Time-Series analysis for wind speed forecasting

Garima Jain

Abstract
In this paper, an enormous amount of study has been made on various weather forecasting models and many specialists developed different models for optimal results. Different models which is taken for implementation and were used for predicting the forecasted data. A technique used for forecasting the given data is defined as a time series data. Box-Jenkins method is a statistical methodology used for prediction of data in time series. During this paper, an ARIMA model is implemented for predicting the data of Wind Speed. Results are compared with respective model i.e., ETS Model. With this paper we like to through some light on ARIMA (Auto-Regressive Integrated Moving Average) and ETS (Exponential Smoothing) models for forecasting the weather conditions in India. ARIMA model is chosen; because of it is acceptable in terms of easiness and extensive of the model. Eight years weather data (from year 2007 to 2014), i.e., wind Speed for time-intervals to forecast (i.e., 1 hours) are used in this research.

Keywords
ARIMA (Autoregressive Integrated Moving Average), ETS (Exponential Smoothing), AIC (Akaike’s Information Criteria), BIC (Bayesian Information Criteria), RMSE (Root Mean Square Error), Mean Absolute percentage Error (MAPE), Mean Absolute Error (MAE) and Box-Jenkins.

1. Introduction
Wind speed has as of late gotten an expanding consideration around the earth due to its inexhaustible nature and also ecological agreeableness. The distinctive advantages of wind vitality are joined by a few difficulties: high capriciousness, restricted likelihood, constrained dispatchability and non-storability. Climate determining is an arrangement worried about future climate. Weather forecasting concerned with various strategies which make use of relatively simple observation of the sky to advance computerized mathematical Models [5]. Application like ARIMA has one of the important uses in time series prediction another or equal technique to the traditional statistical strategiesthat are utilized in the prediction of time-series like MA, AR and ETS Box-Jenkins models [1]. These ancientstrategies are generally known as Time Series strategies. However, meteoricknowledge is unsure (uncertain) in nature and data on weather is mistily outlined [3]. Predicting weather has always been a difficult field of research analysis with a very slow progress rate over years. Weather data consider the noises and outliers therefore; the investigation may not be accurate. Random error is defined as noises (or bug) that comes from the device network, error writing and so on. Weather forecasting is a crucial issue within the field of meteorology all over the world. Weather is taken into accountbecause the most difficult drawback witnessed by the globewithin the last decade. The use of wind energy has been developed considerably throughout the world, in order to get the ideal for the future with electricity without pollution. Wind is one of the weather variables which are very difficult to predict. In wind power forecasting, there are two approaches commonly used [6].

However, reliable and correct wind speed forecasts are a major challenge due to its stochastic nature with high amount of changes, highly nonlinear behavior described with no typical patterns [9], and dependency on elevation, relative humidity, atmospheric pressure, and temperature, which results
in large uncertainties of wind speeds. Extensive efforts have been dedicated to develop efficient wind speed forecasting models. Recent advances toward a better understanding of the hydro dynamical aspects of the difficulties, coupled with the development of high-speed computing tools, promise to create the current area one amongst the foremost exciting in all meteorological history [1]. Instead of requiring separate power forecasts, they can get them by converting the common wind speed forecasts based on their own power curves. Wind speed forecasting can be more precise than wind power forecasting due to the spatial correlation of wind. Instead of demanding separate power forecasts, they can usually get them by transforming the common wind speed forecasts.

A. Objective

One of the recent objective of the weather study is to implement a method that generated an improving idea about the forecasting approximation of different parameters. There are number of struggles which should be made to develop and improve the existing time series weather forecasting models by using different techniques. The role played by statistical methodology for forecasting the weather parameters is defined to be most important for their detailed estimates. The original measurement time-series comprises 54,956 data corresponding to wind speed in nautical-miles per hour (kn), acquired each 1 hour. Kn is a symbol which is illustrated as a ISO Standard. The knot is defined as a unit of speed which is equal to one nautical mile as 1.852 km (or 6076.12 feet) per hour, approximately 1.151 mph internationally which means that 0.0639% should be added to “UK knots” to get “International Knots”.

Weather Data of India from the dates May 03, 2007 to March 06, 2014 was used to provide useful insight about the performance of the algorithms. The recorded data of the years 2007-2014 were used to make predictions. The accuracy of various models is measured and then compared by Moving Absolute Error, Moving Absolute Scaled Error, Moving Absolute Percentage Error and Root Mean Square Error. We also include the criteria Akaike’s Information Criteria and Bayesian Information Criteria. The model which formed the best prediction result will use for comparison and prediction. In this paper, ARIMA model is used in R software for predicting the weather. R is widely used Language not only by scientists but also in many time series applications. It is an additional concept including mathematical and statistical expertise which helps us to groom our knowledge in analytical field that when deployed into existing processes makes them adaptive to improve conclusions [4]. Thus, together with these advantages offered by this software, we can predict the results before they occur. ETS model is also used from R Software with R Studio Tool. These tools apply R techniques to data modeling. This allows our stationary analytical systems to learn from the data they are modeling [5]. These methods are defined briefly by Hyndman and Khandakar (2008) collection of functions for evaluating time series data, as well as many

more interesting time series datasets for different functionality. The package Forecast (Hyndman, 2010) provides additional care for forecasting using ARIMA and an extensive class of exponential smoothing models.

The data are collected from the Indian Meteorological Department which is the head office. In the data set, there we have Wind Speed parameter measurements, and this data are only on Indian weather condition. For analysis and forecast, we applied ETS and ARIMA on this data; finally they are evaluated and compared.

B. Outline

The rest of the paper is organized as follows: Section 2 describes the conceptual study on ARIMA model. Section 3 is used for approaches named as ETS model and data demonstrations, the models that are examined are also described in Section 3. Section 4 presents and analyzes the results. Finally we summarize and give future direction of research in Section 5. Section 6 describes the related references.

2. ARIMA

ARIMA stands for Autoregressive Integrated Moving Average. ARIMA models aim to define the autocorrelation in the parameter value and can be applied to evaluate stationary and non-stationary time series. ARIMA model is the technique which is implemented and designed by the respectable Box and Jenkins (1976). It is an optimal combination of three designed calculated models. It uses auto-regressive, integrated, moving-average (ARIMA) models for time series data. The model is known by ARIMA(p), where $p \geq 0$. After differencing ARIMA Model finitely many times we correspond to an ARIMA Model. The components $p$ and $q$ are defined as the order for AR and MA parts, whereas $d$ is the degree of differencing. Differencing is generally used to eliminate the trend (may be linear or exponential) in a time series. The differencing order $d$ relates to how many times the process $Y_t$ requests to be differenced to become stationary. To see that if the Time Series is Stationary or not we take help of ACF (Auto Correlation Function) and PACF (Partial Auto Correlation Function). Time series is a set of interpretations ordered (sequentially) according to the time they were observed.

An ARIMA $(p, d, and q)$ model can explain for temporal dependence in several ways.

- First, the time series is taken to reduce it stationary, by taking d differences. If $d = 0$, the observations are modelled directly, and in case, if $d = 1$, the differences between successive observations are modelled.

- Second, the time that depend on the stationary process is modelled, by including p Auto-regressive (AR). The equation for $p$ is that:

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + Z_t,$$

(2.1)
where, \( c \) is an unknown constant term, and \( \varphi_i, i = 1, 2, \ldots, p \), are the parameters of the AR model.

- Third, \( q \) is related as the moving-average terms. It takes the observation of previous errors. The equation for \( q \) is that:

\[
Y_t = c + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \quad (2.2)
\]

where, \( c \) is an unknown constant term, and \( \theta_1, \ldots, \theta_q \), are parameters of the MA model.

- Finally, combining these three mathematical models we get ARIMA model. So the integrated form of the ARIMA model is that:

\[
Y_{t-1} = a_0 + (1 + \varphi_1)Y_t - (\varphi_1 - \varphi_2)Y_{t+1-p} + \varphi_p Y_{t-p} + \theta_1 Z_{t} + \cdots + \theta_q Z_{t+1-q} + \varepsilon_t, \quad (2.3)
\]

where, \( Y_t \), a stationary is a stochastic process, \( a_0 \) is the constant, \( \varepsilon_t \) is the error or white noise disturbance term, \( \varphi_p \) means auto-regression coefficient and \( \theta_q \) is the moving average coefficient. If we consider a cyclical time series, following steps can be repeated according to the period of the cycle, it may be viewed whether quarterly or monthly or other time interval. ARIMA models result in extremely flexible for continuous data analysis [4].

3. Experimental Procedure And Result

A. ARIMA model

It includes various steps for forecasting by Automatic forecasting Algorithm.

First we depict to identify the given series to make it stationary for prediction of Wind Speed measurements, after that we look for the value of parameters, then we validate the model by gradually fitting it and at last after completing all steps the series is use for forecasting.

The very first point is to begin the ARIMA model which best fits the time-series behavior. With this objective the autocorrelation coefficient and the partial autocorrelation coefficient are estimated and depicted below.

In this section we explain the procedure to achieve the possible models which recover and explain the time-series behavior. The original dimension time-series comprises 54,956 data corresponding to wind speed in nautical-miles per hour (kn), acquired each 1 hours.

B. The plots of sample ACF and PACF are helpful analytical tool when selecting the order of an ARIMA Model specially when data must originated by an ARMA(p,0) model or it may from an ARMA (0,q) model. For stationary processes, we can look for auto correlation between any two observations which depend on the Time Lag \( h \) between them.

Partial Autocorrelation between \( Y_t \) and \( Y_{t-h} \) after removing any Linear dependence. The Autocorrelation for lag \( h \) is defined as:

\[
\rho_y(h) = \frac{\gamma_y(h)}{\gamma_y(0)}, \quad (3.1)
\]

The PACF of an ARMA Model is the function \( \alpha(\cdot) \) which
can be defined as:
\[ \alpha(0) = 1, \]

**Figure 3.** Partial Autocorrelation for Model Identification

In the estimation process, errors of the model for which the computation should be calculated and assessed, along with the resulted values of the evaluated parameters. The values for the autoregressive coefficients are 0.4807 and 0.2004 and the moving average (MA) coefficient is \(-0.9876\) for \((2, 1, 1)\). The series: Wind Speed \((2, 1, 1)\) model is presented as a simple models with less computation-time, in the forecasting procedure. In that case, a multiple seasonal model such as TBATS is required.

**Figure 4.** Forecasted Multiple Seasonal Model

On taking in concern about seasonal decomposition, let first look on the fact that should it necessary to do that as trend component comprises under the time series, both the components that is irregular as well as seasonal. This is why decomposing the series is must require. Now for this we concluded that there is no irregularity in the shape of plot, so it is optimal not to perform seasonal decomposition for this series. The ARIMA model has some parameters which are commonly defined as defined \(c, \varphi_1, \varphi_2, \ldots, \varphi_p, \theta_1, \theta_2, \ldots, \theta_q\). Using the function `auto.arima` in R, this is often done by the strategy of maximum likelihood estimation assuming that time series is Gaussian. Perform auto arima on that time series data using `auto.arima()` function of “forecast” package. This technique helps us to finds the values of the parameters which maximize the probability of obtaining the information which we have observed previously. For models where \(p > 0\) and \(q > 0\), the sample ACF and PACF are difficult to recognize and value in order selection than in the special cases where \(p = 0\) or \(q = 0\). A logical approach, however, is still available through minimization of the corrected Akaike’s information criteria (AICc) statistic \((3.3)\). AIC is known as theAkaike’s information criteria (AIC) is outlined as:

\[ AIC = -2\ln(L) + 2(p + q + k + 1), \]

The corrected AICc is outlined as:

\[ AICc = AIC + \frac{2(p + q + k + 2)(p + q + k + 2)}{n - p - q - k - 2}, \]

BIC (Bayesian Information Criteria) has mathematical formula. It should provide complexity more than AIC does:

\[ BIC = AIC + \log(T)(p + q + P - 1), \]

Hence, the concluding decision regarding the orders \(p\) and \(q\) that minimize AICc must be based on maximum likelihood estimation. Here \(L\) is the likelihood of the data, and \(k = 1\) if \(c \neq 0\) and \(k = 0\) if \(c = 0\). \(L\), the likelihood function of the data is used to find the maximum likelihood estimates of the parameters of an ARMA process.

**Table 1.** Coefficients of ARIMA Model

<table>
<thead>
<tr>
<th></th>
<th>ar1</th>
<th>ar2</th>
<th>ma1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated</td>
<td>0.4807</td>
<td>0.2004</td>
<td>-0.9876</td>
</tr>
<tr>
<td>Std Error</td>
<td>0.0044</td>
<td>0.0044</td>
<td>0.0013</td>
</tr>
</tbody>
</table>

\(\Sigma^2\) estimated as 3935 and log likelihood=305418

\[ AIC = 610844.1, AICc = 610844.1, BIC = 610879.8. \] Now we do seasonal decomposition using the TBATS model. TBATS model is a time series model for series demonstrating multiple complex seasonality. The TBATS model was introduced by De Livera, Hyndman and Snyder (2011, JASA). "TBATS" is an abbreviation denoting its salient types:

- \(T\) for trigonometric regressors to model multiple-seasonality
- \(B\) for Box-Cox transformations
- \(A\) for ARMA errors
- \(T\) for trend
- \(S\) for seasonality

The generalization of BATS model is TABTS method, which is similar except for lacking the trigonometric regressors. The BATS model can be fitted using the `tbats()` command
in the forecast package for R. The value of the Parameters is evaluated as by using TABTS method. The value of Sigma is 0.608145 and AIC is 962921.8.

### Table 2. Forecasted Parameters using TABTS method

<table>
<thead>
<tr>
<th>Lambda</th>
<th>Alpha</th>
<th>AR coefficient</th>
<th>MA coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000497</td>
<td>0.02107786</td>
<td>1.045348</td>
<td>-0.644123</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.103271</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.052669</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 5.** Forecasts from TABTS Function

Now we use decomposition procedures that are used in time series to describe the trend and seasonal factors in a time series. More wide decompositions may also include long-run cycles, day of week effects and so on. Here, we’ll only consider trend and seasonal decompositions. The main purposes for decomposition are to estimate seasonal effects that can be used to generate and current seasonally adjusted values. A seasonally adjusted value eliminates the seasonal result from a value so that trends can be seen more openly and helpful in predicting the series.

**Figure 6.** Result of Decomposition of Additive Series

4. **ETS Model**

**A. ETS(M,N,N): Simple exponential smoothing with multiplicative error**

The models with multiplicative errors can be indicated by giving the one-step random errors defined as relative errors:

\[ Z_t = y_t - \hat{y}_{t/1}, \]  

Hence the state space model for multiplicative form can be written as:

\[ y_t = l_t - 1 + Z_t l_t = l_t - 1 + aZ_t, \]  

We require initial standards for those modules as well. The unknown parameters and the initial values for any exponential smoothing strategy can be surveyed by limiting the entirety of squared mistakes as in relapse investigation, yet dissimilar to the relapse case, we includes a non-straight an issue for mini-mization and henceforth we require to advance instrument to play out this. Straightforward exponential smoothing has just a single smoothing parameter and it requires just an early incentive for the level, other exponential smoothing techniques that include a pattern or regular parts.

An option technique that is utilized to assess the parameters by limiting the entirety of squared blunders this strategy is utilized as the technique for greatest probability evaluate under the supposition that time arrangement is Gaussian. The probability capacity of the developments state-space display is computed and greatest probability appraise is gained. The exponential smoothing capacity in the conjecture bundle in R and it is accordingly utilized to select for the suitable model automatically accordingly and according to maximum likelihood method and to find the point forecasts as many steps.

— 59/61
We can evaluate the Information criteria $AIC$, $AICc$ and $BIC$ can be used for selecting the best among 30 ETS models. The evaluated AIC is 1003202, AICc is 1003202 and BIC is 1003229.

### B. Automatic Forecasting Algorithm

Automatic forecasting technique is a technique forecasting using exponential smoothing function in forecast package in $R$.

The dynamically and broadly applicable for the automatic forecasting algorithm is acquired for ETS models, the steps followed are:

1. For each given series, we apply all techniques that are suitable, optimizing the parameters (for both the smoothing parameters maintained and the initial state variable) of the model in each case.

2. The best of the models can be selected according to the information criteria $AICc$.

3. Point forecast with the property of best model can be defined with optimized parameters.

4. To obtain forecasting intervals for the best model, systematic results suggested by Hyndman [9] or pretending future aspects with the illustrat paths and finding $\alpha/2$ and $1 - \alpha/2$ percentiles of the accurate concluded simulated data at each forecasting view. In case if we used simulation, the sample paths may be created using the normal distribution for errors. The resulted Smoothing parameters are given as, the value alpha is given 0.0777, Initial states are 176.0555 and sigma is calculated as 0.729.

### 5. Discussion and the Result Analysis

As soon as we have selected the possible models that correspond with our information, much the ARIMA and the ETS, we realize a comparison of all of them to choose which or which are better. In our study, all data, recorded in the Indian Meteorological Department, is used for the analysis in this comparison study. We collected the eight years data from 2007 to 2014, and then did the data pre-processing to clean data such as missing data and inconsistent data. The performance comparisons of ETS and ARIMA(2, 1, 1) model for MAE, MAPE, RMSE and ACF1 are shown below:

<table>
<thead>
<tr>
<th>Performance criteria</th>
<th>ARIMA</th>
<th>ETS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ME</td>
<td>15.60954</td>
<td>-0.01775237</td>
</tr>
<tr>
<td>2. RMSE</td>
<td>67.807</td>
<td>73.1055</td>
</tr>
<tr>
<td>3. MAE</td>
<td>35.69797</td>
<td>46.48531</td>
</tr>
<tr>
<td>4. MPE</td>
<td>-34.79379</td>
<td>-78.01136</td>
</tr>
<tr>
<td>5. MAPE</td>
<td>57.44789</td>
<td>97.21146</td>
</tr>
<tr>
<td>6. MASE</td>
<td>0.4457349</td>
<td>0.5804287</td>
</tr>
<tr>
<td>7. ACF1</td>
<td>0.1896902</td>
<td>0.490238</td>
</tr>
</tbody>
</table>

### 6. Conclusion

In this paper we have identified the comparative study of ARIMA and ETS models which match for the measurement of wind speed time-series. The model(2, 1, 1) exhibits the best performance. The confirmation has been done using three common quality indexes, based in correlation procedures. In this research, the performances of ARIMA and ETS are compared. ARIMA can more efficiently capture the dynamic behavior of the weather property, say, Wind Speed compared to ETS. However we could not explore all the features of ETS due to the time limitation. Therefore, the decision about the performance of ETS model is not complete and final. We need to investigate more in this direction. Our preliminary findings show that ARIMA is better than ETS.

### A. Future Work

The main aim for improving the prediction performance for the time series weather prediction model is designed and developed in this work. The implemented technique is effective and precise for weather prediction but some limitations of the model is also observed thus in near future need to be review before use of the proposed technique. The key extension of the works is the need to be increasing the training samples for collecting the training data as the performance of prediction is rises with the amount of information thus huge amount of data can solve the issues of prediction. In the future, we plan to develop our own fuzzy logic techniques for weather forecasting. Finally, we need to compare those techniques along with ARIMA model to find the best one.

### Acknowledgment

The authors would like to express many thanks to the reviewers for his/her valuable suggestions and special thanks to my teachers and parents.
References