

A Study of Time Series Models

ARIMA and ETS

Garima Jain

Galgotias College of Engineering and Technology/Computer Science , Gr.Noida, 201310, India
Email: jaingarima2011@gmail.com

Dr. Bhawna Mallick

Galgotias College of Engineering and Technology /HOD(Computer Science), Gr.Noida,201310, India
Email: bhawna.mallick@gmail.com

Abstract

The aim of the study is to introduce some approach which might help in improving daily temperature of data. Weather is a natural a phenomenon for which forecasting is a great challenge today. Weather parameters such as Rainfall, Relative Humidity , Wind Speed , Air Temperature are highly non-linear and complex phenomena, which include mathematical simulation and modeling for its correct forecasting. Weather Forecasting is use to simplify the purpose of knowledge and tools that are used for the state of atmosphere at a given place. The prediction is becoming more complicated due to changing weather condition. There are different software and types are available for Time Series forecasting. Our aim is to analyze the parameters and do the comparison of some strategies in predicting these temperatures. Here we tend to analyze the data of given parameters and notice the prediction for few period using the strategy of Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing (ETS) .The data from meteorological centers are taken for comparison of methods using packages such as ggplot2, forecast, time Date in R and automatic prediction strategies are available within the package applied for modeling with ARIMA and ETS methods. On basis of accuracy we tend to attempt the simplest Methodology. Our model will compare on basis of MAE, MASE, MAPE AND RMSE. The identification of model will chromatic inspection of both the ACF and PACF to hypothesize many possible models will estimated by selection criteria AIC, AICc and BIC.

Keywords: ARIMA (Autoregressive Integrated Moving Average), ETS (Exponential Smoothing), AIC (Akaike's Information Criteria), and BIC (Bayesian Information Criteria).

Introduction

Weather Prediction System is having most complex Systems Equation that can only solve by Computer. Weather Forecasting is an inspiration about future weather. There is various techniques concern with weather forecasting from comparatively easy observation of the sky to extremely advance computerized mathematical Models [5]. The main goal of the study is to research a technique that improves the forecasting estimate of various parameters. We would like to match the forecasting of automatic forecasting methods in predicting the weather parameters such as Rainfall, Relative Humidity, Air pressure and Wind Speed etc. However predicting the weather considers many forms, which mainly depend upon the required applications. The forecasting challenge of weather is equal to the very first major difficulty depend on the non-linear equations which leads to numerical means and mathematical physics. However meteoric knowledge is tentative (unsure) in nature and knowledge on weather is mostly outlined. Weather knowledge will have the noises and outliers, and this is why analyze may not be correct. Noises are the random error therefore we need prepossessing of a Weather knowledge to improve the standard of knowledge for precise weather prediction, or to improve the value for exact data weather estimation. This paper gives an idea to compare the two automatic forecasting methods ETS (Exponential Smoothing) and ARIMA (Autoregressive Integrate Moving Average) using the data of various weather parameters such as Rainfall, Relative Humidity, Air Temperature , Pressure , Wind speed etc. For convenience the paper provided with some basic definition and background that designates the strategies followed by the analysis of weather data by estimation and verification of varies forecasting methods. The accuracy of various strategies is compared by MAE (Moving Absolute Error), MASE (Moving Absolute Scaled Error), MAPE (Moving Absolute Percentage Error) and RMSE (Root Mean Square Error). We also include the criteria AIC (Akaike's Information Criteria) and BIC (Bayesian Information Criteria).The methods which gives the best forecast will use for comparison and prediction. As of India Meteorological Department information can even be composed that is over all India climate stations head place of work. In information set we are having the parameters that are: Air Temperature, Rainfall, Relative Humidity, and Wind Speed among which we consider one. In this paper ARIMA (Autoregressive

Integrated Moving Average) and ETS (Exponential Smoothing) is used for analysis and predicting of weather information mentioned above. A Comparative study brings between the ETS and ARIMA Model through SSE, MAE, RMSE, MASE and MAPE also include the criteria AIC and BIC result from the given parameters. Automatic forecasting by ARIMA and ETS is performed by using the forecast, time Date, ggplot2, and Zoo packages in R for Time Series analysis, and also various attributes are predicted by considering Correlations. Efforts are being intense to use statistic Autoregressive Moving Average (ARMA) Model to forecast or predict hydrological information. The use of ARMA Model is even as a result of its theoretical base in hydrological Studies. The paper is organized as follows. Section 2 gives you study on topic as Literature review on related work. Section 3 presents some theory and concepts of time series ARIMA modeling and forecasting, used to analyses the algorithm on which our future implementation will based on. Section 3 provides the detailed analysis of ETS Model and the comparison of proposed models. Section 4 gives some suggestions for further research and areas of interests.

2. Literature Review

Around the world weather predicting is one of the most challenging difficulties, due to its practical value in popular scope for scientific study and meteorology. The different methodologies i.e., Statistical Decomposition Models, Exponential Smoothing Model (ETS) and Autoregressive Integrated Moving Average (ARIMA) Model variable Time series and following informative variables etc., is used for forecasting purposes. Several Trainings have taken place among the analysis of pattern and circulation in various parameters in different regions of the World. Regression Analysis could be applied for statistical technique and it should be widely used in several sciences and many other relevant Areas. Agrawal *et al.* (1980) explained the phenomena for time series regression models for forecasting the yield of rice in Raipur district on weekly data using weather parameters [1]. Box and Jenkins (1994), in early 1970's, pioneered in developing methodologies for statistic representing within the univariate case often referred to as Univariate Box-Jenkins (UBJ) ARIMA modeling in this approach of the author [2]. Akaike, H. (1976) an information criterion (AIC) which is used as a parameter estimation to check the performance of Model which gives history for the development of statistical theory testing in time series analysis was studied concisely and it was pointed out that the theory testing procedure is not effectively defined as the technique for statistical model identification. The most likelihood estimation process was reviewed and a new estimate least information theoretical criterion (AIC) estimate was introduced [4]. Hyndman, R.J., & Kandahar, Y (2007). *Automatic time series for forecasting: the forecast package for R*. Monash University, Department of Econometrics and Business Statistics [5]. Weather forecasts provide critical data about future weather. There are various techniques involved in weather forecasting, from relatively simple observation of the sky to highly complex computerized mathematical models (M. Tektas, 2010) [6]. *An Integrated Approach for Weather Forecasting based on Data Mining and Forecasting Analysis by G.Vamsi Krishna* where in this paper the weather data was considered with attributes, such as wind pressure, humidity, Temperature, Forecast and Type, of Visakhapatnam city for a period of 97days. The forecasting experiment was carried out to test, the weather condition for the following 15 days by enabling the ARIMA model prediction algorithm model to predict the forecasts [7]. *An Introductory Study on Time Series Modeling and Forecasting by Ratnadip Adhikari and R. K. Agrawal* has given an idea about Time series forecasting is a fast growing area of research and which provides many scope for predicting future works. One of them is the *Combining Approach*, i.e. to combine a number of methods to improve forecast accuracy. Combining with other analysis in time series prediction, we have thought to estimate an efficient combining model, in future as possible. With the aim of further studies in time series modeling and forecasting[8]. *Prediction Of Rainfall Using Data Minning Technique Over by Pinky Saikia Dutta* which describe the model by considering temperature, wind speed, Mean sea level as Predictors. They found 63% accuracy in variation of rainfall for our proposed model. The model can predict monthly rainfall. Some predictor like wind direction is not included due to constraints on data collection which could give more accurate result. The work can be extended for multiple stations in future. The resulted rainfall amounts was intended to help farmers in making decision about their crop[9]. Mahmudur Rahman, A.H.M. Saiful Islam, Sahah Yaser Maqnoon Nadvi, Rashedur M Rahman (2013) consider Arima and Anfis Model and explained how ARIMA Model can more efficiently capture the dynamic behavior of the weather property say, Minimum Temperature, Maximum Temperature, Humidity and Air pressure which must be compared by different performance metrics, such as, with Root Mean Square Error (RMSE), R-Square Error and The Sum of the Square Error(SSE) and author can prove that ARIMA would give the more efficient result than other modeling techniques like ANFIS[10]. *Prediction of rainfall using an autoregressive integrated moving average model: Case of Kinshasa city* (Democratic Republic of the Congo), from the period of 1970 to 2009 by Dedetemo Kimilita Patrick¹, Phuku Phuati Edmond², Tshitenge Mbwebwe Jean-Marie², Efoto Eale Louis², Koto-te-Nyiwa Ngbolua³, There aim is to present study of the test a model of simulation on the monthly series of precipitations data from the Binza meteorological station of Kinshasa/Democratic Republic of the Congo. After their stationnarization of the Time series, we applied an Auto-Regressive Integrated Moving Average (ARIMA) model into the starting series. There model also makes it possible to predict implication of rainfall on the lifestyle of the Kinshasa inhabitants [11].

3. ARIMA (Auto Regressive Integrated Moving Average)

ARIMA Stand for Auto Regressive Integrated Moving Average. ARIMA model was popularized by box and Jenkins (1976). It is combination of three statistical models. It uses Autoregressive, Integrated and Moving Average (ARIMA) model for statistical information. The ARIMA Model analyze and Forecasts uniformly spaced univariate statistic information, transmission of function data, and intercession information that is done by using Auto Regressive Integrated Moving Average (ARIMA) and Auto Regressive Moving Average (ARMA). An ARIMA Model forecasts a value in a response Time Series which is a linear combination of its own related past values, past Errors and current and past values of alternative Time Series. ARIMA Model aims to explain the Auto Correlation within the information and may applied to stationary and non-stationary statistic. The Model written as ARIMA (p, d, q) where p, d, q ≥ 1. An ARIMA Model corresponds to ARMA after finitely many times differences the data. The elements p and q are the order for Auto Regressive and Moving average Components, because the degree of differencing written as d. Differencing is usually accustomed eliminated the Trend which may Linear and Exponential in a Time Series. The differencing order d relates that how many times the method yt must be differenced to become stationary. The prediction method was applied by summing last period's value with some constant, this indirectly help to estimate the prediction changes on a mean at specific interval of time. Various Transformation Techniques may be used for variance stabilization, e.g., Box-cox Transformation, Log Transformations. We can also make use of Auto Correlation Function (ACF) to visualize if the statistic is stationary or not. Auto Regressive indicated the Regression of a variable itself. In ARE Models the forecasting variables is a linear combination of its own earlier observations. An AR (p) Model is often delineated as:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + Z_t, \quad (1)$$

Where $Z_t \sim (0, 2)$,

c associate unknown constant term, and $\phi_i, i=1, \dots, p$, are the parameters of the ARE Model.

A MA (q) Model is often described as:

$$Y_t = c + Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \quad (2)$$

3.1 Estimation and Order Selection:

The ACF and PACF can used for the selection of p, d; q. Auto Correlation is the linear dependence of variable with itself at two points in time. For stationary processes, we can define auto correlation between any two observations which can depend on the Time Lag h between them.

The Autocorrelation for lag h is given by:

$$\gamma(h) = \gamma_Y(h) = \text{Corr}(y_t, y_{t-h}) \quad (3)$$

From this equation we define as a $\gamma(0)$ is a lag 0 covariance i.e., the unconditional variance of the process. Confounding can defined as the distortion in the estimated measure of association where two variables can result from a mutual linear dependence on other variables. Partial Autocorrelation between y_t and y_{t-h} after removing any Linear dependence. The partial lag h autocorrelation is denoted as $\phi_{h,h}$. The plots of sample ACF and PACF are helpful diagnostic when selecting the order of an ARIMA Model specifically when data must originated by an ARMA(p,0) model or from an ARMA(0,q) model. This shows that how we can find the model by defining the criteria about plots.

Table 1: Plots of ACF and PACF for ARMA models.

	MA(q)	AR(q)	ARMA(p, q)
ACF	Tail off	Cut off, q lag	Tail off
PACF	Cut off, p lag	Tail off	Tail off

If the ACF and PACF having the positive (large) values that decreased much slowly with time, this means that d is larger than 0, i.e., it should be done by differencing. We use many other ways to define unit tests to see d. Variety of unit root Tests to see where differencing was needed e.g., increased dickey- fuller (ADF) and kwiatkowski-phillips-schmidt-shin(KPSS). Here, we are using auto.arima function in R which automatically selects the ARIMA Model. This function uses the KPSS test with the null hypothesis to measures that if the data is stationary. (The small p-values than 0.5 recommended the differencing was required). After defining the order we have to determine the parameters of

ARIMA Model, i.e., $c, 1, \varphi_2, \dots, \varphi_p, \theta_1, \theta_2, \dots, \theta_q$. `auto.arima` is the function in R, this method will done by maximum likelihood estimation assuming that Time Series is Gaussian.

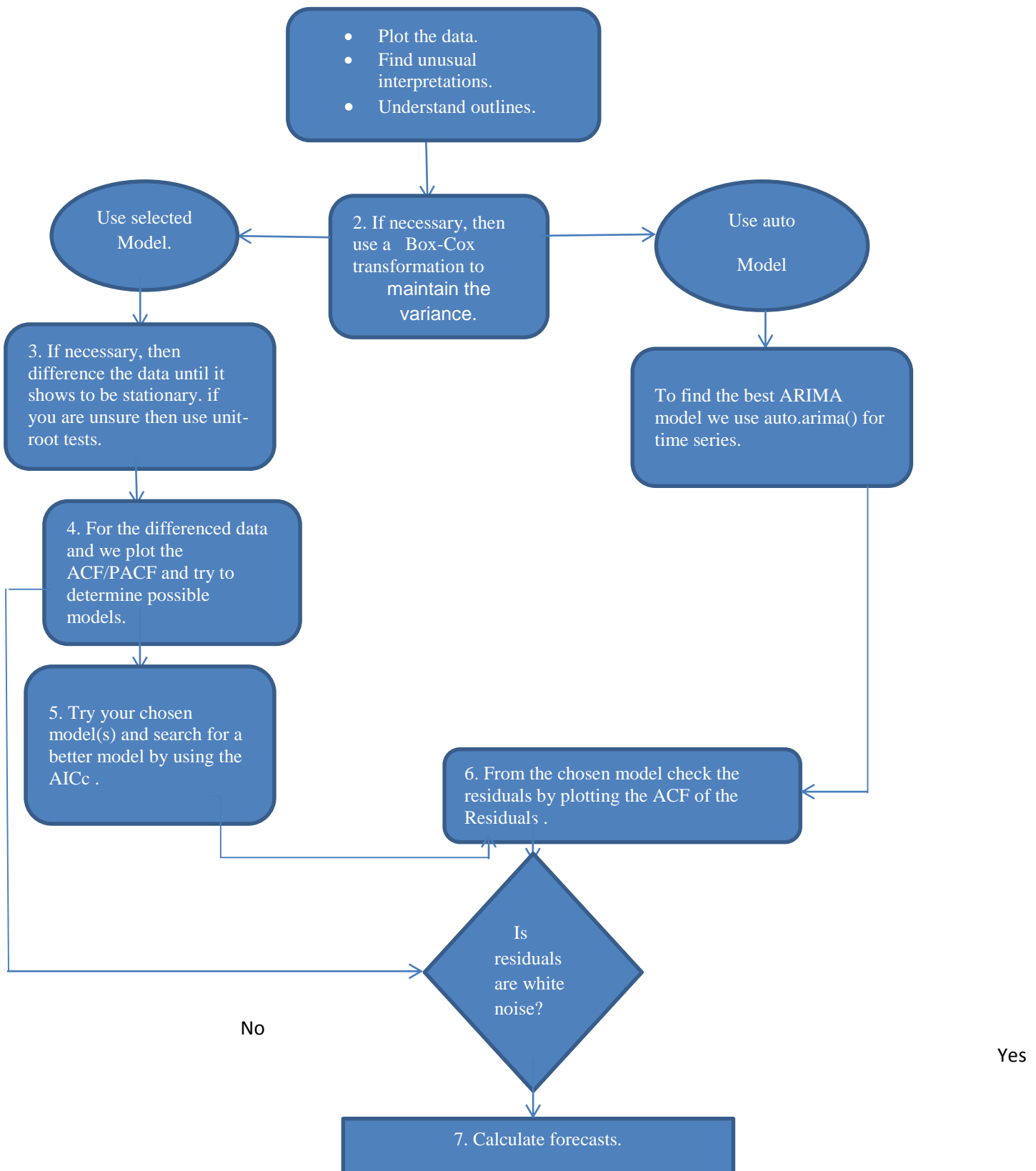


Fig 1: A Flow Chart Of ARIMA Model.

This technique would result in value of parameters in which the value of data was maximized. For model $p > 0, q > 0$, the sample ACF and PACF would difficult to analyze the order selection. An approach will made through minimization of the corrected Akaike's Information Criteria (AICc). If AIC computation was based on Algorithmic Mathematical formula:

$$AIC = -2 \log L + 2m, \text{ where} \quad (4)$$

$$AICc = AIC + (2(p+q+K+1)(p+q+K+2))/(n-p-K) \quad (5)$$

Where $m = p+q+K+1$ is the number of parameters in the model and L is the likelihood function. The best model is one with lowest AIC value. BIC (Bayesian Information Criteria) has mathematical formula. It should penalize complexity more than AIC does:

$$BIC = AIC + \log(T) (p+q+P-1) \quad (6)$$

Hence the final decision about the orders p and q that is probability based on maximum likelihood estimation. It is important to analyze the residuals of the chosen model for white noise. This can done by looking at residuals of ACF and formally using it with the value hypothesis that residuals are white noise or not correlated.

4. Exponential Smoothing

Exponential Smoothing invented by Robert G. Brown's work as an OR it is analyst for the US Navy (**Gass and Harris, 2000**). During the Early 1950s, Brown prolonged simple Exponential Smoothing to discrete data and the method should develop for trend and seasonality. Time Series (Brown, 1963), developed the general exponential smoothing methodology. When Time Series forecasted with Exponential Smoothing (ETS) Method, the first order auto-correlation was founded by Taylor in the residuals. Each Exponential Smoothing (ETS) method is equivalently related to one or more stochastic models. It also follows the property of Robustness and treated as a model of judgmental Extrapolation. The idea behind the exponential Soothing is that the forecasts by this approach are the weighted average of past observations and the weight decays exponentially with time i.e., current observations have larger weights as compare to the earlier ones. Thus produces somehow smart forecasts as compared to the average methods.

We use the simplest exponential smoothing which used to forecast the Time Series with no Trends and seasonal patterns.

$$\hat{y}_{n+1|n} = \alpha y_n + (1-\alpha)\hat{y}_{n-1|n-1} + \alpha(1-\alpha)\hat{y}_{n-2|n-2} + \dots, \quad (7)$$

Where $\alpha, 0 \leq \alpha \leq 1$, is the smoothing parameter.

We must compare out-of-samples RMSE for the comparison.

4.1 ETS (A, N, N)

Observation Equation defined as:

$$e_t = y_t - \hat{y}_t | t-1 \quad (8)$$

State Equation defined as:

$$\ell_t = \ell_{t-1} + \alpha Z_t \quad (9)$$

- $e_t = y_t - \ell_t = \ell_{t-1} + \alpha Z_t$
- Error from single source "because some error process, e_t ".

For estimating the parameter for minimizing SSE, is to use the strategy of most likelihood estimate beneath the idea that Time Series is Gaussian. The innovation State Space has likelihood function which was computed and maximum likelihood estimate is obtained. The ETS obtained within the forecast package in R is employ to pick the acceptable model automatically on the premise maximum likelihood method and to seek out the forecast for several steps ahead as needed. Information criteria AIC, AICc, BIC used for choosing the simplest ETS Model among 30 given state space model. In ETS, the default is AICc. The model that minimizes the standard chosen as acceptable for the information.

AIC (Alkies' Information Criteria) is:

$$AIC = -2(L) + 2k, \quad (10)$$

AICc (Alkies' Information Criteria Corrected) is:

$$AICc = AIC + 2(k+1)(k+2)n - k, \quad (11)$$

BIC (Bayesian Information Criteria):

$$BIC = BIC = AIC + [\ln(n) - 2]. \quad (12)$$

MASE (Mean Absolute Scaled Error<1) defined as:

$$MASE = 1/m \sum |qt| \quad (13)$$

MAPE (Mean Absolute Percentage Error) defined as:

$$MAPE = 1/n \sum |pt|/nt = 1 \quad (14)$$

Where $pt = 100etYt/\%$.

Forecasting Technique is applied to perform forecasting by making the use of ETS function which can make use of R. To get a widely and robust applicable to ETS Model for automatic Forecasting are followed:

1. For Every Series, apply all ways that are acceptable, optimizing (both Smoothing Parameter and therefore the initial state variable) of the model in every case.
2. According to AICc value select the best Model.
3. After Selecting the Model with optimized parameter produce the point forecast.
4. To obtain the prediction intervals for the most effective model, analytic results advised by Hyndan [4] and finding $\alpha/2$ and $\alpha/2-1$ percentiles of the simulated information at every prediction Horizon.

5. Conclusion:

In this paper we presented a study of the use of Weather Forecasting Techniques with Time Series data. This is very useful since it brings better understanding of the field of Research Analysis and it is present as a vital role during this paper. We focus on the ability of ETS Model and ARIMA Model in the prediction of different weather parameters such as Rainfall, Air Temperature, Relative Humidity, Gust etc. The Study of forecasting Models which we consider in our paper is ETS Model and ARIMA Model which use for long-range of Meteorological parameters which forecasts could recognize over geographic region and will illustrate a good performance and reasonable prediction accuracy. We have given a Study for Comparing the ETS (Exponential Smoothing) Model and ARIMA (Autoregressive Integrated Moving Average) Model with estimation of MAE (Moving Absolute Error), MASE (Moving Absolute Scaled Error), MAPE (Moving Absolute Percentage Error) and RMSE (Root Mean Square Error). We also included the criteria AIC (Akaike's Information Criteria) and BIC (Bayesian Information Criteria) values to check which model deploys a best result for Weather parameters.

References

- [1] Agrawal, R. Jain, R.C. Jha, M.P. and Singh, "Forecasting of rice yield using climatic variables", Indian Journal of Agricultural Science, Vol. 50, No. 9, pp. 680-684, 1980.
- [2] G.E.P. Box, G. Jenkins, "Time Series Analysis, Forecasting and Control", Holden-Day, San Francisco, CA, 1970.
- [3] Akaike, H. (1976), "An information criterion (AIC)", Math Sci. 14, 5-9.
- [4] Hyndman, R. J., & Khandakar, "Automatic time series for forecasting: the forecast Package for R (No. 6/07)", Monash University of Department of Econometrics and Business Statistics, 2007.
- [5] Tektas M. Weather forecasting using ANFIS and ARIMA, "A case study for Istanbul. Environmental Research, Engineering and Management", 2010; 1(51):5-10.
- [6] G.Vamsi Krishna (2014), "An Integrated Approach for Weather Forecasting based on Data Mining and Forecasting Analysis", Indian Journal of Computer Science and Engineering (IJCSSE) by Vol. 5 No.2 .
- [7] Agrawal et al., Ratnadip Adhikari, R. K. Agrawal, "An Introductory Study on Time Series Modeling and Forecasting".

- [8] Pinky Saikia Dutta et.al, Hitesh Tahbiller, "Prediction of Rainfall using DataMining Technique over Assam", Indian Journal of Computer Science and Engineering (IJCSE), 2014.
- [9] Mahmudur Rahman, A.H.M. Saiful Islam , Sahah Yaser Maqnoon Nadvi , Rashedur M Rahman , "Comparative Study of ANFIS and ARIMA Model for weather forecasting in Dhaka" ,IEEE, 2013.
- [10] Dedetemo Kimilita Patrick¹, Phuku Phuati Edmond², Tshitenge Mbwebwe Jean-Marie², Efoto Eale Louis², Koto-te-Nyiwa Ngbolua, "Prediction of rainfall using autoregressive integrated moving average model", Case of Kinshasa city (Democratic Republic of the Congo), from the period of 1970 to 2009 Volume 2/ Issue 1 ISSN: 2348 – 7321, 2014.
- [11] Akaike, H., "A New Look at the Statistical Model Identification", IEEE Transaction on Automatic Control, **19**,716-723, 1974.
- [12] Garima Jain, Bhawna Mallick , "A Review on Weather Forecasting Techniques", Vol. 5, Issue 12, 2016.