International Journal of Artificial Intelligence,
ISSN 0974-0635; Int. J. Artif. Intell.
2025, Volume 23, Issue Number 2

Optimization of Wind Energy Planning Through Data Analysis

M. Ríos-Palacio[1], P. A. Sánchez-Sánchez[2], J. R. García-González[3] and **A. Berrones[4]**

[1]Facultad de ingenierías, Universidad Simón Bolívar, Barranquilla, Colombia;
Email: mauricio.rios@unisimon.edu.co

[2]Facultad de ingenierías, Universidad Simón Bolívar, Barranquilla, Colombia;
Email: paola.sanchezs@unisimon.edu.co

[3]Facultad de ingenierías, Universidad Simón Bolívar, Barranquilla, Colombia;
Email: jose.garciag@unisimon.edu.co

[4]School of Mechanical and Electrical Engineering, Universidad Autónoma de Nuevo León, México;
Email: arturo.berronessn@uanl.edu.mx ;

ABSTRACT

Objective: This article presents the implementation of a model for wind energy production planning, based on the hybridization of two components: one focused on data analysis for wind speed prediction, and the other on the optimization of energy production. Method: A hybrid strategy was methodologically structured to enable the effective and practical application of data analysis and optimization techniques across various planning scenarios. Results: The proposed model was evaluated in wind energy planning processes across 13 stations in the metropolitan area of Monterrey, Mexico. Three different budget scenarios were considered to determine optimized calculations of the types and quantities of wind turbines to be installed in specific areas. Conclusions: The proposed hybrid approach leverages advanced data analysis and optimization techniques within a unified framework to enhance wind energy planning. The results achieve energy generation capacities comparable to large-scale wind farms worldwide, positioning the model as a promising alternative for substantial environmental improvement and regional sustainability.

Keywords: Wind Energy Production, Hybrid Model, Wind Speed Prediction, Optimization, Renewable Energy Planning

Mathematics Subject Classification: 68T09, 90C08

Computing Classification System: I.2.6, G.1.6, J.2

1. INTRODUCTION

In a world increasingly shaped by globalization and technological development, the processes associated with energy transformation pose significant challenges for contemporary societies, whose functioning depends heavily on the continuous supply of electricity.

Given the environmental impact of traditional power generation systems, there is a growing urgency to transition toward sustainable models that incorporate renewable energy sources. In this context, wind energy has emerged as a viable and expanding alternative, both in terms of efficiency and reduction of pollutant emissions.

Today's societal and economic dynamics rely heavily on electricity consumption, which is essential for the production and delivery of goods and services. Traditionally, this electricity is generated through the combustion of fossil fuels, which presents two major issues: first, these are non-renewable sources and thus limited in availability over time; and second, they are the leading contributors to global greenhouse gas emissions.

In response to these societal needs, the energy sector has begun shifting toward more sustainable alternatives based on renewable resources. A prominent example is wind energy, which is generated from wind movement, typically using wind turbines or windmills (Banuelos, 2011), (Ackermann & Soder, 2002).

According to the Global Wind Energy Council (GWEC), in 2021 there was a 53% increase in the installed capacity, producing a total of 743 GW of wind energy worldwide. Environmentally, this corresponds to 1100 million tons of CO2 less emissions. Nevertheless, experts are concerned because this growth is insufficient to guarantee the objectives set for climate change and global warming for 2050. For such purposes, the world needs to triple installed wind energy capacities in the next decade (GWEC, 2021).

Despite the high degree of implementation of renewable energy worldwide, new uncertainties and doubts arise, such as whether the amount of energy produced by new sources will be sufficient to cover present needs, the possibility of reducing the consumption of devices connected to the electricity grid, and the feasibility of improving new energy sources. Wind energy raises its own challenges, such as the selection of wind turbines, the accessibility to the geographical area where windmills will be deployed, the best cost-benefit ratio, and the registered wind speed (Blaschke et al., 2013) (Hernández, et al., 2021).

To address these issues, diverse models, software, and techniques based on data analysis have been developed (Mamyrbayev, et al., 2025). The aim is to predict, model and optimize the launching and operation of wind turbines and find optimized parameters so that the best equipment can be chosen depending on the geographical, technological, and economic characteristics available and the energy production needed.

The reliability of the prediction was strengthened by applying Big Data technologies. They are regarded as the most advanced for large-scale data analysis due to their enormous processing capacity resulting from the implementation of massive information sets. It is usually a very complex process to analyses such large sets with traditional systems and algorithms. Additionally, as elements such as wind speed, flow and direction differ over time due to their own nature, they entail an implicit complexity in their behaviors while maximizing electricity generation (Gandoni & Haider, 2015) (Sánchez-Sánchez & García-González, 2018).

In accordance with this, this article describes a theoretical characterization of wind turbine power generation, followed by a proposal for a hybrid model of data analysis and optimization based on information collected by sensors set up in the metropolitan areas of Monterrey (Mexico). The aim of this work is to find a solution to maximize the amount of electricity and minimize production costs.

2. FEATURES AND CURRENT TRENDS OF POWER SYSTEMS

Wind energy is generated by the kinetic energy of air masses. It can be converted into mechanical energy and, subsequently, into electricity or other useful forms of energy for human needs.

Wind energy has been used by humanity for over 3,000 years, mainly in mill, pump and as a ship-driving force. In 1891, Danish engineer Poul LaCour produced the first wind turbine to generate electricity. Based on his design, Brush, an American company, developed the first commercial wind turbine (Talayero & Martínez, 2011). Since then, till date, technological advances in wind turbines continued, improving the aerodynamics and developing wind turbines with increased dimensions that generate more power. The production capacity of turbines increased from 12 kW to 5000 kW in around 60 years of evolution (Ackermann & Soder, 2002).

A wind turbine is an electro-mechanical device that can transform the kinetic energy of wind into electrical energy. At the shaft, by means of the wind rotor, the aerodynamic wind power is converted into mechanical power. This rotor usually has two or three blades attached to the hub. Gearboxes are installed on the said hub. Electrical generators also convert mechanical energy into electrical power that is injected into the grid by a transformer. The alternator and the speed multiplier are housed in a nacelle, placed at the end of a clamping tower (Lopez, 2011). A wind turbine might include an anemometer if the wind speed is excessive, which will supply information to control systems, preventing excessive mechanical stress, and a vein to guide the nacelle toward the wind direction.

There are two types of generators, depending on how and where electricity is generated: the first is the onshore wind energy production model, and the second is the offshore model, where wind turbines are installed at sea. In this study, our focus is on high power (>2 MW) onshore wind turbines, with a horizontal axis, three blades directed in a windward direction and a tubular tower.

Horizontal wind turbines have high energy efficiency and rotation speed. Thus, they need gearboxes with low rotation multiplication ratios. Because of their elevated tower height, they greatly benefit from the wind speed increase. Windward or "facing the wind" orientations mean that the rotor is facing the dominant wind direction, reaching a better use of the wind force (Garrido, 2021).

A wind turbine obtains its input power transforming the force of the wind current (turning force) and acting on its rotor blades. Air density, rotor swept area, and wind speed affect the amount of energy transferred to rotors by the wind. The amount of wind energy that a wind turbine can capture is determined by the rotor area. The area, which is based on the diameter, is determined by its diameter, as well as power generated. For instance, wind turbines with diameters of 27 meters generate approximately 225 kW, while those with diameters of 80 meters generate over 2500 kW (Giner, 2013). The kinetic energy of an air mass m moving at speed v is given by:

$$E_c = \frac{mv^2}{2} \quad \text{(Eq1)}$$

Since air mass is the product of its volume V and density D ($m = V \cdot D$), the kinetic energy becomes:

$$E_c = \frac{V \cdot D \cdot v^2}{2} \quad \text{(Eq2)}$$

For a wind turbine, the air volume passing through the rotor in time t depends on the swept area A and the wind speed v, so $V = A \cdot v \cdot t$. Thus, the kinetic energy in time t is:

$$E_c = \frac{A \cdot D \cdot v^3 \cdot t}{2} \quad \text{(Eq3)}$$

To work with power instead of energy, we consider energy per unit of time, obtaining the wind available to the rotor:

$$P = \frac{E_c}{t} = \frac{A \cdot D \cdot v^3}{2} \quad \text{(Eq4)}$$

3. PROPOSED METHODOLOGY

Methodologically, the proposal presented in this article is structured based on a systemic strategy for the specification and construction of a hybrid model for wind energy planning, whose definition is aligned with the theoretical framework for the development of research proposals and projects as outlined in (García-González &. Sánchez-Sánchez, 2020). The hybrid model includes the interaction of two components: a wind speed predictor that employs data analysis techniques, and an energy production optimizer. The methodological operationalization from a technical perspective is described below and is schematically illustrated in Figure 1. Proposal for a hybrid planning model for wind power.

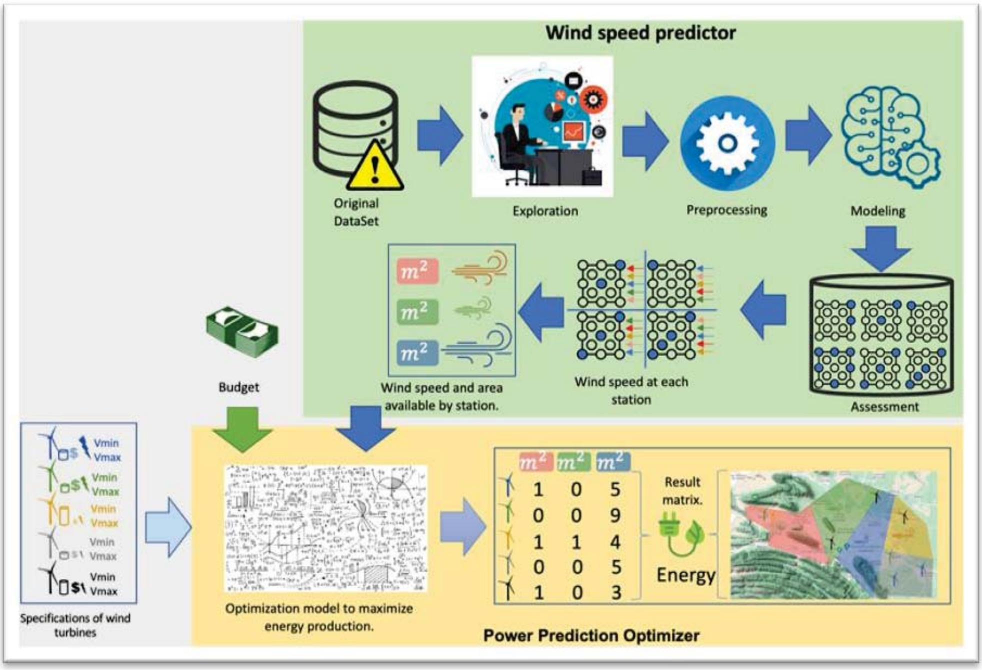

Figure 1. Proposal for a hybrid planning model for wind power

3.1. Wind speed predictor

The wind speed predictor receives information from the Integrated Environmental Monitoring System (SIMA), one of the essential inputs for the process. The original dataset undergoes data analysis processes such as exploration, preprocessing, modelling, and assessment so that the model can predict the wind speed with maximum accuracy. This occurs independently for each station (region).

In turn, wind speed predictions for each station feed the power production optimizers. This also requires information on available wind turbines and the budget; the optimizer overseas maximizing power production for each station.

This component receives, monitors, and analyses environmental information available for each system station, calculating, evaluating, and selecting the best model for wind speed prediction for each station and transferring such prediction to the energy production optimizer.

The aim of a prediction model is behavior (value) anticipation of a variable depending on the values of other explanatory variables based on known examples of previous predictions (Sánchez-Sánchez et al, 2019).

The first step to specify and develop the predictive model is the existence of a dataset with variables that influence wind speed calculation. Subsequent processes are described next (García-González et al., 2019) (Sánchez-Sánchez, et al., 2020):

- Exploration: The data exploration process is focused on a preliminary analysis made by the modeler with the data, its characteristics, and aspects related to the environment from which they are procured. This might be defined as an approximation process to the information available by means of subjective analysis and statistical treatment, which enables the inferring of the hidden data structure.
- Preprocessing: This refers to input data analysis and transformation with the objective of minimizing noise, emphasizing important relationships, and detecting errors to facilitate modelling and prediction. Three types of processes are included in the preprocessing performed in this study: I) a process aimed at noise minimization by converting target data and eliminating non-regular (atypical, poorly typed, blank, inconsistent, and incomplete data, among others) patterns; ii) a process of syntactic transformations of target data, to facilitate their operation, without changing the results; iii) a process focused on selecting variables, features or attributes to be included in the model.
- Modelling: The modelling process includes the use of different representations to detect hidden data behavior patterns and based on this, predict values of unknown instances [8]. This process applies machine learning techniques, such as linear and nonlinear regression models, neural networks, decision trees, and regularized regression algorithms, among others. This study compares the findings from models that were extensively tested in the literature and that offer excellent results: Decision Tree Regressor (Breiman et al, 1984), Elastic Net, Elastic NetCV (Zou & Hasti, 2005), Gradient Boosting Regressor (Friedman, 2001), KNeighbors Regressor (Cover & Hart, 1967), LASSO (Tibshirani, 1996), MLPR (Rumerlhart, 1986), Regresión Lineal Múltiple (Galton, 1894), Ridge Regression (Hoerl & Kennard, 1970), XGBoost Regression (Chen & Guestrin, 2016).
- Assessment: The quality of prediction models is assessed by analyzing and comparing the results of different measurements calculated by the modeming process. The comparison base is error criteria (differences among calculated or predicted values and actual values) to select the model with the lowest error. These criteria require dividing the data sample into at least two sets: adjustment or training, and test or forecast. The training set is applied to calculate the adequate model parameters, while the forecast set is the basis to verify model validity and usefulness. Regarding the training set, several error measures have traditionally been used to determine the adjustment of models to data. This study applies the mean square error (MSE) as a criterion to select the best model.

3.2. Power Prediction Optimizer

This component oversees the maximization of energy production by applying optimization models. Wind energy generation is characterized by limited physical and budgetary resources, and it relies on environmental conditions, which have their own dynamics. To validate optimization benefits, two models with diverse complexity degrees are implemented.

Among the physical resources considered are wind turbines, their references and power capacities, required installation areas, minimum and maximum speed limits admitted by each turbine, and maximum areas offered at each zone (station).

Budgetary investment resources considered in this research are exclusively limited to direct costs for the purchase of wind turbines, without including additional related costs such as land rental, turbine installation, grid and road construction, etc.

Environmental conditions are related to diverse zones (seasons) and impact the calculation of predictions of wind speed.

The optimization model answers the following questions:

- Which references of wind turbines should be bought?
- How many wind turbines for each reference should be bought?
- In which areas (stations) will wind turbines be installed?
- How much total wind energy is produced (what is the maximum power produced)?

Optimization Model – OPT: It considers M meteorological stations, which correspond to 13 different zones where sensors were installed and their respective wind speed (v_j^3) and areas ($area\_zone$) to install wind turbines were measured, along with the corresponding budget.

The objective function is defined as follows:

$$max(z) = \sum_{i=1}^{n} x_{(i,j)} \cdot \left(\frac{8 \cdot A_i \cdot D \cdot v_j^3}{2{,}7 \cdot 10^7} \right) \quad j = 1,2,3, \dots, m$$

Subject to:

$$\sum_{i=1}^{n} x_{(i,j)} \cdot a_i \leq area\_zone_j \quad j = 1,2,3, \dots, m$$

$$\sum_{i=1}^{n} x_{(i,j)} \cdot c_i \leq budget \quad j = 1,2,3, \dots, m$$

Where:

- A_i = Rotor area of the i-th wind turbine. (m^2)
- D = air density. ($1.255\, kg/m^3$)

- $x_{(i,j)}$ = Number of wind turbines to be purchased from an i-th reference that will be installed in a j-th zone (Station).
- v_j = Average annual wind speed for the j-th zone. (m/seg)
- n = Number of existing wind turbine references.
- i = Wind turbine reference index.
- a_i = Area occupied by an i-th wind turbine. (m^2)
- c_i = Cost of an i-th wind turbine.
- $area\_zonas_j$ = Available area for the j-th zone. (m^2)
- m = Number of zones (Stations).
- j = Zone index (Station).

The OPT model, described in Figure 2. OPT model scheme, is powered by wind speed predictions for each zone, developed by wind speed predictors, and subject to a different model for each zone, depending on their own conditions. It also considers different wind turbine references and the characteristics of their installed area capacity.

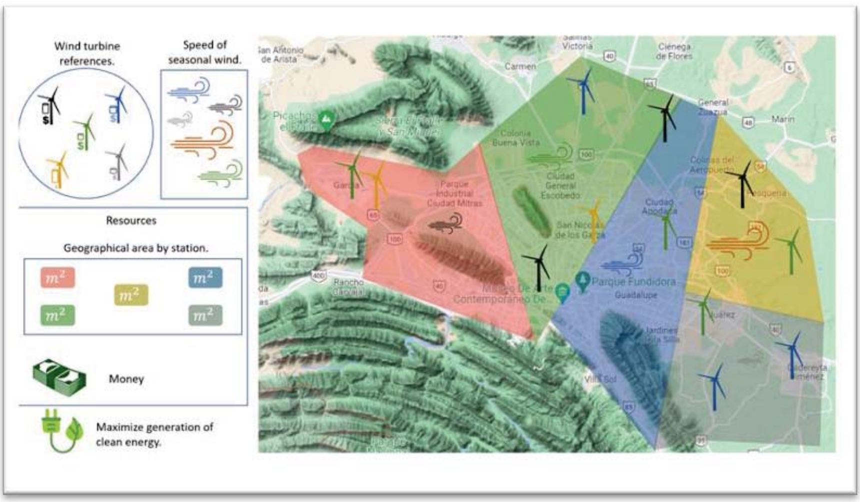

Figure 2: OPT model scheme. Source: Own draft

4. RESULTS AND DISCUSSION

Using the information provided by the Integrated Environmental Monitoring System (SIMA), an analysis of the dataset was conducted employing various statistical techniques and data visualization tools to determine its nature and structure.

Similarly, the dataset was refined through the application of filters and the completion of missing information where possible. This process enabled the subsequent formulation of the wind prediction model and the selection of appropriate wind turbines through the optimization model to maximize electricity generation in each of the regions (stations).

4.1. Specifications of the model

Method: The proposed systemic strategy is used to specify and develop a hybrid model for wind energy planning.

Dataset: The dataset has 1,542,696 records from the measurement of 16 variables as shown in Table 1. List of identified variables. In 13 monitoring stations in Monterrey (Mexico), hourly data (24/7) from 1993 to 2019 was registered and measured by the Integrated Environmental Monitoring System (SIMA). In Figure 3. Distribution of sensors by stations in Monterrey, Mexico, a map of sensor distribution for diverse stations in Monterrey region is shown.

Table 1: List of identified variables.

Description	Variable
Monitoring station	Station
Carbon monoxide level	CO
Nitric oxide level	NO
Nitrogen dioxide concentration	NO2
Nitrogen oxide levels	NOX
Ozone level	O3
Particulate matter in proportion ≤ 10 μm	PM10
Particulate matter in proportion ≤ 2.5 μm	PM 2.5
Atmospheric pressure	Pressure
Rainfall	Rainfall
Humidity level	Humidity
Sulfuric dioxide level	SO2
Solar radiation	Solar
Temperature	Temperature
Wind direction	Direction
Wind velocity (m/s)	Velocity

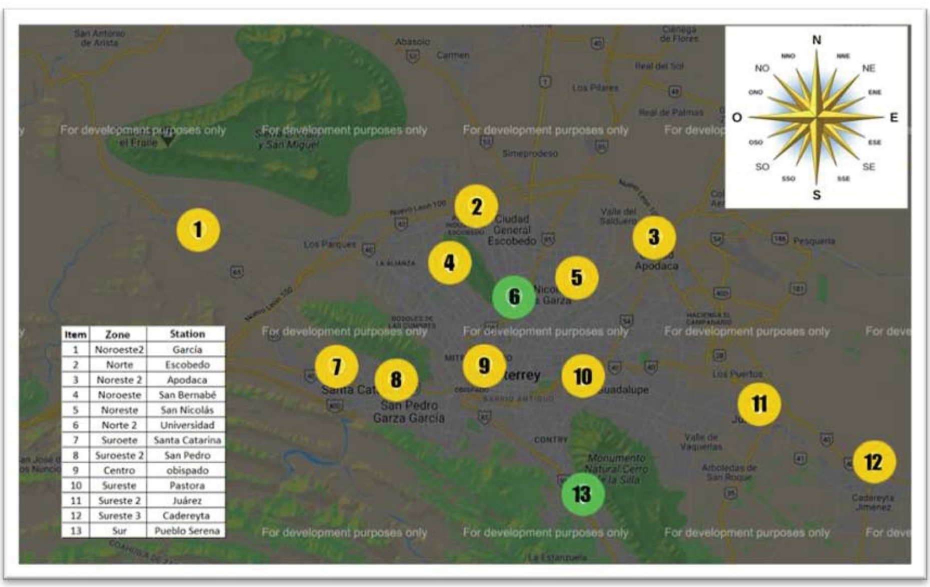

Figure 3. Distribution of sensors by stations in Monterrey, Mexico

4.2. Procedure: Exploration and preprocessing

The database compiled data on environmental and meteorological conditions from 13 different stations (zones) in Monterrey, Mexico. The first 15 variables of Table 1 are explanatory variables and the last one (velocity) is a variable to be predicted. The complete data set is subjected to noise elimination, data transformation, error detection, and variable selection. Data imputation techniques are used to complete null data detected in different variables. Tests are performed using complete data sets (with 15 variables) and applying a variable reduction process, with a recursive elimination of variables and the selection of the best number with cross-validation to eliminate variables with redundant information. To prepare to specify and apply optimization model 1, a general grouping of data is used. In optimization models 2 and 3, they are separated by stations.

4.3. Modelling and assessment

The pre-processed data are inputs for ten different regression models: Decision Tree Regressor, Elastic Net, Elastic Net CV, Gradient Boosting Regressor, KNeighbors Regressor, LASSO, MLPR, Multiple Linear Regression, Ridge Regression, XGBoost Regression. A Python script with the scikit-learn and pandas' libraries is used.

The configuration parameters of the models are the following:

- Input variables: 15 variables that are the first 15 are listed in Table 1. Separate calculations are made for each of the 13 stations.
- Output variable: Wind speed for each of the 13 stations.
- Error measurements: MSE
- Number of periods: 2000
- Number of iterations: 5
- The Decision Tree Regressor model of the Scikit-learn library is an optimized CART version (Classification and Regression Trees) by applying discrete numerical variables. An iterative process is executed to adjust the number of levels according to errors.
- The multiple linear regression applies a traditional additive regression model to 15 explanatory variables and the wind speed response variable.
- In LASSO automatic variable selection is carried out by applying the L1 prior method as a regularized and subsequently adjusted with the application of the coordinate descent model.
- In the Ridge Regression model, the resulting model is regularized by imposing a penalty on the coefficient size of the linear relationship between the explanatory and target variables. The loss function corresponds to the linear least squares function, and the regularization is calculated by L2 prior.
- The Elastic Net model is used to estimate a standardized linear regression that combines L1 and L2 priors to prevent an overadjustment.
- The Elastic Net CV model is used to calculate normalized linear regression. Unlike the Elastic Net, it includes an iterative adjustment by means of a regularization path via cross-validation. It automatically selects the best hyperparameters for model adjustment. A 10-fold model and a randomized restart were applied.

- The Gradient Boosting Regressor develops a generalized model from a set of smoothed decision tree models, with a maximum 4-level depth.
- The XGBoost Regression-extreme gradient boosting model executes a sequential set of decision trees that corrects errors produced, whose growth is performed up to their maximum extent (without a depth limit).
- With the KNeigbors Regressor, the optimal neighbour number is calculated based on their accuracy and the Euclidean uniform weight function applied as distance measures.
- The MLPR model is a multilayer perceptron neural network. The MLP is trained using Backpropagation, with four hidden layers of 200 neurons per layer.
- The dataset is divided into two sets in all cases: training and test sets, representing 80% and 20% of data, respectively.

The best prediction model for each station is selected by means of five performances for each model, changing the training dataset. The model with the lowest MSE is chosen among all such datasets.

In Table 2. Best models to predict seasonal wind speed, the optimum models selected for each station are listed. The associated wind speed prediction is calculated. A preliminary result review shows that the installation of wind turbines for stations in the west, both north and south, is favored, since wind speeds in these areas are greater.

Table 2: Best models to predict seasonal wind speed.

Best model	Station	Wind speed
CART	North 2	7.8
Elastic Net	North	7.41
Gradient Boosting Regressor	Northwest	9.32
	Northeast	8.18
	Southeast	8.82
LGB	Southwest 2	4.13
	Southeast 2	6.8
Multiple linear regression	Center	6.7
Ridge Regressor	South	2.41
XGB Regressor	Northeast 2	8.48
	Northwest 2	9.99
	Southwest	8.46
	Sureste3	2.61

4.4. Optimization

To develop OPT optimization models, the methodology listed in the Wind Energy Production Optimization section, whose inputs are wind speed predictions for each zone, is applied. The limit values of resources applied to execute optimization models are the following:

- Occupied area, using the criteria set forth by (EREN, 2015), which states that wind turbines should be linearly placed, following the peak profile, and directed based on wind conditions at a distance that is 2 and 4 times the blade (vane) diameter. Areas occupied by wind turbines are calculated with this formula:

$$area\_turbine = \pi \cdot \left(\frac{3 \cdot diameter\_vane}{2}\right)^2$$

- Wind turbine cost: An estimated value of US$1.75 million is set for each MW of power capacity. Therefore:

$$coste\_turbine = 1.75 \cdot power$$

- Budget: There are three scenarios for clean wind energy generation investment: Worst-case scenario: US$2 billion; Moderate scenario: US$4.05 billion; Optimistic scenario: US$6.1 billion.
- Available area: for each station, an available area capacity for the installation of wind turbines is set by the Comisión Federal de Electricidad – CFE (Mexico), as shown in Table 3. Area available for wind turbine installation.

Table 3: Area available for wind turbine installation

	Station	Available area (M2)
1	Center	1,500,000
2	Northeast	86,800,000
3	Noreste2	239,000,000
4	Northwest	110,000
5	Noroeste2	1,034,000,000
6	North	207,100,000
7	North 2	150,000
8	South	180,000
9	Southeast	140,000
10	Sureste2	160,000
11	Sureste3	1,004,400,000
12	Southwest	984,500,000
13	Southeast 2	69,000,000

- Wind turbine reference: Considering the commercial availability of onshore wind turbines, study analyses forty different references, with power ratings ranging from 1.7 MW to 5.8 MW. Table 4. Wind turbine specifications by reference show the power specifications, supported minimum and maximum wind speeds, blade swept area, estimated price, and area occupied for each reference analyses.

Table 4: Wind turbine specifications by reference

ID	Reference	Rated power	Min speed	Max speed	Swept area (M2)	Cost (US$)	Occupied area (M2)
1	SG 2.1-114	2.1	8.2	22	10,207	$3,675,000	91,863
2	SG 2.2-122	2.2	8	22	11,690	$3,850,000	105,209
3	SG 2.6-114	2.6	8.8	22	10,207	$4,593,750	91,863
4	SG 2.9-129	2.9	8.4	23	13,273	$5,075,000	119,459
5	SG 3.4-132	3.4	8.8	22	13,685	$6,063,750	123,163
6	SG 5.0-132	5	9.9	23	13,685	$8,750,000	123,163
7	SG 5.0-145	5	9.3	23	16,742	$8,750,000	150,674
8	SG 5.8-155	5.8	9.3	22	19,113	$10,150,000	172,021
9	SG 5.8-170	5.8	8.8	22	22,698	$10,150,000	204,282
10	SWT-DD-120	3.9	9.8	20	11,310	$6,825,000	101,788
11	SWT-DD-130	3.9	9.2	18	13,273	$6,825,000	119,459
12	SWT-DD-142	3.5	8.4	10	15,837	$6,125,000	142,531
13	1.7-100	1.7	8.3	12	7,854	$2,975,000	70,686
14	1.7-103	1.7	8.1	11	8,495	$2,975,000	76,454
15	1.85-82.5	1.85	9.8	14	5,281	$3,237,500	47,529
16	1.85-87	1.85	9.4	13	6,082	$3,237,500	54,739
17	2MW-116	2.7	8.8	38	10,568	$4,725,000	95,115
18	2MW-127	2.8	8.4	40	12,868	$4,900,000	115,812
19	2MW-132	2.7	8.1	35	13,685	$4,725,000	123,163
20	4MW-158	4.8	8.7	37	19,607	$8,400,000	176,460
21	5MW-158	5.3	9	37	19,607	$9,275,000	176,460
22	S111	2.1	8.3	30	9,852	$3,675,000	88,668
23	S120	2.1	7.9	30	11,310	$3,675,000	101,788
24	S128	2.7	8.2	30	13,273	$4,725,000	119,459
25	V150-5.6MW	5.6	9.5	25	17,671	$9,800,000	159,043
26	V162-5.6MW	5.6	9	25	20,612	$9,800,000	185,508
27	V138-3.0MW	3	8.1	24	14,957	$5,250,000	134,614
28	V90-2.0MW	2	9.5	23	6,362	$3,500,000	57,256
29	V100-2.0MW	2	8.8	20	7,854	$3,500,000	70,686
30	V100.1-2.0MW	2	8.3	18	9,503	$3,500,000	85,530
31	V116-2.1MW	2.1	8.1	18	10,568	$3,675,000	95,115
32	V120-2.2MW	2.2	8.1	18	11,310	$3,850,000	101,788
33	V105-3.45MW	3.45	10.2	23	8,825	$6,037,500	79,423
34	v112-3.45MW	3.45	9.8	23	9,852	$6,037,500	88,668
35	v117-3.45MW	3.45	9.5	23	10,936	$6,037,500	98,423
36	v117-4.2MW	4	9.9	25	10,936	$7,000,000	98,423
37	v126-3.45MW	3.45	9.1	22.5	12,469	$6,037,500	112,221
38	v136-3.45MW	3.45	8.6	22.5	14,527	$6,037,500	130,741
39	v136-4.2MW	4.2	9.2	25	14,527	$7,350,000	130,741
40	v150-4.2MW	4.2	8.6	22.5	17,671	$7,350,000	159,043

To execute the optimization models, a Python script with the PuLP library was used.

In Table 5. Results of the OPT model, the results obtained by the OPT model are listed. For each of the budget scenarios, the models show the optimal wind turbine references that will be installed, the number of wind turbines per reference, the station where they should be installed, and wind energy capacities of generated power to be produced. The OPT model promotes the placement of wind turbines at the Northwest2 station, with a greater wind speed and lower rated power.

Table 5: Results of the OPT model

OPT Model				
Budget	Wind turbine reference	Station	Quantity	Power generated
Worst-case scenario	SG 2.2-122	Northwest2	3	2281.44
	S120	Northwest2	541	
Moderate-case scenario	S120	Northwest2	1102	4620.73
Optimistic scenario	SG 2.2-122	Noroeste2	1	6959.54
	1.7-103	Northwest2	1	
	S120	Northwest2	1658	

From the results in Table 5, we can conclude the following:

- For all budget scenarios, the OPT model recommends a majority installation of low power S120 wind turbines (2 MW) along with their combination, in small quantities, with additional references of low nominal power (1.7 MW and 2.2 MW).
- For the OPT model, optimal wind turbines should be exclusively installed at the Northwest2 station, characterized by high wind speeds.
- The values of power generated in the different budget scenarios correspond to the production capacities of large wind farms, of around 1090 MW, which shows a competitive scenario.

Experimental results generated by the wind speed predictor and the energy production optimizer show that the performance of the proposed hybrid model t is satisfactory. A generic methodology is proposed because it is not limited to regional conditions and can be easily adapted to different sets of variables, prediction algorithms, budgets, etc.

The hybrid strategy helps promote competitive techniques such as data analysis and optimization in a joint scenario to achieve a common goal. The values of maximum power generated correspond to capacity levels of vast wind farms worldwide, which is an opportunity for the region of Monterrey. The proposed budgetary approach only includes costs for the purchase of wind turbines, which represent around 75% of total investment.

4.5. Comparison with previous studies

To evaluate the effectiveness of our proposed model for wind energy production, we compare its performance against existing wind farms. The following table summarizes key studies that report maximum power generation, Table 6. Comparison of Maximum Power in Wind Farm Studies.

Comparing the maximum power generation among these studies, Germany's large-scale wind expansion project stands out with a total of 14,000 MW across 2,400 turbines, averaging approximately 5.83 MW per turbine. The offshore wind project in New York achieves 2,800 MW, while the Mudgee Wind Farm in Australia generates 1,300 MW with 185 turbines, averaging 7.03 MW per turbine.

Our proposed OPT model offer competitive results: Worst-case scenario (US$2 billion investment): 2,281.44 MW with SG 2.2-122 (2.2 MW/turbine) and S120 (2.1 MW/turbine) and Optimistic scenario (US$6.1 billion investment): 6,959.54 MW with SG 2.2-122 (2.2 MW), S120, and 1.7-103 (1.7 MW)

*Table 6***:** Comparison of Maximum Power in Wind Farm Studies.

Study Title	Location (State, Country)	Number of Turbines	Turbine Capacity (MW)	Power Generated (MW)
RWE Expands Renewable Energy in West Texas (RWE, 2025)	Texas, USA	109	2.83	308
Germany Wind Expansion Projects (Reuters, 2025)	Various, Germany	2400	5.83	14000
Offshore Wind Project in New York (Associated Press, 2025)	New York, USA	Not specified	Not specified	2800
Mudgee Wind Farm (Daily Telegraph, 2025)	New South Wales, Australia	185	7.03	1300
This study (OPT)	Monterrey, México	1660	2.1	6959.54

The proposed OPT model in the optimistic scenario (6,959 MW) is the second highest among all cases analyzed, behind only Germany's large-scale project (14,000 MW). However, it achieves this with far fewer turbines (1,660 vs. 2,400), indicating a more efficient use of resources. The turbine capacity in our model is lower than that of the Mudgee Wind Farm (2.1 MW vs. 7.03 MW per turbine) but provides a significantly higher total power output.

Compared to the New York offshore wind project (2,800 MW), OPT optimistic scenario exceeded this capacity, demonstrating the potential of the proposed models to maximize wind energy output. The OPT worst-case scenario (2,281.44 MW) is close to New York's project, suggesting that even with a lower investment, the model remains competitive.

Our proposed model, particularly OPT in the optimistic scenario, outperforms the most existing wind projects in terms of power generation. The efficiency of energy production per turbine and cost-effectiveness in different investment scenarios highlight the potential impact of these models on the future of wind energy development.

5. CONCLUSIONS

Wind energy, understood as the use of energy from the wind to produce mechanical energy and its further conversion into electrical energy by means of wind turbines, is a market leader among other renewable energy sources. Its popularization is based on the growing need to control the effects of global climate change and replace fossil fuels, especially oil, with clean energy sources that guarantee energy security and are not affected by the volatility of oil prices or military conflicts. Thus, it is essential to undertake feasibility analysis on the use of wind energy in diverse regions.

In this study, a systemic and systematic methodology is introduced and developed to maximize wind energy production based on a hybrid model with two components. The results of the methodology were derived based on the implementation of optimization models. They are used for the accurate calculation of generated power and provide suggestions of the optimal references of wind turbines that will be installed, quantities and installation areas to guarantee an optimal use of physical, budget and environmental/meteorological resources. Additionally, the values obtained correspond to the capacities installed in vast wind farms worldwide.

For the first component, there are 10 regression models in the literature, which have revealed excellent results. They are validated through data analysis techniques to adjust, for each of the 13 meteorological stations in Monterrey, the most adequate model to predict wind speed, based on the influence of 16 explanatory variables fed with actual data from sensors of the different areas. In each case, the model set was implemented, evaluated, and compared, after which the best model was selected based on error criteria.

For the second component, an optimization model (OPT) was determined and implemented. This offered the installation of wind turbines in the 13 regional stations. Adequate values of maximum generated power were obtained.

The implementation of the hybrid model to maximize wind energy production is a strong tool whose potential has only been tentatively developed in this study. It establishes a meaningful preliminary advancement of the vast problem related to wind energy facility planning. The significance of this work lies in presenting a precise and timely model to estimate the maximum power generated to back decisions related to wind turbine installation.

The main innovations of this study are:

- The development of a systemic methodology to maximize the production of wind energy incorporates the prediction of wind speed for each season and the optimized calculation of references and number of wind turbines that will be installed and in which areas.
- The successful inclusion of actual design features limits the performance of wind turbines such as rated power, minimum and maximum speed, costs of wind turbines, and occupied area.

5. REFERENCES

Ackermann, T., Soder, L., 2002, An overview of wind energy-status. *Renew. Sustain. Energy Rev.* **6**, 67-127. doi: 10.1016/S1364-0321(02)00008-4.

Associated Press, 2025, New York offshore wind project advances amid clean energy push. https://apnews.com/article/20648fdf57184c0d95c4b6fe42b4e8dd.

Banuelos, F., 2011, *Impacto de la generación eléctrica usando fuentes de energía eólica en la red eléctrica nacional*. Universidad Nacional Autónoma de México, Ciudad de México, México.

Blaschke, T., Biberacher, M., Gadocha, S., Schardinger, I., 2013, Energy landscapes: Meeting energy demands and human aspirations. *Biomass Bioenergy*, 6-13. doi: 10.1016/j.biombioe.2012.01.015.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984, *Classification and Regression Trees*. Wadsworth & Brooks/Cole Adv. Books & Software, Monterey, CA, USA.

Chen, T., Guestrin, C., 2016, XGBoost: A scalable tree boosting system. *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, San Francisco, CA, USA, 785-794. doi: 10.1145/2939672.2939785.

Cover, T.M., Hart, P.E., 1967, Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **13**(1), 21-27. doi: 10.1109/TIT.1967.1053964.

Daily Telegraph, 2025, Fewer turbines but taller: Minns Govt approves $2B wind farm in Mudgee. https://www.dailytelegraph.com.au/news/nsw/fewer-turbines-but-taller-minns-govt-approves-2b-wind-farm-in-mudgee/news-story/a7d5f9cedaec705544f75ff6775db148.

EREN, 2015, Parques eólicos. https://energia.jcyl.es/web/es/energias-renovables-ordenacion-energetica/parques-eolicos.html.

Friedman, J.H., 2001, Greedy function approximation: a gradient boosting machine. *Ann. Stat.* **29**(5), 1189-1232. doi: 10.1214/aos/1013203451.

Galton, F., 1894, *Natural Inheritance*, 5th ed. Macmillan, New York, NY, USA.

Gandoni, A., Haider, M., 2015, Beyond the hype: Big data concepts, methods, and analytics. *Int. J. Inf. Manage.*, 137-144. doi: 10.1016/j.ijinfomgt.2014.10.007.

García-González, J.R., Sánchez-Sánchez, P.A., 2020, Theoretical research design: methodological instructions for the development of scientific research proposals and projects. *Inf. Tecnol.* **31**(6), 159-170. doi: 10.4067/S0718-07642020000600159.

García-González, J.R., Sánchez-Sánchez, P.A., Orozco, M., Obredor, S., 2019, Extracción de conocimiento para la predicción y análisis de los resultados de la prueba de calidad de la educación superior en Colombia. *Form. Univ.* **12**(4), 55-62. doi: 10.4067/S0718-50062019000400055.

Garrido, S.G., 2021, *Types of wind turbines*. RENOVETEC, Madrid, Spain.

Giner, D.T., 2013, *Diseño de un aerogenerador de 1.8 Mw*. Univ. Politécnica de Cartagena, Cartagena, Colombia.

Global Wind Energy Council, 2021, *World Wind Report 2021*. Brussels, Belgium. https://gwec.net/global-wind-report-2021

Hernández Vásquez, J.D., Solano De La Rosa, J.E., Ferreira Medina, F.E., 2021, Modelo multicriterio AHP-TOPSIS: una estrategia para evaluar alternativas de generación de energía eléctrica a partir de fuentes renovables. *Investigación Innov. Ing.* **9**(2), 175-191. doi: 10.17081/invinno.9.2.4482.

Hoerl, A.E., Kennard, R.W., 1970, Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12(1), 55-67. doi: 10.2307/1267351.

López, M.V., 2011, *Ingeniería de la Energía Eólica*. Marcombo, Barcelona, Spain.

Mamyrbayev, O., Akhmediyarova, A., Oralbekova, D., Alimkulova, J., Alibiyeva, Z., 2025, Optimizing Renewable Energy Integration Using IoT and Machine Learning Algorithms. *Int. J. Ind. Eng. Manag.* **16**(1), 101-112.

Reuters, 2025, Germany approved 14 GW more wind capacity projects last year, report shows. https://www.reuters.com/sustainability/climate-energy/germany-approved-14-gw-more-wind-capacity-projects-last-year-report-shows-2025-01-15/

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986, Learning internal representations by error propagation. In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1. MIT Press, Cambridge, MA, USA.

RWE, 2025, RWE expands renewable energy in West Texas. *Midland Reporter-Telegram*. https://www.mrt.com/business/oil/article/rwe-west-texas-wind-energy-20245961.php

Sánchez-Sánchez, P.A., García-González, J.R., 2018, Autoregressive Moving Average Recurrent Neural Networks Applied to the Modelling of Colombian Exchange Rate. *Int. J. Artif. Intell.* **16**(2), 194-207.

Sánchez-Sánchez, P.A., García-González, J.R., Rúa, J.M., 2020, Automatic migraine classification using artificial neural networks. *F1000Research* **9**, 618. doi: 10.12688/f1000research.23181.2.

Sánchez-Sánchez, P.A., et al., 2019, Knowledge discovery in musical databases for moods detection. *IEEE Lat. Am. Trans.* **17**(12), 2061-2068. doi: 10.1109/TLA.2019.8986567.

Talayero Navales, A.P., Telmo Martínez, E., 2011, *Wind energy (Renewable Energy Series)*, 2nd ed. Zaragoza University Press, Zaragoza, Spain.

Tibshirani, R., 1996, Regression shrinkage and selection via the LASSO. *J. R. Stat. Soc. B* **58**(1), 267-288. doi: 10.1111/j.2517-6161.1996.tb02080.x.

Zou, H., Hastie, T., 2005, Regularization and Variable Selection via the Elastic Net. *J. R. Stat. Soc. B* **67**(2), 301-320. doi: 10.1111/j.1467-9868.2005.00503.x.