Environment and Ecology 41 (2B) : 1162—1170, April– June 2023
ISSN 0970-0420

Prediction and Projections of Temperature in Western Haryana through ARIMA Model

Mohit Kamboj, Pankaj Dahiya, Pooja Yadav,
Ekta Pathak Mishra, Raj Singh

Received 18 December 2022, Accepted 22 April 2023, Published on 21 June 2023

ABSTRACT

ARIMA model and prediction of temperature using various techniques with the help of temporal techniques. The ARIMA model was developed for the prediction of temperature patterns using SPSS software. The analysis was carried out for diagnostic, calibration, and predicting of temperature for Western Haryana (Hisar and Bhiwani). This study indicates that the best stochastic time series model for temperature analysis in Hisar and Bhiwani is ARIMA (0,01) and (0,1,1) that predicted values are well-fitted through the original data with the lower and upper limits containing majorities of the temperature original data. Forecasting accuracy of ARIMA model using R^2 value for temperature in Hisar region (0.97) and in Bhiwani region (0.99). The ARIMA models are therefore adequate to be used for predicting monthly temperature in Hisar and Bhiwani. Forecasting of temperature helps in the planning and decision-making process and it gives an insight of the future uncertainty using the past and current scenarios of weather parameter.

Keywords ARIMA model, Temperature, Hisar, Bhiwani, Haryana.

INTRODUCTION

On a worldwide level, numerous attempts have been made to predict the behavioral pattern using suitable techniques. In the last few decades, time series forecasting has received tremendous interest by researchers. Conventionally, the researchers have employed traditional methods of time series analysis, modeling, and forecasting, e.g., Box-Jenkins methods of autoregressive (AR), auto-regressive moving average (ARMA), autoregressive integrated moving average (ARIMA), autoregressive moving average with exogenous inputs (ARMAX). The conventional time series modeling methods have served the scientific community for a long time; though, they provide only reasonable accuracy and suffer from the stationary and linear assumptions (Patel *et al.* 2014).

Climate change seems to be one of the most important issues in the recent two decades and temperature has been identified as one of the key ele-

Mohit Kamboj[1]*, Pankaj Dahiya[2], Raj Singh[5]
[1,2]PhD Scholar, [5]Principal Scientist
Department of Agricultural Meteorology, Chaudhary Charan Singh Haryana Agricultural University, Hisar, Haryana 125001, India

Pooja Yadav[3]
[3]PhD Scholar
Department of Agrometeorology, Govind Ballabh Pant University of Agriculture and Technology, Pantnagar, Udham Singh Nagar, Uttarakhand 263145, India

Ekta Pathak Mishra[4]
[4]Assistants Scientist
College of Forestry, Sam Higginbottom University of Agriculture, Technology and Sciences, Allahabad 211007, UP, India

Email: mohitkamboj@hau.ac.in
*Corresponding author

ments that can indicate climate change (Tanusree and Kishore 2012). It is widely believed that the changing temperature due to global warming is permanently changing the entire Earth's climate. People perceive the impacts of global warming differently with some taking the necessary precautions to help reduce the rates of the rising temperature. In the past century alone, studies have shown that the globe's mean temperature has risen between 0.4°C and 0.8 °C. According to a study by IPCC (2007), the temperatures could rise between 1.4°C and 5.8 °C by the end of the 21st century. This increase in temperature may seem to be minute but the impacts are great. Increases in temperatures are likely to lead to a global increase in drought conditions, decreased water supplies due to evapotranspiration and an increase in urban and agricultural demand.

Generally, a time series {x (t), t = 0,1, 2,...} is assumed to follow a certain probability model which describes the joint distribution of the random variable xt. The mathematical expression describing the probability structure of a time series is termed a stochastic process (Adhikari and Agrawal 2013). They differentiate the series, so the seasonal differenced series removes the trend effect and the series becomes stationary. The SARIMA (Seasonal Autoregressive Integrated Moving Average) models for sub-series also show the monthly average temperature has a stable structure. The forecasting results show the SARIMA models fit the data well (Li and Moller 2009).

MATERIALS AND METHODS

Temperature data of 31 years were used for developing ARIMA models. Data were used for diagnostic, calibration, and forecasting (1985-2016) of average monthly temperature. The temperature data was collected from Climatic Research Unit (CRU), University of East Anglia (https://www.uea.ac.uk) on a grid basis.

The statistical software used for analysis is SPSS (Statistical Package for the Social Sciences). ARIMA (Auto-Regressive Integrated Moving Average) model is a type of statistical model that can be used to analyze and forecast time series data. This model was developed by using SPSS software adopting procedure that are discussed above.

Location

Haryana is a landlocked Indian state in the north. Between 27°39' and 30°35' N latitude, and 74°28' and 77°36' E longitude, it is located. Haryana's elevation ranges from 700 to 3600 feet (200 to 1200 meters) above sea level. Hisar is located in western Haryana at 29.09°N 75.43°E. It is located at an average elevation of 215 meters (705 feet) above sea level. Bhiwani has an average elevation of 225 meters and is located at 28.78°N 76.13°E.

Climate

Haryana is hot in the summer and cool in the winter with an average rainfall of 354.5 mm, the climate is arid to semi-arid. The climate of Hisar is characterized by dryness, temperature fluctuations, and a lack of rainfall. During the summer, the highest daytime temperature is between 40 and 46 °C. The average annual maximum and minimum temperatures are 32.3 °C and 15.4 °C, respectively. The relative humidity ranges from 5% to 100%. The average annual rainfall is roughly 429 mm, with July and August being the wettest months. The local steppe climate has an impact on Bhiwani. The temperature in the Bhiwani District varies from 2 °C to 45°C. The average temperature in Bhiwani is 25.2 °C.

Model used

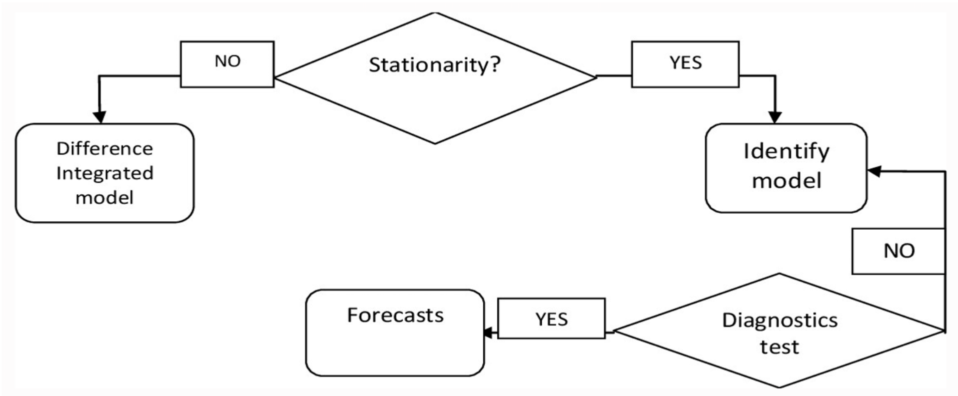
The Box-Jenkins method

This study follows the Box-Jenkins methodology for modelling, steps involved:

a) Model identification- Box and Jenkins 1976.
b) Estimation of the model parameters- Chatfield 2004, Box and Jenkins 1976.
c) Diagnostic checking- Anderson 1977.
d) Forecasting.

Conceptual Framework

It is a scheme of concepts that the researcher operationalizes in order to achieve the set objectives. The following conceptual framework proposed is considered in this study.



Model types

The autoregressive models

An autoregressive model, AR (p), is a model in which a linear combination of previous measurements of the variable and a random error term with a constant term are used to forecasts.

Moving average models

A MA (q) model uses past errors to predict the variable of interest. The general algebraic representation of a moving average of order q, MA (q).

The residuals are assumed to follow a normal distribution. Thus a MA model is a linear regression of the current values of the time series against the residuals of one or more prior observations (Adhikari and Agrawal 2013).

Autoregressive moving average models

The AR model includes the lagged terms on the time series itself while the MA model includes lagged terms on the noise or residuals. If the AR and MA models are effectively combined together, we form the ARMA model. Thus ARMA (p, q), where p is the autoregressive order and q the moving average order. It is important to note that the ARMA models can only be used when time series data is stationary.

Autoregressive inntegrated moving average models

In practice, many time series are always non-stationary but ARMA models are therefore inadequate to effectively describe non-stationary time series which are more frequently encountered in actual practice. Box and Jenkins (1976) proposed the ARIMA model which is a generalization of an ARMA model to include the case of non- stationarity. When using the ARIMA model, finite differencing is applied to the data to remove non- stationarity.

The model is referred to as an ARIMA (p, d, q) and is represented algebraically as:

$$\phi(B)\ (1-B)\ ^{d}X_t=\theta\ (B)\varepsilon_t,\ \varepsilon_t \sim WN\ (0,\ s^2)$$

Where:

i. *WN* stands for white noise.
ii. p represents non-seasonal AR order, d represents non-seasonal differencing, and q represents non-seasonal MA order.

Generally, d=1 is enough in most cases. If d=0, the model reduces to an ARMA (p, q) model.

In case a series has both seasonal and non-seasonal behaviors then the ARIMA model may mislead to the selection of a wrong order for non-seasonal components because it may not be able to capture the behavior along the seasonal part of the series.

Where:

p represents non-seasonal AR order,
P represents seasonal AR order,

d represents non-seasonal differencing,
D represents seasonal differencing,
q represents non-seasonal MA order,
Q represents seasonal MA order,
S represents seasonal order (for monthly data S = 12)
B is the backward shift operator (Bkyt = y t -k),
yt represents time series data at period t, and
εt is the random shock (white noise error).

In this model, non-stationarity is removed from the series using the appropriate order of seasonal differencing. A first-order seasonal difference is a difference between an observation and the corresponding observation from the previous year.

Non-seasonal differencing is also necessary if the trend is present in the data. Often a first non-seasonal difference will "detrend" the data (Abdul-Aziz *et al.* 2013).

Model identification

Stationarity analysis

Stationarity is achieved when a time series has a constant mean, variance, and autocorrelation over time. Stationarity is a necessary and sufficient condition for ARIMA models before performing any analysis. Plotting the series and its autocorrelation is the standard way to check for non-stationarity. The time series graph can be examined through time to determine whether it has any trend or variability over time. For a non-stationary series, the autocorrelation function decays slowly.

For a series characterized by trend, seasonality, or any other non-stationary patterns, we analyze the series after differencing. In order to obtain stationary data from a first-order non-stationarity, we first sieve the observations with ARIMA models by differencing them d times, Using Δdyt instead of yt as the time series.

This is normally done with the transformation

$$\Delta y t = y t y t -1 \quad (1)$$

For the non-seasonal part the above equation results to the values 0,1,2... d = and for the seasonal part D = 0,1,2....

Sometimes a series might need differencing more than once or to be differenced at lags which are greater than one period. In case of a second order non-stationarity, a simple transformation like the log transformation could be helpful.

Autocorrelation and Partial Autocorrelation Functions (ACF and PACF)

When using the ARIMA models, model specification and selection is a crucial step of the analysis process. A proper model for the series is identified by analyzing the ACF and PACF.

They reflect how the observations in a time series are related to each other. It is useful that the ACF and PACF are plotted against consecutive time lags for the purposes of modelling and forecasting. The order of the AR and MA are determined by these plots.

For a time series, the auto-covariance function ACVF at lag k is defined as:

$$c_k = 1/n \, \Sigma_{t=1}^{n-k} (x_t - \mu)(x_{t-k} - \mu) \quad (1)$$

If xt is a stationary process with mean μ, the autocorrelation of order k is simply the relation between xt and xt-k. The ACF estimate for the sample at lag k is thus defined as

$$p_k = \frac{E\{(x_t - \mu)\,(x_{t-1} - \mu)\}}{E\,\{(x_t - \mu)^2\}} \quad (2)$$

The PACF of a stationary process, xt, denoted ϕ_{hh} is

$$\phi 11 = corr\,(x_t+1, x_t) = \mathrm{p}(1) \quad (3)$$

$$\text{And } \phi_{hh} = corr\,(x_{t+h} - \hat{x}_{t+h}, \hat{x}_t - \hat{x}_t) \ \mathrm{h} \geq 2 \quad (4)$$

$x_{t+h} - \hat{x}_{t+h}$ and $\hat{x}_t - x_t$ are not correlated with $\{x_{t+1}.....$
$x_{t+1-h}\}$-

The ACF and PACF plots are used to identify the terms of the SARIMA model. The non-seasonal terms are identified from the early lags 1, 2, 3, …. Non-seasonal MA terms are indicated by spikes in the ACF at low lags while non-seasonal AR terms are indicated by spikes in the PACF at low lags. The seasonal terms are examined from the patterns across lags that are multiples of S.

For monthly data, we look at lags 12, 24, 36 and so on (probably won't need to look at much more than the first two or three seasonal multiples). The ACF and PACF are judged at the seasonal lags in the same way it is done for the earlier lags.

Function or partial autocorrelation function is cutting off or tailing off (Shumway and Stoffer 2006). Models that look different can also be very similar. Precision should therefore not be a major concern at this stage of model fitting.

Forecasting

The selected model does not always necessarily provide the best forecasting therefore it is important to apply other tests such as MAE, MSE and MAPE to confirm the forecasting accuracy of the model.

Forecasting an ARMA process with mean μ_x , m-step-ahead forecasts can be defined as

$$\tilde{X}_{n+m} = \mu_x + \sum_{j=m}^{\infty} \Psi_j \, w_{n+m-j} \qquad (1)$$

The precision of the forecast is assessed with a prediction interval of the form

$$X^{n}_{n+m} \pm C_{\frac{a}{2}} \sqrt{P^{m}_{n+m}} \qquad (2)$$

Where $C_{\frac{a}{2}}$FF is identified such that the desired degree of confidence is achieved. Suppose it is Gaussian process, then having $C_{\frac{a}{2}} = 2$ will yields approximately 95% prediction interval for X_{n+m}.

Mean absolute error

$$MAE = ½ \sum_{t=1}^{n} |e_t| \qquad (3)$$

Where et= Xt - Ft is the error term
Xt is the actual observation for time period t, Ft is the forecast value for period t and n is the number of forecasting values (Spyros *et al.* 1998).

Mean square error

$$MSE = ½ \sum_{t=1}^{n} e_t^2 \qquad (4)$$

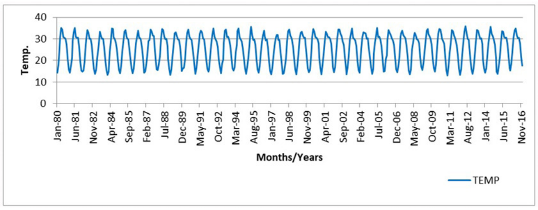


Fig. 1. Time plots of temperature changes for Hisar.

Where et= Xt - Ft is the error term and Xt is the actual observation for time period t, Ft is the forecast value for period t and n is the number of forecasting values (Spyros *et al.* 1998).

Mean absolutes percent error

$$MSE = ½ \sum_{t=1}^{n} |PE_t| \qquad (5)$$

The mean absolute percentage error is the mean or average of the sum of all the percentage errors for a given data set taken without regard to sign (That is, their absolute values are summed and the average computed). It is one measure of accuracy commonly used in quantitative methods of forecasting (Spyros *et al.* 1998).

RESULTS AND DISCUSSION

The statistical parameters of input data were estimated such as, the maximum, minimum, mean, standard error of data sets and stationarity of data series was examined using ljung-box test analysis. The stationarity of data series was also checked through examining the time series plot. Stationary means the data fluctuate around a constant mean. The various plots of temperature viz., Observed temperature, fitted temperature, and predicted temperatures are shown in Figs. 1– 6) for Hisar and Bhiwani (1985-2025).

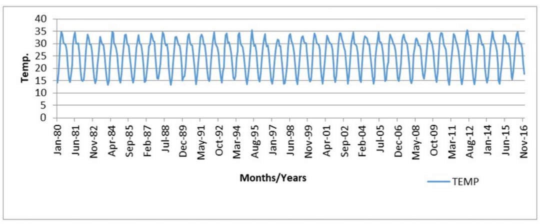


Fig. 2. Time plots of temperature changes for Bhiwani.

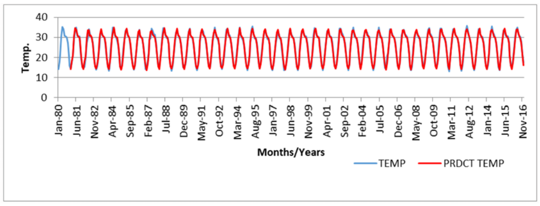


Fig. 3. Observed and fitted values of temperature series for Hisar..

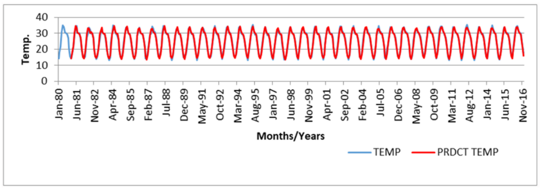


Fig. 4. Observed and fitted values of temperature series for Bhiwani.

Monthly temperature analysis

Summary of temperature are shown for Hisar as well as for Bhiwani from observed data of Hisar and Bhiwani (Figs.1–2). The maximum highest temperature was recorded in May 2013 while the lowest temperature was recorded in June 1989. The lowest temperature was recorded in January 1990 while the highest minimum temperature was recorded in January 2006. The month of May and June recorded the highest temperature was recorded in November and December. The temperature was recorded in December and January while the highest temperature was recorded in March and April. Temperature seem to be unstable throughout the year. However, from April temperature drop significantly till July and start rising again till October. The lowest temperatures are experienced in November with December being the coldest month of the year. High temperatures are experienced in the month of May and June.

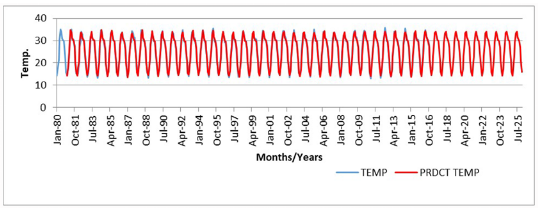


Fig. 5. Observed and forecast values of temperature series for Hisar.

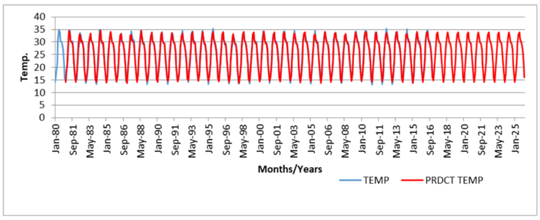


Fig. 6. Observed and forecast values of temperature series for Bhiwani.

Model building for monthly series

According to Takele (2012), the process of model fitting involves data plotting, data transformation if necessary and identification of dependence order, estimation of parameter, diagnostic analysis and choosing appropriate model. In this section, an univariate ARIMA methodology is used to model monthly temperature of Hisar and Bhiwani.

Model identification

ACF and PACF plots are used in the identification of the values p, q, P and Q. For the non-seasonal part, spikes of the ACF at low lags are used to identify the value of q while the value of p is identified by observing the spikes at low lags of the PACF. For the seasonal part the value of Q is observed from the ACF at lags that are multiples of S while for P, the PACF is observed at lags that are multiples of S. Looking at the ACF plots and PACF plots for differenced time series, the following models are suggested;

Best fitted models for temperature of Hisar and Bhiwani

Hisar- ARIMA (0,0,1) (0,1,1)
Bhiwani- ARIMA (1,0,0) (0,1,1)

Diagnostic aSnalysis

For a well fitted model, the standardized residuals estimated from the model should behave as an independent identically distributed sequence with zero

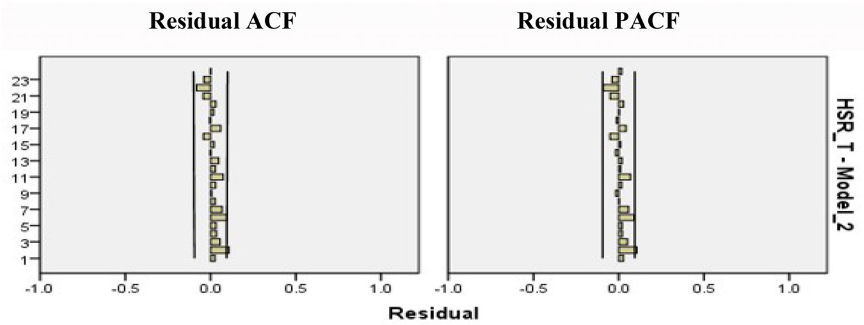


Fig. 7. Residual ACF and PACF temperature model of Hisar.

mean and constant variance.

Diagnostic analysis for temperature model

From the time plot of the residuals against time, we can see that there is no obvious pattern in the plot except for a possible outlier and it looks like an independently and identically distributed sequence of mean zero with a constant variance (Figs. 7–8). The plots of the ACF of the residuals lack enough evidence of significant spikes which clearly shows that the residuals are white noise.

Model validation

In order to test the adequacy and predictive ability of the chosen models, the actual data sets, predicted values, lower and upper limits are plotted and displayed in Fig. (3 and 4). The graphs show that the predicted values are well-fitted through the original data with the lower and upper limits containing majorities of the original data. This indicates that the models chosen for temperature series are the best fitted ones for the data sets.

Forecasting

Forecasting helps in planning and decision making process since it gives an insight of the future uncertainty using the past and current behavior of given observations. From most research studies, the selected model is not always the best for forecasting. Further accuracy tests such as MAE, MAPE and RMSE must therefore be carried out on the model. The (Table 1) shows a summary of ME, RMSE and MAE for temperature models.

The data depicted (Table1) that about the forecasting accuracy by calculating the mean value the ARIMA (0,0,1) (0,1,1) and ARIMA (1,0,0) (0,1,1)

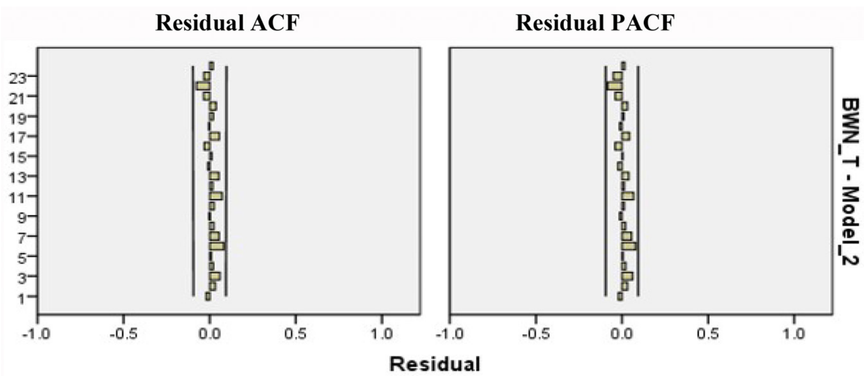


Fig. 8. Residual ACF and PACF temperature model of Bhiwani.

Table 1. Forecasting accuracy statistics.

Parameters	Temp model (Hisar) ARIMA (0,0,1)(0,1,1)	Temp model (Bhiwani) ARIMA (1,0,0)(0,1,1)
Stationary		
R-squared	0.438	0.446
R-squared	0.978	0.997
RMSE	1.014	0.998
MAPE	3.248	3.242
MAE	0.770	0.764

Stationary R-squared (0.438) not much difference. However, ARIMA (0,0,1) (0,1,1) and ARIMA (1,0,0) (0,1,1) RMSE was found large 1.014 in Hisar district. In case of, MAPE calculation showed that accuracy of forecasting model possessed not much difference i.e., (3.248) in comparison of (3.242). While, the calculation in ARIMA (0,0,1) (0,1,1) and ARIMA (1,0,0) (0,1,1) MAE secured higher mean value (0.770) whereas, in Bhiwani secured mean (0.764). It is indicated from the (Table 1) in each parameters of forecasting accuracy were not found highly differentiated in both ARIMA models.

CONCLUSION

ARIMA model offers a good technique for predicting the magnitude of any variable. Its strength lies in the fact that the method is suitable for any time series with any pattern of change and it does not require the forecaster to choose a priori the value of any parameter. It can be successfully used for forecasting long time series data. We found that the performance of ARIMA model is evaluated by forecasting the data from 1985-2016 both the forecasted and observed monthly temperature data fitted on the same plot and indicate the model accuracy/adequacy for performance purpose. The similarity in matching between the forecasted and observed temperature were good. While the ARIMA model catches the correct trend overall and predicts the monthly temperature with accuracy. The analysis revealed that the best model for Hisar temperature is ARIMA (0,0,1) (0,1,1) and for Bhiwani is ARIMA (1,0,0) (0,1,1). The model residuals for both series are near normality as most points fall on the straight line with a few close to it. The residuals are also confirmed to be white noise. From the model validation results, the predicted values are very much well-fitted through the original data with the lower and upper limits containing majorities of the temperature original data.The identified models are therefore adequate to be used for forecasting monthly temperature in Hisar and Bhiwani. According to the similar study it was found that fluctuation of temperature is marked more in case of minimum temperature but slightly steady in case of maximum temperature. This is the signature of increasing global warming in terms of temperature. The fitted forecasted model can be used to generate the forecast value of temperature in others area and the comparison among them would provide better idea about the changing pattern of temperature specifically and would give an idea about the local changing pattern of climate. This is also suggested that this seasonal model can also be used in any other areas of time dependent data with necessary modifications (Sultana and Hasan 2015). Moreover, study also indicated that the Box-Jenkins ARIMA method has proved to be a useful technique which can help decision makers to establish better strategies as well as to set up priorities for equipping themselves against upcoming weather changes (Uchechukwu *et al.* 2014).

ACKNOWLEDGMENT

We are grateful to the HOD and all the faculty members of the Agricultural Meteorology Department, SHUATS, Prayagraj for their kind support, knowledge and direction related to research problem.

REFERENCES

Abdul-Aziz AR, Anokye M, Kwame A, Munyakazi L, Nsowah-Nu amah NNN (2013) Modeling and forecasting rainfall pattern in Ghana as a seasonal ARIMA process: The case of Ashanti region.

Adhikari R, Agrawal RK (2013) An introductory study on time series modeling and for ecasting.arXiv preprint arXiv: 1302-6613.

Anderson OD (1977) Time series AIC procedure. analysis and forecasting: Another look at the Box-Jenkins, pp 288.

Box GEP, Jenkins GM (1976) Time series analysis: forecasting and control. Prentince Hall, Inc, pp 575.

Li X, Moller M (2009) Applying GLM model and ARIMA model to the analysis of monthly temperature of Stockholm. D-level Essay in Statistics in Spring, pp 1-24.

Patel DP, Patel MM, Patel DR (2014) Implementation of ARIMA model to predict rain attenuation for KU-band 12 Ghz

Frequency. *J Electron Commun Engg* 9 : 83-87.
Shumway RH, Stoffer DS (2006) Time series analysis and its applications with R examples 2nd edn. Springer Science +Business Media, LLC, NY, USA, pp 1-19.
Sultana N, Hasan MM (2015) Forecasting temperature in the Coastal area of Bay of Bengal-an application of Box-Jenkins seasonal ARIMA model. *Civ Environ Res* 7 : 241-249.
Spyros M, Wheelwright SC, Hyndman RJ (1998) Forecasting: Methods and applications. 3rd edn, USA, John Wiley and Sons, Inc. 1-334.
Tanusree DR, Kishore DK (2012) Time series analysis of Dibrugarh air temperature. *J Atmospheric Earth Environ* 1 : 30-34.
Takele (2012) "Statistical analysis of rainfall pattern in Dire Dawa", Eastern Ethiopia (Doctoral dissertation, MSc thesis, submitted to the Department of Statistics, Addis Ababa University).
Uchechukwu G, Chisimkwuo J, Okezie C (2014) Time Series Analysis and Forecasting of Monthly Maximum Temperatures in South Eastern Nigeria. *Int J Innov Res Dev* 3 : 165-171.