Residual Analysis

This section presents evaluation of residuals using histograms and probability plots to determine their normality. In addition, residual plots are generated to identify outliers, which helps verify quality of polynomial model [10]. To understand behavior of two developed polynomials, experimental and predicted coefficients of performance corresponding each one are graphically represented. Both polynomials include different variables, contour plots are generated for each of them. Graphical representations are shown in Figures 4, 5, 6, 7, 8, and 9. Agreement between predicted and experimental values is observed.

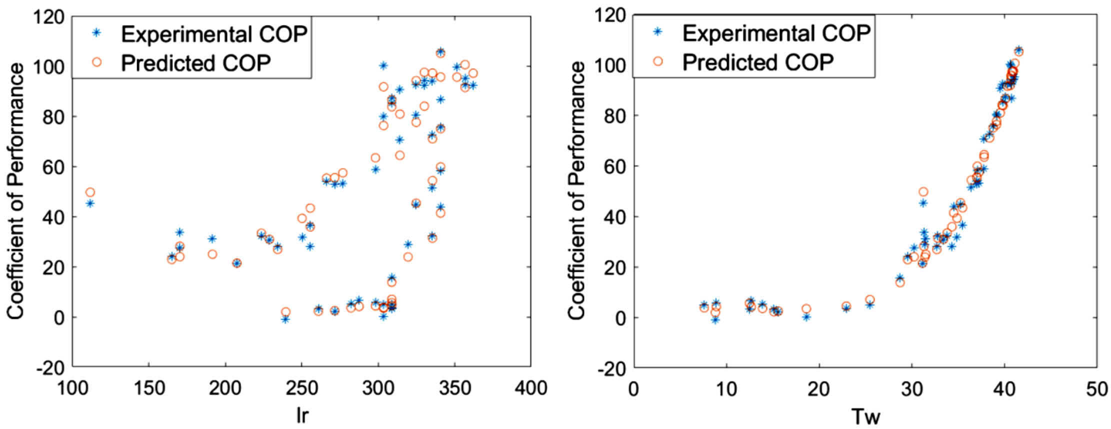


Figure 4. Level surfaces. Comparison of experimental and predicted coefficient of performance for each variable significant in the polynomial model Energy Efficiency

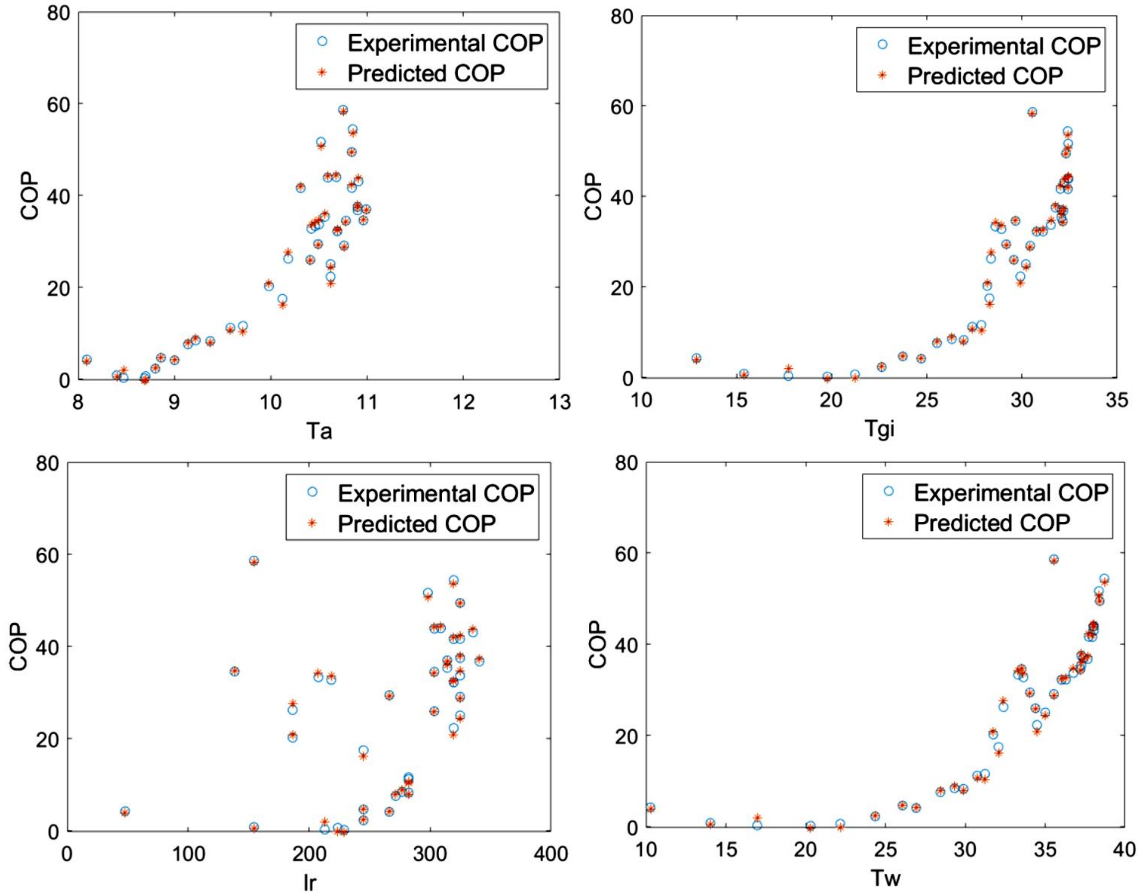


Figure 5. Level surfaces. Comparison of experimental and predicted coefficient of performance for each variable significant in the polynomial model Exergy

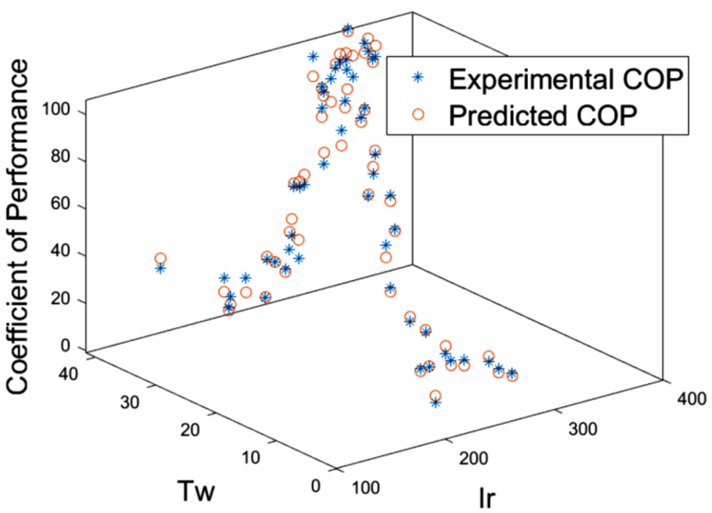


Figure 6. Level surfaces. Comparison of experimental and predicted coefficient of performance for two variables significants in the polynomial model Energy Efficiency.

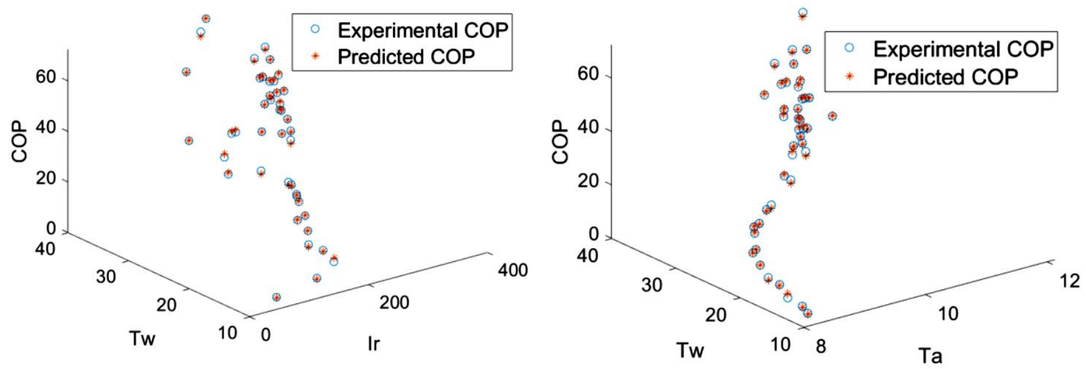


Figure 7. Level surfaces. Comparison of experimental and predicted coefficient of performance for Tw/Ir and Tw/Ta in the polynomial model Exergy.

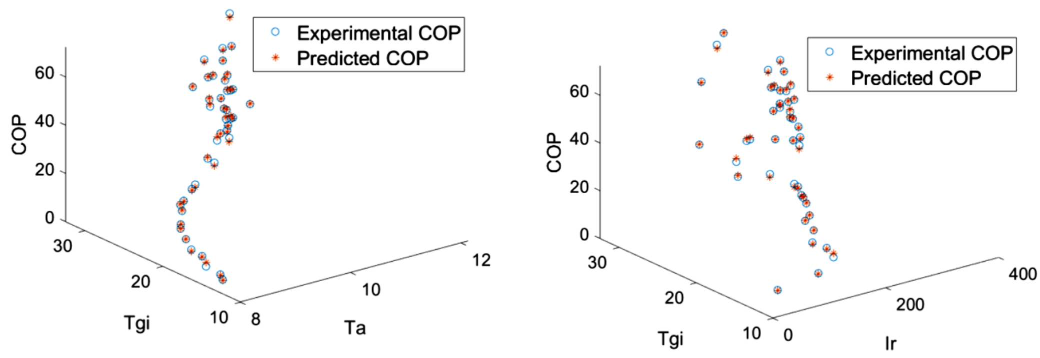


Figure 8. Level surfaces. Comparison of experimental and predicted coefficient of performance for Tgi/Ta and Tgi/Ir in the polynomial model Exergy.

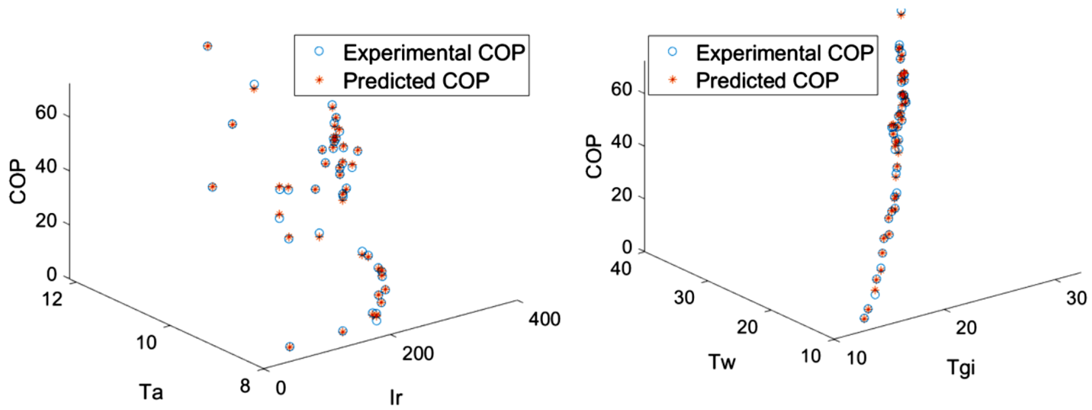


Figure 9. Level surfaces. Comparison of experimental and predicted coefficient of performance for Ta/Ir and Tw/Tgi in the polynomial model Exergy.

Assuming that models described in Equations 3 and 5 follows a Gaussian distribution, residual histograms were constructed for each coefficient of performance, following methodology outlined by Montgomery [10]. Histograms have a bell-shaped form, as shown in Figure 10. However, in Exergy case , a notorious bar appears at end of histogram, which may indicate the indicate of outliers (see Figure 11).

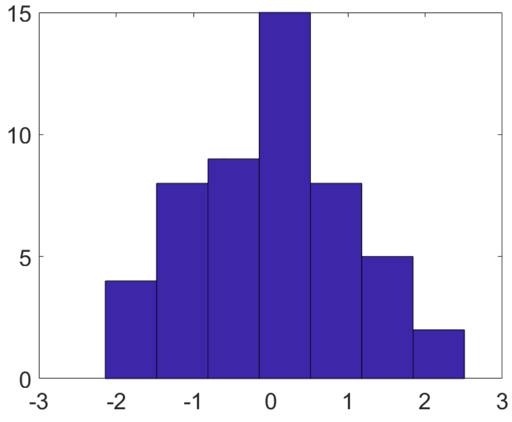


Figure 10. Histogram of residuals for Energy Efficiency.

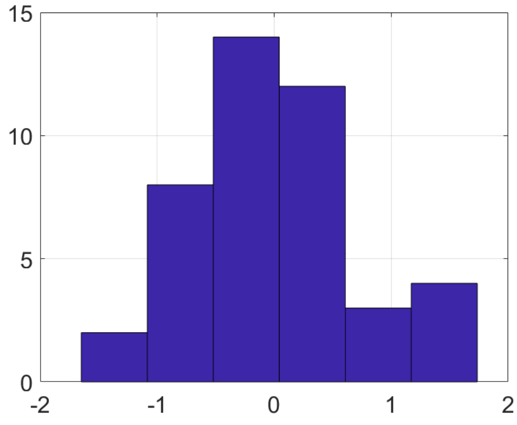


Figure 11. Histogram of residuals for Exergy

Figure 12 was generated to show fitting goodness of proposed models. Displays standardized confidence intervals (green lines) corresponding to each residual (represented by circles), with outliers highlighted in red. This indicates that predicted values closely match observed ones. Furthermore, Figures 13 and 14 show that mean of residuals is approximately zero, as their distribution remains balanced on either side of y-axis.

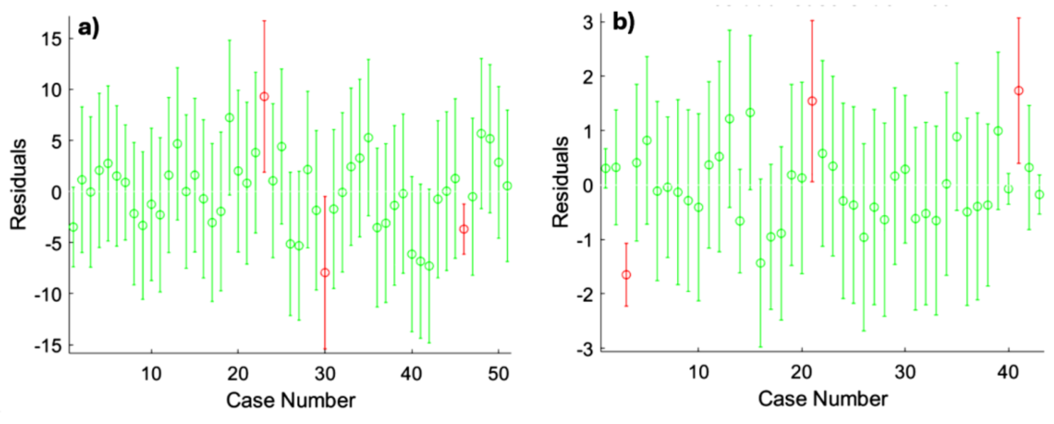


Figure 12. Confidence intervals of residuals with a 95% confidence level: a) Energy Efficiency and b) Exergy.

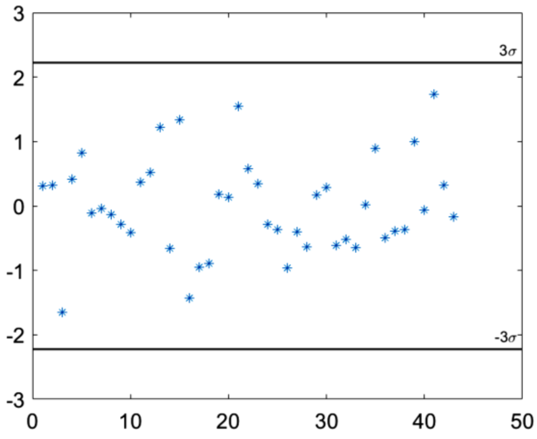


Figure 13. Predicted coefficient of performance for Energy Efficiency versus standardized residuals.

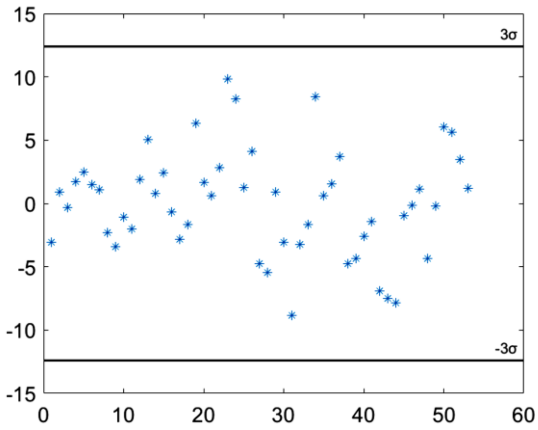


Figure 14. Predicted coefficient of performance for Exergy versus standardized residuals.

.

3.2 Anova

An analysis of variance (ANOVA) was carried on both proposed models (Table 2 y 3), demonstrating that each is highly significant. Null hypothesis (H_0) states that no independent variables have a significant effect on dependent variables. Alternative hypothesis (H_1) states that at least one independent variable has a significant effect on dependent variable, meaning model explains a significant part of observed variability. For Energy Efficiency model, Error accounts for less than 1.5% of total data variability (887.6 out of 61,068), which indicates an excellent fit to experimental data. A p-value below 0.001 confirms that the model is highly significant, suggesting the observed results are very unlikely under null hypothesis. Based on F-distribution tables, for a 5% significance level and with 7 degrees of freedom in numerator (independent variables) and 45 degrees of freedom in denominator (total number of observations minus the number of estimated parameters), critical value is:

$$F_{critical} = 3.065$$

so,

$$F_{calculated} = 485.87 > F_{critical}l = 3.065$$

Therefore, null hypothesis is rejected, and model is statistically significant. Is accepted that at least one variable has effect on Energy Efficiency.

Table 2: Analysis of variance (ANOVA) of polynomial model for Energy Efficiency

Source	Sum of Squares (SS)	DF	Mean Square (MS)	F	p-value
Model	60,181	7	8 597.2	435.87	0
Error	887.6	45	19.72		
Total	**61,068**	**52**			

Table 3. Analysis of variance (ANOVA) of polynomial model for Exergy.

Source	Sum of Squares (SS)	DF	Mean Square (MS)	F	p-value
Model	15 214	14	1 086.7	1313.6	0
Error	23.164	28	0.82729		
Total	**15238**	**42**			

4. Conclusions

During implementation of a solar still with a mobile wick, data collection can be affected by climatic limitations, as well as difficulties in handling equipment due to weight. The development of a model for estimating system efficiency offers valuable support for obtaining operational data, reducing costly experimental procedures. Two polynomial models were successfully developed to predict coefficient of performance for Energy and Exergy Efficiency in a solar still with mobile wick and passive external condenser, with coefficients of determination exceeding 98% is possible to estimate system behaviour using easily measurable variables, thus avoiding complex instrumentation.

References

1. S. F., E. M., & A. T. Modelo matemático y solución analítica del funcionamiento de un destilador solar de bandeja. *Tecnología Química 32* (Aug. 2012), pp. 180–190. ISSN: 2224-6185.

2. D., S., S. M., F., S., B., P., & S. O. Data on the experimental characterization of a mobile wick solar still with external condenser tested under the climatic conditions of Rennes, France. In: Data in Brief 54 (2024), p. 110395. ISSN: 2352-3409. DOI: https://doi.org/10. 1016/j.dib.2024.110395.

3. M. E. Solar stills: A review of the latest developments in numerical simulations. *Solar Energy* 135 (2016), pp. 897–922. ISSN: 0038-092X. DOI: https://doi.org/10.1016/j.solener.2016.03.005.

4. C.T. M., P. F., I., S.J., Á., & F. Z., J. L. Experimental study on the air velocity effect on the efficiency and fresh water production in a forced convective double slope solar still. *Applied Thermal Engineering* 75 (2015), pp. 1192–1200. ISSN: 1359-4311. DOI: https : / / doi . org / 10 . 1016 / j . applthermaleng . 2014 . 10 . 032.

5. D. M., H. P., & K. K. S. Investigation on evacuated tubes coupled solar still with condenser and fins: Experimental, exergo-economic and exergo-environment analysis.: *Case Studies in Thermal Engineering* 27 (2021), p. 101217. ISSN: 2214-157X. DOI: https://doi.org/10.1016/j.csite.2021 . 101217.

6. M. K., M. E., & K. J. Effect of design parameters on performance of passive basin solar stills considering instantaneous ambient conditions: A transient CFD modeling. *Solar Energy* 201 (2020), pp. 884–907. ISSN: 0038-092X. DOI: https://doi.org/10.1016/j.solener.2020.03.068.

7. J.C. Análisis en los residuos de redes altimétricas mínimamente condicionadas. *Revista de Topografía AZIMUT* 12.1 (2021), pp. 01–07.

8. H. S. Mathematical modeling of an inclined solar water distillation system. *Desalination* 190.1 (2006), pp. 63–70. ISSN: 0011-9164. DOI: https:// doi.org/10.1016/ j.desal.2005.07.015.

9. D. C., E.T., B. A., C.M., I., F. A., & H.R., J. V. Selection of the best coefficient of performance prediction by artificial neural network model considering uncertainty. *Desalination and Water Treatment* 92 (2017), pp. 60–71. ISSN: 1944-3986. DOI: https://doi.org/10.5004/dwt.2017.21504.

10. D. C. M., A. P., & G. G. Introduction to linear regression analysis. John Wiley & Sons, 2021.

11. B. A. E. T., F. A. A. H., D. C. & J. S. Coefficient of performance prediction by a polynomial model of absorption heat transformer. *Revista Mexicana de Ingeniería **Q**uímica.* 13 (2014), pp. 907-917.

PREDICCIÓN DE ARSÉNICO REMANENTE EN SU EXTRACCIÓN CON QUITOSANO MEDIANTE UN MODELO POLINOMIAL OBTENIDO CON EL MODELO DE RESPUESTA DE SUPERFICIE (RSM)

F. Guzmán-Recino[1,*], F. A. Alaffita-Hernández [2], y B. A. Roger-Fouconnier [1]

[1] Facultad de Ciencias Químicas, Universidad Veracruzana, Avenida Universidad, Col. Santa Isabel, Coatzacoalcos, Veracruz, C. P. 96538, MÉXICO; feguzman@uv.mx, broger@uv.mx
[2] Centro de Investigación en Recursos Energéticos y Sustentables, Universidad Veracruzana, Col. Santa Isabel, Coatzacoalcos, Veracruz, C. P. 96538, MÉXICO, falaffita@uv.mx
* Correspondencia: feguzman@uv.mx

Resumen

Se propone un modelo polinomial, expresado en variables codificadas y naturales, para predecir la cantidad de arsénico remanente al efectuar la extracción de dicho contaminante en medio acuoso empleando quitosano como medio adsorbente. Se aplica el Método de Respuesta de Superficie (*RSM)* considerando las condiciones de pH, tiempo de contacto, concentración de arsénico y concentración del adsorbente (quitosano) para proponer las condiciones óptimas en las que aumenta la remoción de arsénico. El modelo polinomial propuesto presenta una buena aproximación, pues el coeficiente de determinación obtenido es de 0.9983. Se presentaron curvas de nivel y análisis residuales para la validación del modelo. Este trabajo pretende ser de ayuda para la predicción de arsénico remanente y así disminuir la cantidad de experimentos a realizar en el análisis de extracción de arsénico mediante quitosano.

Palabras clave: RSM, arsénico, contaminante, modelo polinomial, adsorción

1. Introducción

El arsénico es un metal pesado que se puede encontrar de manera natural en rocas y sedimentos; sin embargo, las concentraciones de éste tienden a aumentar en el medio natural debido a las actividades humanas, tal como la quema de carbón que usualmente es efectuada en centrales eléctricas, industrias de fundición de cobre y procesamiento de minerales [1]. En algunas ocasiones, este compuesto inorgánico es difícil de extraer del medio en el que se encuentra debido a las diversas condiciones, como lo puede ser el pH [2]. Por otro lado, el arsénico es tóxico al igual que los demás metales pesados, por lo que merece atención especial por su volatilidad, su capacidad de bioacumularse en el ambiente y a que es potencialmente cancerígeno [3].

Debido a ello, se ha estado estudiando la implementación de diversos materiales y métodos que ayuden a su extracción, siendo el quitosano uno de los más empleados para dicha extracción, el cual suele extraerse del caparazón de mariscos. Esto se debe a que aunque existen diversos métodos que pueden emplearse para remover el arsénico, como: oxidación/precipitación, coagulación, intercambio iónico y tecnología de membranas, el quitosano presenta bajo costo, se puede conseguir fácilmente, es biodegradable, no es tóxico y posee propiedades estructurales que favorecen la extracción del arsénico, ya que se lleva a cabo mediante el método de adsorción, el cual es uno de los métodos más simples y efectivos en remoción de contaminantes [4].

Generalmente, este tipo de análisis de extracción de contaminantes suele modelarse mediante el Método de Respuesta de Superficie (*RSM*, por sus siglas en inglés), pues es empleado para el análisis de procesos químicos. No obstante, también puede utilizarse a nivel ingeniería, ya que analiza los efectos de diversos parámetros de entrada (ξ_1, ξ_2, …, ξ_k) que puedan influir en una respuesta de salida, lo cual ayuda a reducir el número de experimentos realizados para el análisis de remoción de contaminantes mediante predicciones, tal como se ha hecho para el análisis de remoción de amoniaco [5]. En este método se utilizan variables codificadas (x_1, x_2, …, x_k) en lugar de las variables naturales, las cuales son las variables que se encuentran en las unidades en las que se realizaron las mediciones [6]; en este caso, en unidades de concentración (mg/L), pH y tiempo (min). Las variables codificadas consisten en variables adimensionales que facilitan el análisis sobre qué variables tienen mayor impacto en la respuesta del proceso, ya que de esta forma se pueden estimar los términos del modelo de forma independiente.

En el presente trabajo se busca proponer un polinomio en variables codificadas y otro en variables naturales que sea capaz de predecir la cantidad de arsénico remanente al aplicar quitosano como medio adsorbente en un medio acuoso, tomando en cuenta la variación de pH, tiempo de contacto, concentraciones de arsénico y concentraciones de adsorbente (quitosano), y, al mismo tiempo, analizar la relación que presenta el tiempo de contacto con la cantidad de adsorbato inicial presente en el proceso de remoción.

2. Metodología

Se obtuvieron datos experimentales de Dehghani y colaboradores [7], quienes analizan los efectos que poseen la cantidad de arsénico inicial, la cantidad de adsorbente (quitosano), el pH y el tiempo de contacto en un proceso de extracción de arsénico. Los datos se presentan en el Apéndice A.

Para encontrar el modelo de predicción de arsénico remanente se utilizó el software RStudio 2021.09.1 con el Método de Respuesta de Superficie, el cual relaciona condiciones de entrada con la variable de salida analizada; en este caso, la variable de salida es el arsénico remanente (Ecuación 1), donde ε incluye los efectos de errores de medición y efectos externos que no pueden controlarse, como los niveles de humedad, cambio de las propiedades y envejecimiento del material [6] y se espera que presente una distribución normal con varianza constante y media cero.

$$y = f(\xi_1, \xi_2, \dots, \xi_k) + \varepsilon \tag{1}$$

Sin embargo, el *RSM* analiza las variables codificadas (x_1, x_2, …, x_k) en lugar de las variables naturales (ξ_1, ξ_2, …, ξ_k). Para realizar la transformación de variables se utiliza la Ecuación 2, en la cual se consideran los valores máximos, mínimos y centrales de los datos experimentales:

$$x_i = \frac{\xi_i - [\max(\xi_i) + \min(\xi_i)]/2}{[\max(\xi_i) - \min(\xi_i)]/2} \tag{2}$$

Una vez efectuado el cambio de variables, la aproximación se representa con la Ecuación 3:

$$y = f(x_1, x_2, \dots, x_k) + \ \varepsilon \tag{3}$$

Así mismo, se asignaron factores para identificar las condiciones experimentales estudiadas (Tabla 1).

Tabla 1. Asignación de factores

Factor	Variables
A	Cantidad de arsénico
B	Cantidad de quitosano
C	pH
D	Tiempo de contacto

[1] Variables asignadas a cada factor empleado en el Método de Respuesta de Superficie

Los niveles de las variables de operación (A, B, C y D) se establecieron en tres niveles: -1 (mínimo), 0 (centro), 1 (máximo), tal como se muestra en la Tabla 2.

Tabla 2. Rango experimental y codificado de las variables de operación

Factor	Rango y niveles		
	-1	**0**	**1**
A	194	295.5	397
B	2	3	4
C	4	5	6
D	45	65	75

* Se colocan datos codificados para facilitar la manipulación de datos durante el análisis estadístico.

Para validar el modelo polinomial se realizaron análisis residuales mediante métodos analíticos y gráficos. Por otro lado, se empleó el *RSM* para analizar la interacción que tiene la cantidad de arsénico inicial con el tiempo de contacto en cuanto a la adsorción de arsénico con quitosano.

3. Resultados y discusión

3.1. Modelo polinomial

Se presenta el modelo polinomial (Ecuación 4) propuesto en variables codificadas, el cual presenta unidades de concentración en mg/L:

$$As_{remanente} = 215.51 + 82.389\boldsymbol{A} - 6.2633\boldsymbol{B} + 1.7633\boldsymbol{C} - 16.871\boldsymbol{D} - 4.1386\boldsymbol{AB} - 3.4787\boldsymbol{AD} - 2.2633\boldsymbol{BD} + 3.7744\boldsymbol{A}^2 + 3.232\boldsymbol{D}^2 \quad (4)$$

En el polinomio se utilizaron las cuatro variables de investigación (pH, tiempo de contacto y concentración de adsorbente y adsorbato), el cual posee un coeficiente de determinación de 0.9983, siendo el valor de 1 el que representa la mejor aproximación, lo cual indica una buena correlación [8,9]. Por otro lado, al comparar los valores de arsénico remanente predichos y experimentales (Figura 1) se obtuvo que los valores predichos coinciden con la mayor parte de los valores objetivo, pues coinciden tanto que se obtiene algo parecido a la recta identidad, la cual posee una pendiente de 1 y un intercepto de 0.

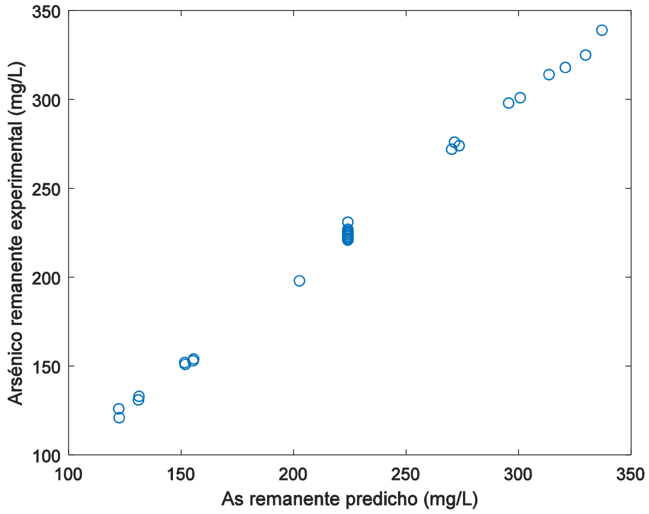


Figura 1. Arsénico remanente predicho contra el arsénico remanente experimental.

La comparación entre el arsénico remanente experimental y predicho con respecto a cada variable de análisis se muestra en la Figura 2, en la cual se puede apreciar que la mayoría de los valores aproximados coinciden con los experimentales, siendo esto otro indicativo de que el modelo polinomial propuesto presenta buena aproximación; dicha figura surge de un seccionamiento de superficies de nivel en 3 dimensiones (Figura 3).

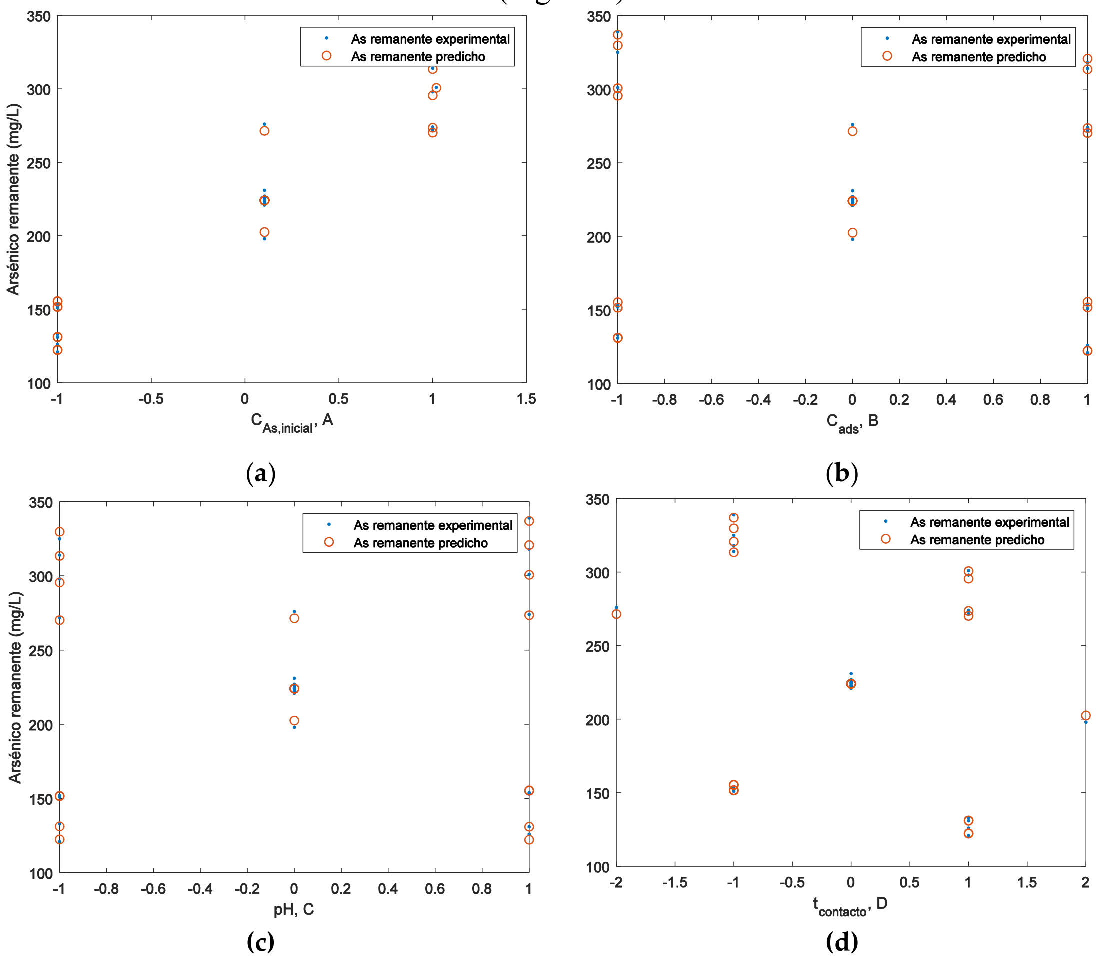
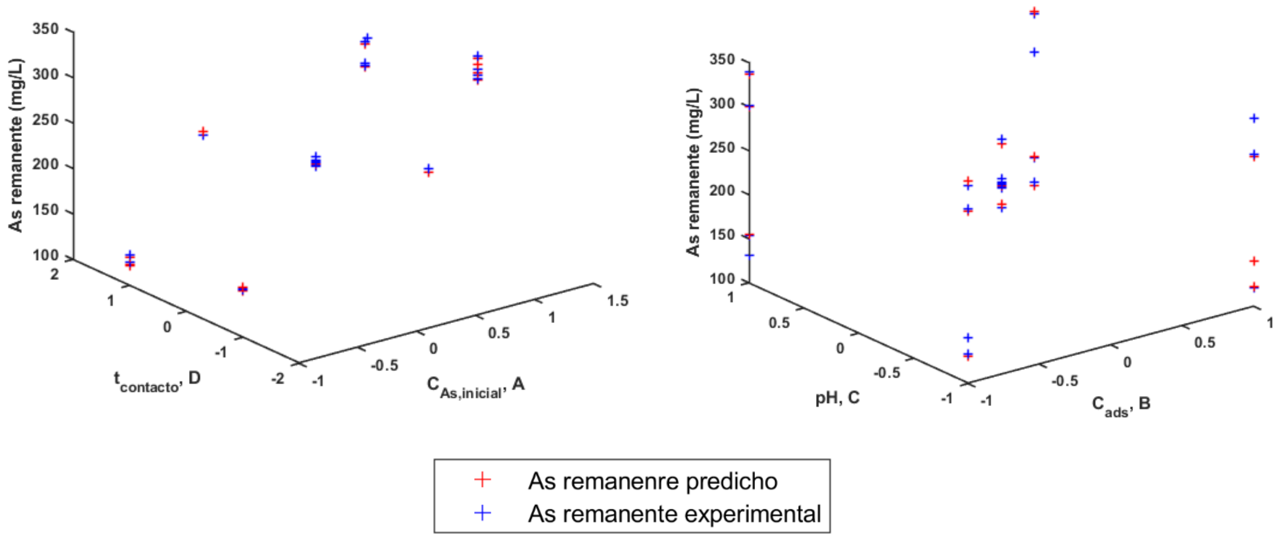


Figura 2. Curvas de nivel. Comparación del arsénico remanente experimental y el arsénico predicho para cada variable utilizada en el polinomio: (a) concentración inicial de arsénico; (b) concentración del adsorbente; (c) pH; (d) tiempo de contacto.



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

Figura 3. Superficies de nivel. Comparación del arsénico remanente experimental y el arsénico remanente predicho para las variables significativas del modelo polinomial: (a) tiempo de contacto y concentración de arsénico inicial; (b) pH del medio y la concentración del adsorbente.

3.2. Análisis estadístico

El histograma de residuos estandarizados (Figura 4) presenta una distribución similar a la campana de Gauss, la cual una distribución normal con media cero y varianza constante.

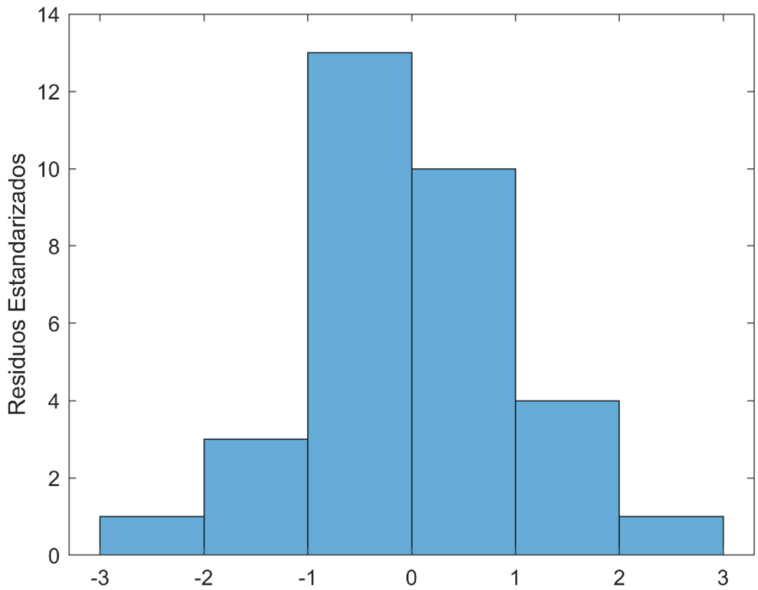


Figura 4. Histograma de residuos estandarizados.

Se empleó el método 3-sigma, con líneas horizontales de 3 y -3 (Figura 5) para identificar alguno de los *outliers* que parecen estar presentes según el histograma de residuos estandarizados (Figura 4), ya que no presenta una simetría perfecta. Sin embargo, no se pudo apreciar ningún valor atípico, pues el 100% de valores se encuentra dentro de los valores de 3 y -3.

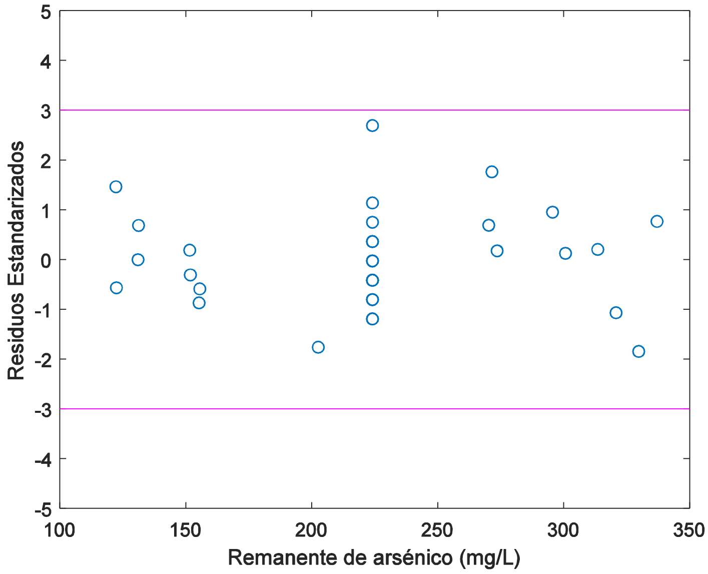


Figura 5. Residuos estandarizados del arsénico remanente predicho y experimental.

Se obtuvieron los intervalos de confianza de los residuos con un nivel de confianza del 95 % (Figura 6), donde los residuos se representan con puntos amarillos y los intervalos de confianza con líneas verdes. En dicha figura se aprecian 4 valores atípicos, los cuales pueden ser los que se ven reflejados en el histograma de residuos estandarizados. Al mismo tiempo, se observa que los residuos (ε) presentan media cero y su varianza es constante, pues se logra un "efecto espejo" en los intervalos de confianza, debido a que presentan una anchura similar, lo cual indica que el valor es consistente y aleatorio en todo el rango de datos observados.

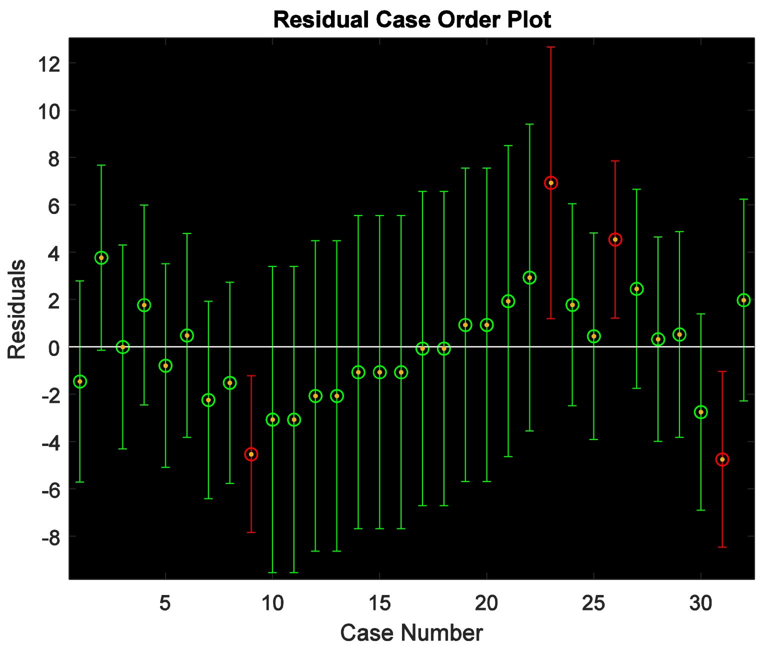


Figura 6. Intervalos de confianza en los residuos con un nivel de confianza del 95%.

El aporte de cada variable en el modelo se analizó mediante el Análisis de Varianza (*ANOVA*), Tabla 3. En dicho análisis se puede observar que, con base al *P-value,* todas las variables elegidas para el modelo polinomial son importantes y altamente importantes para lograr una buena aproximación. El *P-value* ayuda a diferenciar resultados que son producto del azar del muestreo de resultados que son estadísticamente significativos.

Tabla 3. Coeficientes estimados del modelo polinomial de segundo grado y resultados de ANOVA.

	Modelo cuadrático		
Factor	**Coeficiente estimado**	**Error estándar**	**P-*value***
A	82.3887	0.8136	<0.0001
B	-6.2633	0.8149	<0.0001
C	1.7633	0.8149	0.041593
D	-16.8712	0.6659	<0.0001
AB	-4.1386	0.8139	<0.0002
AD	-3.4787	0.8124	<0.0003
BD	-2.2633	0.8149	0.01098
A^2	3.7744	1.2052	0.004852
D^2	3.232	0.6160	<0.0001

* $P<0.001$ altamente importante; $0.001\leq P \leq 0.05$ importante; $P > 0.05$ no importante.

El análisis de la interacción del tiempo de contacto y el arsénico inicial durante el proceso de adsorción se presenta en la Figura 7. En dicho análisis se utilizaron una concentración de adsorbente de 3 mg/L y un pH de 5; con base a ello, la mayor cantidad de arsénico remanente se obtiene cuando el tiempo de contacto del adsorbato y el adsorbente se encuentra entre los 30 y 40 min con grandes concentraciones de arsénico inicial, mientras que la cantidad óptima de arsénico remanente se obtiene cuando se trabaja con un tiempo de contacto aproximado de 60 y 90 min con concentraciones de arsénico inicial entre 200 y 250 mg/L, pues se pretende obtener la menor cantidad de arsénico remanente posible.

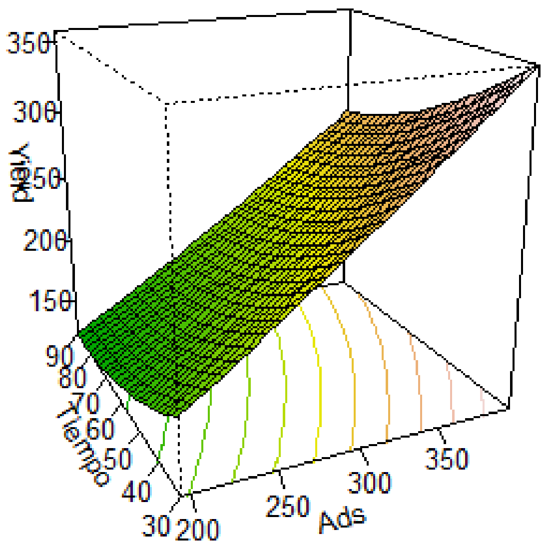


Figura 7. Respuesta de superficie de la interacción entre el tiempo de contacto y la concentración de arsénico inicial

El modelo polinomial en términos de variables naturales (Ecuación 6) es el siguiente:

$$\begin{aligned} As_{remanente} = {} & 32.99 + 0.85461(C_{As,inicial}) + 14.839(C_{ads}) + 1.7633(pH) - 1.7206(t) \\ & - 0.040775(C_{As,inicial} * C_{ads}) - 0.0022849(C_{As,inicial} * t) - 0.15089(C_{ads} * t) \\ & + 0.00036636({C_{As,inicial}}^2) + 0.014364(t^2) \end{aligned} \quad (6)$$

Donde: $C_{As,inicial}$, es la concentración de arsénico inicial, mg/L, C_{ads}, es la concentración de adsorbente empleado, mg/L, pH, es el del medio en el que se efectuó la adsorción, t, es el tiempo de contacto, min.

4. Conclusiones

Se obtuvo un polinomio con un coeficiente de determinación de 0.9983, y con base al análisis residual realizado, se puede decir que el polinomio obtenido con 4 variables puede predecir correctamente el arsénico remanente en un rango de pH (4 a 6), tiempo de contacto (45 a 75 min), concentración de arsénico inicial (194 a 397 mg/L) y concentración de adsorbente (2 a 4 mg/L). Asimismo, se determinó que la concentración de arsénico inicial y el tiempo de contacto tienen un impacto importante en el proceso de adsorción, pues mientras mayor sea el tiempo de contacto menor es la cantidad de arsénico remanente, siendo un rango óptimo de operación un tiempo de contacto entre 60 y 90 min con arsénico inicial entre 200 y 250 mg/L.

Contribución de los autores: Manejo de variables para análisis estadístico, metodología, validación, interpretación de resultados, investigación, escritura, visualización: Fernanda Guzmán Recino Conceptualización y supervisión del proyecto: Francisco Alejandro Alaffita Hernández. Conceptualización: Benoit Auguste Roger Fouconnier.

Financiamiento: La autora agradece a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) por el otorgamiento de la beca de doctorado.

Declaración del Comité de Ética: No aplicable.

Declaración de Consentimiento Informado: No aplicable.
Declaración de Disponibilidad de Datos: Las contribuciones originales de este estudio están incluidas en el artículo; cualquier consulta adicional puede dirigirse a los autores de correspondencia.

Conflictos de Interés: Los autores declaran que no hay conflictos den interés.

Abreviaturas

RSM	Método de Respuesta de Superficie
As	Arsénico
$t_{contacto}$	Tiempo de contacto
C_{ads}	Concentración adsorbida
$C_{AS, inicial}$	Concentración inicial de arsénico

Apéndice A

Tabla A. Datos utilizados para la obtención del modelo polinomial.

Corrida	$C_{As,inicial}$ (mg/L)	C_{ads} (mg/L)	pH	Tiempo (min)	Resultado	As removido	% Removido
1	306	3	5	60	231	75	24.51
2	397	4	6	75	274	123	30.98
3	397	4	6	45	318	79	19.9
4	397	2	4	75	298	99	24.94
5	194	4	6	75	126	68	35.05
6	306	3	5	60	221	85	27.78
7	306	3	5	60	225	81	26.47
8	306	3	5	60	223	83	27.12
9	194	4	4	75	121	73	37.63
10	397	4	4	45	314	83	20.91
11	194	4	6	45	154	40	20.62
12	399	2	6	75	301	98	24.56
13	306	3	5	60	227	79	25.82
14	194	2	6	75	131	63	32.47
15	194	2	4	45	152	42	21.65
16	397	4	4	75	272	125	31.49
17	397	2	4	45	325	72	18.14
18	306	3	5	60	224	82	26.8
19	194	4	4	45	151	43	22.16
20	397	2	6	45	339	58	14.61
21	306	3	5	60	222	84	27.45
22	194	2	4	75	133	61	31.44
23	194	2	6	45	153	41	21.13
24	306	3	5	60	226	80	26.14
25	306	3	5	30	276	30	9.8
26	306	3	5	90	198	93	35.29
27	306	3	5	60	223	83	27.12
28	306	3	5	60	225	81	26.47
29	306	3	5	60	224	82	26.8
30	306	3	5	60	221	85	27.78
31	306	3	5	60	222	84	27.45
32	306	3	5	60	223	83	27.12

* Los datos fueron obtenidos de: Experimental dataset on adsorption of Arsenic from aqueous solution using Chitosan extracted from shrimp waste; optimization by response surface methodology with central composite design. Dehghani, M. H., Maroosi, M., & Heidarinejad, Z. (2018). *Data in Brief*, *20*, 1415-1421. https://doi.org/https://doi.org/10.1016/j.dib.2018.09.003

Referencias

1. Choi, H.; Park, S.-K.; Kim, D.-S.; Kim, M. Determination of 6 Arsenic Species Present in Seaweed by Solvent Extraction, Clean-up, and LC-ICP/MS. *Food Sci. Biotechnol.* **2011**, *20*, 39–44, doi:10.1007/s10068-011-0006-9.
2. Kwok, K.C.M.; Koong, L.F.; Chen, G.; McKay, G. Mechanism of Arsenic Removal Using Chitosan and Nanochitosan. *J. Colloid Interface Sci.* **2014**, *416*, 1–10, doi:10.1016/j.jcis.2013.10.031.
3. Wang, C.; Liu, H.; Zhang, Y.; Zou, C.; Anthony, E.J. Review of Arsenic Behavior during Coal Combustion: Volatilization, Transformation, Emission and Removal Technologies. *Prog. Energy Combust. Sci.* **2018**, *68*, 1–28, doi:10.1016/j.pecs.2018.04.001.
4. Horzum, N.; Demir, M.M.; Nairat, M.; Shahwan, T. Chitosan Fiber-Supported Zero-Valent Iron Nanoparticles as a Novel Sorbent for Sequestration of Inorganic Arsenic. *RSC Adv.* **2013**, *3*, 7828, doi:10.1039/c3ra23454a.
5. Moradi, M.; Fazlzadehdavil, M.; Pirsaheb, M.; Mansouri, Y.; Khosravi, T.; Sharafi, K. Response Surface Methodology (RSM) and Its Application for Optimization of Ammonium Ions Removal from Aqueous Solutions by Pumice as a Natural and Low Cost Adsorbent. *Archives of Environmental Protection* **2016**, *42*, 33–43, doi:10.1515/aep-2016-0018.
6. Myers, R.H..; Montgomery, D.C..; Anderson-Cook, C.M. *Response Surface Methodology : Process and Product Optimization Using Designed Experiments*; 3rd ed.; Wiley, 2009; ISBN 9780470174463.
7. Dehghani, M.H.; Maroosi, M.; Heidarinejad, Z. Experimental Dataset on Adsorption of Arsenic from Aqueous Solution Using Chitosan Extracted from Shrimp Waste; Optimization by Response Surface Methodology with Central Composite Design. *Data Brief* **2018**, *20*, 1415–1421, doi:10.1016/j.dib.2018.09.003.
8. Du, J.; Li, Y.; Jin, G.; Wang, J.; Chen, L. Determination and Correlation of Solubility of 2,4,6,8,10,12-Hexacyclopropyl-2,4,6,8,10,12-Hexaazaisowurtzitane in 12 Pure Solvents at Temperatures Ranging from 278.15K to 318.15K. *J. Chem. Eng. Data* **2025**, *70*, 4302–4314, doi:10.1021/acs.jced.5c00434.
9. Imame, S.; Atlas, H.; Sadoq, M.; Loukili, K.; Legsair, M.; El Bakkali, M.K.; Boukhlifi**, F. Utilization of Maize Pith Powder as a Cost-Effective Adsorbent for the Removal and Analysis of Crystal Violet from Aqueous Solutions before Determination by UV-Vis Spectrophotometry. *Analytical Methods in Environmental Chemistry Journal* **2025**, *8*, 5–25, doi:10.24200/amecj.v8.i02.1010.

TEMPERATURE PREDICTION IN A MINI SPLIT AIR CONDITIONING SYSTEM OF 3.5 KW USING MULTIPLE LINEAR REGRESSION

A. Cortés-García [1] y F. A. Alaffita-Hernández[2]

[1] Centro de Recursos Energéticos y Sustentables, Universidad Veracruzana, Av. Universidad km 7.5 Col. Santa Isabel, C.P.96535, Coatzacoalcos, Veracruz, México; zs24019639@estudiantes.uv.mx
[2] Centro de Recursos Energéticos y Sustentables, Universidad Veracruzana, Av. Universidad km 7.5 Col. Santa Isabel, C.P.96535, Coatzacoalcos, Veracruz,
* Correspondence: zs24019639@estudiantes.uv.mx.

Abstract

A polynomial was developed using multiple linear regression technique to predict final indoor air temperature at the outlet, of a 3.5 kW Mini split air conditioning system that uses R410A as refrigerant. The model was based on an experiment that utilized 21 sensors to measure variables such as temperature, pressure, humidity, power at different points and components of the system. Three input control variables were held constant throughout the experiment, which was conducted in a tropical weather condition of Barranquilla, Colombia. Results showed a coefficient of determination r^2 of 0.91, demonstrating model´s high predictive capability

Keywords: Multiple Linear Regression, HVAC Systems, Temperature Prediction.

1. Introduction

Study of HVAC systems (Heating, Ventilation and Air Conditioning) has gained relevance because they represent 30% to 40% of global energy consumption [1], [2], [3]. Energy management and energy consumption have become crucial for developed countries [4], this aligned with the 2030 sustainable development goals, that seek to reduce the carbon footprint and mitigate the effect of climate change.

Another area of interest in study of HVAC systems is comfort and indoor air quality in buildings and residential homes; this has caused an increase in the global demand of air conditioning [5].

HVAC Systems are typically modeled using three approaches, the one defined as white box [6] which consists of using the laws of physics to describe the process, the second type called gray box [7] which is made up of physical models but with parameters calculated by algorithms and the black box model [8] where data is collected from the system under operating conditions and a mathematical relationship is found between the input and output variables (usually performance) using statistical method such as linear regression or artificial neural networks(ANN).

Although efforts have been made to change energy use in buildings, such as use of new materials, selection of HVAC equipment type, and operation mode [9], energy consumption depends mainly on its control, maintenance, and use by occupants [10], [11], so it is important to develop statistical and non-statistical models for predicting indoor temperature in order to develop better control systems. In turn, these models must contain a greater number of input variables of internal and external to increase the accuracy in estimating the inferred variable [12], [13], [14].

Polynomial regression is a statistical tool that allows for the creation of a simple, fast, and precise empirical model to estimate a variable of interest, reducing computational costs in control system simulations and real-time estimations [15][16]17][18].

Objective of this work was to develop a polynomial using multiple linear regression tool to predict final temperature of a Mini split type of HVAC system operating in Colombia's weather.

Independent variables with the greatest impact on final temperature were identified, polynomial's fit was assessed using coefficient of determination r^2 ,and residuals were analyzed, additionally surface and contour plots were generated to visualize relationship between predicted and experimental temperature.

2. Materials and Methods

In this study, we predicted indoor temperature at the evaporator outlet, considering final temperature of a 3.5 kW air conditioning system operating with R410A as refrigerant.

Dataset were obtained from Ramírez-León et al., 2023 [5], where effects of inlet temperature and humidity, and fan speed were evaluated through 16 interactions scenarios.

Table 1 and Figure 1 show the instruments and their distribution in the HVAC system.

Table 1. Variables measured in the experiment [5].

Diagram Reference	Description
T1	*Refrigerant temperature at the compressor inlet in °C*
T2	*Compressor surface temperature in °C*
T3	*Refrigerant temperature at the compressor outlet in °C*
T4	*Outside air temperature at the condenser inlet in °C*
T5	*Outside air temperature at the condenser outlet in °C*
T6	*Coolant temperature in the condenser in °C*
T7	*Refrigerant temperature at the condenser outlet in °C*
T8	*Capillary tube surface temperature in °C*
T9	*Indoor air temperature at the evaporator inlet in °C*
T10	*Air temperature inside the evaporator outlet in °C*
P1	*Refrigerant pressure at the compressor inlet in PSI*
P2	*Refrigerant pressure at the condenser outlet in PSI*
P3	*Refrigerant pressure at the condenser outlet in PSI*
P4	*Refrigerant pressure at DX outlet in PSI*

H1	*Indoor air relative humidity at air hood in %*
H2	*Relative humidity of the indoor air at the evaporator outlet in %*
H3	*Relative humidity of the indoor air at the evaporator inlet in %*
I1/ PO_1	*Compressor current consumption in A/Compressor voltage in V*
I2 /PO_2	*Condenser fan current draw in A/Condenser fan voltage in V*
I3/ PO_3	*Evaporator fan current draw in A/Evaporator fan voltage in V*
F1	*Flow rate in the refrigerant liquid flowmeter at the condenser outlet in mL/min*

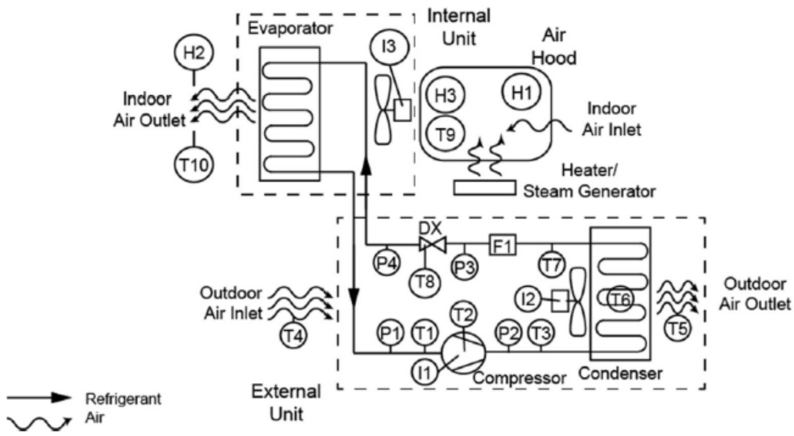


Figure 1. Mini-split diagram [5]

All data from 16 iterations were used to propose the following model according to equation (1):

$$T_{10} = p(T_1, \dots . T_{10}, P_1, \dots P_4, H_1, \dots H_3, PO_1, \dots PO_3, F_1) + \epsilon \qquad (1)$$

Where p is a polynomial ϵ the error, we assumed to follow standard normal distribution.

Model residuals were determined; residuals defined as difference between observed values and estimated values. were analyzed with two criterions, the three-sigma rule and the 95% confidence interval criterion that were used to identify outlier (atypical values) which were subsequently removed. A least squares regression was then reapplied to polish polynomial.

Polynomial performance was evaluated by determinate the coefficient of determination r^2 , additionally surface levels and contour lines plots were generated to visualize relationship between predicted and experimental temperature T10 and independent variables.

3. Results

When applying multiple linear regression the polynomial has an r^2 0.9153 which indicates that the model explains at least 91% of the variability of the internal temperature at the outlet of condenser T_{10} [19],[20], [21], [22] by removing outliers the parameter $r^2 = 0.9576$ increased, and the polynomial is according to equation (2):

$$T_{10} = -60.7028 + 1.8178 * T_1 + 1.7566 * T_4 + 1.3654 * T_7 + 1.0400 * T_9 + 0.0238 * H_3 + \ -1.0681 * P_1 + \ -0.1885 * T_5 + \ -0.5032 * T_8 - 0.0662 * T_1 * T_9 + \ -0.0038 * T_4 * T_7 + 0.0018 * T_7 * P_4 + 0.0247 * T_9 * T_8 + 0.0104 * T_1 * T_2 \quad (2)$$

It indicates that the temperature T_{10} only depends on variables T_1, T_4, T_7, T_9, H_3, P_1, T_5, T_8, which are the significant terms for the correlation coefficient, we can infer that the most important variables determining the final temperature delivered by the mini-split indoors depend on the conditions of the outside air (T_4, T_5, T_9, H_3,), refrigerant (T_1, T_7, P_1) and capillary tube (T_8). Polynomial included:

- ✓ Linear terms
- ✓ Iteration terms between variables

This structure allows capturing both linear and nonlinear behavior of the variables with respect to T_{10}

In Figure *2,* we are comparing the experimental T_{10} values against estimated by polynomial, showing that most of data points fit well with regression model

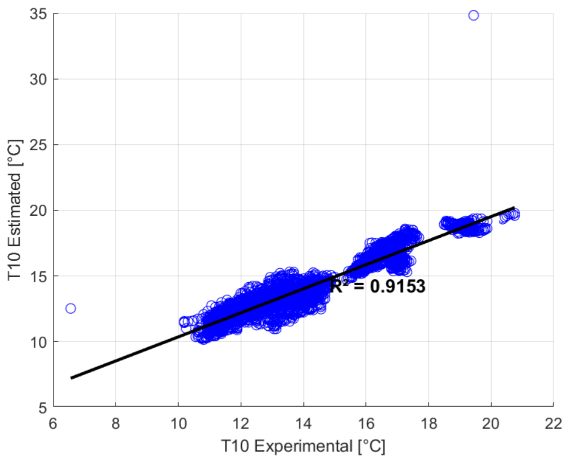


Figure 2. Comparing estimated versus experimental value.

Contour lines were generated to compare value of T10 both experimental value and one predicted by polynomial against independent variables, as shown in next figures Figure 3, Figure 4, Figure 5 y Figure 6, we can notice a close match between experimental and predicted values, demonstrating a good fit of polynomial.

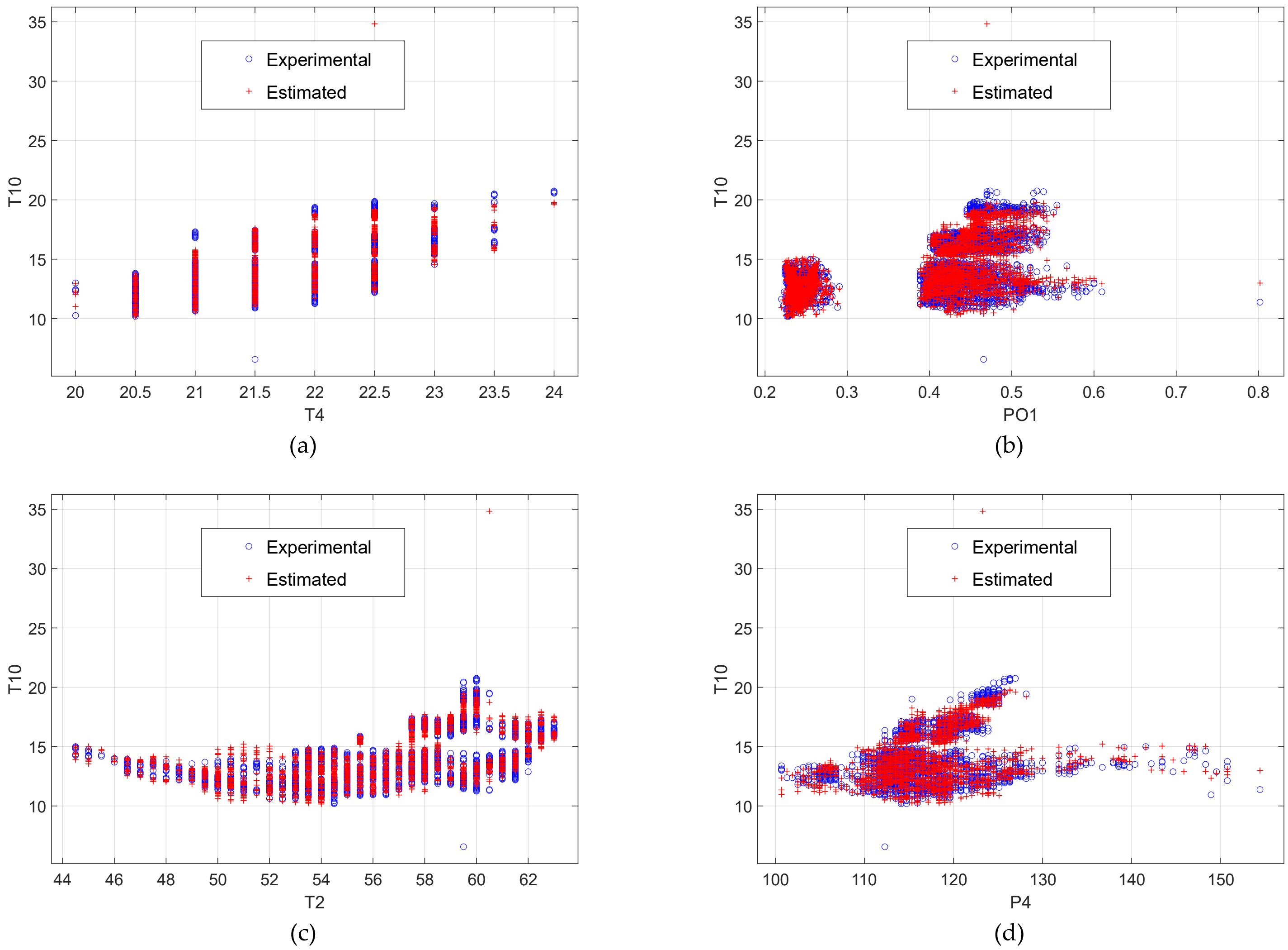
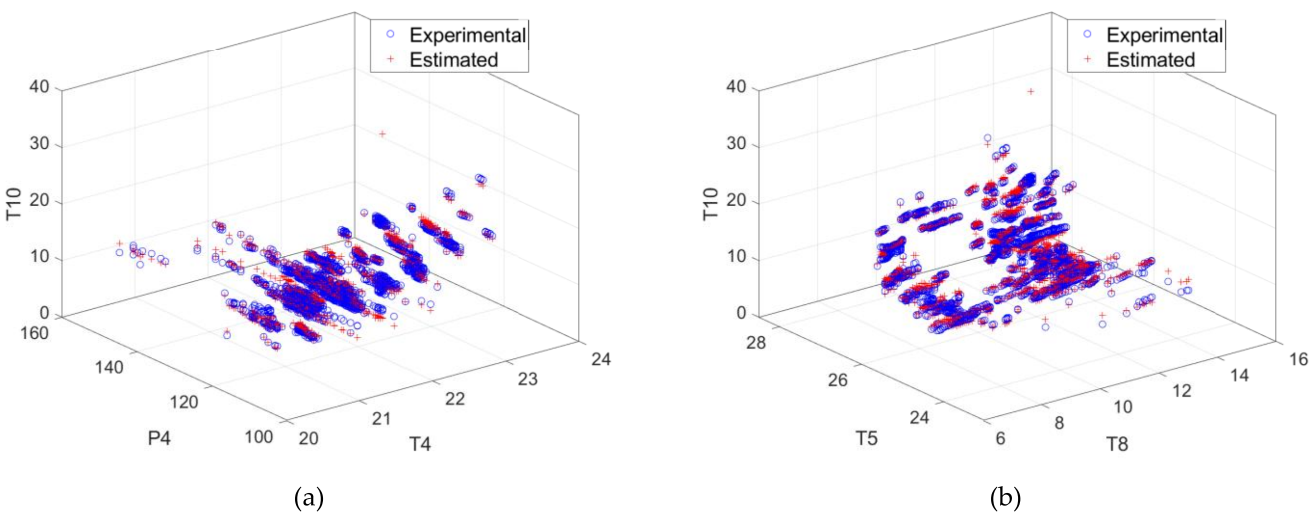


Figure 3. Contour lines: (**a**) *T10 & T4*; (**b**) *T10 & PO1*; (**c**) *T10 & T2;* (**d**) *T10 & P4.*

Level surfaces are present in figures: Figure 7, Figure 8, Figure 9, Figure 10, Figure 11, Figure 12, Figure 13 y Figure 14, in which the predicted surface can be observed closely follows the shape of experimental data, indicating that polynomial captures the overall behavior of the system effectively.



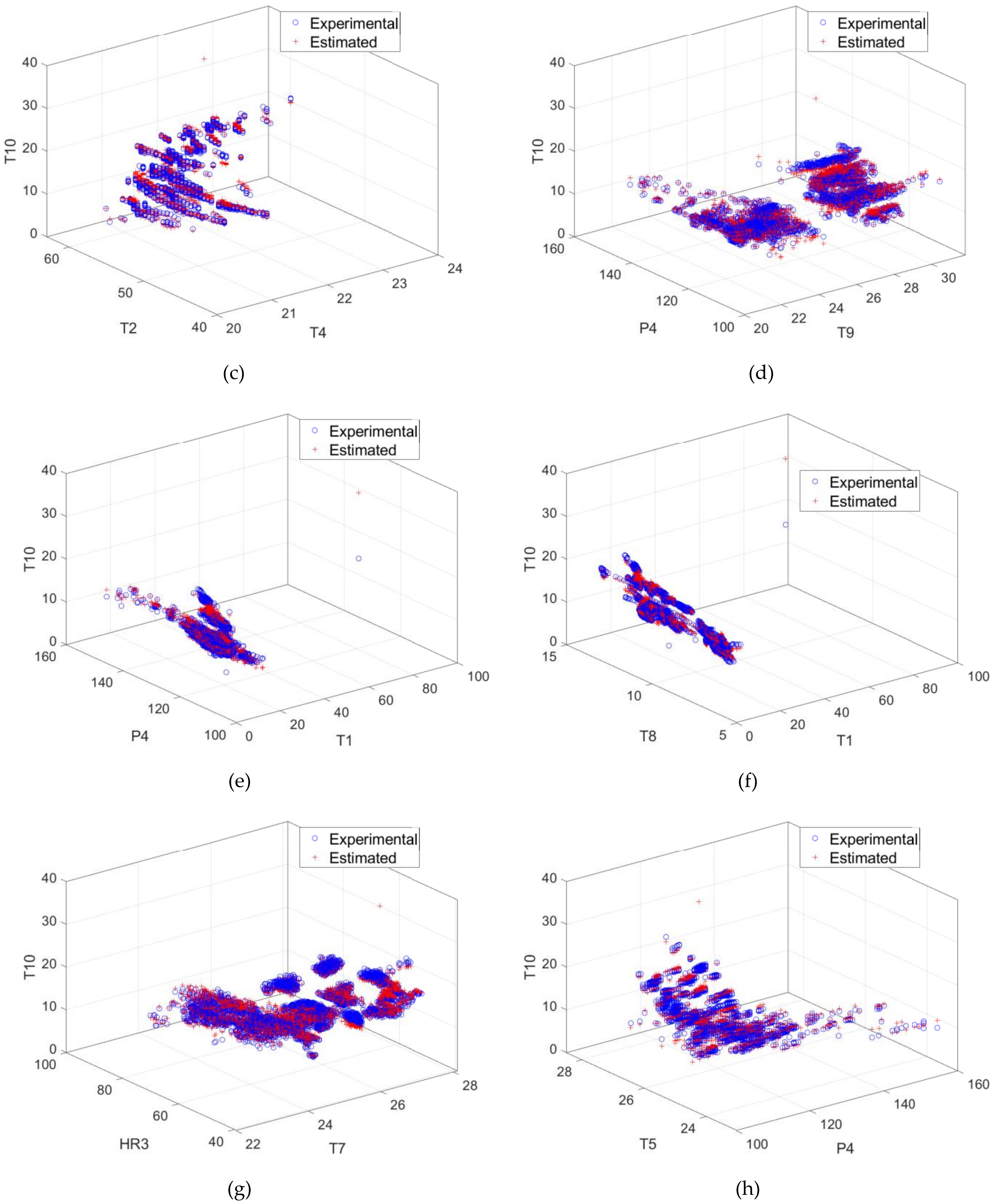
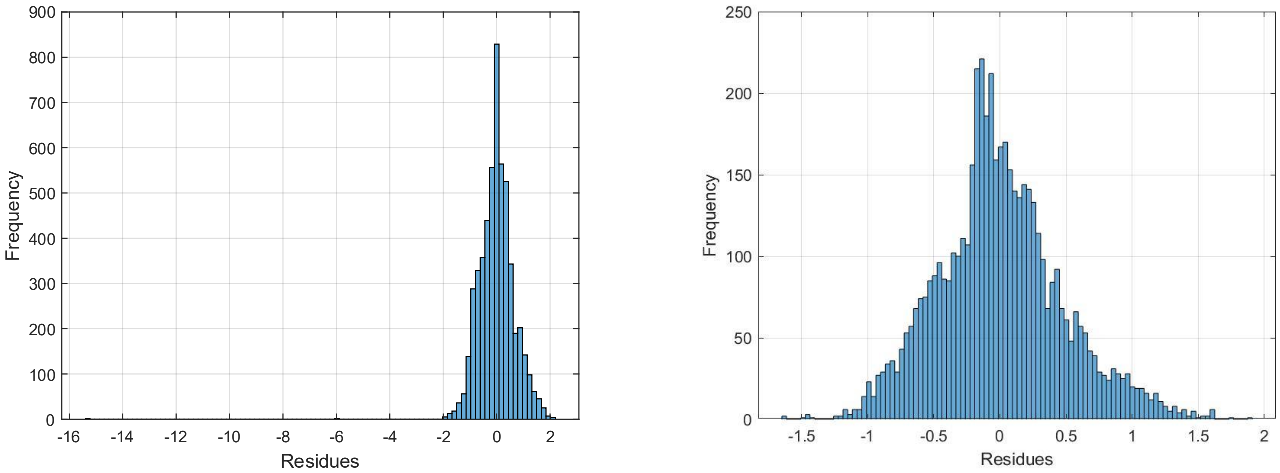


Figure 4. Level surfaces: (**a)** *T10 against P4 & T4*; (**b**) *T10 against T5 & T8*; (**c**) *T10 against T2 & T4 ;* (**d**) *T10 against P4 & T9;* (**d**) *T10 against P4 & T9;* (**e**) *T10 against P4 & T1;* (**f**) .

Finally, to check the assumption of random error, the residual histogram was graphed in Figure *14* with outliers in one we can observe a normal distribution with peak centered in zero with a peak in this value in *figure 15* show that distribution of residues continue in a shape bell gauss before outliers was eliminated, so we can affirm that errors of the model are small and symmetries

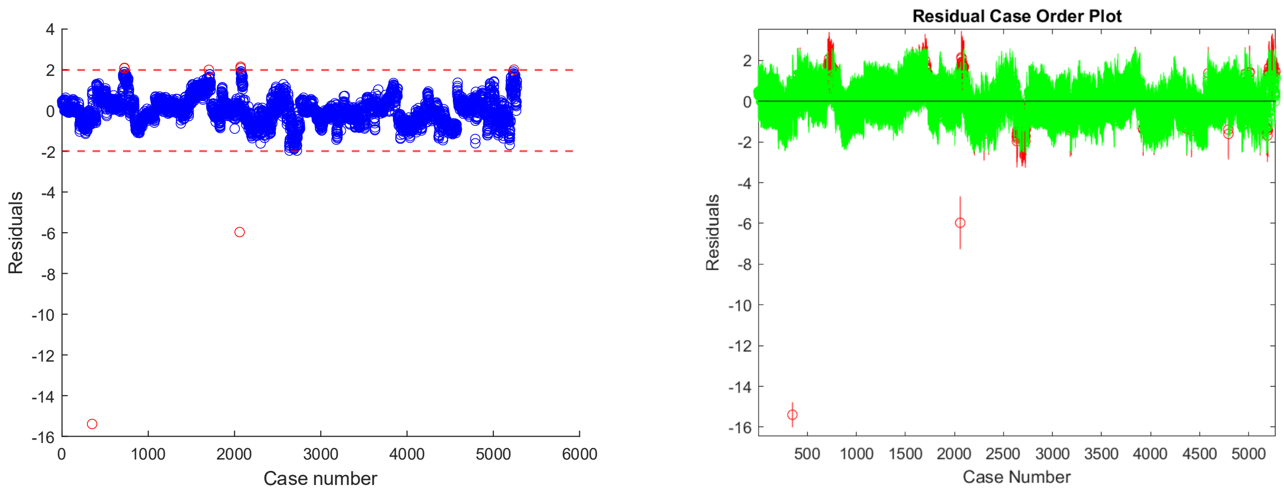


(a) (b)

Figure 5. (**a**) *Histogram with outliers*; (**b**) *Histogram without outliers*.

We also constructed a residual plot shown in *Figure 17* based on the 3$\boldsymbol{\sigma}$ rule criterion that covers at least 97% of the sample and some outliers are observed

In *Figure 18,* we analyze the residuals with the 95% confidence interval criterion in what we use the function *RCOPLOT* of *MATLAB*. We can observe a mayor number of outliers, that's indicates that observations exist in our data out of the range that won't fit to our model, maybe due to error in measurement.



(a) (b)

Figure 6. (**a**) *Sigma Rule Residues*; (**b**) *Residues with a 95% Confidence Interval*.

5. Conclusions

This work shows that indoor temperature is strongly influenced by outdoor environmental air conditions (T4, T5, T9 and HR3), as well internal thermodynamic conditions that involve equipment like refrigerant (T1 , T7), compressor (PO1) and Capillary Tube(T8). These findings can be applied in futures studies of mini split systems exploring alternatives capillary tube designs or heat exchanger configurations and in fault detection methods for predictive maintenance.

Polynomial obtained a good fit to experimentally values. In addition, reducing number of variables from 21 to 8, it can explain up to 91% of variation in T10, furthermore polynomial incorporates not only linear terms but also interaction terms between variables to enhance predictive accuracy. Through residual analysis and the generation of surface and contour plots, model's adequacy has been effectively assessed. Result is a simple but efficient polynomial, based on real data, that holds potential for use in more complex models for optimization, control, real-time prediction, or energy management in HVAC systems.

5. References

[1] Y. Li, W. Wang, Y. Wang, Y. Xin, T. He, and G. Zhao, "A review of studies involving the effects of climate change on the energy consumption for building heating and cooling," Jan. 01, 2021, MDPI AG. doi: 10.3390/ijerph18010040.

[2] L. Pérez-Lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information," Energy Build, vol. 40, no. 3, pp. 394–398, 2008, doi: 10.1016/j.enbuild.2007.03.007.

[3] A. Foucquier, S. Robert, F. Suard, L. Stéphan, and A. Jay, "State of the art in building modelling and energy performances prediction: A review," 2013. doi: 10.1016/j.rser.2013.03.004.

[4] E. E. O. Opoku, N. K. Kufuor, and S. A. Manu, "Gender, electricity access, renewable energy consumption and energy efficiency," Technol Forecast Soc Change, vol. 173, Dec. 2021, doi: 10.1016/j.techfore.2021.121121.

[5] H. Ramirez-Leon, M. A. Gonzalez-Carmona, J. Jiménez-Cabas, J. Escorcia-Gutierrez, and A. Bula, "Experimental air conditioning energy evaluation under Caribbean climate conditions," Heliyon, vol. 9, no. 3, Mar. 2023, doi: 10.1016/j.heliyon.2023.e14464.

[6] R. Z. Homod, "Review on the HVAC System Modeling Types and the Shortcomings of Their Application," Journal of Energy, vol. 2013, pp. 1–10, 2013, doi: 10.1155/2013/768632.

[7] A. Afram and F. Janabi-Sharifi, "Gray-box modeling and validation of residential HVAC system for control system design," Appl Energy, vol. 137, pp. 134–150, Jan. 2015, doi: 10.1016/j.apenergy.2014.10.026.

[8] A. Afram and F. Janabi-Sharifi, "Black-box modeling of residential HVAC system and comparison of gray-box and black-box modeling methods," Energy Build, vol. 94, pp. 121–149, May 2015, doi: 10.1016/j.enbuild.2015.02.045.

[9] J. Hossain et al., "A Review on Optimal Energy Management in Commercial Buildings," Feb. 01, 2023, MDPI. doi: 10.3390/en16041609.

[10] G. Fajilla, M. De Simone, L. F. Cabeza, and L. Bragança, "Assessment of the impact of Occupants' Behavior and climate change on heating and cooling energy needs of buildings," Energies (Basel), vol. 13, no. 23, Dec. 2020, doi: 10.3390/en13236468.

[11] Y. He, Y. Chen, Z. Chen, Z. Deng, and Y. Yuan, "Impacts of Occupant Behavior on Building Energy Consumption and Energy Savings Analysis of Upgrading ASHRAE 90.1 Energy Efficiency Standards," Buildings, vol. 12, no. 8, Aug. 2022, doi: 10.3390/buildings12081108.

[12] P. Norouzi, S. Maalej, and R. Mora, "Applicability of Deep Learning Algorithms for Predicting Indoor Temperatures: Towards the Development of Digital Twin HVAC Systems," Buildings, vol. 13, no. 6, Jun. 2023, doi: 10.3390/buildings13061542.

[13] K. Huang, K. P. Hallinan, R. Lou, A. Alanezi, S. Alshatshati, and Q. Sun, "Self-learning algorithm to predict indoor temperature and cooling demand from smart wifi thermostat in a residential building," Sustainability (Switzerland), vol. 12, no. 17, Sep. 2020, doi: 10.3390/su12177110.

[14] F. Zamora-Martínez, P. Romeu, P. Botella-Rocamora, and J. Pardo, “Towards energy efficiency: Forecasting indoor temperature via multivariate analysis,” Energies (Basel), vol. 6, no. 9, pp. 4639–4659, 2013, doi: 10.3390/en6094639.

[15] H. Yang, M. Ran, and C. Zhuang, “Prediction of Building Electricity Consumption Based on Joinpoint−Multiple Linear Regression,” Energies (Basel), vol. 15, no. 22, Nov. 2022, doi: 10.3390/en15228543.

[16] I. Ridwana, N. Nassif, and W. Choi, “Modeling of building energy consumption by integrating regression analysis and artificial neural network with data classification,” Buildings, vol. 10, no. 11, pp. 1–14, Nov. 2020, doi: 10.3390/buildings10110198.

[17] S. D. Atalay, G. Calis, G. Kus, and M. Kuru, “Performance analyses of statistical approaches for modeling electricity consumption of a commercial building in France,” Energy Build, vol. 195, pp. 82–92, Jul. 2019, doi: 10.1016/j.enbuild.2019.04.035.

[18] H. Sha, P. Xu, C. Hu, Z. Li, Y. Chen, and Z. Chen, “A simplified HVAC energy prediction method based on degree-day,” Sustain Cities Soc, vol. 51, Nov. 2019, doi: 10.1016/j.scs.2019.101698.

[19] J. C. Lam, K. K. W. Wan, D. Liu, and C. L. Tsang, “Multiple regression models for energy use in air-conditioned office buildings in different climates,” Energy Convers Manag, vol. 51, no. 12, pp. 2692–2697, Dec. 2010, doi: 10.1016/j.enconman.2010.06.004.

[20] B. Chi, Y. Li, and D. Zhou, “A Hybrid Method of Cooling and Heating Consumption Prediction for Six Types of Buildings Based on Machine Learning,” Sustainability (Switzerland), vol. 16, no. 24, Dec. 2024, doi: 10.3390/su162411200.

.