Forecasting of Short-Term and Long-Term Wind Speed of Ras-Gharib Using Time Series Analysis

Osama. A. El-Kashty (iD), Ahmed. A. Daoud (iD), E. E. El-Araby (iD)

Department of Electrical Engineering, Faculty of Engineering, Port Said University, Egypt

(osama.alkashty@eng.psu.edu.eg, a.daoud@eng.psu.edu.eg, elsaid.elaraby@eng.psu.edu.eg)

Corresponding Author: E. E. El-Araby, Department of Electrical Engineering, Faculty of Engineering, Port Said University, Egypt

Tel: +20 1061460065,

elsaid.elaraby@eng.psu.edu.eg

Abstract- The integration of renewable energy sources (RES) into the power grid has significantly grown in response to the rise in energy demand. Since wind energy tends to be intermittent and is continuously incorporated into the grid, it is crucial to properly forecast wind speed to maintain the reliability of the power system and the balance between supply and demand. In this study, using R-Studio software, daily and monthly data of wind speed of Ras Gharib-Egypt were modeled using statistical models (time series forecasting methods). The best parameters related to each method that result in the best model are identified through a careful examination of various values for each parameter. Then, using various error metrics calculated for each model, a process of model assessment is used to determine the best models among the investigated forecasting techniques. The precise models selected are extremely important because they can be employed to address both short-term operating issues and long-term planning concerns in the power system to obtain more trustworthy results over various timescales.

Keywords- Wind speed, modelling, forecasting, statistical models, time-series analysis, R-Studio.

NOMENCLATURE

Abbreviations

RES	Renewable Energy Sources	MA	Moving Average
BOO	Build-Own-Operate	ARMA	Auto-Regressive Moving Average
NWP	Numerical Weather Predictions	ARIMA	Auto-Regressive Integrated Moving Average
AI	Artificial Intelligence	SARIMA	Seasonal Auto-Regressive Integrated Moving Average
ANN	Artificial Neural Networks		
SVM	Support Vector Machine	UAE	United Arab Emirates
AR	Auto Regressive	RF	Random Forests
SSA	Singular Spectrum Analysis	LM	Lagrange Multiplier
MAPE	Mean Absolute Percentage Error	ARCH	Auto-Regressive Conditionally Heteroscedastic
RMSE	Root Mean Square Error	ANFIS	Adaptive Neuro-Fuzzy Inference Systems
MAE	Mean Absolute Error	ACF	Autocorrelation Function
ME	Mean Error	MLR	Multiple Linear Regression
MSE	Mean Square Error	PACF	Partial Autocorrelation Function
NARX	Non-linear Auto-Regressive network with Exogenous inputs	STL	Seasonal-Trend-Level
		ES	Exponential Smoothing
MODWT	Maximal Overlap Discrete Wavelet Transform	SES	Single Exponential Smoothing
GARCH	Generalized Auto-Regressive Conditionally Heteroscedastic	DES	Double Exponential Smoothing
		TES	Trible Exponential Smoothing
ETS	Error-Trend-Seasonal	AIC	Akaike Information Criterion
TBATS	Trigonometric-Seasonal, Box-Cox Transformation ARMA residuals, Trend, and Seasonality	SSE	Sum of Squared Errors
		ADF	Augmented Dickey-Fuller

KPSS	Kwiatkowski-Philips-Schmidt-Shin
PR-value	Probability value
MARS	Multivariate adaptive regression splines
WMA	Weighted Moving Average
SEATS	Seasonal Extraction in ARIMA Time Series
ARIMAX	Autoregressive integrated moving average with exogenous variables
SARIMAX	Seasonal autoregressive integrated moving average with exogenous variables

Variables and Greek Letters

p	Order of autoregression of the ARIMA method
q	Order of moving average of the ARIMA method
d	Order of differentiation of the ARIMA method
P	Number of previous observations of the seasonal part of the SARIMA method
Q	Number of previous errors of the seasonal part of the SARIMA method
D	Order of differentiation of seasonal part of the SARIMA method
m	Length of the period of the seasonality of a time series
$y(t)$	Current observation of a given time series a time t
φ_i	Autoregression coefficient of the ARIMA method
θ_j	Moving average coefficient of the ARIMA method
$\varepsilon(t)$	White noise term of a forecast equation at time t
α_l, β_l	Fourier coefficients of the Dynamic harmonic regression method

k	Number of Fourier sin and cos pairs Dynamic harmonic regression method
c	Intercept of time-series of the Dynamic harmonic regression method
Ut	ARIMA error of the Dynamic harmonic regression method
$\hat{S}(t)$	Seasonal component of the STL decomposition method
$\hat{A}(t)$	Seasonally adjusted component of the STL decomposition method
$\hat{y}_{t+h\|t}$	Forecasted value at an h-step-ahead forecast of time t of the ES method
l_t	Level smoothing equation of the ES method
b_t	Trend smoothing equation of the ES method
s_t	Estimate of the seasonal part at time t of the ES method
k_1	Integer part of (h-1)/m to ensure that the seasonal part used for forecasting comes from the final period of sample of the ES method
λ	Box-cox transformation value in the TBATS method
Db	Binary variable which values donated if the trend is damped or not, in the TBATS method
T	Number of observations used for estimation of AIC
z	Number of predictors in the model for calculating AIC
N	Number of fitted points used to calculate error metrics
$y_t, \hat{y}_t$	Actual and forecasted values used to calculate error metrics

1. Introduction

As the power demand is increasing, innovation and expansion of renewable sources of energy are important keys to maintaining a sustainable level of energy. Producing electricity using (RES) rather than fossil fuels (oil, coal, and natural gas) decreases greenhouse gas emissions and helps treat climate change. Egypt has an ambitious plan to raise RES to 42% of the total power by 2035, with wind and solar generation projected to grow at the greatest rates. According to the International Trade Administration, U.S. Department of Commerce [1], with a wind speed of 10.5 m/sec on average, Egypt receives great wind in the Gulf of Suez. Egypt's capacity to produce wind energy is anticipated to increase to 7 GW by 2022, making it a significant component of the country's renewable energy. In the Gulf of Suez, a 540 MW project is being built, and a 580 MW project is in the funding stage. Additionally, the 262.5 MW Ras Gharib wind farm project, which will power approximately five thousand homes, was completed in December 2019. In this wind farm project, the Build-Own-Operate (BOO) paradigm was employed for the first time. Furthermore, Lekela's West Bakr Wind Farm will produce 250 MW of sustainable energy. Currently, the Egyptian government is working to upgrade and repower the wind farms in Zaafarana and Gabal El Zeit in the Gulf of Suez. According to the Minister of Electricity and Renewable

Electricity, preliminary estimates showed that Zaafarana projects may produce much more energy and have increased efficiency at a far cheaper cost with the rehabilitation of some turbines.

Since the outputs of wind turbines are intermittent, integrating them into the electrical grid becomes challenging due to their uncertainty and variability, which could have negative impacts on the network's performance (power balance and operating constraints). This situation might lead the utilities to generate additional power at the expense of costs [2]. With the use of storage systems, demand response programs, and operating reserves, the impact of uncertainty and fluctuation of wind speed can be alleviated. All these techniques, however, depend on the precise prediction of wind speed. Since wind speed and the outputs of wind energy are directly correlated, accurate forecasting of wind speed is essential [3].

For predicting wind speeds, several techniques are used. These techniques include the persistence method, physics-based techniques, statistical techniques, spatial correlation models, artificial intelligence (AI) techniques, and hybrid techniques [4]. Persistence methods assume that the variation between the present value and the future value is very tiny and may be ignored, so the forecast value is the same as the current value.

Physical techniques rely on numerical weather predictions (NWP), which use weather-predicted data like temperature and pressure. In this approach, a lot of physical considerations are used for the most accurate forecasts [5]. Statistical models directly use historical data to make forecasts by finding the characteristics of the measured data. As a result, because historical data are used as the base for forecasting, it might generate better and more accurate predictions. In general, physical strategies (meteorological methods) are technologies based on orography, real-time numerical weather forecasts (NWP), and atmospheric data to predict wind speed. Based on initial conditions, these models produce predictions that help solve the complicated numerical system. However, the application of these models to short-term wind speed predictions is constrained by two fundamental issues: (1) It's not always easy to obtain comprehensive information about the characteristics of wind farms, and (2) Increasing geographical resolution and continuously updating environmental data demand significant computational time. In contrast to physical strategies, statistical strategies, also called time series strategies, are superior in the field of short-term forecasting due to the sufficient use of historical data. Statistical models are not dependent on physical understanding because they are based on the internal relationships between historical data. Based on the assumption that there are no interfering components within the input variables, prediction errors can be reduced when statistical and mathematical models are built. Spatial correlation models are based on estimating the wind speed at a specific location using the wind speed measurements of a neighboring site [6]. AI approaches employ a combination of machine learning techniques to generate forecasts, it includes artificial neural network (ANN), fuzzy logic methods, and support vector machine (SVM) [7-12]. Hybrid methods combine several different methods to take full advantage of each approach's obvious benefits and enhance forecasts [13]. Using several time-series models, statistical methods are employed in this paper. This technique is employed because it is efficient and has been applied to numerous sites worldwide on various forecast horizons [14–29].

-Literature Review

In this article, the statistical method is applied using different time-series methods. Time-series method is used because of its effectiveness and because it has been applied in several locations in the world at different time scales, including short-term and long-term. Several publications categorize wind power forecasts on different time scales. The vast majority of publications suggest that forecasting models can be categorized into four groups based on their time horizons: very short-term (0–30 minutes), short-term (30–6 hours), medium-term (6–1 day), and long-term (>1 day). However, these four categories can be further divided into short-term and long-term, as short-term wind power prediction occurs between a few minutes and a day in advance, whereas long-term forecasting occurs multiple days to a year or longer in advance. References [14], [17–18], [20–24], and [26], studied short-term wind speed forecasting time-series methods; references [16], [19], and [25] studied long-term wind speed forecasting time-series methods; and references [15], [27–28], and [29] studied long-term wind speed forecasting time-series methods.

In [14], a detailed statistical analysis is done using the moving average (MA), auto-regressive (AR), auto-regressive moving average (ARMA), auto-regressive integrated moving average (ARIMA), and seasonal auto-regressive integrated moving average (SARIMA). These methods are used to forecast wind speeds over a period of 15 years (2000-2014) in two locations in Rajasthan. This demonstrates empirically why SARIMA is the most effective model for predicting wind speed. In [15], four forecasting models are used and compared to forecast hourly and weekly wind speeds in United Arab Emirates (UAE). The first two of these models, ARIMA and SARIMA models, are traditional statistical techniques. The other two models, ANN, and Singular Spectrum Analysis (SSA) models are taken from the field of machine learning. In order to forecast monthly wind speed across the southern coast of the Mexican state of Oaxaca, the ARIMA and ANN techniques are compared in [16]. In [17], the ARIMA model for forecasting hourly and daily wind speeds in Riga (the capital of Latvia) is discussed. MAPE (mean absolute percentage error), RMSE (root mean square error), and MAE (mean absolute error) are used to check the model's forecasting performance. In [18], a comparison is made between the linear (ARIMA) and the nonlinear autoregressive exogenous artificial neural network (NARX) for hourly wind speed forecasting in two different sites in Mexico. The primary objective was to examine how different meteorological variables affected the performance of the multivariate model for predicting wind speed compared to the highly effective univariate linear model. The NARX model gave better results. In [19], monthly wind speed forecasting at nine meteorological stations in five different regions of Turkey has been studied. Forecasting is done using various time series analysis methods (the moving average method, the exponential smoothing method, the exponential smoothing with trend method, and the exponential smoothing with trend and seasonality method). The best forecasting model is then chosen based on statistics such as MAE, RMSE, MAPE, mean error (ME), and mean square error (MSE). In [20], this study assessed the effectiveness of a variety of time series models for predicting daily wind speed in England, including BATS, TBATS, Holt's Linear Trend, and ARIMA models. To construct and assess the reliability of the forecasting models, the wind data set is divided into train and test subsets. Using the ME, RMSE, MAE, MASE, and autocorrelation functions, the models developed for training data are evaluated for their goodness of fit. The BATS model was chosen because it had the smallest RMSE value for testing data. Moreover, the RMSE values of the ARIMA model and the BATS model are extremely similar. In [21], a novel hybrid model for short-term wind speed forecasting is presented in New Zealand, integrating the maximal overlap discrete wavelet transform (MODWT) with an ARIMA. The proposed method was compared with various methods using MAE, RMSE, and MAPE error metrics. In [22], a hybrid approach based on SARIMA and neural networks is proposed for hourly wind speed forecasting. The proposed approach is examined using historical data from two real-world locations in Brazil and compared with different popular methods to check its

accuracy. In [23], the daily wind speed and the probability density function of generated wind power at five U.K. wind farm locations are predicted using the ARMA-generalized autoregressive conditionally heteroscedastic (GARCH) model and the autoregressive fractionally integrated moving average (ARFIMA)-GARCH model. In [24–25], MA, weighted moving average (WMA), ARMA, and ARIMA time series methods are used for wind speed forecasting at different time horizons (very short-term, short-term, medium-term, and long-term scales). The first part of the research [24] develops 10-min and 1-h time series forecasting models for 10-min and 2-h ahead forecasting. In [25], the second part of the study, 3-h and 6-h time series forecasting models are applied to forecast the next nine and twenty-four hours. The daily wind speed series is predicted using the ARIMA approach in [26]. Across Malaysia, 18 meteorological stations' wind data are analyzed. The Engle's Lagrange Multiplier (LM) is used to check the autoregressive conditionally heteroscedastic (ARCH) influence on the residuals, and the Ljung-Box test is used to evaluate whether autocorrelation exists through the residuals. The predicting accuracy was based on the value of RMSE and MAPE error metrics. In [27], hybrid models, ARIMA, and ANN models were used to develop multi-time scale (ultra-short, short, medium, and long-term) wind speed forecasting models. The variable speed is used to assess the predictability of the wind's mean speed MAE, RMSE, and MAPE. Three forecasting techniques for short- and long-term wind power are evaluated in [28]. In particular, there are several time-series-based approaches, such as ARMA, ANNs, and Adaptive Neuro-Fuzzy Inference Systems (ANFIS). Only RMSE and MAE are used to compare the performance of various models due to their popularity and ease of usage. The research indicates that, for short-term forecasting, the ARMA model outperformed other approaches, in contrast to long-term forecasting. Three different linear time series models, including AR, MA, and ARMA, are investigated in [29]. Furthermore, two machine learning models called random forests (RF) and multivariate adaptive regression splines (MARS) are used to estimate daily and monthly wind speeds over the short and long term, respectively.

The primary goal of this paper is to evaluate time series of wind speed data in order to derive important statistics and other data aspects (Trend, seasonality, noise, etc.) for a specific location. A step-by-step procedure and an accurate, simple-to-use model for predicting the short-term (daily) and long-term (monthly) wind speeds of Ras Gharib are presented in this work. The collected wind speed data will be trained and forecasted using a variety of time series analysis techniques. Eventually, several realistic comparisons of the associated models' prediction accuracy will be realized in terms of their effectiveness in making predictions. Each model's forecast performance is assessed using MAPE, RMSE, and MAE to select the best model for the gathered wind speed data. The method outlined in this study for selecting the best forecasting model for the location under investigation can be easily applied elsewhere if historical wind speed data are available.

In this study, wind speed historical data of Ras Gharib (Latitude: 28° 21′ 33.9′′ & Longitude: 33° 4′ 30.5′′) was collected from DATA.NASA.GOV data sets. To predict short-term and long-term wind speed, data were gathered daily and monthly. To estimate future values at a specific time point and help analysts to choose the appropriate forecasting algorithm, a variety of time series analysis models are used. To evaluate whether the data displays a trend or a seasonal pattern, as well as to establish the frequency of seasonality to be employed within various forecasting methodologies, data analysis is conducted using the autocorrelation function (ACF) and partial autocorrelation function (PACF). The presented forecasting techniques are used to fit and predict time series data, and the results of these tools will then be compared using various error metrics to determine which technique yields the best forecasting and fitting results. In this way, the most accurate model for both short-term and long-term studies can be identified. Given that it is anticipated that wind energy integration will increase significantly in the grid of the future power system, a precise model is essential for both the short-term operational problem and the long-term planning problem in power systems. Although the selected models are intended for Ras Gharib in Egypt, it is important to note that if historical wind speed data are available for other locations, then similar procedures can be easily carried out there as well. Finally, we should mention that all simulation procedures have been done using R-Studio software [30].

2. Material and Methods

2.1. Time Series Forecasting Methods

This section discusses the different time series forecasting methods. Time series forecasting is a statistical method that is necessary for prediction problems with time component data. The historical data used as the input for time series forecasting is examined and analyzed to draw out the characteristics of the data, such as trends, seasonality, and outliers. The objective is to forecast a future value at a particular point in time to assist analysts in selecting the best forecasting algorithm. The forecast package in R-Studio software contains most time-series forecasting techniques [31]. The time series forecasting techniques employed in this study are summarized in the following subsections.

2.1.1. ARMA Model

ARMA is a forecasting framework that employs both the autoregression analysis (AR) and moving average (MA) techniques mentioned below on well-behaved time-series data.

➤ **Auto-Regressive (AR) model:** In AR, the subsequent time step in the series is modeled as a linear function of the preceding time steps, and it appears in Eq. (1) as $AR(p)$

$$AR(p): y(t) = \sum_{i=1}^{p} \varphi_i \, y(t-i) + \varepsilon(t) \qquad (1)$$

➤ **Moving Average (MA) model:** In MA, the time step in the series is represented as a linear function of the prior residual errors from the mean process, and it is donated in Eq. (2) as $MA(q)$

$$MA(q): y(t) = \sum_{j=1}^{q} \theta_j \varepsilon(t-j) \qquad (2)$$

Since the ARMA model combines $AR(p)$ and $MA(q)$ as mentioned above, it may be used to characterize the current observation $y(t)$ of a given time series by Eq. (3) using $ARMA$ of order (p, q).

$$ARMA\ (p, q): y(t)$$
$$= \sum_{i=1}^{p} \varphi_i\, y(t-i) + \varepsilon(t)$$
$$+ \sum_{j=1}^{q} \theta_j \varepsilon(t-j) \qquad (3)$$

Moreover, the ARMA model is only applied if the time series is stationary. The autoregressive integrated moving average (ARIMA) model is used if the time series has a trend "not stationary", and an extra term (d) is added to the equation. This additional term (d) represents the order of differentiation needed to render the time series stationary (to eliminate any trend). The Seasonal Autoregressive Integrated Moving Average (SARIMA) model is developed if the time series has a seasonal pattern, and a seasonal term is added to the ARIMA model. The SARIMA model is capable of handling time series data with seasonality by employing a linear combination of seasonal past values and forecast errors [32]. Both seasonal and non-seasonal parameters are considered by the SARIMA approach, and the model is categorized as described in the following: Eq. (4)

$$SARIMA\ \underbrace{(p, d, q)}\ \underbrace{(P, D, Q)_m} \qquad (4)$$

Non-seasonal part of the model   Seasonal part of the model

Because the ARIMA/SARIMA models are intended to use data from univariate time series (time series with only one variable changed over time), they have the limitation of not working if there are more than one variable in the time series (multivariate time series). In this case, ARIMAX and SARIMAX models could be used. Both ARIMA and SARIMA methods require large amounts of input data to produce an accurate forecast model.

2.1.2. Dynamic Harmonic Regression

A nonstationary time series technique called dynamic harmonic regression is used to identify irregular elements of time series data, such as trends, seasonality, and cyclical components. It has many applications, including the analysis of economic and environmental data [33]. When there are large seasonal periods (daily, weekly, half-hourly, and hourly data), dynamic regression with Fourier terms may be preferable to conventional models. In this case, the short-term time series dynamics are handled by an ARIMA error, and the seasonal pattern is simulated using Fourier terms. Reference

[34] lists the following advantages of this approach: It accepts seasonality of any length and for data with more than one seasonal period, Fourier terms with different frequencies can be included. The smoothness of the seasonal pattern can be adjusted using k, which is mentioned in Eq. (5) that describes the dynamic harmonic regression model:

$$y(t) = c + \sum_{l=1}^{k} \left[\alpha_l\, sin\frac{2\pi lt}{m} + \beta_l cos\frac{2\pi lt}{m} \right] + Ut \qquad (5)$$

The disadvantage of dynamic harmonic regression is that unlike ARIMA and ETS models, it does not allow the model to change over time (for example, it assumes that the seasonal pattern remains constant over time).

2.1.3. Forecasting Using STL Decomposition

Decomposition of data refers to the process of breaking down the data into its basic constituents (seasonal, trend, and reminder components), it can be additive or multiplicative. Decomposition techniques include Classical, SEATS, X11, and STL. The following are some advantages of STL over conventional "SEATS" and "X11": In addition to handling monthly and quarterly data, it can also handle data with any form of seasonality. The rate of change can be controlled, and the seasonal component can change over time. The trend cycle's motion can also be altered by the user. To employ the STL decomposition method for forecasting, the time series is expressed by its seasonal term and seasonal-adjusted term (series after eliminating the seasonal term), as given by Eq. (6) for additive decomposition and Eq. (7) for multiplicative decomposition.

$$y(t) = \hat{S}(t) + \hat{A}(t) \qquad (6)$$

$$y(t) = \hat{S}(t) * \hat{A}(t) \qquad (7)$$

In this technique, we independently anticipate the seasonal component, $\hat{S}$, and the seasonally adjusted component, $\hat{A}_t$. Since it is often assumed that the seasonal component is either constant or varies very slowly, forecasting it simply requires utilizing the most recent year for the estimated component. Any non-seasonal forecasting method, such as Holt's approach (explained below) or a non-seasonal ARIMA model, may be used to forecast the seasonally adjusted component.

The main drawback to STL is that it does not automatically manage a trading day or calendar variance.

2.1.4. Exponential Smoothing Methods

One of the best time-series forecasting techniques is exponential smoothing (ES). With this approach, new values are predicted using a weighted average of historical data. As the observation data get older over time and the relevance of these values decreases exponentially, this method lends more weight to the most recent values in the series. There are three main types of ES as follows:

2.1.4.1. Single Exponential Smoothing (SES)

It is a method for forecasting time-series data that lacks a trend or seasonality. It is composed of a forecast equation and one smoothing equation for level, as given by Eq. (8)

$$\hat{y}_{t+h|t} = l_t \qquad (8)$$

2.1.4.2. Double Exponential Smoothing (DES)

This type of ES is also frequently referred to as the Holt linear approach. It can handle time series data without any seasonal characteristics but with a linear trend. It has two smoothing equations, one for level and the other for the trend in the data, as in Eq. (9).

$$\hat{y}_{t+h|t} = l_t + hb_t \qquad (9)$$

2.1.4.3. Trible Exponential Smoothing (TES)

TES is also commonly known as Holt-Winters Exponential Smoothing. By extending ES, Holt and Winters made it feasible to predict data that exhibit a trend and seasonal pattern. In this method, a forecast equation and three smoothing equations-one for the level, one for the trend, and the third for the seasonal component-are used. The smoothing equations are classified into additive and multiplicative techniques as given by Eq. (10) and Eq. (11).

$$\hat{y}_{t+h|t} = l_t + hb_t + s_{t+h-m(k_1+1)} \qquad (10)$$

$$\hat{y}_{t+h|t} = (l_t + hb_t)s_{t+h-m(k_1+1)} \qquad (11)$$

It is important to mention that the Holt-Winter exponential smoothing approach can handle frequencies up to 24, but above that, any other method that can handle huge frequencies can be employed.

The l_t, b_t, and s_t are described in detail in Ref [35].

2.1.5. ETS Method

Even though ES approaches have been employed for a long time, recent methodological developments have allowed for the integration of these models into a new dynamic nonlinear model framework. Reference [36] describes the ETS (Error-Trend-Seasonal or Exponential Smoothing) framework, which defines an extended class of Exponential Smoothing methods and offers a theoretical framework for analyzing these models using state-space-based calculations with support for model selection and forecast standard error calculation. Furthermore, by incorporating the typical ES models, the ETS framework gives a theoretical basis for what was previously a collection of ad hoc approaches (such as the Holt and Holt-Winters additive and multiplicative methods). In this approach, each model consists of a measurement equation that describes the observed data and a few state equations that describe how the hidden components or states (level, trend, and seasonality) change over time. These are therefore referred to as state space models, where each state space model is designated as ETS (x, y, z) for (Error, Trend, Seasonal). The possible values for each component are Error = {A, M} Trend = {N, A, Ad}, and Seasonal = {N, A, M}. Where "A" refers to an additive equation, "Ad" stands for an additive dampening equation, "M" is a multiplicative equation, and "N" is none [36].

As mentioned, ES methods and their extended ETS method lend more weight to the most recent values in the series; these methods are best for very close period predictions. As the forecast period increases, the accuracy of the forecast results decreases.

2.1.6. TBATS Method

A fully automated alternative method called TBATS, which stands for Trigonometric Seasonal, Box-Cox Transformation, ARMA residuals, Trend, and seasonality, was discovered by [37]. This technique is primarily used to handle time series data with complex and numerous seasonality patterns. The optimum model for the data selected by TBATS depends on if it requires a Box-Cox transformation to make it linear, whether it has a damped trend or not. Additional aspects include, whether it is necessary to model residuals in ARMA (p, q) and whether the harmonics were used to simulate seasonal effects. The notion of the TBATS is as Eq. (12):

$$TBATS(\lambda, \{p, q\}, Db, \{< m, k >\}) \qquad (12)$$

Although the TBATS approach is fully automated and capable of handling time series data with complex and numerous seasonality patterns, it may need a lengthy computation time and does not support the addition of an external variable to improve forecasts.

2.2. Validation of The Process

To determine whether the model is adequate and whether any changes will make it better or worse, the process must be validated. The process validation in this paper includes two steps. First, using a particular information criterion, the optimal parameters are selected from a range of possible orders and coefficients of various models. The Akaike Information Criterion (AIC) represented by Eq. (13) has been chosen for this study as it is one of the information criteria that is most usually employed in the literature. This criterion assesses the trade-off between the model's complexity and the quality of fitting.

$$AIC = T \log\left(\frac{SSE}{T}\right) + 2(z + 2) \qquad (13)$$

SSE is the sum of squared errors (ε_t) determined by the following Eq. (14).

$$SSE = \sum_{t=1}^{N} \varepsilon(t)^2 \qquad (14)$$

The second step is to assess the goodness of the method for forecasting, not just for fitting. To achieve this, the overall data set is split into two subsets called "train data" and "test data". The train data are used to fit the model and generate forecasts based on it. The test data are employed to be compared with the resulting forecasting data that the train set generated to evaluate forecasting accuracy.

2.3. Evaluation and Comparing Forecasting Models

The evaluation of forecasting models is a crucial phase in forecasting studies since it allows researchers to determine the model's accuracy and compare the performance of various forecasting models. In most forecasting studies, achieving high prediction accuracy is the main goal. Since the Akaike information criterion (AIC) has the limitation that it cannot be used to compare models from different model families (ARMA and Holt-Winter's methods), many other metrics have been used to evaluate and compare forecasting models, which will be covered in this section.

Root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) are three of the most common direct error metrics. Because of their mathematical convenience and popularity in time-series wind speed forecasting purposes, these three metrics used in this study to evaluate and compare models for fitting and forecasting wind speed data. The minimization of these measures is often the most crucial aspect to take into account when comparing forecasting models. Higher forecast accuracy is indicated by lower values for any of these metrics.

2.3.1 The Root Mean Squared Error (RMSE)

RMSE which is the standard deviation of the residuals (errors of prediction), as given by Eq. (15) demonstrates how tightly the data is clustered around the line of best fit.

RMSE is particularly important for assessing forecasts since it quantifies the spread of forecast error values, regardless of the direction or sign of the error.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_t - \hat{y}_t)^2} \qquad (15)$$

2.3.2 The Mean Absolute Error (MAE)

MAE, as computed by Eq. (16), measures the absolute difference between predictions and actual values. Therefore, its value is always positive, even if the difference between the actual and predicted values is negative. Unlike the RMSE, the MAE does not give greater or lesser weight to errors, but when errors increase, the MAE value rises linearly.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|(y_t - \hat{y}_t)| \qquad (16)$$

2.3.3 The Mean Absolute Percentage Error (MAPE)

A forecasting system's accuracy is evaluated using MAPE shown by Eq. (17). It is computed by dividing the total of all absolute errors by the actual values related to each time separately. MAPE is the most widely used metric to assess forecast error since the variable's units are scaled to percentage units, which makes it simpler to understand. The accuracy of the model can be calculated using Eq. (18).

$$MAPE\% = \frac{1}{N}\sum_{i=1}^{N}|\frac{y_t - \hat{y}_t}{y_t})| * 100 \qquad (17)$$

$$Accuracy\ \% = (100 - MAPE) \qquad (18)$$

3. Results and Discussions

This section examines historical wind speed data for Ras Gharib, Egypt (latitude: 28° 21' 33.9" and longitude: 33° 4' 30.5"), which were gathered from DATA.NASA.GOV data sets [38]. The study's data were collected on a daily and monthly basis. To evaluate the precision of model fitting and forecasting, the collected data are separated into train and test subsets as previously described. After that, data analysis is carried out by plotting ACF and PACF to ascertain whether the data exhibits a trend or a seasonal pattern, as well as to determine the frequency of seasonality to be used within various forecasting techniques [39]. The various time series forecasting techniques covered in Section 2 will be employed to fit and predict time series data, and the results of these tools will then be compared using the different error metrics previously mentioned (MAE, RMSE, and MAPE) to ascertain which technique produces the best forecasting and fitting outcomes. All simulation procedures have been done using R-Studio software.

3.1 Case 1 (Daily Wind Data)

Daily wind speed data that was collected from 2017 to 2021 is shown in Fig.1 The data is then split into train data and test data, as depicted in Fig.2. The train data (from 2017 to 2020) is utilized to identify the best model fitting the data, while the test data (from 2021) is applied to assess the model's forecasting accuracy. It should be noted that the time axis in Figs. 1 and 2 are expressed as the number of years (1 represents "1 January 2017," 2 represents "1 January 2018," etc.).
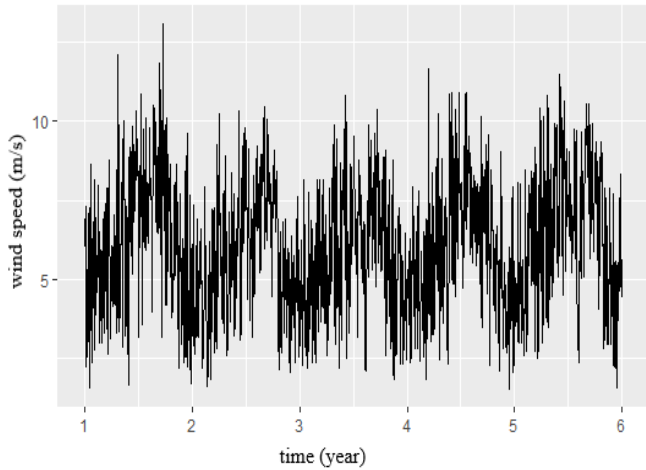
Fig. 1. Plot of daily wind speed for 2017-2021
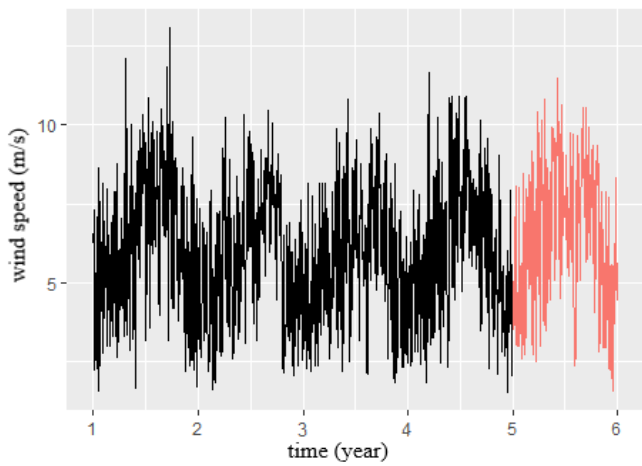
Fig. 2. Train and test daily wind speed data

3.1.1. Daily Data Analysis

Fig.3 and Fig.4 display the plots of the train data's ACF and PACF, respectively. Note that Fig.3 and Fig.4 cover only two years of the train data; however, the subsequent years exhibited a similar pattern. To compare and validate the output results in determining the stationarity of the train data over time and whether or not the data exhibits a trend, two statistical tests are carried out in this study. The statistical tests employed are the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Philips-Schmidt-Shin (KPSS) tests [40,41]. In both tests, a crucial PR-value, or probability value, is identified and compared with the stated null hypothesis of each test. Accordingly, if the PR-value meets a specified level, it can be concluded that there are no trends in the train data (stationarity). The ADF test's null hypothesis is that the data have a trend if the PR-value is greater than 5%, whereas the PR-value for the KPSS test is less than 5%.

The outcomes of PR-values for the ADF test and KPSS test for the simulated train data (from 2017 to 2020) are 0.01 and 0.1, respectively. These results demonstrate that the series is stationary, i.e., devoid of any trend, with a PR-value for the ADF test less than 5% and, for the KPSS test greater than 5%. However, as can be seen in the ACF plot in Fig.3, it is evident

that the train data exhibit seasonality with a 365-day frequency. Thus, seasonal stationarity is observed in the simulated train data.

3.1.2. Daily Wind Speed Forecasting Methods

In this section, various time series forecasting techniques are used to fit and forecast the daily wind speed data under the findings from the daily analysis in the preceding subsection. We select techniques that can accommodate the data's 365-day seasonality, which was mentioned above. AIC, represented by Eq. (13), is used to identify the optimal parameters of each forecasting model under consideration. Different forecasting methods are implemented using the forecasting package in R-Studio software.
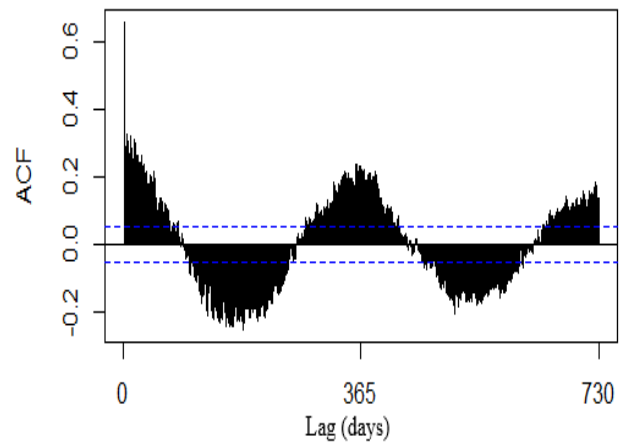
Fig. 3. ACF of daily train data
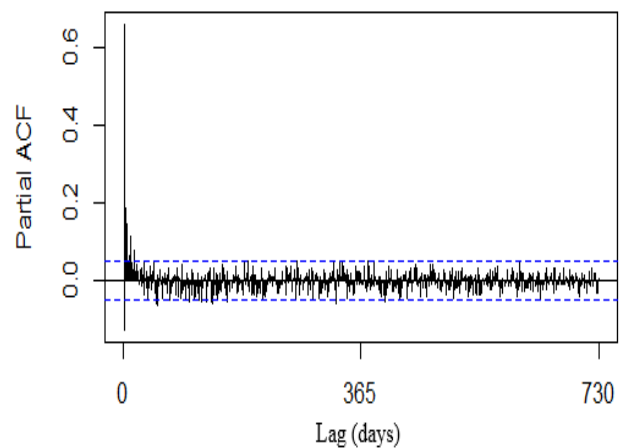
Fig. 4. PACF of daily train Data

3.1.2.1. SARIMA Model

SARIMA method is one of the most effective time series forecasting methods. It is necessary to acquire the appropriate main numbers of $(p, d, q) - (P, D, Q)_m$ for the model described by Eq. (4) to find the best model. Using the results

of the aforementioned ADF and KPSS tests, it is possible to conclude that there is no trend and that no differencing is required. In this situation, d is eliminated, leaving $d = 0$. There is seasonality with a frequency of 365, so $m = 365$, as seen in Fig.3 and Fig.4 (the ACF and PACF plots). To establish the proper primary number order of $(p, q) - (P, Q)$ associated with the best model, several values of them are examined and assessed by AIC. The order that satisfies the lowest AIC is the best model. $SARIMA$ $(1,0,1)$ $(0,1,0)_{365}$ has the least AIC, with a value of 4646.06, according to the findings of attempting numerous different orders. To confirm the necessity of using the SARIMA model for time series data with seasonality, the normal ARIMA model is used to fit the data. After attempting several orders, as was done above, the best model that fits the data is $ARIMA$ $(3, 0, 2)$ with a non-zero mean model, where the least AIC value is 5302.83 and is higher than the SARIMA model's AIC value. Fig.5 indicates the forecasting result obtained by ARIMA, where the actual data are shown in black, and the actual forecasts are in the blue line. Note that, the prediction intervals for 80% and 95% of the actual forecasted points are shown as shaded light and dark blue areas, respectively. The forecasting result obtained by SARIMA is shown in Fig.6 where black and blue colors represent the actual data and the actual forecasts, respectively. The model was developed using the R-Studio function Arima (), with the argument "Seasonality" set to FALSE for normal ARIMA and TRUE for SARIMA.
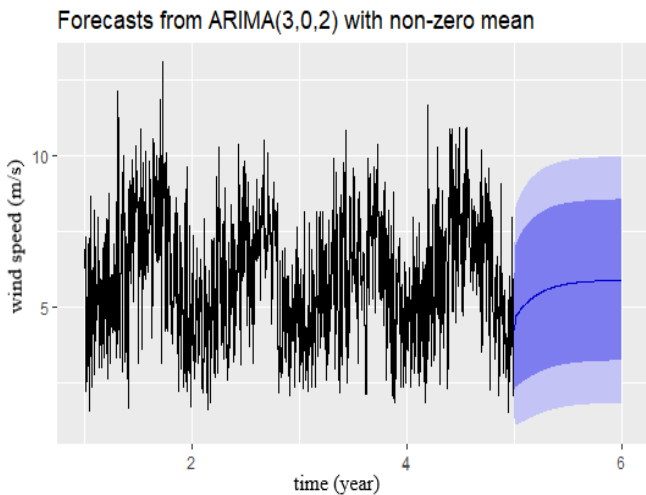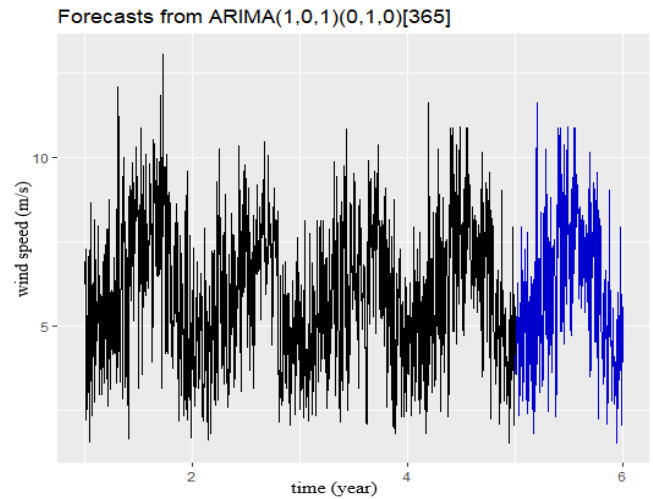

Fig. 6. Forecasts from SARIMA Model

3.1.2.2. Dynamic Harmonic Regression

By experimenting with various models using various ARIMA orders (p, d, q) to model residuals and varying k "number of Fourier sin and cos pairs to capture seasonality", $ARIMA$ $(2,0,1)$ $errors, with$ $k = 3$, is selected to fit the data that provides the least AIC of 5255.147. The R-Studio function Arima (), with the argument "xreg" standing for Fourier series of the data with k changed to obtain the best model, was used to develop this model. The Forecasting result of the dynamic regression model is depicted in Fig.7.
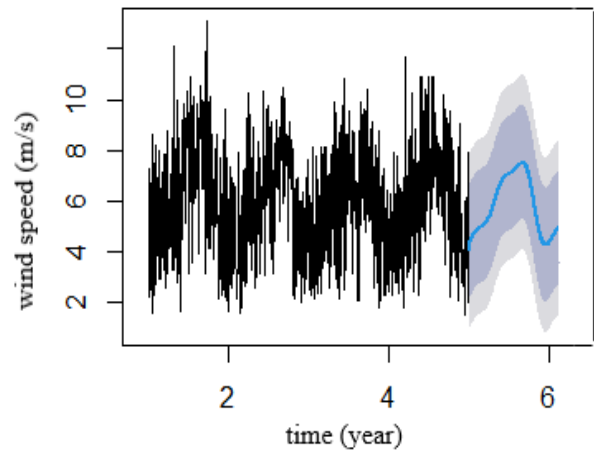

Fig. 5. Forecasts from normal ARIMA model

Fig. 7. Forecasts from dynamic harmonic regression model

3.1.2.3. STL Method

By using AIC to choose the best STL model, $STL +$ ETS (A, N, N) (the naïve method for seasonal parts and simple exponential smoothing with additive error for seasonal-adjusted parts) is chosen to fit the data with the least AIC of 11654.54. The model was created using R-Studio's stlf () function, The Forecasting result of the dynamic regression model is depicted in Fig.8.
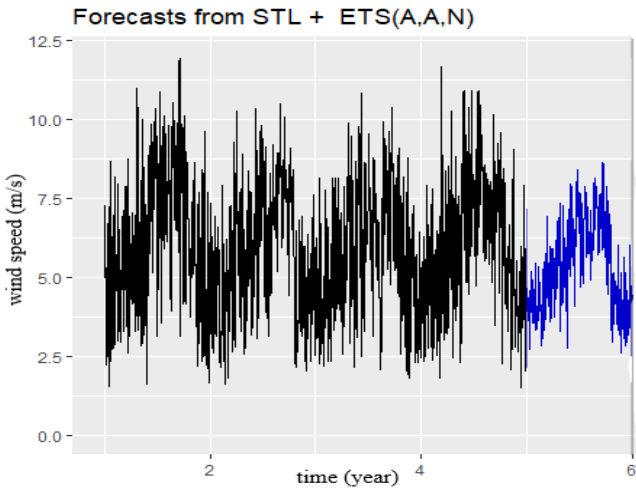
Fig 8. Forecasts from STL model

3.1.2.4. TBATS Method

After experimenting with various TBATS models for the various λ, and ARMA orders $\{p, q\}$, as well as the number of Fourier's series k. The $TBATS$ $(1, \{3,4\}, -, \{< 365,4 >\})$ model with the lowest AIC of 11767.48 is chosen to fit the data. The model's outcomes are displayed in Fig.9 using the R-Studio function tbats().
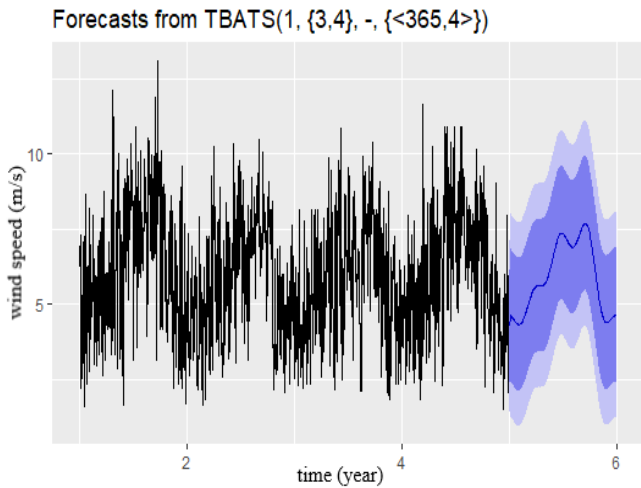

Fig. 9. Forecasts from TBATS model

3.1.3. Evaluating and Comparing Forecasting Models

An essential step in assessing the forecast model is looking at and analyzing the residuals. Residuals, which represent the portion of the data that the model cannot explain, are differences between the model's one-step predicted output and the measured output from the data set. There must be no correlation between the estimated residuals. If there are correlations between them, the residuals still contain information that can be used to create forecasts. There are some statistical tests used to assess the correlation of the residuals [42], and the Ljung-Box test is employed here. In this test, if the test's PR-value is greater than 0.05, which

denotes that there is no correlation between the residuals, then the residuals pass the test. Analyzing the residuals in this paper is performed using R-Studio function checkresiduals (). To compare various models, RMSE, MAE, and MAPE are computed for each forecasting technique. The different error metrics are produced by R-Studio function accuracy (). Table 1. displays the RMSE, MAE, MAPE, and Ljung-Box test outcomes for the generated models that fit the train data in the previous subsection. The results show that, on fitting the train data, although the STL method has the least error metrics (RMSE, MAE, and MAPE), it does not pass the Ljung-Box test, where the PR-value is 2.8 e-10, which is less than 0.05, and therefore the STL model is rejected. The TBATS model is the second-least error metric, but in this case, the PR-value is 0.7792, which is greater than 0.05, and passed the Ljung-Box test. TBATS shows better results than other models, and consequently, it is accepted as the best model for fitting the train data. To evaluate the goodness of the forecast at the time horizon of the test data (365 days), the error metrics and PR-value of each forecasted model are calculated as shown in Table 2. According to the least error metrics and the PR-value that satisfies Ljung-Box test, TBATS model is selected for the test data. The above results for both train and test data confirm that TBATS is the most accurate model to fit and forecast the daily wind speed data of Ras Gharib, with an accuracy of 76.4% on fitting the data and 73.1% on forecasting at h=365. The parameters associated with the selected TBATS model were as follows:

$TBATS$ $(1, \{3,4\}, -, \{< 365,4 >\})$

with λ =1(no transformation needed), p of ARMA errors =3, q of ARMA errors=4, frequency=365, the number of Fourier pairs =4, AR coefficients= [0.0815; 0.6013; 0.1506], and MA coefficients= [0.55735;0.5413; -0.2586; 0.000134].

**Table. 1.** RMSE, MAE, MAPE, and residuals test of models on the train data

Model	RMSE	MAE	MAPE	PR-value
ARIMA (3,0,2) with non-zero mean	1.47	1.18	24.06	0.40
SARIMA $(1,0,1)$ $(0,1,0)_{365}$	1.74	1.21	24.74	$3e^{-10}$
STL+ETS (A, N, N)	1.41	1.11	21.87	$2e^{-16}$
Dynamic harmonic ARIMA (2,0,1) with errors, K=3	1.45	1.15	23.58	0.56
TBATS (1, {3,4}, -, {<365,4>})	1.44	1.14	23.55	0.78

**Table. 2.** RMSE, MAE, MAPE, and residuals test of models on the test data

Model	RMSE	MAE	MAPE	PR-value
ARIMA (3,0,2) with non-zero mean	2.28	1.89	32.38	0.40
SARIMA (1,0,1) (0,1,0)$_{365}$	2.64	2.12	36.68	$3e^{-10}$
STL+ETS (A, N, N)	2.27	1.83	30.45	$2e^{-16}$
Dynamic harmonic ARIMA (2,0,1) with errors, K=3	1.93	1.55	27.55	0.56
TBATS (1, {3,4}, - , {<365,4>})	1.93	1.55	26.89	0.78

3.2. Case 2 (Monthly Wind Data)

In this case, the monthly wind speed data gathered from Ras Gharib for the years 2010 to 2020 is shown in Fig.10. As in the previous instance, the data are split into train data and test data, with the train data making up 80% of the data (about 106 months) and the test data comprising 20% of the data (or roughly 26 months), as shown in Fig.11. It should be noticed that in Fig. 10 and Fig. 11, the time axis is expressed as the number of years (1 represents "1 January 2010," 2 represents "1 January 2011," etc.).
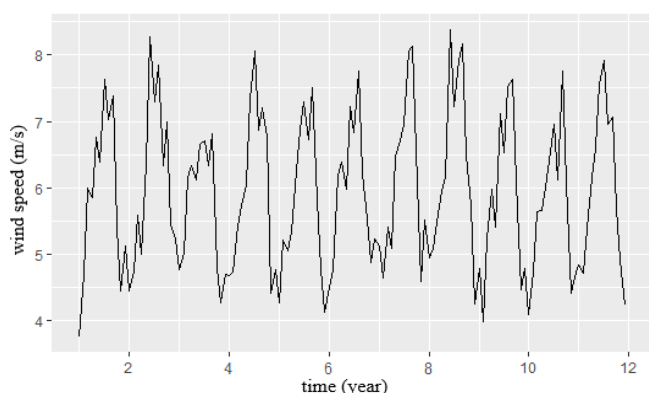

**Fig. 10.** Monthly wind speed data (2010-2020)
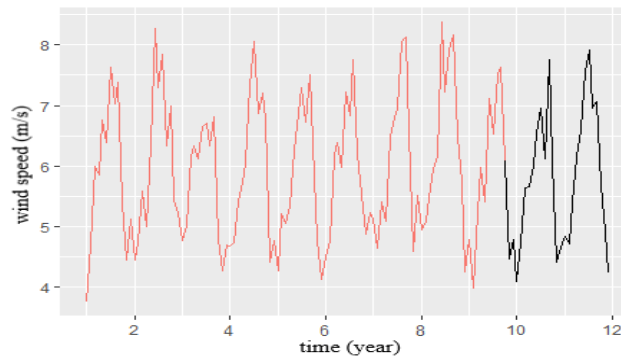
Fig. 11. Monthly wind speed train and test data

3.2.1 Data Analysis

The plots of ACF and PACF of the train data are shown in Fig.12 and Fig.13, respectively. In these graphs, the Time axes represent the number of months (1 denotes "January 2010," 2 denotes "February 2010," etc.). For the simulated train data (80% of the data), the results of the ADF test and KPSS test are 0.01 and 0.1, respectively. With a PR-value for the ADF test less than 5% and for the KPSS test greater than 5%, these results show that the series is stationary, or free of any trend. However, it is clear from the ACF plot in Fig.12 that the train data displays a seasonality pattern with a frequency of 12 (Monthly data). As a result, seasonal stationarity is seen in the simulated train data.
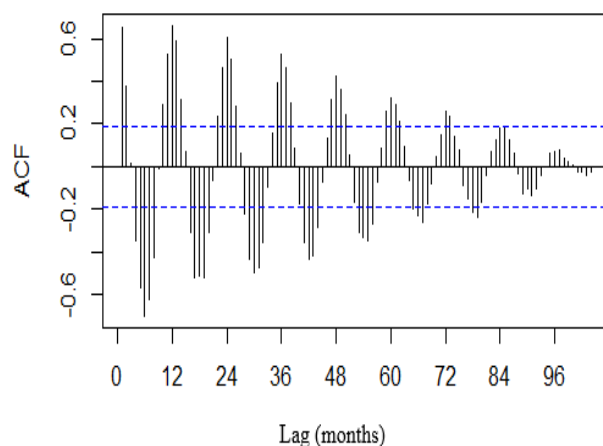
Fig. 12. ACF of monthly train data
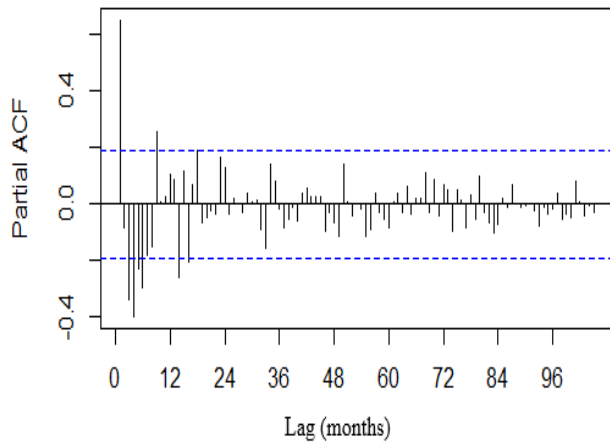
Fig. 13. PACF of monthly train data

3.2.2 Monthly Wind Speed Forecasting Methods

In this section, two extra methods are used for monthly wind data in addition to the various time series forecasting techniques that were previously used to fit and forecast the daily wind speed. Holt-and winter's ETS are the other two methods, and they can both model data with seasonal periods of less than 24. Since the examination of the monthly data reveals a seasonality pattern with a frequency of 12, Holt-Winter's and ETS methods are also used to investigate further techniques for discovering the best model. The same as in the prior instance, AIC is used to determine the best settings for each forecasting model under consideration.

The following list provides a summary of the best parameters for each forecasting technique that are implemented using the forecasting package in R-Studio software.

3.2.2.1. SARIMA Model

Using the results from ADF and KPSS tests, there isn't trend, so $d = 0$, and from ACF, PACF plots $D = 1$, at frequency m = 12. Applying different orders of $(p, q) - (P, Q)$ and using AIC to determine the best model, $SARIMA\ (0,0,0)\ (0,1,2)12$ is the one that satisfies the lowest AIC, which is 176.11. If a normal ARIMA model is used to fit the data, $ARIMA\ (4,0,1)$ with a non-zero mean model was chosen to fit the data with AIC=231.18.

3.2.2.2. STL Method

By using AIC to choose the best STL method, STL+ETS (M, N, N) (naïve method for the seasonal part and simple exponential smoothing with multiplicative errors for the seasonal-adjusted part) is chosen to fit the data with AIC = 344.76.

3.2.2.3. Dynamic Harmonic Regression

Several models with various ARIMA orders and Fourier pairs numbers were tried, and $ARIMA\ (2,0,2)$ with errors, $k=4$ was selected to best fit the data with an AIC of 173.03.

3.2.2.4. TBATS Method

Following testing several TBATS models for different λ, and ARMA orders $\{p, q\}$ as well as the number of Fourier's series k, $TBATS\ (0.032, \{0,0\}, -, \{< 12,4 >\})$ model is chosen to fit the data with AIC=370.3.

3.2.2.5. Holt-Winter's Method

As the frequency of data is 12 (less than 24) Holt-winter's method is used to fit and forecast data. In this examination, AIC is 392.6619.

3.2.2.6. ETS Method

The ETS method can also be used as Holt-Winter's method to model the data with a seasonal period of less than 24. So, it has been used here for monthly data with $f = 12$. By trying different models, we found that $ETS\ (M, N, A)$ is the best model with the lowest AIC of 376.37.

3.2.3. Evaluating and Comparing Forecasting Models

The RMSE, MAE, MAPE, and PR-value of the Ljung-Box test for the different models on fitting train data are shown in Table 3. To evaluate the goodness of the forecast, the length of the test data is set to 26 months, where the obtained results are shown in Table 4. Results of the model evaluation show that the dynamic harmonic technique has the lowest error metrics and passes the Ljung-Box test with a PR-value of 0.1405 (> 0.05). (On fitting the data). Results of model evaluation on forecasting (26 months) show that the SARIMA model has the least RMSE, MAE, and MAPE and passes the Ljung-Box test with a PR-value of 0.1599. So, it can be concluded that the dynamic harmonic regression model can be used for fitting data with an accuracy of 93.73% but not for forecasting. However, the $SARIMA\ (0,0,0)\ (0,1,2)12$ model can be used for forecasting with an accuracy of 92.83% at h = 26. The moving average coefficients (θ_j) of the seasonal part of SARIMA model are $SMA1 = -0.9884$ and $SMA2 = 0.3259$, respectively.

Table. 3. RMSE, MAE, MAPE, and residuals test of models on the train data

Model	RMSE	MAE	MAPE	PR-value
ARIMA (4,0,1) with non-zero mean	0.87	0.72	13.32	$6e^{-3}$
SARIMA (0,0,0) (0,1,2)$_{12}$	0.57	0.41	7.17	0.16
STL+ETS (M, N, N)	0.46	0.33	5.78	$1e^{-3}$
Dynamic harmonic ARIMA (2,0,2) with errors, K=4	1.06	0.84	14.53	0.14
Holt-Winter(additive)	0.45	0.36	6.50	$4.3e^{-8}$
Holt-Winter (Multiplicative)	0.46	0.37	6.58	$3.2e^{-6}$
ETS (M, N, A)	0.44	0.34	5.96	$2.6e^{-6}$
TBATS (0.032, {0,0}, -, {<12,4>})	0.42	0.33	5.72	$3e^{-3}$

Table. 4. RMSE, MAE, MAPE, and residuals test of models on the test data

Model	RMSE	MAE	MAPE	PR-value
ARIMA (4,0,1) with non-zero mean	0.67	0.54	9.29	$6e^{-3}$
SARIMA (0,0,0) (0,1,2)$_{12}$	0.53	0.40	6.72	0.16
STL+ETS (M, N, N)	0.48	0.39	6.55	$1e^{-3}$
Dynamic harmonic ARIMA (2,0,2) with errors, K=4	0.48	0.37	6.27	0.14
Holt-Winter(additive)	0.53	0.43	7.30	$4.3e^{-8}$
Holt-Winter (Multiplicative)	0.53	0.43	7.27	$3.2e^{-6}$
ETS (M, N, A)	0.51	0.41	6.90	$2.6e^{-6}$
TBATS (0.032, {0,0}, -, {<12,4>})	0.50	0.40	6.62	$3e^{-3}$

4. Conclusions and Future Work

In this article, different time series forecasting methods are used to fit and forecast the wind speed of Ras-Gharib in Egypt using historical daily and monthly data for short-term and long-term forecasting. Data sets are separated into train and test sets to evaluate forecast accuracy. Based on the minimum value of AIC, the best parameters are chosen from a range of potential orders and coefficients of different models. Then, the RMSE, MAE, and MAPE are calculated, and a Ljung-Box test that evaluates the correlation of the residuals is carried out for all methods to evaluate and compare forecast outcomes. According to the results of RMSE, MAE, MAPE, and PR-value of Ljung-Box test, the best model was selected among the investigated forecasted techniques. Forecasting and statistical tests are executed using R-Studio software and the results show the following:

4.1. For Daily Wind Speed Data

$TBATS$ $(1, \{3,4\}, -, \{< 365,4 >\})$ is the most accurate model to fit and forecast the daily data with an accuracy of 76.4% on fitting the data and 73.1% on forecasting at h=365.

4.2. For Monthly Wind Speed Data

The dynamic harmonic regression model ($ARIMA$ $(2,0,2)$ $with$ $errors, K = 4$) can be used for fitting data with an accuracy of 93.73%, and the $SARIMA$ $(0,0,0)$ $(0,1,2)_{12}$ model can be used for forecasting with an accuracy of 92.83% at h = 26.

The abovementioned results can provide us with the knowledge that aids in developing strategies for power system operation and planning issues. Finally, we think that extending the current study to handle the variance of residuals forecasts using the ARIMA-GARCH model may produce better predicting results in subsequent work.

References

[1] "Department of Commerce, Country Commercial Guides, Egypt-Electricity and Renewable Energy," Country Commercial Guides, Egypt -Electricity and Renewable Energy, 2022.

[2] G. Sideratos and N. D. Hatziargyriou, "An advanced statistical method for wind power forecasting," IEEE Trans. Power Syst., vol. 22, no. 1, pp. 258–265, 2007.

[3] J. Lerner, M. Grundmeyer, and M. Garvert, "The importance of wind forecasting," Renew. Energy Focus, vol. 10, no. 2, pp. 64–66, 2009.

[4] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, and Z. Yan, "A review on the forecasting of wind speed and generated power," Renew. Sustain. Energy Rev., vol. 13, no. 4, pp. 915–920, 2009.

[5] M. Negnevitsky and C. W. Potter, "Innovative short-term wind generation pre-diction techniques," in Proceedings of the Power Systems Conference and Exposition, 2006, pp. 60–65.

[6] M. C. Alexiadis, P. S. Dokopoulos, H. S. Sahsamanoglou, and I. M. Manousaridis, "Short-term forecasting of wind speed and related electrical power," Sol. Energy, vol. 63, no. 1, pp. 61–68, 1998.

[7] A. Sfetsos, "A novel approach for the forecasting of mean hourly wind speed time series," Renew. Energy, vol. 27, no. 2, pp. 163–174, 2002.

[8] R. Al-Hajj, M. M. Fouad, A. Assi, and E. Mabrouk, "Short-term wind energy forecasting with independent daytime/nighttime machine learning models," in 2022 11th International Conference on Renewable Energy Research and Application (ICRERA), 2022

[9] "Forecasting of wind speed using feature selection and neural networks," Int. J. Renew. Energy Res., no. v6i3, 2016

[10] M. R. Wani, M. A. Wani, and R. Riyaz, "Cluster based approach for mining patterns to predict wind speed," in 2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA), 2016, pp. 1046–1050

[11] K. L. Jorgensen and H. R. Shaker, "Wind power forecasting using machine learning: State of the art, trends and challenges," in 2020 IEEE 8th International Conference on Smart Energy Grid Engineering (SEGE), 2020, pp. 44–50.

[12] M. A. Wani and H. F. Bhat, "Multiclass SVM algorithms for wind speed prediction," in 2017 IEEE 6th International Conference on Renewable Energy Research and Applications (ICRERA), 2017, pp. 1139–1143.

[13] J. Shi, J. Guo, and S. Zheng, "Evaluation of hybrid forecasting approaches for wind speed and power generation time series," Renew. Sustain. Energy Rev., vol. 16, no. 5, pp. 3471–3480, 2012.

[14] S. Shukla, R. Ramaprasad, S. Pasari, and S. Sheoran, "Statistical analysis and forecasting of wind speed," in 2022 4th International Conference on Energy, Power, and Environment (ICEPE), 2022, pp. 1–6.

[15] K. Al Dhaheri, W. L. Woon, and Z. Aung, "Wind speed forecasting using statistical and machine learning methods: A case study in the UAE," in Data Analytics for Renewable Energy Integration: Informing the Generation and Distribution of Renewable Energy, Cham: Springer International Publishing, 2017, pp. 107–120.

[16] E. Cadenas and W. Rivera, "Wind speed forecasting in the South Coast of Oaxaca, México," Renew. Energy, vol. 32, no. 12, pp. 2116–2128, 2007.

[17] E. Grigonytė and E. Butkevičiūtė, "Short-term wind speed forecasting using ARIMA model," Energetika/Ehnergetika/Power Eng., vol. 62, no. 1–2, 2016.

[18] I. Tyass, A. Bellat, A. Raihani, K. Mansouri, and T. Khalili, "Wind Speed Prediction Based on Seasonal ARIMA model," E3S Web Conf., vol. 336, p. 00034, 2022.

[19] E. Cadenas, W. Rivera, R. Campos-Amezcua, and C. Heard, "Wind speed prediction using a univariate ARIMA model and a multivariate NARX model," Energies, vol. 9, no. 2, p. 109, 2016.

[20] M. Abotaleb, "Wind speed in England using BATS, TBATS, Holt's Linear and ARIMA model", MAUSAM, vol. 73, no. 1, pp. 129–138, Mar. 2022.

[21] M. U. Yousuf, I. Al-Bahadly, and E. Avci, "Short-term wind speed forecasting based on hybrid MODWT-ARIMA-Markov model," IEEE Access, vol. 9, pp. 79695–79711, 2021.

[22] D. B. Alencar, C. M. Affonso, R. C. L. Oliveira, and J. C. R. Filho, "Hybrid approach combining SARIMA and neural networks for multi-step ahead wind speed forecasting in Brazil," IEEE Access, vol. 6, pp. 55986–55994, 2018.

[23] J. W. Taylor, P. E. McSharry, and R. Buizza, "Wind power density forecasting using ensemble predictions and time series models," IEEE Trans. Energy Convers., vol. 24, no. 3, pp. 775–782, 2009.

[24] I. Colak, S. Sagiroglu, M. Yesilbudak, E. Kabalci, and H. I. Bulbul, "Multi-time series and -time scale modeling for wind speed and wind power forecasting part I: Statistical methods, very short-term and short-term applications," in 2015 International Conference on Renewable Energy Research and Applications (ICRERA), 2015.

[25] I. Colak, S. Sagiroglu, M. Yesilbudak, E. Kabalci, and H. I. Bulbul, "Multi-time series and -time scale modeling for wind speed and wind power forecasting part II: Medium-term and long-term applications," in 2015 International Conference on Renewable Energy Research and Applications (ICRERA), 2015.

[26] N. H. Hussin, F. Yusof, 'aaishah Radziah Jamaludin, and S. M. Norrulashikin, "Forecasting wind speed in Peninsular Malaysia: An application of ARIMA and ARIMA-GARCH models," Pertanika J. Sci. Technol., vol. 29, no. 1, 2021.

[27] D. B. de Alencar, C. de Mattos Affonso, R. C. L. de Oliveira, J. L. M. Rodríguez, J. C. Leite, and J. C. R. Filho, "Different models for forecasting wind power generation: Case study," Energies, vol. 10, no. 12, pp. 1–27, 2017.

[28] Q. Chen and K. A. Folly, "Comparison of three methods for short-term wind power forecasting," in 2018 International Joint Conference on Neural Networks (IJCNN), 2018.

[29] S. Mehdizadeh, A. Kozekalani Sales, and M. J. S. Safari, "Estimating the short-term and long-term wind speeds: implementing hybrid models through coupling machine learning and linear time series models," SN Appl. Sci., vol. 2, no. 6, 2020.

[30] J. Chambers, Software for data analysis: Programming with R. New York, NY: Springer New York, 2008.

[31] R. J. Hyndman and Y. Khandakar, "Automatic time series forecasting: The forecast package for R," Journal of Statistical Software, vol. 26, no. 3, pp. 1–22, 2008.

[32] T. Dimri, S. Ahmad, and M. Sharif, "Time series analysis of climate variables using seasonal ARIMA approach," J. Earth Syst. Sci., vol. 129, no. 1, 2020.

[33] P. C. Young, D. J. Pedregal, and W. Tych, "Dynamic harmonic regression," J. Forecast., vol. 18, no. 6, pp. 369–394, 1999.

[34] A. J. Zavala and A. R. Messina, "Dynamic harmonic regression approach to wind power generation forecasting," in 2016 IEEE PES Transmission & Distribution Conference and Exposition-Latin America (PES T&D-LA), 2016.

[35] E. S. Gardner, "Exponential smoothing: The state of the art," J. Forecast., vol. 4, no. 1, pp. 1–28, 1985.

[36] R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose, "A state space framework for automatic forecasting using exponential smoothing methods," Int. J. Forecast., vol. 18, no. 3, pp. 439–454, 2002.

[37] A. M. De Livera, R. J. Hyndman, and R. D. Snyder, "Forecasting time series with complex seasonal patterns using exponential smoothing," J. Am. Stat. Assoc., vol. 106, no. 496, pp. 1513–1527, 2011.

[38] P. Stackhouse, "Nasa power," Nasa.gov. [Online]. Available: http://power.larc.nasa.gov/data-access-viewer/. [Accessed: 23-Feb-2023]

[39] G. M. Tinungki, "The analysis of partial autocorrelation function in predicting maximum wind speed," IOP Conf. Ser. Earth Environ. Sci., vol. 235, p. 012097, 2019.

[40] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," J. Am. Stat. Assoc., vol. 74, no. 366, p. 427, 1979.

[41] D. Lee and P. Schmidt, "On the power of the KPSS test of stationarity against fractionally-integrated alternatives," J. Econom., vol. 73, no. 1, pp. 285–302, 1996.

[42] H. Hassani and M. R. Yeganegi, "Sum of squared ACF and the Ljung–Box statistics," Physica A, vol. 520, pp. 81–86, 2019.