

Advanced Time Series Forecasting Models for Wind Power Prediction and Grid Stability

Vinitha Varghese ^{a*}, Dr.P.S. Divya ^b

^{a*} *Research Scholar, Department of Mathematics, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India. E-mail: vinithavarghese@karunya.edu.in, Orcid: <https://orcid.org/0009-0003-5644-513X>*

^b *Assistant Professor, Department of Mathematics, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India. E-mail: divya_deepam@karunya.edu, Orcid: <https://orcid.org/0000-0002-3836-2432>*

Abstract

Reliable wind power forecasts are necessary for ensuring the electric grid's reliability and for integrating the intermittent renewable power sources into it. Although several models for forecasting time series are already available, a methodical and empirical analysis of these models' comparison under the same operating conditions is not available. In this paper, compare Autoregressive (AR), Autoregressive Integrated Moving Average (ARIMA), ARIMA with Exogenous variables (ARIMAX), and Seasonal ARIMA with Exogenous variables (SARIMAX), four models by means of a rigorous multi-site evaluation approach with 1795 daily averaged data points for each location from four sites dispersed across a geographic area. For performance evaluation, employs Normalized Mean Absolute Error (NMAE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Squared Error (MSE). The results show that incorporating a meteorological exogenous variable provides a boost in prediction accuracy to the models. By adding meteorological information, prediction accuracy gains in stable meteorological regimes can reach 50%-70% in the reduction of the forecast error. From the relative error comparison, the SARIMAX model achieved better prediction accuracy and reliability, and it also captured the weekly cycle and fluctuations in high-value states, compared to baseline models, which achieved NMAE 0.264 at Location 4. The presented clear hierarchy in prediction accuracy may allow grid operators to make better decisions to commit units and dispatch economically based on observed evidence. In summary, linear time series models work quite well under the scenario with stable meteorological conditions, whereas site-specific volatilities demand adaptive, high-accuracy forecast models.

Keywords: Autoregressive (AR) Model, ARIMA Model, ARIMAX Model, Mean Absolute Error (MAE), Mean Squared Error (MSE), Normalized Mean Absolute Error (NMAE), Root Mean Square Error (RMSE), SARIMAX Model.

Introduction

The shift towards renewable forms of energy has been gaining momentum at a significantly quicker pace across the globe in recent decades, with wind power taking one of the most salient and fastest-growing renewable technologies. The International Energy Agency reports that the wind power potential has grown rapidly, thus playing a significant role in the global electricity supply (Castro et al., 2024). In spite of that, the irregularity and natural variability of wind power poses significant challenges to power systems operators, especially in the stability of the grid and constant supply of electricity.

Wind-power-curve forecasting models offer essential aids to grid operators because allows them to predict the trends in the power generation, unit commitment, and economic dispatch decisions, the effective and efficient use of spinning reserves, and facilitate real-time balancing activities. Besides, the enhanced accuracy of the prediction directly translates into lower operation costs (including reducing the cost of maintaining costly standby reserves and curtailment measures), and also the economic feasibility of wind-energy integration and the acceleration of the transition to sustainable low-carbon electricity systems without jeopardizing the reliability and stability of the modern power grids (Zhu et al., 2023).

Although many studies applied these models to forecast wind power individually, only a few comparative studies have been conducted so far to compare their performance under identical conditions (Bickici et al., 2025; Lydia et al., 2014). The majority of the available research considers individual applications of the models or compares only a few of them, and thus, practitioners find it challenging to identify the most suitable modeling method to use in their applications. This paper fills this research gap by conducting a systematic comparison of the AR model, the ARIMA model, the ARIMAX model, and the SARIMAX model in wind power prediction. The discussion will involve analyzing the performance of four time series models on wind power data in terms of fitting and predicting.

This paper makes the following unique contributions to the field:

- By using the same dataset, pipeline, and test intervals for all models, this research eliminates performance biases common in studies with inconsistent conditions.
- The study quantifies the accuracy gains from integrating meteorological variables, highlighting the specific value of ARIMAX and SARIMAX over temporal-only models.
- It demonstrates that SARIMAX improves residual distribution and the prediction of critical high-value states, even when aggregate error metrics are similar.
- The research establishes a location-specific model hierarchy, offering actionable recommendations for selecting models based on meteorological stability or regime volatility.

The paper consists of six sections. In Section I, emphasize the need for an accurate forecast of wind power for the stability of the grid. The literature review of wind power forecasting is done in Section II. Section III describes the methodology, where the theoretical foundation of the four models and data development are included. Section IV represents the comparative result and analyzes the impact of integration on the grid in Section V. Section VI gives concluding findings and future research areas.

Literature Review

Recent work on the integration of renewable energies has highlighted the importance of accurate wind forecasting for the purpose of stable operation and optimized dispatch. Existing literature on technical issues of integration has drawn attention to the intermittent and variable nature of wind resources, which makes it difficult to provide steady electricity power. In the past, a lot of effort had been put into time series modeling, but it is a current trend in the academic field to search for models that not only take into account linear temporal dependence but also deal with strong exogenous effects such as meteorological data.

Recent research has looked into the suitability of Autoregressive (AR) and Integrated Moving Average (ARIMA) methodologies for time series analysis. The analysis established that simplistic AR models perform well in identifying short-term patterns within stationary data, yet fall short in replicating the high variability associated with the fluctuating wind energy output. The incorporation of various input variables (e.g., temperature, pressure, wind speed) within more complex extensions, such as ARIMAX, proved to further improve prediction capability through the inclusion of relevant physical drivers influencing the environment. In addition, the inclusion of seasonal features within SARIMAX models further acknowledges the cyclical variations in wind generation; a finding that current

literature corroborates as the periodic characteristics associated with the 24h / 7 days cycles are essential modeling aspects.

Even though advancements have been made, a void exists in the literature as far as the models have been analyzed and compared side by side under similar scenarios. Most of the work done either analyzed a single model in isolation or analyzed only a few subsets of models, leaving the practitioners in the dark on what the best method is to apply in different geographical locations. In this paper, the analysis carried out would fill this vacuum. AR, ARIMA, ARIMAX, and SARIMAX models are analyzed under this study using recent methodologies on the role of preprocessing data, checking stationarity, and selection of the appropriate model using information criteria AIC and BIC for model selection.

The existing literature applies different time series models to wind power forecasting; however, these research studies seem to be dispersed and have shortcomings since models are separately examined and examined using dissimilar conditions, thereby making it hard for practitioners to make a choice on a model when it involves realistic operational deployment. A systematic, comprehensive, and comparative review among these 4 statistical models (AR, ARIMA, ARIMAX, and SARIMAX) with the same operating conditions from 4 different geographical locations is a contribution of this paper.

Table 1: Comparative Analysis of Recent Wind Power Forecasting Studies

Models Compared	Key Focus/Methodology	Limitations/Research Gap
AR, ARIMA, ARIMAX, SARIMAX	Systematic comparison under identical conditions	Lack of localized regime-specific analysis
ARIMA, SARIMA	Forecasting COVID-19 case dynamics	Non-wind power domain application
Hybrid LSTM, Fractional Time Series	Fractional modeling of inflation	Does not address wind power intermittency
Linear/Non-linear Autoregressive	Wind speed forecasting	Lacks multivariate exogenous variable integration
ARIMA	Short-term wind prediction	Limited model comparison scope

Table 1 illustrates the literature from previous work on forecasting techniques used. It shows how the wide variety of techniques has been used for wind power forecasting. The individual models, like ARIMA or the hybrid approaches, have been extensively used; however, the majority of these studies used either different testbeds or too narrow model scopes. By comparing these previous studies, it explains the choice of performing a strict, systematic, and standard approach to evaluate these four different models in the same setting and provides definite benchmarks for grid stability.

In summary, the literature appears to conclude that there is no one best-performing model but that the relative performance depends on the prevailing stationarity and weather regimes at the particular site. These latest studies validate the argument for SARIMAX performing better reliably in a wide variety of regimes and that for this to work at an operational level, it will need to be better than competing models at adapting to sudden, non-stationary changes in regime, which are still being actively explored in terms of grid reliability.

Time Series Models for Wind Power Forecasting

This study involves a comparative analysis of wind power forecasting based on four time series models, namely Autoregressive (AR), Autoregressive Integrated Moving Average (ARIMA), Autoregressive Integrated Moving Average with Exogenous variables (ARIMAX), and Seasonal Autoregressive Integrated Moving Average with Exogenous variables (SARIMAX) (Shrivastava et al., 2025; Acquah, 2017). To ascertain a statistically meaningful forecast, needed at least 30 samples; took 1459 observations on average at four distinct locations (Fanoodi et al., 2019). The suggested methodological framework has a systematic procedure comprising pre-processing of data, model specification, parameter estimation, and performance evaluation within standardized conditions (Arif et al., 2024).

Autoregressive (AR) Model: The most basic model in time series modeling is the autoregressive (AR) model, which was first introduced by Udney Yule, whereby the present observation is represented by a linear relation of the previous observations (Avrilia et al., 2026; Athiyarath et al., 2020).

The AR(p) model is defined as in equation (1):

$$w_t = \phi_0 + \phi_1 w_{t-1} + \phi_2 w_{t-2} + \dots + \phi_p w_{t-p} + \varepsilon_t \quad (1)$$

Where: w_t is the wind power at time t , ϕ_0 is the constant term, ϕ_i , $i = 1, 2, 3, \dots, p$ are autoregressive parameters, p is the order of the autoregressive model, ε_t is the error term with zero mean and constant variance.

Autoregressive (AR) model assumes that the time series is stationary, which is used for forecasting non-seasonal data, and it explains the short-term dependencies by using lagged observations. The information criteria used to arrive at the most appropriate model include Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) (Hu & Wang, 2024).

ARIMA Model: The ARIMA (p, d, q) model was introduced by Box and Jenkins, is an expansion of the AR model that includes the use of differencing to model non-stationary data, and the use of moving-average terms to model the error structure (Lydia et al., 2016).

The general form is in equation (2):

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)(1 - L^s)^D(1 - L)^d w_t = c + (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \varepsilon_t \quad (2)$$

Where: L is the lag operator, defined as $L^k w_t = w_{t-k}$, d is the degree of differencing, c is the constant term, θ_i ($i = 1, 2, \dots, q$) are moving average parameters, q is the order of the moving average component.

The model uses the AR model, Integration (I) (the reciprocal of differencing to obtain forecasts), and the MA model.

ARIMAX Model. In the ARIMAX model, exogenous variables are incorporated in order to improve the accuracy in forecasting using external information that influences the wind power generation (Rezaie-Balf et al., 2019).

The model is expressed as in equation (3):

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)(1 - L)^d w_t = c + (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \varepsilon_t + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt} \quad (3)$$

Where: x_{k_t} ($i = 1, 2, \dots, k$) are exogenous variables at time t , c is the constant term, β_i are the coefficients of exogenous variables, k is the number of exogenous variables. Exogenous variables in this study include the Wind speed, Wind direction, Atmospheric pressure, Temperature, and Humidity.

SARIMAX Model: The SARIMAX model includes seasonal components in *ARIMAX* that can capture periodic patterns of the wind power data (Ahn & Hur, 2023). The general SARIMAX (p, d, q) (P, D, Q) s model is in equation (4):

$$\phi(L)\Phi(L^s)(1 - L)^d(1 - L^s)^D w_t = c + \theta(L)\Theta(L^s)\varepsilon_t + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt} \quad (4)$$

Where: $\phi(L)$ and $\theta(L)$ are non-seasonal AR and MA polynomials, $\Phi(L^s)$ and $\Theta(L^s)$ are seasonal AR and MA polynomials, c is the constant term, P, D, Q are seasonal AR order, the seasonal differencing, and the seasonal MA order, s is the seasonal period.

Analytical Tools

After data collection and data preprocessing time series data stationarity check is done using Augmented Dickey Fuller (ADF) Test and Kwiatkowski- Philips – Schmidt – Shin (KPSS) Test. The differencing technique is applied if the data is not stationary. Once the data have been made stationary, consider the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) plot. In the case of stationary data, the values of ACF and PACF decrease rapidly. The optimal parameters for each model are evaluated using the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Hannan-Quinn Information Criterion (HQIC), and Log-Likelihood (LL). The closer the value of LL is too higher, the better fit (Babaie-Kafaki et al., 2022; Hu & Wang, 2024). Also, smaller AIC, BIC, and HQIC indicate that it fits better. Once the model has been selected, Residual plots are prepared in order to test the hypothesis that the residuals are white noise (i.e., mean = 0, constant variance) by plotting ACF and PCF.

If the ACF and PACF show no significant spikes (all auto-correlations within 95% confidence limit), then the selected model is the best. The coefficients or hyperparameters are estimated using Ordinary Least Squares (OLS) for AR Model and Maximum Likelihood Estimate (MLE) for ARIMA, ARIMAX, and SARIMAX models (Aljeddani & Mohammed, 2023). The model diagnostics or validation is done with Q-Q (Quantile Quantile) Plots (compares the observed and theoretical quantiles and tests the normality of the residuals) and Histograms (shows overall shape and spread, hence checks normality of the residuals). The evaluation of the forecast accuracy is found using Mean Squared Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Normalized Mean Absolute Error (NMAE). Lower error value means better forecast.

Mean Squared Error (MSE)

In Mean Squared Error, evaluate the average of the squared difference between actual power output and predicted power output (Huang et al., 2024). The lesser MSE is a sign of improved performance of the model. The formula for MSE is given in equation (5).

$$\text{Mean Squared Error, MSE} = \frac{\sum (w_t - \hat{w}_t)^2}{n} \quad (5)$$

Root Mean Square Error (RMSE)

RMSE is the square root of *MSE*, which is reverting the error to its original unit (watts or kilowatts). This renders it interpretable as compared to *MSE* (Cadenas & Rivera, 2010). The formula for *RMSE* is given in equation (6).

$$\text{Root Mean Square Error, RMSE} = \sqrt{\frac{\sum (w_t - \hat{w}_t)^2}{n}} \quad (6)$$

Mean Absolute Error (MAE)

The absolute difference between the predictions and the actual values are computed in MAE. It provides a direct avenue of mean prediction accuracy in wind forecasting (Li & Shi, 2010). The formula for MAE is given in equation (7).

$$\text{Mean Absolute Error, MAE} = \frac{\sum |w_t - \hat{w}_t|}{n} \quad (7)$$

Normalized Mean Absolute Error (NMAE)

NMAE expresses the error proportionately, and it is therefore easy to compare the forecasting results of different wind farms or even different turbines with different capacities (Cadenas & Rivera, 2010). The formula for NMAE is given in equation (8).

$$\text{Normalized Mean Absolute Error, NMAE} = \frac{MAE}{w_{max} - w_{min}} \quad (8)$$

Data Description

The data set consists of hourly measurements of wind power from four different geographical location, with varied weather characteristics. The collected data was pre-processed, in order to include thorough quality assessments, aimed to Detect missing values, outliers and errors. Stationary tests on the time series were done with using Augmented Dicky-Fuller test, followed by log transform and differencing on some series when needed. This ensures consistency of data with good quality features for training and robust model comparisons.

A statistical data with time stamps or arranged in a chronological manner is called a Time Series which is classified into two: stationary and non-stationary data (Fanoodi et al., 2019). The mean and variance of a stationary time series is constant over time (ArunKumar et al., 2021). The stationarity helps the model to learn relationships effectively and hence accurate forecast can be done. Also, the forecasting reliability will be the most as past patterns are more likely to recur, making predictions using historical data meaningful. So, if the data is not stationary, it should be converted into stationary data using the technique of differencing. Differencing is of two kinds: Ordinary Differencing and Seasonal Differencing. First-order ordinary differencing is denoted as w'_t , second-order ordinary differencing as w''_t and so on. Equations (9), (10) and (11) are the mathematical representation.

$$w'_t = w_t - w_{t-1} \quad (9)$$

$$w''_t = w_t - 2w_{t-1} + w_{t-2} \quad (10)$$

$$w'_s = w_t - w_{t-m} \quad (11)$$

Where, w_t is the variable value at time t

Similarly, the first-order seasonal differencing is given in equation (11), where, ' m ' is the seasonal period (which can be monthly, quarterly, yearly, etc.)

The current study uses the wind power generation information, from four different locations, which was obtained from the Data Science and machine learning platform. It provides a detailed list of hourly data from 1 January 2017 (Python Software Foundation, 2023). The data set has 43801 observations at an hourly rate over a span of 5 years. These are then summed up into daily values using a daily average, thus lowering the number to 1795 observations on an average.

Model Development Process

The key steps involved in the model development process of wind power forecasting include the collection of input data, data processing, check of stationarity, differencing where necessary, analysis of ACF- PACF data plots, determination of the best parameters, estimation of the parameters, followed by training and forecasting of wind power, lastly model diagnostics or validation is done by comparing the predictions with the test data and finally,

evaluation of the accuracy of the forecasts is carried out. Fig. 1 provides a well-defined systematic process between raw data and precise wind power forecasts.

The selection of exogenous variables, particularly wind speed, direction, pressure, temperature, and humidity, is chosen because they exhibit fundamental physical relationships to power generation. Among these variables, wind speed has a dominant influence on power generation, while other factors like temperature, pressure, and humidity affect the air density, thereby affecting the kinetic energy the turbine can harness. The inclusion of these meteorological parameters makes the models (ARIMAX and SARIMAX) capture the underlying physical environmental variables influencing fluctuations in power volatility.

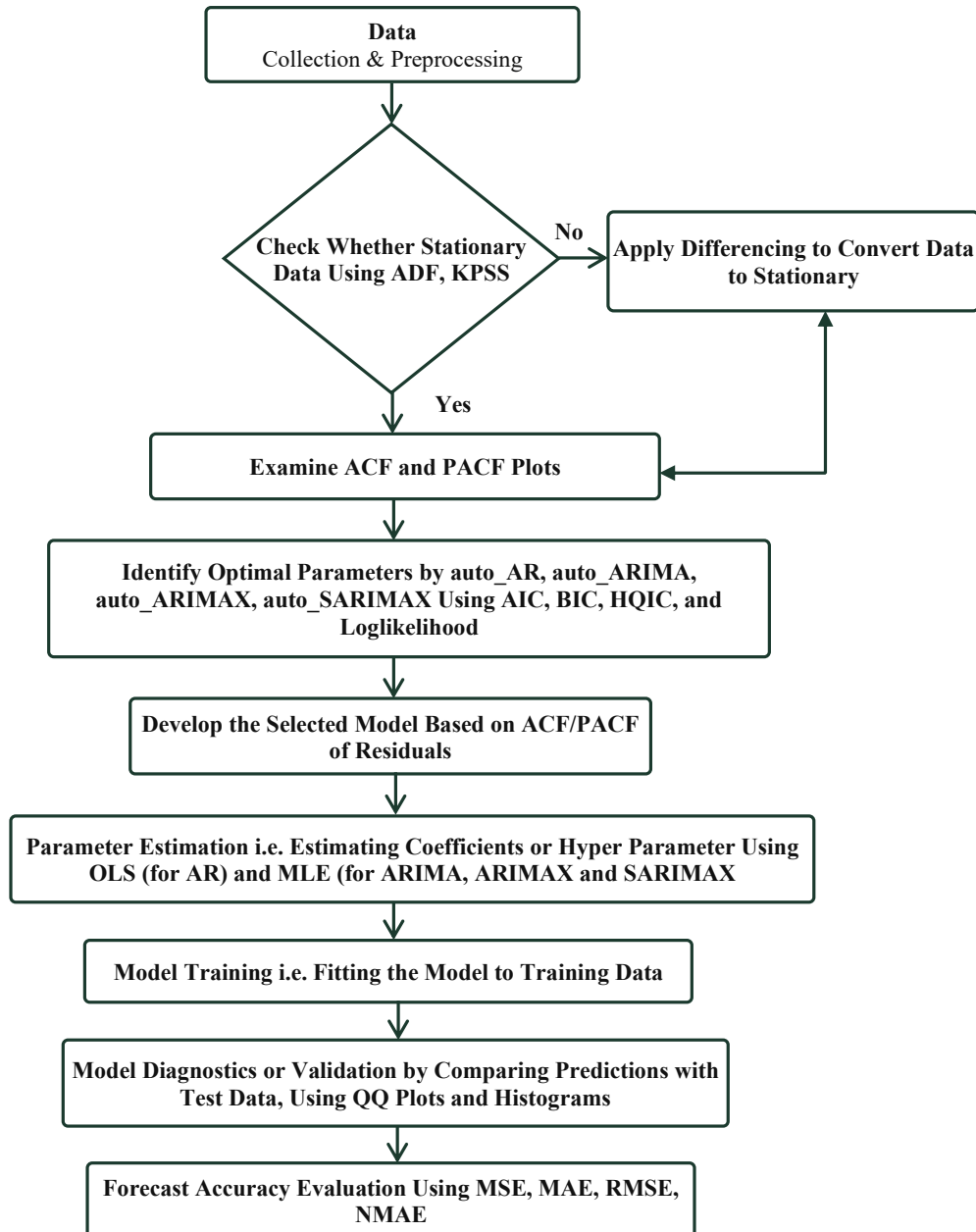


Fig. 1: Model Development Process

All models used a one-step-ahead forecasting mechanism in to generate the forecasts (Hamzaoui et al., 2024). The forecast standard errors are used to construct forecast intervals that form prediction intervals that capture the uncertainty of the predictions. 80% of the data is taken for training (01 January 2017- 31 December 2020 (1430 days)), and 20% of the data is taken for testing (01 January 2021- 31 December 2021 (365 days)).

Wind Power Forecast of Results

Implementation Details

The models were built in Python 3.12 and relied on the rich and specialized time series libraries available in the language. Data management and pre-processing were carried out in *Pandas* and *NumPy*, statistical model fitting, parameter tuning, and diagnostic testing in the *statsmodels* library, while model validation was done with *Scikit-learn* to calculate error measures. The calculation was carried out on an Intel Core i7 with 16 GB of RAM, which is adequate to run all necessary steps of parameter optimization and cross-validation on the four geographies used in this study.

The analysis was conducted on wind power time series data collected from four distinct geographical locations, say, Location 1, Location 2, Location 3, and Location 4. The independent assessment of the stationarity of each location was done using both the Augmented Dickey-Fuller (ADF) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test (Li, 2023).

Table 2: ADF & KPSS Test Report for Locations 1, 2, 3 & 4

Test	Metric	Location 1	Location 2	Location 3	Location 4
ADF	Test Statistic	-5.180	-4.360	-2.968	-8.301
	p-value	9.65x 10 ⁻⁶	0.000349	0.037965	4.02x 10 ⁻¹³
	Used Lags	23	25	24	11
	Number of observations	1801	1799	1800	1813
	Critical Values	level of significance: 1%: -3.433 5%: -2.863 10%: -2.567	level of significance: 1%: -3.433 5%: -2.863 10%: -2.567	level of significance: 1%: -3.433 5%: -2.863 10%: -2.567	level of significance: 1%: -3.433 5%: -2.863 10%: -2.567
KPSS	Test Statistic	0.0789	3.343	1.109	0.105
	p-value	0.1	0.01	0.01	0.1
	Used Lags	20	22	23	17
	Number of observations	1801	1799	1800	1813
	Critical Values	level of significance: 1%: 0.739 5%: 0.463 10%: 0.347	level of significance: 1%: 0.739 5%: 0.463 10%: 0.347	level of significance: 1%: 0.739 5%: 0.463 10%: 0.347	level of significance: 1%:0.739 5%: 0.463 10%: 0.347

Table 2 shows the results of the stationarity test in Location 1; the test statistic of the ADF test was -5.18026577, which implies that it is stationary. The same can be said of Location 2, which had ADF values of -4.36019. The KPSS test at Location 2, with a value of 3.343 and a p-value of 0.01, is a strong rejection of the null hypothesis of non-stationarity (Li, 2023). Location 3 KPSS test has a test value of 1.110, indicating that Location 3 has non-stationary or nearly non-stationary behavior. Location 4 has the most compelling evidence of stationarity of all four locations, with an ADF test value of -8.301.

After the confirmation that the location had reached stationarity, the temporal dependence structure of each location was analyzed by ACF and PACF. Fig. 2 displays the autocorrelation and partial autocorrelation functions for the stationary series at all four locations.

The slowly decaying ACF and a very sharp cut-off PACF beyond lag 1 give a sufficient indication of an AR (1) process that controls the wind power dynamics of Location 1. In Location 2, the AR (1) character is stronger and more persistent (autocorrelation of 0.55). The slightly high PACF at lag 2 justifies the testing of AR (2) specifications. Location 3 has the highest temporal persistence of all locations, with a lag-1 autocorrelation of 0.60 and a very slow decay of the ACF. This is absolutely in line with the results of the stationarity test, which marginal ADF stationarity (nearly rejecting non-stationarity at the 5% level). Location 4 has the lowest temporal persistence and the simplest correlation structure compared to any other location (Song et al., 2024; Rahim, 2017). This is right in line with the outcomes of the stationarity test.

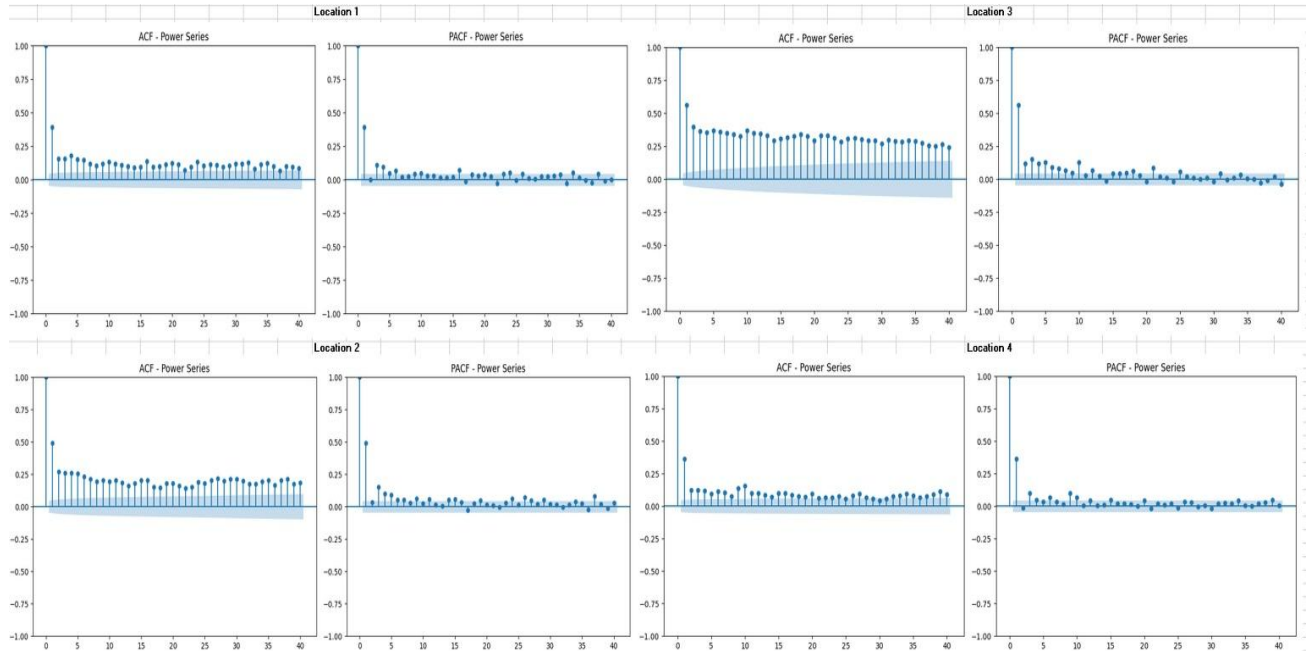


Fig. 2: ACF and PACF Plots of Wind Power Time Series for Locations 1, 2, 3 & 4

The forecasting model that gives the best fit was systematically tested with each location using Information Criteria- Loglikelihood, AIC, BIC and HQIC. The information criteria values of each model at four different locations are given in table 3, and it is evident that SARIMAX makes a relative decrease in information criteria values compared to AR, ARIMA, and ARIMAX (Li et al., 2021). The models were estimated through Ordinary Least Squares (OLS) for the AR model and Maximum Likelihood Estimation (MLE) for ARIMA, ARIMAX, and SARIMAX models in all the locations. The test period forecast has been generated on all four models in each location, and it was compared with the value of the actual power generation of wind. Fig. 3 shows the detailed comparisons of the predictions of all models with actual measurements of each location.

Table 3: Information Criteria Values of AR Model, ARIMA Model, ARIMAX Model, SARIMAX Model for Locations 1, 2, 3 & 4

Location	Metric	AR	ARIMA	ARIMAX	SARIMAX
Location 1	Log-likelihood	-4513.646	-4515.5464	-3658.3602	-3637.9575
	AIC	9033.293	9039.0928	7334.7205	7295.9151
	BIC	9049.149	9060.2376	7382.2962	7348.7152
	HQIC	9039.208	9046.9806	7352.4681	7315.6177
	No.of Observations	1459	1460	1460	1460
	Degrees of Freedom of residuals	1457	∞	∞	∞
Location 2	Log-likelihood	-4011.849	-4013.0046	-3413.2564	-3377.796
	AIC	8029.698	8034.0091	6844.5128	6775.592
	BIC	8045.554	8055.1539	6892.0885	6828.3921
	HQIC	8035.613	8041.897	6862.2604	6795.2947
	No.of Observations	1459	1460	1460	1460
	Degrees of Freedom of residuals	1457	∞	∞	∞
Location 3	Log-likelihood	-4258.855	-4258.6365	-3781.2633	-3746.0590
	AIC	8523.710	8525.273166	7580.5266	7512.1180
	BIC	8539.566	8546.4179	7628.1023	7564.9181
	HQIC	8529.625	8533.1610	7598.2742	7531.8207
	No.of Observations	1459	1460	1460	1460
	Degrees of Freedom of residuals	1457	∞	∞	∞
Location 4	Log-likelihood	-4216.127	-4218.2333	-3015.8329	-3005.4926
	AIC	8438.254	8444.4665	6049.6657	6030.9851
	BIC	8454.111	8465.6113	6097.2415	6083.7852
	HQIC	8444.169	8452.3544	6067.4134	6050.6878
	No.of Observations	1459	1460	1460	1460
	Degrees of Freedom of residuals	1457	∞	∞	∞

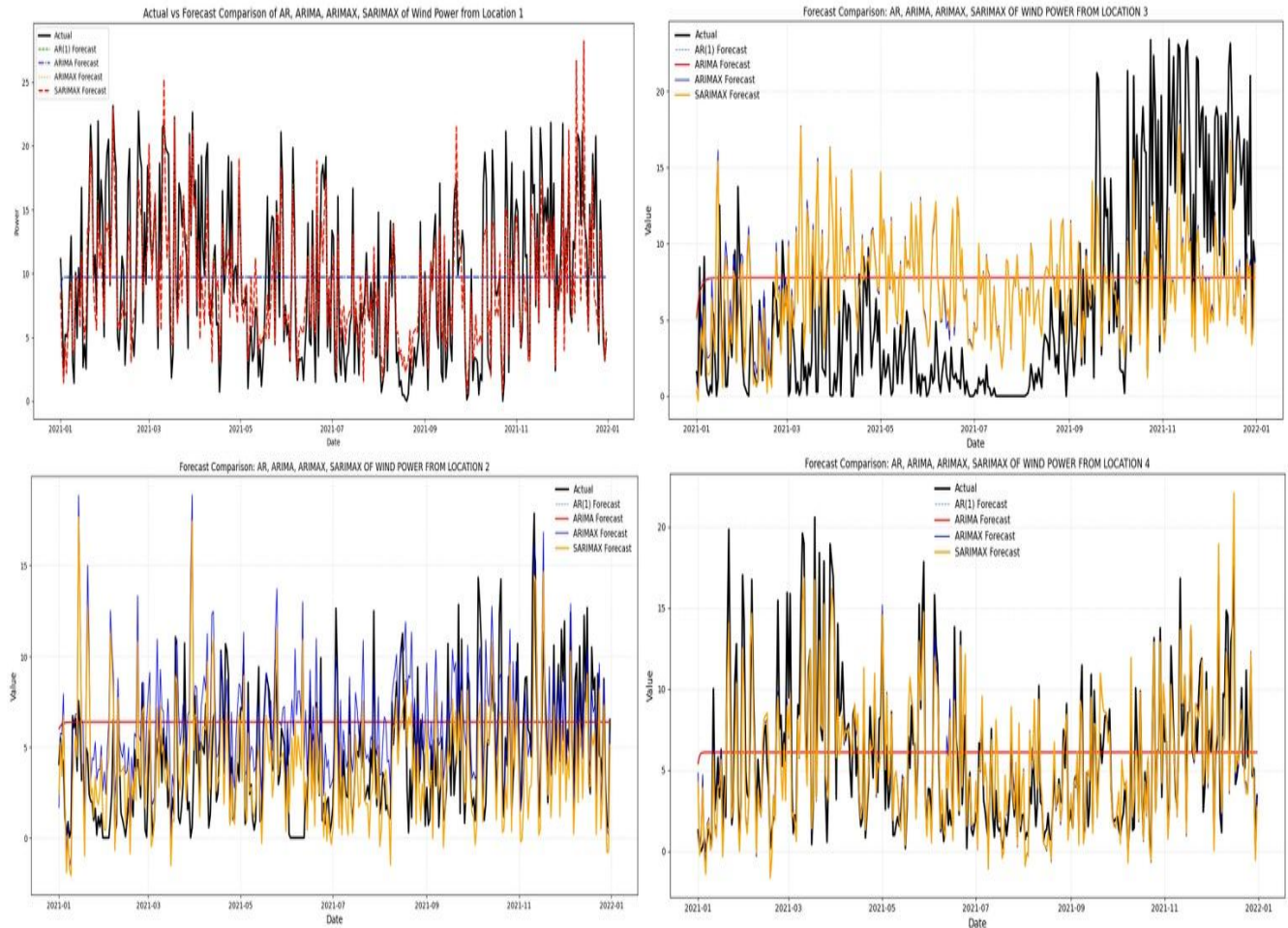


Fig. 3: Actual vs Forecast Comparison of AR, ARIMA, ARIMAX, SARIMAX of Wind Power from Locations 1, 2, 3 & 4

Fig. 3 (time series plot) of the actual power generation in the wind (black line), forecasts of AR (green line), ARIMA (blue line), ARIMAX (orange line) and ARIMAX (red dashed line) models at Location 1 shows that there are marked performance differences between the forecasts across the forecasting period. The real wind power has a high level of variability and oscillates widely between a range of 0-5 MW at the low-generation stage and 20-27 MW at the high-generation stage which reveals the nature of intermittency and volatility of wind resources. Comparative Model Ranking: Under visual analysis of the forecast accuracy over the time series, the models can be ranked as follows in Location 1: SARIMAX has the highest overall performance, highest peak, best responsiveness to rapid changes, and best tracking over high periods of variance. It has an obvious added value in the weekly seasonality. ARIMAX is the second-best, the performance is much better in terms of amplitude matching without a major difference in time-only models. Well able to model weather-driven dynamics but not as seasonal as SARIMAX. ARIMA model is average and with enhanced error correction as compared to AR does not forecast extremes sharply because of the lack of meteorological data. AR shows least robust performance on the one hand, too much smoothing, always underestimates the peaks, and responsiveness to the rapid change is poor.

The Location 2 time series plot, of actual wind power (black line), AR (green line), ARIMA (red line), ARIMAX (blue line), and SARIMAX (orange line) indicates entirely different dynamics than that of Location 1. Real wind power at Location 2 is characterized by significantly lower values of overall generation, which intermittently switch between 0-10 MW over most of the time series with periodic variants of 12-17MW. The most notable one is the dramatic growth in the generation and volatility in the last period (since the end of November 2021), in which actuals rapidly rise to record levels, often reaching 15 MW and even up to 17-18 MW. This change of regime in the late period forms a dire forecasting problem as can be observed in the difference between the actual and predicted traces over this critical period.

Comparison of the models based on a visual evaluation of the whole time series: According to visual inspection, the models are ranked in the following way in Location 2: ARIMA and ARIMAX (tie) have best performance, both with good tracking of the early/middle periods, then some, and partial tracking of the late-period peak. The fact that the performance is almost identical shows that there is very little contribution of the exogenous variables, making ARIMA perhaps the best choice. SARIMAX shows maximum complexity, but highly unpredictable behaviour and very large errors around the critical late period, yet overfitting in all periods by being too volatile or unstable seasonal behaviour. AR is totally inadequate and disastrous failures in the high-generation events and failure to capture the time dynamics.

Patterns of bifurcated generation exhibiting distinct forecasting problems are dramatically varied in the time series plot of Location 3. The actual wind power (black line) shows the extreme regime change that is around the middle of the year (around July-August 2021), and the sharp transition between the continuous flow generation (mainly 0-10 MW with some sporadic peaks to 15-18 MW during January-August 2021) and the radically high generation (mainly 10-25 MW during September 2021-December 2021). This sudden and extended change of the regime is the strongest non-stationarity of all sites and poses the greatest challenge to model malleability and predictive power. According to the visual analysis, with more weight on the operationally important high-generation late period, the relative ranking of the models is as follows in Location 3: SARIMAX is the best performer in the critical late period with better tracking properties of the high-generation regime, the best amplitude convergence, responsive fluctuation capturing, and quick adaptation to regime change. The weekly seasonal aspects are valuable beyond measure to this location. ARIMA and ARIMAX (tie), mediocre performance with good early-period tracking and terrible regime transition failures and idiosyncratic late-period underestimation. The virtually equal performance implies the zero contribution of the exogenous variables. AR, prejudiced performance according to the common AR behavioural patterns.

The time series plot of Location 4 indicates actual wind power (black line) with a moderate to a high level of variability through the forecasting period and with the generation mostly in the range of 0-15 MW and a high frequency of exceeding 15-22 MW. The real series is very volatile over a short time frame with quick changes in value, indicating that it is very sensitive to the changing conditions. Unlike in Location 3, where there is dramatic regime change, or in Location 2, where there is a spurt in the late period, Location 4 shows relatively steady statistical characteristics across the entire time series, with the same variability characteristics over the series. This stationarity makes the forecasting task simpler, and it also points out the variation in the volatility but not unstable dynamics in different model specifications. Visual assessment of the whole time series, with the focus on difficult high-variability intervals, has the following rankings of the models in Location 4: SARIMAX is the best with the best tracking of most of the periods, the great amplitude tracking following during the peaks, the responsive tracking of the sharp changes, and the definite advantage during the operationally critical high-volatility late period. Weekly seasonal products are contributing a significant value. ARIMAX performs moderately, but apparently it offers a gain over pure ARIMA, but difficult to display on a graph. It reacts more rationally than ARIMA to the middle-period systematic bias episode. ARIMA, the performance of the variable is reasonable plus tracking in certain periods, but has problematic over-forecasting using systems in the middle period, which implies specification or parameter stability problems. The horizontal movement pattern indicates basic modeling issues. AR was unstable with continued smoothing over and under estimating of peaks along the time series.

Table 4: Evaluation Metrics for Locations 1, 2, 3 & 4

Location	Model	MSE	MAE	RMSE	NMAE
Location 1	AR	38.292	5.359	6.188	0.551
	ARIMA	38.294	5.360	6.188	0.551
	ARIMAX	11.324	2.656	3.365	0.273
	SARIMAX	11.334	2.651	3.366	0.272
Location 2	AR	15.618	3.416	3.952	0.770
	ARIMA	15.603	3.414	3.950	0.769
	ARIMAX	13.313	2.767	3.648	0.623
	SARIMAX	9.549	2.256	3.090	0.508
Location 3	AR	44.530	5.862	6.673	1.011
	ARIMA	44.531	5.865	6.673	1.011
	ARIMAX	40.579	5.224	6.370	0.901
	SARIMAX	40.456	5.202	6.360	0.897
Location 4	AR	20.987	3.719	4.581	0.651
	ARIMA	20.984	3.718	4.580	0.651
	ARIMAX	4.1876	1.511	2.046	0.267
	SARIMAX	4.112	1.529	2.027	0.264

Comparison of the evaluation metrics of all the places shows the obvious differences in performance of the place and its particular features. In the case of Location 1, AR and ARIMA are not suitable to do operational forecasting, where gives forecast errors that are nearly twice that of models informed by meteorology. Nevertheless, with the error reduction of 50-70 percent being dramatic with the addition of meteorological variables, to obtain the acceptable level of forecast accuracy, operational systems at Location 1 will need to incorporate weather information. SARIMAX demonstrated a little higher peak tracking compared to ARIMAX, indicating that SARIMAX can possibly have slight advantages in extreme values which are not adequately captured by aggregate error measures.

In case of Location 2, the metrics of evaluation should be viewed carefully. AR and ARIMA are obviously insufficient, with the same bad performance that supports simple AR instead of complex ARIMA, provided that meteorological variables are not accessible. SARIMAX has the highest aggregate performance values and it will be chosen solely due to MSE/MAE.

The metrics of evaluation of Location 3 show the most impressive patterns of model performance in comparison with all other locations with the vivid differences in the performance of both models (temporal and seasonal). The ideal match of AR and ARIMA suggests that sophistication in temporal modeling offers no practical value at Location 3, which confirms the results of the visual analysis according to which the dynamics of Location 3 can hardly be explained through pure time structure. The extreme values of NMAE support the fact that the forecasting environment in Location 3 is the most difficult, and the relative performance of SARIMAX is operationally critical, although absolute performance may still be imperfect.

Location 4, table 4, evaluation metrics reflect trends between the high levels of dramatic meteorological variables and the low levels of differentiation of the temporal models of Locations 1 and 3, respectively. In the case of Location 4, the evaluation metrics give clear-cut information that AR and ARIMA are not suitable, with equally poor performance (NMAE = 0.651, which is almost 2.5 times lower than ARIMAX/SARIMAX). Simple AR (1) is validated by the perfect AR-ARIMA equivalence when conditions are not meteorological variables. The great result in Location 4 shows that the meteorological-power relationships must be stable and of strong strength to result in highly accurate operational forecasts by standard ARIMAX/SARIMAX frameworks, which proves the underlying method and the location-specific forecasting success.

Discussion

Grid Stability/Integration

The involvement of wind power in contemporary electrical grids creates several tough issues for the operators. A significant part of the trouble comes from the natural randomness and also the time variation of the wind. These findings show that precise prediction is critical for optimal financial dispatching as well as system commitment in order to stabilize the power system, reduce the requirement for standby power, and manage the outcomes of rapid power variation. The study has concluded that meteorological exogenous factors included in a time-series model (e.g. ARIMAX as well as SARIMAX) make highly accurate prognoses, which support real-time balancing actions and spinning reserve usage.

Practical Recommendations

The practical solution for grid managers and operators of wind farms should focus on using site-specific meteorologically sensitive forecast modeling schemes. For locations where the meteorology does not change regimes suddenly (which have a predictable pattern), traditional ARIMAX or SARIMAX will offer enough accuracy to operate effectively. But at the sites where sudden regime shifts or volatile generation profiles may occur, location-specific adaptive models that focus more on extremes and regime shifts rather than just average forecast error are beneficial. The seasonal improvements that found indicate that it is also important to incorporate weekly patterns to enhance overall system reliability and residual distribution.

Limitations and Future Research

Despite these time series linear models demonstrating impressive performance under ideal conditions, there are known limitations that this paper implicitly accepts. Predicting these so-called outliers (e.g., extreme events) as well as dealing with non-stationary rapid shifts in regimes is an aspect that traditional models cannot adequately tackle, especially where the dependency between meteorological and power generation values is not strong. However, future research can explore these challenges by developing ensemble regressor/deep neural net hybrid models that integrate statistical frameworks into powerful ML models. Research should also investigate the benefits of high-resolution wind speed data and how changes in climate may affect power curve stability in relation to the grid's long-term integration.

Conclusion

The research presents an extensive comparison of different time series models, AR, ARIMA, ARIMAX, and SARIMAX, to forecast wind power at four geographically distinct locations. Analyzing 1795 observations through rigorous testing, the research illustrated that meteorological stability and the characteristics of the observed regime impact the reliability of model results. The most crucial difference from a statistical point of view among the models is the presence or absence of external weather parameters. In this case, models with weather conditions are able to beat time-only models significantly, with SARIMAX reaching the best forecasting result. SARIMAX results achieved impressive accuracy and were highly reliable from the point of view of average accuracy in stable weather conditions, as best shown in location 4, where a NMAE of 0.264 was attained, and from a statistical point of view, with its result for residual distributions, forecasting of high state value, which for a grid operator is highly important when managing reserves. The models' ability to maintain low RMSE values even during fluctuating atmospheric conditions underscores their reliability for operational-grade forecasting. Maintaining low RMSE across varying atmospheric conditions testifies to the suitability of models for real-time operations and forecasts. The results demonstrated a hierarchy for the best modeling practice, validating that SARIMAX is the most predictable and robust model for the grid. This study identified the limitation of statistical linear models with rapid non-stationary changes, where the model does not mitigate the loss of accuracy well; thus, future research ought to take the leap toward hybrid models integrating classical statistical concepts with Machine learning techniques, such as ensemble regressors and deep learning network (LSTM, Transformers) which are able to capture complex weather behaviors, including non-linearity and high variability. The future research also aims to explore the climate variations on the long-time scale's effect to the power curve stability as well as incorporating high-resolution satellite met data in the forecasts horizon to get more accurate results; with such works, grid managers can obtain an empirical guideline to deploy the best models that will minimize the costs of renewable integration, and secure technical and economic stability of renewable energy integration.

References

- Ahn, E., & Hur, J. (2023). A short-term forecasting of wind power outputs using the enhanced wavelet transform and arimax techniques. *Renewable Energy*, 212, 394-402. <https://doi.org/10.1016/j.renene.2023.05.048>
- Avrilia, K., Yuniarti, D., Nurmayanti, W. P., Fathurahman, M., & Wahyuningsih, S. (2026). Implementasi Model Hybrid Autoregressive Fractionally Integrated Moving Average-Neural Network (ARFIMA-NN) pada Peramalan Indeks Harga Saham Gabungan. *VARIANSI: Journal of Statistics and Its application on Teaching and Research*, 8(1), 31-48. <https://doi.org/10.35580/variansiunm487>
- Aljeddani, S. M., & Mohammed, M. A. (2023). Parameter estimation of a model using maximum likelihood function and Bayesian analysis through moment of order statistics. *Alexandria Engineering Journal*, 75, 221-232. <https://doi.org/10.1016/j.aej.2023.05.079>
- Bickici, B., Beyaztas, B. H., Yaseen, Z. M., Beyaztas, U., & Kahya, E. (2025). Streamflow Intervals Prediction Using Coupled Autoregressive Conditionally Heteroscedastic with Bootstrap Model. *Journal of Flood Risk Management*, 18(1), e70009. <https://doi.org/10.1111/jfr3.70009>
- ArunKumar, K. E., Kalaga, D. V., Kumar, C. M. S., Chilkoor, G., Kawaji, M., & Brenza, T. M. (2021). Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Average (SARIMA). *Applied soft computing*, 103, 107161. <https://doi.org/10.1016/j.asoc.2021.107161>
- Babaie-Kafaki, S., Aminifard, Z., & Ghafoori, S. (2022). Nonmonotone diagonally scaled limited-memory BFGS methods with application to compressive sensing based on a penalty model. *Applied Numerical Mathematics*, 181, 618-629. <https://doi.org/10.1016/j.apnum.2022.07.008>
- Cadenas, E., & Rivera, W. (2010). Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA-ANN model. *Renewable Energy*, 35(12), 2732-2738. <https://doi.org/10.1016/j.renene.2010.04.022>
- Castro, V., Georgiou, M., Jackson, T., Hodgkinson, I. R., Jackson, L., & Lockwood, S. (2024). Digital data demand and renewable energy limits: Forecasting the impacts on global electricity supply and sustainability. *Energy Policy*, 195, 114404. <https://doi.org/10.1016/j.enpol.2024.114404>
- Fanoodi, B., Malmir, B., & Jahantigh, F. F. (2019). Reducing demand uncertainty in the platelet supply chain through artificial neural networks and ARIMA models. *Computers in biology and medicine*, 113, 103415. <https://doi.org/10.1016/j.combiomed.2019.103415>

- Hamzaoui, H., Moussa, F. D., & El Matouat, A. (2024). Order estimation for autoregressive models using criteria based on stochastic complexity. *Kuwait Journal of Science*, 51(4), 100251. <https://doi.org/10.1016/j.kjs.2024.100251>
- Hu, Y., & Wang, Y. (2024). Evaluating statistical homogeneity of cone penetration test (CPT) data profile using auto-correlation function. *Computers and Geotechnics*, 165, 105852. <https://doi.org/10.1016/j.compgeo.2023.105852>
- Huang, J., Yuan, C., Boland, J., Guo, S., & Liu, W. (2024). One-step ahead short-term hourly global solar radiation forecasting with a dynamical system based on classification of days. *Renewable Energy*, 237, 121639. <https://doi.org/10.1016/j.renene.2024.121639>
- Arif, E., Herlinawati, E., Devianto, D., Yollanda, M., & Permana, D. (2024). Hybridization of long short-term memory neural network in fractional time series modeling of inflation. *Frontiers in big Data*, 6, 1282541. <https://doi.org/10.3389/fdata.2023.1282541>
- Li, G., & Shi, J. (2010). On comparing three artificial neural networks for wind speed forecasting. *Applied Energy*, 87(7), 2313-2320. <https://doi.org/10.1016/j.apenergy.2009.12.013>
- Li, S. H. (2023). Effect of nonstationary extreme wind speeds and ground snow loads on the structural reliability in a future Canadian changing climate. *Structural Safety*, 101, 102296. <https://doi.org/10.1016/j.strusafe.2022.102296>
- Li, Y., Zhang, Q., Wang, L., & Liang, L. (2021). An AIC-based approach to identify the most influential variables in eco-efficiency evaluation. *Expert Systems with Applications*, 167, 113883. <https://doi.org/10.1016/j.eswa.2020.113883>
- Lydia, M., Kumar, S. S., Selvakumar, A. I., & Kumar, G. E. P. (2014). A comprehensive review on wind turbine power curve modeling techniques. *Renewable and Sustainable Energy Reviews*, 30, 452-460. <https://doi.org/10.1016/j.rser.2013.10.030>
- Lydia, M., Kumar, S. S., Selvakumar, A. I., & Kumar, G. E. P. (2016). Linear and non-linear autoregressive models for short-term wind speed forecasting. *Energy conversion and management*, 112, 115-124. <https://doi.org/10.1016/j.enconman.2016.01.007>
- Song, J., Cheng, Q., Bai, X., Jiang, W., & Su, G. (2024). LSTM-based deep learning model for financial market stock price prediction. *Journal of Economic Theory and Business Management*, 1(2), 43-50. <https://doi.org/10.5281/zenodo.10940654>
- Rahim, M. (2017). Wind Power Generation Data—Forecasting. <https://www.kaggle.com/datasets/mubashirrahim/wind-power-generation-data-forecasting>
- Python Software Foundation, *Python*, ver. 3.12 [Computer software], 2023. <https://www.python.org>
- Rezaie-Balf, M., Fani Nowbandegani, S., Samadi, S. Z., Fallah, H., & Alaghmand, S. (2019). An ensemble decomposition-based artificial intelligence approach for daily streamflow prediction. *Water*, 11(4), 709. <https://doi.org/10.3390/w11040709>
- Shrivastava, R., Iyer, I. S., & Batra, A. (2025). Short term wind prediction with an ARIMA model. *Wind Energy and Engineering Research*, 100020. <https://doi.org/10.1016/j.weer.2025.100020>
- Acquah, D. G. H. (2017). The effect of outliers on the performance of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship. *Russian Journal of Agricultural and Socio-Economic Sciences*, (5), 32-37. <https://doi.org/10.18551/rjoas.2017-05.05>
- Athiyarath, S., Paul, M., & Krishnaswamy, S. (2020). A comparative study and analysis of time series forecasting techniques. *SN Computer Science*, 1(3), 175. <https://doi.org/10.1007/s42979-020-00180-5>
- Zhu, C., Bao, G., Xu, R., Song, Z., & Liu, Y. (2023). Low-carbon economic analysis of a virtual power plant with wind and solar power considering the integrated flexible operation mode of a carbon capture thermoelectric unit. *International Journal of Greenhouse Gas Control*, 130, 104011. <https://doi.org/10.1016/j.ijggc.2023.104011>