RESEARCH ARTICLE

MODELING OF MEAN TEMPERATURE OF FOUR STATIONS IN ASSAM.

Tanusree Deb Roy[1] and Kishore K. Das[2].
1. Department of Statistics, Assam University, Silchar-788011, Assam, India.
2. Department of Statistics, Gauhati University, Guwahati-781014, Assam, India.

……....

Manuscript Info	*Abstract*
…………………….	………………………………………………

Climate change is one of the most important issues in the recent decade. Many factors contribute towards the alarming trend of climate change. Temperature is one of the important elements of climate change. Temperature forecast can be applied in decision making and environment management fields. In this paper, an attempt has been made to study the temperature trend in four station of Assam, India viz. Guwahati, Tezpur, Silchar and Dibrugarh for the period 1981-2010. For this temperature series traditional seasonal ARIMA model has been fitted.

……....

Introduction:-
Forecasting temperature is an important and challenging problem that has received a lot of attention. One important cause is the worldwide interest in the issues of global warming and climate change. In the literature, various methods and models were investigated for forecasting temperature time series.

In this article we focus on modeling the temperature data using the seasonal autoregressive integrated moving-average model which we denote here as SARIMA. The investigation is carried out on the temperature data of four stations of Assam, India viz. Guwahati, Tezpur, Silchar and Dibrugarh. The temperature data have been obtained from Regional Meteorological Centre, Guwahati, India for a period of 30 years from 1981-2010. From the monthly mean maximum and mean minimum temperature we calculated the monthly mean temperature separately for each month.

Time series methods are used to study climate mainly to evaluate temperature. Some of the time series studies are related to Jahanbakhsh and Basseri (2003) [5] research that used the mean monthly temperature of Tabriz station based on Box & Jenkins ARIMA model to examine the monthly temperature of Tabriz for a 40 year statistical period (1959-98) based on autocorrelation and partial autocorrelation methods as well as controlling the normality of residues. Smadi (2009) [2] research was based on time series modeling of the average monthly maximum temperature data in Jordan. They fitted the traditional seasonal ARIMA model. Asmari et al. (2011) [4] studied all primary and secondary components of Mashhad, air temperatures in Mashhad, using a few simple graphical and mathematical techniques. Having these purposes, air temperature records (from 1961 to 2005) taking from Mashhad, Synoptic Station were analyzed. Shamsnia (2011) [6] did research on 20 years statistics of relative humidity and monthly average temperature and precipitation of Abadeh Station by ITSM time series analysis software. Thus, considering the importance of climatic parameters and their modeling and forecasting by stochastic methods is a necessity and could be a basic role in decision making and environment management fields.

Corresponding Author:- Tanusree Deb Roy.
Address:- Department of Statistics, Assam University, Silchar-788011, Assam, India.

Materials and Methods:-

In time series analysis, the Box-Jenkins methodology is used to find the best fit of a time series to past values of this time series, in order to make forecasts. This approach possesses many appealing features. To identify a perfect ARIMA model for a particular time series data, Box and Jenkins (1976) [9] proposed a methodology that consists of four phases viz. i) Model identification; ii) Estimation of model parameters; iii) Diagnostic checking for the identified model appropriateness for modeling and iv) Application of the model (i.e. forecasting). The first step in developing a Box-Jenkins model is to determine if the time series is stationary and if there is any significant seasonality that needs to be modeled. The data was examined to check for the most appropriate class of ARIMA processes through selecting the order of the consecutive and seasonal differencing required making series stationary, as well as specifying the order of the regular and seasonal autoregressive and moving average polynomials necessary to adequately represent the time series model. Stationary of a data set can be identified by performing Portmanteau test, which is used to test whether a data set is significantly different from a zero set. A common portmanteau test is the Box-Pierce test, designed by Box and Pierce (1970) [8]. This residual from a forecast model test is based on the Box-Pierce statistic:

$$Q = n \sum_{k=1}^{h} r_k^2$$

where, h is the maximum lag being considered, n is the number of observations in the series and r_k is the autocorrelation at *lag k*. If the residuals are white noise, the statistic Q has a chi-square (χ^2) distribution with degrees of freedom (h - m) where m is the number of parameters in the model which has been fitted to the data. An alternative portmanteau test is the Ljung-Box due to Ljung and Box [3]. They argued that the alternative statistic:

$$Q^* = n(n+1) \sum_{k=1}^{n} \frac{r_k^2}{n-k}$$

has a distribution closer to the chi-square distribution with (h - m) degrees of freedom than does the Q statistic. It is normal to conclude that the data are not white noise if the value of Q or Q^* lies in the extreme 5% of the right-hand tail of the (χ^2) distribution. To choose the best model among the class of plausible models we use the Akaike's Information Criterion (AIC), proposed by Akaike (1974) [1]. The model which has the minimum AIC value is our model of interest.

To test whether the chosen model satisfies the normality test ShapiroWilk test has been used. The test statistic for ShapiroWilk test [7] is given by

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)} \right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

where $x_{(i)}$ is the *i-th* order statistic, $\bar{x}$ is the sample mean, the constants a_i are given by

$$a_1, a_2, \dots, a_n = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}$$

where m = $(m_1, m_2, \dots, m_n)^T$ and $m_1, m_2, \dots, m_n$ are the expected values of the order statistics of independent and identically distributed random variables sampled from the standard normal distribution, and V is the covariance matrix of those order statistics.

Results and Discussion:-

The summary of the descriptive statistics of the mean temperature data is given in Table (1). It can be concluded that Silchar received the highest mean of mean temperature during the last 30 years. The coefficient of variation for Dibrugarh followed by Guwahati is found to be higher as compared to other stations. This may indicate that the amount of mean temperature in those two stations is relatively more spread as compared to other stations.

Table 1:- Mean temperature characteristics

Stations	Years	Mean	Standard Deviation	Coefficient of variation
Guwahati	30	24.71	4.06	0.16418
Tezpur	30	24.25	3.95	0.16295
Silchar	30	25.24	3.54	0.14033
Dibrugarh	30	23.53	4.22	0.17938

For modeling by ACF and PACF methods, examination of values relative to auto regression and moving average were made and at last, an appropriate model for estimation of precipitation values for the stations Guwahati, Tezpur, Silchar and Dibrugarh were found. To prevent from excessive fitting errors, Akaikie's (AIC) statistic was used.
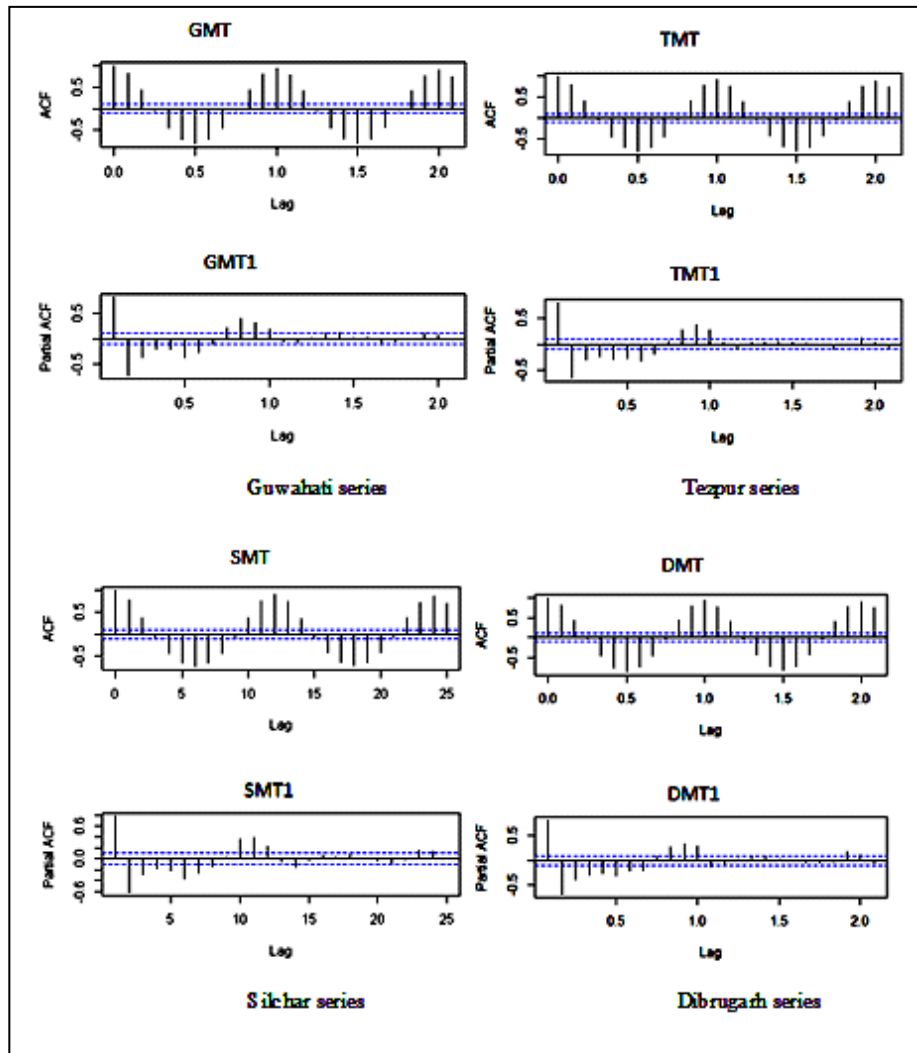

Fig. 1:- Plot of ACF and PACF for temperature series

The ACF and PACF plots of the original data, as shown in Figure 1, show that the temperature data is not stationary. To make the data stationary, seasonal first difference, $D = 1$, of the original data have been used for all the stations. Therefore, an ARIMA (p, 0, q) (P, 1, Q) ^{12} model could be identified for the differenced and de-seasonalized temperature data for all the stations.
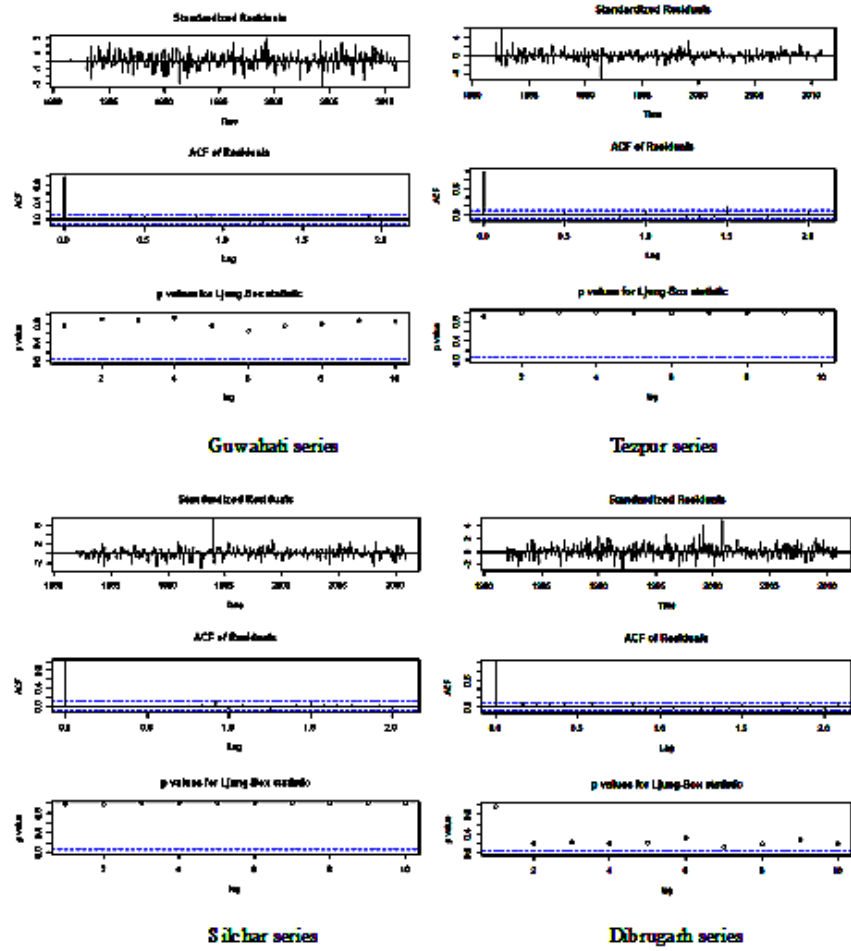
Fig. 2:- Diagnostics for the SARIMA fit on the temperature series

Based on the automatic ARIMA forecasting our selected model are ARIMA $(1,0,0)(0,2,2)^{12}$, $(12,0,0)(0,1,1)^{12}$, $(0,0,10)(0,1,1)^{12}$ and $(0,0,1)$ $(0,1,1)^{12}$ for the stations Guwahati, Tezpur, Silchar and Dibrugarh which are adequate to represent the temperature data which could be used to forecast the upcoming temperature data. Figure 2 displays a plot of the standardized residuals, the ACF of the residuals and the p-values associated with the Q-statistic; it appears this model fits the data well. The ACF of the standardized residuals shows no apparent departure from the model assumptions, and the Q-statistic is never significant at the lags shown.

Table 2:- Mean temperature characteristics

Station	Model	Ljung–Box test	Shapiro Wilk Normality test	AIC	RMSE
Guwahati	$(1,0,0)(0,2,2)^{12}$	$\chi^2 = 9.91$, df = 20, p-value = 0.9697	W = 0.9872, p-value = 0.00286	857.54	0.70358
Tezpur	$(12,0,0)(0,1,1)^{12}$	$\chi^2 = 21.55$, df = 20, p-value = 0.3655	W = 0.956, p-value = 6.626e-09	938.12	0.84802
Silchar	$(0,0,10)(0,1,1)^{12}$	$\chi^2 = 11.09$, df = 20, p-value = 0.9438	W = 0.9432, p-value = 1.633e-10	931.58	0.85188
Dibrugarh	$(0,0,1)(0,1,1)^{12}$	$\chi^2 = 23.33$, df = 20, p-value = 0.2729	W = 0.9819, p-value = 0.00017	874.7	0.81309

The values of the Ljung-Box test for the models are given with degree of freedom and p- value, giving further indication that the model has captured the dependence in the time series. The Shapiro-Wilk test of normality for all

the stations proves that normality is not rejected at any significance level. The AIC and root mean square error values are also depicted in the table for selection of the model as given in Table (2).
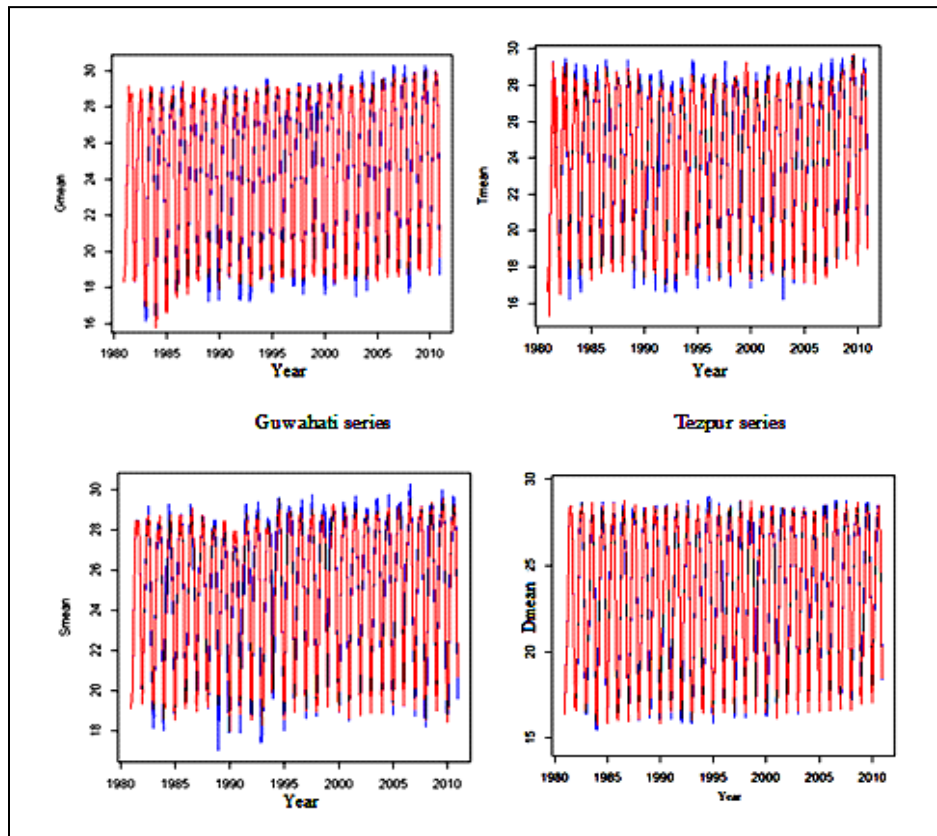

Figure 3:- Comparison graph of actual vs. fitted values of the Temperature series

Figure 3 shows a comparison between the real values and the ones resulted from the developed ARIMA model for the period 1981 to 2010. Using the seasonal ARIMA models, we obtain the graphical plot for the actual temperarure series versus the fitted temperature series. From the visual inspection of the plot, it is quite evident that the chosen model is good enough as the predicted series is very close to the observed series and this model could be used to forecast the upcoming temperature for the stations Guwahati, Tezpur, Silchar and Dibrugarh, Assam, India.

References:-

1.   A. Hirotugu. "A new look at the statistical model identification." *IEEE Transactions on Automatic Control*, vol. 19(6), pp. 716–723, 1974.
2.   A. Smadi. "Periodic Auto-Rregression Modeling of the Temperature Data of Jordan." *Pak. J. Statist.*, vol. 25(3), pp. 323-332, 2009.
3.   Ljung-Box test, http://en.wikipedia.org/wiki/Ljung%E2%80%93Box_test.
4.   M. Asmari, F. Abbasi2 and H. Arabshahi, "Preliminary Time Series Analysis of Mashhad Air Temperatures (1961-2005)", *International Journal of Science and Advanced Technology,* vol.1(6), Aug, 2011.
5.   S. Jahanbakhsh, E.A. Babapour Basseri. "Studying and forecasting of the mean monthly temperature of Tabriz,
6.   using ARIMA model."*Journal of Geographic Researches*, 15(3), pp. 34-46, 2003.
7.   S. A. Shamsnia, N. Shahidi, A. Liaghat, A. Sarraf and S. F. Vahdat. "Modeling of Weather Parameters Using Stochastic Methods (ARIMA Model)(Case Study: Abadeh Region, Iran)." *International Conference on Environment and Industrial Innovation,* vol.12, 2011.
8.   Shapiro-Wilk Normality test, http://en.wikipedia.org/wiki/Shapiro%E2%80%93Wilk_test.
9.   G. E. P. Box, and , D. A. Pierce. "Distribution of the Autocorrelations in Autoregressive Moving Average Time Series Models." Journal of the American Statistical Association, vol. 65, pp. 1509–1526, 1970.
10. G.E.P. Box, and G.M. Jenkins. *Time Series Analysis: Forecasting and Control* , Revised Edn., Holden-Day, San Francisco, 1976.