Usage of Statistical Techniques to Monitor the Performance of Wind Turbines

Aditya Dubey
Center for Internet of Things
Madhav Institute of Technology and Science, India
dubeyaditya65@gmail.com

Pradeep Yadav
Department of Computer Science and Engineering,
Institute of Technology and Management, India
pradeep.yadav@itmgoi.in

Subhash Chandra Patel
School of Computing Science and Engineering,
VIT Bhopal University, India
subhash.patel@vitbhopal.ac.in

Abstract: *Calculating the yearly energy output, which keeps the balance between both the generation and consumption of electricity, is made easier with the use of wind energy production estimates for grid interfaces. Effective wind speed forecasting is crucial for achieving this goal. In this research, linear statistical models of prediction Generalized Autoregressive Score (GAS), GAS model with exogenous variable x (GASX), and Autoregressive Integrated Moving Average (ARIMA) are the models used to estimate wind speed accurately. Additionally, the modeling of non-linear time-series data has been done using the Non-Linear GASX (NLGASX) statistical predictive modeling method. Additionally, the REctified Linear Unit (RELU), Softmax, Hyperbolic Tangent (TANH), and Sigmoid modeling approaches are used to optimize the suggested NLGASX model. In comparison to existing models, the suggested optimized NLGASX approach performs significantly better. In order to anticipate wind power, the wind power curve modeling additionally takes wind speed as an input. The estimated wind power has been used to determine the yearly energy.*

Keywords: *AEP, sigmoid, RELU, softmax, TANH, wind power curve model, statistical model.*

Received August 31, 2023; accepted February 11, 2024
https://doi.org/10.34028/iajit/21/2/15

1. Introduction

Utilizing wind power is a significant aspect of the energy producing businesses. Transmission System Operators (TSO) keep the generation and usage of electricity in harmony. Wind power is a variable and non-stationary energy source. It is anticipated that a specialized power-saving method will account for these variations [10]. The cost of integrating power supply into the wind power, system is reduced by the use of wind power forecasts. For wind power forecasting, two techniques have been proposed. Create a wind farm physical model firstly to determine how climatic data and wind power are related. The second approach is a mathematical model that defines the link between weather forecast data and power output based on historic datasets using artificial intelligence and statistics [15]. Numerous innovative approaches have been proposed recently to increase prediction accuracy, classifying the prediction models into three groups: neural networks-based techniques [5, 19, 20], statistical models [14], and physical models [13]. The goal of the physical approaches, also known as analytical approaches, is to create a mathematical function that accounts for every predictor component [23]. Precise values of all predictions are hard to determine as the predictor values are complicated and computationally expensive. The application of statistical modeling approaches is driven by challenges in the model of analysis. A number of defined statistical approaches for forecasting, including Autoregressive Integrated Moving Average (ARIMA) model and their derivates models are suitable when the modeling data that are given are of type linear [14]. Artificial Neural Network (ANN) [5], Support Vector Regression (SVR) [20], and Support Vector Machine (SVM) [19], are examples of artificial intelligence models that may be used efficiently for data of the type non-linear in nature. The performance of models from all three groups is good when dealing with homogeneous data. These mathematical models, though, fail to function well when the data are heterogeneous because they do a poor job of managing heteroscedasticity. Hybrid models, also known as heterogeneous estimation tools, which include the benefits of many forecasting techniques have been created for further improving the accuracy of wind power and wind speed predictions [6, 24]. In the same dataset, heterogeneous estimators show diverse characteristics of the parameter. These estimators assist in identifying a striking discrepancy in the parameter estimations and generate unreachable evaluations. Whereas homogenous estimators display the same variable’s character throughout the same dataset. These approaches provide statistically similar short-run predictions that only vary in the long-run outcome. Non-Linear GASX (NLGAS), GAS model with exogenous variable x (GASX), Generalized Autoregressive Score (GAS) models are utilized and described in this research to address the heteroscedasticity. The entire density structure, together with variance and mean, are used with the GAS approach. It is employed for representing all types of time-series data, including bounded, integer-valued, and real-valued data. While the scoring function is well explained, the values have to guarantee conditional density [2]. Evaluating the score and the highest likelihood estimator is the fundamental difficulty in applying the GAS model to non-linear data. The GAS method currently has a shortcoming in that it does not accurately show updates for the correlation parameters for multivariate GAS models with more than four variables. In order to get over this drawback, a hybrid approach is suggested in this paper. In order to enhance wind speed prediction accuracy and get beyond the constraints of a single forecasting process, the statistical and ANN models are combined to create a hybrid approach.

In this research, statistical models and a few modeling approaches based on neural networks that are essential for modeling and wind energy forecasting are described. By merging the Non-Linear GASX (NLGASX) algorithm with modeling methods utilizing the neural networks activation function (REctified Linear Unit (RELU), Softmax, Sigmoid, and Hyperbolic Tangent (TANH) to attain greater accuracy, this research proposes a hybrid approach to forecast wind speed. Through modeling of the measured wind speed, the suggested model's accuracy is confirmed. Through modeling of the observed wind speed, the suggested model's accuracy is confirmed. Maximum Likelihood Estimation (MLE) is employed to train the models, as given in Equation (1):

$$\hat{\theta} = argmax\ l(\theta; x) \tag{1}$$

Here, θ represent the vector comprising the coefficients of model that can be determined by the MLE and $\hat{\theta}$ represents the predicted value of h. The MLE is referred to as a technique for calculating a value of h in order to achieve maximum $l(\theta;x)$.The models for statistics are learned using MLE.

Annual Energy Production (AEP), as illustrated in Equation (2), as another measure to assess the efficiency of the proposed method.

$$AEP = \sum_{i=1}^{N} \bar{p}\,(v_i) \tag{2}$$

here $\bar{p}$ denotes the average hourly power production, v denotes the wind speed, and a year, total number of hours are represented as N, which is determined using a model of the power curve of a wind turbine.

These examples highlight the uniqueness and contributions of this work.

a) The created model NLGASX uses modeling approaches to increase wind speed prediction accuracy. Using the NLGASX approach, the wind speed is initially predicted; afterwards, modeling approaches use this projected wind speed as an input to increase forecast accuracy.
b) The power curve wind turbine computational approach, that might be employed to wind speed estimated using the hybrid approach, are used to forecast both wind power and wind speed.
c) The yearly generation of energy is determined utilizing wind energy generated through the wind turbine power curve method.

The research paper is structured as: Different time series methods for predicting speed of wind are discussed in section 2. The power curve method for wind turbine that was used to forecast wind power are shown in section 3. Several modeling methods for predicting wind speed are given in section 4 in order to maximize the forecasts. NLGASX technique and the method for creating hybrid models employing optimizations based on neural networks are both presented in section 5. The evaluation of several generated hybrid models is done, and results are shown in section 6. The future work and conclusion are discussed in last section.

2. Wind Speed Prediction Using Time Series Model

The forecast accuracy, forecast horizon, and prediction of the data have all been described for the time-series model. Utilizing variables like density, temperature, wind speed, etc., forecasting is done. Wind speed measurements are taken at regular intervals in order to create precise predictions about the amount of energy produced at a particular area. It is possible to make utilization of the wind speed data that is gathered at regular or predetermined intervals as a timeseries data that may be utilized for forecasting using time-series data. The GASX, GAS, ARIMA model are three statistical models that are used to estimate wind speed and are discussed in the following sections:

2.1. ARIMA

The ARIMA approach, that Maatallah *et al*. [14] implemented, may describe a variety of time series kinds, such as ARMA series and their derivatives such as Moving Average (MA), and Auto-Regressive (AR) series. The actual constraint is the model's presumed linear shape, which assumes a straight related structure amongst the time-series data and prevents the ARIMA model from detecting nonlinear patterns. The model containing q representing the MA terms, d represents the degree of a differencing term, and p AR terms is denoted by the expression ARIMA (p, d, q). The MA(q) and AR(p) models that are described in Equation (3) are included in this model.

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p} + \epsilon_t + \alpha_1\epsilon_t + \alpha_1\epsilon_{t-1} + \cdots + \alpha_q\epsilon_{t-q} \tag{3}$$

Here ϵ_t is assumed to be white noise, $\alpha_1,\alpha_2,\ldots,\alpha_q$ whereas q represents the MA coefficients, and $\beta_0,\beta_1,\ldots,\beta_p$ represent the AR coefficients.

The ARIMA approach only works with stationary time series and has superior performance for linear time-series data. For non-stationary series, it must first become stationary before the ARIMA model can be used.

2.2. GAS model

The GAS method has been proposed by Harvey [8], and Creal *et al*. [3] and it can deal with the density time series of varying nature. The GAS approach is expressed by the conditioned observing density $p(y_t/\theta_t)$, where θ_t represents a parameter of latent time-varying nature. The recursion is stated in Equation (4):

$$\theta_t = \mu + \sum_{(i=1)}^{p} \varphi_i \theta_{t-i} + \sum_{(j=1)}^{q} \alpha_j s(\theta_{j-1}) \frac{\partial log\, p(y_{t-j})/(\theta_{t-j}))}{\partial \theta_{t-j}} \qquad (4)$$

For one measurement at time j, s is the conditioned density contribution's first derivative multiplied by a factor of scaling that must be specifically positive. α specifies the scale factor. φ is an indicator of AR. In the AR procedure for order 1, μ represents an intercept. The GAS method takes into consideration all time-series data kinds. Whatever the data type (0, 1)-bounded, integer-valued, or real valued as long as it has a conditional density that is properly defined by the scoring function and the Hessian, it is of no significance. Determining scores and performing the MLE for the data of non-linear nature are the two most challenging aspects of employing GAS models.

2.3. GASX

The GASX approach is the GAS model that has been optimized through the addition of exogenous elements X. Here, the observation y_t and a latent time-varying parameter h_t are used to determine the conditioned observed density $p(y_t/\theta_t)$, When the recursion is followed by the parameter ht, as illustrated in Equation (5):

$$\theta_t = \mu + \sum_{(k=1)}^{K} \beta_k X_{t,k} + \sum_{(i=1)}^{p} \varphi_i \theta_{t-i} + \sum_{(j=1)}^{q} \alpha_j s(\theta_{j-1}) \frac{\partial log\, p(y_{t-j})/(\theta_{t-j}))}{\partial \theta_{t-j}} \qquad (5)$$

where A scaling parameter represented as α. An AR coefficient is φ. β is an exogenous factor coefficient. A exogenous variable is X. l represents a first-order AR procedure intercept. For a single measurement at time j, s is the conditional density distribution's first derivative multiplied by a scaling factor that is strictly positive. The GASX model's benefit is that it incorporates an additional component to increase the model's accuracy. The inclusion of more than four components restricts the GASX model, which does not provide an accurate updating for the correlation. The Pseudocode of the proposed techniques is described below:

Pseudocode: NLGASX Technique.

```
# Import necessary libraries
import statsmodels.api as sm
# Function to simulate NGAS model
def simulate_ngas_model(y, x, order, exog_order):
# Create lagged variables for the autoregressive component
        y_lagged = lag_variables(y, order)
# Create lagged variables for the exogenous component
        x_lagged = lag_variables(x, exog_order)
# Combine lagged variables for autoregressive and exogenous components
    X = np.column_stack((y_lagged, x_lagged))
    # Add a constant term to the model
    X = sm.add_constant(X)
    # Fit NGAS model
    model = sm.OLS(y, X)
    results = model.fit()
    # Display model summary
    print(results.summary())
# Function to create lagged variables
def lag_variables(series, order):
lags = [series.shift(i) for i in range(1, order + 1)]
return pd.concat([series] + lags, axis=1).dropna()
# Example usage
# Assuming y and x are your time series data
order = 2# Autoregressive order
exog_order = 1# Exogenous order
simulate_ngas_model(y, x, order, exog_order)
```

3. Power Curve Techniques for Wind Turbine

Power curve techniques are essential in the context of wind turbines as they play a crucial role in assessing and optimizing the performance of these renewable energy systems. The relationship between the electrical power output produced by a wind turbine and wind speed is represented by the power curve. Power curve techniques for wind turbines are indispensable for assessing, optimizing, and ensuring the reliable operation of wind energy systems. They provide valuable insights into the turbine's performance under various wind conditions, facilitating efficient energy production and contributing to the ongoing advancement of wind energy technology. A wind turbine maker will offer a tool to assess the performance of the turbine in ideal circumstances. Typically, the wind power measurement or the turbine's power rating are used for releasing the power curve of a commercially accessible wind turbine. The cut-out speed, rated speed, and cut-in speed are all included in this turbine power curve. Similar to this, under typical testing circumstances, a nominal wind power value is collected [21]. An experimental power curve that can forecast the wind turbine's output power for all likely wind speeds during cut-out and cut-in is required to determine the AEP of a specific wind turbine. Using the cubic and quadratic laws, respectively, Jangamshetti and Rau [9] and Pallabazzer [18] implemented experimental power curve approximations. Equation (6) provides the Quadratic Power Curve (QPC) method.

$$P(v) = \frac{p_r(v^2 - v_{cin}^2)}{v_{rate}^2 - v_{cin}^2} \quad (6)$$

Equation (7) describes the Cubic Power Curve (CPC) method:

$$P(v) = \frac{p_r(v^3 - v_{cin}^3)}{v_{rate}^3 - v_{cin}^3} \quad (7)$$

Wind speed is demonstrated by v, rated power is presented by pr, cut-out and cut-in wind speed is demonstrated by v_{cout} and v_{cin} respectively, Rated speed is represented by v_{rate}.

Despite the simplicity of these techniques, these do not accurately reflect how wind turbines actually behave. Kazemi and Goudarzi [11] implemented the CPC and QPC generalized techniques, which may be created utilizing the nominal wind power data made available by the maker, in order to improve forecast accuracy. Equation (8) defines the generalized QPC concept.

$$P(v) = a_1 v_i^2 + a_2 v_i + a_3 \quad (8)$$

Equation (9) provides the generalized CPC concept in the same way.

$$P(v) = b_1 v_i^3 + b_2 v_i^2 + b_3 v_i + b_4 \quad (9)$$

In which the coefficients are b_1, b_2, b_3, b_4, a_1, a_2, a_3. The wind power $P(v)$ determined between the rated wind speed and the cut-in is shown below.

Wind power tends to be consistent between cut-out and rated wind speeds, in addition to that it is inclined to become zero after cut out and prior cut-in wind speed. The Piecewise Polynomial Power Curve (PPPC) approach, that is utilized to compute the wind power by splitting the domain of wind speed into contiguous intervals that may be expressed as a different polynomial in each interval, has also been proposed by Ai *et al*. [1] and Thapar *et al*. [22]. The least squares approach was used in each region to apply a third-degree polynomial to three locations, sometimes achieving 100% accuracy. Equation (10) represents an expression for the generalized PPPC approach.

$$p(v) = \begin{cases} 0, & (v < v_c, v > v_f) \\ a_1 + b_1 v + c_1 v^2 + d_1 v^3, & (v_c \le v < v_1) \\ a_2 + b_2 v + c_2 v^2 + d_2 v^3, & (v_1 \le v < v_2) \\ a_3 + b_3 v + c_3 v^2 + d_3 v^3, & (v_2 \le v < v_f) \end{cases} \quad (10)$$

In this case, the cubic equation's coefficients are d_1, c_1, b_1, a_1. The breakpoints, or knots, in this case are represented by v_f, v_c, v_2, and v_1. Although on an ideal number of data points, by finding the optimum fitting lower degree polynomial, the generalized model may be created. The technique is employed for estimating wind power utilizing the provided wind speed.

4. Methodologies for Forecasting of Wind Speed

For estimating the wind speed, a variety of modeling methods are utilized, including optimization approaches, data mining techniques, and activation functions for neural networks. In this research, a hybrid approach to forecast wind speed is designed using a variety of neural network activation functions. Many of the modeling methods utilized to develop the hybrid approach are discussed here, including RELU, Softmax, TANH, and Sigmoid.

4.1. Sigmoid

A Sigmoid function has a distinctive Sigmoid curve or S-form curve [7]. Sigmoid function often pertains to the particular case of the logistic function. Applying Equation (11) we derive the function:

$$S(x) = \frac{1}{1 + e^{-x}} \quad (11)$$

The NLGASX approach's non-linear feature is stored using the Sigmoid function. This function possesses a domain of all real values and a return variable that gradually increases often from *-1* to *1* or, in other situation, from 0 to *1* [12].

4.2. TANH

A Sigmoid function substitute is the TANH [4]. The drawback of Sigmoid function states that it may become stuck while training. The result is close to zero provided the dataset contains values that are highly negative. Though less prevalent while in the TANH training phase, this feature nevertheless produces negative outputs when given negative inputs and zero outputs when given zero inputs. Hyperbolic sine to hyperbolic cosine ratio, which is provided in Equation (12), is the definition of the TANH function.

$$T(x) = \frac{Sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (12)$$

4.3. Softmax

The logistic function has been generalized to create the Softmax function [16]. The outcome of the Softmax function may be used for clear-cut dissemination in likelihood theory, which is a probability dispersion across K different possible outcomes. In reality, it is the clear-cut likelihood dissemination's angle log-normalizer. Equation (13) describes the function.

$$S(x_j) = \frac{e^{x_j}}{\sum_{k=1}^{K} e^{x_k}} \quad (13)$$

Various multi-class grouping methods, such as multiclass direct discriminant analysis, naive Bayes classifiers, ANN, and multinomial logistic regression (also known as Softmax regression), all make use of the Softmax function.

4.4. RELU

The rectifier, an activation function that has a wide range of definitions, is responsible for the positive portion of an ANN argument [11]. Equation (14)'s definition of it is as follows:

$$R(x) = \begin{cases} x, x \geq 0 \\ 0, x < 0 \end{cases} \quad (14)$$

while the input value is x. The widely used logistic Sigmoid function was replaced with a convolutional network that made use of that actuation function. Its widespread use as an actuation function with deep neural networks is a result of its successful training. The RELU is the name of the component used in the rectifier. The analytical function specified in Equation (15) is a straightforward rectifier approximation.

$$F(x) = log(1 + e^x) \quad (15)$$

Another name for it is a function of soft plus. Sigmoid function is the same as the soft plus function's first derivative. For complex and large datasets, the RELU function performs more quickly compared to Sigmoid or corresponding functions.

5. Proposed Methodology

The following subsection discusses proposed NLGASX and hybrid approaches that are used to estimate wind speed. Hybrid approaches work well for both non-linear and linear datasets, whereas the NLGASX approach works best for non-linear datasets. The NLGASX approach is used with modeling approaches to create the hybrid approaches.

5.1. NLGASX

By combining linear and nonlinear blocks in parallel, the nonlinear structures when NLGASX model builds. Using Equation (16), the NLGASX function is specified as follows:

$$y_t = G(y_{t-1}, y_{t-1}, \ldots\ldots, x_{t-1}, x_{t-2}, \ldots\ldots) + \epsilon_t \quad (16)$$

where ϵ_t is an error term, x representing exogenous variable, and y is the variable of interest. The nonlinear function that uses the Poisson distribution to represent it is called the function G. The Poisson distribution is a discrete probability distribution that provides the likelihood of such occurrences whenever a count of events are supplied that take place in a specified time period or location at a constant rate that's irrelevant of the time of the previous event. There are several categories other than the Poisson variety that may be used to calculate the nonlinear block, including the exponential, skewed, normal Student-t, Student-t, etc. It performs poorly with data of linear time-series nature.

5.2. Hybrid Technique

A hybrid approach for predicting speed of wind that combines the NLGASX approach with modeling approaches that utilize the activation function of neural networks (RELU, Softmax, Sigmoid, and TANH). The definition of a hybrid approach is given in Equation (17):

$$y_t = F(G(y_t)) \quad (17)$$

where F represents a neural network function and G represents the previously described NLGASX function. The hybrid approaches that are created fall into four categories: RELU+NLGASX, Softmax+NLGASX, Sigmoid+NLGASX+, and TANH+NLGASX. The detailed organization of proposed work can be examined in Figure 1.

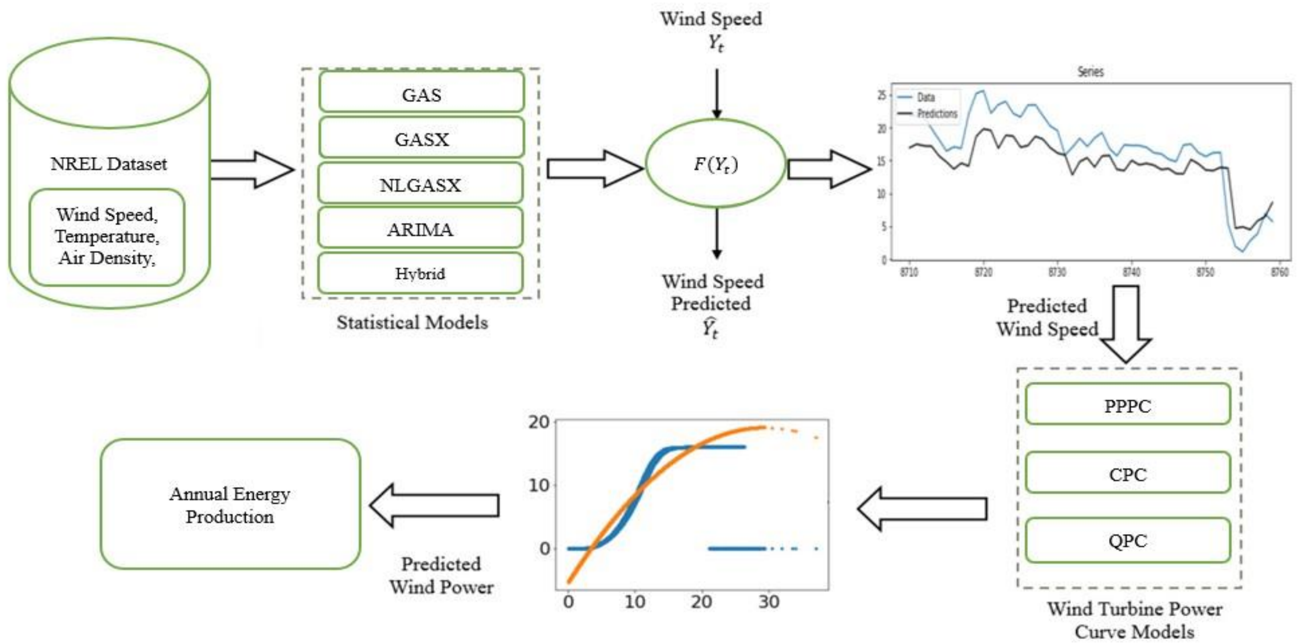


Figure 1. Proposed technique for assessing the performance of wind turbines.

6. Results and Discussions

6.1. Statistical Test for the Comparison of the Examined Methods

Comparing the performance of different models, including nonlinear GAS models with exogenous variables, often involves statistical tests or metrics to assess their goodness of fit, predictive accuracy, or overall performance. Here are some common statistical tests and metrics used for comparing models:

- **Likelihood Ratio Test (LRT)**: this test is often used for comparing nested models. For a nonlinear GAS model with exogenous variables, you might compare a simpler model (e.g., without exogenous variables) to a more complex model (e.g., with exogenous variables) using the LRT.
- **Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE)**: these are common metrics for assessing the accuracy of model predictions. Lower values indicate better predictive performance. You can compare the *RMSE* or *MAE* of different models to evaluate their prediction quality. After designing a model that accurately captures the behavior of the real data, it is important to choose appropriate criteria for evaluating a model's generalizability. The efficiency of the wind speed and power curve of wind turbine is often obtained utilizing the *R-square*, *RMSE*, and *MAE*. Equations (18), (19), and (20), respectively, describe the *R-square*, *RMSE*, and *MAE*:

$$R - square = 1 - \frac{\sum_{i=1}^{N}\left(y(i) - x(i)\right)^2}{\sum_{i=1}^{N}\left(x(i) - \bar{x}(i)\right)^2} \tag{18}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y(i) - x(i)\right)^2} \tag{19}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N} y(i) - \bar{x}(i) \tag{20}$$

$\bar{x}$, represents the input variables average, x denotes the input variable, y denotes the variable that is estimated, and counts of input data is represented by N.

Variable density of air, temperature of the air, and speed of the wind are used for evaluating the wind power. The approach that works better has lesser RMSE and MAE values in addition to the highest R-squared value. Python (3.6) is used for the implementation of models of statistics (NLGASX, GASX, GAS, and ARIMA) and optimization methods (Sigmoid, RELU, Softmax, and TANH). Other metrics, including as the Bayesian Information Criterion (BIC), and the Akaike Information Criterion (AIC) are additionally employed to determine which approach fits a given dataset the best. It may be determined that a certain model performed superior to the others based on the BIC and AIC values.

- **BIC and AIC**

The *BIC* and *AIC* are statistical selection parameters for models from a limited number of models [1, 9]. They have a likelihood estimate basis. Parameter addition would raise the probability while model fitting but might potentially lead to overfitting. By using the term of penalty relevant to number of parameters, the BIC and *AIC* are able to resolve the overfitting issue. In comparison to *BIC*, the penal term in *AIC* is lower. *BIC* and *AIC* are shown, respectively, in Equations (21) and (22):

$$BIC = ln\left(\frac{\sum_{t=1}^{T} e_t^2}{T}\right) + \left(\frac{pln(T)}{T}\right) \tag{21}$$

$$AIC = ln\left(\frac{\sum_{t=1}^{T} e_t^2}{T}\right) + \left(\frac{2p}{T}\right) \tag{22}$$

Here, the residual generated by the model-fitting procedure in period T is called e_t. Data from T time periods were utilized to fit a model having p parameters. Whenever the *BIC* and the *AIC* penalized the sum of squared residuals, more parameters may be added to the model. Models are seen to be effective if their *BIC* or *AIC* scores are low.

6.2. Dataset Description

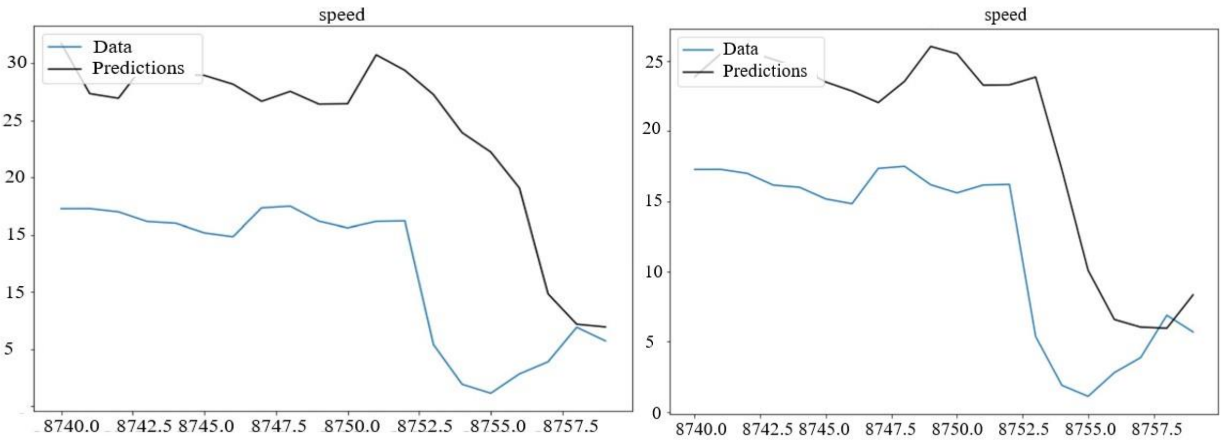
The use of statistical techniques in real-world situations is covered in this subsection. This paper utilizes the National Renewable Energy Laboratory (NREL) resource file datasets here, having site_ID as 72509. The site of the event is located at and latitude 41.775928 and longitude-106.259064; for Datasets-2007 9.24 m/s is the average speed of wind whereas for the dataset-2008, 9.83 m/s is the average speed of wind [17]. A wind turbine dataset involves providing information about the data's origin, structure, variables. This dataset captures the operational parameters and performance metrics of wind turbines over a specific period. The data is collected to analyze the relationship between wind speed and turbine performance. The measurements were taken at regular intervals of 10 minutes. The timestamp indicating when the measurements were taken. Wind Speed (m/s): The data type is Continuous. The speed of the wind at the turbine location at the time of measurement. Turbine Power Output (kW): The data type is Continuous. The electrical power output generated by the wind turbine at the time of measurement. Rotor Speed (rpm): The data type is Continuous. The rotational speed of the turbine's rotor at the time of measurement. Generator Speed (rpm): The data type is Continuous. The rotational speed of the turbine's generator at the time of measurement. Ambient Temperature (°C): The data type is Continuous. The temperature of the surrounding air at the turbine location at the time of measurement. Wind Direction (degrees): The data type is Continuous. The direction from which the wind is blowing at the turbine location at the time of measurement.

6.3. Models for Predicting Wind Speed

The Capacity Factor (CF) for the NREL wind farm is 0.359, and the 5-min average power production and 5-min average wind speed statistics were collected from the Supervisory Control and Data Acquisition (SCADA) architecture. The wind farm statistics were obtained at the height of 100 m. The CF measures the difference between actual and maximum energy output during that time. It is employed for comparing various methods of producing power. Between Jan 2007 and Dec 2007, as well as in the following year of 2008, over lakh observations of data were collected. In this research, the MLE technique in its two variants, MLE-202 and MLE-404, has been employed to apply the NLGASX, GASX, GAS, and ARIMA techniques. The real and expected wind speed for the NLGASX framework is shown in Figure 2 and for the GASX approach, it is shown in Figure 3. The GAS approach, which outperforms the model developed by ARIMA in terms of prediction, is shown in Figure 4. For Dataset-2007, Figure 5 depict the actual and estimated wind speeds for the ARIMA framework employing MLE-202 and MLE-404, respectively. The model was trained from the training set, for testing set, the comparison of predicted to the available values is represented in Figures 2, 3, 4, and 5. Performance-wise, the MLE-404 variant outperforms the MLE-202 variant. Comparing the GAS approach to the GASX approach, the GASX approach exhibits superior prediction. Additionally, the predicted wind speed for the NLGASX, GASX, GAS, and ARIMA approaches utilizing MLE-404 and MLE-202, respectively, is presented in Figures 6 , 7, 8, and 9.

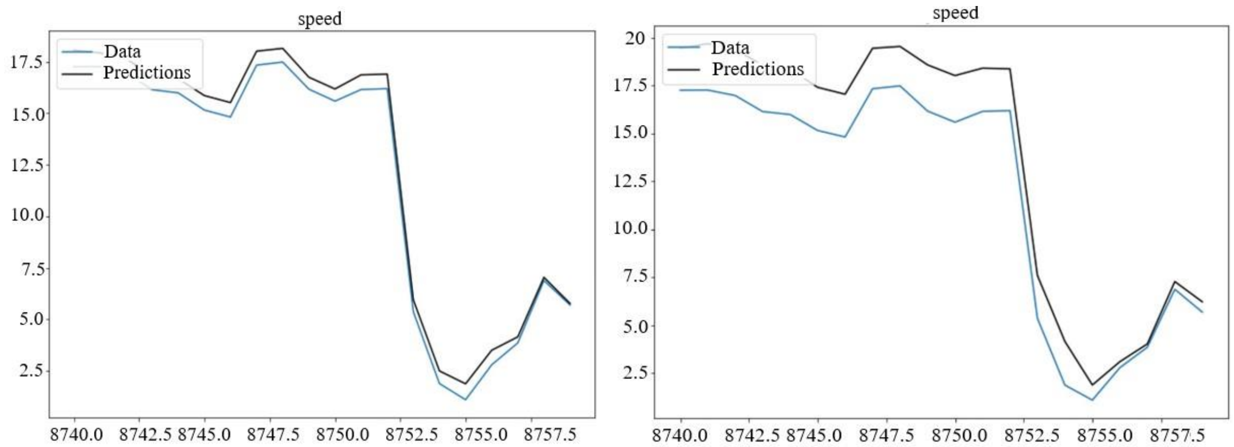
Cross-validation techniques are generally utilized for assessing the model performance by splitting the dataset into testing and training sets. For data of time series nature, traditional cross-validation techniques like k-fold cross-validation may not be directly applicable due to the temporal nature of the data.

However, for nonlinear models with exogenous variables, cross-validation techniques can still be tailored for time series data. Some common approaches include: Time series split cross-validation, walk-forward validation, expanding window cross-validation, rolling origin validation, and block cross-validation. Time series split cross-validation involves splitting the time series data into training and testing sets sequentially. The training set maintains data from the beginning up to a certain point in time, and the testing set has data from that point forward.



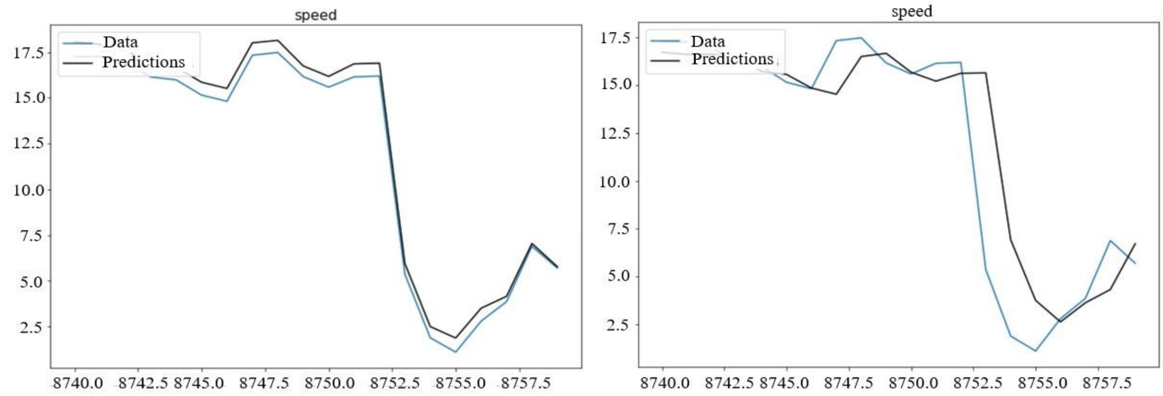
a) NLGASX (404). b) NLGASX (202).

Figure 2. For Dataset 2007 usage of NLGASX for predicting wind speed.



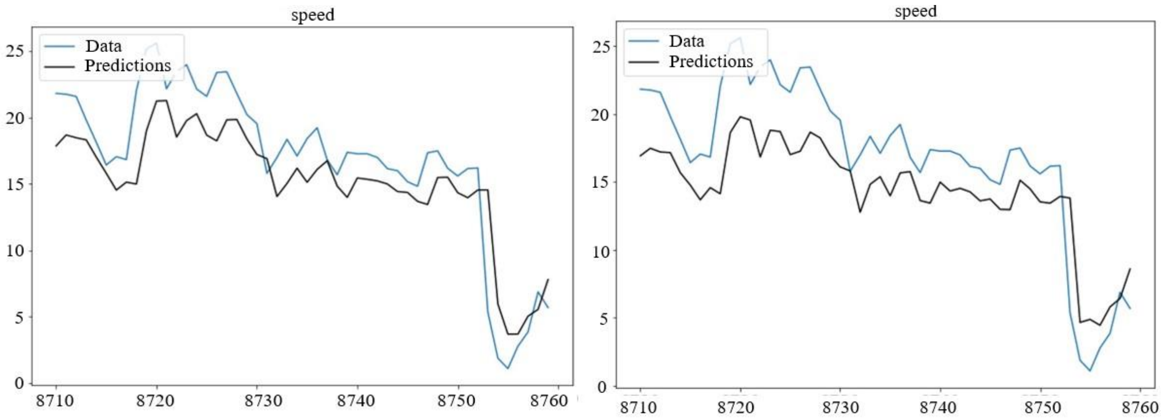
a) GASX (404). b) GASX (202).

Figure 3. For Dataset 2007 usage of GASX for predicting wind speed.



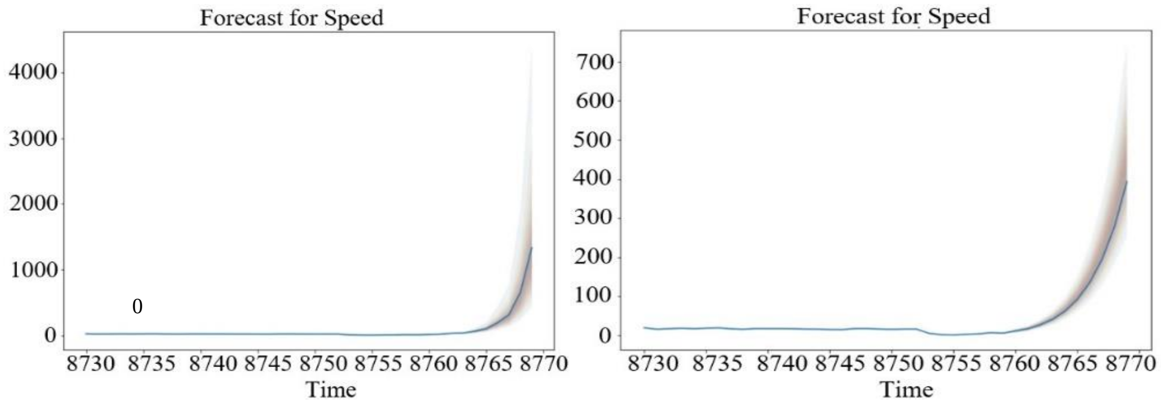
a) GAS (404). b) GAS (202).

Figure 4. For Dataset 2007 usage of GAS for predicting wind speed.



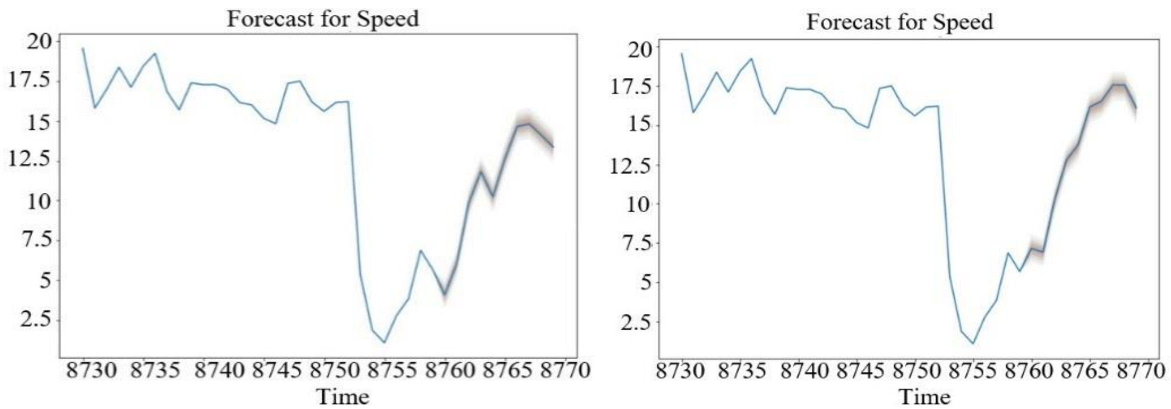
a)ARIMA (404). b) ARIMA (202).

Figure 5. For Dataset 2007 usage of ARIMA for predicting wind speed.



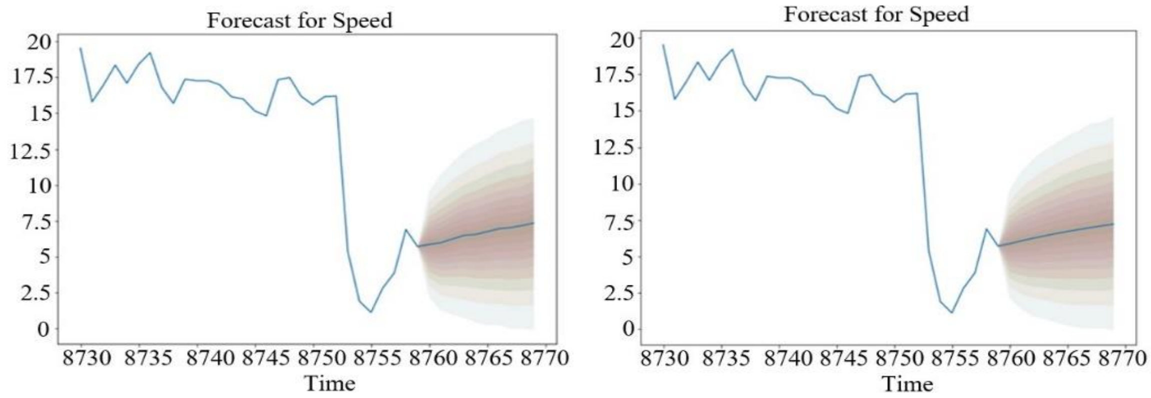
a) NLGASX (404). b) NLGASX (202).

Figure 6. For Dataset 2007 usage of NLGASX for predicting wind speed's next 10 values.



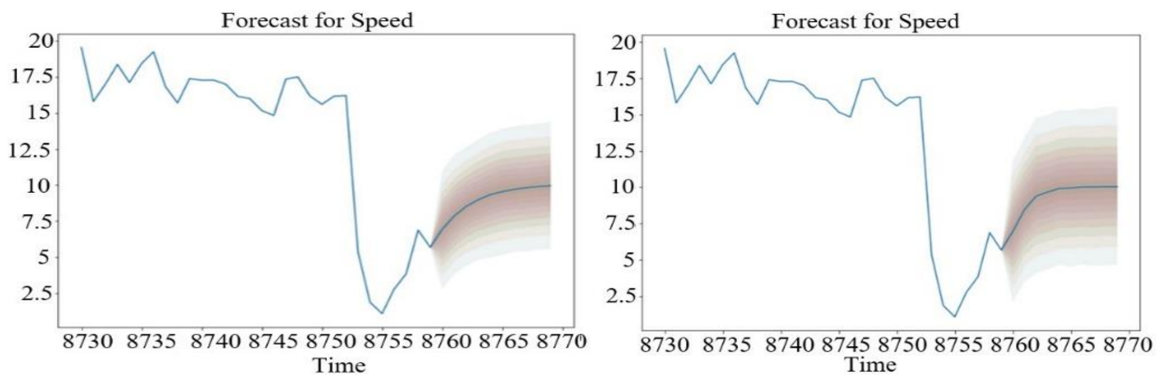
a) GASX (404). b) GASX (202).

Figure 7. For Dataset 2007 usage of GASX for predicting wind speed's next 10 values.



a) GAS (404). b) GAS (202).

Figure 8. For Dataset 2007 usage of GAS for predicting wind speed's next 10 values.



a) ARIMA (404). b) ARIMA (202).

Figure 9. For Dataset 2007 usage of ARIMA for predicting wind speed's next 10 values.

Hub density, air temperature, and Wind speed are taken into account while creating the non-linear and linear models for forecasting. Density and temperature are regarded as exogeneous factors. Particularly temperature is employed as an exogenous factor in the experiments for the suggested model. The Sigmoid+ NLGASX hybrid approach for Dataset-2007 provides the smallest highest R-square, RMSE, and MAE values. Its highest lowest R-square, MAE, and RMSE results were obtained by the ARIMA model employing MLE (202) in comparison. As a result, a hybrid framework that employs the Sigmoid modeling approach outperforms other models. Table 1 shows that the RMSE/MAE values for the GAS approach are lower than those for the ARIMA approach, indicating that the GAS approach outperforms the ARIMA approach. GASX outperforms the GAS method when compared to the GAS method. The (Sigmoid+NLGASX) hybrid approach is the best-performing approach for Dataset-2008 and reports the lowest MAE, while the ARIMA approach yields the greatest MAE. The GASX approach has the smallest RMSE and greatest R-square values, making it the most efficient approach overall. Because of this, the GASX approach and hybrid approaches that employ the Sigmoid modeling approach outperform the rest of the models.

Table 2 pertaining to the Dataset-2007 Dataset-2008 both show the results of the metrics for the chosen model. Gas models are often associated with time series analysis, and they typically involve nonlinear modeling of time series data. Exogenous variables, denoted as x, are often included in these models to capture external influences on the time series.

Table 1. Analysis of the efficiency of forecasting methods of wind speed for Dataset 2007.

Technique	MLE Criterion	R-square	RMSE	MAE	BIC	AIC
NLGASX	Sigmoid	0.810	2.271	1.568	14391.982	17456.559
NLGASX	RELU	0.810	2.276	1.573	14359.779	17456.566
NLGASX	Softmax	0.809	2.279	1.611	14422.086	17456.552
NLGASX	TANH	0.808	2.290	1.577	14500.079	17456.534
NLGASX	404	0.198	6.637	5.345	52620.305	52528.297
NLGASX	202	0.152	6.864	5.912	57876.962	57813.262
GASX	404	0.807	2.302	1.654	2299.511	2284.441
GASX	202	0.806	2.319	2.212	2373.905	2317.283
GAS	404	0.795	2.584	2.584	1.941	39241.742
GAS	202	0.786	2.979	2.538	39220.130	39177.664
ARIMA	404	0.430	2.895	2.472	40823.511	40752.736
ARIMA	202	0.109	3.528	3.145	42813.914	42771.448

Table 2. Analysis of the efficiency of forecasting methods of wind speed for Dataset 2008.

Technique	MLE Criterion	R-square	RMSE	MAE	BIC	AIC
NLGASX	Sigmoid	0.833	2.162	1.485	13497.118	17456.764
NLGASX	RELU	0.832	2.170	1.499	13563.068	17456.749
NLGASX	Softmax	0.833	2.161	1.497	13486.932	17456.766
NLGASX	TANH	0.832	2.170	1.529	13561.498	17456.749
NLGASX	404	0.145	6.335	4.741	51793.811	51701.804
NLGASX	202	0.175	5.503	4.457	62699.138	62635.439
GASX	404	0.902	1.834	1.749	3821.824	3736.894
GASX	202	0.795	1.946	1.778	3893.248	3949.869
GAS	404	0.651	3.098	2.086	38487.066	38416.291
GAS	202	0.650	3.116	2.089	38469.951	38427.485
ARIMA	404	0.376	3.522	2.721	40249.917	40179.142
ARIMA	202	0.121	4.093	3.338	42380.262	42337.796

6.4. Power Curve Modelling for Wind Turbines

The PPPC, CPC, and QPC techniques are used for calculating wind power. Equations (8), (9), and (10) are used to determine wind power for the QPC, CPC, and PPPC models, respectively. Tables 3 and 4 for the corresponding Datasets 2007 and 2008 give the efficiency metrics for each model. Here, the standards of R-square, RMSE, and MAE value are used to evaluate efficiency across various datasets. The approach that performs best is the one with a high R-square value and a small MAE and RMSE values. The PPPC method has the smallest RMSE and MAE, as well as the higher value of R-square in both datasets, based to the results. Although the QPC approach has the smallest Rsquare value, the greatest RMSE and MAE. On considering this, the PPPC approach outperforms the CPC and QPC approaches. Wind speed from Dataset-2007 and -2008 is used to compute wind power. Wind power is computed in MW and wind speed expressed as in m/s.

Table 3. Wing power prediction techniques performance for the NREL Dataset 2007.

Technique	Square	RMSE	MAE	BIC	AIC
PPPC	0.9738	0.9929	0.6277	1481.5823	210216.9029
CPC	0.9679	1.0989	0.6891	19864.0456	210216.4968
QPC	0.8617	2.2822	1.7850	173499.8325	210213.5737

Table 4. Wing power prediction techniques performance for the NREL Dataset 2008.

Technique	R-square	RMSE	MAE	BIC	AIC
PPPC	0.9718	1.0411	0.6487	8494.2321	210216.7131
CPC	0.9667	1.1311	0.7224	25924.7641	210216.3815
QPC	0.8647	2.2800	1.7785	173301.2935	210213.5775

6.5. AEP

Table 5 with the unit of MWh shows the calculated AEP for two distinct datasets utilizing Equation (3). As stated in earlier sections, three distinct wind power curve approaches the PPPC approach, CPC approach, and QPC model are utilised for determination of the wind power. The yearly energy determined using Dataset-2007 is highest when the PPPC approach is used and lowest when the QPC approach is used. The yearly energy determined using Dataset-2007 is highest when the PPPC approach is used and lowest when the QPC approach is used. As a result, when examined alongside QPC and CPC approaches, the efficiency of the PPPC approach is improved. All datasets had improved power calculations when employing the PPPC approach, making AEP utilizing the PPPC approach superior to CPC and QPC approaches. In compared to the other approaches, the AEP employing QPC approach yields the lowest value.

Table 5. Annual energy generation performance.

Technique	Dataset 2008	Dataset 2007
PPPC	65317.6775	59465.0909
CPC	65317.3461	59464.8145
QPC	65713.1751	59464.2598

The proposed technique might be considered time-efficient in certain cases:

- **Adaptability to Nonlinear Patterns**: nonlinear GAS models are designed to capture and model nonlinear patterns in time series data. In situations where the underlying relationships are genuinely nonlinear, a model that explicitly accounts for this nonlinearity may require fewer parameters to achieve a good fit compared to linear models.
- **Score-Driven Approach**: GAS models are often score-driven, meaning that the model parameters are updated in response to the score (gradient) of the log-likelihood function. This approach can lead to efficient parameter estimation, particularly when dealing with time-varying volatility or complex dynamic patterns.
- **Effective Handling of Exogenous Variables**: including exogenous variables (x) in a model allows for the incorporation of external influences on the time series. In cases where exogenous variables have a significant impact, the inclusion of x can enhance the model's explanatory power without necessitating overly complex parameterization.
- **Flexibility in Model Specification**: nonlinear GAS models provide flexibility in model specification, allowing for the inclusion of various nonlinear functions to capture complex relationships. This flexibility can lead to more accurate representations of the underlying data dynamics.

7. Conclusions

In conclusion, proposed technique stands as a versatile and powerful technique in the realm of time series analysis. Its ability to capture complex nonlinear relationships within a time series, coupled with the inclusion of exogenous variables, adds a layer of sophistication to modeling dynamic systems. The statistical approach for predicting speed of wind is crucial for the design of the appropriate approach for wind energy. For predicting wind speed, the suggested hybrid and statistical approach is built, and for predicting wind power, the power curve approach has been designed. This paper presents a first-ever NLGASX method. The intended work concentrated on choosing the best approach to forecasting wind power and wind speed. These suggested hybrid approaches were examined for the selection using two datasets from 2008 and 2007. The model with the greatest accuracy was chosen to reliably predict the following 100 values of wind speed. It was demonstrated that a hybrid approach, which was dependent on our NLGASX concept and used the Sigmoid approach, did an excellent task of predicting wind speed. A commercial wind farm is advised to apply the proposed hybrid approaches to evaluate the effectiveness of its operations. The proposed technique represents a valuable addition to the toolkit of time series analysts. Its flexibility, adaptability, and capacity to incorporate external factors make it a promising approach for a wide range of applications. As research progresses and the methodology evolves, it is likely to continue contributing to advancements in our understanding of dynamic systems and improving the accuracy of time series modeling.

Future work for the proposed technique could focus on addressing various aspects to enhance its applicability, accuracy, and usability. Here are several potential avenues for future research: Explore the extension of nonlinear GAS models with exogenous variables to handle multivariate time series data. Investigate how the model can be adapted to capture dependencies and interactions across multiple variables. Secondly, extend the model to handle exogenous variables that exhibit time-varying patterns. This could involve introducing dynamic structures for the exogenous variables, allowing the model to adapt to changes in external influences over time. Thirdly, Develop Bayesian versions of the nonlinear GAS model with exogenous variables. Bayesian approaches can provide uncertainty estimates for model parameters and improve the handling of parameter priors, especially in situations with limited data. Fourthly, investigate the robustness of the nonlinear GAS model to outliers and structural breaks in the data. Develop methods to enhance the model's ability to handle anomalies and sudden changes in the time series. Another future work is exploring ways to integrate nonlinear GAS models with machine learning techniques, such as neural networks or ensemble methods. This could improve the model's ability to capture complex patterns and dependencies in the presence of high-dimensional exogenous variables. Sixthly, develop methodologies for real-time forecasting and monitoring using the nonlinear GAS model with exogenous variables. Explore how the model can be efficiently updated as new data becomes available and assess its performance in dynamic forecasting environments. Seventhly, working on improving the interpretability of the model results, especially when dealing with complex nonlinearities. Develop methods to extract meaningful insights from the estimated parameters and understand the contribution of exogenous variables to the overall model dynamics. Another future work is conducting comprehensive comparative studies with other state-of-the-art nonlinear time series models, both with and without exogenous variables. Assess the strengths and weaknesses of the nonlinear GAS model in different data scenarios. Another future work is investigating strategies for handling missing or irregularly sampled data within the context of the nonlinear GAS model with exogenous variables. Develop techniques to impute missing values or adapt the model to unevenly spaced time series observations. Lastly, the future work is to develop user-friendly software packages or libraries for implementing the nonlinear GAS model with exogenous variables, making it accessible to a broader audience of researchers and practitioners.

References

[1] Ai B., Yang H., Shen H., and Liao X., "Computer-Aided Design of PV/Wind Hybrid System," *Renewable Energy*, vol. 28, no. 10, pp. 1491-1512, 2003. https://doi.org/10.1016/S0960-1481(03)00011-9

[2] Ardia D., Boudt K., and Catania L., "Generalized Autoregressive Score Models in R: The GAS Package," *Journal of Statistical Software*, vol. 88, no. 6, pp. 1-28, 2019. https://doi.org/10.18637/jss.v088.i06

[3] Creal D., Koopman S., and Lucas A., "Generalized Autoregressive Score Models with Applications," *Journal of Applied Econometrics*, vol. 28, pp. 777-795, 2013. https://onlinelibrary.wiley.com/doi/pdfdirect/10.1002/jae.1279

[4] Goodfellow I., Bengio Y., and Courville A., *Deep Learning*, The MIT Press, 2016. https://www.deeplearningbook.org/

[5] Guo Z., Zhao W., Lu H., and Wang J., "Multi-Step Forecasting for Wind Speed Using a Modified EMD-Based Artificial Neural Network Model," *Renewable Energy*, vol. 37, no. 1, pp. 241-249, 2012. https://doi.org/10.1016/j.renene.2011.06.023

[6] Guo Z., Zhao J., Zhang W., and Wang J., "A Corrected Hybrid Approach for Wind Speed Prediction in Hexi Corridor of China," *Energy*, vol. 36, no. 3, pp. 1668-1679, 2011. https://doi.org/10.1016/j.energy.2010.12.063

[7] Han J. and Moraga C., "The Influence of the Sigmoid Function Parameters on the Speed of Backpropagation Learning," *in Proceedings of the International Workshop on Artificial Neural Networks*, Malaga, pp. 195-201, 1995. https://doi.org/10.1007/3-540-59497-3_175

[8] Harvey A., *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*, Cambridge University Press, 2013. https://doi.org/10.1017/CBO9781139540933

[9] Jangamshetti S. and Rau V., "Site Matching of Wind Turbine Generators: A Case Study," *IEEE Energy Conversion*, vol. 14, no. 4, pp. 1537-1543, 1999. DOI:10.1109/60.815102

[10] Jursa R. and Rohrig K., "Short-Term Wind Power Forecasting Using Evolutionary Algorithms for the Automated Specification of Artificial Intelligence Models," *International Journal of Forecasting*, vol. 24, no. 4, pp. 694-709, 2008. https://doi.org/10.1016/j.ijforecast.2008.08.007

[11] Kazemi M. and Goudarzi A., "A Novel Method for Estimating Wind Turbines Power Output Based on Least Square Approximation,"

International Journal of Engineering and Advanced Technology, vol. 2, no. 1, pp. 97-101, 2012. file:///C:/Users/user/Downloads/A0704092112.pdf

[12] Krizhevsky A., Sutskever I., and Hinton G., "ImageNet Classification with Deep Convolutional Neural Networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017. https://doi.org/10.1145/3065386

[13] Lei M., Shiyan L., Chuanwen J., Hongling L., and Yan Z., "A Review on the Forecasting of Wind Speed and Generated Power," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 4, pp. 915-920, 2009. https://doi.org/10.1016/j.rser.2008.02.002

[14] Maatallah O., Achuthan A., Janoyan K., and Marzocca P., "Recursive Wind Speed Forecasting Based on Hammerstein Auto-Regressive Model," *Applied Energy*, vol. 145, pp. 191-197, 2015. https://doi.org/10.1016/j.apenergy.2015.02.032

[15] Mao M., Ling J., Chang L., Hatziargyriou N., Zhang J., and Ding Y., "A Novel Short-Term Wind Speed Prediction Based on MFEC," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 4, no. 4, pp. 1206-1216, 2016. DOI:10.1109/JESTPE.2016.2590834

[16] Nair V. and Hinton G., "Rectified Linear Units Improve Restricted Boltzmann Machines," *in Proceedings of the 27th International Conference on Machine Learning*, Haifa, pp. 1-8, 2010. https://www.cs.toronto.edu/~fritz/absps/reluICML.pdf

[17] National Renewable Energy Laboratory (NREL) 2007-2008 Western dataset. Site-id 72509, http://wind.nrel.gov/Web_nrel/2007, Last Visited, 2024. https://www.nrel.gov/grid/western-wind-data.html

[18] Pallabazzer R., "Evaluation of Wind-Generator Potentiality," *Solar Energy*, vol. 55, no. 1, pp. 49-59, 1995. https://doi.org/10.1016/0038-092X(95)00040-X

[19] Pinto T., Ramos S., Sousa T., and Vale Z., "Short-Term Wind Speed Forecasting Using Support Vector Machines," *in Proceedings of the IEEE Computational Intelligence Dynamic Uncertain Environments Symposium*, Orlando, pp. 40-46, 2014. DOI: 10.1109/CIDUE.2014.7007865

[20] Santamara-Bonfil G., Reyes-Ballesteros A., and Gershenson C., "Wind Speed Forecasting for Wind Farms: A Method Based on Support Vector Regression," *Renewable Energy*, vol. 85, pp. 790-809, 2016. https://doi.org/10.1016/j.renene.2015.07.004

[21] Shohoni V., Gupta S., and Nema R., "A Comparative Analysis of Wind Speed Probability Distributions for Wind Power Assessment of four Sites," *Turkish Journal of Electrical Engineering and Computer Science*, vol. 24, no. 6, pp. 4724-4735, 2016. https://journals.tubitak.gov.tr/elektrik/vol24/iss6/14/

[22] Thapar V., Agnihotri G., and Sethi V., "Critical Analysis of Methods for Mathematical Modelling of Wind Turbines," *Renewable Energy*, vol. 36, no. 11, pp. 3166-3177, 2011. https://doi.org/10.1016/j.renene.2011.03.016

[23] Uqaili I. and Ahsan S., "Machine Learning Based Prediction of Complex Bugs in Source Code," *The International Arab Journal of Information Technology*, vol. 17, no. 1, pp. 26-37, 2020. DOI:10.34028/iajit/17/1/4

[24] Zhang W., Wang J., Wang J., Zhao Z., and Tian M., "Shortterm Wind Speed Forecasting Based on a Hybrid Model," *Applied Soft Computing*, vol. 13, no. 7, pp. 3225-3233, 2013. https://doi.org/10.1016/j.asoc.2013.02.016

Aditya Dubey is Assistant Professor in Center for Internet of Things, Madhav Institute of Technology and Science, Gwalior, India. He has over 05 years of teaching experience. He has done his Ph.D. in the area of Big Data from MANIT Bhopal. His key research areas are Big Data, Data Mining and Image Processing. He has authored several national and international research publications. He has also authored and edited books in Computer Science.

Pradeep Yadav working as an associate professor in ITM Gwalior. He published more than 40 research papers in national and international conferences and journals. He published 6 books and 8 patents. He has an experience of more than 16 years. His research area is Image Processing, Cyber Security, Machine Learning, Big Data and Software Engineering.

Subhash Chandra Patel received his PhD degree in CSE from IIT (BHU), Varanasi in 2018 and M. Tech. degree in Information Security from the Guru Gobind Singh Indraprashtha University, New Delhi in 2010. B. Tech in CSE in 2006. Currently he is working as a Senior Assistant Professor in the School of Computer Science and Engineering at VIT University Bhopal. He has Seven years of Experience in Teaching. He published various research papers in National and international conferences and Journals also. He reviewed various paper for Transactions on Cloud Computing Journal. His research interests include Cloud Computing Security, Information Security, Internet of Things, Software Engineering and Machine Learning.