Journal of Advanced Engineering Trends
ISSN : 2682 - 2091
Vol.43, No.1. January 2024
http://jaet.journals.ekb.eg

Wind Forecasting Based on Hybrid Stochastic Scheme

Tamer Mohamed El-Saied[1], Ramadan Mohamed Mostafa Elwan[2], Omar Mohamed Salim[1]
Electrical Engineering Department Benha Faculty of Engineering Benha, Egypt Tamer.alkhodhary@bhit.bu.edu.eg[1]
Operation engineer in Bilbies, Canal Electricity Distribution, Bilbies, Egypt r.elwan50915@beng.bu.edu.eg [2]
Electrical Engineering Department Benha Faculty of Engineering Benha, Egypt omar.salim@bhit.bu.edu.eg[1]

A B S T R A C T

Wind forecasting has gained considerable interest due to the abundance of renewable energy and the rapid advancement of wind energy extraction technologies. Wind forecasting is the process of extracting one or more features from time series data to increase prediction accuracy. The various forecasting models for wind speed and power include physical, statistical, computer, and hybrid models. The steps involved in forecasting wind speed and energy are preprocessing the raw data, feature extraction, and prediction. In this work, hybrid model prediction algorithms are combined to obtain better forecasting accuracy and maintain model efficacy and simplicity. The proposed model combines either autoregressive or autoregressive integrated moving average with cumulative Weibull distribution. The results demonstrated an improvement in short- and medium-term prediction when compared to other computational techniques such as Weibull, (AR), and autoregressive integrated moving average (ARIMA). Numerical error evaluation approaches such as Mean Absolute Percentage Error Mean Square Error, and Mean Absolute Error were used to forecast the model's correctness. The results indicated that the hybrid model's projected error is signification less than that of the AR and ARIMA models independently.

Key words: *Wind forecasting, renewable energy, data preprocessing, Weibull, Autoregressive, and Autoregressive moving average.*

ABBREVIATIONS:

AR	*Autoregressive*	*MAPE*	*Mean Absolute Percentage Error*
ARIMA	*Autoregressive Integrated Moving Average*	*ENREA*	*Egyptian National Renewable Energy Authority*
ARMA	*Autoregressive Moving Average*	*MSE*	*Mean Square Error*
CDF	*Cumulative Distribution Function*	*MAE*	*Mean Absolute Error*
NWP	*Numerical Weather Prediction*	*RMSE*	*Root Mean Square Error*
PDF	*probability density function*	*MPF*	*Main Probabilistic Features*
MA	*Moving Average*	*Std*	*Standard Deviation*
WPF	*wind power forecasting*	*Var*	*Variance*
UQ	*uncertainty quantification*	*K, C*	Two factors of Weibull distribution

1. Introduction

It is commonly recognized that forecasting wind energy accurately can help mitigate the hazards associated with high wind energy penetration. Historically, wind energy forecasting has produced deterministic forecasts (i.e., point forecasts). Numerous academics have concentrated on minimizing predicting error using various statistical or physical models [1]. More precisely, point projections indicate the anticipated value of future wind energy. Wind power in the future, on the other hand, is a random variable with a probability density function (PDF), and point predictions nearly invariably omit this random variable's uncertainty information. This restricts the use of point forecasts in the investigation of the security and stability of electricity systems. The computational study of demonstrated that the developed hybrid models yield better performance contrast with those of other models involved in terms of both (wind speed, pressure, and temperature) deterministic and probabilistic forecasting. [2].

These categories have underlined the significance of forecast uncertainty and its facets for forecasters and decision-makers alike. Consequently, they have been increasing their efforts in uncertainty quantification (UQ) for the planning of wind farms (e.g. [3]), wind farm performance. (e.g. [4], [5]) and its application to operational forecasting and marketing practices [6], [7], [8], [9], [10]). However, many end-users' awareness and understanding of probabilistic forecasts, as well as their application of UQ to such forecasts, are not (yet) widespread enough to support uncertainty mitigation and improved use of uncertainty in WPF. (See e.g. [11], [12]).

According to the Egyptian National Renewable Energy Authority (NREA), Egypt generated approximately 1385 MW of wind energy in 2022 and its primary wind farm, Wind farms include 542.3 MW in Zafarana, 580 M W in Gabal El-Zeit, and 250 MW in Ras Ghareb. The Ministry of Electricity & Renewable Energy aims to maximize utilizing renewable energy in Egypt to reach about 20% of the total peak load by 2022, and up to 42% of the total generated energy by 2035, by adopting policies that encourage private sector investments in electricity production projects from renewable energies (wind and solar). [13]:

1. Egypt's wind resource is one of the best in the world due to its high and consistent wind speed.

2. Sufficient land is available at low economic rates.

Revised:15 June, 2022, Accepted: 20 August , 2022

3. Demand for electricity and other non-renewable energy sources is increasing significantly.

Forecasting is generally classified into several categories based on the time period involved: very short-term forecasting (a few seconds to 30 minutes ahead), short-term forecasting (30 minutes to several hours ahead), medium-term forecasting (several hours to one week ahead), and long-term forecasting (several hours to one week ahead) (from 1 week to 1 year or more).

In many cases, a substantial volume of data is required during the testing phase, as well as the development of adequate feasibility studies for future expenditures.

Numerous statistical methods and clever algorithms for deterministic wind power forecasting are described in the literature. The ARIMA model has been used in conjunction with the persistence approach to forecast short-term wind energy. [14]. Numerical weather prediction (NWP) is always used in conjunction with a physical model to improve forecasting accuracy [15].

The recent approaches are classified as follows: (i) physical algorithms, (ii) traditional statistical algorithms, (iii) spatial correlation algorithms, and (iv) machine-learning algorithms.

Physical algorithms primarily make use of meteorological environment data, which includes information on temperature, speed, density, and topography.

The primary objective of short-time wind speed forecast correction with the aim of improving decision support systems for traffic control in dangerous wind situation is also the problem of wind farm power prediction. For example, in the study [16] about wind farm NWP wind speed correction methods, measured time series were decomposed into different bands by wavelet multi-resolution analysis. Correction premise was verified using moment correlation coefficient, and then the linear correction method was used to stationary NWP wind speed. [17]. For example, Cheng et al. [18] combined current numerical weather forecast data with assimilation, with the result that prediction accuracy is significantly increased. Nonetheless, given the disadvantages associated with short-term wind speed forecasting, as well as the high cost in terms of computing time and resources, it is evident that this category is unsuitable for wind farm short-term wind speed forecasting.

Statistical methods generally refer to the application of mathematical statistics, probability theory, and stochastic processes to forecasting problems. They typically use a large amount of historical data for model training or error fitting, establish a mapping relationship between input variables and output variables, and predict the future wind power value (or interval of such). Established techniques include exponential smoothing, auto-regressive moving average (ARMA) models, and auto-regressive integrated moving average (ARIMA) models. These can be applied to both deterministic and probabilistic forecasting [19]. The authors suggested a modified ARIMA method for predicting wind speed, and the results indicate that it is more accurate than the AR model. However, statistical algorithms continue to have some flaws.

To begin, the majority of statistical techniques assume a normal distribution for time series, which is not necessarily the case for wind speed time series. Second, these models have a linear correlation structure, which results in low accuracy when nonlinear wind speed data is used. To address these issues, spatial correlation methods are used that take into account the geographical link between wind speeds measured at various locations. For example, developed a unique wind speed prediction model based on a wavelet transform and a spatio-temporal technique that outperformed existing benchmark models in terms of short-term wind speed forecasting. However, in the interim, this model [19].

2. METHODOLOGY AND DATA SOURCE

To determine an ARIMA model's adaptability, we must first ensure that the stationary and invertibility assumptions are met. While all ARIMA (p, d, 0) models are stationary, in order to determine whether the model was chosen correctly, we must examine whether the time series satisfy the other criteria of invertibility. Due to the fact that ARIMA (p, d, q) models are invertible and dependent on the parameter values, they may not be stationary. As a result, the model is represented in a variety of ways. That is why it is prudent to seek out the simplest representations for wind speed estimation.

Granger and Andersen proposed a broader definition of inversion in 1978, which they applied to linear, nonlinear, and bilinear models [21]. As can be shown, some non-linear models are not invertible, however this criterion can be satisfied by combining them with another model. To describe the criteria for a general Movable Average (MA) process of order q, the input data must be invertible (accordingly, process borderline should be non-invertible). Acceptability conditions must be used to refer to the conditions.

The dependency can be found on the magnitude of the final moving average parameter, θ_q. If $|\theta_q| < 1$, the process is not acceptable. The process should reach the conditions $|\theta_q| = 1$ for any particular q meaning and is expected to run smoothly. If $|\theta_q| < 1$, the conditions need to be established. Simultaneously, the stationarity of autoregressive processes is examined. In 2008, Ojo compared subset autoregressive integrated moving average models with full autoregressive integrated moving average models [22].

The purpose of this work is to conduct a trial to determine the most efficient autoregressive integrated moving average (ARIMA) model structure that achieves the lowest errors when comparing forecast and real-time series scenarios. The research was conducted to anticipate daily wind speeds in Hurghada for June, July, August, and September 2021 using this model. It is chosen to evaluate the model's predicting performance using MAPE, RMSE, and MAE (mean absolute error).

2.1 Time series analysis with ARIMA models

The majority of modeling techniques, including Box-Jenkins [23], are applicable to stationary time series. ARIMA models are a class of time series that are based on statistical models and are frequently used for short-term forecasting. A typical ARIMA model, represented by ARIMA (p, d, q), is as follows:

$$ARMA(p,q):\ Y_t = C + \sum_{i=1}^{p} \emptyset_i Y_{t-1} + \sum_{j=1}^{q} \theta_j \epsilon_{t-1} + \epsilon_t \quad (1)$$

This equation can be written as:

$$Y_t = c + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \ldots\ldots + \emptyset_p Y_{t-p} + \epsilon_t \quad (2)$$

where C is the constant term (i. e. the mean of the underlying stochastic process), $\emptyset_i$ is the i-th autoregressive parameter, θ_j is the j- the moving average parameter, ϵ_t is the error term at time t, and Y_t is the value of the wind speed observed at the time t [18]. The autoregressive parameters represent the lags of differenced series, and the moving average terms show the lags of the prediction errors. It is possible that the time series data is non-stationary (or seasonal), and in this case it needs to be differenced to become stationary [22]. The result is an "integrated" version of a stationary series, and the model becomes an ARIMA model, denoted by ARIMA (p, d, q), where p, d, and q are the numbers of autoregressive terms, non-seasonal differences, and lagged prediction errors, respectively. Clearly, if d is zero, the ARIMA model becomes an ARMA (p, q) model. If both d and q are zero, then the ARIMA model becomes an AR (p) model. If both d and p are zero, then the ARIMA model becomes a MA (q) model.[23]

In this study, we employ the general procedure of ARIMA, AR and Weibull distribution modelling for the prediction of wind speed. Based on the data obtained, a suitable model structure and model parameters will be obtained.

2.2 Model precision analysis

The root mean-square error (RMSE), mean absolute percentage error (MAPE) and mean absolute error (MAE) are adopted to evaluate the prediction accuracy of the approaches [24]. MAE is a common measure of the forecast error in time series analysis, which measures the average magnitude of the errors in a set of forecasts:

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}\left|\left(y_{t-f_t}\right)\right| \quad (3)$$

where n is the number of observations in the total evaluation period, y_t is the value of observation at time t, and f_t is the forecast value.

Equation (3) shows that MAE is the average over the absolute values of deviations between the forecast and the corresponding observation [18]. MAPE is calculated as the average absolute percentage error:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\left\{\left(y_{t-f_t}\right)\right)}{y_t}\right| \quad (4)$$

As seen in equation (4), the main purpose of MAPE is to show if the data is stable (variation is small). That is why MAPE is important in wind power prediction.

$$\text{RMSE} = \sqrt{\left\{\frac{1}{N}\sum_{t=1}^{n} y_t - f_t\right)^2} \quad (5)$$

Equation (5) indicates that RMSE is a quadratic scoring rule, which measures the average magnitude of the error [24]. The difference between forecasts and corresponding observed values are squared, summed, and then averaged over the sample number. Finally, the square root of the average is taken. Since the errors are squared before they are averaged, RMSE gives relatively high weights to large errors. This means RMSE is most useful when large errors are undesirable.

2.3 Wind data

Meteorological data was received from weather underground site. They can be denominated as all-Hurghada city averages. All variables are presented as hourly data. The investigation period covers the period of 2021. Real data from 1st June 2021to 29th September 2021 and forecasting data from 29th September to 2nd October 2021. for all experiments.

The RMSE, MAE, and MAPE criteria were used to compare actual and forecasted results for two days+ period (53hr). For ARIMA and AR models, the forecasted data calculated for three different sample size, small size up to 400 sample (200 and 400 used), medium size up to 1500 sample and large size for more than 1500 sample (2900 used).

And the number of samples divided is a smaller number of samples (200, 400), a medium number of samples (1000 and 1500), and a large number of samples (2900). and calculate the error in predicted values as the number of samples increases. because with increase the numbers of samples increase the error.

Different forecast errors are shown after 53 hours in two models. When the ARIMA (12, 0, 4), and AR (12,0,0) model is selected, the input data period (200,400,1000,1500, and 2900) should be analyzed. to show the main results of analyzed input periods. Errors made from the beginning until the end of the period can reach 30% of the worst accuracy of forecast values, especially in the last analyzed period. As mentioned before, to find the best model

structure and input periods, the MAE, RMSE, and MAPE were used. Each error shows different changes in analyzed data. It is clear that the fifth period causes at the highest RMSE errors in 2 models, and the first input period is the best. Moreover, in the second addition more error, the prediction of the first input period is better than the prediction of the second, but the errors increase with the number of entries of the model.

models ARIMA (12,0,4) and AR (12,0,0) show that RMSE and MAE errors have very similar results. The 53-hour (two days+) actual and predicted wind speed values of (200,400,1000,1500 and 2900) were analyzed. The difference in error value was the lowest between I and II input periods in the beginning of the forecasted 53 hours ahead. However, for 53-hour wind speed prediction, we received the lowest RMSE and MAE errors when the model was trained.

3. EXPERIMENTS AND RESULTS

The research consists of two parts: in the first model used the ARIMA, AR model individually and in the second model used a hybrid with Weibull distribution model.

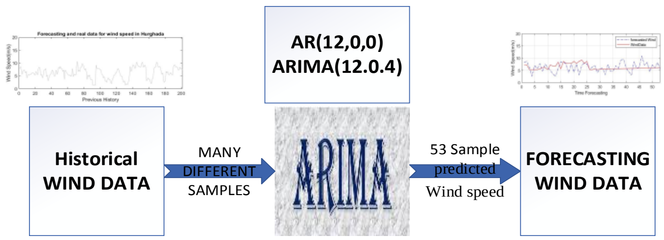

Figure 1 explain first model used the ARIMA, AR model individually.

In Figure 1 The model in the illustration employs Arima (12,0,4) and AR (12,0,0). With five alternative sample sizes (200,400,1000,1500,2900) historical data entered the ARIMA (12.0.4) and AR (12,0,0) models, to predict wind speed and compare forecasting data with real data for 53 samples (two days+).

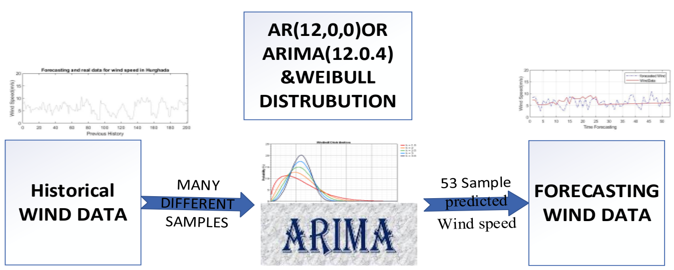

Figure 2 explain the second model used a hybrid model with Weibull distribution.

In Figure 2 The model in the figure employs an ARMIA (12,0,4) hybrid with a Weibull distribution. and the AR (12,0,0) hybrid with the Weibull distribution. Input historical data and use 5 samples (200,400,1000,1500,2900) into the Arima and AR model with fixed order ARIMA (12.0.4) and AR (12,0,0), predict wind speed and compare forecasting data with real data for 53 samples. at five samples.

3.1 model-I: 53-Hrs Ahead Based on Autoregressive moving average. (12,0,4).

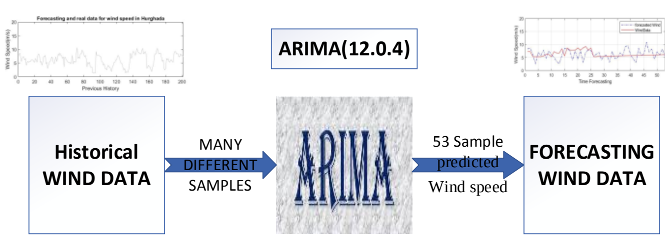

Figure 3 explain first model used the ARIMA for all samples (200,400,1000,1500 and 2900) and fixed orders (12,0,4).

In Figure 3 The model in the illustration employs ARIMA (12,0,4), Input historical data and use 5 samples (200,400,1000,1500,2900) into the Arima and AR model with fixed order ARIMA (12.0.4) and AR (12,0,0), predict wind speed and compare forecasting data with real data for 53 samples. at five samples.

In Figure 4 using the ARIMA model (12, 0, 4), and 200-hour sample input before the 53 sample output forecasts, wind speed predictions were made for the next 53 hours (2.2 days). Error computed, RMSE, MAPE, MSE.

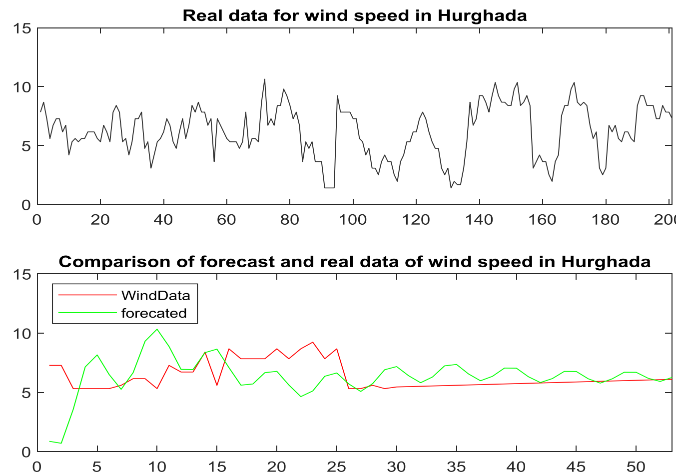

Figure 4 ARIMA (12,0,4) With 200 samples and 53-hrs forecasting.

Table 1: ARIMA (12,0,4) With 200 samples and 53-hrs forecasting.

Real data	forecasting	error	RMSE	MAPE	MSE
6.3639	6.3595	0.0044	2.0852	0.2962	4.3481

In table 1 the first sample is 200 hours. This indicates that the data from September 20 to 29, 2021 in the ARIMA model (12,0,4) to predict wind speed in the autumn season. The forecasted wind speed is close to the real wind speed and the error is approximately 0.0044 m/s. MAPE is approximately 0.3. It is the best model for short term forecasting.

In Figure 5 using the ARIMA model (12, 0, 4), and 400-hour sample input before the 53 sample output forecasts, wind speed predictions were made for the next 53 hours (2.2 days). Error computed, RMSE, MAPE, MSE.

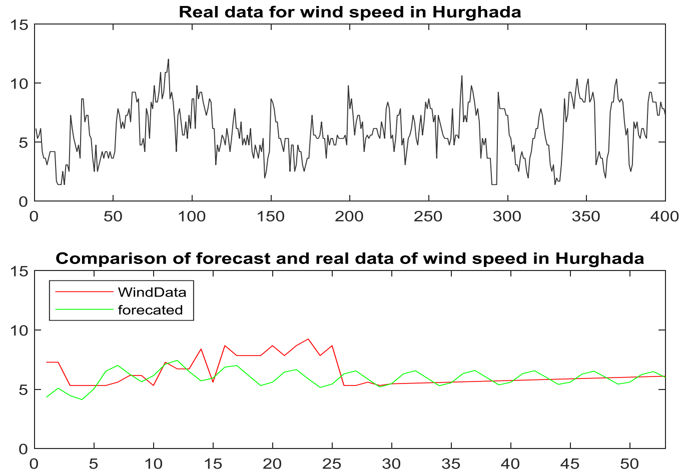

Figure 5 ARIMA (12,0,4) With 400 samples and 53-hrs forecasting.

Table 2: ARIMA (12,0,4) With 400 samples and 53-hrs forecasting.

Real data	**forecasting**	**error**	**RMSE**	**MAPE**	**MSE**
6.3639	**5.9479**	**0.4160**	**1.3347**	**0.20**	**1.7814**

In table 2 the second sample is 400 hours. This indicates that the data is from September 11 to 29. 2021 must be used in the ARIMA (12,0,4) form for the Autumn season. The forecasted wind speed is close to the real wind speed and the error is approximately 0.4160 m/s. The error rate increases by 94% when the input period is doubled MAPE is approximately 0.2.

In Figure 6 using the ARIMA model (12, 0, 4), and 1000-hour sample input before the 53 sample output forecasts, wind speed predictions were made for the next 53 hours (2.2 days). Error computed, RMSE, MAPE, MSE

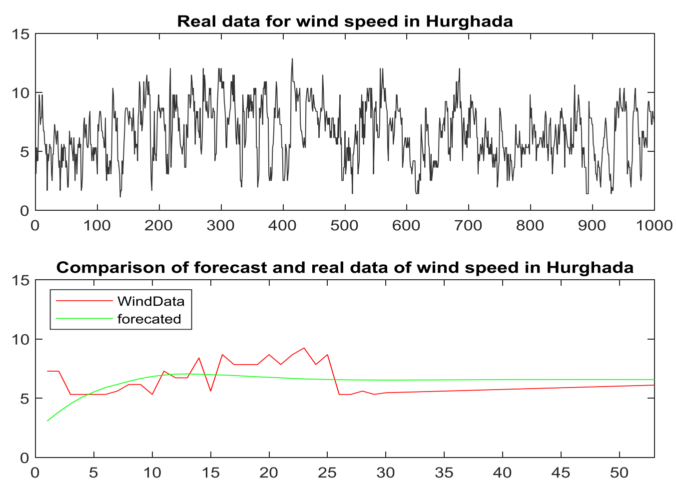

Figure 6 ARIMA (12,0,4) With 1000 samples and 53-hrs forecasting.

Table 3: ARIMA (12,0,4) With 1000 samples and 53-hrs forecasting

Real data	**forecasting**	**error**	**RMSE**	**MAPE**	**MSE**
6.3639	**6.4197**	**-0.0558**	**1.3347**	**0.20**	**2.7814**

In table 3 the third sample is 1000 hours. This indicates that the data is from august 18 to September 29, 2021. used in the ARIMA (12,0,4) form for the Autumn season. The forecasted wind speed is great than the real wind speed and the error is approximately -0.0558 m/s. MAPE is approximately 0.26.

In Figure 7 using the ARIMA model (12, 0, 4), and 1500-hour sample input before the 53 sample output forecasts, wind speed predictions were made for the next 53 hours (2.2 days). Error computed, RMSE, MAPE, MSE.

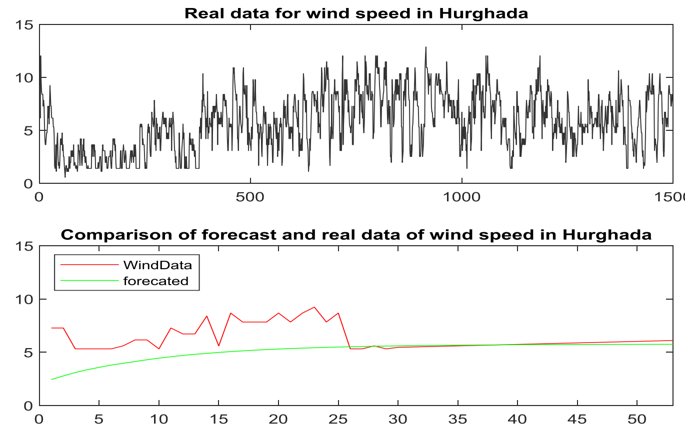

Figure 7 ARIMA (12,0,4) With 1500 samples and 53-hrs forecasting.

Table 4: ARIMA (12,0,4) With 1500 samples and 53-hrs forecasting.

Real data	**forecasting**	**error**	**RMSE**	**MAPE**	**MSE**
6.3639	**5.1305**	**1.2334**	**1.8951**	**0.26**	**3.5915**

In table 4 the fourth sample is 1500 hours. This indicates that the data is from July 29 to September 29, 2021. used in the ARIMA (12,0,4) form for the Autumn season. The forecasted wind speed is low than the real wind speed and the error is approximately 1.2334 m/s. MAPE is approximately 0.26

In Figure 8 using the ARIMA model (12, 0, 4), and 2900-hour sample input before the 53 sample output forecasts, wind speed predictions were made for the next 53 hours (2.2 days). Error computed, RMSE, MAPE, MSE.

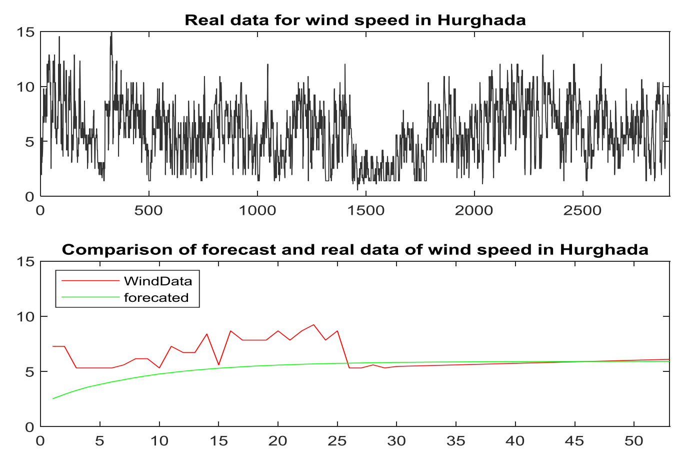

Figure 8 ARIMA (12,0,4) With 2900 samples and 53-hrs forecasting.

Table 5: ARIMA (12,0,4) With 2900 samples and 53-hrs forecasting.

Real data	**forecasting**	**error**	**RMSE**	**MAPE**	**MSE**
6.3639	**5.3618**	**1.0021**	**1.7364**	**0.2421**	**3.0151**

In table 5 the fifth sample is 2900 hours. This indicates that the data is from June 1, 2021 to September 29, 2021. Used in the

ARIMA model (12,0,4) for the fall season. The forecasted wind speed is lower than the real wind speed and the error is about 1.0021 m/s. MAPE is about 0.2421.

3.2 model-II: 53-Hrs Ahead Based on Autoregressive (12,0,0).

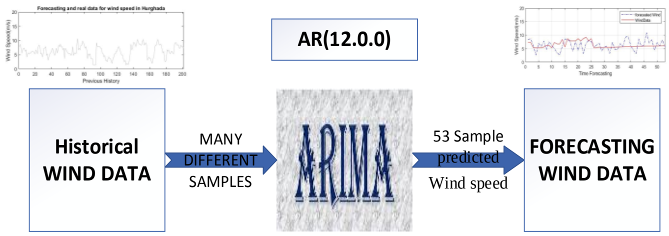

Figure 9 explain part 2 of first model used the AR for all samples (200,400,1000,1500 and 2900) and fixed orders (12,0,0).

In Figure 10 using the AR model (12, 0,0), and 200-hour sample input before the 53 sample output forecasts, wind speed predictions were made for the next 53 hours (2.2 days). Error computed, RMSE, MAPE, MSE.

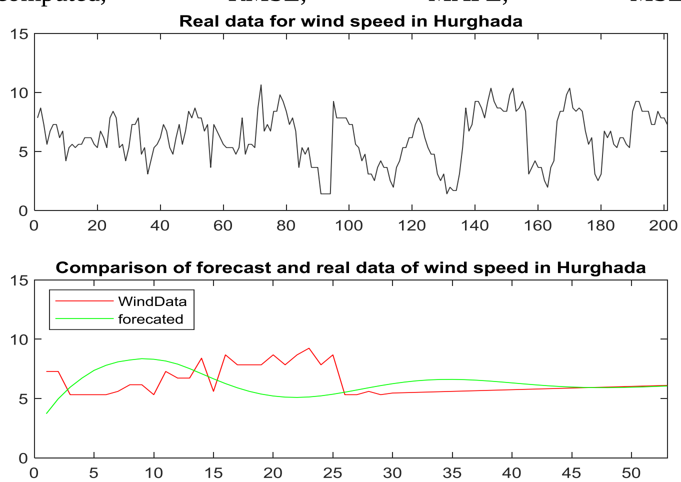

Figure 10 AR (12,0,0) With 200 samples and 53-hrs forecasting.

Table 6: AR (12,0,4) With 200 samples and 53-hrs forecasting.

Real data	forecasting	error	RMSE	MAPE	MSE
6.3639	6.2962	0.0677	1.7298	0.27	2.9921

In table 6 the first sample is 200 hours. This indicates that the data from September 20 to 29, 2021 in the AR model (12,0,0) to predict wind speed in the autumn season. The forecasted wind speed is close to the real wind speed and the error is approximately 0.0677m/s. MAPE is approximately 0.27.

In Figure 11 using the AR model (12, 0,0), and 400-hour sample input before the 53 sample output forecasts, wind speed predictions were made for the next 53 hours (2.2 days). Error computed, RMSE, MAPE, MSE.

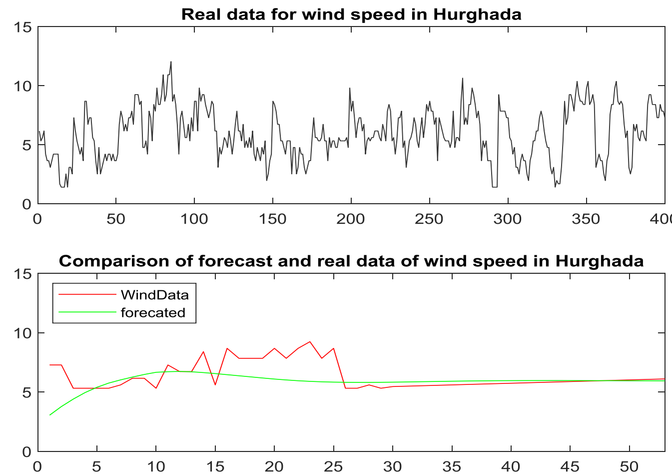

Figure 11 AR (12,0,0) With 400 samples and 53-hrs forecasting.

Table 7: AR (12,0,0) With 400 samples and 53-hrs forecasting.

Real data	forecasting	error	RMSE	MAPE	MSE
6.3639	5.8999	0.4640	1.3223	0.17	1.7486

In table 7 the second sample is 400 hours. This indicates that the data is from September 11 to 29. 2021 must be used in the AR (12,0,0) form for the Autumn season. The forecasted wind speed is close to the real wind speed and the error is approximately 0.4640 m/s. MAPE is approximately.0.17

In Figure 12 using the AR model (12, 0,0), and 1000-hour sample input before the 53 sample output forecasts, wind speed predictions were made for the next 53 hours (2.2 days). Error computed, RMSE, MAPE, MSE.

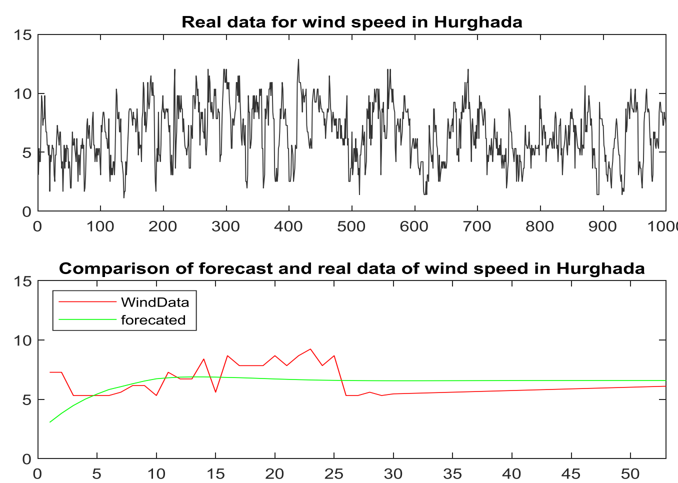

Figure 12 AR (12,0,0) With 1000 samples and 53-hrs forecasting.

Table 8: AR (12,0,0) With 1000 samples and 53-hrs forecasting.

Real data	forecasting	error	RMSE	MAPE	MSE
6.3639	6.3954	-0.0315	1.2869	0.21	1.6562

In table 8 the third sample is 1000 hours. This indicates that the data is from august 18 to September 29, 2021. used in the AR (12,0,0) form for the Autumn season. The forecasted wind speed is great than the real wind speed and the error is approximately -0.0315 m/s. MAPE is approximately 0.21.

In Figure 13 using the AR model (12, 0,0), and 1500-hour sample input before the 53 sample output forecasts, wind speed predictions were made for the next 53 hours (2.2 days). Error computed, RMSE, MAPE, MSE.

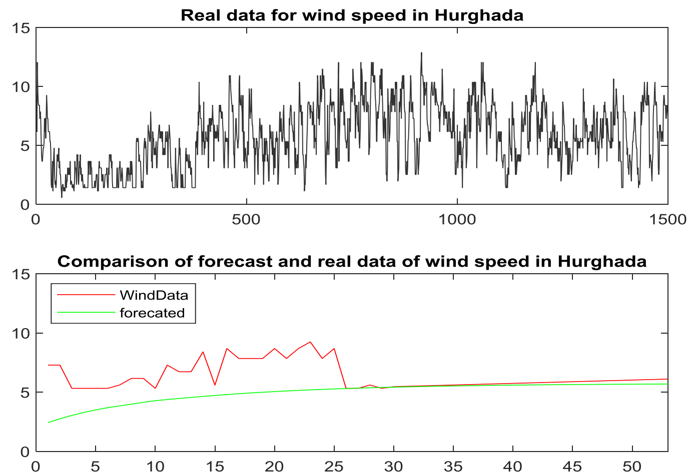

Figure 13 AR (12,0,0) With 1500 samples and 53-hrs forecasting.

Table 9: AR (12,0,0) With 1500 samples and 53-hrs forecasting

Real data	forecasting	error	RMSE	MAPE	MSE
6.3639	4.9771	1.3868	2.0160	0.28	4.0641

In table 9 the fourth sample is 1500 hours. This indicates that the data is from July 29 to September 29, 2021. used in the AR (12,0,0) form for the Autumn season. The forecasted wind speed is low than the real wind speed and the error is approximately 1.3868 m/s. MAPE is approximately 0.28.

In Figure 14 using the AR model (12, 0,0), and 2900-hour sample input before the 53 sample output forecasts, wind speed predictions were made for the next 53 hours (2.2 days). Error computed, RMSE, MAPE, MSE.

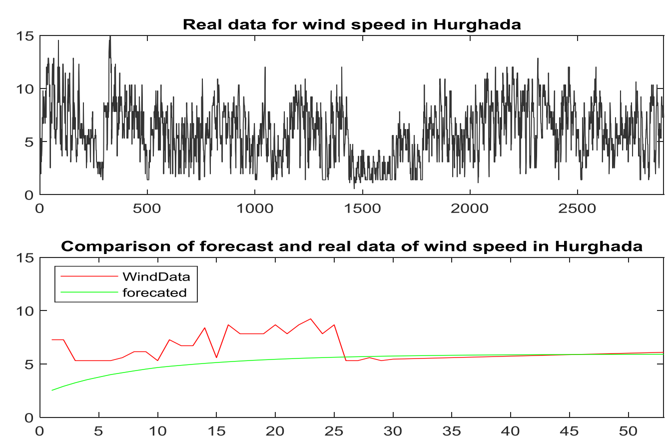

Figure 14 AR (12,0,0) With 2900 samples and 53-hrs forecasting.

Table 10: AR (12,0,0) With 2900 samples and 53-hrs forecasting.

Real data	forecasting	error	RMSE	MAPE	MSE
6.3639	5.2955	1.0684	1.7955	0.25	3.2240

In table 10 the fifth sample is 2900 hours. This indicates that the data is from June 1, 2021 to September 29, 2021. Used in the AR model (12,0,0) for the autumn season. The forecasted wind speed is lower than the real wind speed and the error is about 1.7955 m/s. MAPE is about 0.25.

After using ARIMA and AR individually, why is the Weibull distribution used? The Weibull distribution is a probability distribution function of wind speed characterizing the frequency of occurrence of wind speed at a site. The Weibull distribution function is described by two factors: K which is the form factor and C which is the scale factor (some references use A instead of C).[25] Thus, the Weibull distribution represents the wind speed distribution. The Weibull distribution describes the distribution of wind speed because the distribution of wind speed is well fitted by the Weibull distribution.[26] Due to the random nature of wind, it could be represented by a probabilistic model. Usually, Weibull and Rayleigh models are used to describe the probability distribution for a measured wind speed in a specific location during a certain time [27]. Estimating wind Weibull for a location could be used in distributed wind power generation planning. Moreover, it can be used to predict Weibull for cyber-physical system sensor network as in [27,28], In other words, the Weibull distribution is proportional to the wind speed distribution close to the actual wind speed distribution, so using the help of the Weibull distribution with the hybrid model with ARIMA and AR to get more accurate results.

3.3 model-II: 53-Hrs Ahead Based on Autoregressive moving average With Weibull distribution (12,0,4).

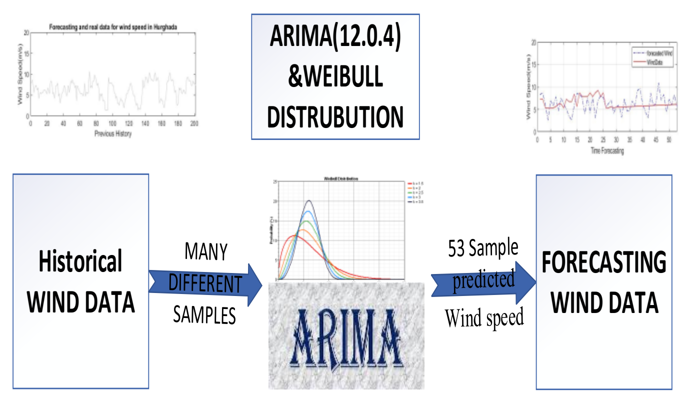

Figure 15 ARIMA (12,0,4) and Weibull distribution with 53-hrs forecasting.

In Figure 16 using the ARIMA model (12,0,4) and Weibull distribution model, 200-hour sample input before the 53 sample output forecasts, wind speed predictions were made for the next 53 hours (2.2 days). Error computed, RMSE, MAPE, MSE.

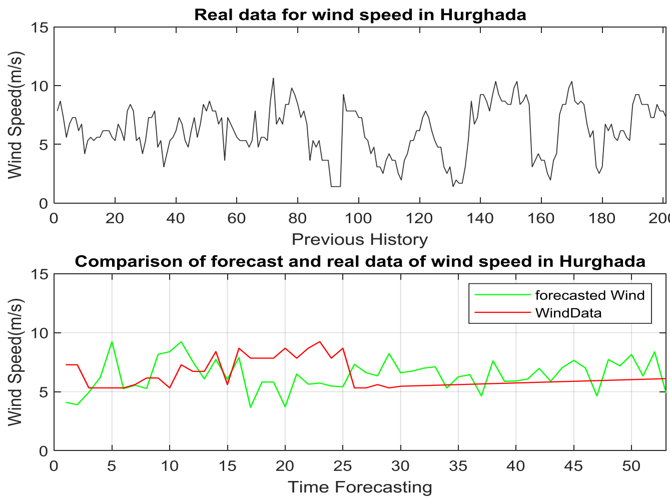

Figure 16 ARIMA (12,0,4) and Weibull distribution with 200 samples with 53-hrs forecasting.

Table 11: ARIMA (12,0,4) and Weibull distribution with 200 samples and 53-hrs forecasting.

Real data	forecasting	error	RMSE	MAPE	MSE
6.3639	6.3197	0.0443	1.6563	0.27	1.6563

In table 11 the first sample is 200 hours. This indicates that the data from September 20 to 29, 2021 in the ARIMA model (12,0,4) and the Weibull distribution model to predict wind speed in the autumn season. The forecasted wind speed is close to the real wind speed and the error is approximately 0.0443m/s. MAPE is approximately 0.27.

In Figure 17 using the ARIMA model (12,0,4) and Weibull distribution model, 400-hour sample input before the 53 sample output forecasts, wind speed predictions were made for the next 53 hours (2.2 days). Error computed, RMSE, MAPE, MSE.

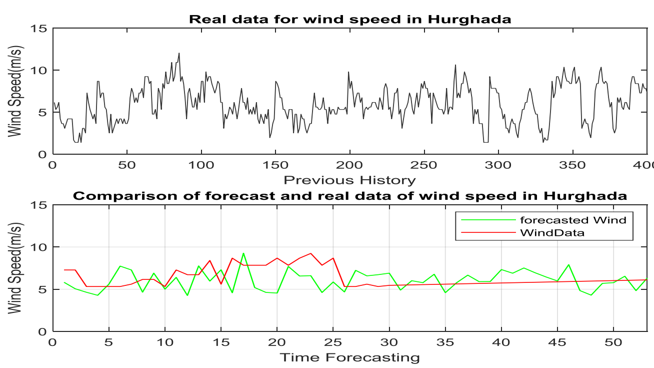

Figure 17 ARIMA (12,0,4) and Weibull distribution with 400 samples and 53-hrs forecasting.

Table 12: ARIMA (12,0,4) and Weibull distribution with 400 samples and 53-hrs forecasting.

Real data	forecasting	error	RMSE	MAPE	MSE
6.3639	6.0365	0.3275	1.6924	0.28	2.8642

In table 12 the second sample is 400 hours. This indicates that the data is from September 11 to 29. 2021 must be used in the ARIMA (12,0,4) and the Weibull distribution model form for the Autumn season. The forecasted wind speed is close to the real wind speed and the error is approximately 0.3275 m/s. MAPE is approximately 0.28.

In Figure 18 using the ARIMA model (12,0,4) and Weibull distribution model, 1000-hour sample input before the 53 sample output forecasts, wind speed predictions were made for the next 53 hours (2.2 days). Error computed, RMSE, MAPE, MSE.

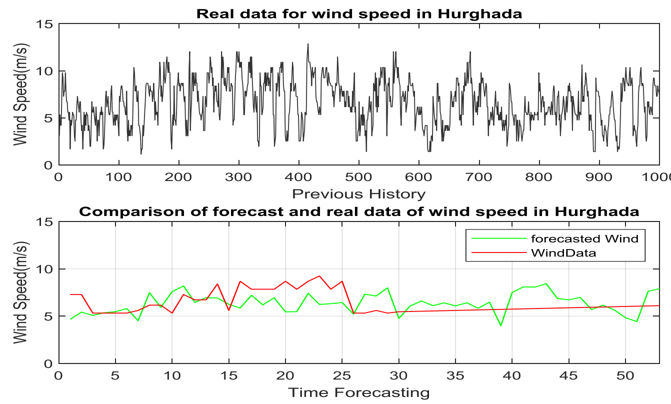

Figure 18 ARIMA (12,0,4) and Weibull distribution with 1000 samples and 53-hrs forecasting.

Table 13: ARIMA (12,0,4) and Weibull distribution with 1000 samples and 53-hrs forecasting.

Real data	forecasting	error	RMSE	MAPE	MSE
6.3639	6.3545	0.0094	1.5199	0.26	2.3102

In table 13 the third sample is 1000 hours. This indicates that the data is from august 18 to September 29, 2021. used in the ARIMA (12,0,4) and the Weibull distribution model form for the Autumn season. The forecasted wind speed is great than the real wind speed and the error is approximately 0.0094 m/s. MAPE is approximately 0.26.

In Figure 19 using the ARIMA model (12,0,4) and Weibull distribution model, 1500-hour sample input before the 53 sample output forecasts, wind speed predictions were made for the next 53 hours (2.2 days). Error computed, RMSE, MAPE, MSE.

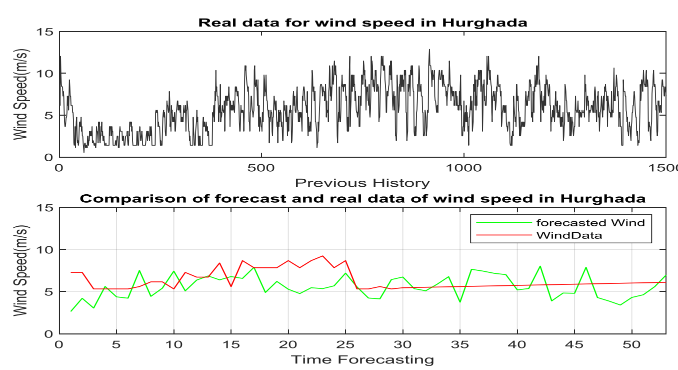

Figure 19 ARIMA (12,0,4) and Weibull distribution with 1500 samples and 53-hrs forecasting.

Table 14: ARIMA (12,0,4) and Weibull distribution with 1500 samples and 53-hrs forecasting.

Real data	forecasting	error	RMSE	MAPE	MSE
6.3639	5.5852	0.7787	1.8780	0.32	3.5267

In table 14 the fourth sample is 1500 hours. This indicates that the data is from July 29 to September 29, 2021. used in the

ARIMA (12,0,4) and the Weibull distribution model form for the Autumn season. The forecasted wind speed is low than the real wind speed and the error is approximately 0.7787 m/s. MAPE is approximately 0.32.

In Figure 20 using the ARIMA model (12,0,4) and Weibull distribution model, 2900-hour sample input before the 53 sample output forecasts, wind speed predictions were made for the next 53 hours (2.2 days). Error computed, RMSE, MAPE, MSE

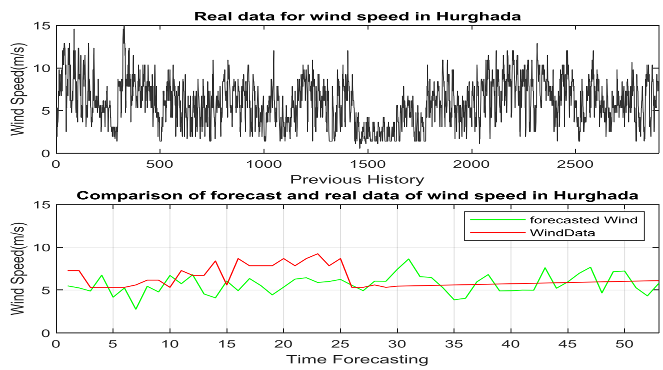

Figure 20 ARIMA (12,0,4) and Weibull distribution with 2900 samples and 53-hrs forecasting.

Table 15: ARIMA (12,0,4) and Weibull distribution with 400 samples and 53-hrs forecasting.

Real data	forecasting	error	RMSE	MAPE	MSE
6.3639	5.6879	0.6761	1.7563	0.29	3.0847

In table 15 the fifth sample is 2900 hours. This indicates that the data is from June 1, 2021 to September 29, 2021. Used in he ARIMA model (12,0,4) and the Weibull distribution model from for the autumn season. The forecasted wind speed is lower than the real wind speed and the error is about 0.6761 m/s. MAPE is about 0.29.

3.4 model-III: 53-Hrs Ahead Based on Autoregressive with Weibull distribution (12,0,0).

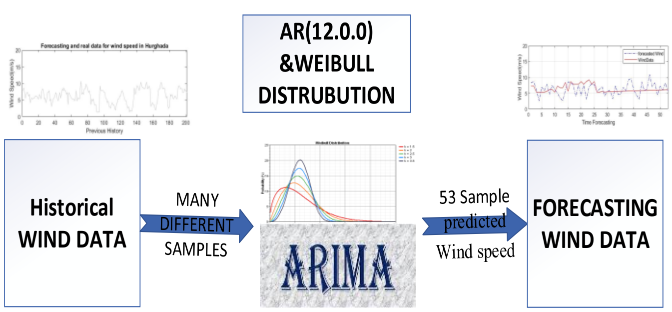

Figure 21 AR (12,0,0) and Weibull distribution and 53-hrs forecasting.

In Figure 22 using the AR model (12,0,0) and Weibull distribution model, 200-hour sample input before the 53 sample output forecasts, wind speed predictions were made for the next 53 hours (2.2 days). Error computed, RMSE, MAPE, MSE.

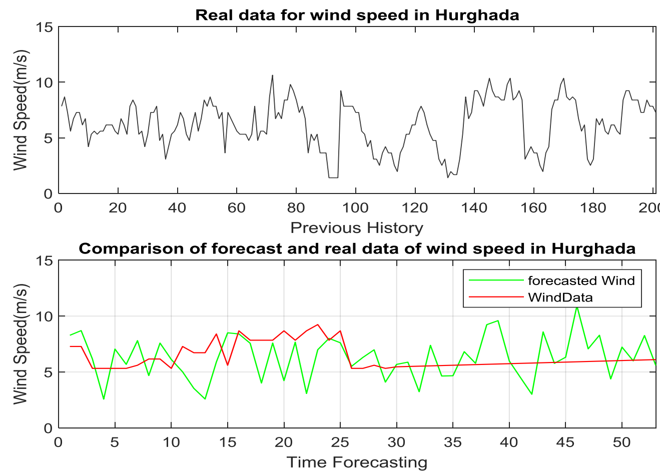

Figure 22 AR (12,0,0) and Weibull distribution with 200 samples and 53-hrs forecasting.

Table 16: AR (12,0,0) and Weibull distribution with 200 samples and 53-hrs forecasting.

Real data	forecasting	error	RMSE	MAPE	MSE
6.3639	6.2880	0.0759	1.5801	0.25	2.4968

In table 16 the first sample is 200 hours. This indicates that the data from September 20 to 29, 2021 in the AR model (12,0,0) and the Weibull distribution model to predict wind speed in the autumn season. The forecasted wind speed is close to the real wind speed and the error is approximately 0.0759m/s. MAPE is approximately 0.25.

In Figure 23 using the AR model (12,0,0) and Weibull distribution model, 400-hour sample input before the 53 sample output forecasts, wind speed predictions were made for the next 53 hours (2.2 days). Error computed, RMSE, MAPE, MSE.

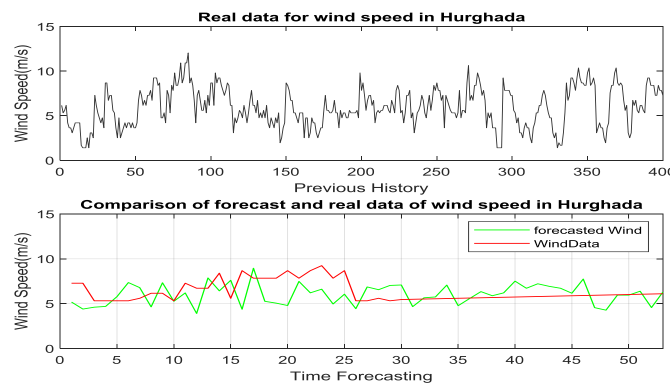

Figure 23 AR (12,0,0) and Weibull distribution with 400 samples and 53-hrs forecasting.

Table 17: AR (12,0,0) and Weibull distribution with 400 samples and 53-hrs forecasting.

Real data	forecasting	error	RMSE	MAPE	MSE
6.3639	6.0125	0.3514	1.6916	0.28	2.8617

In table 17 the second sample is 400 hours. This indicates that the data is from September 11 to 29. 2021 must be used in the AR (12,0,0) and the Weibull distribution model form for the Autumn season. The forecasted wind speed is close to the real wind speed and the error is approximately 0.3514 m/s. MAPE is approximately 0.28.

In Figure 24 using the AR model (12,0,0) and Weibull distribution model, 1000-hour sample input before the 53 sample output forecasts, wind speed predictions were made for the next 53 hours (2.2 days). Error computed, RMSE, MAPE, MSE.

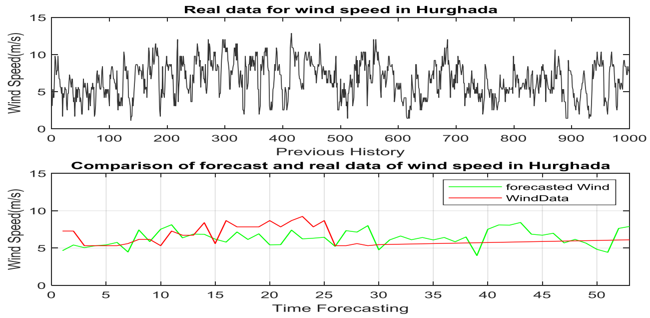

Figure 24 AR (12,0,0) and Weibull distribution with 1000 samples and 53-hrs forecasting.

Table 18: AR (12,0,0) and Weibull distribution with 1000 samples and 53-hrs forecasting.

Real data	forecasting	error	RMSE	MAPE	MSE
6.3639	6.3424	0.0216	1.5260	0.26	2.3286

In table 18 the third sample is 1000 hours. This indicates that the data is from august 18 to September 29, 2021. used in the AR (12,0,0) and the Weibull distribution model form for the Autumn season. The forecasted wind speed is great than the real wind speed and the error is approximately 0.0216 m/s. MAPE is approximately 0.26.

In Figure 25 using the AR model (12,0,0) and Weibull distribution model, 1500-hour sample input before the 53 sample output forecasts, wind speed predictions were made for the next 53 hours (2.2 days). Error computed, RMSE, MAPE, MSE.

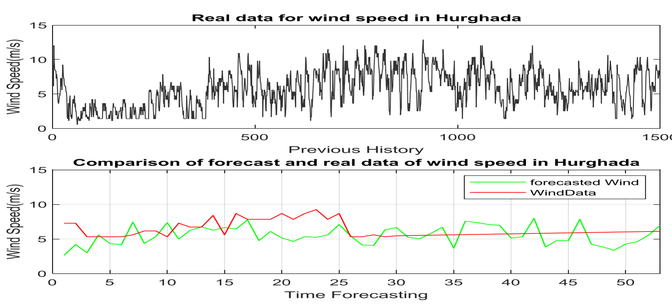

Figure 25 AR (12,0,0) and Weibull distribution with 1500 samples and 53-hrs forecasting.

Table 19: AR (12,0,0) and Weibull distribution with 1500 samples and 53-hrs forecasting.

Real data	Forecasting	Error	RMSE	MAPE	MSE
6.3639	5.5085	0.8555	1.9141	0.33	3.6433

In table 19 the fourth sample is 1500 hours. This indicates that the data is from July 29 to September 29, 2021. used in the AR (12,0,0) and the Weibull distribution model form for the Autumn season. The forecasted wind speed is low than the real wind speed and the error is approximately 0.8555 m/s. MAPE is approximately 0.33.

In Figure 26 using the AR model (12,0,0) and Weibull distribution model, 2900-hour sample input before the 53 sample output forecasts, wind speed predictions were made for the next 53 hours (2.2 days). Error computed, RMSE, MAPE, MSE.

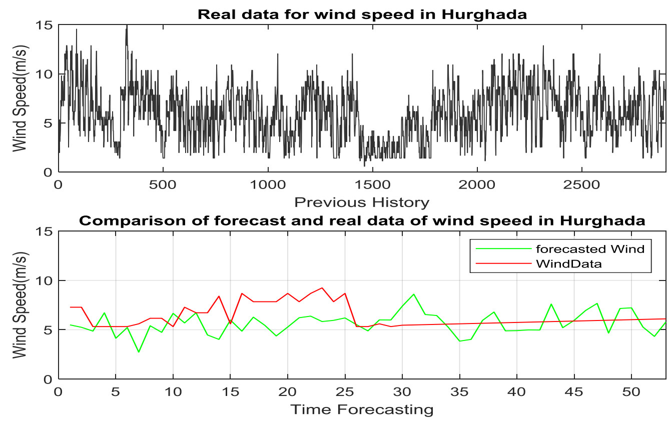

Figure 26 AR (12,0,0) and Weibull distribution with 2900 samples and 53-hrs forecasting.

Table 20: AR (12,0,0) and Weibull distribution with 2900 samples and 53-hrs forecasting.

Real data	forecasting	error	RMSE	MAPE	MSE
6.3639	5.8441	0.7092	1.7797	0.30	3.1672

In table 20 the fifth sample is 2900 hours. This indicates that the data is from June 1, 2021 to September 29, 2021. Used in the AR model (12,0,0) and the Weibull distribution model from for the autumn season. The forecasted wind speed is lower than the real wind speed and the error is about 0.7092 m/s. MAPE is about 0.30.

4 RESULT DISCUSSION.

Table 21 Mean 53-hrs forecasting Based on 5 model (ARIMA-AR-WEIBULL).

6.3639 m/s Real Data	ARIMA (12.0.4)	AR (12.0.0)	ARIMA & WEIBULL	AR & WEIBULL
200 sample	6.3595	6.2962	6.3197	6.2880
400 sample	5.9479	5.8999	6.0365	6.0125
1000 sample	6.4197	6.3954	6.3845	6.3424
1500 sample	5.1305	4.9771	5.5825	5.5085
2900 sample	4.4673	5.2955	5.6879	5.6547

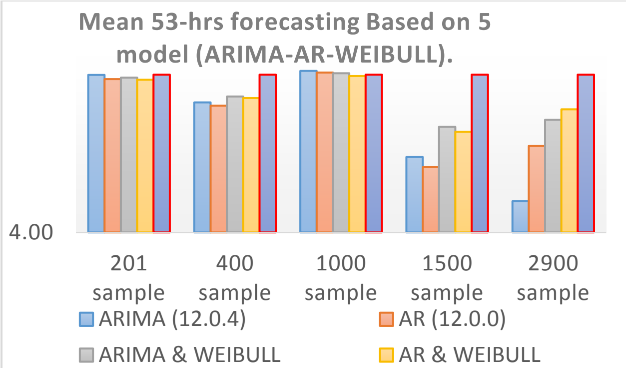

Figure 27 Mean 53-hrs forecasting Based on 5 model (ARIMA-AR-WEIBULL).

Table 22: Error 53-hrs forecasting Based on 5 model (ARIMA-AR-WEIBULL).

No. of Samples	ARIMA (12.0.4)	AR (12.0.0)	ARIMA & WEIBULL	AR & WEIBULL
200	0.0044	0.0677	0.0443	0.0759
400	0.4160	0.4640	0.3275	0.3514
1000	-0.0558	-0.0315	-0.0094	0.0216
1500	1.2334	1.3868	0.7787	0.81533
2900	1.8967	1.0684	0.6761	0.7092

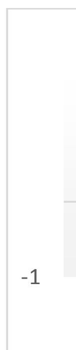

Figure 28 Error 53-hrs forecasting Based on 5 model (ARIMA-AR-WEIBULL).

Table 23: RMSE53-hrs forecasting Based on 5 model (ARIMA-AR-WEIBULL).

No. of Samples	ARIMA (12.0.4)	AR (12.0.0)	ARIMA & WEIBULL	AR & WEIBULL
200	2.0852	1.7298	1.6563	1.5801
400	1.3347	1.3223	1.6924	1.6916
1000	1.2739	1.2869	1.5199	1.5260
1500	1.8951	2.0160	1.8780	1.9141
2900	2.4664	1.7955	1.7563	1.7797

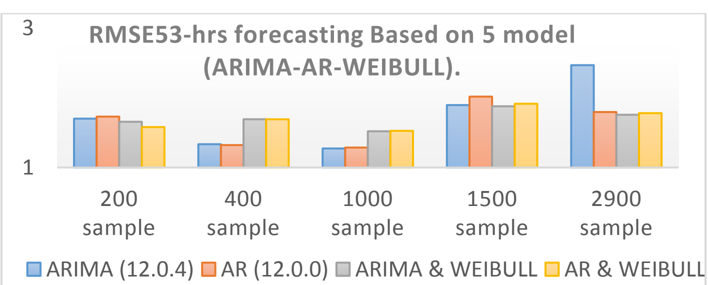

Figure 29 RMSE53-hrs forecasting Based on 5 model (ARIMA-AR-WEIBULL).

Table 24: MAPE 53hrs forecasting Based on 5 model (ARIMA-AR-WEIBULL).

No. of Samples	ARIMA (12.0.4)	AR (12.0.0)	ARIMA & WEIBULL	AR & WEIBULL
200 sample	0.2962	0.27	0.27	0.25
400 sample	0.20	0.17	0.28	0.28
1000 sample	0.20	0.25	0.26	0.26
1500 sample	0.26	0.28	0.32	0.33
2900 sample	0.39	0.25	0.29	0.30

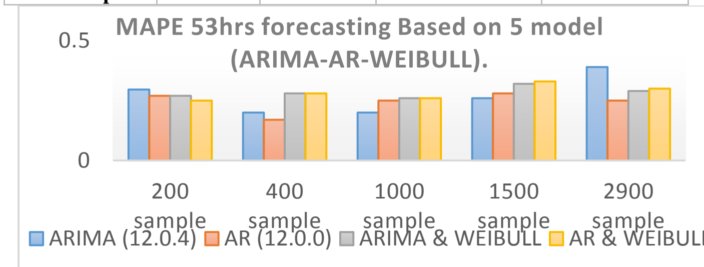

Figure 30 MAPE 53hrs forecasting Based on 5 model (ARIMA-AR-WEIBULL).

In this paper, two methods were presented to predict wind speed in the Hurghada region: the first using the ARIMA structure (p, 0, q) and AR (p, 0, 0) separately, the orders were chosen as a separation in the average wind speed, and the second using the ARIMA structure (p, 0, 0) and an AR hybrid model (p, 0, 0) with a Weibull distribution over five samples, 200, 400, 1000, 1500, and 2900. Wind speed was high during the autumn forecast season.

Comparison of AR, ARIMA, and the hybrid with Weibull distribution in 5 examples 200,400,1000,1500,2900 samples with orders fixed of AR and ARIMA and application of mean wind speed prediction, error, MAPE, and MSE to derive the best model for forecasting wind speed as the number of samples increases.

When employing 200 samples, the best results were obtained by using ARIMA (12,0,4) where mean wind forecast be 6.3595 and real wind speed 6.3639 and error equals 0.0044 and best between all methods, which requires less samples and produces better results.

When employing 400 samples, the best results were obtained by using ARIMA (12,0,4) where mean wind forecast be 6.0365 and real wind speed 6.3639 and error equals 0.3275 and best between all methods, which requires more samples and produces better results.

When employing 1000 samples, the best results were obtained by using ARIMA (12,0,4) where mean wind forecast be 6.3845 and real wind speed 6.3639 and error equal -0.0094 and best between all methods, which requires more samples and produces better results.

When employing 1500 samples, the best results were obtained by using ARIMA (12,0,4) where mean wind forecast be 5.5825 and real wind speed 6.3639 and error equal 0.7787 and best between all methods, which requires more samples and produces better results.

When employing 2900 samples, the best results were obtained by using ARIMA (12,0,4) where mean wind forecast be 5.6878 and real wind speed 6.3639 and error equal 0.6761 and best between all methods, which requires more samples and produces better results.

In comparison to proposed model in [29] which forecast very short to short-term wind speed based on a hybrid approach that combines maximal overlap discrete Wavelet transform with ARIMA and adjusted a dynamic moving window with Markov chains. This model showed improved performance to forecast wind speed with a self-adaptive state categorization for equal/unequal intervals. However, the proposed model in our manuscript combines either autoregressive or autoregressive integrated moving average with cumulative Weibull distribution to obtain better forecasting accuracy and maintain model efficacy and simplicity

5 CONCLUSION AND FUTURE WORK.

This paper proposed a wind Weibull distribution with AR and ARIMA for improve forecasting wind speed and increased of numbers of samples incoming and decreased of error. For each case, the average prediction of wind speed and error rate is calculated with the increase in the number of samples

When employing 200 samples, the best results were obtained by using ARIMA (12,0,4), which requires less samples and produces better results.

The best results were obtained while utilizing 400, 1000,1500 and 2900 samples with a hybrid model ARIMA (12,0,4) with Weibull distribution. As the input time lengthens, so does the difference between real data and predictions. MAPE values differ from MAE and RMSE values.

Because the studied wind speed data is unstable, the error from the first input increases for MAPE, MAE, and RMSE values when predicting wind speed. As the input time lengthens, so does the difference between real data and predictions. The prediction error will be increased with increased samples. MAPE values differ from MAE and RMSE values. It is safe to say that a longer time of windspeed forecast results in bigger errors. Furthermore, wind speed forecast grows progressively wrong over time.

Our next research will use Weibull distribution order 3 in conjunction with ARIMA and AR to overcome the substantial uncertainty in wind speed variation and eliminate data outliers. We also intend to expand the proposed models to include wind-speed forecasts.

6 REFERENCES

[1] Salim, O.M., Zohdy, M.A., Dorrah, H.T, *et al.*: 'Adaptive neuro-fuzzy short-term wind-speed forecasting for Egypt's East-Coast', *Int. J. Sustain. Energy 2014,**33*** 1), pp. 16–34

[2] Hu, Jianming & Heng, Jiani & Wen, Jiemei & Zhao, Weigang,DEC 2020. "Deterministic and probabilistic wind speed forecasting with de-noising-reconstruction strategy and quantile regression based algorithm," Renewable Energy, Elsevier, vol. 162(C), pages 1208-1226.

[3] Murcia Leon J.P.Uncertainty quantification in wind farm flow models(Ph.D. thesis)DTU Wind Energy, Denmark (2017).

[4] Foti D., Yang X., Sotiropoulos F.Uncertainty quantification of infinite aligned wind farm performance using non-intrusive polynomial chaos and a distributed roughness model Wind Energy, 20 (6) (2017), pp. 945-958.

[5] Craig A., Optis M., Fields M.J., Moriarty P. Uncertainty quantification in the analyses of operational wind power plant performance J Phys Conf Ser, 1037 (2018), Article 052021.

[6] Khosravi A., Nahavandi S.An optimized mean variance estimation method for uncertainty quantification of wind power forecasts Int J Electr Power Energy Syst, 61 (2014), pp. 446-454.

[7] Fernández-González S., Martín M.L., Merino A., Sánchez J.L., Valero F. Uncertainty quantification and predictability of wind speed over the Iberian Peninsula J Geophys Res: Atmos, 122 (7) (2017), pp. 3877-3890.

[8] Jin J., Che Y., Zheng J., Xiao F. Uncertainty quantification of a coupled model for wind prediction at a wind farm in Japan Energies, 12 (8) (2019), p. 1505.

[9] Quan H., Khosravi A., Yang D., Srinivasan D. A survey of computational intelligence techniques for wind power uncertainty quantification in smart grids IEEE Trans Neural Netw Learn Syst, 31 (11) (2020), pp. 4582-4599.

[10] Wang J., AlShelahi A., You M., Byon E., Saigal R. Integrative density forecast and uncertainty quantification of wind power generation IEEE Trans Sustain Energy, 12 (4) (2021), pp. 1864-1875.

[11] Richter P., Wolters J., Frank M. Uncertainty quantification of offshore wind farms using Monte Carlo and sparse grid Energy Sources B (2021), pp. 1-27.

[12] Wei W., Wu J., Yu Y., Niu T., Deng X. Uncertainty quantification analysis of wind power: A data-driven monitoring-forecasting framework IEEE Access, 9 (2021), pp. 84403-84416.

[13] *EEHC,* Egyptian electricity holding company annual reports *2020/2021*.

[14] M. Templ, Statistical disclosure control for microdata: Methods and applications in R, no. 1993. 2017.

[15] Büyükşahin, Ü. Ç., & Ertekin, Ş. (2019). Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition. Neurocomputing, pp. 361, 151-163.

[16] Luo, J.; Hong, T.; Fang, S.-C. Benchmarking robustness of load forecasting models under data integrity attacks. Int. J. Forecast. 2018, 34, 89.

[17] Zhang, Y.; Lin, F.; Wang, K. Robustness of Short-Term Wind Power Forecasting against False Data Injection Attacks. Energies 2020, 13, 3780.

[18] W. Cheng，Y. Liu, A. J. Bourgeois，Y. Wu，S．E．Haupt，"Short-term wind forecast of a data assimilation/weather forecasting system with wind turbine anemometer measurement assimilation"．Renew Energy，vol．107，pp．340- 351，Jul．2017．

[19] R．Li，Y．Jin，"A wind speed interval prediction system based on multi-objective optimization for machine learning method"．Appl Energy，vol．228，pp．2207-2220，2018.

[20] Ma, Z., Chen, H., Wang, J., Yang, X., Yan, R., Jia, J., & Xu, W. (2020). Application of hybrid model based on double decomposition, error correction and deep learning in short-term wind speed prediction. Energy Conversion and Management, 205, 112345.

[21] Granger C. W. J., Andersen A. On the invertibility of time series models. Stochastic Processes and Their Applications. 1978. Vol. 8. Issue 1. P. 87–92.

[22] Ojo J. F. Identification of optimal models in higher order of integrated autoregressive models and autoregressive integrated moving average models in the presence of subsets. Journal of Modern Mathematics and Statistics. 2008. Vol. 2(1). P. 7–11.

[23] Nash J., Sutcliffe J. V. River flow forecasting through conceptual models' part I – A discussion of principles. Journal of Hydrology. 1970. Vol. 10.

[24] Zheng, D., Eseye, A. T., Zhang, J., & Li, H. (2017). Short-term wind power forecasting using a double-stage hierarchical ANFIS approach for energy management in microgrids. Protection and Control of Modern Power Systems, 2(13), pp1–10.

[25] James F. Manwell, Jon G. McGowan, Anthony L. Rogers. Wind Energy Explained: Theory, Design and Application, 2nd Edition. p59,2010

[26] Mohamed I. Awaad , Omar M. Salim , Ossama E. Gouda, "Improved Kalman Filtered Neuro-Fuzzy Wind Speed Predictor For Real Data Set Collected At Egyptian North-Western Coast," *Ijmer*, vol. 4, no. 9, pp. 9–18, 2014.

[27] Omar M. Salim, Hassen Taher Dorrah, Mahmoud Adel Hassan,“ Wind speed estimation based on a novel multivariate Weibull distribution,”pp.9 2019.

[28] Peng, C., Sun, H., Yang, M., *et al.* “ A survey on security communication and control for smart grids under malicious cyber-attacks, IEEE. *trans Syst.*, 2019, 49, (8), pp. 1554–1569.

[29] Muhammad Uzair Yousuf, Ibrahim Al-Bahadly, Ebubekir Avci, “Short-Term Wind Speed Forecasting Based on Hybrid MODWT-ARIMA-Markov Model”, IEEE Access, VOLUME 9, 2021.