TEMPERATURE FORECAST CONSIDERING THE IMPACT OF RAINFALL: A CASE STUDY ON BETWA RIVER BASIN AT BAIRAGARH STATION (BHOPAL)

A DISSERTATION

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE

OF

MASTER OF TECHNOLOGY

IN

HYDRAULICS & WATER RESOURCE ENGINERING

SUBMITTED BY

REEMA KASERA

2K18/HFE/10

Under The Supervision Of

Prof. VIJAY K. MINOCHA

CIVIL ENGINEERING DEPARTMENT

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Bawana Road, Delhi – 110042

AUG 2020

DEPARTMENT OF CIVIL ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

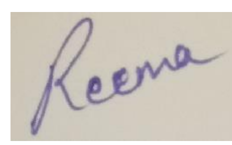
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, REEMA KASERA, Roll No. 2K18/HFE/10 student of M.Tech. (Water Resource Engineering), hereby declare that the project Dissertation titled "**TEMPERATURE FORECAST CONSIDERING THE IMPACT OF RAINFALL: A CASE STUDY ON BETWA RIVER BASIN AT BAIRAGARH STATION (BHOPAL)**" which is submitted by me to the Department of Civil Engineering Department, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associate ship, Fellowship or other similar title or recognition.

Place: Delhi

Date: 26-08-2020



(**REEMA KASERA)**

DEPARTMENT OF CIVIL ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the major project titled "**TEMPERATURE FORECAST CONSIDERING THE IMPACT OF RAINFALL: A CASE STUDY ON BETWA RIVER BASIN AT BAIRAGARH STATION (BHOPAL)**" which is submitted by REEMA KASERA, Roll No-2K18/HFE/10 CIVIL ENGINEERING DEPARTMENT, Delhi Technological University, Delhi, in partial fulfilment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the students under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree to this University or elsewhere.

.

Place: Delhi
Date: 26/08/2020

Prof. Vijay K Minocha
(PROJECT SUPERVISOR)

Professor

Department of Civil Engineering

Delhi Technological University

DEPARTMENT OF CIVIL ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

ACKNOWLEDGEMENT

I am highly grateful to the Department of Civil Engineering, Delhi Technological University (DTU) for providing this opportunity to carry out this project work.

The constant guidance and encouragement received from my supervisor **Prof. VIJAY K. MINOCHA** of Department of Civil Engineering, DTU, has been of great help and inspiration in carrying my present work and is acknowledged with reverential thanks.

Finally, I would like to expresses gratitude to all faculty members of Civil Engineering Department, DTU for their intellectual support in my M.tech study at DTU.

Place: Delhi

Date: 26-08-2020

REEMA KASERA

M.Tech (Hydraulics & Water Resource Engineering)

(2K18/HWE/10)

ABSTRACT

In climate impact studies, temperature forecasting has been considered as one of the most important factors on the sector of vegetation, irrigation, water resources and tourism. The main objective of this study is to forecast daily maximum and minimum temperature in Bairagarh station, Bhopal by employing auto-regressive integrated moving average (ARIMA) and the auto-regressive integrated moving average with exogenous variables (ARIMAX) models. This study compares the two models and provide the best-fit prediction with the observed actual data.

The daily maximum and minimum temperature observations between 1982 and 2012 were collected from NASA's POWER data access viewer. ARIMA model was applied to the daily maximum and minimum temperature series to have the best-fit model however, ARIMA can capture the effect of rainfall by itself but introducing rainfall as exogenous variable will improve the efficiency. That is why ARIMAX was considered in this study.

The Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and residual ACF serves as the error measures in evaluating the forecast ability of the models. The effect of (AIC) was tested. As compared to the ARIMA models, ARIMAX model performed well with lower error matrices, this effect was more significant in maximum temperature series. Which indicates that the rainfall factor was influential in the model. The results shows that the model with parameter greater than 1 and less than 3 would work better for Bairagarh station there is a significant correlation between rainfall and temperature, which is evident by the reduced error from ARIMAX modelling.

For maximum temperature, ARIMA models (1, 1, 1), (2, 1, 1) and (1, 1, 2) have performed well compared to the other models at training stage with RMSE value of 1.325144, 1.319782, and 1.320031 and the AIC values were 12423.88, 12396.3, and 12397.67 and MAPE value of 2.953642, 2.939284 and 2.940243 respectively. At testing stage, model (1, 1, 2) have shown best results with RMSE=1.379335, AIC=1273.74, MAPE= 3.1285561.

Whereas for minimum temperature, the ARIMA models (1, 1, 2), (2, 0, 2), (2, 1, 1), (3, 0, 0), and (3, 1, 0) at training stage with RMSE value of 1.287539, 1.286019, 1.287261, 1.303937, and 1.299155 and AIC values were 12215.74, 12216.6, 12214.16, 12315.47,and 12281.22 and MAPE value of 6.068753, 6.08717, 6.731284, 6.147745, and 6.090528 respectively. At validation stage

model (2, 1, 1) have shown better results among all, with AIC = 1236.45, RMSE = 1.310559, and MAPE = 6.731284.

For maximum temperature, ARIMAX models obtained for the considered station at training stage were (1, 1, 1), (2, 1, 1) and (1, 1, 2) have performed well compared to the other models with RMSE value of 1.3769, 1.3716, and 1.3714 and the AIC values were 1272.47, 1269.68, and 1269.57 respectively. At testing stage, model (1, 1, 2) have shown best results with RMSE=1.3714, AIC=1269.57, MAPE= 3.115147.

Whereas for minimum temperature, the ARIMAX models (1, 1, 2), (2, 0, 2), (2, 1, 1), (3, 0, 0), and (3, 1, 0) at training stage with RMSE value of 1.285462, 1.283925, 1.285178, 1.301806, and 1.297123 and AIC values were 12205.94, 12206.69, 12204.34, 12305.52,and 12271.79 and MAPE value of 6.054431, 6.071131, 6.053243, 6.134001, and 6.077045 respectively. At validation stage model (2, 1, 1) have shown better results among all, with AIC = 1236.45, RMSE = 1.310558, and MAPE = 6.730814.

CONTENT

TABLE OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS, ABBRIVIATIONS

K = number of estimated parameter

L= maximum value of likelihood function for model

X_t = is actual value of a point for a given time period t

n = total number of fitted points

$\hat{x}$ = fitted value for time period t.

B = backward shift operator

AIC = Akaike Information Criteria

RMSE = Root Mean Square Error

MAPE = Mean Absolute Percentage Error

SSR = Sum of Square of Residual

ACF = Autocorrelation Function

PACF = Partial Autocorrelation Function

ARIMA = Autoregressive Integrated Moving Average

ARIMAX= Autoregressive Integrated Moving Average with exogenous variabl

CHAPTER 1

INTRODUCTION

1.1. GENERAL

Prediction is a difficult art, especially when it includes the future"- Neils Bohr (Nobel Laureate Physicist)

"Forecasting is the process of making projections for the future events based on the past and present observed data. The art of forecasting is very important to analyze the risk that are likely to be happened and pre-prepared for the circumstances.

Climate is a long-term average of meteorological parameters (precipitation, wind, temperature and others) in a given region. A period of over 30 years is typically required to average the climate. Change in Climate will have a major impact on the socio-economic, and environment related sectors, including water resources, agriculture and food security, human health and forest diversity, and tourism. Therefore, it is required to forecast temperature accurately in order to intercept un-expected hazards caused by temperature variation.

Temperature and precipitation are the most important parameter of climate and variation in these variables can affect the economic growth, development and health of human. There is a direct impact of rise and fall in earth's temperature on evaporation, snow melting, frost and an indirect impact on stability of atmosphere and conditions of rainfall. The rise in temperature results in an increase in evaporation and cause cloud formation, which increases precipitation, whereas during rain event the temperature of that particular region decrease. It shows that temperature and precipitation are interconnected. Therefore, it becomes necessary to consider the effect of rainfall pattern in temperature forecasting.

Temperature change can have a remarkable impact on water resources by evolving change in the hydrological cycle. The rise in temperature increases the evaporation rate of water into the atmosphere and increase the atmosphere's holding capacity of water. Increased evaporation may dries out some regions and fall as excess rain on other region. Rising temperature can have a significant effect on agriculture sector also. It give rise to changes in crop seasons that affect food safety and changes in the spread of diseases that increase the risk for people. It also led to reduce access to food because it effect the changing patterns of precipitation in extreme events and reduce water availability may all result in reduction of agricultural productivity. Thus, it is necessary to forecast the temperature to deal with the uncertainty of precipitation, and evaporation.

Auto Regressive Integrated Moving Average model (ARIMA) and Auto Regressive Integrated Moving Average with Explanatory variable (ARIMAX) model are used in this study to model and forecast the maximum and minimum temperature time series of Bairagarh station in Bhopal. ARIMA is the most common time series model to develop seasonal forecast, to identify seasonality, trend analysis and forecasting in time series. The time series models used for forecasting requires historical sequence of observation of variables. These observations are statistically dependent and time series modelling is dealt with techniques to analyze such dependencies. In time series modelling through ARIMA, forecast of future event is based on the historical values of the variable. ARIMA is suitable for univariate dataset. Whereas ARIMAX is used for multivariate data set. It can easily establish the cause and effect relation between the variables and then forecast. In this study an explanatory variable is rainfall.

Modelling and forecast of minimum and maximum temperature time series is done by ARIMA model individually and then regression was done to analyze the relation between rainfall events and maximum and minimum temperature The error developed in ARIMA modelling is then minimized by ARIMAX model to have the best-fitted model

1.2. BACKGROUND OF STUDY

Thesis will concentrate on the use of technical research in forecasting future minimum and maximum temperature in Bairagarh station, Bhopal. The station was selected by convenience and availability of daily temperature and precipitation data

1.3. RESEARCH OBJECTIVES

The main objective of this study is to build a model to forecast daily maximum and minimum temperature time series

The main objectives of this study are as follow:

1).To compare the forecasting efficiency of ARIMA and ARIMAX model.

2).To investigate which forecasting model under consideration gives minimum forecasting error.

3).To have the best-fit forecasting model for Bairagarh station.

4).To analyze the significant effect of rainfall in temperature prediction.

5).To check the stationarity of time series of temperature for considered station.

1.4. STRUCTURE OF THESIS

Chapter I gives an overview of the methodology involved for quantifying the objectives and the relevance of this study.

Chapter II deals with the scientific rationale related to correlation of rainfall and temperature, use of ARIMA and ARIMAX models for meteorological forecasting.

Chapter III deals with the methodology (time series, ARIMA, and ARIMAX) involved and gives description about study area, Bairagarh station in Bhopal district, topographic information, rainfall temperature, about the data used in the study.

Chapter IV deals with the model formation using ARIMA and ARIMAX

Chapter V describes and discuss the results obtained from the two different methodologies used in the present study.

Chapter VI gives the conclusion of the present study resulting from two different methodologies and gives the need for future work.

CHAPTER 2

LITERATURE REVIEW

2.1. TEMPERATURE

Temperature is a vital meteorological parameter of climate after precipitation.

Lobell et al (2013) observed that increasing temperature and limited precipitation cause drought incidence as a result of global warming, are posing serious threats to food security.

Pirttioja et al (2015) stated that all the models of crop production are sensitive to climate and environmental variation. Hence, it becomes important to predict temperature by statistical models for optimal crop growth, development and yield.

Intergovernmental panel on climate change (IPCC) in 2007 stated that in past 100 years from 1906 to 2005 the temperature shown a rising tendency of about 0.74℃ of average temperature, and upward trend has been observed in seven sub regions of Asia. Many researchers have also found the rising tendency of temperature.

Reiter et al. (2012) observed the temperature rise during summer in Upper Danube basin of about 0.8℃ per decade and have constant rise during spring and winters.

Luterbacher et al., (2004) observed that the surface temperature in Europe has never been as high as is increasing in 21st century.

2.2. RAINFALL AND TEMPERATURE

Many researchers investigated the relation between rainfall and temperature.

Weining Zhao and M.A.K. Khalil (1992) investigated nearly 100 stations in contiguous United States and found the negative correlation between summer rainfall and temperature.

Tinyiko R. Nkuna, John O. Odiyo. (2016) investigated the relationship between Temperature and variation in rainfall in the Levubu sub-catchment, South Africa and found the positive cross correlation between rainfall and temperature for annual and monthly time scale.

Kalidoss Radhakrishnan et al., (2017) studied the trend analysis of temperature and rainfall in India and found the declining rainfall and rapid warming trend, in last 30 years.

Gabriel Gerard Rooney et al (2018) estimated the effect of rainfall on temperature of tropical lake and found that the day with heavy rainfall cause reduction in lake surface temperature by 0.3K. This shows that it is important to consider the impact of rainfall for temperature prediction.

2.3. TIME SERIES

Time series modelling has been beneficially used in many areas of study to explain, forecast and control processes. Many time series forecasting methods are based on analysis of historical data. They assume that past patterns in the data can be used to forecast future events. A time series represent a set of observations that measure the variation in time of some dimension of a phenomenon, such as, evaporation, precipitation, wind, humidity, river flow etc. various researchers used time series for modelling meteorological parameters.

Sinha Ray et al (1997) used time series for modelling meteorological parameters and suggest its applicability for medium range weather forecast.

Sudipta Sarkar (2004) shown the significant impact of monsoon precipitation and land surface temperature over Indian sub-continental vegetation distribution.

2.4. ARIMA

The traditional Box-jenkins's ARIMA model is widely used in the time series analysis. Most of the methods of forecasting time series rely on analysis of past data, assuming the historical patterns can be used to forecast future events. In recent years, the traditional Box-Jenkin's ARIMA model has been widely used in time series analysis. ARIMA models have been widely used for various applications such as medical area, business, economics, finance and engineering.

Muhammet (2012) used the ARIMA model to predict the temperature and rainfall in Afyonkarahisar province, Turkey, until the year 2025 and found an increase in temperature according to quadratic and linear trend models.[11]

Suteanu et al (2013) used daily maximum and minimum temperature records from Canada stations in Atlantic region and suggest a new approach to study surface pattern of air temperature.

Balyani et al. (2014) used ARIMA model in a time (1955-2005) for Shiraz, south of Iran. They found ARIMA model as the optimal model for modelling temperature.[5]

EI-Mallah and Elsharkawy (2016) also showed that the linear ARIMA model and the quadratic ARIMA model had the best overall performance in making short-term predictions of annual absolute temperature in Libya.[7]

khedhiri (2014) studied the statistical properties of historical temperature data in Canada for the period 1913-2013 and determined a seasonal ARIMA model to predict future temperature records.[10]

Anitha et al. (2014) used SARIMA model to forecast the monthly mean of maximum temperature of India and observed a trend in the data.[2]

2.5. ARIMAX

ARIMAX is an extension of ARIMA model. It includes other independent variables. When ARIMA model consist of an additional input variable, the model is known as ARIMAX.

Pankratz (2012) used ARIMX model in his study and referred ARIMAX as dynamic regression model.[13]

Jalalkamali et al. (2015) compared the predictive ability of several artificial intelligence and ARIMAX models for predicting drought using standard precipitation index (SPI). Several artificial intelligence models like adaptive neuro-fuzzy inference system (ANFIS), multilayer perceptron artificial neural network (MLP-ANN), and support vector machine (SVM). The results shows that the accuracy of ARIMAX in predicting drought is more compared to SVM, ANFIS, and MLP-ANN models [8].

Peter and silvia (2012) compared the predictive ability of ARIMA and ARIMAX model in the analysis of a microeconomic time series data. They observed that predictive ability of ARIMA was slightly better than ARIMAX model.

Anggraeni et al (2015) did similar study and observed that the predictability of the ARIMAX was superior to that of the ARIMA model [1].

2.6. CONCLUSION

The review of the above literature shows that dynamic changes of temperature is necessary to be determined and forecasted, as these changes will help in determining the extreme events. The related effects can be removed or lessened by knowing the appropriate action for these extreme events. In recent years, the traditional Box-Jenkin’s ARIMA model has been widely used in time series analysis by investigators. ARIMA model have become, in last decades, a major tool in meteorological applications to understand the phenomenon of air temperature and rainfall. Most of the investigators have admitted that the accuracy of ARIMAX is more as compared to other models. Therefore, present work is an attempt to use ARIMA and ARIMAX model to forecast the maximum and minimum daily temperature and compare the accuracy of both the models, which has become the objective of this study.

CHAPTER 3

METHODOLOGY

This chapter describes the methodology used in this study to assess the best-fit model for temperature of Bairagarh station. This chapter examines the basic plots definitions and concepts of time series analysis, assumptions, conditions, principles and processes involved in the application of ARIMA and ARIMAX, specifically applied in this research work.

3.1. Basic definitions

3.1.1. Time Series

It is defined as a set of statistics typically gathered sequentially or in a uniform set of period, usually daily, weekly, monthly, quarterly, annually, and so on. Time series data occurs naturally in many areas.

Economics-e.g., yearly data for unemployment, college admission

Finance- e.g., daily share price.

Environmental- e.g., monthly rainfall, air temperature

Medicine- e.g., ECG brain wave activity.

3.1.2. Time Series Analysis

It includes procedures that separate a series into components and explainable portions that permits sequence to be distinguish, estimates and forecast to be made. Basically time series analysis allows to comprehend the hidden meaning of the data points through the use of a model to forecast future values based on known past values. Such time series models incorporate GARCH, TARCH, EGARCH, FIGARCH, CGARCH, ARIMA, etc but the focus of this study is on ARIMA model and extended ARIMAX model.

3.1.3. Time Series Graph

Time series plot is a graph, which show observations on the y-axis and equally spaced time intervals on x- axis. The time series plot explicitly consist of-Time scale (index, schedule, clock) on the x-axis; data scale on the y-axis; and lines displaying each time series. The plot are typically used to detect trends in data over time; detect seasonality; and compare trends across groups.

3.1.4. Stationary and Non- Stationary Time Series

In a time series, if the mean, variance, autocorrelation etc. are all constant are termed as stationary time series. Stationary series does not shows trend and seasonality.

The series whose properties like mean, variance, autocorrelation etc. depends on time are termed as non- stationary time series and it consist of trend and seasonality. If the series is non-stationary, it can be converted in to stationary series by applying differencing or standardizing.

3.1.5. Autocorrelation function (ACF)

It is also known as serial correlation. It detect dependence in series. Instead of correlation between two variables, it is a correlation within the series of single variable. If measurements are given as X_1 , X_2 , …. X_N. Than at lag k autocorrelation is given as-

$$\boldsymbol{r_k = \frac{\sum_{i=1}^{N-k}(X_i - \bar{X})(X_{i+k} - \bar{X})}{\sum_{i=1}^{N}(X_i - \bar{X})^2}}$$

It’s value varies between -1 to +1. If autocorrelation is +1, it is said to be perfectly positive correlation and if the value of autocorrelation is -1, it is said to be as perfectly negative correlation.

3.1.6. Partial Autocorrelation Function (PACF)

It is the correlation of two observations at different time spots considering that both the observations are correlated to the observations at other time spot. It gives partial correlation with its own lagged values**.** PACF graph is a plot of partial correlation coefficient between the series and lag of it self. By looking at the graph of ACF and PACF one can easily tentatively define the order of AR and MA terms.

3.2. COMPONENTS OF TIME SERIES

A vital step in selecting appropriate modeling and forecasting technique is to consider the kind of data patterns exhibited from the graphs of the time series plots. The sources of variation in terms of pattern in time series data are characterized into four main components. These components incorporate seasonal variation; trend variation; cyclic changes; and the remaining “irregular” fluctuation.

3.2.1. The Trend (T)

The trend is the long term behavior or pattern of the data or series. The Australian Bureau of Statistics (ABS, 2008) defined trend as the ‘long term’ movement in a time series without calendar related and irregular effects, and is a reflection of the underlying level. It is the result of impacts such as growth in population, inflation in price and general changes in economy.

3.2.2. Seasonal Variations (S)

A seasonal effect is a regular and periodic effect. Some examples include the sharp growth in most retail series in December due to Christmas, or an increase in water demand during summers. Other seasonal effect is moving holidays-the time of holiday such as Easter varies, so the impact of the holiday will be experienced in different periods each year.

Seasonal adjustment is the process of estimating systematic and periodic influences and then removing from a time series. Observed data needs to be seasonally adjusted as seasonal effects can hide both the real underlying movements in the series, as well as certain non-seasonal characteristics. Seasonality in a time series can be identified by systematically spaced crest and troughs, which are consistent in direction and have approximately the same magnitude every year.

Other techniques to detect seasonality include;

i. A seasonal subseries plot for showing seasonality.
ii. Multiple box plots as an alternative to the seasonal subseries plot..
iii. The autocorrelation plot.

3.2.3. Cyclic variations(C)

Cyclic variations are the short-term variations (rises and falls) exist in data that are not of a fixed period. They are generally due to events that are not expected such as those associate with the stock price, etc. the difference between the seasonal and cyclic variation is the fact that the seasonal variation is of a constant length, while the latter varies in length. The length of a cycle is averagely longer than that of seasonality and the magnitude of a cycle usually being more variable than that of seasonal variation.

3.2.4. Irregular variations (I)

The irregular component (also known as the residual) is what remains after the removal of seasonal trend component. It results from short term fluctuation that are neither systematic nor predictable.

3.3. A COMMON ASSUMPTION IN TIME SERIES TECHNIQUES

A common assumption in many time series is that the series is stationary. Being stationary series, mean, variance, and autocorrelation structure do not change with time.

Time series can be transformed into stationary series if it is not stationary. we can transform it to stationary with following techniques.

i. Given series is $X_{t,,}$ we can create the new series by differencing the data.
ii. If there is trend in data, fit some type of curve to data then model the residuals from that fit.
iii. If variance is not constant, stabilize the variance by taking logarithm or square root of the series. For negative data, add constant to make series positive. Then apply transformation. Subtract this constant from the model to obtain predicted values and forecast.

3.4. UNIVARIATE TIME SERIES MODELS

These are model with only one series of observation or variable. Basic univariate time series model and their processes are discussed below.

3.4.1. Common Approaches to univariate time series

There are various approaches to model time series. Most common approaches are given below.

i. Decomposition

Time series is decomposed into trend, seasonal, and residual component, to leave only irregular effects, which is the focus of time series analysis, decomposition may be linked to de-trending and de-seasonalizing data.

ii. The spectral plot

Analyze the time series in the frequency domain, is the another approach. It is the primary tool for analyzing the frequency of time series.

iii. Autoregressive model (AR)

Another approach to model time series is autoregressive model. The equation for an AR(P) model is-

$$X_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_P x_{t-P} + \varepsilon_t$$

Where ϕ_1 ϕ_2 ... ϕ_P are parameters of model, X_t is time series and ε_t is noise.

An AR model is a linear regression of the present value against one or more previous values of series (AR_P). p denotes the order of AR model. It can be analyzed by various methods including standard linear least square technique.

iv. Moving Average model (MA)

$$X_t = \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_P e_{t-q} + e_t$$

Where θ_1 θ_2 ... θ_q are parameters of model, X_t is time series and e_t is noise.

The q denotes the order of MA model. It is basically a linear regression of the value against the white noise or random shocks of one or more previous values of series. Random shocks are assumed to be from same distribution, typically a normal distribution. In this model random shocks are propagated to future values of time series. Sometimes ACF and PACF suggest the MA and AR terms used in the model.

Box and Jenkins proposed an approach that combines the moving average and autoregressive approach. This resulted in ARMA model.

Box- Jenkins model assumes stationary time series. They recommend differencing non-stationary series one or more times to have a stationary series. Thus, used ARIMA model, with “I” stands for “integrated”.

3.5. BOX- JENKINS ARIMA PROCESS

It is a strategy for identifying, estimating and forecasting ARIMA models. An autoregressive integrated moving average (ARIMA) model is a generalization of ARMA model. Box- jenkin methodology, names after the statisticians George Box and Gwilym Jenkins, applies ARIMA models to find the best fit of a time series to historical values of this time series, in order to make forecasts. ARIMA (p, d, q) model where p, d, q are non-negative integers, refers the order of autoregressive, integrated, and moving average parts of model respectively.

3.5.1. Modeling Approach

There are iterative three stage modelling approach, which are:

1. Model identification and model selection: identify seasonality in series and remove it by differencing (seasonal differencing if necessary) the series. Decide AR and MA terms from autocorrelation and partial autocorrelation plot.
2. To have the coefficients that best fit the ARIMA model estimate the parameters from computation algorithm. Common methods are maximum likelihood or nonlinear least square.
3. Test the model whether it conforms to the specifications of a stationary univariate process. Residual should be independent of each other and mean and variance should be constant, if not so than repeat the procedure from step one and build a better model.

1) **Model identification**

a. Stationarity and seasonality

The primary step is to determine stationarity and seasonality in series that needs to be modeled.

Detecting Stationarity: It can be analyzed by sequence plot. If sequence plot shows constant scale and location, then it is stationary. Autocorrelation plot can also detect stationarity, especially non-stationary series with very slow decay.

If a series is non-stationary, it can be made stationary by differencing the series. To achieve stationarity, Box-jenkin recommend differencing the series. However, fitting a curve and substracting the fitted value from original value can also be used in context of box-Jenkin model.

Detecting Seasonality: It can usually be assessed by ACF plot, a seasonal subseries plot, or a spectral plot.

For Box-jenkin model, seasonality is not removed explicitly before fitting model. Instead, one includes the order of seasonal term to ARIMA estimation software. However it would be helpful to apply seasonal difference to data and regenerate ACF and PACF plots. This may help in model identification of non-seasonal component of model. Seasonal differencing removes most or all the seasonality effect.

b. Identify p and q

Identify the order of p and q. ACF and PACF plots determine p and q.

	ACF	**PACF**
AR(p)	Consists of sine waves- dies out exponentially or damped exponential	Is zero after p lags
MA(q)	Is zero after q lags	Consists of mixtures of damped exponential or sine terms- dies out exponentially
ARMA(p, q)	Eventually dominated by AR(p) part- then dies out exponentially	Eventually dominated by MA(q) part- then dies out exponentially

Order of autoregressive process (p): For AR (1), ACF plot should have an exponentially decreasing pattern. Higher order AR process are often a mixture of exponentially decreasing and damped pattern

For higher order AR, consider PACF plot also. For AR (p) process, partial autocorrelation becomes zero at lag p + 1 and greater. Analyze the partial autocorrelation function to examine the evidence of a departure from zero, determined by placing a 95% confidence limit on PACF plot.

Order of Moving- average process (q): The ACF of a MA (q) becomes zero at lag q + 1 and greater, so examine the ACF plot to check where it becomes zero. Apply 95% confidence limit for

ACF plot. ACF plot generally not helpful in identifying the order of MA.

2. Model Estimation

After identifying the order of model, parameters of model are estimated by maximum likelihood to determine AR and MA parameters, and other parameters.

The penalty function statistic namely Akaike information criteria (AIC), is explained in penalizing fitted models based on the principle of parsimony. Models with smallest AIC are deemed to have residuals as white noise process.

$$\text{AIC} = 2k - 2\ln(\hat{L})$$

Where, K = number of estimated parameter

$\hat{L}$= maximum value of likelihood function for model

The estimated AR and MA must also conform to boundary condition of -1 and 1. If AR and MA parameters do not lie within these boundaries, then re-estimate parameters of the model or consider a different candidate model.

3. Diagnostic checking

This is to examine whether fitted model is adequate or not. All the relevant information from the data should be extracted by the model. Residuals should be small and no systematic and predictable pattern should be left.

Residual diagnostics: Residuals of model should exhibit white noise-like behavior: departure from this assumption means that some information can still be exploited in the modelling. There are two methods of residual diagnostic.

Graphical Method

Plot of residuals and examine for systematic patterns

Use the SACF and SPACF of the residuals and examine for significant elements

Testing Method

Autocorrelation tests: one problem with checking the significance of individual autocorrelation is that each element might be individually insignificant, but all of the element may be jointly significant.

3.6. FORECAST PERFORMANCE MEASURES

RMSE

Root mean square error (RMSE) (Steiger and Lind 1980; Willmott and Matsuura 2005) is an absolute error measure that squares the deviation and keep the positive and negative deviation from cancelling one another out.it is arguably one of the most used goodness of fit indices in most modelling applications. This measure also exaggerate large errors, which can help when comparing methods.

$$RMSE = \sqrt{\frac{\sum_{t=2}^{n}(x_t - \hat{x}_t)^2}{n}}$$

Where, x_t is actual value of a point for a given time period t, n is total number of fitted points, and $\hat{X}$ is the fitted value for time period t.

MAPE

Mean absolute percentage error (MAPE) is a relative error measure that uses absolute value to keep the positive and negative errors from cancelling one another out and uses relative errors to enable you to compare forecast accuracy between time series models. It does not show direction of error.

$$MAPE = \frac{\sum_{t=1}^{n} | \hat{x}_t - x_t |}{\sum_{t=1}^{n} \left(|\hat{\imath}/_t| + |x_t|\right)/\ 2}$$

Where X_t is forecast value for time period

3.7. FORECASTING WITH ARIMA MODEL

G.E.P.BOX and G.M. Jenkins proposed ARIMA model. In general, most of the time series are not stationary. Some series have particular trend; therefore, a differencing the series might transform them into a stationary series. For example, first order differencing can transform a time series with constant slope to a stationary time series with constant mean. The ARIMA model expresses a time series using ARMA model and differencing. Which includes auto regressive (AR) and moving average (MA) model

1**. Autoregressive (AR) model of order p-AR (p)**

It represent that the current value of time series is a combination of previous values of the series**.**it shows the dependency of the value on its one or more previous values if X_1 X_2 ….X_n represent the time series, AR model of order p without a constant term can be given as-

$$X_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_P x_{t-P} + \varepsilon_t$$

2. Moving Average (MA) model of order q- MA (q)

In Moving Average model, the current value is a combination of current and previous values of white noise ($e_{t)}$. This noise series can be obtained by residual and forecast errors. MA model of order q without a constant term can be written as-

$$X_t = \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_P e_{t-q} + e_t$$

3. Autoregressive Moving Average (ARMA) model of order p,q- ARMA(p,q)

It is a combination of AR model of order p and MA model of order q. autoregressive Moving Average model of order p and q can be written as-

$$X_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_P x_{t-P} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_P e_{t-q} + e_t$$

ARMA model based on the assumption that series is stationary but in actually not all the series are stationary in nature. In that case series is made stationary by adopting differencing technique. When time series is made stationary by differencing the data of order d then the integrated term is introduced in ARMA model and named as Autoregressive Integrated Moving Average model.

3.8. Forecast with ARIMAX Model

ARIMAX model can be simply represented by adding covariate on right side of ARIMA equation. If y_t is the covariate at time t and β is its coefficient, than ARIMAX equation can be represented as-

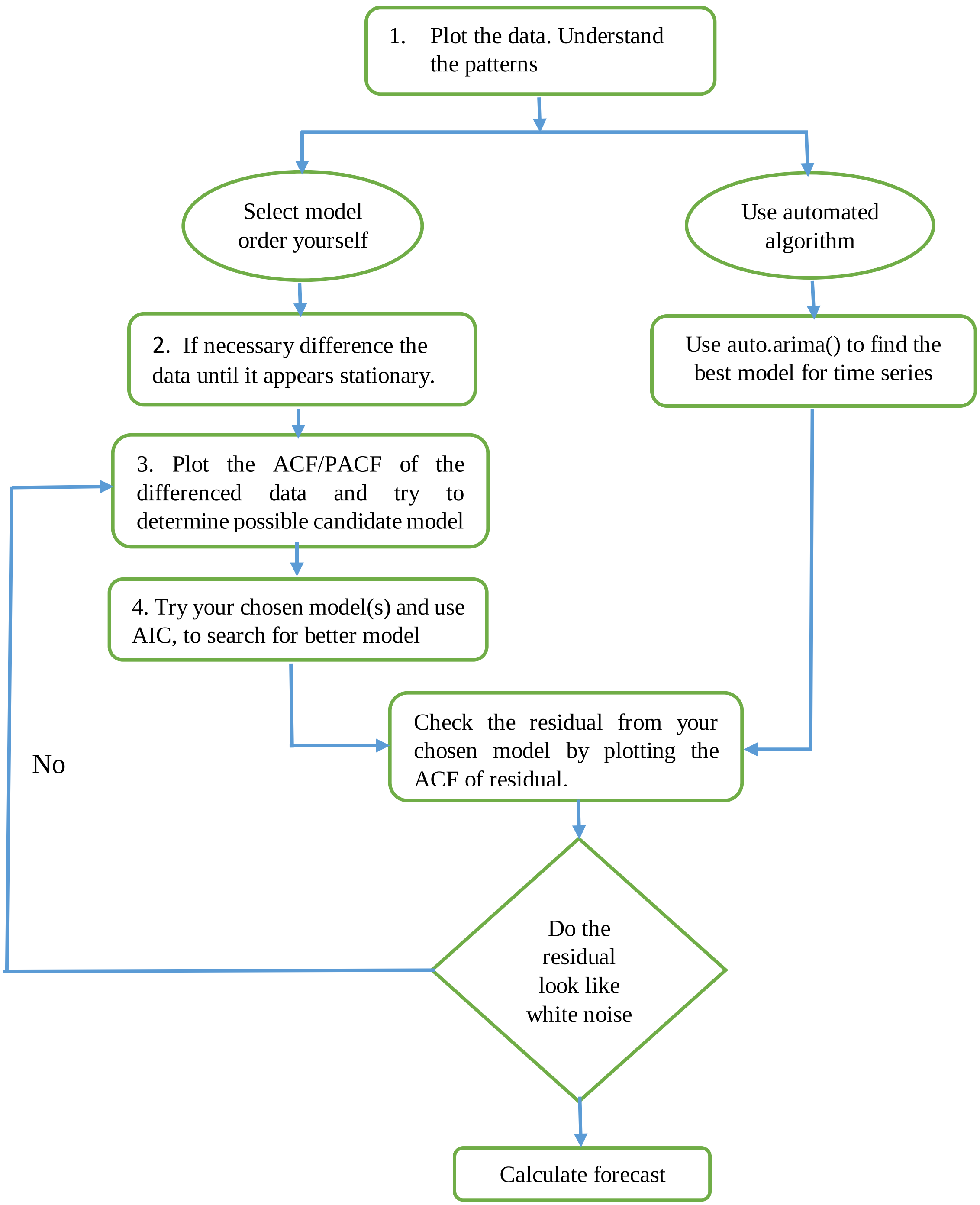
$$X_t = \beta \mathrm{y_t} + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_P x_{t-P} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_P e_{t-q} + \ e_t$$

1. Test the stability of data series. if the series is not stationary, make it stationary by applying initial differencing step.
2. Plot ACF and PACF, to determine p and q parameters.
3. Estimate parameters of model and test the performance
4. Apply the same procedure to series of exogenous variable.
5. Estimate the cross correlation coefficient between the input series and exogenous series to determine configuration of ARIMAX model.
6. Establish diagnostic analysis to verify that the model correspond to the characteristics of data.

3.9. R SOFTWARE

The estimation of ARIMA model in R is done by using maximum likelihood estimation. To maximize the probability of getting the data that we have observed for given value of p, d, and q it maximize the log-likelihood while finding the parameters. Workflow of the software is shown in Fig. 3.1.

Fig. 3.1. Work Flow diagram of R software



.

3.10. DATA COLLECTION AND ANALYSIS

3.10.1. General Study

Bairagarh is a village in Berasia tehsil of Bhopal district, Madhya Pradesh. It comes under Betwa river basin, shown in figure 3.1[15].

Betwa basin having latitude (77 10’-80 20’E) and longitude (22 54’-26 05’ N), located in central India. The basin has catchment area of 43895 sqkm; the elevation range from 106 to 706 m above mean sea level. The Betwa River originates in the Raisen district of Madhya Pradesh near Barkhera village south-west of Bhopal. The Betwa River is an interstate river between Madhya Pradesh and Uttar Pradesh. It flows through Northeastern district of Madhya Pradesh and enters near Banhawan village of Jhansi district in Uttar Pradesh. The total length of the river is 590 km up to its confluence with Yamuna River, out of which 232 km lies in Madhya Pradesh and rest 358 km in Uttar Pradesh.

The areas covered by Betwa River are-Bundelkhand uplands, the Malwa plateau and the Vindhyan scrap lands in the districts of Shivpuri, Guna, Bhopal, Vidisha, Raisen, Tikamgarh, Sagar and Chhatarpur of Madhya Pradesh and in Uttar Pradesh it covers region of Hmirpur, Jalaun, Jhansi, and Banda district.

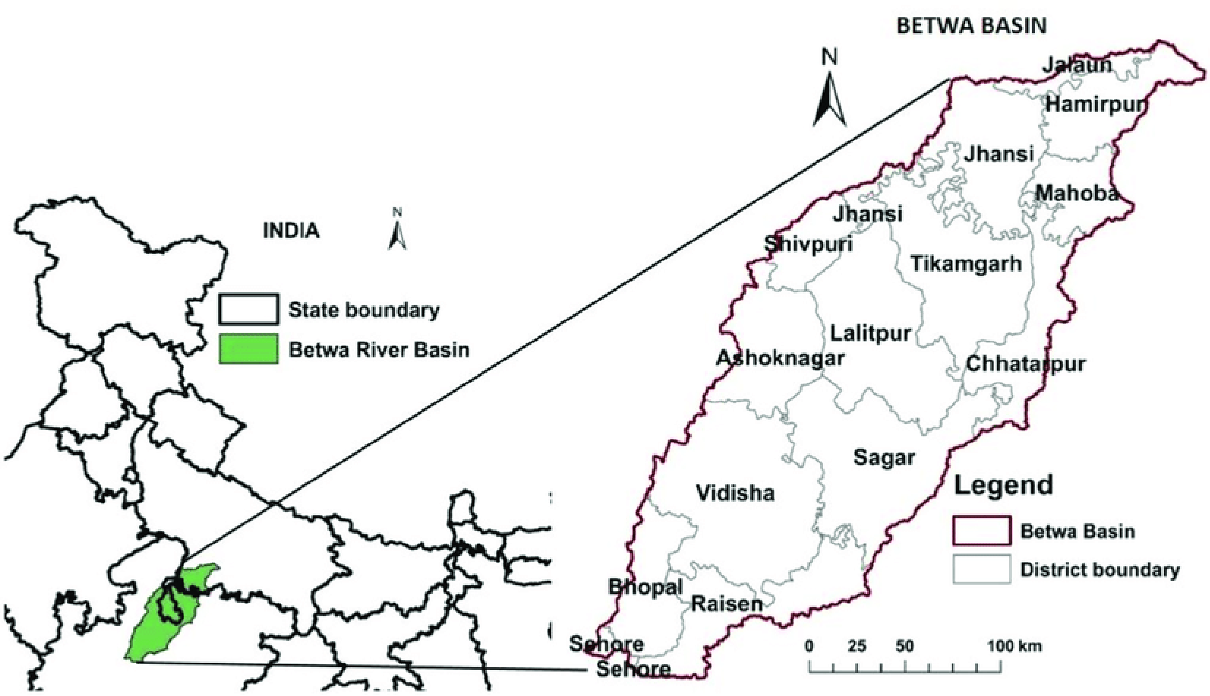


Fig. 3.2 location of Betwa Basin

3.10.2. Bhopal

Bhopal is the capital city of Madhya Pradesh. It is one of the greenest cities in India. There are two lakes namely upper lake and lower lake having surface area of 36 km^2 and 1.29 km^2 respectively with catchment area of 361 km^2 and 9.6 km^2 respectively. Average elevation of Bhopal is 500 m (1401 ft.) .It is situated in central part of India [16].

Climate

The climate of Bhopal is humid subtropical, with cool and dry winters, hot summer and a humid monsoon season. The average temperature in summer (starts in late march and ends in mid-June) is around 30°C, which regularly exceeds 40°C during mid of May. The monsoon starts in late June and go on till late September. The precipitation during these months is about 40 inches and the average temperature is about 25°C with high humidity. The average daily temperature during winters is around 16°C, and winter peak is in January

3.11. DATA COLLECTION

To get the accurate predictions of future events, it is important to study the behavior of historical data. Long-term historical data of temperature and rainfall from 1982 to 2012 were taken from NASA's POWER data access viewer [12]. The data used are daily maximum and minimum temperature and daily rainfall. The data was collected for Bairagarh station, Bhopal. Bairagarh station comes under Betwa basin. Model building and forecast was done in R software. The summary of data collected is given in table. 3.1 and maximum temperature , minimum temperature and precipitation data are shown in figure 3.3, 3.4 and 3.5 respectively.

Parameter	Minimum	Mean	Maximum	St deviation
Max Temperature	15.7	32.75	47.61	5.52
Min Temperature	2.25	18.96	33.55	6.22
Rainfall	0	3.112	211.02	9.79

Table3.1 summary of observed variables

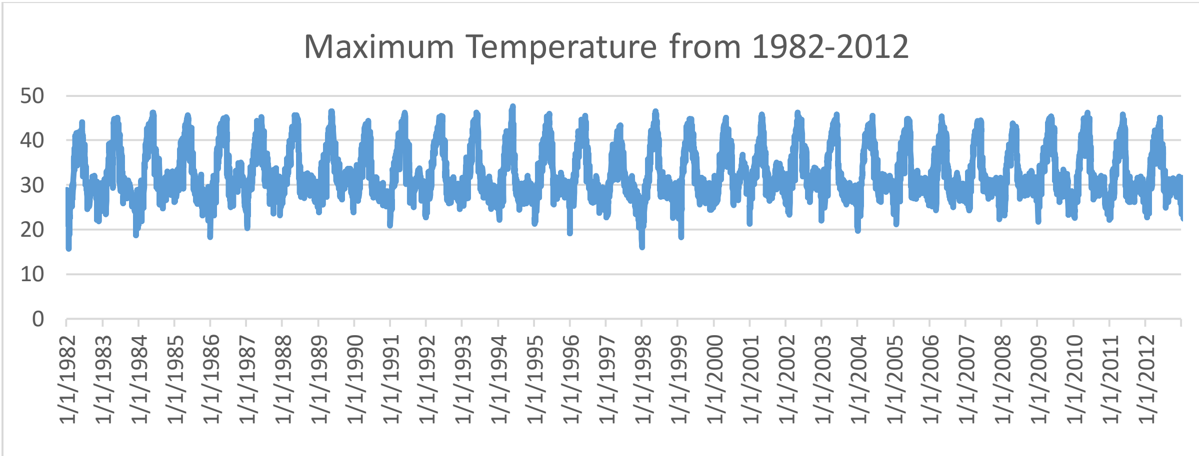


Fig. 3.3. Daily observed maximum temperature from (1982-2012)

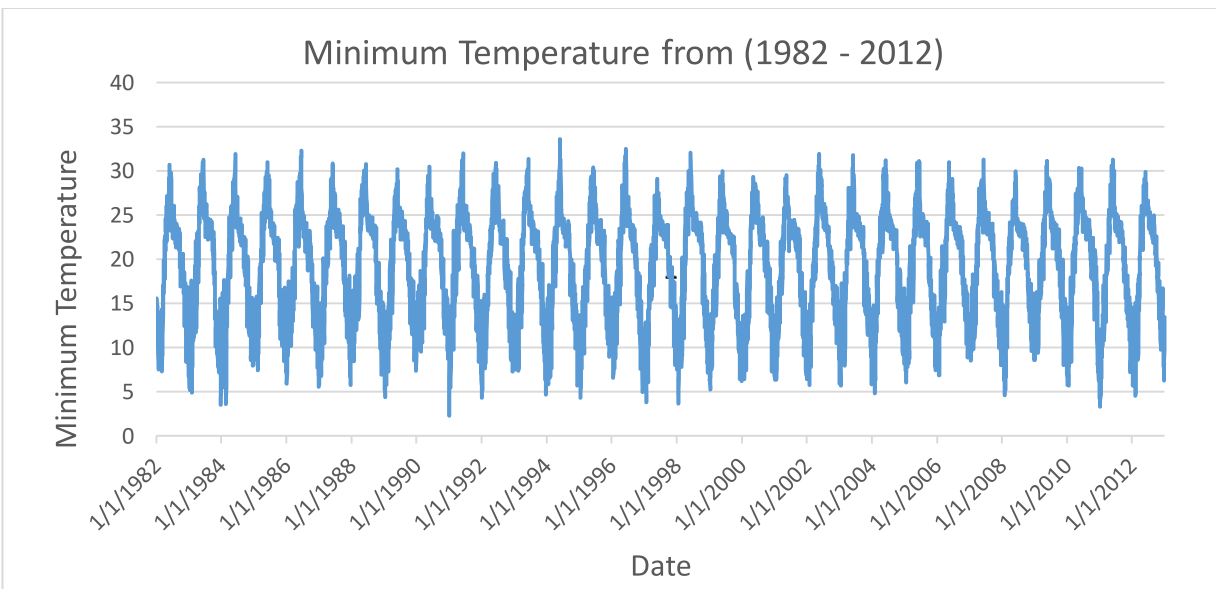


Fig. 3.4 Daily Minimum Temperature from (1982- 2012)

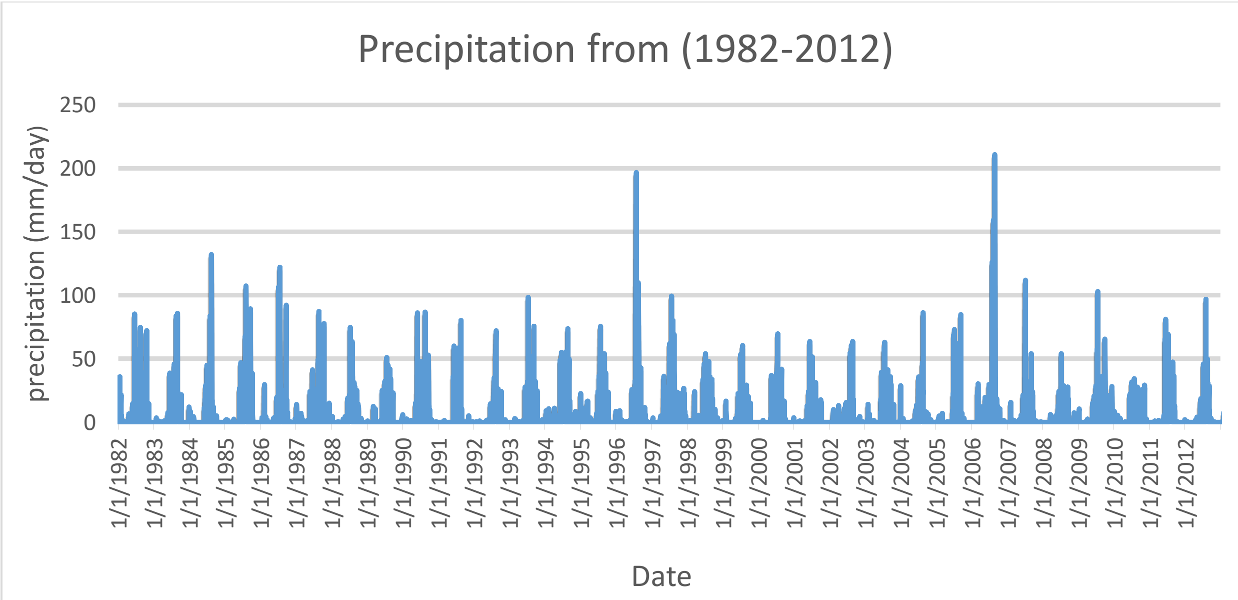


Fig. 3.5. Daily observed precipitation from (1982-2012).

CHAPTER 4

DEVELOPMENT OF MODELS

4.1. INVESTIGATING STATIONARITY IN SERIES

The daily maximum and minimum temperature data from year 1982 to 12012 was divided into 3 parts. All the three parts were plotted individually against time to investigate the stationarity in series. The plots are shown in fig. 4.1, 4.2, and 4.3. From figures, it is clear that the series is stationary as the mean and variance are constant in all the three sets. Since all the three sets have shown same behavior, data set from (2002 – 2011) was selected for further modelling process.

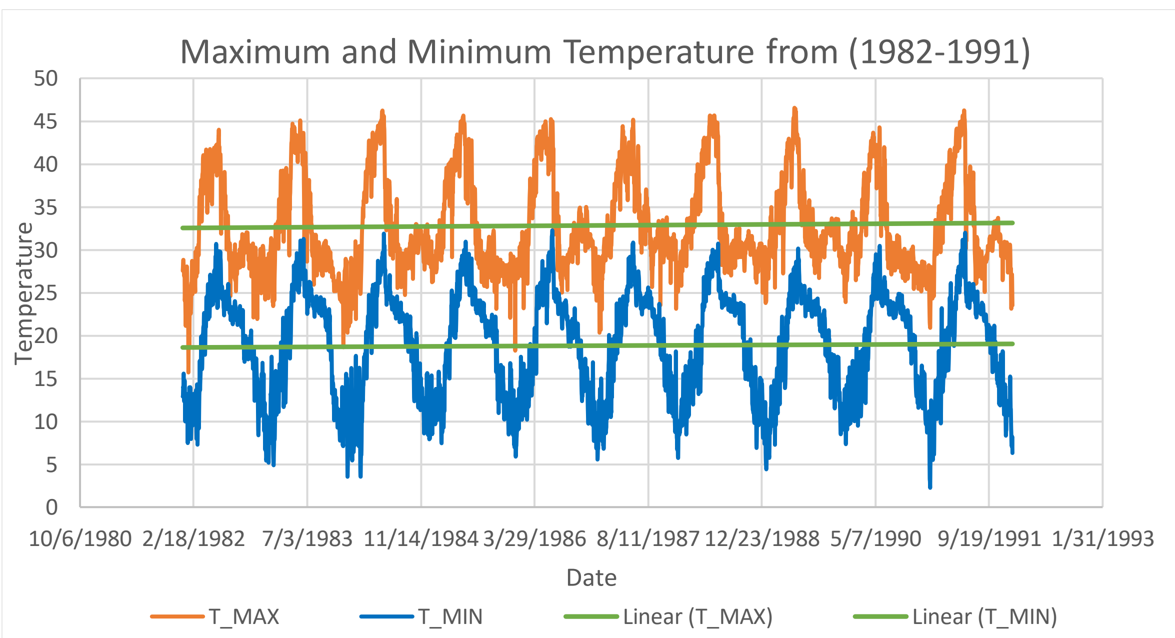


Fig. 4.1. Daily maximum and minimum Temperature split data from (1982-91)

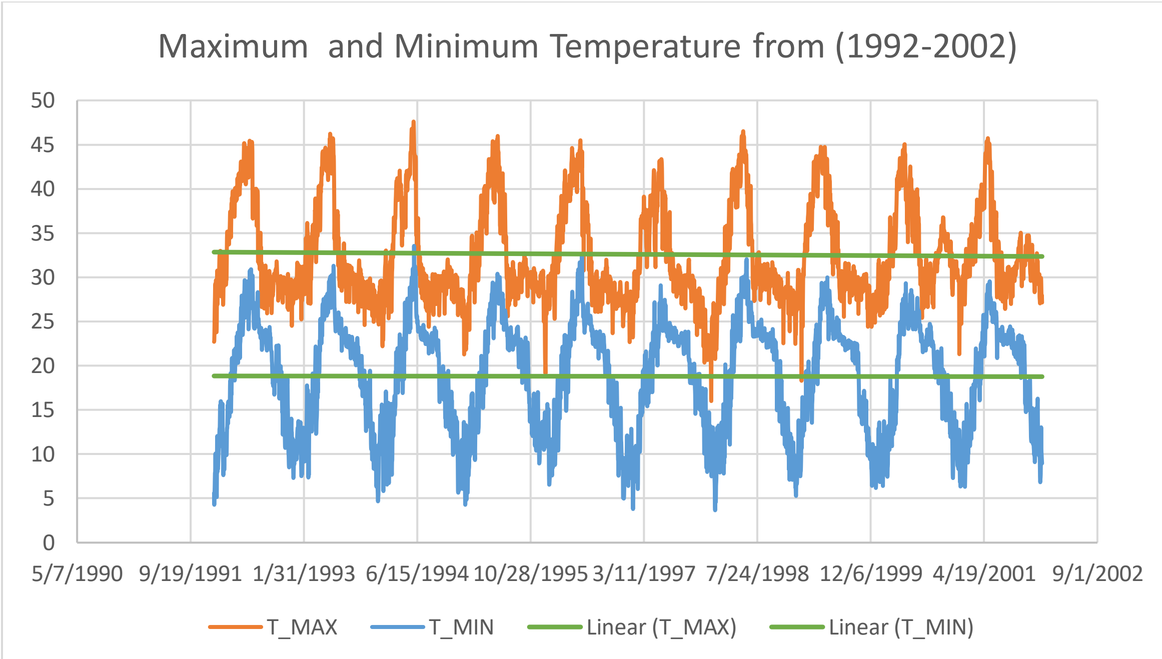


Fig. 4.2 daily Maximum and Minimum Temperature split data from (1992-2002)

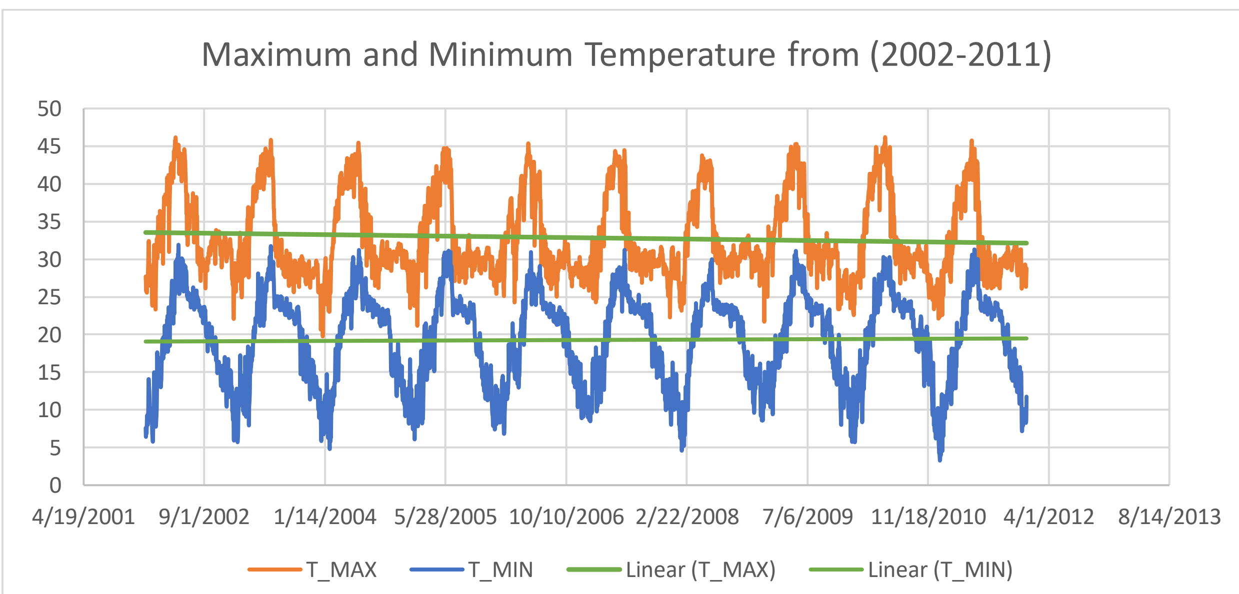


Fig. 4.3 Daily Maximum and Minimum temperature split data from (2002 -2011)

4.2. FINDING p AND q

To determine the values for p and q, ACF and PACF were plotted. ACF and PACF plot helps in determining the initial values of p and q. Followed by few iterations, final model will be based on minimum value of AIC, errors matrices and residual ACF.

4.3. ARIMA MODEL

4.3.1. Training Stage

We have investigated the performances of several models for maximum and minimum temperature data series and selected the one with least AIC, residual ACF and minimum error matrices like RMSE and MAPE. Few of the iterative models are shown in table 4.1 for maximum temperature, and in table 4.2 for minimum temperature. Residual ACF and PACF are also shown in figure 4.6 to 4.11 for maximum temperature and in figure 4.14 to 4.18 for minimum temperature.

4.3.1.1 Parameters and summary of iterative models for Maximum Temperature

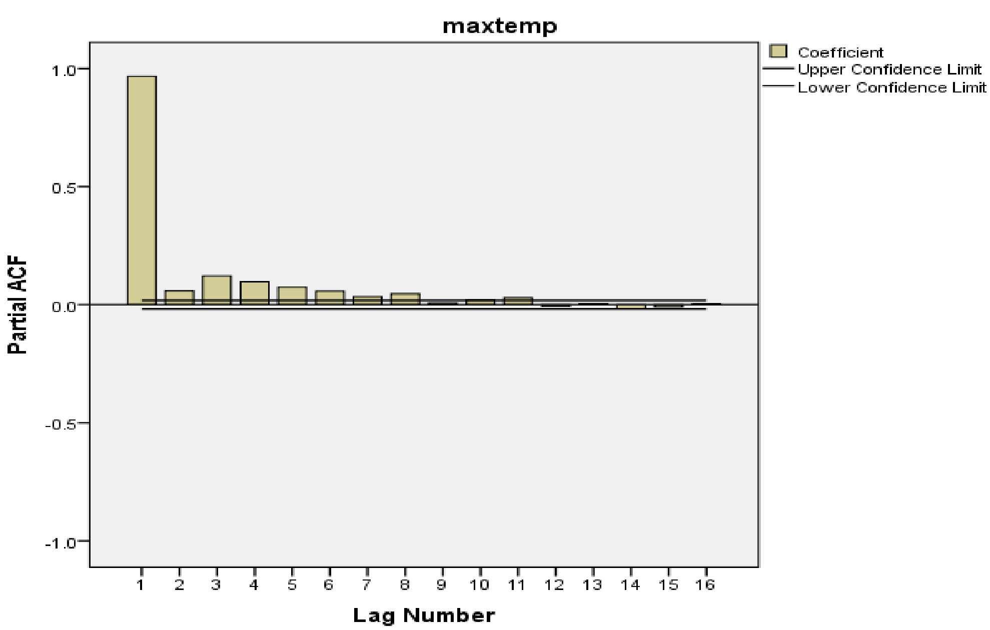


Fig. 4.4. Partial autocorrelation function for maximum temperature

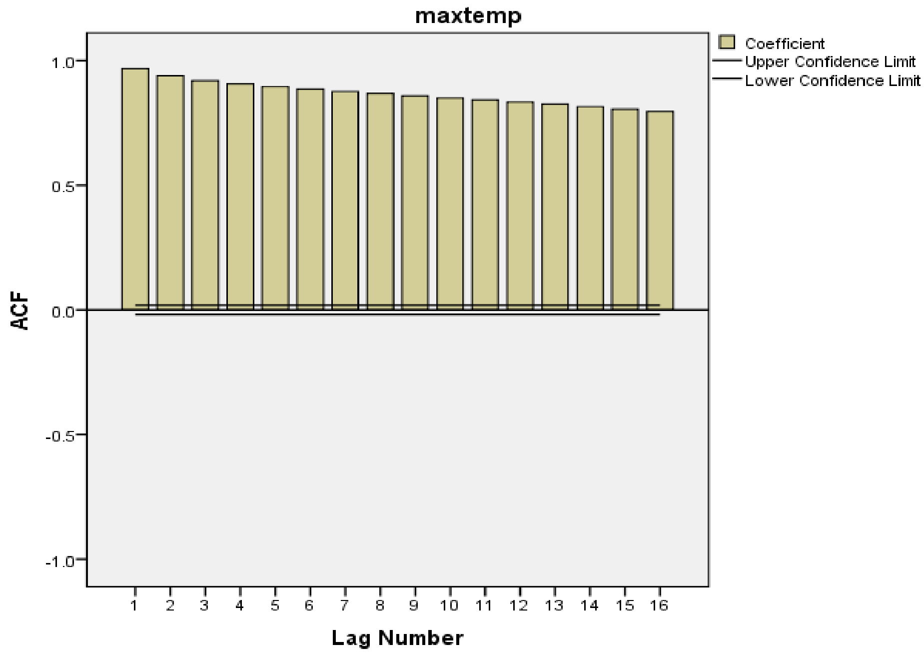


Fig.4.5. Autocorrelation function for maximum temperature

	ARIMA(1,0,0)	ARIMA(1,1,0)	ARIMA(1,1,1)	ARIMA(1,0,1)
Coefficient	ar1 mean 0.9695 32.8371 s.e 0.0040 0.7212	ar1 -0.0560 s.e. 0.0165	ar1 ma1 0.6750 -0.8252 S.e 0.0276 0.0202	ar1 ma1 mean 0.973 -0.0587 32.8374 s.e 0.004 0.0202 0.7647
AIC	12522.05	12558.71	12423.88	12515.52
RMSE	1.34214	1.350233	1.325144	1.340573
MAPE	2.993457	2.998419	2.953642	2.995612
ACF1	-0.03993024	-0.007711732	0.05222951	0.007820855

Table. 4.1. Statistics of models for Maximum Temperature.

	ARIMA(1, 1, 2)	ARIMA (2, 1, 1)
Coefficient	ar1 ma1 ma2 0.5210 -0.6204 -0.1169 Se. 0.0468 0.0476 0.0210	ar1 ar2 ma1 0.655 -0.1027 -0.7526 Se. 0.034 0.0186 0.0308
AIC	12397.67	12396.3
RMSE	1.320031	1.319782
MPE	-0.1261774	-0.1258042
ACF1	0.001262538	-0.0007996728

Continued.. Table 4.1

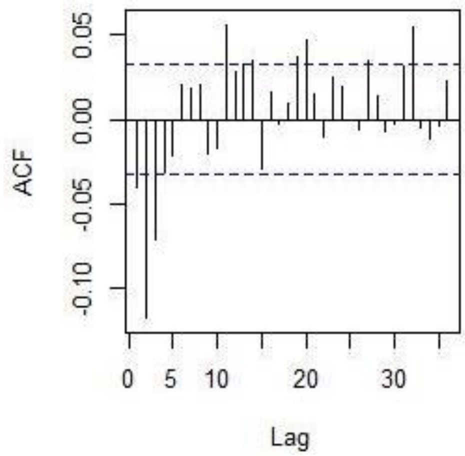


Fig.4.6. Plot of residual ACF for ARIMA(1, 0, 0)

Parameters of model (1, 0, 0) are okay, AIC, error measures and not very good as compared to model (1, 1, 1) ,(2, 1, 1) and (1, 1, 2) .Residual ACF have several spikes outside the significance level, which can not be ignored. Hence, this model was not considered for further process.

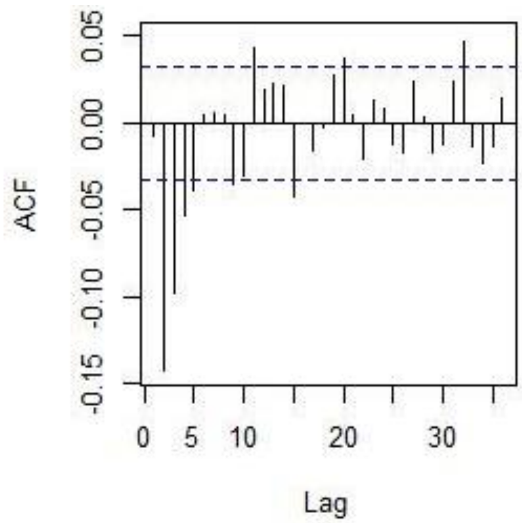


Fig. 4.7. Plot of residual ACF for ARIMA (1, 1, 0)

Parameters in model (1, 1, 0) are good. RMSE ,MAPE and ACF are not as good as in model (1,1,1) (2,1,1) and (1,1,2). In the plot of residual ACF, several spikes are outside the limit, in which few are not to great extent and can be ignored, but at initial lag of 2, 3 and 4 ACF are outside the lower significance level, which are not acceptable. Hence, this model is not good to go.

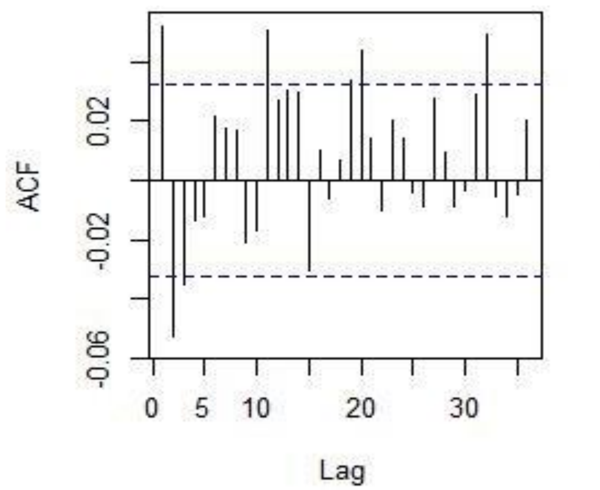


Fig. 4.8. Plot of residual ACF for ARIMA (1, 1, 1)

Lag of order 1, 2, 11, 20, 32are outside the limit, some are of great extent that cannot be ignored and although there is no pattern in these lags. However, the parameters, AIC and error measures are good and as compared to models (1, 0, 1), (1, 0, 0) and (1, 1, 0), model (1, 1, 1) have shown better results. Hence, the model was selected for testing in validation stage.

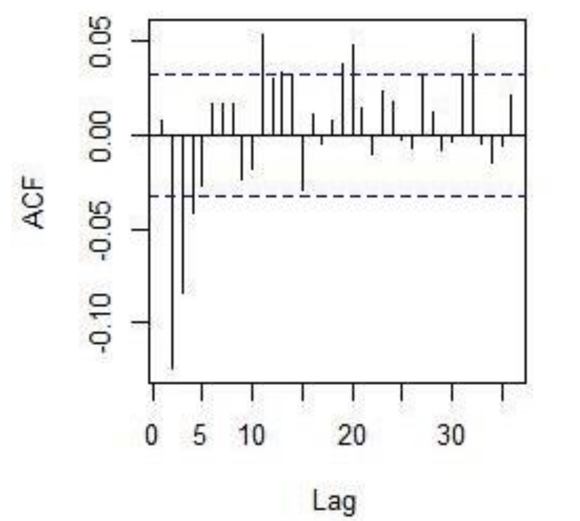


Fig. 4.9. Plot of residual ACF for ARIMA (1, 0, 1)

Parameters of model (1, 0, 1) are good but AIC value and error matrices are not better than (1, 0, 0) and (1, 1, 0). Residual ACF at lag 2, 3, 4, 11, 20, and 31 are outside the limit there is no pattern in these lags. Residual ACF plot are not much satisfactory as the ACF at initial lag of 2, 3, and 4 are outside the lower significance level, which is not preferred. Hence, this model was not considered for further testing.

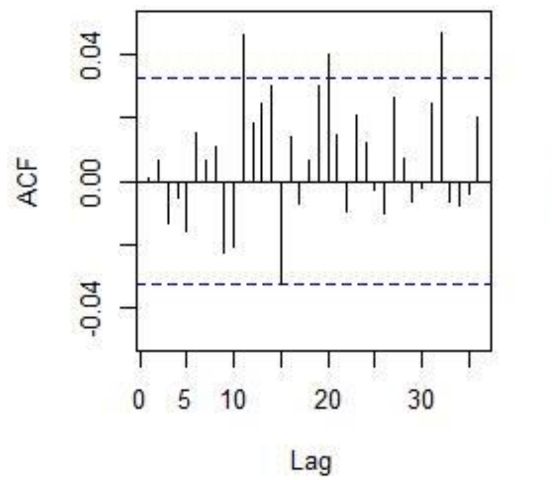


Fig.4.10. Plot of residual ACF for ARIMA (1, 1, 2)

ARIMA (1, 1, 2) have performed well in training stage as AIC value is low, RMSE, MAPE are also low. ACF of residual are also good, only 3 bars are outside the limit but these are not much outside and can be accepted

.

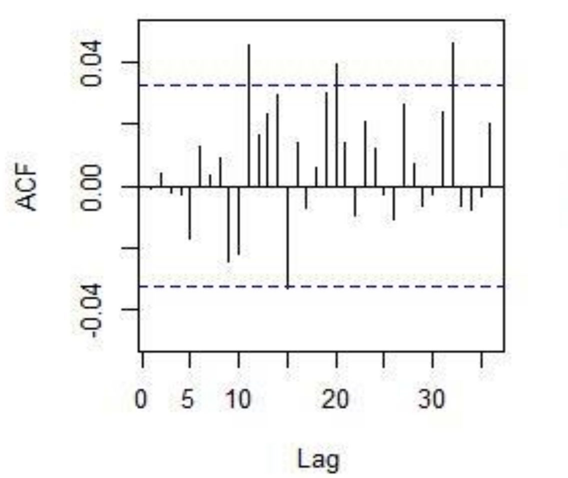


Fig.4.11. Plot of residual ACF for ARIMA (2, 1, 1)

Parameters of model (2, 1, 1) are good, error measures, and AIC value are minimum compared to all the models. 11th 20th and 31st order lag in residual ACF plot are violating the limits, but not to great extent, hence the overall results are good. This model has performed better than rest of the models

4.3.1.2. Parameters and summary of iterative model for Minimum Temperature

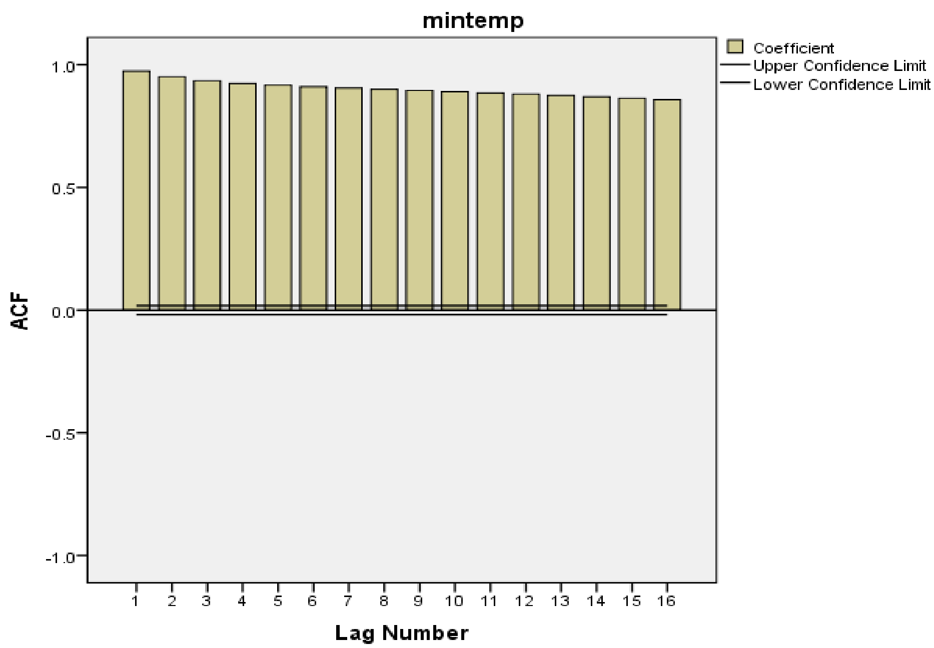


Fig.4.12. Autocorrelation function for minimum temperature

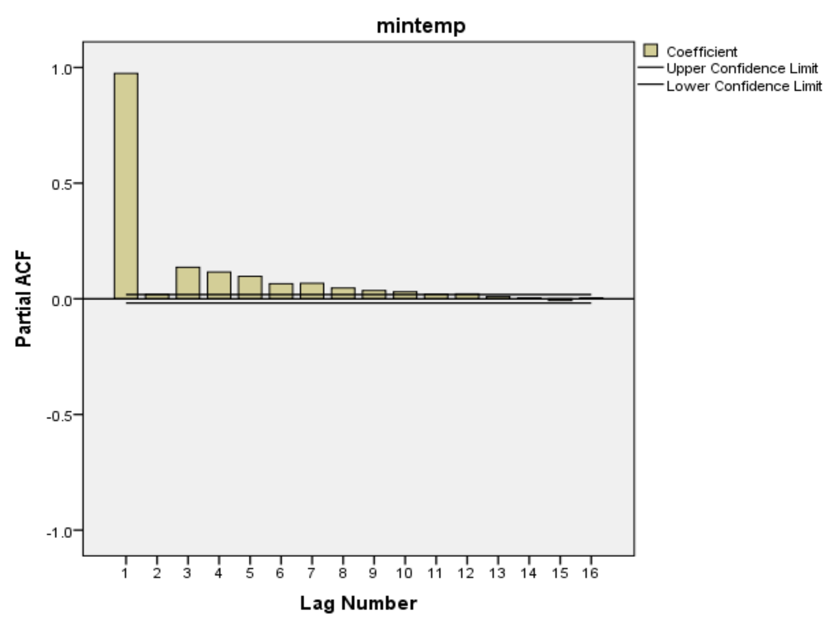


Fig.4.13. Partial autocorrelation function for minimum temperature

	ARIMA (1, 1, 1)	ARIMA (3, 1, 0)	ARIMA (1, 1, 2)	ARIMA (2, 1, 1)	ARIMA (3, 0, 0)
Coefficient	ar1 ma1 0.6657 -0.8341 Se.0.0245 0.017	ar1 ar2 -0.0820 -0.1647 Se. 0.0164 0.01630 ma3 -0.1205 Se. 0.0164	ar1 ma1 0.4995 0.6074 Se.0.042 0.0434 ma2 -0.1397 Se. 0.0211	ar1 ar2 0.6584 -0.1197 Se.0.0305 0.0184 ma1 -0.7645 0.0269	ar1 ar2 0.928 -0.095 Se 0.0164 0.0224 ar3 mean 0.1478 19.1295 Se. 0.0164 1.1215
AIC	12253.32	12281.22	12215.74	12214.16	12315.47
RMSE	1.294545	1.299155	1.287539	1.28726	1.303937
MAPE	6.797668	6.090528	6.068753	6.067791	6.147745
ACF1	0.1137952	-0.01088259	0.001420547	-0.0003633838	-0.01699626

Table. 4.2. Statistics of models for Minimum Temperature

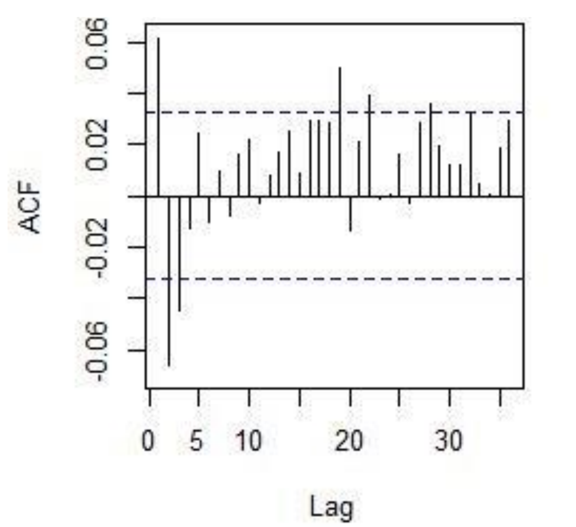


Fig.4.14. plot of residual ACF for ARIMA (1, 1, 1)

The parameters of model (1, 1, 1) are good, error matrices and AIC value are also good. Plot of residual ACF have only few numbers of spikes violating the significance level and rest are within the limit. There is no pattern in these lags so overall the result is good.

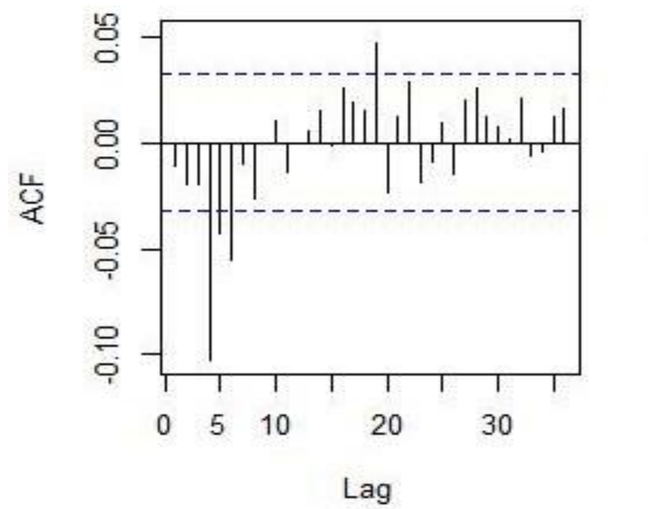


Fig. 4.15. Plot of residual ACF for ARIMA (3, 1, 0)

The residual ACF at lag 4 and 19 are violating the significance level, but the number of violating spikes are less. The parameter of model (3, 1, 0) are good, statistics are also good but not better than model (1, 1, 1).

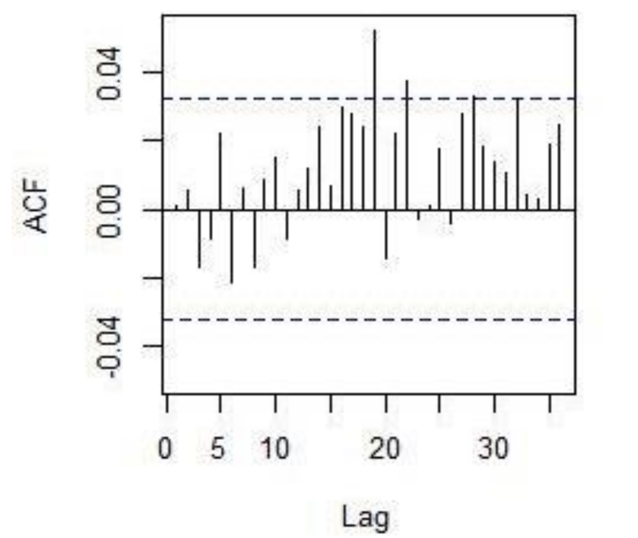


Fig.4.16. Plot of residual ACF for ARIMA (1, 1, 2)

Parameters of the model are good, AIC value is less, RMSE and MAPE are reduced compared to the (1, 1, 1) and (3, 1, 0). Plot of residual ACF are also very good, as all the spikes are within the significance level except lag at 19, which is acceptable.

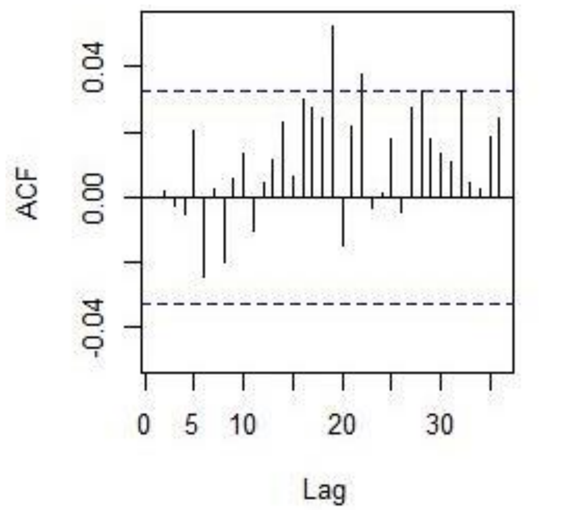


Fig. 4.17. Plot of residual ACF for ARIMA (2, 1, 1)

Results of the model is similar to the model (1, 1, 2), and in fact it is better. Residual ACF are good, as there is only spike violating the significance level, which is acceptable. Parameters and statistics of model are also good.

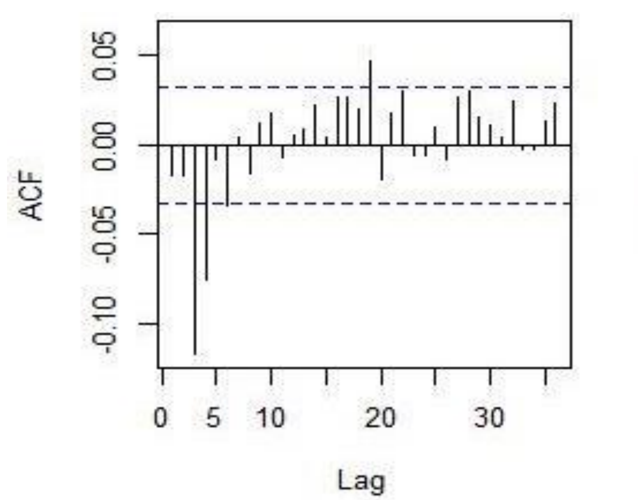


Fig.4.18 Plot of residual ACF for ARIMA (3, 0, 0)

Error measures like RMSE, MAPE are okay. Residual ACF is also okay. Test the model in validation stage and analyze the performance.

4.3.1.3. Goodness of fit at Training stage

Decision of best model among these four models were done based on the principle of parsimony, least AIC value and minimum measures of error like RMSE, MAPE and ACF1. Among all the models, ARIMA (2, 1, 1) shown best results .AIC value and error matrices are minimum compared to other models. The difference between ARIMA (1, 1, 2) and ARIMA (2, 1, 1) is marginal as AIC=12397.67 for ARIMA (1, 1, 2), whereas for ARIMA (2, 1, 1) it is 12396.3, there is minor difference of 1.37. If we talk about RMSE, than there is a difference of 0.000249 and for MAPE, it is 0.00037. Results from ARIMA (1, 1, 1) are also good. Some of the lags of residual ACF are outside the confidence limit but not to great extent. Therefore model (1, 1, 1), (1, 1, 2), and (2, 1, 1) were selected for Maximum Temperature series.

Similarly, for minimum temperature series, model (2, 1, 1) have shown best results as the AIC value is minimum, RMSE, MAPE are less. For minimum temperature series, we are considering all the models for further validation process. Then, all the good to go models for both maximum and minimum temperature series were tested for the consistency and performance of model at validation stage.

4.3.2. Testing stage

Models, which have performed better at training stage, were implemented on the test data to analyze the performance and consistent behavior of the model.

4.3.2.1. For Maximum Temperature

ARIMA (1, 1, 1)

Training set	Test set
Coefficients: ar1 ma1 0.6750 -0.8252 S.e. 0.0276 0.0202 t-value 24.4 40.8	Coefficients: ar1 ma1 0.675 -0.8252 S.e. 0.000 0.0000
AIC=12423.88	AIC=1276.38
RMSE= 1.325144	RMSE=1.384389
MAPE= 2.953642	MAPE= 3.152277
ACF1= 0.05222951	ACF1= -0.006287662

SSR_=6412.935	SSR_=701.4508

Table. 4.3. Statistics of ARIMA (1, 1, 1) at validation stage for maximum temperature.

At validation stage, the error was increased, which is obvious. value of ACF was 0.0522951 and becomes -0.006287662 at testing stage , this value is under confidence limit. Performance of model was consistent and good at testing stage

ARIMA (2, 1, 1)

Training set	Test set
Coefficients: ar1 ar2 ma1 0.655 -0.1027 -0.7526 s.e. 0.034 0.0186 0.0308 t-stat 19.2 -5.5 -24.4	Coefficients: ar1 ar2 ma1 0.655 -0.1027 -0.7526 s.e. 0.000 0.0000 0.0000
AIC=12396.3	AIC=1273.89
RMSE= 1.319782	RMSE= 1.379608
MAPE= 2.939284	MAPE=3.125651
ACF1= -0.0007996728	ACF1=-0.06198128
SSR=6361.145	SSR=696.6141

Table.4.4. Statistics of ARIMA (2, 1, 1) at validation stage for maximum temperature

Model was consistent, performed better than model (1, 1, 1).

ARIMA (1, 1, 2)

Training set	Test set
Coefficients: ar1 ma1 ma2 0.5210 -0.6204 -0.1169 s.e. 0.0468 0.0476 0.0210	Coefficients: ar1 ma1 ma2 0.521 -0.6204 -0.1169 s.e. 0.000 0.0000 0.0000
AIC= 12397.67	AIC=1273.74
MAPE= 2.940243	MAPE= 3.128556
RMSE= 1.320031	RMSE= 1.379335
ACF1= 0.001262	ACF1= -0.058815
SSR=6363.543	SSR=696.3391

Table 4.5. Statistics of ARIMA (1, 1, 2) for maximum temperature data at validation stage

At testing stage, the model was consistent errors were slightly increased. Value of ACF obtained as -0.05. Overall, the performance was better compared to other two models.

4.3.2.2. For Minimum Temperature

ARIMA (1, 1, 1)

Training set	Test set
Coefficients: ar1 ma1 0.6657 -0.8341 s.e. 0.0245 0.0170	ar1 ma1 0.6657 -0.8341 s.e. 0.0000 0.0000
AIC=12253.32	AIC=1240.73
RMSE= 1.294545	RMSE=1.318332
MAPE= 6.112144	MAPE= 6.797668
ACF1= 0.06158152	ACF1=0.1137952
SSR=6120.8	SSR=636.1074

Table. 4.6. Statistics for ARIMA (1, 1, 1) at validation stage on minimum temperature data

Note: training data is of 10 years i.e. 3652 entries, whereas test data is of 1 year i.e. 366 entries. The value of ACF, obtained at testing stage was very high rest statistics were good.

ARIMA (2, 1, 1)

Training set	Test set
Coefficients: ar1 ar2 ma1 0.6584 -0.1197 -0.7645 s.e. 0.0305 0.0184 0.0269	ar1 ar2 ma1 0.6584 -0.1197 -0.7645 s.e. 0.0000 0.0000 0.0000
AIC=12214.16	AIC=1236.45
RMSE=1.287261	RMSE=1.310559
MAPE=6.067791	MAPE=6.731284
ACF1= -0.0003633838	ACF1= 0.0449708
SSR=6051.509	SSR=628.6289

Table. 4.7. Statistics of ARIMA (2, 1, 1) for minimum temperature data at validation stage.

The residual ACF value was increased to 0.0449708, which is slightly above the confidence limit. Error matrices were slightly increased at testing stage. Overall performance of the model was consistent and good.

ARIMA (1, 1, 2)

Training set	Test set
Coefficients ar1 ma1 ma2 0.4995 -0.6074 -0.1397 s.e. 0.0426 0.0434 0.0211	ar1 ma1 ma2 0.4995 -0.6074 -0.1397 s.e. 0.0000 0.0000 0.0000
AIC=12215.74	AIC=1237.54
MAPE=6.068753	MAPE=6.752003
RMSE=1.287539	RMSE= 1.312518
ACF1=0.001420547	ACF1= 0.0490354
SSR=6054.131	SSR=630.5096

Table 4.8. Statistics of ARIMA (1, 1, 2) for minimum temperature series at validation stage

The statistics of the model were similar to the model (2, 1, 1). The model was consistent at testing stage

ARIMA (2, 0, 2)

Training set	Test set
ar1 ar2 ma1 ma2 mean 1.4836 -0.4866 -0.5939 -0.1404 19.0569 s.e. 0.0446 0.0442 0.0452 0.0210 1.8058	ar1 ar2 ma1 ma2 mean 1.4836 -0.4866 -0.5939 -0.1404 19.0569 s.e. 0.0000 0.0000 0.0000 0.0000 0.0000
AIC=12216.6	AIC=1243.48
RMSE=1.286019	RMSE=1.313249
MAPE=6.08717	MAPE=6.80077
ACF1=0.0009850433	ACF1=0.05465441
SSR=6039.837	SSR=631.2117

Table 4.9. Statistics of ARIMA (2, 0, 2) on minimum temperature series at validation stage.

Model (2, 0, 2) have performed well at testing stage, selection of the best model among these tentative models would be based on the comparison of the statistics of these models.

ARIMA (3, 0, 0)

Training set	Test set
ar1 ar2 ar3 mean 0.9283 -0.0951 0.1478 19.1295 s.e 0.0164 0.0224 0.0164 1.1215	ar1 ar2 ar3 mean 0.9283 -0.0951 0.1478 19.1295 s.e 0.0000 0.0000 0.0000 0.0000
AIC=12315.47	AIC=1258.19
RMSE= 1.303937	RMSE=1.34025

MAPE=6.147745	MAPE=6.880539
ACF1=-0.01699626	ACF1=0.04798551
SSR=6209.316	SSR=657.4353

Table. 4.10. Statistics of ARIMA (3, 0. 0) for minimum temperature series at validation stage.

ARIMA (3, 1, 0)

Training set	Test set
ar1 ar2 ar3 -0.0820 -0.1647 -0.1205 s.e. 0.0164 0.0163 0.0164	ar1 ar2 ar3 -0.082 -0.1647 -0.1205 s.e. 0.0000 0.0000 0.0000
AIC=12281.22	AIC=1253.26
RMSE=1.299155	RMSE=1.34127
MAPE= 6.090528	MAPE=6.811527
ACF1= -0.01088259	ACF1= 0.04583181
SSR=6163.863	SSR=658.4355

Table 4.11 Statistics of ARIMA (3, 1, 0) for minimum temperature series at validation stage

4.3.2.3. Goodness of fit at testing stage

For maximum temperature series, ARIMA (1, 1, 2) have best results. Whereas in training stage, best model was ARIMA (2, 1, 1). There was a minor difference in all the statistics of ARIMA (1, 1, 2) and ARIMA (2, 1, 2) in training stage and so here is. The difference in AIC value between both the models is 0.15. All the three models performed well in testing stage. The errors are little higher in testing stage as compared to training stage which is obvious.

For minimum temperature series, all the models performed well in testing stage except ARIMA (1, 1, 1). ACF value is high in testing stage so reject this model. Rest all are good.

CHAPTER 5

DEVELOPMENT OF TIME SERIES MODELS WITH EXOGENOUS VARIABLES

5.1. ARIMAX MODEL

ARIMAX is an extension of ARIMA model. It includes other independent variables. When ARIMA model consist of an additional input variable, the model is known as ARIMAX. To implement ARIMAX model, exogenous variable is required. In this study, independent variable-daily rainfall is exogenous variable. Apply ARIMAX model to those iterative models, which were already tested in ARIMA modelling, to compare the performance of two (ARIMA and ARIMAX) models.

5.1.1. Training stage

Plot of ACF and PACF, determining p and q, iterations to choose best-fitted models has already done in ARIMA modelling. The models, which have already been calibrated in ARIMA modelling will now be considered for ARIMAX modelling to compare the performance of both the models. The AIC value has reduced, error measures are also improved in both training and testing stage as compared to ARIMA model.

5.1.1.1. Parameters and summary of models for Maximum Temperature

ARIMAX (1, 1, 1)

Parameters of the model are good as it comply t-value.Value of residual ACF is approx 0.04,which is beyond the confidence limit and also there are few spikes outside the limit especially first and second order lag. Statistics and residual ACF plot are shown in table 4.12 and in figure 4.19 respectively.

ARIMAX(1, 1, 1)			
	Ar1	Ma1	Xreg
Coefficient	0.6656	-0.8200	-0.0274
S.E.	0.0283	0.0208	0.0024
t-stat	23.5	39.4	11.41
AIC =12294.05			
RMSE =1.301434			
MAPE =2.909321			
ACF =0.04176108			

Table. 5.12. statistics of ARIMAX(1,1,1) for Maximum temperature at training stage.

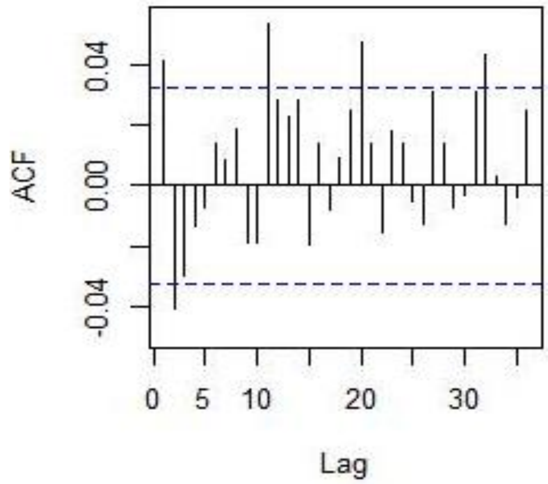


Fig. 5.19. Residual ACF for ARIMAX (1, 1, 1)

ARIMAX (2, 1, 1)

All the Parameters of model are good as t value is greater than 2. Reasidual ACF plot is also good as its value is -0.0008, within the confidence limit. Number of spikes outside the significant level are also less. Hence the performance of model is good.

ARIMAX(2, 1, 1)				
	Ar1	Ar2	Ma1	Xreg
Coefficient	0.6490	-0.0841	-0.7605	-0.0265
S.E.	0.0339	0.0188	0.0304	0.0024
t-stat	19.1	4.4	25	11
AIC=12276.49				
RMSE=1.297949				
MAPE=2.899442				
ACF=-0.000840491				

Table 5.13. statistics of ARIMAX(2,1,1) for maximum temperature at training stage

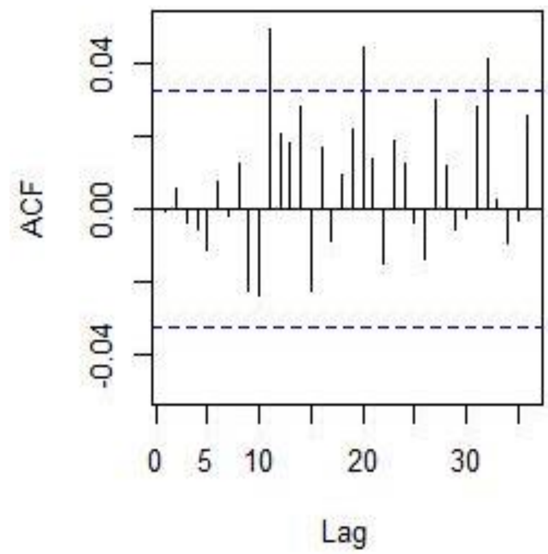


fig.5.20 Residual ACF for ARIMAX(2,1, 1)

ARIMAX (1, 1, 2)

Parameters of the model are good AIC, error matrices like RMSE and MAPE are also better than previous models. Number of spikes outside the limit are also less and value of residual ACF is also low

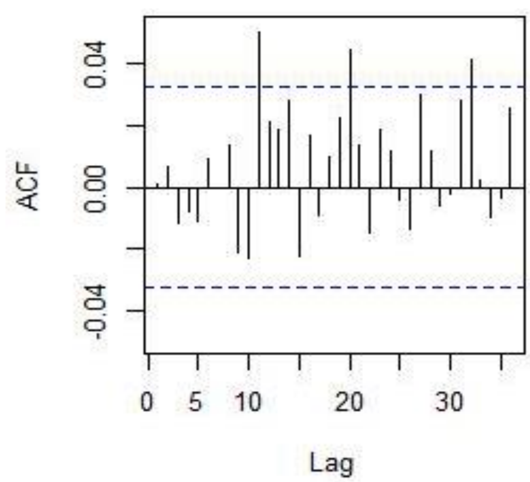


Fig. 5.21 Residual ACF for ARIMAX(1,1,2)

	ARIMAX(1, 1, 2)
Coefficient	Coefficients: ar1 ma1 ma2 xreg 0.5356 -0.6489 -0.0968 -0.0265 Se 0.0476 0.0488 0.0217 0.0024
AIC	12277.46
RMSE	1.298123
MAPE	2.89992
ACF	0.00099994

Table 5.14 statistics for ARIMAX (1, 1, 2) for Maximum temperature at training stage

5.1.1.2. Goodness of fit for maximum temperature at training stage

All the models have performed well at training stage. Parameters of all the models comply t- value. AIC for model (1, 1, 2) was lowest. The value of RMSE for model (1, 1, 1), (2, 1, 1) and (1, 1, 2) was 1.301434, 1.297949 and 1.298123 respectively. MAPE for model (1, 1, 1), (2, 1, 1), and (1, 1, 2) was 2.909321, 2.899442, and 2.89992 respectively. The residual ACF plot for model (2, 1, 1) and (1, 1, 2) were very good. All the three models were tested at validation stage to examine the consistent behavior of model.

5.1.1.3. Parameters and summary of models for Minimum Temperature

ARIMAX (1, 1, 2)

Parameters comply t value. AIC value is also good. AS it is clear from the plot of residual ACF that only one spike is outside the limit rest all spikes are within the band, so ACF plot is also good. The model have shown overall good results hence taken for further process of validation.

Since this model have shown good results, few more nearby models were considered to have the best-fit model.

ARIMAX (1,1, 2)				
Training set				
	Ar1	Ma1	Ma2	Xreg
Coeff	0.5004	-0.6090	-0.1383	0.0081
S.E.	0.0427	0.0435	0.0211	0.0024
t-stat				
AIC = 12205.94				
RMSE = 1.285462				
MAPE = 6.054431				
ACF = 0.001422478				
SSR = 6034.608				

Table.5.15 statistics for ARIMAX (1, 1, 2) for Minimum temperature at training stage.

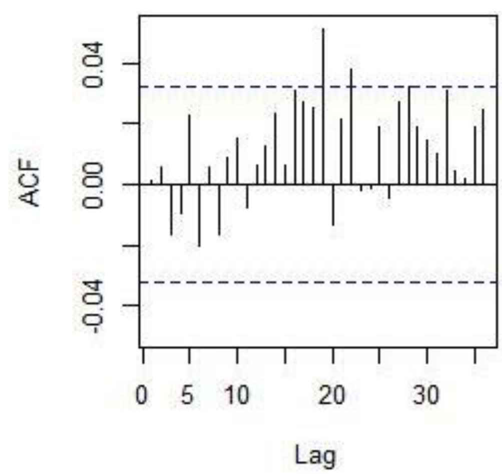


Fig.5.22. Residual ACF for ARIMAX (1,1,2)

ARIMAX (2, 0, 2)

t-value of all the parameters are above 2,so the parameters are good. ACF plot have also shown significant result. Value of residual ACF is 0.0013, which lies within the band and all other spikes are within the band except 19th order lag.

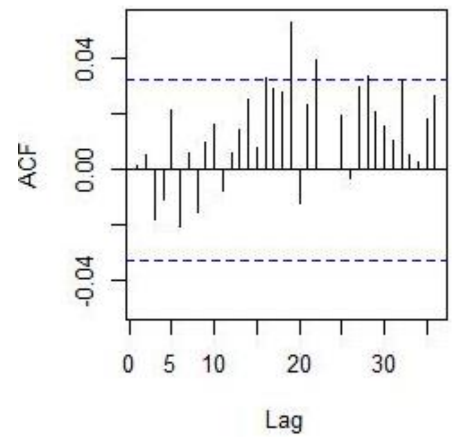


Fig.5.23. Residual ACF for ARIMAX (2,0,2)

ARIMAX (2,0,2)						
Training set						
	Ar1	Ar2	Ma1	Ma2	Intercept	Xreg
Coeff	1.4855	-0.4886	-0.5969	-0.1381	18.8444	0.0081
S.E.	0.0446	0.0442	0.0452	0.0210	1.7869	0.0024
t-stat	33.3	11.05	13.2	6.5	10.5	3.5
AIC=12206.69			ACF1=0.001316086			
RMSE=1.283925		MAPE=6.071131		SSR=6020.187		

Table.5.16 statistics for ARIMAX (2, 0, 2) for minimum temperature at training stage.

ARIMAX (2, 1, 1)

Parameters of model are good. Value of AIC and error matrices are better han above two models. Residual ACF plot is also very good and its value is -0.0004403349, lies within the confidence limit. Overall the performance of the model is better than better than the previous two iterative models

Further to have the better model than this,tried few more models.

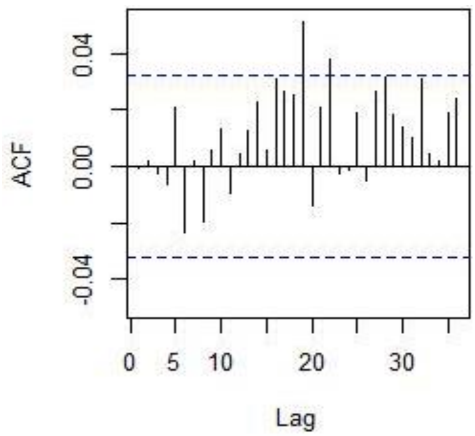


Fig. 5.24. Residual ACF for ARIMAX (2, 1, 1)

ARIMAX(2, 1, 1)				
Training set				
	Ar1	Ar2	Ma1	Xreg
Coeff	0.6578	-0.1185	-0.7647	0.0081
S.E.	0.0305	0.0184	0.0268	0.0024
t-stat	21.5	6.4	28.5	3.3
AIC = 12204.34				
RMSE = 1.285178				
MAPE = 6.053243				
ACF = -0.0004403349				
SSR = 6031.949				

Table.5.17 statistics for ARIMAX (2, 1, 1) for Minimum temperature at training stage.

ARIMAX (3, 0, 0)

performance of the model was okay, as the parameters comply t value,residual ACF is also good and error matrices are okay. Performance of the model is not better than previous three models but all the measures are fulfilling the criteria.

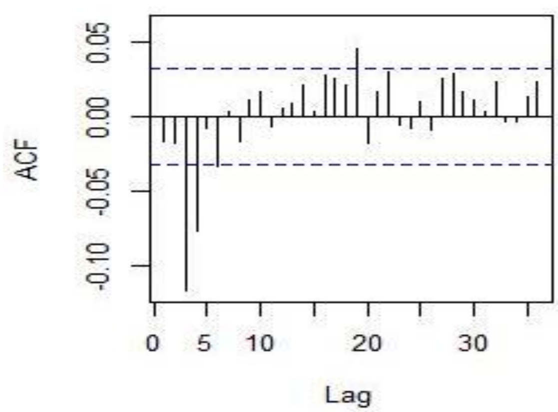


Fig. 5.25. Residual ACF for ARIMAX (3, 0, 0)

ARIMAX(3, 0, 0)					
Training set					
	Ar1	Ar2	Ar3	Intercept	Xreg
Coeff	0.9274	-0.0931	0.1468	18.9836	0.0081
S.E.	0.0164	0.0224	0.0164	1.11701	0.0023
t-stat					
AIC=12305.52			ACF-0.01677992		
RMSE=1.301806		MAPE=6.134001		SSR=6189.037	

Table.5.18 statistics for ARIMAX (3, 0, 0) for minimum temperature at training stage

ARIMAX (3, 0, 1)

Parameters of the model are good. AIC is 12271.79, which is better than model (3, 0, 0) and error measures, RMSE and MAPE are also better. Residual ACF is -0.0109, which lies within `the confidence limit and number of spikes outside the limit are also less.

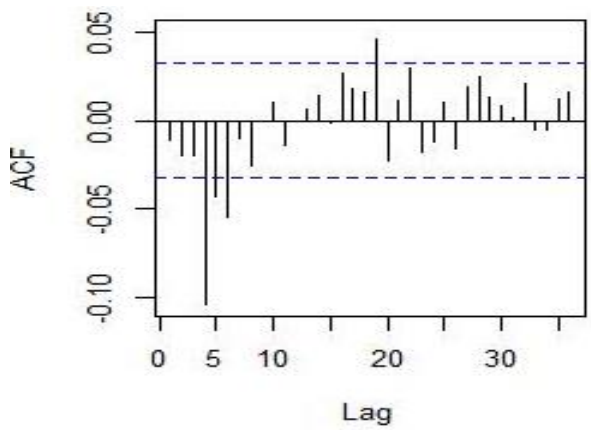


Fig. 5.26. Residual ACF for ARIMAX (3, 0, 1)

ARIMAX(3, 0, 1)				
Training set				
	Ar1	Ar2	Ar3	Xreg
Coefficient	-0.0827	-0.1638	-0.1200	0.0079
S.E.	0.0164	0.0163	0.0164	0.0023
t-stat				
AIC=12271.79		ACF=-0.01098977		
RMSE=1.297123	MAPE=6.077044	SSR=6144.588		

Table.5.19 statistics for ARIMAX (3, 0, 1) for minimum temperature at training stage.

5.1.2. Testing stage

5.1.2.1. Parameters and summary of models for maximum temperature

ARIMAX(1, 1, 1)	**ARIMAX(2, 1, 1)**	**ARIMAX(1, 1, 2)**
ar1 ma1 xreg 0.6656 -0.82 -0.0274 s.e. 0.0000 0.00 0.0000	ar1 ar2 ma1 xreg 0.649 -0.0841 -0.7605 0.0265 Se 0.000 0.0000 0.0000 0.0000	ar1 ma1 ma2 0.5356 -0.6489 -0.0968 se. 0.0000 0.0000 0.00 xreg -0.0265 0.0000
AIC 1272.47	AIC 1269.68	AIC 1269.57
RMSE 1.376979	RMSE 1.371692	RMSE 1.371489
MAPE 3.133679	MAPE 3.113423	MAPE 3.115147
ACF -0.007636464	ACF -0.0520484	ACF -0.04924913

SSR	SSR 688.6434	SSR=688.4391

Table.5.20 statistics for iterative ARIMAX models for maximum temperature at testing stage.

5.1.2.2. Goodness of fit at testing stage

All the models at testing stage were consistent. For model (1, 1, 1) RMSE is 1.376979 and MAPE is 3.133679. Whereas at training stage, this value was 1.301434 and 2.909321 respectively. ACF value is -0.007, which is very small and within the significance level.

Value of RMSE and MAPE for model (2, 1, 1) was obtained as 1.371692 and 3.113423 respectively. Whereas, at training stage this value was 1.297949, and 2.899442 respectively.

For model (1, 1, 2), value of RMSE and MAPE obtained as 1.371489 and 3.115147 respectively. Whereas, at testing stage this value was 1.298123 and 2.89992 respectively. So the values at testing stage have increased but not to great extent. Hence, the performance of models were consistent.

5.1.2.3. Parameters of models for minimum temperature at testing stage

ARIMAX(1, 1, 2)	ARIMAX (2,0,2)	ARIMAX (2, 1, 1)
Test set	**Test set**	**Test set**
ar1 ma1 ma2 0.5004 -0.609 -0.1383 Se 0.0000 0.000 0.0000 xreg 0.0081 0.0000	ar1 ar2 ma1 1.4855 -0.4886 -0.5969 se.0.0000 0.0000 0.0000 ma2 intercept xreg -0.1381 18.8444 0.0081 0.0000 0.0000 0.0000	ar1 ar2 0.6578 -0.1185 se.0.0000 0.0000 ma1 xreg -0.7647 0.0081 0.0000 0.0000
AIC=1237.53	AIC=1243.37	AIC=1236.45
RMSE= 1.312508	RMSE=1.313068	RMSE= 1.310558
MAPE=6.751069	MAPE=6.799312	MAPE=6.730814
ACF1= 0.04841968	ACF1=0.05418438	ACF1=0.04436953
SSR=630.4995	SSR=631.038	SSR=628.6279

Table.5.21 statistics for iterative ARIMAX models for minimum temperature at testing stage.

ARIMAX(3, 0, 0)	ARIMAX(3, 1, 0)
Test set	**Test Set**
ar1 ar2 ar3 intercept 0.9274 -0.0931 0.1468 18.9836 se.0.0000 0.0000 0.0000 0.0000 xreg 0.0081 0.0000	ar1 ar2 ar3 -0.0827 -0.1638 -0.12 s.e.0.0000 0.0000 0.00 xreg 0.0079 0.0000
AIC=1258.04	AIC=1253.15
RMSE=1.339997	RMSE= 1.341068
MAPE=6.879085	MAPE=6.812926
ACF1=0.0471509	ACF1=0.04523215
SSR=657.1866	SSR=658.2378

Continued Table.5.21..

5.1.2.4. Goodness of fit at testing stage

All the models have performed well at testing stage. Error have little increased. RMSE and MAPE value of model (1, 1, 2) at training stage were 1.285462 and 6.054431. Whereas at testing stage it was obtained as 1.312508 and 6.751069 respectively. The residual ACF have increased from 0.001422 to 0.048. For model (2, 0, 2), RMSE, MAPE and residual ACF at training stage were 1.283925, 6.071131 and 0.001316 whereas at testing stage these values have increased to 1.283925, 6.071131 and 0.0541 respectively. The best results were obtained from model (2, 1, 1) compared to all the models at testing stage, RMSE have increased from 1.285178 to 1.310558 at testing stage, value of residual ACF at testing stage was obtained as 0.044, which is within confidence limit. Model (3,0,0) and (3,1,0)have also performed well at testing stage. The best model would be selected based on the minimum AIC, error, and residual ACF within confidence limit.

CHAPTER 6

RESULT & DISCUSSION

6.1. ARIMA MODEL

The order of AR and MA in the range of 1 to 3 with no or regular differencing was applied and among various iterative ARIMA models, candidate models with significant results are and shown here. Performance of the candidate models for maximum and minimum temperature are shown in table 6.1 and table 6.2.

6.1.1. For Maximum Temperature

For maximum temperature series, 3 models with regular differencing (d) and (p), (q) in the range 1 to 2 have shown some notable results. Performance of these tentative models in both training and validation stage for maximum temperature are shown in table 6.1.

Table.6.1 summary of estimation results of tentative 3 ARIMA models for Maximum Temperature

ARIMA(1,1,1)				**ARIMA(2,1,1)**				**ARIMA(1,1,2)**			
Training set				**Training set**				**Training set**			
	Coeff.	S.E.	t-stat		Coeff	S.E.	t-stat		Coeff	S.E.	t-stat
Ar1	0.6750	0.027	24.4	Ar1	0.655	0.034	19.2	Ar1	0.5210	0.0468	11.1
Ma1	-0.8252	0.020	-40.8	Ar2	-0.1027	0.0186	-5.5	Ma1	-0.6204	0.0476	-13
				Ma1	-0.7526	0.0308	-24.4	Ma2	-0.1169	0.0210	5.56
AIC=12423.88				AIC=12396.3				AIC=12397.67			
RMSE=1.325144				RMSE=1.319782				RMSE=1.320031			
MAPE=2.953642				MAPE=2.939284				MAPE=2.940243			
Test set				**Test set**				**Test set**			
Coefficients: ar1 ma1 0.675 -0.8252 s.e. 0.000 0.0000				Coefficients: ar1 ar2 ma1 0.655 -0.1027 -0.7526 s.e. 0.000 0.0000 0.0000				Coefficients: ar1 ma1 ma2 0.521 -0.6204 -0.1169 s.e. 0.000 0.0000 0.0000			

AIC=1276.38	AIC=1273.89	AIC=1273.74
RMSE=1.384389	RMSE=1.379608	RMSE=1.379335
MAPE=3.152277	MAPE=3.125651	MAPE=3.1285561
ACF1= -0.006287662	ACF1= -0.06198128	ACF1= -0.058815
SSR = 701.4508	SSR= 696.6141	SSR =696.3391

The parameter of model (1, 1, 1) are good as t value is within the range and also follows law of parsimony, all other statistics like AIC and error measures of like RMSE and MAPE are also ok. Value of ACF in validation stage is -0.0062, which is also very good. Hence the performance of model (1, 1, 1) was good at both training and validation stage. The two other models (2, 1, 1) and (1, 1, 2) (marginal difference in their statistics) have performed better than model (1, 1, 1) as AIC value of both the models are 12396.3 and 12397.6 respectively, which is less than the AIC value of model (1, 1, 1). Similarly, the error is also less as compared to the model (1, 1, 1), as the RMSE value of the models are 1.319782, and 1.320031 respectively whereas for model (1, 1, 1) it is 1.325144 And the same is in validation stage also

6.1.2. For Minimum Temperature

Model with regular or no differencing (d) and, (p), (q) in the range of 1 to 3 have shown significant results. Among various iterative models, 5 models have shown some notable results and are listed in table 6.2.

Table.6.2 summary of estimation results of tentative 5 ARIMA models for Minimum Temperature

	ARIMA (1, 1, 2)	**ARIMA (2,0,2)**	**ARIMA (2, 1, 1)**	**ARIMA (3,0,0)**	**ARIMA (3, 1, 0)**
	Training set	**Training set**	**Training set**	**Training set**	**Training set**
Ar1	0.4995	1.4836	0.6584	0.9283	-0.0820
(S.E.)	0.0426	0.0446	0.0305	0.0164	0.0164
t-stat	11.7	33.2	21.5	56.6	5
Ar2		-0.4866	-0.1197	-0.0951	-0.1647
(S.E.)		0.0442	0.0184	0.0224	0.0163
t-stat		11	6.5	4.2	10.1
Ar3				0.1478	-0.1205
(S.E.)				0.0164	0.0164

t-stat				9.0	7.3
Ma1	-0.6074	-0.5939	-0.7645		
(S.E.)	0.0434	0.0452	0.0269		
t-stat	13.9	13.1	28.4		
Ma2	-0.1397	-0.1404			
(S.E.)	0.0211	0.0210			
t-stat	6.6	6.6			
Mean	19.0569	1.8058		19.1295	
				1.1215	
AIC	12215.74	12216.6	12214.16	12315.47	12281.22
RMSE	1.287539	1.286019	1.287261	1.303937	1.299155
MAPE	6.068753	6.08717	6.067791	6.147745	6.090528
ACF					
SSR	6054.131	6039.837	6051.509	6209.316	6163.863
	Test Set	**Test set**	**Test set**	**Test set**	**Test set**
Ar1	0.4995	1.4836	0.6584	0.9283	-0.082
Ar2		-0.4866	-0.1197	-0.0951	-0.1647
Ar3				0.1478	-0.1205
Ma1	-0.6074	-0.5939	-0.7645		
Ma2	-0.1397	-0.1404			
Mean		19.0569		19.1295	
AIC	1237.54	1243.48	1236.45	1258.19	1253.26
RMSE	1.312518	1.313249	1.310559	1.34025	1.34127
MAPE	6.752003	6.80077	6.731284	6.880539	6.811527
ACF	0.0490354	0.05465441	0.0449708	0.04798551	0.04583181
SSR	630.5096	631.2117	628.6289	657.4353	658.4355

Among all the 5 tentative models, model (2, 1, 1) have shown best results, as error is least, AIC Is 12214.16, which is also minimum. SSR is 6051.509 which is not least but the second least among the rest of the models so it is acceptable. The minimum value of SSR is for model (2, 0, 2) but rest of the statistics are not as good as in model (2, 1, 1). At validation stage also, the performance of the model (2, 1, 1) was best. SSR value is minimum, ACF is also least compared to rest of the models.

6.2. ARIMAX MODEL

By introducing rainfall as exogenous variable, the results were expected to be improved, and the same was found. ARIMAX was simply applied to those models, which have shown appropriately good results in both training and testing stage of ARIMA.

56.2.1. For Maximum Temperature

The performance of all the models have improved in both training and testing stage and are shown in table 6.3.

Table.6.3 Summary of estimation results of tentative 3 ARIMAX models for Maximum Temperature

ARIMAX(1,1,1)	ARIMAX(2,1,1)	ARIMAX(1,1,2)
Training set	**Training set**	**Training set**
Coefficients: ar1 ma1 xreg 0.6656 -0.8200 -0.0274 s.e. 0.0283 0.0208 0.0024	Coefficients: ar1 ar2 ma1 xreg 0.6490 -0.0841 -0.7605 -0.0265 Se.0.0339 0.0188 0.0304 0.0024	Coefficients: ar1 ma1 ma2 xreg 0.5356 -0.6489 -0.0968 - 0.0265 Se 0.0476 0.0488 0.0217 0.0024
AIC=12294.05	AIC=12276.49	AIC= 12277.46
RMSE=1.301434	RMSE=1.297949	RMSE=1.298123
MAPE=2.909321	MAPE= 2.899442	MAPE=2.89992
Test set	**Test set**	**Test set**
ar1 ma1 xreg 0.6656 -0.82 -0.0274 s.e. 0.0000 0.00 0.0000	ar1 ar2 ma1 xreg 0.649 -0.0841 -0.7605 -0.0265 Se 0.000 0.0000 0.0000 0.0000	ar1 ma1 ma2 xreg 0.5356 -0.6489 -0.0968 - 0.0265 se 0.0000 0.0000 0.0000 0.0000
AIC=1272.47	AIC=1269.68	AIC=1269.57
RMSE=1.376979	RMSE=1.371692	RMSE=1.371489
MAPE=3.133679	MAPE=3.113423	MAPE=3.115147
ACF1= -0.007636464	ACF1= -0.0520484	ACF1= -0.04924913
	SSR=688.6434	SSR=688.4391

ARIMAX model has performed well in both training and validation stage, error measures and AIC values have slightly increased in testing stage. Parameters of all the models are good as t value is greater than 2. Variation in results have been observed at training and testing stages as minimum

AIC, RMSE and MAPE was obtained from ARIMAX (2, 1, 1) at training stage. Whereas minimum AIC and RMSE was found For ARIMAX (1, 1, 2) and minimum MAPE was obtained for model (2, 1, 1) at testing stage. Model (2, 1, 1) and (1, 1, 2) both are better than model (1, 1, 1).

6.2.2 For Minimum temperature

Table.6.4 summary of estimation results of tentative 5 ARIMAX models for Minimum Temperature

Model	ARIMAX (1, 1, 2)	ARIMAX (2,0,2)	ARIMAX (2, 1, 1)	ARIMAX (3,0,0)	ARIMAX (3, 1, 0)
	Training set	Training set	Training set	Training set	Training set
Ar1	0.5004	1.4855	0.6578	0.9274	-0.0827
S.E.	0.0427	0.0446	0.0305	0.0164	0.0164
t-stat	11.7	33.3	21.5	56.5	5.0
Ar2		0.4866	-0.1185	-0.0931	-0.1638
S.E.		0.0442	0.0184	0.0224	0.0163
t-stat		11	6.4	4.1	10
Ar3				0.1468	-0.1200
S.E.				0.0164	0.0164
t-stat				8.9	7.3
Ma1	-0.6090	-0.5969	-0.7647		
S.E.	0.0435	0.0452	0.0268		
T-stat	14	13.2	28.5		
Ma2	-0.1383	-0.1381			
S.E.	0.0211	0.0210			
t-stat	6.5	6.5			
Xreg	0.0081	0.0081	0.0081	0.0081	0.0079
	0.0024	00024	0.0024	0.0023	0.0023
Intercept		18.8444		18.9836	
		1.7869		1.1170	
AIC	12205.94	12206.69	12204.34	12305.52	12271.79
RMSE	1.285462	1.283925	1.285178	1.301806	1.297123
MAPE	6.054431	6.071131	6.053243	6.134001	6.077044
ACF	0.001422478				

SSR	6034.608	6020.187	6031.949	6189.037	6144.588
Test Set		**Test set**	**Test set**	**Test set**	**Test set**
Ar1	0.5004	1.4855	0.6578	0.9274	-0.0827
Ar2		-0.4886	-0.1185	-0.0931	-0.1638
Ar3				0.1468	-0.12
Ma1	-0.609	-0.5969	-0.7647		
Ma2	-0.1383	-0.1381			
Xreg	0.0081	0.0081	0.0081	0.0081	0.0079
Intercept		18.8444		18.9836	
AIC	1237.53	1243.37	1236.45	1258.04	1253.15
RMSE	1.312508	1.313068	1.310559	1.339997	1.341068
MAPE	6.751069	6.799312	6.730814	6.879085	6.812926
ACF	0.04841968	0.05418438	0.04436953	0.0471509	0.04523215
SSR	630.4995	631.038	628.6279	657.1866	658.2378

The performance of ARIMAX model (3, 1, 0) and (3, 0, 0) were not good. As model (3, 1, 0) does not comply t value and for model (3, 0, 0), error was large. For minimum temperature series, minimum AIC and MAPE was found from ARIMAX (2, 1, 1) and minimum RMSE was obtained from ARIMAX (2, 0, 2) at training stage. Whereas at validation stage, all the three measures were minimum for (2, 1, 1).Table 6.20. Shows the minimum values of error measures for maximum and minimum temperature series

Table 6.5. Minimum error measures for maximum and minimum temperature

Series	Coefficients	AIC	RMSE	MAPE
Maximum Temperature	4 IN (111) 6 in others	1269.57 in model (112)	1.3714 in model (112)	3.1134 in model (211)
Minimum Temperature	6 in (2,1,1)	1236.45 in model (211)	1.310558 in model (211)	6.7308 in model (211)

6.3 FORECAST

6.3.1 ARIMAX

Table 6.6 Comparison of forecast values (up to 10 steps ahead) and observed values of maximum temperature series for selected model ARIMAX (1, 1, 2)

S. no.	Time series no.	Actual data	Predicted data	% Forecast error
1	5653	23.95	23.9260	-0.1
2	5654	25.09	24.02723	-4.42
3	5655	26.6	25.04388	-6.21
4	5656	28.37	26.34658	-7.68
5	5657	26.98	27.87175	3.91
6	5658	27.34	26.61801	-2.71
7	5659	26.78	27.15171	1.36
8	5660	26.16	26.65020	1.83
9	5661	24.58	26.17997	6.11
10	5662	23.93	24.81770	3.57

Table 6.7 Comparison of forecast values (up to 10 steps ahead) and observed values of minimum temperature series for selected model ARIMAX (2, 1, 1)

S. no.	Time series sequence No.	Actual data	Predicted data	% Forecast error
1	5653	14.52	14.505496	-0.09
2	5654	11.9	14.426187	17.51
3	5655	9.21	11.963049	23.01
4	5656	9.6	9.791742	1.95
5	5657	9.83	10.315126	4.70
6	5658	12.7	10.315356	-23.11
7	5659	11.92	12.744183	6.46
8	5660	9.67	11.691484	17.29
9	5661	7.84	9.823403	20.19

10	5662	7.64	8.417570	9.23

6.3.2 ARIMA

Table 6.8 Comparison of forecast values (up to 10 steps ahead) and observed values of maximum temperature series for selected model ARIMA (1, 1, 2)

S. no.	Time series no.	Actual data	Predicted data	% Forecast error
1	5653	23.95	23.92605	-0.1
2	5654	25.09	23.97692	-4.43
3	5655	26.6	25.05677	-5.80
4	5656	28.37	26.35307	-7.10
5	5657	26.98	27.87955	3.33
6	5658	27.34	26.57833	-2.78
7	5659	26.78	27.16130	1.42
8	5660	26.16	26.63467	1.81
9	5661	24.58	26.17404	6.48
10	5662	23.93	24.79963	3.63

Table 6.9 Comparison of forecast values (up to 10 steps ahead) and observed values of maximum temperature series for selected model ARIMA (2, 1, 1)

S. no.	Time series sequence No.	Actual data	Predicted data	% Forecast error
1	5653	14.52	14.505480	-0.1
2	5654	11.9	14.441315	21.35
3	5655	9.21	11.961598	29.87
4	5656	9.6	9.793061	2.01
5	5657	9.83	10.318245	4.96
6	5658	12.7	12.748141	0.37
7	5659	11.92	11.690443	-1.92
8	5660	9.67	9.821920	1.57
9	5661	7.84	8.417757	7.36

10	5662	7.64	8.320733	8.91

6.4. COMPARISON BETWEEN ARIMA AND ARIMAX

ARIMAX model have shown better results in almost all the selected models except in model (3, 1, 0) for minimum temperature modelling. Significant difference in error measures were observed between both the ARIMA and ARIMAX models in maximum temperature series whereas this difference was minor for minimum temperature series. In fact, the value of MAPE in model (3, 1, 0) for ARIMAX model was higher than ARIMA model. This difference can be seen in Table 5.20 clearly, negative value of MAPE for (3, 1, 0) indicates that the error was increased. Whereas no change was observed in the AIC value of model (2, 1, 1). Overall, the results were improved compared to the ARIMA

Table 6.10 Increased value of error measures for ARIMAX compared to ARIMA at validation stage

Series	Maximum Temperature			Minimum temperature				
Model	(1,1,1)	(1,1,2)	(2,1,1)	(1,1,2)	(2,0,2)	(2,1,1)	(3,0,0)	(3,1,0)
AIC	3.91	4.17	4.21	0.01	0.11	0	0.15	0.11
RMSE	0.00741	0.007846	.007916	0.00001	0.000181	0.000001	0.000253	0.000202
MAPE	0.018598	0.013409	0.012228	0.000934	0.001458	0.00047	0.001454	-0.001399

6.4.1 Best Model

6.4.1.1 For maximum temperature

For maximum temperature series, the best fit model was (1, 1, 2) for both ARIMAX and ARIMA, but the difference analyzed from both the models can be clearly seen in the table 5.11 AIC value in ARIMAX(1, 1, 2) compared to ARIMA (1, 1, 2). RMSE was 1.298123 in ARIMAX whereas it was 1.320031 in ARIMA. Similarly, MAPE was 2.89992 in ARIMAX and for ARIMA it was 2.940243. The results from ARIMAX was improved and error was less.

Similar results were obtained at testing stage also, ARMAX model performed better than ARIMA. The model was consistent in the performance. The ACF was -0.04924913 for ARIMAX whereas foe ARIMA it was -0.058815, -0.04 is better than 0.05 as smaller spikes are preferred. SSR value was 688.4391 for ARIMAX, less than the value of SSR from ARIMA model.

Table 6.11 comparison of statistics of best-fit model from ARIMAX and ARIMA for maximum temperature

ARIMAX(1,1,2)					**ARIMA(1,1,2)**			
Training set					Training set			
	Ar1	Ma1	Ma2	Xreg		Ar1	Ma1	Ma2
Coeff	0.5356	-0.6489	-0.0968	-0.0265	Coeff	0.5210	-0.6204	-0.1169
S.E.	0.0476	0.0488	0.0217	0.0024	S.E.	0.0468	0.0476	0.0210
t- stat	11.2	13.2	4.4	11.04	t-stat	11.13	13.0	5.56
AIC= 12277.46					AIC=12397.67			
RMSE=1.298123					RMSE=1.320031			
MAPE=2.89992					MAPE=2.940243			
Test set					**Test set**			
Coefficients: ar1 ma1 ma2 xreg 0.5356 -0.6489 -0.0968 -0.0265 se 0.0000 0.0000 0.0000 0.0000					Coefficients: ar1 ma1 ma2 0.521 -0.6204 -0.1169 s.e. 0.000 0.0000 0.0000			
AIC=1269.57					AIC=1273.74			
RMSE=1.371489					RMSE=1.379335			
MAPE=3.115147					MAPE=3.1285561			
ACF1= -0.04924913					ACF1= -0.058815			
SSR=688.4391					SSR =696.3391			

6.4.1.2. For minimum temperature

For minimum temperature series, model (2, 1, 1) performed better than the rest of the iterative models for both ARIMA and ARIMAX. Results from ARIMAX was better than ARIMA model, as the value of AIC, error matrices and ACF were less in ARIMAX(2, 1, 1) compared to the AIC, error matrices, and ACF in ARIMA (2, 1, 1).

Similarly, at testing stage also the results from ARIMAX was improved. However, the improvement is minor, but better than ARIMA. The value of AIC and RMSE for both the models were same. Whereas there was a minute improvement in the values of MAPE, ACF and SSR from ARIMAX model.

Table 6.12 comparison of statistics of best-fit model from ARIMAX and ARIMA for minimum temperature

ARIMAX (2, 1, 1)					ARIMA (2, 1, 1)			
Training set					**Training set**			
	Ar1	Ar2	Ma1	Xreg		Ar1	Ar2	Ma1
Coeff	0.6578	-0.1185	-0.7647	0.0081	Coeff	0.6584	-0.1197	-0.7645
S.E.	0.0305	0.0184	0.0268	0.0024	S.E.	0.0305	0.0184	0.0269
t-stat	21.5	-6.4	-28.5	3.375	t-stat	21.5	-6.5	-28.4
AIC=12204.34					AIC=12214.16			
RMSE= 1.285178					RMSE=1.287261			
MAPE=6.053243					MAPE=6.067791			
SSR=6031.949					SSR=6051.509			
Test set					**Test set**			
	ar1	ar2	ma1	xreg		ar1	ar2	ma1
	0.6578	-0.1185	-0.7647	0.0081		0.6584	-0.1197	-0.7645
se.	0.0000	0.0000	0.0000	0.0000	s.e.	0.0000	0.0000	0.0000
AIC=1236.45					AIC=1236.45			
RMSE= 1.310558					RMSE=1.310559			
MAPE=6.730814					MAPE=6.731284			
ACF1=0.04436953					ACF1=0.0449708			
SSR=628.6279					SSR=628.6289			

CHAPTER 7

CONCLUSION & FURTHER IMPROVEMENT

1. Among various candidate models for maximum temperature, ARIMA models (1, 1, 1), (2, 1, 1) and (1, 1, 2) have performed well at training stage. These models were examined further at validation stage to have the best-fit model and ARIMA model (1, 1, 2) have shown best results with RMSE=1.379335, AIC=1273.74, MAPE= 3.1285561. Similarly for minimum temperature, among various candidate models (1, 1, 2), (2, 0, 2), (2, 1, 1), (3, 0, 0), and (3, 1, 0) at training stage, ARIMA model (2, 1, 1) have shown best results, with AIC = 1236.45, RMSE = 1.310559, and MAPE = 6.731284.
2. ARIMAX models (1, 1, 1), (2, 1, 1) and (1, 1, 2), for maximum temperature were considered at training stage. At testing stage, model (1, 1, 2) have shown best results with statistics RMSE=1.3714, AIC=1269.57, MAPE=3.115147. Whereas for minimum temperature, among various ARIMAX models (1, 1, 2), (2, 0, 2), (2, 1, 1), (3, 0, 0), and (3, 1, 0) model (2, 1, 1) have shown best results with AIC = 1236.45, RMSE = 1.310559, and MAPE = 6.731284.
3. The result revealed that both ARIMA and ARIMAX models were suitable for estimating maximum and minimum temperature. For Maximum temperature, values from ARIMAX model were better than the values from ARIMA model while for minimum temperature, ARIMAX was slightly better than ARIMA. It has seen from the comparison of AIC, RMSE and MAPE that the performance of ARIMAX was better than ARIMA model. All the error matrices from ARIMAX were less than ARIMA. Accuracy of forecasting from ARIMAX is more than ARIMA.
4. However, ARIMA is capable of capturing rainfall effects but considering rainfall as an exogenous variable has improved the efficiency of model.
5. In this study, daily rainfall data was considered because of which the exact timing of rainfall event was not known and due to this lack of information, the results were not accurate though the results were improved from ARIMAX compared to ARIMA model.

6. Model (1, 1, 2) for maximum temperature and model (2, 1, 1) for minimum temperature were observed as the best fit model for Bairagarh station, Bhopal.
7. Resullts might be improved if seasonality is considered instead of regular differencing.
8. Further work should expand like use of recursive adaptive models based on artificial intelligence method.

References

[1]. Anggreni W ,Vinarti RA, Kurniawati YD (2015) Performance comparison between arima and arimax method in Muslim Kids clothes demand forecasting: case study; ARIMA model. Civ Environ Res 7:69-77

[2]. Anitha K., Boiroju N.K., and Reddy P.R., 2014. Forecasting of monthly mean of maximum surface air temperature in India. Int. J. Statistika Mathematika, 9(1), 14-19.

[3] ARIMAX software. Available: https://www.r-exercises.com/2017/05/05/forecasting-arimax-model-exercises-part-5.

[4] B. Agyemang."Autoregressive Integrated Moving Average (ARIMA) Intervention analysis model for the major crimes in Ghana,M.Phil" Dept. Applied Mathematics, Kwame Nkrumah Univ., Kumasi, Ashanti, 2012.

[5] Balyani Y., Niya G.F., and Bayaat A., 2014. A study and prediction of annual temperature in shiraz using ARIMA model. J. Geographic Space, 12(38), 127-144.

[6] Box GE, Jenkins GM, Reinsel GC, Ljung GM (2015) Time series analysis: forecasting and control. John Proc Comput Sci 72:630-637

[7] EI-Mallah E.S. and Elsharkawy S.G., 2016. Time series modelling and short term prediction of annual temperature trend on coast Libya using box-jenkins ARIMA model. Advances res., 6(5), 1-11.

[8] Jalalkamali A, Moradi M, Moradi N (2015) Application of several artificial intelligence models and ARIMAX model for forecasting drought using the standardized precipitation index. Int Environ Sci Technol 12:1201-1210.

[9] K.Goswami. (). " Monthly temperature prediction based on Arima Model: A case study n Dibrugarh station of Assam, India International journal of advanced research in computer science vol. 8, No. 8. Available: https://www.researchgate.net/publication/320691511_MONTHLY_TEMPERATURE_PREDICTION_BASED_ON_ARIMA_MODEL_A_CASE_STUDY_IN_DIBRUGARH_STATION_OF_ASSAM_INDIA.

[10] Khediri S., 2014. Forecasting temperature record in PEI, Canada. Letters in Spatial and Resources Sciences, 9, 43-55, doi 10.1007/s12076-014-0135-x.

[11] Muhammet B., 2012. The analyse of precipitation and temperature in Afyonkarahisar (Turkey) in respect of box-jenkins technique. J. Academic Social Sci. Studies, 5(8), 196-212.

[12] NASA's POWER data access viewer. Available: https://power.larc.nasa.gov/data-access-viewer.

[13] Pankratz A (2012) Forecasting with dynamic regression models. Wiley, New York

[14] Time Series Analysis, "Lecture notes on Univariate Time Series Analysis and Box Jenkins Forecasting".

[15] Wikipedia. Available: https://en.wikipedia.org/wiki/Bairagarh

[16] Wikipedia. Available: https://en.wikipedia.org/wiki/Bhopal

APPENDIX I

Models for Maximum Temperature

Models	Coefficients	AIC	RMSE	MAPE	ACF1
ARIMA(1,0,0)	ar1 mean 0.9695 32.8371 s.e 0.0040 0.7212	12522.05	1.34214	2.993457	-0.03993024
ARIMA(1,1,0)	ar1 -0.0560 s.e. 0.0165	12558.71	1.350233	2.998419	-0.00771173
ARIMA(1,1,1)	ar1 ma1 0.6750 -0.8252 S.e 0.0276 0.0202	12423.88	1.325144	2.953642	0.05222951
ARIMA(1,0,1)	ar1 ma1 mean 0.973 -0.0587 32.8374 s.e 0.004 0.0202 0.7647	12515.52	1.340573	2.995612	0.00782085
ARIMA(2,1,1)	ar1 ar2 ma1 0.655 -0.1027 -0.7526 Se0.034 0.0186 0.0308	12396.3	1.319782	2.939284	-0.000799672
ARIMA(1,1,2)	ar1 ma1 ma2 0.5210 -0.6204 -0.1169 s.e. 0.0468 0.0476 0.0210	12397.67	1.320031	2.940243	0.001262538
ARIMA(4,1,4)	Ar1 ar2 ar3 -0.5339 0.3490 0.4031 s.e. 0.2971 0.2322 0.0736	12398.3	1.31833	2.939418	0.001413687

APPENDIX II

Models for Minimum Temperature

Models	Coefficients	AIC	RMSE	MAPE	ACF1
ARIMA (1,1,1)	ar1 ma1 0.6657 -0.8341 s.e. 0.0245 0.0170	12253.32	1.294545	6.112144	0.06158152
ARIMA(2,1,1)	ar1 ar2 ma1 0.6584 -0.1197 -0.7645 s.e. 0.0305 0.0184 0.0269	12214.16	1.287261	6.067791	-0.0003633838
ARIMA(1,1,2)	ar1 ma1 ma2 0.4995 -0.6074 -0.1397 s.e. 0.0426 0.0434 0.0211	12215.74	1.287539	6.068753	0.001420547
ARIMA(2,0,2)	ar1 ar2 ma1 1.4836 -0.4866 -0.5939 s.e. 0.0446 0.0442 0.0452 ma2 mean -0.1404 19.0569 0.0210 1.8058	12216.6	1.286019	6.08717	0.0009850433
ARIMA(3,0,0)	ar1 ar2 ar3 0.9283 -0.0951 0.1478 s.e 0.0164 0.0224 0.0164 mean 19.1295 1.1215	12315.47	1.303937	6.147745	-.01699626
ARIMA(3,1,0)	ar1 ar2 ar3 -0.0820 -0.1647 -0.1205 s.e. 0.0164 0.0163 0.0164	12281.22	1.299155	6.090528	-0.0108259
ARIMA(3,1,1)	Ar1 ar2 ar3 0.6544 -0.1183 -0.0039 s.e. 0.0375 0.0198 0.0204 ma1 -0.7609 0.0337	12216.12	1.287254	6.067962	-0.00005292
ARIMA (2,1,2)	ar1 ar2 ma1 0.6852 -0.1385 -0.7917 s.e.0.1489 0.1033 0.1501 ma2 0.0234 0.1271	12216.12	1.287254	6.067972	-0.00004862
ARIMA(4,0,0)	Ar1 ar2 ar3 0.9115 -0.0843 0.0427 s.e. 0.0164 0.0223 0.0223 ar4 mean 0.1134 19.1037 0.0165 1.2614	12270.32	1.295536	6.111718	-0.009776578
ARIMA(4,0,1)	Ar1 ar2 ar3 8.6369 -0.7576 0.1121 s.e. 0.0409 0.0468 0.0332	12216.71	1.285684	6.085585	-0.000164288

	ar4 ma1 mean 0.0057 -0.7460 18.9484 0.0205 0.0375 1.7827				
ARIMA(4,1,1)	Ar1 ar2 ar3 0.6566 -0.1179 -0.0044 s.e. 0.0431 0.0203 0.0210 ar4 ma1 0.0019 -0.7630 s.e. 0.0198 0.0398	12218.11	1.287253	6.067714	-0.000082638