Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/090115

Temporal Variation of Temperature in Guwahati, Assam: An Application of Seasonal ARIMA Model

Utpal Dhar Das[1], *Brijesh P. Singh*[1*] *and Tanusree Deb Roy*[2]

[1]Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi, India
[2]Department of Statistics, Assam University, Silchar, India

Received: 4Jan. 2019, Revised: 20 Sep. 2019, Accepted: 24 Sep. 2019.
Published online: 1 Mar. 2020.

Abstract: Analysis of time series has become a major tool in various applications in environmental areas to understand phenomena, like rainfall, temperature, humidity, draught and so forth. Here, SARIMA model has been used to carry out short-term predictions of monthly maximum and minimum temperature in Guwahati city of Assam. The present study aimed at analysing the temporal variation in temperature in Guwahati during the period 1981-2018. The study of linear trend indicated increasing trends in mean and maximum temperatures and decreasing trends in minimum temperatures, Mann-Kendall test has been used to know the variability in trends in annual and seasonal temperature. The analysis reveals an increasing pattern in temperature both seasonally and annually during the period 1981-2018. The analysis of moving average indicated two periods when annual mean temperature increased significantly. These periods are closely associated with the post-independence period of industrialization (1955-1965) and recent phase of urbanization (after 2000). The analysis also indicates that the rate of change in temperatures brought by post-independence period of industrial activities is higher than that of recent year's urbanization. Again, the monthly and seasonal analysis of temperature revealed a sharp decline in minimum temperature during winter season and especially in the coldest month (January) than any other seasons.

Keywords: Box-Pierce Test, Climate Change, Forecasting, Mann-Kendall (MK) Test, SARIMA, Shapiro-Wilk Test.

1 Introduction

Climate change is one of the biggest environmental threats to water availability, food production, forest biodiversity and livelihoods for many countries in the world. Moreover, it is widely believed that developing countries in tropical regions of the world, e.g. India, will be affected more severely than the developed ones [1,2]. Understanding the nature and scale of possibleclimate changes is of utmost importance to the policy makers and people of North-Eastern India as it gives them vital information to prepare efficient mitigation and adaptivemeasures. Time series analysis ofweather data can be a very valuable tool for this purpose. It can be considered as a significant tool in detecting and developing variability pattern and, maybe, even to predict short as well as long-term variations in the time series. Even though a single extreme climate-related event cannot be attributed unequivocally to climate change, the probability of high temperature events will increase if there is an underlying trend of rising mean temperature. In fact, according to data from the reinsurance industry, the number of climate-related disasters has increased significantly over time [3,4].

Weather data are generally classified as either synoptic data (meteorological data) or climate data. Synoptic data is the real time data provided which is used mainly in aviation safety protocols and forecast modeling. Climate data is the official data record, usually provided after some quality control tests performed on it. Climate change plays a major role in environmental change. Environmental change is defined as a change or disturbance of the environment most often caused by natural ecological processes and human influences [5]. Climate change not only affects ecosystems species directly, but also interacts with other human stressors like infrastructural development. Moreover, logged forested areas may become

*Corresponding author e-mail: brijesh@bhu.ac.in

vulnerable to erosion if climate change leads to increase in heavy rain storms, which may include a number of factors such that human interference, natural disasters or animal interaction. Climate change is a long-term alteration in weather conditions identified by changes in temperature, magnitude of precipitation, wind and so forth. Climate change is also brought into effect by factors such as variations in solar radiation received by Earth, shifting of plate tectonics, biotic processes and volcanic eruptions. Certain human activities, notably carbon dioxide and methane emission, contribute to trapping the sun's heat in the atmosphere and are also known as greenhouse gases, they exist normally in the atmosphere and keep the earth's surface warm enough to sustain life. Science has made enormous inroads in understanding climate change and its causes, and is beginning to help develop a strong understanding of current and potential impacts of climate change that will affect people of today and the generation of the upcoming decades. This understanding is crucial to life on earth as it allows decision makers to place climate change in the context of other large challenges being faced by the nations and the world as a whole. There are and always will be some uncertainties, in understanding a complex system like Earth's climate [6, 7, 8].

There are many natural factors for climate change. Temperature is one of the major factors of climate change. Temperature is an objective comparative measure of hot or cold. It is measured by temperature gauges such as thermometers, which may work through the bulk behavior of a thermometric material, detection of thermal radiation, or particle kinetic energy [9, 10, 11]. Several scales and units exist for measuring temperature, the most common being Celsius (denoted by °C; formerly called centigrade) Fahrenheit (denoted by °F), and especially in science, Kelvin (denoted by K). Climate change is seldom predictable. By using different climate model calculations, scientists can state that the earth's climate is unstable and the human beings have played an important role in bringing forward this change. The very rapid development of technology and exponential growth of population has brought many ecological crises to different regions of the earth. As a result of industrial growth and deforestation, levels of carbon dioxide, methane, CFCs, aerosols and particulate matter have escalated in the atmosphere to such an extent that they are negatively affecting the Earth's climate. According to IPCC, "The magnitude and timing of climate change due to all types of activities (natural and man-made activities) will depend on the ultimate concentrations of greenhouse gases and aerosols (particles) and their rates of growth and on the detailed response of the climate system". Greenhouse gas levels, especially that of CO_2, have risen considerably after the industrial revolution. Global warming (due to increase in absorption of the solar radiation) is a direct consequence of the increase in the levels of Greenhouse gases in the environment.

The pollution of the air by dust, particulate, fumes and aerosols, created either by natural or human activities, provides the basis of another critical problem. That is the cooling of the Earth's ambient temperature. Approximately one third of incoming solar radiation is reflected back into the atmosphere and the remainder is absorbed by the earth. A positive radioactive force tends to warm the surface; but a negative one tends to cool the surface. To observe this radiation change throughout the years, meteorological air temperatures are studied and this information form the basis of different scenarios for climate change. Earth's climate is changing and that leads to rapid change in regional climate as well. Weather and climate have elements or variables which act separately, but are still interrelated in some ways. One of such variables is temperature. Temperature could be that of the sea (water), soil, body, ambient or air. The temperature described here is the air temperature which is described as how hot, warm or cold a place is at a particular point of time. Temperature is controlled by the amount, duration and intensity of solar output which in turn is controlled mostly by the seasons and the position of the Earth with respect to the Sun. Temperature is very crucial because it influences health, agriculture, evaporation, transport, rainfall and human comfort among others [12,13]. Many studies have been conducted on temperature [14, 15, 16]. A study on climate change done by various researchers found that earth's temperature has been gradually increasing over the last century. As such an attempt has been made to study the temporal variation in temperature over Guwahati city for a period of 38 years.

2 Objectives

The present research has the following objectives.
1. To study the temperature trend for the maximum and minimum temperature of Guwahati city for the period 1981-2018.
2. To develop a suitable time series model for maximum and minimum temperature of Guwahati city.
3. To forecast, monthly, the maximum and minimum temperature of Guwahati city by SARIMA model.

3 Methodologies

This chapter provides the details on the source of data and study area for the study. It provides the details of the methodologies which are used for the computation of the results. For this study, we have used Mann-Kendall rank statistics test to confirm the significance of the observed trend. SARIMA model, SARIMA Box-Jenkins method to find seasonal autoregressive integrated moving average (SARIMA) models for the best fit of a time-series model to past values of a time series, Shapiro-Wilk Normality test and Box-Pierce test.

3.1 Study Area and Source of Data

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. Monthly temperature data covering the Guwahati and the neighboring district of Assam has been collected from the Indian Meteorological Department (IMD), New Delhi which is the principal organization gathering meteorological data in India. Temperature station of Guwahati city is situated in North-east India at latitude $26^0$11'N, longitude $91^0$44'E. The temperature data covers a period of 38 years, from 1981 to 2018.

3.2 Time Series

A time series is a series of data points indexed (or listed or graphed) in time order. Most commonly, a time series is a sequence taken at successive equally spaced points in time. Thus it is a sequence of discrete-time data. Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values [10,11]. Data, which are obtained from the observations of a phenomenon over time, are extremely common and named as Time Series. Time series analysis can be applied to real valued, continuous data, discrete numeric data, or discrete symbolic data. Time series have some fluctuations trend, seasonal variations, cyclical variations, irregular fluctuations.

Trend estimation is a statistical technique to aid in the interpretation of data. When a series of measurements of a process are tested against a time series, trend estimation can be used to make and justify statements about tendencies in the data, by relating the measurements to the times at which they occurred. This model can be used to describe the behavior of the observed data. In particular, it may be useful to determine if measurements exhibit an increasing or decreasing trend which is statistically distinguished from random behavior. Seasonal variations in a time series are due to the rhythmic forces which operate in a regular and periodic manner over a span of less than a year i.e. during a period of 12 months and have the same or almost same pattern year after year. Thus, seasonal variations in a time series will be there if the data are recorded quarterly, monthly, weekly, daily, hourly and so on. The oscillatory movements in a time series with a period of oscillation more than one year are termed as cyclic fluctuations. One complete period is called a 'cycle'. Cyclic fluctuations, though more or less regular, are not periodic. Irregular fluctuations are purely random, erratic, unforeseen, and unpredictable and are due to numerous non-recurring and irregular circumstances which are beyond the control of human hand. Methods for time series analysis may be divided into two classes: frequency-domain methods and time-domain methods. The former include spectral analysis and wavelet analysis; the latter include auto-correlation and cross-correlation analysis. Time series analysis techniques may be divided into parametric and non-parametric methods. The parametric approaches assume that the underlying stationary stochastic process has a certain structure which can be described using a small number of parameters.

3.3 Autocorrelation Function

Autocorrelation, also known as serial correlation, is the correlation of a signal with a delayed copy of itself as a function of delay. Informally, it is the similarity between observations as a function of the time lag between them. The analysis of autocorrelation is a mathematical tool for finding repeating patterns, such as the presence of a periodic signal obscured by noise, or identifying the missing fundamental frequency in a signal implied by its harmonic frequencies. It is often used in signal processing for analyzing functions or series of values, such as time domain signals.

3.4 Partial Autocorrelation Function

In time series analysis, the partial autocorrelation function (PACF) gives the partial correlation of a time series with its own lagged values, controlling for the values of the time series at all shorter lags. It contrasts with the autocorrelation function, which does not control for other lags. The partial auto-correlation function plays an important role in data analyses. Also the aim of the function is to identify the extent of the lag in an autoregressive model. The Box–Jenkins approach to time series modeling uses this function,where graph of the partial autocorrelation determines the appropriate lags p in an AR(p) model or in an extended ARIMA (p,d,q) model. In a time series Z_t the partial autocorrelation of lag k, between Z_t and Z_{t+k} denoted by $\alpha(k)$.

$$i.e\ \alpha(1) = corr(Z_{t+1}, Z_t), \tag{1}$$

$$\alpha(k) = (Z_{t+k} - P_{t,k}(Z_{t+k}), Z_t - P_{t,k}(Z_t)),\ k \geq 2 \tag{2}$$

3.5 Mann-Kendall Test

For n consecutive observations of a time series $Z_t, t = 1, 2, ..., n$, Mann (1945) suggested using the Kendall rank correlation of Z_t with $t, t = 1, 2, ..., n$ to test for monotonic trend. No trend in null hypothesis assumes that Z_t are independently distributed. The S-Plus function, Mann-Kendall (z) implements the Mann-Kendall test using Kendall (x, y) to compute τ and its significance level under the null hypothesis. The Mann-Kendall trend test has some desirable features. Here we have used Mann-Kendall rank statistics to confirm the significance of the observed trend. Mann-Kendall test had been formulated as a non-parametric test for trend detection and the test statistic distribution had been given by [17] for testing non-linear trend and turning point. The purpose of the Mann-Kendall test is to statistically assess if there is a monotonic upward or downward trend of the variable of interest over time. A monotonic upward (downward) trend means that the variable consistently increases (decreases) through time, but the trend may or may not be linear. The Mann-Kendall Statistic is computed as follows:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sign(T_j - T_i) \tag{3}$$

Where T_j and T_i are the annual values in years j and i, $j > i$, respectively. If $n < 10$, the value of $|S|$ is compared directly to the theoretical distribution of S derived by Mann and Kendall. For $n \geq 10$, the statistic S is approximately normally distributed with the mean and variance as follows

$$E(S) = 0$$

$$\sigma^2 = \frac{n(n-1)(2n+5) - \sum t_i(i)(i-1)(2i+5)}{18} \tag{4}$$

in which t_i denotes the number of ties to extent i. The summation term in the numerator is used only if the data series contains tied values. The standard test statistic Z_s is calculated as follows

$$Z_s = \begin{cases} \frac{S-1}{\sigma}; S > 0 \\ \frac{S+1}{\sigma}; S > 0 \end{cases} \tag{5}$$

The test statistic Z_s is used as a measure of significance of the trend. Another statistics obtained on running the Mann-Kendall test is Kendall's tau. It is a measure of correlation and therefore measures the strength of the relationship between the two variables. The Mann-Kendall test is essentially limited to testing the null hypothesis that the data are independent and identically distributed. Our time series data may diverge from this assumption in two ways. First, there may be autocorrelation and second, there may be a seasonal component. To eliminate these factors, we can use annual data but this reducing the power of test. For strong positive autocorrelation in the series, the effect of using annual totals will reduce the effect of this autocorrelation substantially and the loss of power is, perhaps, not expected to be too much.This is something we will investigate further in a methodological study.

3.6 Shapiro-Wilk test

The Shapiro-Wilk test is a test of normality in frequentist statistics. It was published in 1965 by Samuel Sanford Shapiro and Martin Wilk. The Shapiro-Wilk test tests the null hypothesis that a sample $x_1, x_2, ..., x_n$ came from a normally distributed population. The test statistic is given by

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{6}$$

where and $x_1, x_2, ..., x_n$ are the order statistics of independent and identically distributed random variables sampled from the standard normal distribution. The coefficients a_i are given by

$$(a_1, a_2, ..., a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{\frac{1}{2}}} \tag{7}$$

where $m = (m_1, m_2, ..., m_n)^T$ is the expected values of order statistics of independent and identically distributed random variables sampled from the standard normal distribution and V is the covariance matrix of these order statistics.
The null hypothesis of this test is that the population is normally distributed. Thus, if the p-value is less than the chosen alpha-level, then the null hypothesis is rejected and there is evidence that the data tested are not from a normally distributed population; in other words, the data are not normal.

3.7 Box-Pierce test

The Box-Pierce formula is as follows: $Q = n\sum_{k=1}^{h} r_k^2$ which we will compare against the χ^2 distribution; n is the total number of observations and h is the maximum lag we are considering. Essentially, the Box-Pierce test indicates that if residuals are white noise, the Q-statistic follows a χ^2 distribution with $(h-m)$ degrees of freedom. If a model is fitted, then m is the number of parameters. However, no model is fitted here, so $m = 0$. If each r_k value is close to zero, then Q will be very small; otherwise, if some r_k values are large - either negatively or positively - then Q will be relatively large. We will compare Q to the χ^2 distribution, just like any other significance test.

3.8 SARIMA(Seasonal ARIMA)

In time series possess a seasonal component that repeats after 's'observations. For monthly observations,s= 12 (12 in 1 year), for quarterly observations, s= 4 (4 in 1 year). In order to deal with seasonality, ARIMA processes have been generalized and it is called as SARIMA model and it is formulated as

$$\Phi(\text{B})\Delta^t X_t = \theta(\text{B})\alpha_t \quad (8)$$

Autoregressive Integrated Moving Average (ARIMA) models in forecasting were popularized by [3] in the 1970s, and were traditionally known as Box-Jenkins analysis. The purpose of ARIMA methods is to fit a stochastic (randomly determined) model to a given set of time series data, such that the model can closely approximate the process that is actually generating the data. There are three main steps in ARIMA methodology; identification, estimation and diagnostic checking, and then application. Before undertaking these steps, however, an analyst must be sure that the time series is stationary. That is, the covariance between any two values of the time series is dependent upon only the time interval between those particular values and not on their absolute location in time.

To determinethat a time series is stationary, it is require to use autocorrelation function (ACF), which is also called a correlogram. For climate data which usually follows a seasonal, i.e. an annual cycle, it is more appropriate to use a seasonal ARIMA $(p,d,q)(P,D,Q)S$ (p, d, q) (P, D, Q)S model, whereby P is the order of the seasonal AR-model; D is the order of the seasonal differencing (for monthly data, usually, D = 12) and Q is the order of the seasonal MA-model and s is the number of periods in the season (S =12, for an annual cycle). When two out of the three terms are zeros, the model may be referred to based on the non-zero parameter, dropping "AR", "I" or "MA" from the acronym describing the model. For example, ARIMA (1,0,0) is AR(1), ARIMA(0,1,0) is I(1), and ARIMA(0,0,1) is MA(1).The ARIMA is of the form,

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + ... + \alpha_p X_{t-p} + Z_t + ... + \beta_1 Z_{t-1} + ... + \beta_q Z_{t-q} \quad (9)$$

4 Results and Discussion

In this study, 38 years monthly data have been considered to study the temporal variation in Guwahati. December, January and February are considered for the analysis of winter temperature as these three months record lower temperatures. While computing the mean for winter season, December of the previous year is included. March, April and May are the months which represent the summer season. June to September months constitute monsoon season and October and November form the post monsoon season. From the monthly mean maximum and mean minimum temperature, we calculated the monthly mean temperature separately for each month. Accordingly, the yearly totals were calculated for each year.

4.1 Analysis of Maximum Temperature

To get a view of temperature trend for maximum temperature of Guwahati, we plot the Figure 4.1, it is apparent that the mean maximum temperature is showing a steady increase with time. There is also a clear cycle in the data that has a one-year period, i.e. there is clear seasonal variation in the data.

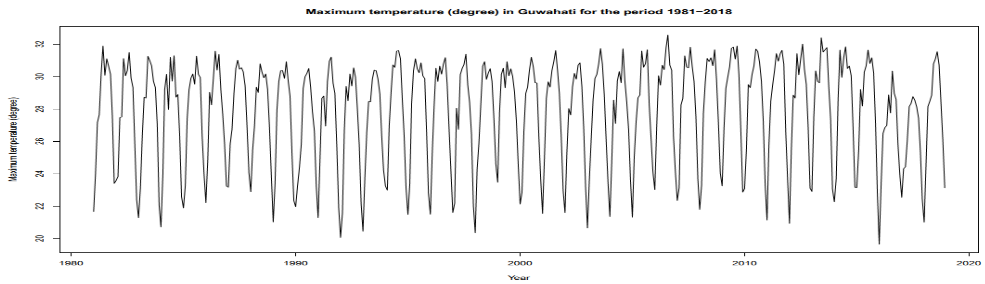


Fig. 4.1: Maximum monthly temperatures of Guwahati from 1981-2018.

Figure 4.2 gives the correlation for the temperature data over the 30 year period 1981-2018, from which the following features are observed. The x-axis indicates the lag (k) and the y-axis for the autocorrelation (ρ_k) at each lag, i.e. the correlation is a plot of ρ_k against k . Here, the units on the x-axis are 'years', so lags 0, 1, 2 . . . (in months) appear at times 0, 1/12, 2/12 . . . (in years).

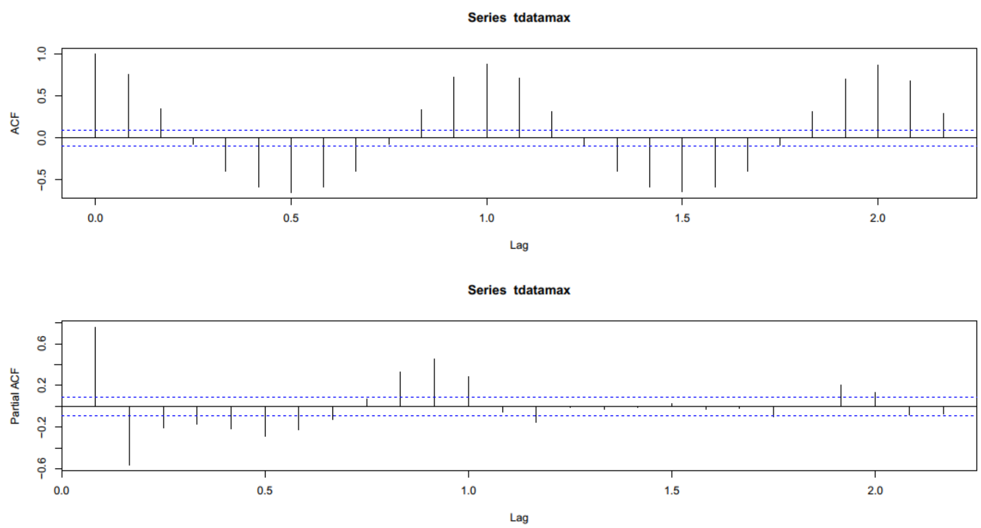


Fig.4.2: ACF and PACF of the maximum temperature.

Correlation is dimensionless, so there are no units for the y-axis. The dotted lines represent the 5% significance level for the statistical test $\rho_k = 0$. Any correlations that fall outside these lines are 'significantly 'different from zero. In particular, as they represent multiple significance tests, 5% of the values would show up as 'significant', due to sampling variation, even when their underlying values are zero.

The lag 0 autocorrelation is 1 and is shown on the plot. It provides an indicator of the relative values of the other autocorrelations; if a correlation turns out to be statistically significant, but is actually very small in magnitude, it may be of little or no practical consequence, and will look 'insignificant' alongside ρ_0 . For example, a lag 1 autocorrelation of 0.1 implies that a linear dependency of x_t on x_{t+1} would only explain 1% of the variability between the two variables. The annual cycle appears in the correlation as a cycle of the same period, with a significant positive correlation at period 1-year. This reflects a positive linear relationship between pairs of variables, separated by 12-month periods. Conversely, values separated by a period of 6 months will tend to have a negative relationship, because, for example, higher values tend to occur in the monsoon months followed by lower values in the winter months. A negative correlation therefore occurs at lag 6 months (or 0.5 years). The significant autocorrelation at lag 1 month is probably due to the increasing trend over the period of the data. However, the correlation, exhibits a gradual decay in the autocorrelations, due to the trend, as expected. To get a clearer view of the trend, the seasonal effect is removed by aggregation of the data to the annual level. A summary of the values for each season is viewed using a boxplot in Figure 4.3. The seasonal effects are clearly revealed in the boxplot, from which we can observe that summer and monsoon season have high mean maximum temperature. This is showing an increasing pattern with time temperature that has increased during the summer and monsoon seasons.

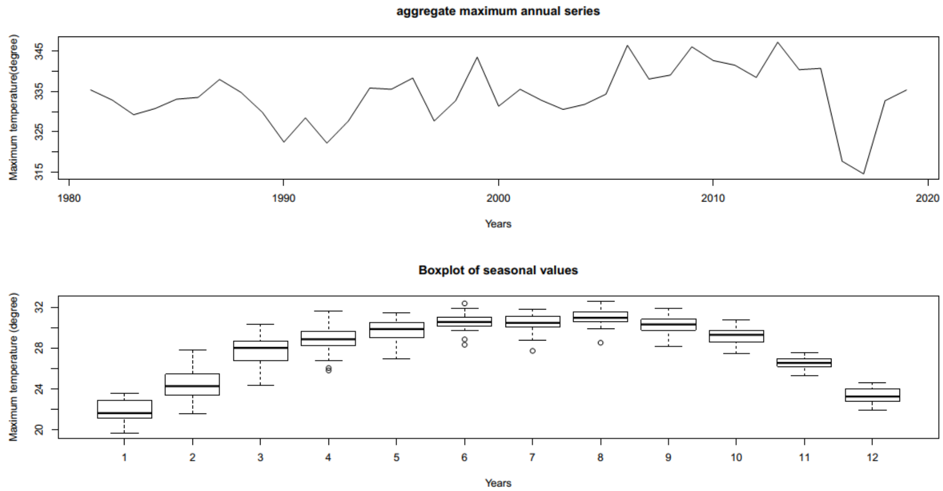


Fig.4.3: Boxplot of maximum monthly temperature.

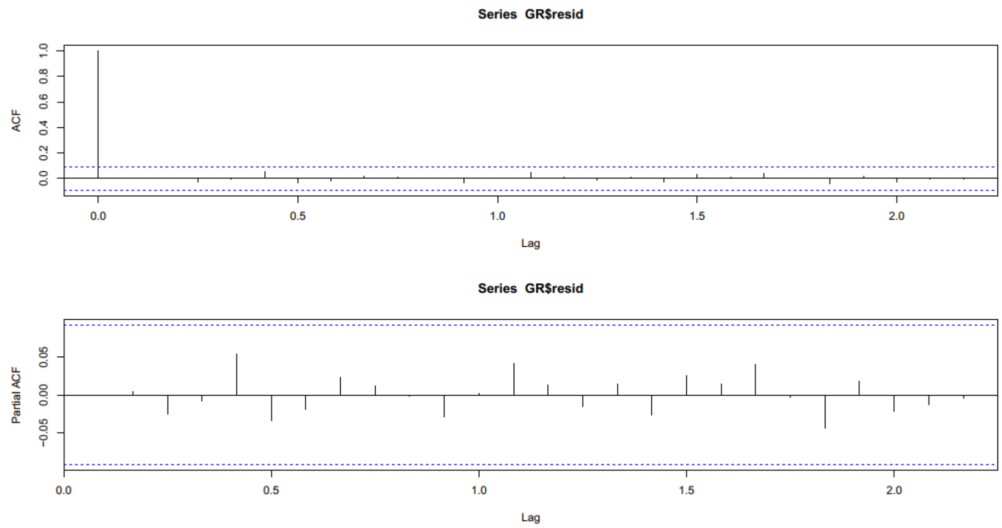


Fig. 4.4: ACF and PACF of the maximum temperature series.

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 is found. To prevent excessive fitting errors, Akaikie's (AIC) statistic was used. Based on the automatic ARIMA forecasting, our selected model is SARIMA. $(3,0,3)(1,1,2)^{12}$ for the stations Guwahati. This shows that 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. It is clearly evident from the ACF plot above that none of the autocorrelation coefficients between lag 1 and 20 are breaching the significant limits i.e. all the ACF values are well within the significant bounds and all the PACFs or partial autocorrelation coefficients of residuals of fitted ARIMA for lag 1 to lag 20 are within the significant limits. This means ACF and PACF concluded that there is no non-zero autocorrelations in the forecast residuals (or standard errors) at lag 1 to 20 in the fitted model.

4.2 Analysis of Minimum Temperature

From Figure 4.5 we can view that the mean minimum temperature is showing a steady increase with time. There is a clear cycle in the data that has a one-year period, i.e. there is seasonal variation in the data.

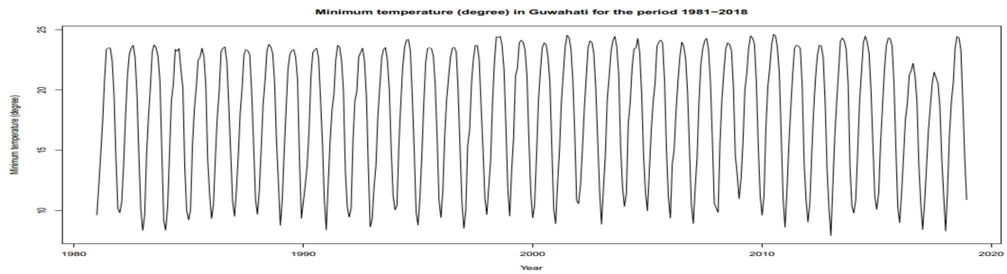


Fig.4.5 Minimum monthly temperatures for Guwahati from 1981-2018

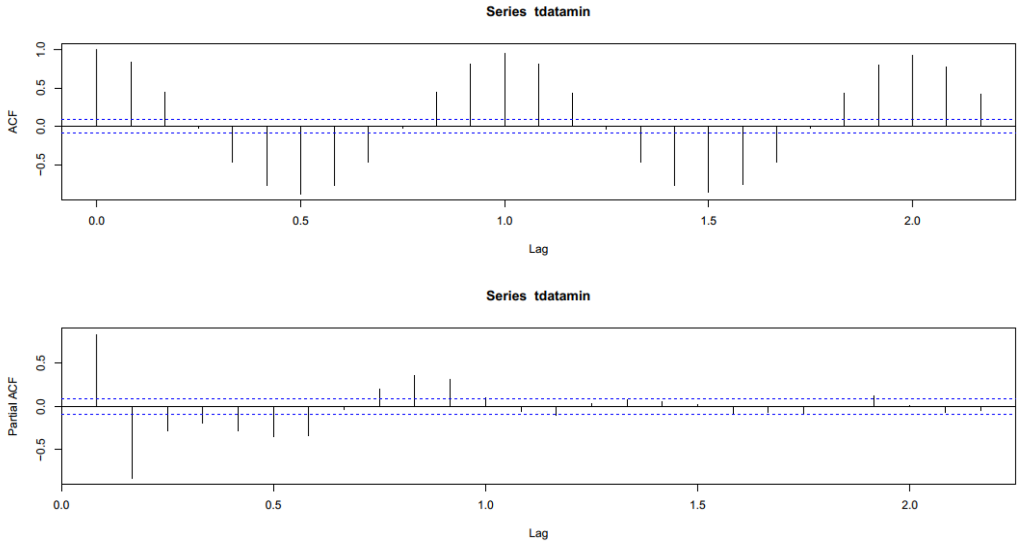


Fig. 4.6: ACF and PACF of the minimum temperature.

Figure 4.6, gives the correlation of the temperature data over the 38 year period 1981-2018, from which the following results are observed. The x-axis gives the lag (k) and the y-axis gives the autocorrelation (ρ_k) at each lag, i.e. the correlation is a plot of ρ_k against k. Here, the units on the x-axis are 'years', so lags 0, 1, 2 . . . (in months) appear at times 0, 1/12, 2/12 . . . (in years). The lag 0 autocorrelation is 1 and is shown on the plot. It provides an indicator of the relative values of the other autocorrelations; if a correlation turns out to be statistically significant. The annual cycle appears in the correlation as a cycle of the same period, with a significant positive correlation at period 1-year. This reflects a positive linear relationship between pairs of variables, separated by 12-month periods. Conversely, values separated by a period of 6 months will tend to have a negative relationship, because, for example, higher values tend to occur in the monsoon months followed by lower values in the winter months. A negative correlation therefore occurs at lag 6 months (or 0.5 years). The significant autocorrelation at lag 1 month is probably due to the increasing trend over the period of the data. However, the correlation, exhibits a gradual decay in the autocorrelations, due to the trend, as expected.

From the boxplot, Figure 4.7, we can observe that winter season has low mean minimum temperature and also we can observe that this is showing an increasing pattern with time and has increased during the summer and monsoon months. The 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 is found. To prevent excessive fitting errors, Akaikie's (AIC) statistic was used. Based on the automatic ARIMA forecasting, our selected model is SARIMA. $(3,0,2)(1,1,0)^{12}$ for the stations Guwahati .

This shows that 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. It is clearly evident from the ACF plot that none of the autocorrelation coefficients between lag 1 and 20 are breaching the significant limits i.e. all the ACF values are well within the significant bounds and all the PACFs or partial autocorrelation coefficients of residuals of fitted ARIMA for lag 1 to lag 10 are within the significant limits. This means ACF and PACF concluded that there is no non-zero autocorrelations in the forecast residuals (or standard errors) at lag 1 to 20 in the fitted model.

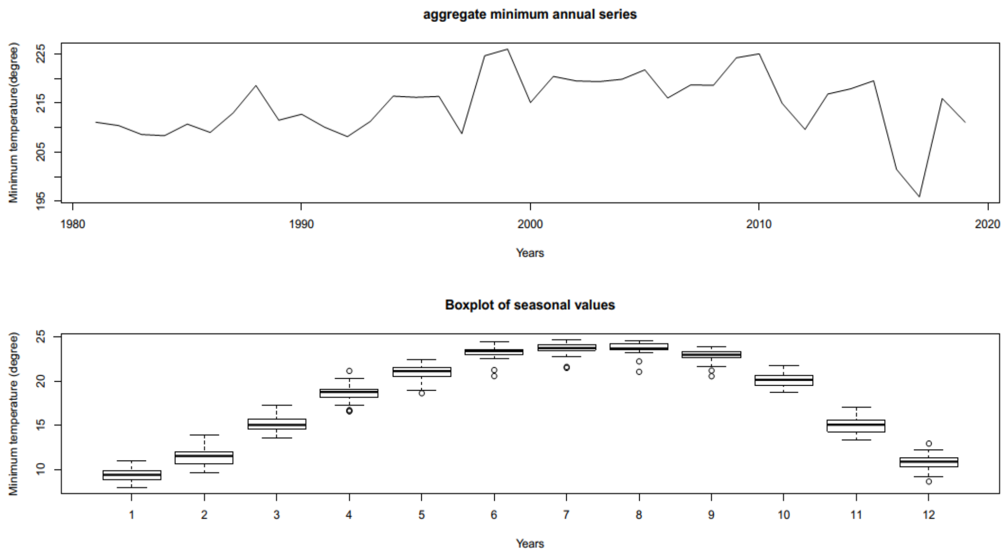


Fig. 4.7: Boxplot of the minimum temperature data

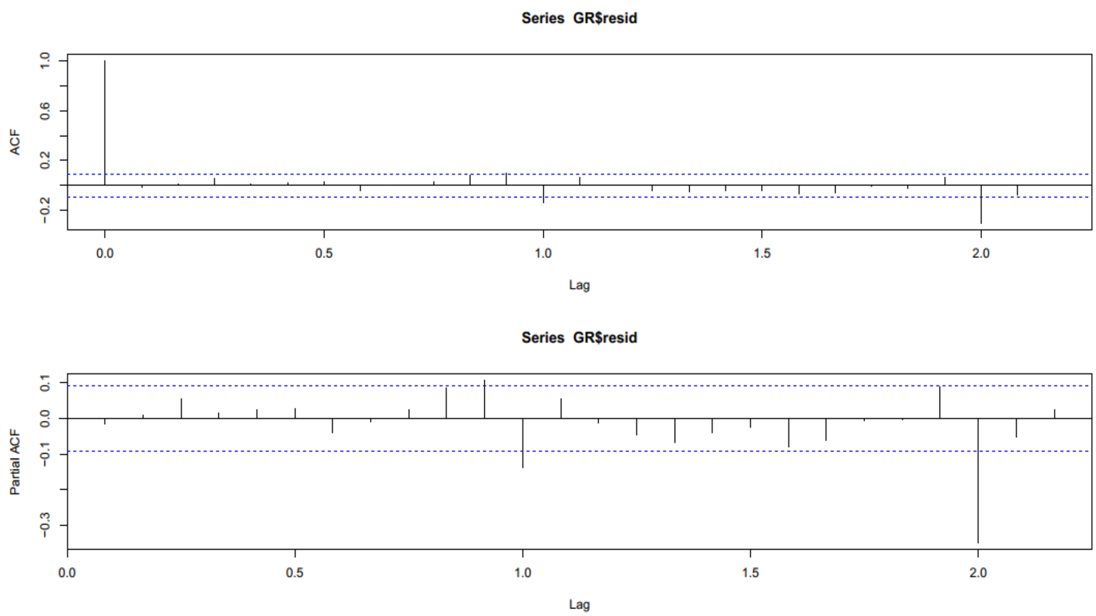


Fig.4.8: ACF and PACF of the minimum temperature series.

4.3 Analysis using Mann-Kendall Test

Temporal changes in the annual and seasonal values were analyzed by Mann-Kendall rank statistics to confirm the significance of the observed trend. Table 4.1 indicates that the Mann-Kendall test for average minimum temperature is increasing and average maximum temperature is decreasing in winter but not statistically significant. The trend of T_{min} is significant in March, April and June to September (monsoon season). The monsoon season, depicts a significant increase in the mean temperature. This indicates that the night temperatures during recent years have gone up during monsoon in Guwahati. The post monsoon mean temperature also shows an increasing trend. It can be inferred that in Guwahati city, daytime temperature in post monsoon season is significantly increasing. In majority of the months the minimum and maximum temperatures show significant trend. The beginning of winter shows increasing trend in minimum temperature, statistically significant at 5 percent level of significance. The later part of winter though shows an increasing trend that is not statistically significant. The increasing trend of maximum temperature is seen during the monsoon months and is statistically significant. However, the month of June shows a decreasing trend.

Table 4.1: Mann-Kendall rank statistics.

Monthly τ value for Mann-Kendall Test				
Month	**T$_{max}$(τ)**	**p-value**	**T$_{min}$(τ)**	**p-value**
January	0.016	0.90	-0.025	0.84
February	0.203	**0.08**	0.139	0.23
March	0.186	0.11	0.213	**0.06**
April	0.023	0.85	0.206	**0.08**
May	0.059	0.61	0.151	0.19
June	-0.017	0.89	0.196	**0.09**
July	0.271	**0.02**	0.353	**0.00**
August	0.089	0.45	0.359	**0.00**
September	0.308	**0.01**	0.311	**0.01**
October	0.092	0.43	0.011	0.94
November	0.068	0.56	0.050	0.70
December	0.042	0.72	0.170	0.14
Seasonal τ value for Mann-Kendall Test				
Season	**T$_{max}$(τ)**	**p-value**	**T$_{min}$(τ)**	**p-value**
Winter	0.201	**0.08**	0.302	**0.01**
Summer	0.293	**0.01**	0.338	**0.00**
Monsoon	0.129	0.26	0.019	0.87
Post Monsoon	0.174	0.13	0.133	0.25

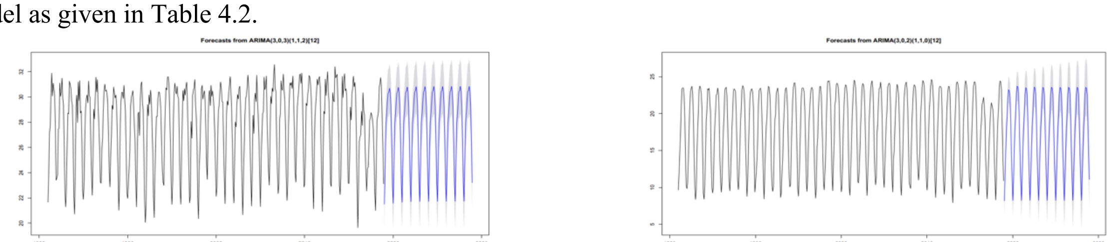
4.4 Analysis using Shapiro-Wilk test and Box-Pierce test

Based on the automatic ARIMA forecasting, our selected models are ARIMA $(3,0,3)(1,1,2)^{12}$, $(3,0,2)(1,1,0)^{12}$ for the maximum and minimum temperature series in Guwahati city which are adequate to represent the temperature data which could be used to forecast the upcoming temperature data. Table 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 4.2: Mean temperature characteristics.

Temperature	**Model**	**Box-Pierce test**	**Shapiro-Wilk normality test**	**AIC**
Maximum	$(3,0,3)(1,1,2)^{[12]}$	$\chi^2 = 0.00017341$ df = 1, p-value = 0.9895	W = 0.99129, p-value = 0.0088	1179.46
Minimum	$(3,0,2)(1,1,0)^{[12]}$	$\chi^2 = 0.0026589$ df = 1, p-value =0.9589	W = 0.98386, p-value = 4.474e-05	1134.26

The values of the Box-Pierce 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 is also depicted in the table for selection of the model as given in Table 4.2.



Forecasted average monthly maximum temperature   Forecasted average monthly minimum temperature

Fig 4.9: Forecasted average monthly temperature.

Forecasting or predicting future is one of the important reasons to develop a time series model. Thus we are using the proposed model for prediction purpose and based on the proposed SARIMA models. Figure 4.9 displays the forecasted average monthly maximum and minimum temperature along with 95% forecast limits for upcoming ten years. The proposed procedure of forecasting indicates that average maximum and minimum temperature is slightly increasing over the time. The predicted average maximum and minimum temperature shows very good approximation for the available data on temperature. Thus we can suggest SARIMA model is a proper model for forecasting temperature or any environmental phenomenon.

5 Conclusions

The present study analyzed 38 years data of temperature taken from Guwahati over a period of 1981-2018. The study revealed that the mean maximum temperature as well as the mean minimum temperature is showing a steady increase with time. High mean maximum temperature has been observed in summer and monsoon season while, low mean minimum temperature has been observed in winter season which is obvious. The results obtained from Mann-Kendall rank statistics test reveal that the variation of maximum temperature during the monsoon months is statistically significant. The night temperatures during recent years have gone up during monsoon in Guwahati and the post monsoon mean temperature also shows an increasing trend. Box-Pierce methodology, the identified ARIMA model must be diagnostically checked for its appropriateness by looking at the ACF and PACF of the model residuals. These are shown for the monthly maximum temperatures at Guwahati station. As the spikes at the different lags in the ACF and PACF plots in the figure are within the statistical confidence bands. Similarly good results have been obtained for the other time series analyzed.

Study of climate change found that Earth's climate is always changing. Climatic variability, including changes in the frequency of extreme events has always had a large impact on humans. The study of past climate change helps us understand how humans influence the Earth's climate system. Since temperature is one of the most indicative factors of climate change, then the study of temperature gives us an overall idea how it's affecting the climate of an area. Mann-Kendall test is used in the study for comparing trend pattern of several regions.Other method of trend determination may also be used. The L-moments and L-moment ratios are useful for summarizing statistical properties of hydrological data. This method facilitates distribution parameters estimation process and selection of best fit distribution. Further study can be done by taking the temperature data to identify the distribution which best fits the rainfall pattern. Suitable model can be developed for studying the temperature trend of Guwahati, India. The selected ARIMA $(3,0,3)(1,1,2)^{12}$ for maximum mean temperature and ARIMA $(3,0,2)(1,1,0)^{12}$give us ten years predicted monthly temperature along with 5% significant level that can help decision makers to predict for Guwahati city. The both maximum and minimum forecasted temperature data came in a very good agreement with the recorded previous data.

Acknowledgement: Authors would like to thank the editor and the referees for their support and assistance to improve the manuscript.

References

[1] D. Machiwal and M.K.Jha, Time series analysis of hydrologic data for water resources planning and management: A Review. *J. Hydrol. Hydromech*, **54(3)**, 237-257, (2006).

[2] S.A. Shamsnia, N. Shahidi, A. Liaghat, A. Sarraf and S.F. Vahdat, *Modeling of weather parameters (temperature, rainfall and humidity) using stochastic methods*. International Conference on Environment and Industrial Innovation, IPCBEE, Singapore., 282-285, (2011).

[3] S.B.Cheema, G.Rasul, G. Ali and D.H. Kazmi, A comparison of minimum temperature trends with model projections. *Pakistan Journal of Meteorology*, **8(15)**, (2011).

[4] C.Serra, A. Burguenoand X. Lana, Analysis of maximum and minimum daily temperatures recorded at Fabra Observatory (Barcelona, NE Spain) in the period 1917-1998. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **21(5)**, 617-636, (2001).

[5] E.S.Chung, K. Park, and K.S. Lee, The relative impacts of climate change and urbanization on the hydrological response of a Korean urban watershed. *Hydrological Processes*, **25(4)**, 544-560, (2011).

[6] E.B.Audu, An analytical view of temperature in Lokoja, Kogi State, Nigeria. *International Journal of Science and Technology*, **2(12)**, 856-859,(2012).

[7] Sindhu P.Menon, R. Bharadwaj, P. Shetty, P. SanuandS. Nagendra, *Prediction of temperature using linear regression*. "International Conference on Electrical, Electronic Communication, Computer and Optimization Techniques (ICEECCOT)", (2017).

[8] S.B. Cheema, G. Rasul, G. Ali and D.H. Kazmi, A comparison of minimum temperature trends with model

projections. Pakistan Journal of Meteorology,**8**, 39-52, (2011).
[9] G.E.P. Box and G.M. Jenkins, *Time series analysis: forecasting and control, (revised edition)*Holden&Day. San Francisco, (1976).
[10] C. Chatfield, *Time-series forecasting*, Chapman and Hall/CRC, (2000).
[11] R.J. Hyndman and K. Yeasmin, Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, **27(3)**, 1-22, (2008).
[12] A.Mondal, S.Kundu, and A. Mukhopadhyay, Rainfall trend analysis by Mann-Kendall test: A case study of north-eastern part of Cuttack district, Orissa. *International Journal of Geology, Earth and Environmental Sciences*, **2(1)**, 70-78, (2012).
[13] L.S.Hingane, K.R. Kumar, and B.V.R. Murty, Long-term trends of surface air temperature in India. *Journal of Climatology*, **5(5)**, 521-528, (1985).
[14] T.D. Roy, and K.K. Das, Temperature trends at four stations of Assam during the period 1981-2010. *Int. J. Sci. Res. Publ*, **3(6)**, 1-3, (2013).
[15] S.I.Ahmed, R. Rudra, T. Dickinsonand M. Ahmed, Trend and periodicity of temperature time series in Ontario. *American Journal of Climate Change*, **3(03**), 272-288, (2014).
[16] T.D. Roy, and K.K. Das,Time series analysis of dibrugarh air temperature.*Journal of Atmospheric and Earth Environment*, **1(1)**, 30-34, (2012).
[17] M.G.Kendall, *Rank Correlation Methods*, 4th edition. Charles Griffin, London, U.K., (1975).

Utpal Dhar Das is presently working as research scholar in the Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi, India. He is a bright fellow in Mathematics and Statistics and was awarded gold medal in M. Sc. (Statistics). He has two papers in his credit.

Brijesh P. Singh is currently working as Associate Professor in the Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi, India. He has obtained Ph. D. degree in Statistics form Banaras Hindu University, Varanasi and has more than 16 years' experience of teaching and research in the area of Statistical Demography. He has published 120 research papers in the refereed journals and books of national and international repute. He edited 2 books containing research papers and presently supervising four students for their research work. His research interests are in statistical modelling and analysis of demographic data specially fertility, mortality, reproductive health and domestic violence with its reason and consequences.

Tanusree Deb Roy is Assistant Professor in the Department of Statistics, Assam University, Silchar, Assam. She did her M. Sc. and Ph. D. in Statistics from Gauhati University, Guwahati. Her area of interest of research is distribution theory and time series modelling. She published more than 10 papers in the reputed journal and completed a DST, India funded project on time series modelling.