

A Review on Weather Forecasting Techniques

Garima Jain¹, Bhawna Mallick²

Student, Department of Computer Science and Engineering, Galgotias Educational Institutions, Greater Noida, Uttar Pradesh, India¹

Head of Department (Computer Science), Galgotias Educational Institutions, Greater Noida, Uttar Pradesh, India²

Abstract: A brief summary of Daily meteorology s employed for multiple reasons in many areas like agriculture, transportations, etc. Accuracy of weather shown in forecast reports is very necessary. During this paper, the review is conducted to analyze a higher approach for prediction that is used for various sorts of prediction. Among which Time series with the ARIMA MODEL performs prediction with stripped-down error. For seasonal statistic prediction, Box and Jenkins [5] had projected a comparatively successful variation of ARIMA model, viz. the seasonal ARIMA (SARIMA) .The popularity of the ARIMA model is especially as a result of its flexibility to represent many sorts of statistic with simplicity still because the related Box-Jenkins[5] methodology for optimum model building ARIMA MODEL that is self-adaptive in nature which propose higher efficiency and reliability. It studied by itself in the training data and generates a lot of relative techniques that are helpful for forecasting the weather. This paper reviews varied techniques and focuses mainly on ARIMA MODEL technique for daily meteorology. The technique uses different parameters to forecast the daily weather in terms of rainfall, humidity, temperature, cloud condition, and weather of the day. The prime contribution of this paper is to compare the present meteorology model and to select the precise model to support their predictive ability.

Index Terms: Weather forecasting, Time Series, ARIMA (Auto Regressive Integrated Moving Average), Correlation.

I. INTRODUCTION

Weather forecasting has been one amongst the foremost scientifically and technologically difficult issues round the world within the last century. Weather Forecast systems are among the foremost advanced equation systems that computer needs to solve. This most advances scientifically and technological modification is attributable to two factors: 1st, it's used for several human activities and second, attributable to the resourcefulness created by the varied technological advances that are directly associated with this concrete analysis field, just like the evolution of computation and therefore the improvement in activity systems.

Weather forecasting provides critical stimulation about future weather .There are various techniques that include in weather forecasting, from easy sighting of the sky to highly complicated computerized mathematical stimulation[10]. To form associate correct prediction is one in every of the key challenges facing meteorologist everywhere on planet. Since precedent days, weather prediction has been one in every of the foremost attention-grabbing and interesting domain. Scientists have tried to forecast meteoric characteristics employing arrange of ways, a number of these ways being a lot of correct than others. Data recorded by a Weather Station at the outstanding scientific discipline centers are used for the analysis. The techniques ARIMA that is capable of handling non - stationary statistic knowledge, uses variable's lag values combining with previous and present random residual item to reveal the variation of variables. That is, at the same time it take into account all to

regressive with distinction variables. It improves the static multivariate analysis& thus it's additional wide use to predict forecasting. Meteorological data mining is a form of Data mining concerned with discovering hidden designs inside largely existing meteorological data, so that the information retrieved can be converted into usable knowledge [12]. Weather forecasting is a vital issue within the field of science on the globe. Many factors contribute considerably to extend the forecasting accuracy one in all that the development of applied statistic always for enhancing the scope and accuracy of model prediction. Several efforts are created by researchers to spot the simplest precise forecasting models .The combination of linear and nonlinear models are one in every of the foremost well-liked and was wide used in forecasting models for the forecasting accuracy.

The prime contribution of this paper is to compare the present weather forecasting model and to choose the precise model based on their analytical ability. There are several application domains of time series data, being that one important application domain is weather prediction. This will be the focus of this paper. The analysis may not be accurate if we find weather data that have the noises and outliers. Noises are the random error or parasite that comes from the sensor network, error handling and so on. On other hand the outliers is an observations so much. Therefore we need preprocessing of the weather data to improve the quality of data for precise weather prediction. The main contribution of this paper is to deliver brief literature survey of the use of statistical methods for time

series weather predictions, along with the modern trends in the use of DM for time series forecasting [2].

2. LITERATURE REVIEW ON RELATED WORK

Weather forecasting has been one of the most challenging difficulties around the world because of both its practical value in popular scope for scientific study and meteorology. Weather is a continuous, dynamic, multi-dimensional chaotic process, and data-intensive and these properties make weather forecasting a stimulating challenge. It is one of the most imperious and demanding operational responsibilities that must be carried out by many meteorological services all over the globe. Various organizations / workers in India and abroad have done demonstrating using supported time series data manipulation. The various methodologies viz. statistic decomposition models, Exponential smoothing models, ARIMA models and their dissimilarities like seasonal ARIMA models, vector ARIMA models using flexible time series, ARMAX models i.e. ARIMA with following informative variables etc., which has been used for forecasting purposes. Many trainings have taken place within the analysis of pattern and circulation of rainfall in many regions of the world. Totally altered time series methods with different purposes are used to investigate weather information in many different literatures. Accurate and timely weather forecasting is a major challenge for the scientific research. Weather prediction modelling involves a combination of many computer models, observations and acquaintance of trends and designs. Using these methods, practically accurate forecasts can be made up. Regression is a statistical experimental technique and it must be widely used in many business, the behavioural sciences, social and climate recasting and many other 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]. In [2] the author Kuo and Sun, (1993) was used to Associate in having intervention model for average 10 days stream flow forecast and synthesis that was investigated by to effect the extraordinary phenomena caused by typhoons and different serious irregularities of the weather of the Tanshui geographical area in Taiwan. In [3] Chiew et al, (1993) conducted a comparison of six rainfall-runoff modeling approaches to pretend daily, monthly and annual flows in eight tolerant catchments. They concluded that time-series approaches will agreement adequate estimates of monthly and annual yields within the water capitals of the catchments. In [4] Langu, (1993) is another approach which used statistical analysis to observe changes in weather and runoff patterns to go and look for important changes inside the parts of variety of weather statistic. Box and Jenkins (1994), in early 1970's, pioneered in developing methodologies for statistic indicating within the univariate case often known to Univariate Box-Jenkins (UBJ) ARIMA modeling in this approach of the

author[5]. The variables which describes weather conditions vary continuously with time in this explanation we found that the author [6], describing time series of each and every parameter and can be used to develop a forecasting model either statistically or using some new means that uses this time series data (**Chatfield 1994; Montgomery and Lynwood 1996**). Several authors have discussed the fuzziness associated with the weather systems.

In [7] Chaotic features are associated with the atmospheric phenomena also have fascinated the attention of the modern scientists (**Sivakumar 2001; Sivakumar et al. 1999; Men et al. 2004**). At present, the valuation of the nature and causes of seasonal climate variability is still formation. Since, it is a complicated phenomenon that includes many specialized fields of know-how to work for weather prediction (**Guhathakurata, 2006**); therefore, in the field of meteorology all assumptions are to be taken in the visage of uncertainty connected with local of and global climatic variables. Different scientists over the world have developed stochastic weather models. It is mainly used to predict and warn about how natural disasters that are caused by abrupt modification in climate conditions and has been approached using Climatic means [8]. In [9] Seyed, A., Shamsnia, M., Naeem, S. and Ali, L., (2011) was explained that the modelled weather parameter using some of the random methods (ARIMA Model) It include the Case Study: Abadeh station, Iran. Mahsin et al. (2012) used Box-Jenkins methodology to form seasonal ARIMA model for monthly weather information taken for Dhaka station, Bangladesh, for the significant amount from 1981-2010. In their paper, ARIMA (0, 0, 1) (0, 1, 1) model was found suitable and also the model is working for forecasting the monthly climate prediction. Marmara University, Vocational School of Technical Sciences, Turkey (2010) comparative study of statistical and neuro-fuzzy network models for forecasting the weather of Göztepe, İstanbul, Turkey Adaptive Network Based Fuzzy Inference System and Auto Regressive Moving Average models have been applied that ensure the efficiency of ARIMA and ANFIS techniques, different models using a different training and test data set have been tested [10] and the criteria of performance evaluation which are calculated for estimating and then comparing the performances of ARIMA and ANFIS models for forecasting has mentioned the clear review .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, for example, with Root Mean Square Error (RMSE), R-Square Error and The Sum of the Square Error (SSE) [11] and author can prove that ARIMA would give the more efficient result than other modeling techniques like ANFIS.

3. TIME SERIES AND CORRELATION MODELS

Time series modeling may be a dynamic analysis space that has attracted attentions of researchers community over previous few decades. The main aim of time series modeling is to carefully collect and strictly study the past observations of a statistic to develop associate applicable model which describes the inherent structure of the series. This model is then accustomed generate future values for the series, i.e. to create forecasts. Statistic forecasting therefore are often termed because the act of predicting the long run by understanding the past values.

A time series is non-deterministic in nature, i.e. we cannot predict with certainty what will occur in future. Generally a time series $\{x(t), t = 0, 1, 2, \dots\}$ is assumed to follow certain probability model [12] which describes the joint distribution of the random variable. Time series analysis includes strategies for analyzing statisticknowledge related to Time Series so as to extract meaningful statistics and different characteristics of the information. Time Series Models and forecastingstrategiesare unit classified into twocategories, Univariate Model primarily {based} and variable based (multivariate based).Time series modeling are very useful models when we have serially correlated data. By convention, there are two types of series which are called continuous and discrete time series. A time series in overall is supposed to be affected by four main modules, which can be separated from the observed data. These components viewed are: Trend, Cyclical, Seasonal and Irregular components.

- (i) The general tendency of a Time Series to extend, decrease or stagnate over a protracted amount of time is termed as secular Trend or just Trend.
- (ii) Seasonal variations in a Time Series measure fluctuations among a year throughout the season. The important factors inflicting seasonal differences are: climate and Weather conditions, customs, traditional habits, etc
- (iii) The alternate variation in a time series describes the medium-term changes within the series, caused by circumstances, that repeat in cycles.
- (iv) Irregular or random variations in a Time Series measure caused by unpredictable influences, which are not regular and additionally don't repeat in a very specific pattern.

(i) AUTO REGRESSIVE MODEL

An **autoregressive (AR) model** is a representation of a type of random model; as such, it is used to describe assured time-varying processes in nature, etc. In an AR(p) model the future value of a variable is expected to be a linear combination of p past observations and a constant together with a random error term. Mathematically, the AR(p) model can be expressed as process :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + Z_t$$

Equation....(1)

Where $\{Z_t\}$ denotes a only arbitrary process with zero mean and variance $\sigma^2 Z$.

(ii) MOVING AVERAGE MODEL

In time series analysis, the **moving-average (MA) model** is an approach for modeling univariate time series. The moving-average model states that the output variable regularly depends linearly on the current and several past values of a imperfectly parameters. Mathematically, the MA(q) model can be expressed as process :

$$X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$$

Equation....(2)

Where $\{Z_t\}$ denotes a only random process with zero mean and variance $\sigma^2 z$.

(iii) AUTO REGRESSIVE MOVING AVERAGE MODEL

A combine autoregressive moving average model with p autoregressive terms and q moving average terms is abbreviated ARMA (p, q) and it may be written as

$$\phi(B)X_t = \theta(B)Z_t$$

Equation.... (3)

Where $\phi(B)$, $\theta(B)$ are polynomials in B of finite order p, q, respectively.[14]

The befits of ARMA model is that numerous real data sets may be estimated in a more mean way by a mixed ARMA model rather than by a pure AR or pure MA model. Autoregressive-moving-average (ARMA) models are mathematical models of the autocorrelation, in a time series. ARMA models are widely used in dendrochronology, hydrology and other fields. There are many possible reasons for fitting ARMA models to data. [14] ARMA models can be defined as a series of equations.

(iv) AUTO REGRESSIVE INTEGRATED MOVING AVERAGE

The Box-Jenkins research to modeling and estimating time series data is one of a large family of computable forecasting methods which have been developed in the fields of processes research, management science, and statistic. Box-Jenkins models are also known as "ARIMA" models, the abbreviation standing for Autoregressive Integrated Moving Average.

An ARIMA (p, d, q) model of the nonstationary random method Y(t) is may be written as

$$\Phi(B)(1-B)^d X_t = \theta(B)Z_t$$

Equation. (5)

with an AR operator

$$\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

and a MA operator

$$\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$$

Where ϕ_p is pth AR coefficient, θ_q the qth MA coefficient, Z_t noise, X_t Rain attenuation.[2]"

(v) AUTO CORRELATION AND PARTIAL CORRELATION MODELS

As described by Box and Jenkins [5], the sample ACF is useful in determining the type of

Model to fit to a time series of length N . Since ACF is said to be symmetrical about lag zero, it is simply required to plot the sample ACF for positive lags, from lag one forwards to a maximum lag of about $N/4$. The sample PACF plot helps in identifying the maximum order of an AR process.

To conclude a proper model for a given time series data, it is necessary to transmit out the ACF and PACF analysis. These statistical methods reveal how the observations in a time series are related to each other. For forecasting and modeling purpose it must be often useful to plot the ACF and PACF against many consecutive time lags.”

4. CONCLUSION

This paper presented an overview of weather forecasting techniques with time series data describing the main contributions in this field. Papers were reviewed to emphasize the diversity of forecasting methods and the time-scales of forecasting methods. It is difficult to evaluate the performance of various methods, as the existing applications were in different time-scale and different way. This review is very useful, since it brings a better understanding of the field of analysis, and this is an important role in this paper. From the review it can be concluded that this field invites a great deal of interest by researchers. However, numerous research issues remain unexplored. One of the idea that were identified during this research is related with the combined use of various climatic issues. We focused on the capabilities of ARIMA model in the prediction of numerous weather phenomena such as humidity, temperature, atmospheric pressure etc. Future research directions include the study of ways to select the best features for time series models, with special reference to weather forecasting techniques with time series data analysis. The existence of features with different frequencies is a concern, and methods that will help how to predict this problem will be made use for future research study.

REFERENCES

- [1] Agrawal, R., Jain, R.C., Jha, M.P. and Singh, D. (1980): Forecasting of rice yield using climatic variables. *Indian Journal of Agricultural Science*, Vol. 50, No. 9, pp. 680-684.
- [2] Lee, S., Cho, S. & Wong, P.M., (1999) : Rainfall prediction using artificial neural network. — *J. Geog. Inf. Decision Anal.* 2, 233–242 1998. [10] Wong, K. W., Wong, P. M., Gedeon, T. D. & Fung, C. C. —Rainfall Prediction Using Neural Fuzzy Technique.
- [3] C. Hamzacebi, “Improving artificial neural networks’ performance in seasonal time Series Forecasting”, *Information Sciences* 178 (2008), pages: 4550-4559.
- [4] Lin, Long-Ji. (1993): Scaling up reinforcement learning for robot control. *Proceedings of the tenth International Conference on Machine Learning*.
- [5] G.E.P. Box, G. Jenkins (1970), “Time Series Analysis, Forecasting and Control”, Holden-Day, SanFrancisco,CA.
- [6] Chatfield, C. (1994): *The Analysis of Time Series-An Introduction*. Chapman and Hall.
- [7] Sivakumar, B. (2001): Rainfall dynamics in different temporal scale: A Chaotic perceptive. *Hydrology and Earth System Science*, Vol.5, pp. 645-651.
- [8] Guhathakurta, P. (2006). Long range of monsoon rainfall prediction of 2005 for the districts and sub-division in kerala with artificial neural network. *Current science*, Vol. 90, pp. 773-779.
- [9] Saima, H., Jaafar, J., Belhaouari, S. and Jillani, T.A. (2011): ARIMA based Interval Type-2 Fuzzy Model for Forecasting. *International Journal of Computer Applications*, Vol. 28, No. 3, pp. 17-21.
- [10] M.Tektas, “Weather Forecasting Using ANFIS and ARIMA (2010): A Case study of Istanbul,” *Environment Research , Engineering and Management* , vol. 1(51), pp.5-10.
- [11] Mahmudur Rahman, A.H.M. Saiful Islam , Sahah Yaser Maqnoon Nadvi , Rashedur M Rahman (2013) : Comparative Study of ANFIS and ARIMA Model for weather forecasting in Dhaka” *IEEE*.
- [12] Sarah N. kohail, Alaa M. El-Halees(2011): “Implementation of Data Mining Technique for Metrological Data Analysis”, *IJICT Journal Volume 1 No.3*, july .
- [13] S.A. Shamsnia, N. Shahidi, A. Liaghat, A.Sarraf and S.F. Vahdat. 2011. Modeling Of Weather Parameters Using Stochastic Methods. *Internat. Conference on Environment and Industrial Innovation, IPCBEE, Singapore*, 282-285.
- [14] R Development Core Team (2008). *R: A language and environment for analytic computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.Rproject.org>.
- [15] Montgomery, D.C., and Lynwood A.J. (1996): *Forecasting and Time Series Analysis*, Mc. Graw-Hill.