CHAPTER 5

Linear Statistical Processes in Studying the Behavior of Selected Meteorological Variables in NE India

Chapter 5

Linear Statistical Processes in Studying the Behavior of Selected Meteorological Variables in NE India

Contents

5.1 Introduction

Model fitting to a time series data is necessary when it comes to time series analysis. It is established well that the methods involved in fitting a specific model to such data series relies on two assumptions: viz. i) the time series is assumed to be stationary, i.e., the statistical properties of it being constant over time (or, to be converted to stationary if not by transformations such as differencing), and ii) that the time series follows a linear model, i.e., the values of it can be represented as linear combinations of present and past values of a strictly random or independent series [1]. Thus, the state of the statistical properties of a time series manifests the behavior of the said time series. This can be described with the help of the time series processes, e.g., white noise processes

(e_t), that merely consists of a sequence of independent random variables, with constant mean (μ) and variance (σ^2) over each time point in stationary condition; autoregressive (AR) processes, where the current observation (x_t) of a time series (X) is assumed to be regressed over the previous observations (x_{t-1}, x_{t-2},…, x_{t-n}); moving average (MA) processes, where the x_t of a X is assumed to be dependent on the current and past error terms (ε_t, ε_{t-1}, ε_{t-2},…, ε_{t-n}) of X, or a mixture of both the AR and MA processes, called as mixed autoregressive moving average (ARMA) process, that is known to lead to more parsimonious model (model with the ability to adequately describe X) than can be achieved by the pure AR or MA processes [1,2]. Time series may be non-stationary in nature, i.e., exhibiting mean and variance that changes over different time points of the time series. In such cases, applying differencing and incorporating the differenced term *d* into an ARMA process gives rise to a new approach of time series analysis called autoregressive integrated moving average, ARIMA [2, 3]. The ARIMA processes become weakly stationary after *d* is applied [4].

Often the meteorological variables in the climate system such as temperature, wind speed etc. produce non-stationary time series. The non-stationarity in the meteorological time series is induced by the characteristics such as trend and variability [5]. Research around the globe has been incorporating trend and variability assessments to the weather forecasting of meteorological variables to co-operate in decision and policy making for an area. In such scenario, ARIMA is widely used in different regions of the globe. For example, Edwards [6] developed ARIMA models to study the variability in temperature and rainfall, by simulating trends in these meteorological parameters on a regional basis over selected regions of New Zealand. In this approach the ARIMA generated data was found to fit good with both the actual temperature and rainfall datasets for most of the regions. The different best fit ARIMA models were found to be able to predict rainfall and temperature trends likely for the next hundred years, with no significant difference in variability between the observed vs predicted data. Kim et al. [7] applied ARIMA in studying the spatio-temporal characteristics of rainfall over Mongolia. In this study, $1^o \times 1^o$ gridded precipitation data for summer season were fitted and predicted. The summer precipitation mean intensity for the period 2008-2029 was predicted to be 28.2 mm. From the predicted projections it was found that the summer (June-August) rainfall followed a declining trend in the North-East, North-West, South-East and South-West regions, with a decrease in rainfall by

1.3 mm y^{-1}, 2.7 mm y^{-1}, 1.7 mm y^{-1} and 2.7 mm y^{-1} respectively. Cadenas et al. [8] used hourly average and a ten-minute average meteorological data, such as, wind speed (ms^{-1}), wind direction (^{o}), barometric pressure (mbar), air temperature (^{o}C), solar radiation (W m^{-2}), relative humidity (%) over two places of Mexico, to compare the impact of different meteorological variables on the performance of multivariate model (nonlinear autoregressive exogeneous artificial neural network, NARX) of wind speed forecasting and high performance univariate linear model (ARIMA). In this study the univariate ARIMA model was found to bring reasonable one-step ahead prediction of wind speed. However, NARX model was found to be better at performance than that of the ARIMA model. In another study by Balasmeh et al. [9], ARIMA model was used to predict the changes in rainfall over Wadi Shueib catchment area in Jordan. The best fit models, validated with a ten-year data for the period 2007-2016 were used in rainfall forecasting up to the year 2026, which was followed by trend projection in the precipitation records. Based on the monthly and seasonal analysis, the ARIMA(3,1,3), ARIMA(4,1,3) and ARIMA(4,2,4) were the best fitted models for monthly, average and seasonal data records for different stations. The ARIMA models were statistically significant at 5% significance level.

Large water bodies such as dams may lead to an increase in the levels of atmospheric moisture, thus effecting the local climate, in terms of increasing or decreasing trends in the relative humidity around the dam. This importance of study pertinent to relative humidity over a dam area was investigated by Eymen and Köylü [10]. They tried to identify significant trends in seasonal meteorological time series (maximum, minimum and average relative humidity, and average wind speed) for the area nearby the Yamula Dam, Turkey. The trend identification was followed by prediction of the average relative humidity over the region using ARIMA. For the prediction, the time series was split into two parts: pre dam (1970-2004) and post dam (2005-2014) period. The pre-dam series was used in predicting the post dam time series and validating the model performance by ARIMA. In this study, ARIMA was found to be a strong technique in predicting relative humidity.

Statistical time series forecasting model was developed based on ARIMA by Lai and Dzombak [11] for near-term prediction of temperature and precipitation as an efficient alternative tool to GCM. For this purpose, daily time series of different climate indices of precipitation and temperature starting earlier than the year 1900 for 93 cities of U.S.

were used in model building as well as validating. Further on, quantitative analyses on comparison of ARIMA based forecasts with that of the other statistical techniques such the linear trend method were also performed. It was observed that though ARIMA based model could not surpass them, but the ARIMA based forecasting model proved to be effective, explainable, and steadfast for obtaining near-term (2-20 years) forecasts for regional temperature and precipitation. In another study, the seasonal ARIMA (SARIMA) was applied on a 25 year monthly averaged relative humidity data for the period 1984-2010 over North-West Iran by Shiri et al. [12] to study and forecast the changes in the selected parameter up to year 2014. On the basis of the parsimonious model SARIMA$(1,0,1)(1,1,1)_{12}$, the results of the study were indicative of increasing trends in the relative humidity for the months May, June and September.

Literatures are available on ARIMA in studying the behaviour of meteorological variables and forecasting over East Asia [13-15]. Han et al. [13] utilized ARIMA model in forecasting drought over Guanzhong, China. The standardized precipitation index (SPI), data was used as a drought quantifying tool for this region. Here, data from 1966-2003 was utilized for model development and the year 2004 data was used for model validation. It was noticed in this study that ARIMA was proven as a strong tool in short-term forecasting of drought. Likewise, while forecasting the monthly precipitation time series, the influences brought about by the inter-monthly variations within each year is often ignored. An improved seasonal ARIMA model was developed, considering both inter-monthly and inter-annual variations by Wang et al. [14] for the Lanzhou region of China. The seasonal model was improvised by applying cluster analysis on the monthly data series for classifying the data series and subsequent extraction of the characteristic features (maximum, minimum and truncated mean of each series), followed by the build-up of linear regression models to determine the associated parameters for each monthly series. Finally, the seasonal ARIMA model was built for each characteristic feature as identified from the cluster analysis. From the results it was noticed that the accuracy of the improved model was significantly higher than that of the conventional seasonal model (forecasting precision increased by 21%). Li et al. [15] applied comparison study on ARIMA and Long Short-Term Memory (LSTM; a deep learning approach) in forecasting relative humidity for China, taking 300 days daily average relative humidity data from August 1952 to June 1953 in model building, and ten days data from June 19-28, 1953, for checking the model validation. Here, ARIMA(1,0,0)

was found to be the best -fit among the selected ARIMA models, which was used to predict daily average relative humidity over the next ten days. Subsequently, the same was performed by LSTM and Back Propagation (BP) neural network also. The RMSE values of the predicted models were found to be the least in case of ARIMA (5.1) than LSTM (7.1) and BP neural network (6.4). Thus, it was concluded that the prediction effect by ARIMA on relative humidity was better than the selected deep learning methods. However, the shortcomings such as limited experimental environment and problems in choosing proper parameters in neural network were also listed.

In South-East Asia, monthly rainfall records for the period 1980-2010 were used to build a SARIMA model for the prediction of long-term rainfall over the Sylhet region of Bangladesh, by Bari et al. [16]. In this study, data from 1980-2006 were utilized in model building, while data from 2007-2010 were used for checking the model performances. Here, ARIMA$(0,0,1)(1,1,1)_{12}$ was found as the most effective model in forecasting rainfall at 95% confidence level. Another study over Bangladesh, on prediction of atmospheric pollutant was investigated by Shahriar et al. [17]. He assessed the performance of two hybrid models, viz. ARIMA-Artificial Neural Network (ARIMA-ANN) and ARIMA-Support Vector Machine (ARIMA-SVM) along with Decision Tree (DT) and CatBoost deep learning model (Tree based soft computing models) in forecasting ambient $PM_{2.5}$ concentrations over selected regions of Bangladesh from January 2013 to May 2019. The CatBoost deep learning model was found out to be the most effective in prediction, while ANN in combination with ARIMA showed accuracy in prediction for the selected stations. The ARIMA-ANN and deep learning techniques were revealed to be efficient in delivering useful manifestation for early alerts of the $PM_{2.5}$ pollution over the region. In another study, the characteristics of rainfall in terms of variability, anomaly and trend were studied by Dawood et al. [18] for the Hindu Kush region using ARIMA(1,0,0). The analysis on annual average and mean monthly rainfall data displayed the presence of both increasing and decreasing trends at different locations of this region.

In case of the research related to India, Somvanshi et al. [19] used a hundred- and four-years rainfall data (1901-2003) over Hyderabad, India to model the behaviour of the rainfall time series. The first 93 years of average annual rainfall were used for model training, while the rest ten years of data were utilized in forecasting by ARIMA(4,2,1) and ANN(4 0 0). From the comparison of the results, ANN was found to be

outperforming ARIMA and the concluding remark was drawn that ANN could be used as an appropriate prediction method for rainfall in this region. Chattopadhyay and Chattopadhyay [20] developed a univariate model to forecast the monsoonal rainfall (June-August) over India using ARIMA. Here, three models, i.e., ARIMA(0,1,1), ARIMA(0,2,2) and ARIMA(1,1,1) were built, among which the ARIMA(0,1,1) was identified as a suitable representative model. Finally it was concluded that the ARIMA(0,1,1) might be an substitute model for predicting rainfall in monsoon season over India, provided there's no scope for neural network modelling. In another case, Chaudhuri, and Dutta [21] utilised ARIMA in their study with the purpose to provide accurate daily forecasted data on concentration of the meteorological parameters (surface temperature and relative humidity) and air pollutants (SO_2, NO_2, PM_{10}, CO and O_3) over Kolkata. In this study, three ARIMA models, viz. ARIMA(1,1,1), ARIMA(2,1,2) and ARIMA(0,2,2) were built and the performance was assessed at 95% confidence level. The best fit model selection was accomplished with the help of different information criteria. The selection through the validation with the observed data for the year 2012 revealed that ARIMA(0,2,2) was the best among the three models in predicting both pollutant and meteorological time series over Kolkata during 2002 to 2012. In recent past, Hosamane et al. [22] (2020) used ARIMA to predict PM_{10} concentration over SG Hali area of Bangaluru, using daily PM_{10} concentration data from July 2019 to January 2020. The results showed that ARIMA (2,1,3)(1,0,0) was the best-fit model for the prediction over the selected area of Bangalore city. It was concluded that the use of ARIMA was beneficial in reducing the measurement uncertainty associated with PM_{10} concentration.

The seasonal behaviour of rainfall over India has been studied in recent past [23-25]. Narayanan et al. [23] used ARIMA in forecasting rainfall for the period 2010-2030, following the detection of rainfall trends at 10% significance level in pre-monsoon season over the western region of India- a semi-arid region, characterized by intense dust-storms. Here, the forecasted rainfall series showed the signs of significant increase in the pre-monsoonal rainfall over the region. On the other hand, Narayanan et al. [24] used sixty-year continuous rainfall data from 1949-2009 for twenty IMD stations over India to forecast the pre-monsoon rainfall for the period 2010-2030. The results upon comparison with the report by the Indian Network of Climate Change Assessment (INCCA, 2010) showed agreeability in forecasting rainfall for 2030, i.e., the projection

of increase in pre-monsoon rainfall (~5% increase in ensemble average) for stations in a Himalayan region (North-West, West, and entire Indo-Gangetic Plain). ARIMA(4,1,1) and ARIMA(1,2,2) were found to be the parsimonious models in forecasting the pre-monsoon rainfall at two stations of NER of India, namely, Cherrapunji and Guwahati respectively.

Studies are available on the application of SARIMA modelling approach in developing forecasting model for monthly prediction of meteorological variables in India also. In this context, Nirmala and Sundaram [26] utilised monthly total rainfall for a period of 136 years, from 1871 to 2006 over Tamilnadu, India. The authors' concluding remarks was that SARIMA could give better prediction accuracy for monthly rainfall over Tamilnadu, provided that more input parameters such as El Nino Southern Oscillation (ENSO), land surface temperature were available. In another study, an SARIMA model ($SARIMA(1,2,1)(1,0,1)_{12}$) was developed for the prediction of monthly rainfall by Swain et al. [27] over Khordha district, Odisha, India. The model was trained by monthly rainfall data for over 80 years (1901-1982), and the forecasting efficiency was tested for 20 years (1983-2002) by Nash-Sutcliffe efficiency and coefficient of determination, R^2. The results revealed an excellent consistency of the forecasted values for 20 years monthly rainfall with respect to the observed rainfall. Dimri et al. [25] utilised a hundred years (1901-2000) monthly averaged data on precipitation and temperature (maximum and minimum) in forecasting for the next twenty years (2001-2020) using SARIMA over the Bhagirathi River basin, Uttarakhand, India. The best fit models for the rainfall and temperature were found to be $SARIMA(0,1,1)(0,1,1)_{12}$ and $SARIMA(0,1,0)(0,1,1)_{12}$ respectively. The forecasted data seemed to fit well with the trend upon comparison. However, over-prediction was observed in case of extreme rainfall events. In a recent finding, Shad et al. [28] executed SARIMA and ANN with multilayer perceptron (MLP) methods to forecast the monthly relative humidity over Delhi, India for the period 2017 to 2025 using time series data of relative humidity from 2000 to 2016. From the AIC scores the best fit model was found to be $SARIMA(1,0,0)(0,1,1)_{12}$. However, it was concluded that ANN with MLP was more effective in predicting relative humidity for the region.

While talking to the NER, India, rather limited research has been observed pertinent to the behaviour of meteorological variables using statistical approaches such as ARIMA as per the literature survey. Most of the studies were found to be focused on Assam

only. Goswami et al. [29] used SARIMA models were developed in prediction of long-term monthly temperature data over Dibrugarh, India for a period of fifty years (1966-2015). (2017). They developed two seasonal ARIMA models, SARIMA$(2,1,1)(0,1,1)_{12}$ and SARIMA$(2,1,1)(0,1,1)_{12}$ to forecast temperature (maximum and minimum temperature respectively) over Dibrugarh, which seemed effective in forecasting temperature for this region. In 2018, Murthy et al. [30] adopted SARIMA in forecasting monsoonal rainfall in NER and concluded that SARIMA$(0,1,1)(1,0,1)_4$ was appropriate in forecasting rainfall during monsoon except for some extreme values. Das et al. [31] studied the temporal variation of temperature in Guwahati, Assam with the application of SARIMA. The selected models SARIMA$(3,0,3)(1,1,2)_{12}$ and SARIMA$(3,0,2)(1,1,0)_{12}$ for maximum and minimum temperature respectively. The models forecasted the respective time series for ten years at 5% significance level, which was concluded as in very good agreement with the previously recorded findings. Among the recent findings, Kumar et al. [32] and Barman et al. [33] applied SARIMA in monthly analysis of air pollution and rainfall over Assam respectively. The study by Kumar et al. [32] was a comparative account of performance of models on SO_2, NO_2 and RSPM, generated by a number of machine learning algorithms and SARIMA$(3,1,3)(1,1,1)_{12}$ for 17 years (2003-2019) over selected districts of Assam. It was seen that among ARIMA approach outperformed the machine learning approaches. However, it was concluded that the prediction was poor by all the methods. On the other hand, Barman et al. [33] utilised monthly rainfall data for the period 1901-2017 in forecasting by both ARIMA and SARIMA. The performance of model by SARIMA was found to be best fit for Assam in this study, with the suggestion for the scope of support vector regression (SVR), ANN and recurrent neural network (RNN) in the same. Thus, the studies pertinent to ARIMA is limited with lack of focus on other meteorological variables governing the meteorology of NER, in a changing climate scenario.

Therefore, in this chapter, we've tried to study the behaviour of rainfall as well as the other meteorological variables as mentioned in Chapter 2, using a linear statistical method, ARIMA as follows. All the analyses were carried out using EasyReg package under R programming language.

5.2 Methods

In the following sub-sections, a detailed account on the methods of ARIMA modelling is described. For the whole analysis, monthly time series of rainfall, temperature, relative humidity, sea level pressure and wind speed were considered and analysed as per the following.

5.2.1 Stationarity testing

The Augmented Dickey-Fuller (ADF) test [34] was applied on the time series to check for the presence of unit roots. This test is used as a tool for checking stationarity in time series [35-38]. The ADF test uses a high order autoregressive process to examine the unit roots present in the time series data. The testing procedure is like Dicky Fuller test except lags with respect to the variable that were incorporated in the model [39-41].

$$\Delta Y_t = \alpha + \beta t + \gamma_i Y_{t-1} \sum_{i=1}^{n} \delta_i \Delta Y_{t-i} + \varepsilon_t$$

Here,

ΔY_t = difference operator

α = intercept constant

β = coefficient on a time trend

γ = coefficient presenting process root

n = optimal lagged length

δ_i = time trend

ε_t = independently and identically distributed sequence of random variables

The chances to reject the null hypothesis of presence of a unit root in the time series becomes more with the more negative values of the test statistics of the ADF test applied on that time series.

5.2.2 Selection of lag order

Autocorrelation Function and Partial autocorrelation function

The correlation between two random variables, W and Z, is defined as [2]:

$$\rho_{WZ} = \frac{Cov(W,Z)}{\sqrt{V(W)V(Z)}}$$

The autocorrelations at lag k refers to the correlation between any two observations in a time series that are k periods apart. Thus, autocorrelations at lag k can be defined as,

$$\rho_k = \frac{Cov(x_t, x_{t+k})}{\sqrt{(V(x_t) \cdot V(x_{t+k})}} = \frac{\gamma_k}{\gamma_0}$$

The graphical representation of ρ_k vs. the lag k is called autocorrelation function (ACF) of the process and is denoted by $\{\rho_k\}$. The autocorrelation functions characterize its respective time series models. It is dimensionless and always $-1 \leq \rho_k \leq 1$.

Another concept in the description of time series models is partial correlation.

Let's consider three random variables *W, Y* and *Z*. The conditional distribution of *W* and *Y* given *Z*, when the joint density function of *W, Y* and *Z* be *f* (w, y, z), can be defined as,

$$h(w, y \mid z) = \frac{f(w, y, z)}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(w, y, z) dw\, dy}$$

The correlation coefficient between *W* and *Y* in the conditional distribution *h* (*w, y* | *z*) is defined as the partial (or conditional) correlation coefficient.

In case of a time series, it is convenient to think of the partial autocorrelation at lag k as the correlation between x_t and x_{t+k} with the effects of the intervening observations ($x_{t+1}, x_{t+2}, \ldots, x_{t+k-1}$) removed. A graphical representation of the partial autocorrelation ϕ_k vs. lag k is called partial autocorrelation function (PACF) and is denoted by $\{\phi_k\}$. It should be noted that $\phi_{00} = \rho_0 = 1$ and $\phi_{11} = \rho_1$.

5.2.3 Autoregressive Integrated Moving Average model

The Autoregressive Integrated Moving Average model (ARIMA) has been used in forecasting hydro-meteorological variables [20][23][42-43].

An autoregressive process combined with a moving average process of order (*p,q*) is called an autoregressive-moving average process of order (*p,q*) [ARMA(*p,q*)], can be expressed as follows:

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \cdots + \theta_p y_{t-p} + v_t + \alpha_1 v_{t-1} + \cdots + \alpha_q v_{t-q} \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Equation (1) can be rewritten alternatively as

$$(1 - \theta_t L - \cdots - \theta_p L^p) y_t = (1 + \alpha_1 L + \cdots + \alpha_q L^q) v_t \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

To ensure both stationarity and invertibility of this process it is required that

$$\theta_p(z) = 1 - \theta_1 z - \cdots - \theta_p z^p \neq 0 \;\; for \; |z| \leq 1$$

and

$$\alpha_q(z) = 1 + \alpha_1 z + \cdots + \alpha_q z^q \neq 0 \;\; for \; |z| \leq 1$$

In case of a non-stationary AR process of the form $y_t = y_{t-1} + v_t$, it is easy to transform y_t such that a stationary process results by simply considering the first differences $z_t = y_t - y_{t-1} = (1 - L) y_t$. Generally saying, if the AR operator *θ(L)* of the ARMA process y_t has unit roots [i.e., $\theta(1) = 0$], these can be removed by differencing, thereby making the time series stationary. Considering *d* unit roots, a general process is obtained:

$$\theta_p(L)(1 - L)^d y_t = \alpha_q(L) v_t \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

where, v_t is the zero mean white noise.

Equation (3) is called an autoregressive integrated moving average process of order (*p, d, q*), abbreviated as ARIMA(*p, d, q*) [44]. It is a general model capable of representing a wide class of nonstationary time series. The AR part of ARIMA (*p*) depicts that the time series is regressed on its own past data value, while the MA part (*q*) represents that the forecast error is a linear combination of respective past errors. Lastly, the I part (*d*) integrates these two processes together by differencing (replacing the original series y_t by a new series y'_t, calculating the change between two consecutive data points in the y_t, so that $y'_t = y_t - y_{t-m}$), smoothing the series and making it stationary. Thus, the prediction is the differenced y_t in the d^{th} order [45].

In our study, ARIMA modelling approach was applied on the meteorological variables rainfall, maximum (MaxT) and minimum temperature (MinT), relative humidity (RH), sea level pressure (SLP) and wind speed (WS). A flow chart showing the model construction is depicted in Figure 5.1.

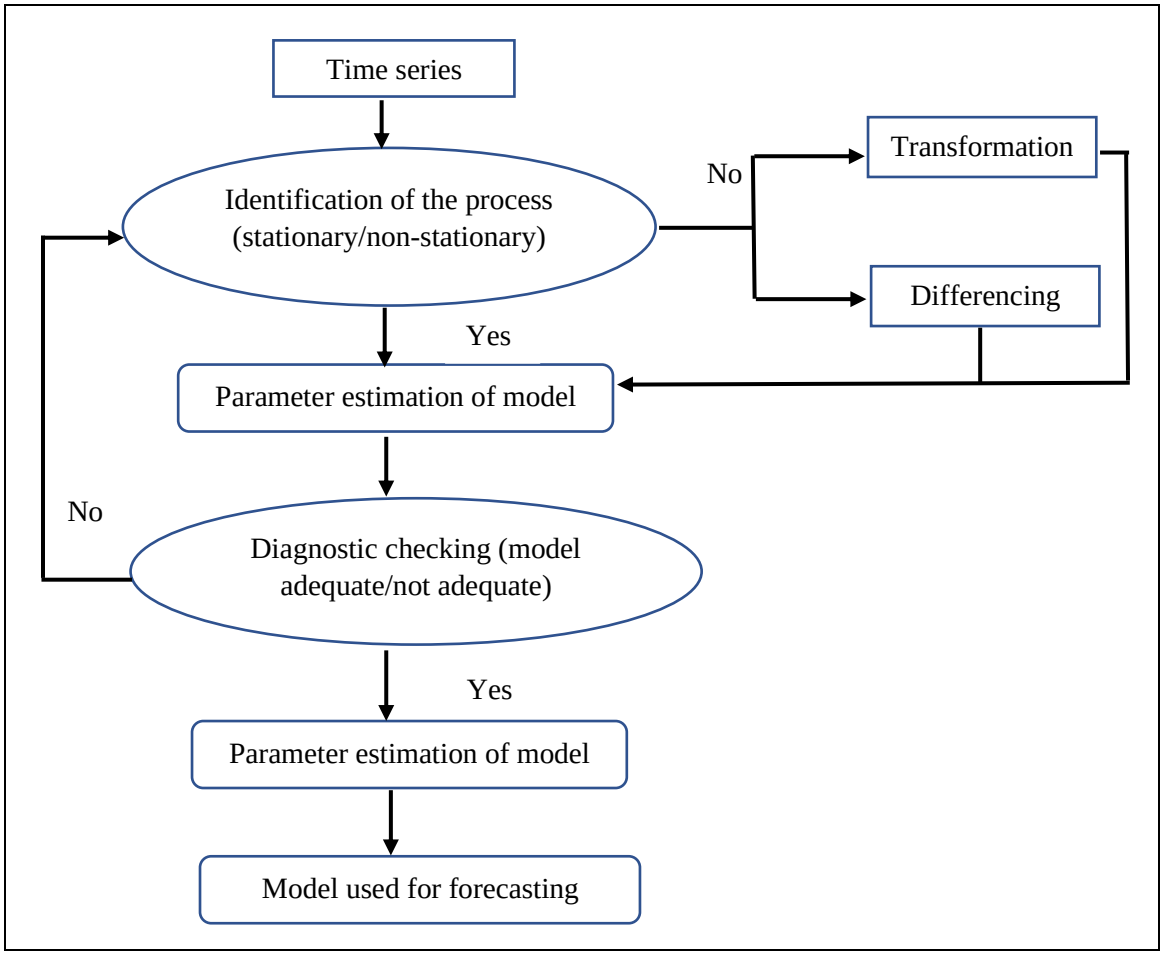


Figure 5. 1 Flow chart of the steps involved in ARIMA model implemented in the study

Process identification

The prior requisite to fit an ARIMA model in a time series is the data to be stationary (constant variance and mean). Stationarity could be attained in the variance by having log transformation and differencing of the original data to attain stationarity in the mean. In case of a non-stationary data series, a seasonal first difference (D = 1) of the original data is to be performed to obtain stationarity.

Models' parameter estimation

Since the orders *p, d,* and *q* are necessary to adequately model for a time series process, it is required to determine the model that best fits the data based on the observations of the ACF and the PACF plots of the differenced data. Alternatively, different information criteria can also be utilized for this task. Based on these, several models may be estimated for the specific time series dataset.

In our study, the stationarity tests were applied and a few ARIMA models were fitted to the data based on the ADF and PP test as well as from the ACF and PACF results. A

total of 24 models with different combinations of (*p, d, q*) were fitted to the test data and the parameters were estimated.

Diagnostic Checking of the fitted models

Initialization of several diagnostic checking of the fitted models is done once the models were fitted to the time series. If the model fits well, the residuals should be uncorrelated with constant variance. Moreover, in developing model this is often assumed that the errors are normally distributed. Therefore, it is expected for the residuals to be normally distributed. Standard checks for ARIMA are to compute the ACF and the PACF of the residuals. Further diagnostics checking can be done by looking at the residuals in various ways. If the residuals are normally distributed, they should all lie on a straight upward sloping line [46]. A fitted model transforms the observations to a white noise process at last, if it is suitable [2]. Different information criterions may be used as a tool for diagnostic checking also, among which the popular ones are the Akaike Information Criterion (AIC), Hannan-Quinn Information Criterion (HIC) and Schwarz Information Criterion (SIC, or otherwise known as Bayesian Information Criterion, BIC). In case that all the models under evaluation if fit poorly with respect to a given set of observations or data, the lowest values of these information criterions are taken into consideration to indicate the best suitable among them.

The mathematical equations governing these information criterions are as follows [47]:

$$AIC = 2 \cdot \frac{k}{n} - 2 \cdot \frac{l}{n}$$

$$HIC = 2 \cdot \frac{k \cdot \ln(\ln n)}{n} - 2 \cdot \frac{l}{n}$$

$$SIC = \frac{k \cdot \ln n}{n} - 2 \cdot \frac{l}{n}$$

Where,

n = number of observations,

k = number of estimated parameters in the model

l = the log likelihood function (assuming normally distributed errors), which is determined as follows:

$$l = -\frac{n}{2} \cdot \left(1 + \ln(2 \cdot \pi) + \ln\left(\frac{1}{n} \cdot \sum_{i=1}^{n} (y_i - \hat{y}_i)^2\right)\right)$$

$$l = -\frac{n}{2} \cdot \left(1 + \ln(2 \cdot \pi) + ln\left(\frac{1}{n} \cdot \sum_{i=1}^{n} (y_i - \hat{y}_i)^2\right)\right)$$

In this study, the model with the lowest value of among AIC, HIC or SIC was selected to be the best fitted model for the selected meteorological time series. The performance of the modelled forecasted series was checked with the help of scatter plots. All the statistical analyses in this study were performed on RStudio and EasyReg software.

5.3 Results and discussions

The monthly time series of different meteorological variables were used in this study. The results are described in the following sub-headings:

5.3.1 Stationarity checking

The results of ADF test are elaborated in table 5.1. It was seen from the results that the ADF test statistics in majority of the meteorological variables across the selected locations of NER were positive. Only rainfall time series at CHR. DBR, GHY and TUL, MinT at CHR, RH at CHR and DBR, SLP at TUL and WS at CHR, GHY, KSH and TUL were displaying negative ADF statistics. After calculation of the p values, it was seen that none of the ADF statistics values of the time series were significant (at 5% significance level). Therefore, from the results it can be said that all the time series of meteorological variables across the selected locations of NER were non-stationary in nature.

Table 5. 1 The ADF test results performed on the monthly time series of meteorological variables in different locations of NER. Here, Variable_site denotes the meteorological variable per studied location in NER.

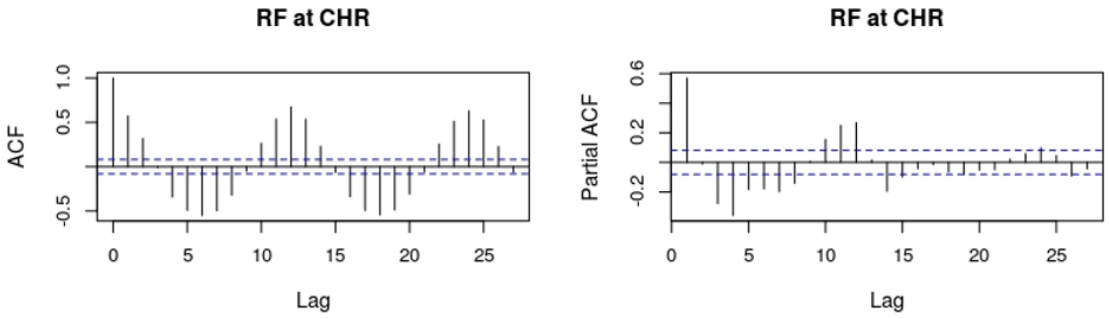
Variable_site	ADF		Variable_site	ADF	
	Statistics	p		Statistics	p
RF_CHR	-0.847	0.367	***RH_CHR***	-0.254	0.400
RF_DBR	-0.576	0.600	***RH_DBR***	-0.226	0.400
RF_GHY	-0.558	0.500	***RH_GHY***	0.217	0.800
RF_KSH	0.112	0.800	***RH_KSH***	0.303	0.400
RF_TUL	-0.121	0.700	***RH_TUL***	0.575	0.500
MaxT_CHR	0.224	0.600	***SLP_CHR***	0.358	0.500
MaxT_DBR	0.224	0.600	***SLP_DBR***	0.159	0.400
MaxT_GHY	0.468	0.700	***SLP_GHY***	0.359	0.800
MaxT_KSH	0.330	0.700	***SLP_KSH***	0.128	0.700
MaxT_TUL	0.287	0.700	***SLP_TUL***	-0.149	0.400
MinT_CHR	-0.243	0.600	***WS_CHR***	**-3.651**	**0.100**
MinT_DBR	0.382	0.700	***WS_DBR***	0.346	0.800
MinT_GHY	0.733	0.800	***WS_GHY***	-0.784	0.400
MinT_KSH	0.733	0.900	***WS_KSH***	-1.086	0.200
MinT_TUL	0.467	0.700	***WS_TUL***	-0.860	0.400

5.3.2 Inferences from the ACF and PACF plots (/Lag order detection)

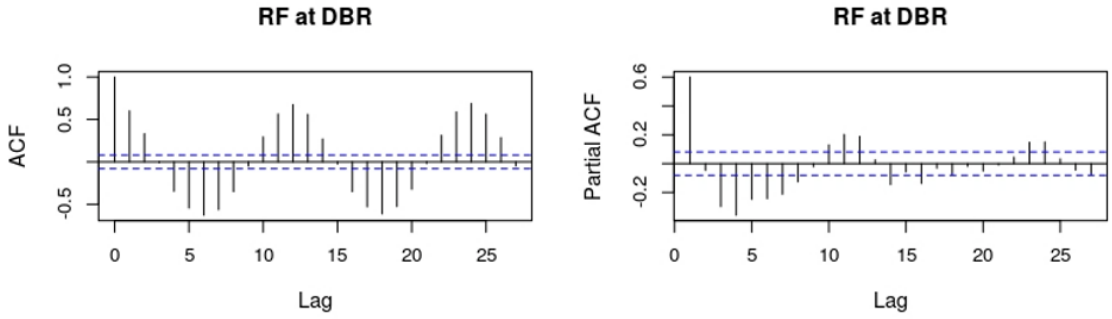
Further on, ACF and PACF plots were prepared for the better understanding of the time series process of the selected variables of interest.

Rainfall

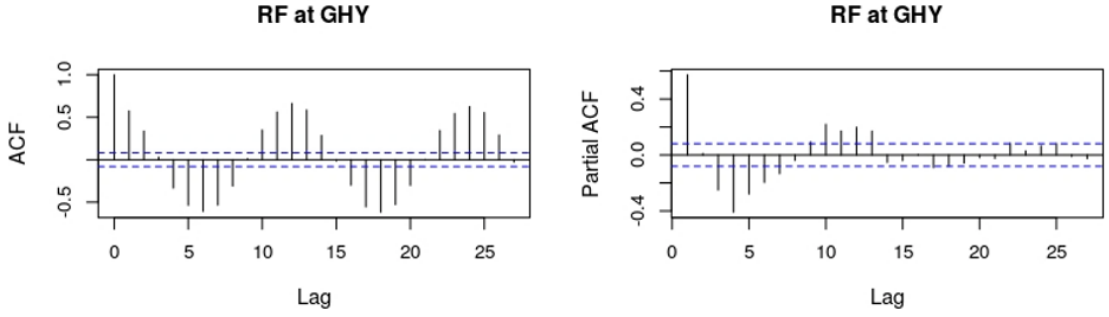
As the ACF and PACF plots of the rainfall series (Figure 5.2) suggest, the autocorrelation persisted for two months after lag 0 at all the selected locations of NER. Two cyclic periodicities of six and 12 months were also observed in the ACF plots. This clearly indicates the rainfall time series at these locations as non-stationary. Additionally, on the other hand, the PACF becomes significantly negative at lag 3 which persists up to lag 8 before reversing back to positive. The tails of the PACF plots geometrically declines repeating this pattern. Thus, these whole observations indicated towards the need for differencing prior to fitting ARIMA model of appropriate order to the rainfall data series for NER.


RF at CHR
ACF
Lag
RF at CHR
Partial ACF
Lag

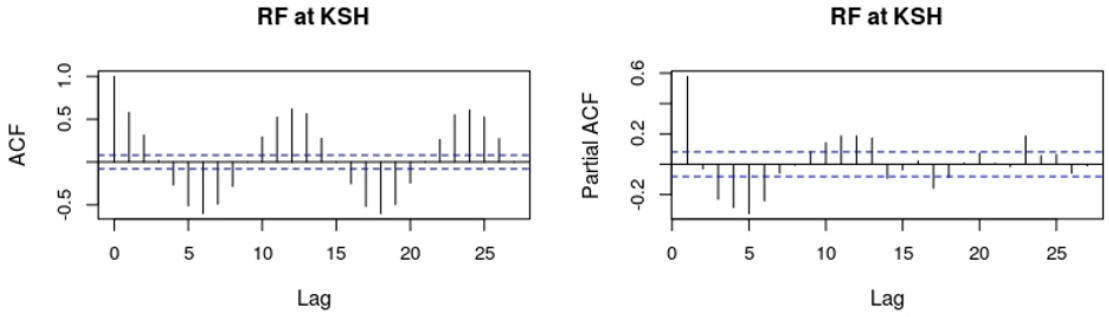
a)


RF at DBR
ACF
Lag
RF at DBR
Partial ACF
Lag

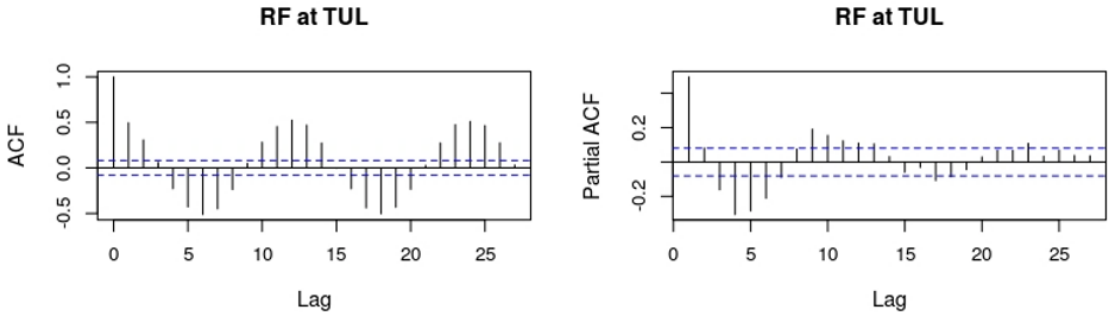
b)


RF at GHY
ACF
Lag
RF at GHY
Partial ACF
Lag

c)


RF at KSH
ACF
Lag
RF at KSH
Partial ACF
Lag

d)

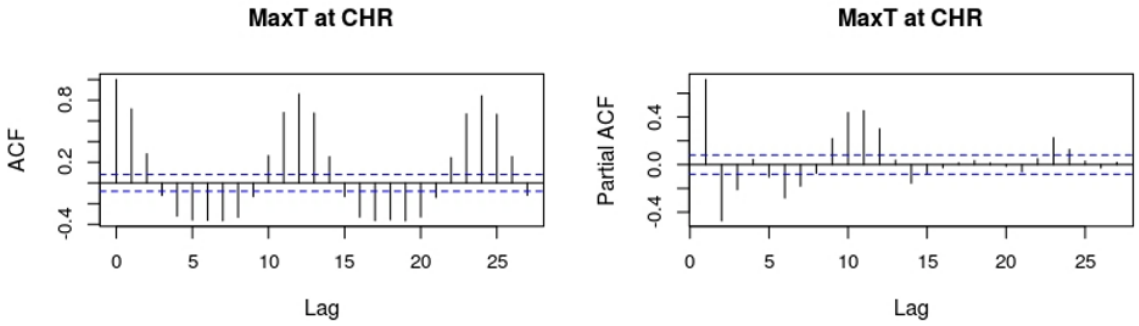


e)

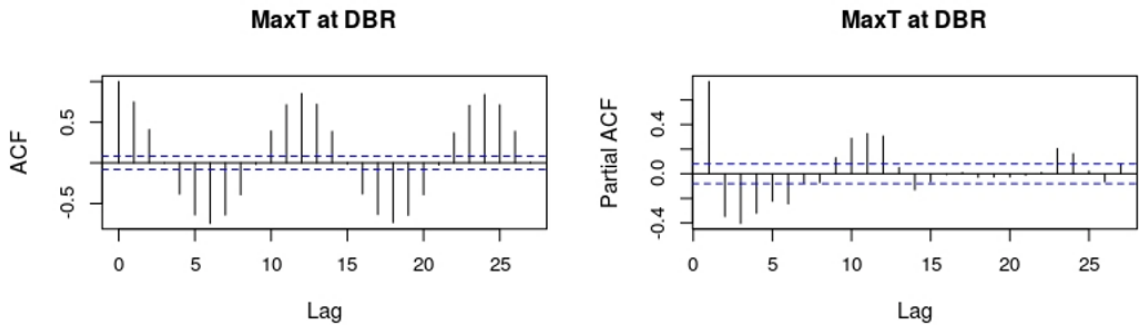
Figure 5. 2 ACF and PACF plots of the original rainfall time series at a) CHR, b) DBR, c) GHY, d) KSH and e) TUL. Here x- axis denotes monthly lags. The confidence bands are depicted in dashed lines.

Temperature

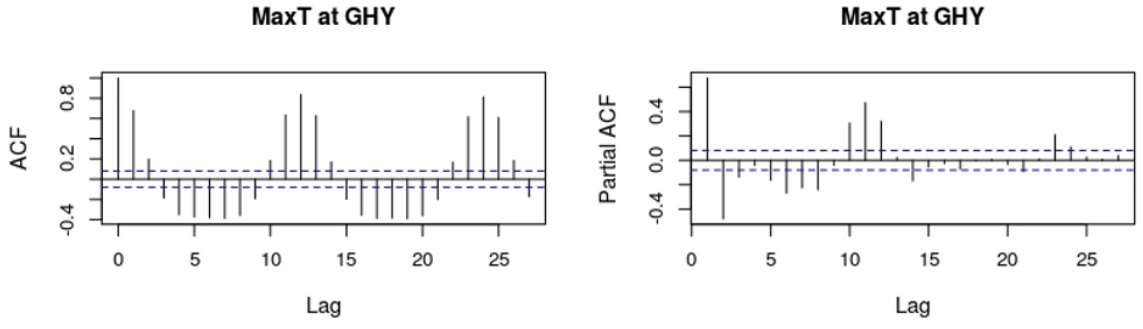
As evident from Figure 5.3, the ACF and PACF plots of the temperature series (both MaxT and MinT) over the NER display similar non-stationary features in the time series as observed in case of rainfall. Here also the autocorrelation persisted for two months after lag 0 at all the selected locations of NER and the six- and 12-months cyclic periodicities were also prominent in the ACF plots. In case of PACF, the temperature at all the locations sharply cuts off significantly after lag 1. The tails of the PACF plots geometrically declines in all the cases. Thus, here too these whole observations indicated towards the need for differencing before going for fitting ARIMA model of appropriate order to the data series.



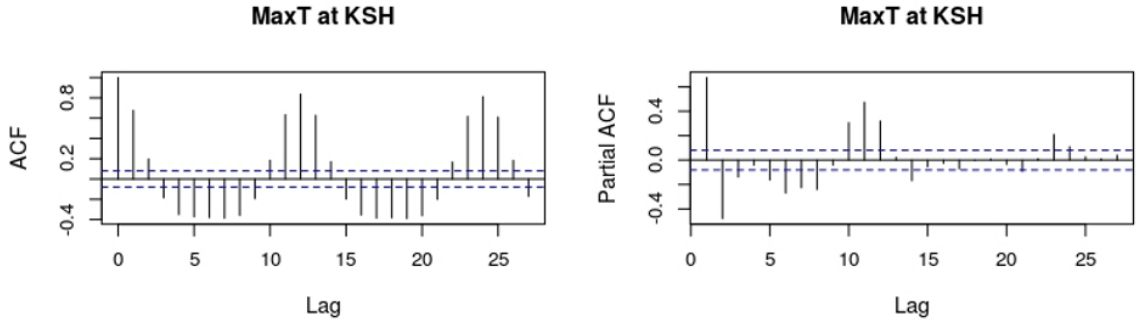
a)


MaxT at DBR
ACF
Lag
MaxT at DBR
Partial ACF
Lag

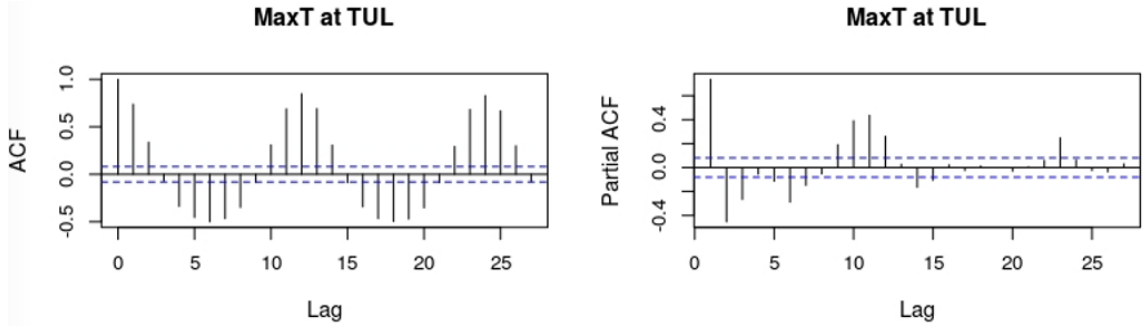
b)


MaxT at GHY
ACF
Lag
MaxT at GHY
Partial ACF
Lag

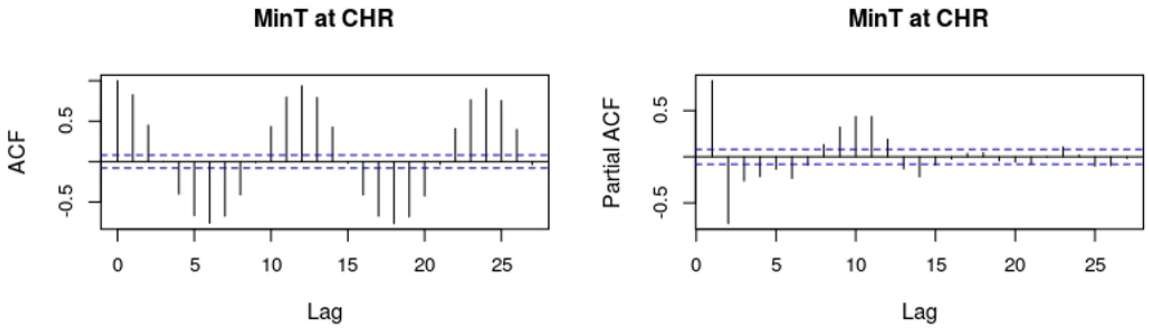
c)


MaxT at KSH
ACF
Lag
MaxT at KSH
Partial ACF
Lag

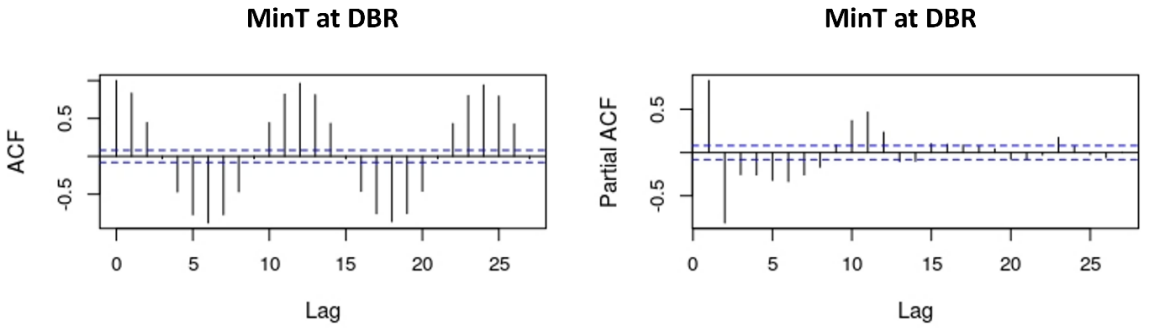
d)


MaxT at TUL
ACF
Lag
MaxT at TUL
Partial ACF
Lag

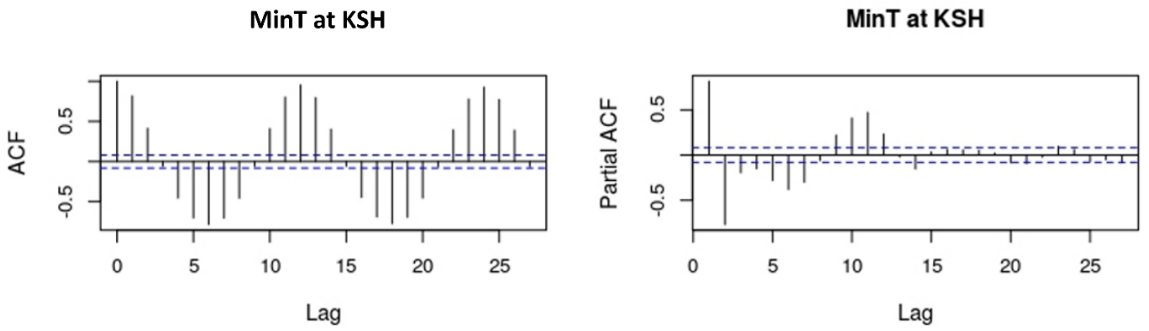
e)


MinT at CHR
ACF
Lag
MinT at CHR
Partial ACF
Lag

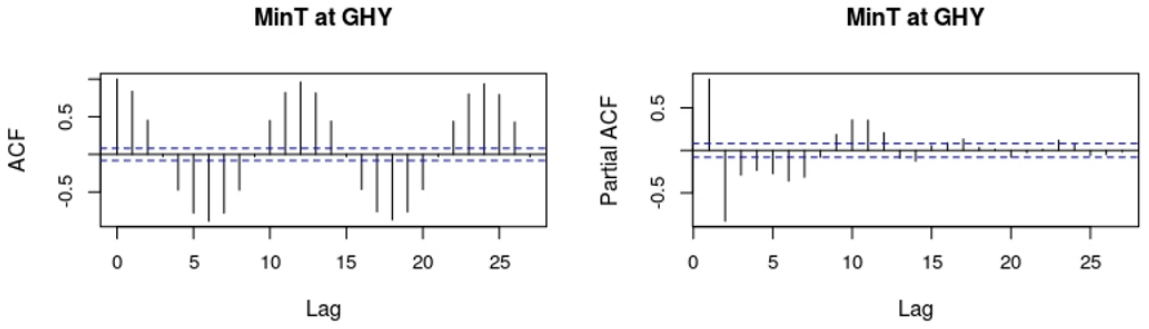
f)


MinT at DBR
ACF
Lag
MinT at DBR
Partial ACF
Lag

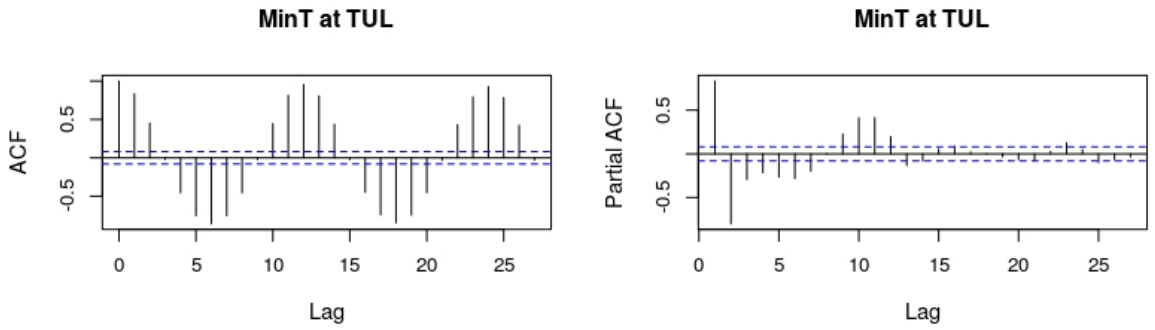
g)


MinT at KSH
ACF
Lag
MinT at KSH
Partial ACF
Lag

h)


MinT at GHY
ACF
Lag
MinT at GHY
Partial ACF
Lag

i)

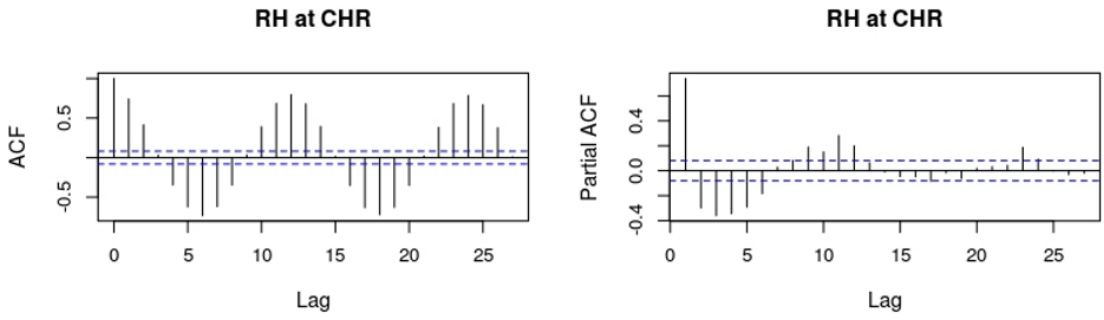


j)

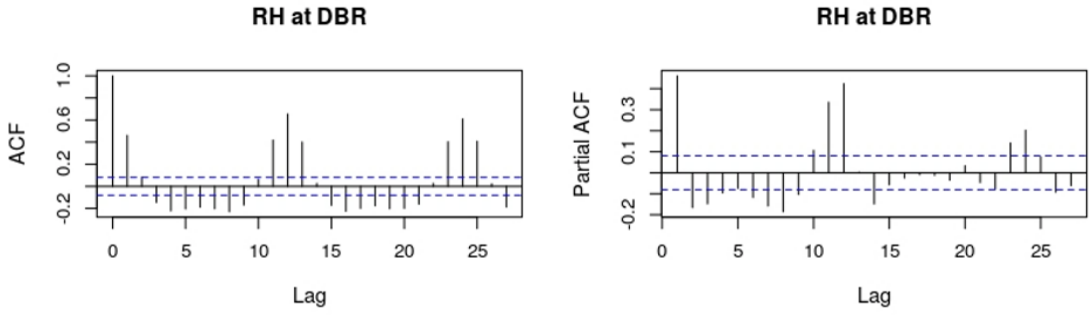
Figure 5. 3 ACF and PACF plots of the original MaxT time series at a) CHR, b) DBR, c) GHY, d) KSH and e) TUL and of MinT time series at f) CHR, g) DBR, h) GHY, i) KSH and j) TUL. Here x- axis denotes monthly lags. The confidence bands are depicted in dashed lines.

RH

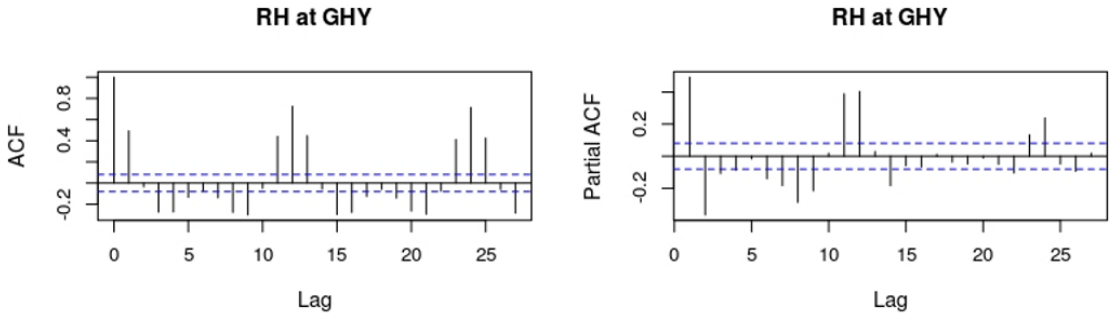
In case of RH, the autocorrelation persisted for two lags after zero at CHR and TUL only (Figure 5.4). In other three locations of NER, the autocorrelation became negative sharp after lag 1. The twelve-months cyclicity was observed in RH at all selected locations of NER too. The six-months cycle was prominent in RH at CHR only. No inference could be drawn from the PACF plots. Thus, the raw data series of RH also suggested differencing prior to the fitting of ARIMA model in the RH data.



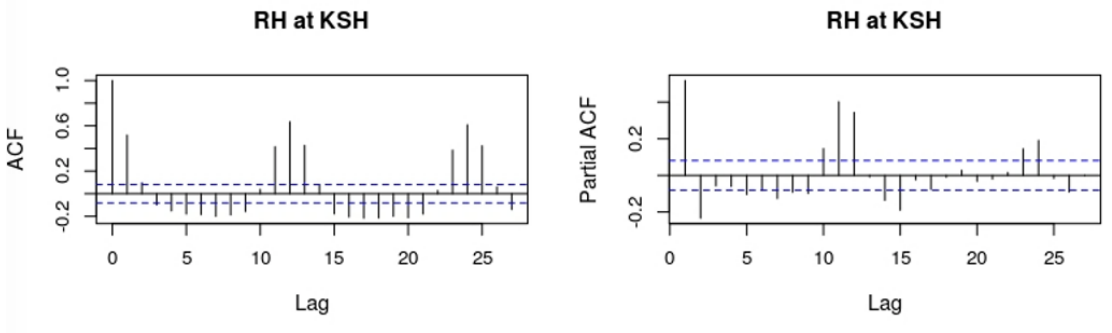
a)



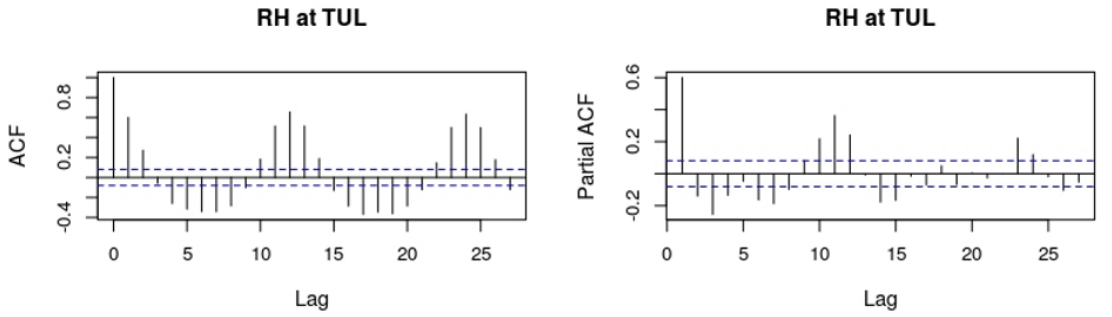
b)



c)



d)



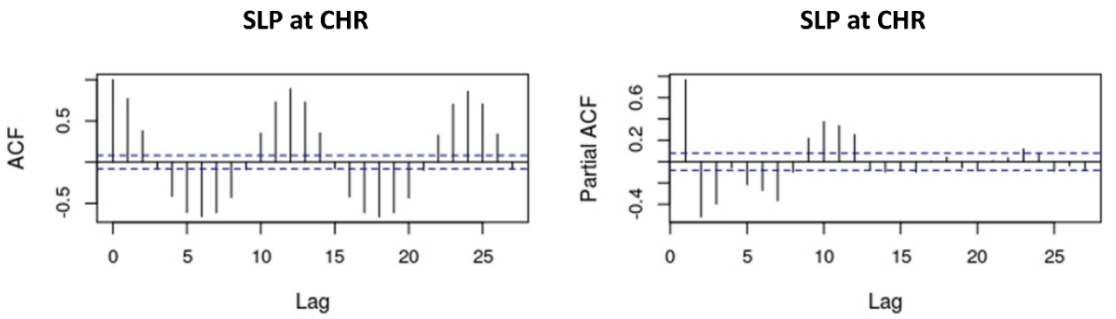
e)

Figure 5. 4 ACF and PACF plots of the original relative humidity time series at a) CHR, b) DBR, c) GHY, d) KSH and e) TUL. Here x- axis denotes monthly lags. The confidence bands are depicted in dashed lines.

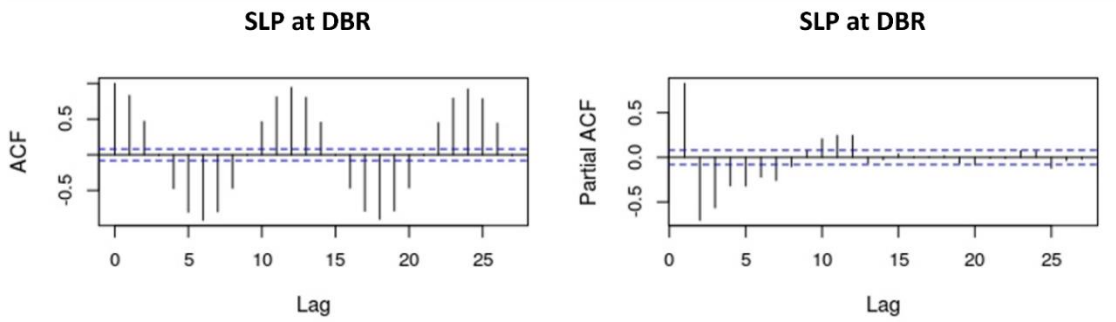
SLP

The autocorrelation in the SLP time series (Figure 5.5) revealed the same feature regarding persistence as that in case of rainfall and the temperature time series. Here too, the persistence of two months autocorrelation was observed. The six- and twelve-months cyclical behaviour was also exhibiting in the time series, depicting non-stationarity in the time series. The PACF plots were gradually declining, thus suggesting towards an underlying AR process of order 2. However, from the existing

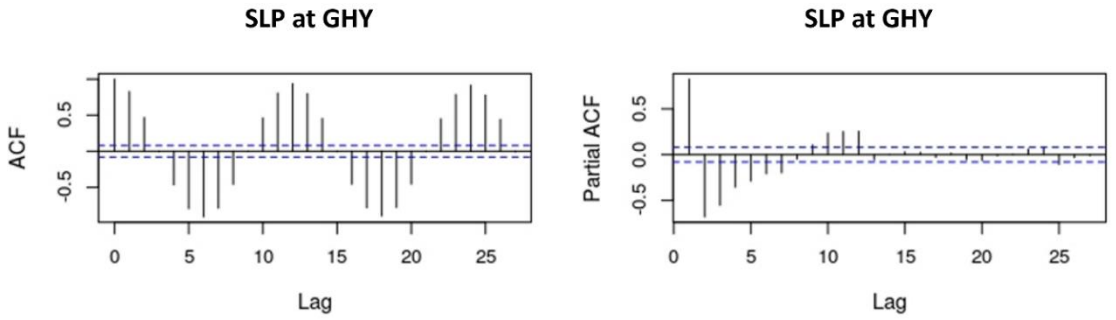
cyclicities in the ACF plots, rather differencing seemed to be the necessity at first and that fitting an ARIMA model to the data was needed.



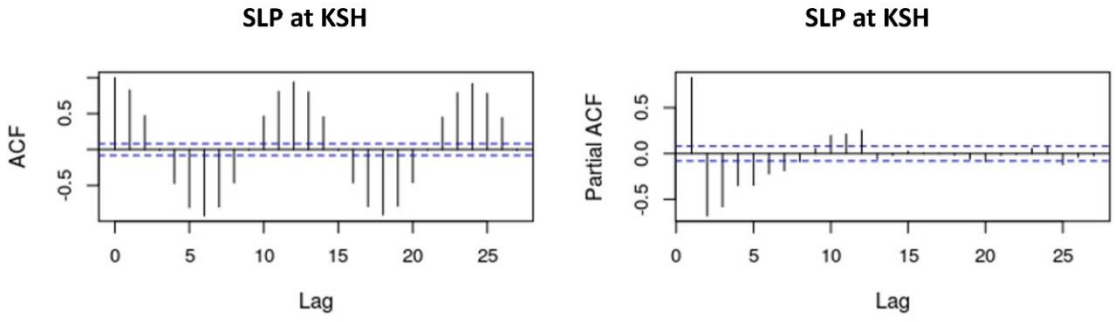
a)



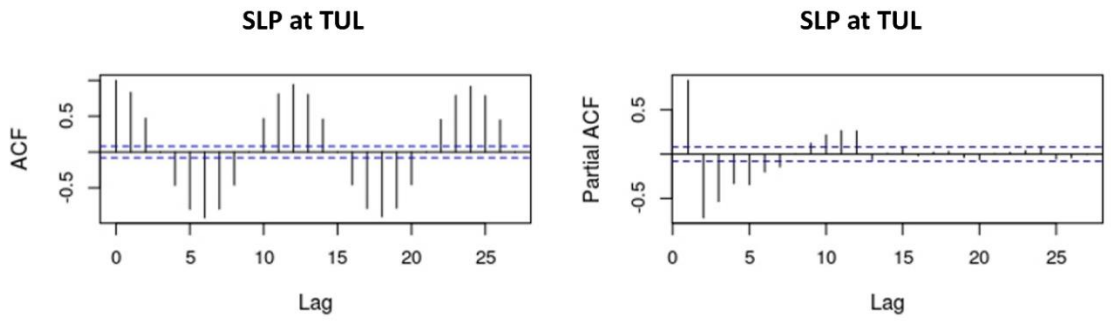
b)



c)



d)

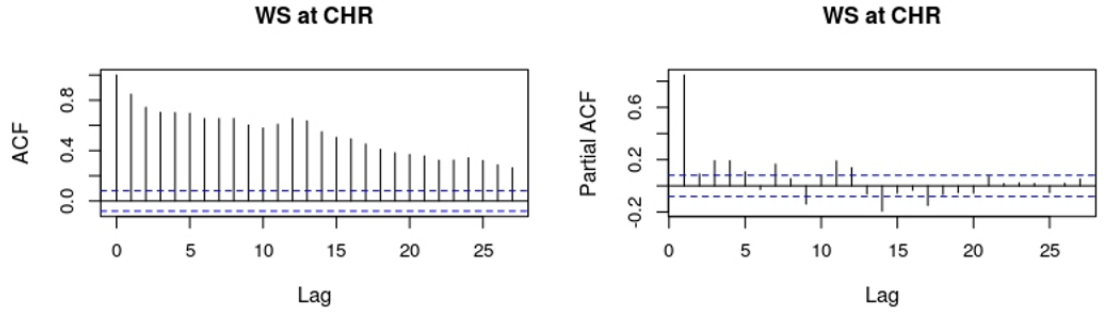


e)

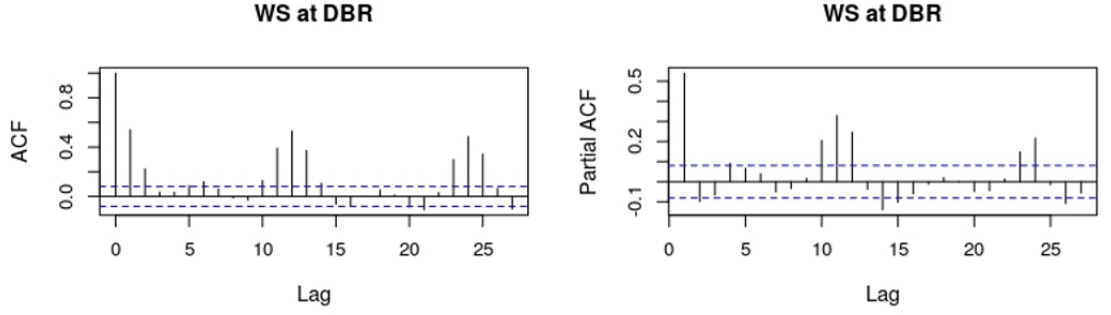
Figure 5. 5 ACF and PACF plots of the original sea level pressure time series at a) CHR, b) DBR, c) GHY, d) KSH and e) TUL. Here x- axis denotes monthly lags. The confidence bands are depicted in dashed lines.

WS

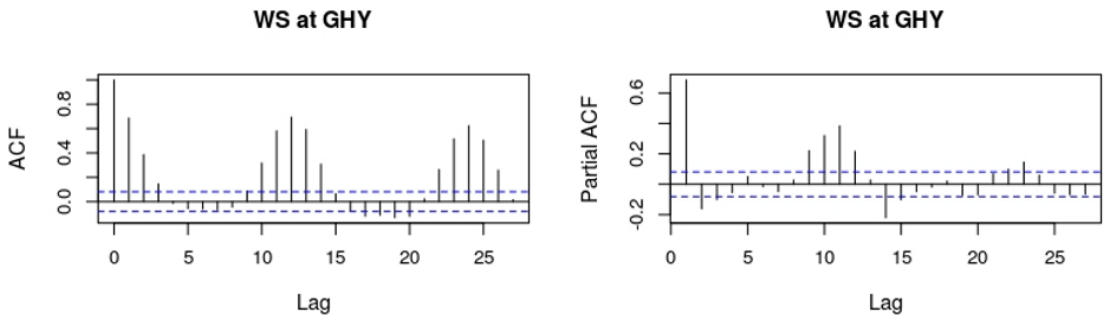
From the Figure 5.6, WS at CHR showed gradual decline in both ACF and PACF values. The persistence of autocorrelation for two (at DBR), three (at GHY and KSH) and four (TUL) months were detected in the ACF plots. The twelve months cycle was prominent in the ACF plots at all the selected location of NER except CHR, otherwise no definite pattern was seen in the ACF and PACF at these locations. No clear-cut information about the lags could be obtained from the PACF plots also. Thus, all the WS time series pointed towards non-stationarity existing within them. Thus, in the further steps differencing was performed.



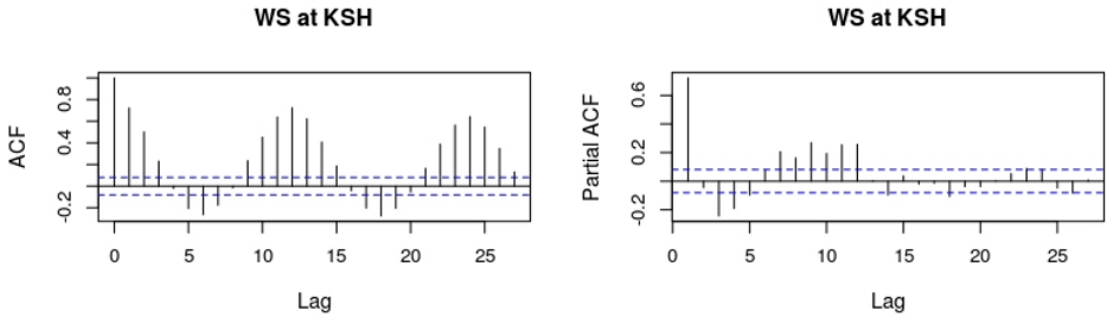
a)



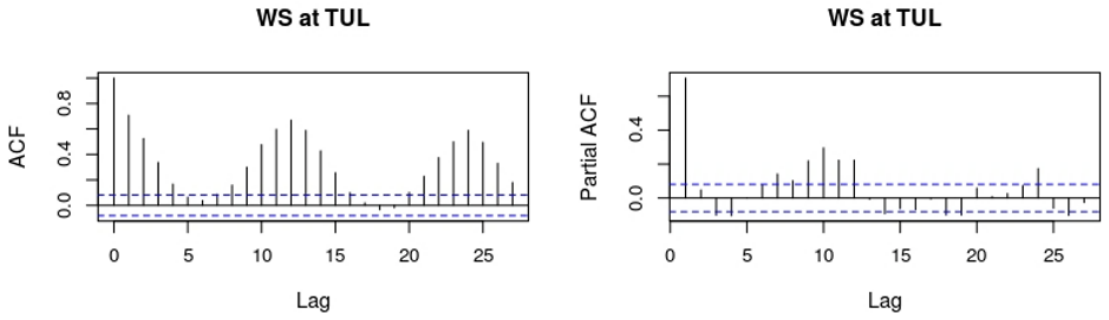
b)



c)



d)



e)

Figure 5. 6 ACF and PACF plots of the original wind speed time series at a) CHR, b) DBR, c) GHY, d) KSH and e) TUL. Here x- axis denotes monthly lags. The confidence bands are depicted in dashed lines.

5.3.3 The best fit ARIMA model

Keeping in view of the presence of 1-4 months autocorrelation (from the ACF plots) in the time series of the selected meteorological variables, we subjected the time series to differencing, followed by ARIMA model built up. 80% of the total data were used for model building, while the rest of the data (20%) were used for forecasting. A total of 24 models were tested for best fitting. The results are presented in Table 5.2.

As depicted in Table 5.2, a varied number of ARIMA models were found to be the best fit for different meteorological variables per selected locations in NER. Across all the

selected locations in our study, the model 22 [ARIMA(2,1,4)] was fitted as the parsimonious one for the maximum number of times (7); for MaxT at CHR, MinT at DBR and GHY, SLP at DBR, GHY, KSH and TUL. Contrary to model 22, model 12 [ARIMA(2,1,2)] was the best fitted model that appeared for only once, in case of MaxT at KSH. In all the cases, the AIC values were the determining factor governing the selection of the best fitted model than the other two criterions (HIC and SIC).

In case of rainfall, models 19 [ARIMA(4,1,3)], 14 [ARIMA(4,1,2)] and 23 [ARIMA (3,1,4)] were found to be best fitted with the time series for CHR, GHY and KSH respectively. Model 24 was detected as the best fit for DBR and TUL. It was found to be best fit for MaxT at DBR and TUL also. MaxT at the other sites, viz., CHR, GHY and KSH showed that among the tested combinations, models 22, 18 and 12 were the best fit for this meteorological variable respectively. In case of MinT, performance of model 22 was found to be better than the other models at DBR and GHY. The model 17 [ARIMA(2,1,3)] seemed to be best fitted for KSH and TUL, while at CHR, model 23 showed the best performance in MinT. In case of RH, both model 23 and 24 [ARIMA(4,1,4)] showed better performances for a pair of locations; model 23 for CHR and DBR, while model 24 for KSH and TUL. At KSH however, model 19 was fitted as the best fit. Model 22 among the other ones was the best fit in predicting SLP over all the studied locations except CHR, where model 23 showed the best performance. In case of WS over the locations of NER, model 19 was found to be best fit at DBR, KSH and TUL. At the other two locations- CHR and GHY, models 18 [ARIMA(3,1,3)] and 23 were detected to be the best among the fitted models in WS.

Table 5. 2 The best fit ARIMA model for different meteorological variables in the selected locations of NER. Here, the best fit model is shown in the arrangement ARIMA (p, d, q)[model number]

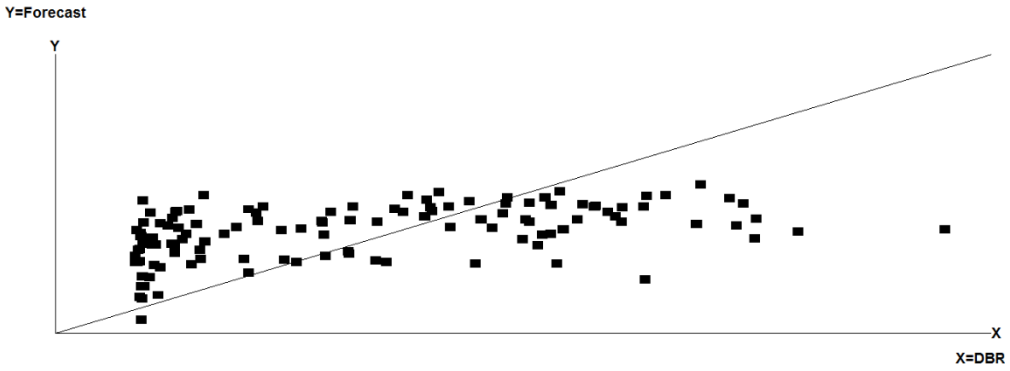
Variable_site	Best Fit Model	AIC	Variable_site	Best Fit Model	AIC	Variable_site	Best Fit Model	AIC
RF_CHR	ARIMA (4,1,3) [19]	13.566	***MinT_CHR***	ARIMA (3,1,4) [23]	0.608	***SLP_CHR***	ARIMA (3,1,4) [23]	5.301
RF_DBR	ARIMA (4,1,4) [24]	9.866	***MinT_DBR***	ARIMA (2,1,4) [22]	0.859	***SLP_DBR***	ARIMA (2,1,4) [22]	0.879
RF_GHY	ARIMA (4,1,2) [14]	9.330	***MinT_GHY***	ARIMA (2,1,4) [22]	0.710	***SLP_GHY***	ARIMA (2,1,4) [22]	0.967
RF_KSH	ARIMA (3,1,4) [23]	10.079	***MinT_KSH***	ARIMA (2,1,3) [17]	1.164	***SLP_KSH***	ARIMA (2,1,4) [22]	0.721
RF_TUL	ARIMA (4,1,4) [24]	8.695	***MinT_TUL***	ARIMA (2,1,3) [17]	1.181	***SLP_TUL***	ARIMA (2,1,4) [22]	0.901
MaxT_CHR	ARIMA (2,1,4) [22]	0.874	***RH_CHR***	ARIMA (3,1,4) [23]	3.906	***WS_CHR***	ARIMA (3,1,3) [18]	1.851
MaxT_DBR	ARIMA (4,1,4) [24]	0.838	***RH_DBR***	ARIMA (3,1,4) [23]	3.305	***WS_DBR***	ARIMA (4,1,3) [19]	0.940
MaxT_GHY	ARIMA (3,1,3) [18]	1.147	***RH_GHY***	ARIMA (4,1,3) [19]	3.345	***WS_GHY***	ARIMA (3,1,4) [23]	0.933
MaxT_KSH	ARIMA (2,1,2) [12]	1.010	***RH_KSH***	ARIMA (4,1,4) [24]	2.526	***WS_KSH***	ARIMA (4,1,3) [19]	1.057
MaxT_TUL	ARIMA (4,1,4) [24]	0.971	***RH_TUL***	ARIMA (4,1,4) [24]	3.355	***WS_TUL***	ARIMA (4,1,3) [19]	0.188

5.3.4 Performance of the best fit ARIMA model in forecasting

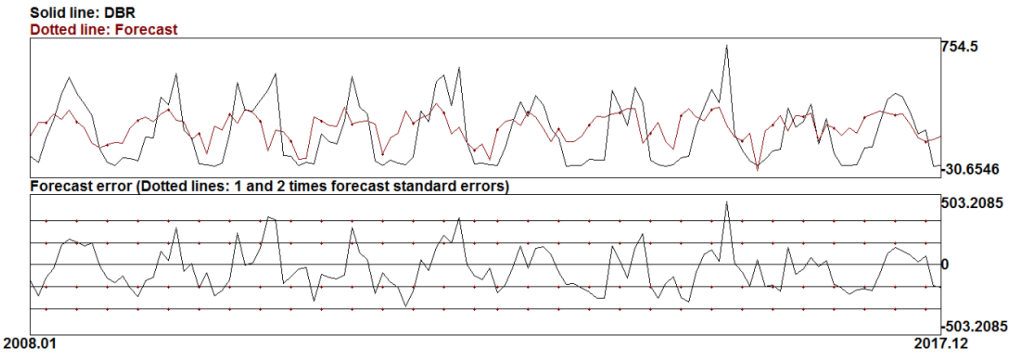
The performance evaluation of the best fitted ARIMA model to the time series of the selected meteorological variables over NER could be explained with the help of the comparison graphs (scatter plots as well as the respective time series plots) of forecasted vs original series in the following sub-sections.

Rainfall

The performance of ARIMA modelling in rainfall is elaborated with the help of the Figures 5.7 a-e (i and ii). In a close look on scatter plots of original vs. forecasted series of rainfall at all the five studied locations in NER revealed a rather weak resemblance of forecasted series with the original one. A typical representative time series plot of the original series (solid black line) and forecasted series (solid red line) in plot (ii) of Figure 5.7 also reveal the same in forecasting in rainfall.



(i)



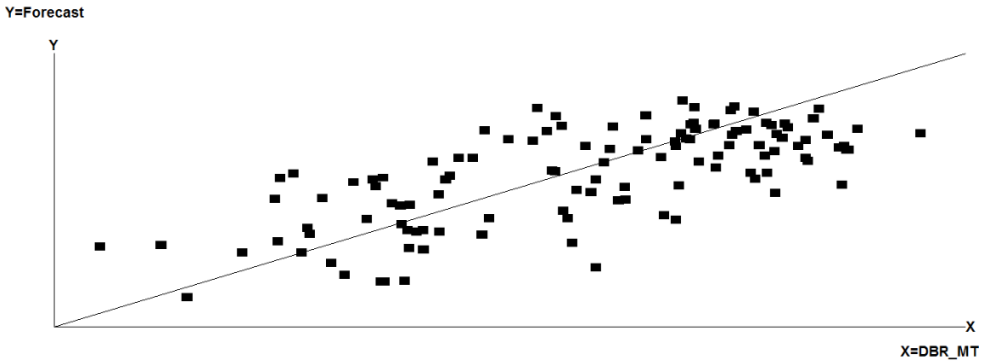
(ii)

Figure 5. 7 Details of original vs. forecasted time series of rainfall at a representative location of NER (DBR), (i) denotes scatter plot of original vs. forecasted series, (ii) denotes time series plot of original and forecasted series at each site (a-e).

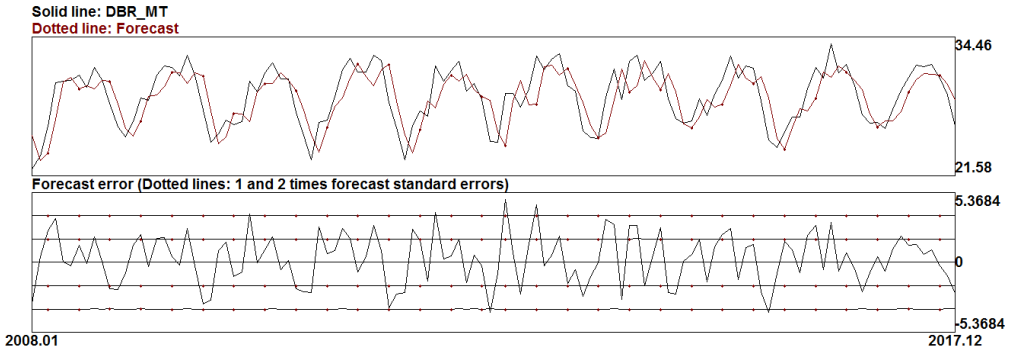
MaxT and MinT

The performance of ARIMA in forecasting MaxT and MinT time series can be described with the help of Figure 5.8 and Figure 5.9 respectively.

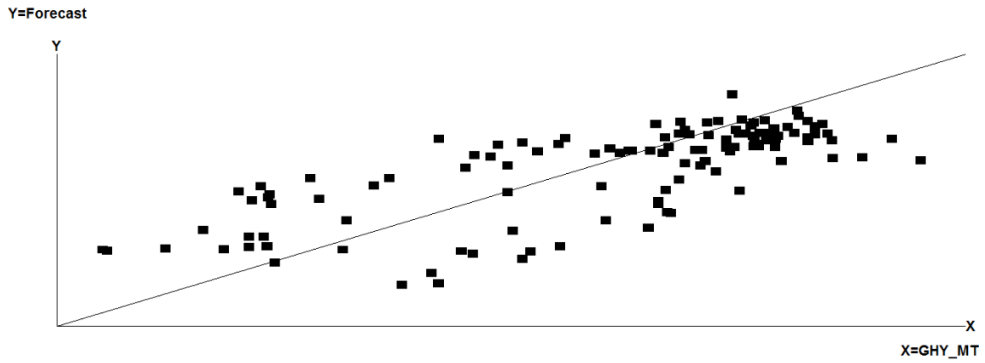
As evident from the Figure 5.8 (i and ii), the scatter plot (i) revealed a relatively positive association of the predicted and original time series at DBR and GHY. This association was evident in the time series plots of the forecasted and original series also, with similar pattern and inherent lag structure to be present in the forecasted series in red. The other locations in NER showed no specific association in the original and forecasted MaxT series.



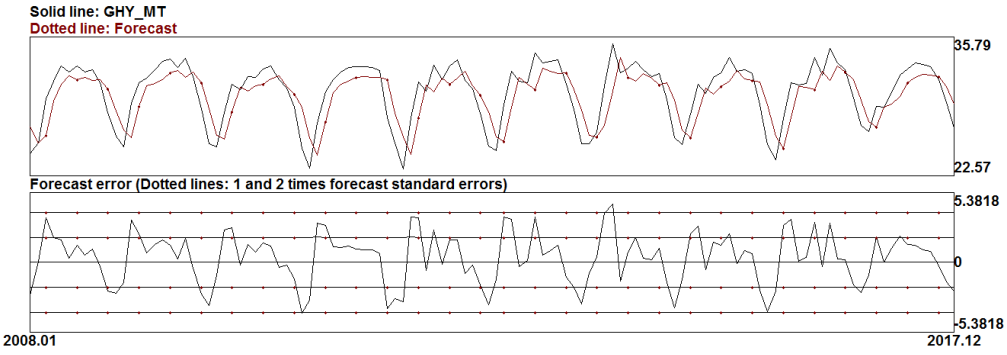
a-(i)



a-(ii)



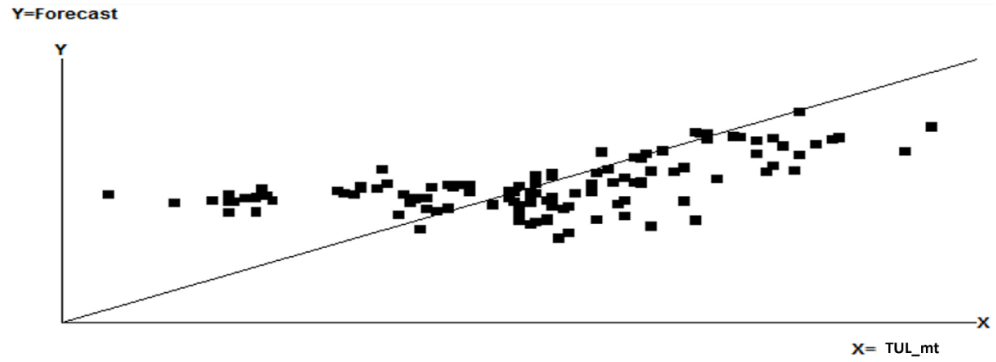
b-(i)



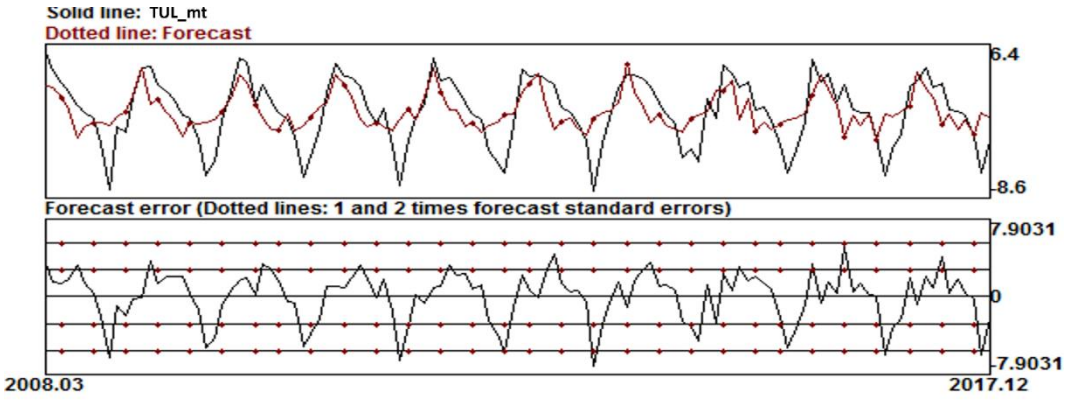
b-(ii)

Figure 5. 8 Details of original vs. forecasted time series of MaxT at a) DBR and b) GHY. (i) denotes scatter plot of original vs. forecasted series, (ii) denotes time series plot of original and forecasted series at each site (a-e).

A relatively better performance of ARIMA was evident in the prediction of MinT time series than rainfall also. (Figure 5.9). Among all the selected study sites, the prediction by ARMIA at TUL was the best as evident in the representative plots (Figure 5.9).



(i)



(ii)

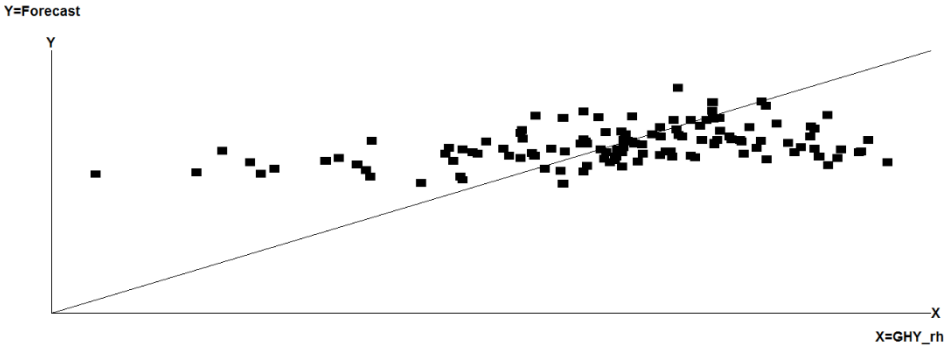
Figure 5. 9 Details of original vs. forecasted time series of MinT at TUL. (i) denotes scatter plot of original vs. forecasted series, (ii) denotes time series plot of original and forecasted series at each site (a-e).

RH

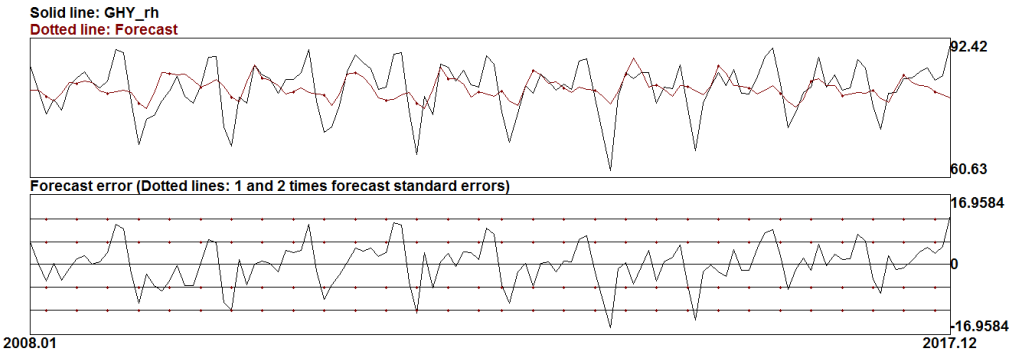
In case of RH, the ARIMA modelling approach adopted in this study didn't show a descent result in predicting the time series at all the studied locations in NER accurately. A detailed view of the scatter plots from a typical representative scatter plot and time series plot of original Vs. forecasted RH at GHY is presented in Figure 5.10.

SLP

Among all the selected locations of NER, prediction of SLP was somewhat better only at TUL as shown in Figure 5.11. The time series of the forecasted SLP depicted seasonal behaviour intact as in the other meteorological variables.

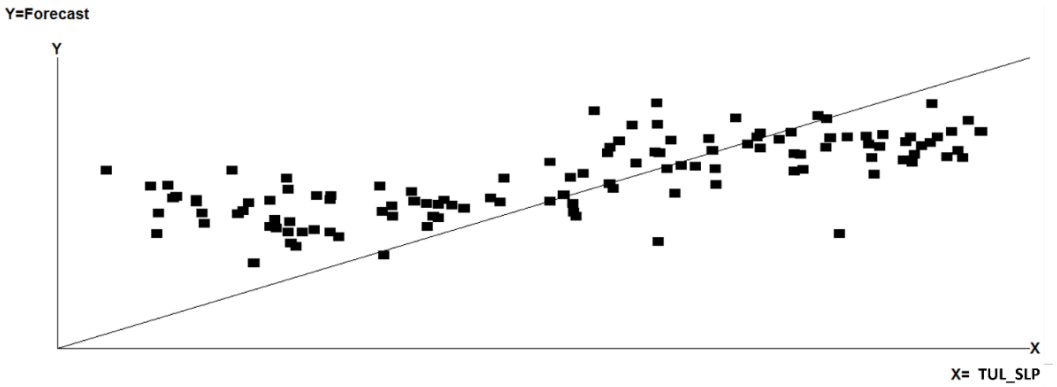


(i)

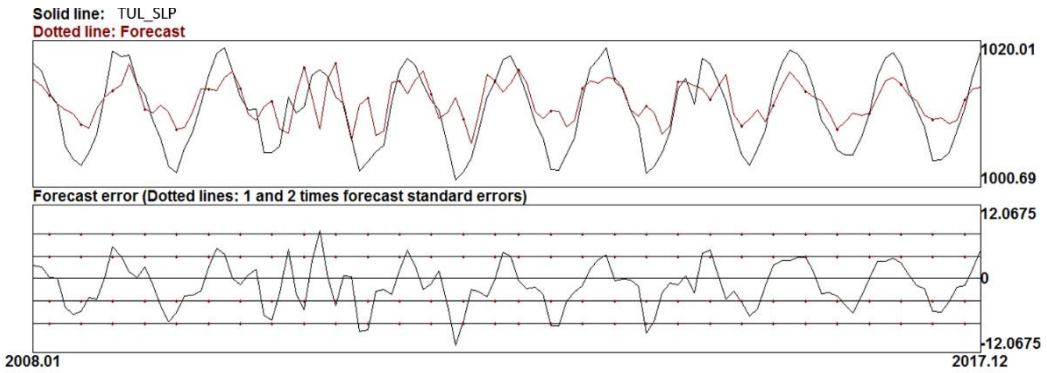


(ii)

Figure 5. 10 Details of original vs. forecasted time series of RH at GHY (i) denotes scatter plot of original vs. forecasted series, (ii) denotes time series plot of original and forecasted series at each site (a-e).



e-(i)

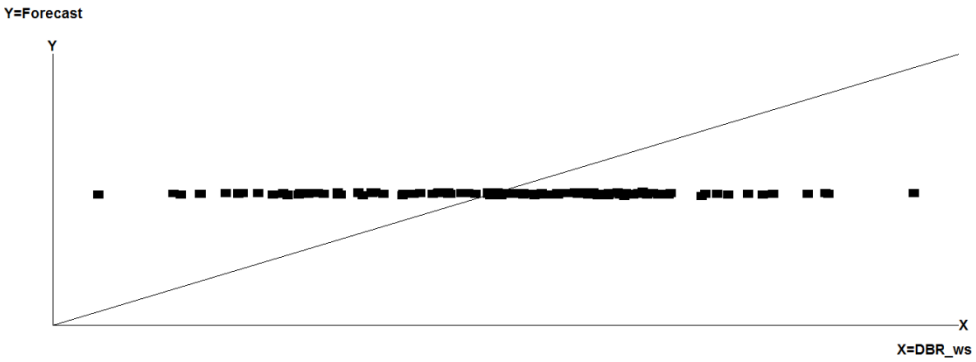


e-(ii)

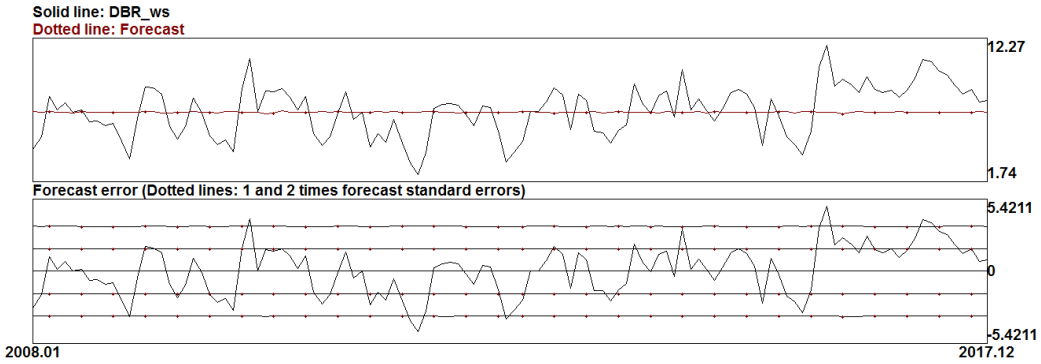
Figure 5. 11 Details of original vs. forecasted time series of SLP at TUL. (i) denotes scatter plot of original vs. forecasted series, (ii) denotes time series plot of original and forecasted series at each site (a-e).

WS

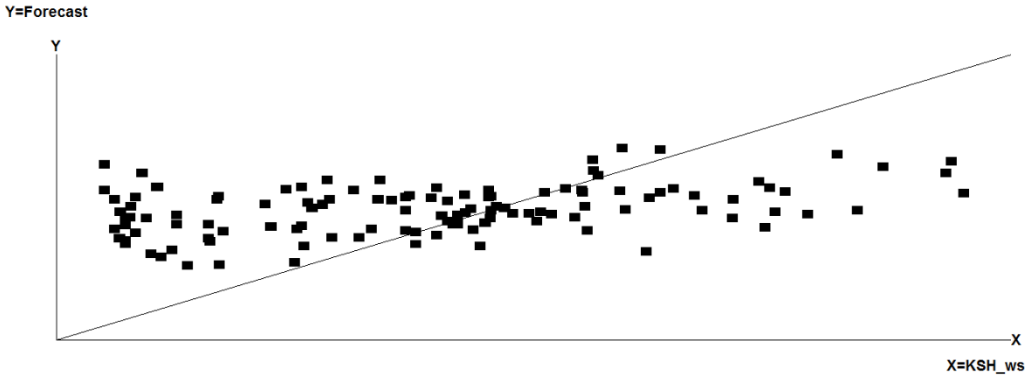
The results of WS forecasting by ARIMA for two representative locations of NER are presented here as the worst and relatively best prediction among all (Figure 5.12 a-b (i) and (ii) for DBR and KSH respectively). As evident from the scatter plots, and time series plots, the performance of ARIMA model in WS was not satisfactory at all in case of DBR.



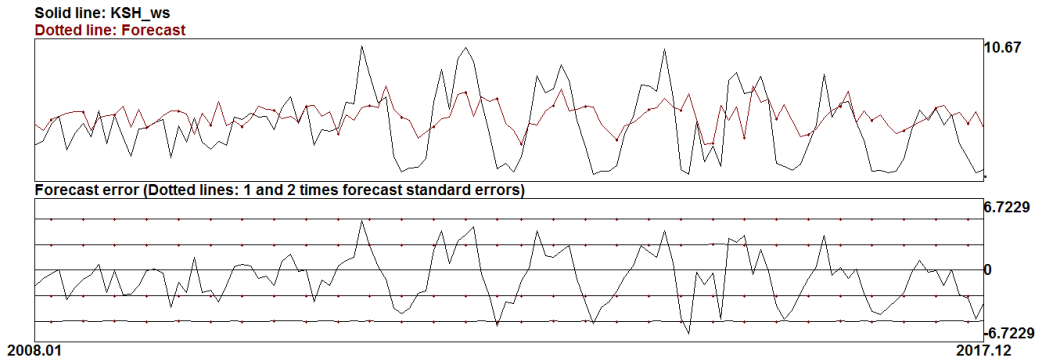
a-(i)



a-(ii)



b-(i)



b-(ii)

Figure 5. 12 Details of original vs. forecasted time series of WS at a) DBR and b) KSH. (i) denotes scatter plot of original vs. forecasted series, (ii) denotes time series plot of original and forecasted series at each site (a-e).

5.4 Summary

Regional climate is influenced by the conditions of the meteorological variables. Therefore, studying the behaviour of a meteorological time series becomes imperative as any fluctuations or changes in their statistical properties may lead to fuel the climate crisis that's happening over the globe at regional scale. In this concern, here in this study we've tried to analyze time series of meteorological variables rainfall, temperature (both MaxT and MinT), RH, SLP and WS over selected locations of NER to understand the individual behaviour of the variables in terms of their stationarity, persistence, autoregressive behaviour, seasonality and subsequently used them as input variables in building ARIMA models. As per the results suggested, it was revealed that all the original time series were non-stationary in nature, exhibiting seasonality. Two cyclic patterns- of six months and a year were observed in the ACF plots except for RH (at DBR, GHY and KSH, where only twelve months' cyclicity was observed) and WS (twelve months' cyclicity was present across all sites except at CHR and six months' cyclicity was observed only at KSH). The persistence of two months autocorrelation was noticed in the ACF plots of rainfall, temperature, RH (at CHR and TUL only), SLP and WS (at DBR only), as an indicative of two months lagged dependence of each point of time upon its previous value in each of the said meteorological variable. Likewise, three months (at GHY and KSH) and four months (at TUL) persistence as noticed in the ACF plots of WS indicated three and four months lagged association of each point of time upon its previous value in WS, respectively. Among the meteorological variables, RH at DBR, GHY and KSH exhibited one month's persistence of autocorrelation.

Keeping in view of the lagged autocorrelation of 1-4 months observed in the meteorological variables, 24 different ARIMA models were built, and the meteorological variables showed different best fit models per site of NER. The model ARIMA(2,1,4) was best fit in most of the cases (7), followed by ARIMA(4,1,4). ARIMA(2,1,4) was best fit for MaxT at CHR, MinT at DBR and GHY, SLP at all sites except CHR. ARIMA(4,1,4) on the other hand, was found as the best fit for rainfall and MaxT at both DBR and TUL, RH at KSH and TUL. As evident in the table 5.2, it can be noticed from the best fit ARIMA models, that, rainfall was auto regressed over 3 and 4 past months observation in NER. Autoregression of 2, 3 and 4 months were prominent

in MaxT. Likewise, MinT and SLP for most of the sites was auto regressed over 2 past months observation. 3 and 4 months autoregression were prominent in RH and WS.

The ARIMA best fit models were used for predicting the meteorological variables. It was noticed from the time series plots that the association of original and forecasted series was negligible in rainfall (Figure 5.7 (i)). The similar feature was prominent in case of temperature also, except for MaxT and MinT at both DBR and GHY, where a weak positive association was seen between the original and the predicted time series. In case of SLP the forecasted series showed relatively a weak positive association with the original series at TUL of NER. The prediction of WS by the best fit ARIMA was also not satisfactory. It is to be noted that quantitative estimation of prediction efficiency by ARIMA is excluded in this study. In most of the studies, ARIMA model gave better performance in short-term forecasting [11]. It is established that the Box-Jenkins seasonal ARIMA forecasting model provides more statistical information than other techniques of analysis [7]. However, the addition of ANN to ARIMA has been suggested as a measure to further improvement of forecasting by the proposed ARIMA model [14]. Recent findings also suggest that the efficiency of combined statistical and neural network models is higher than either of the single models [20][48-49]. Therefore, it is to be noted that there's a scope in making better performance by our best fit ARIMA models in prediction with the incorporation of such artificial intelligence-based modification. Also, preference of SARIMA is suggested over ARIMA as the best fit models in our study detected the presence of seasonal signatures in the studied meteorological time series for the NER.

5.5 References

[1] Priestly, M.B. *Non-linear and non-stationary time series analysis*. Academic Press, Harcourt Brace Jovanovich, ISBN 0-12-564911-8, 1991.

[2] Montgomery, D.C. and Johnson, L.A. *Forecasting and time series analysis*. McGraw-Hill, New York, 1976.

[3] Box, G.E.P. and Jenkins, G.M. *Time series analysis: forecasting and control*. Holden-Day, San-Francisco, 1970.

[4] Storch, H. and Zwiers, F.W. *Statistical analysis in climate research*. Cambridge University Press, ISBN 0 521 45071 3, 2003.

[5] Duchon, C.E. and Hale, R. *Time series analysis in meteorology and climatology*. John Wiley & Sons Ltd., ISBN 978-0-470-97199-4, 2012.

[6] Edwards, S. Regionally dissected temperature and rainfall models for the South Island of New Zealand. Master's thesis, Department of Applied Science, Lincoln University, 2011.

[7] Kim, B.S., Hossein, S.Z. and Choi, G. Evaluation of Temporal-spatial Precipitation Variability and Prediction Using Seasonal ARIMA Model in Mongolia. *KSCE Journal of Civil Engineering*, 15(5):917-925, 2011. DOI: 10.1007/s12205-011-1097-9.

[8] Cadenas, E., Rivera, W., Campos-Amezcua, R. and Heard, C. Wind Speed Prediction Using a Univariate ARIMA Model and a Multivariate NARX Model. *Energies*, 9(109):1-15, 2016.

[9] Balasmeh, O.A., Babbar, R. and Karmakar, T. Trend analysis and ARIMA modeling for forecasting precipitation pattern in Wadi Shueib catchment area in Jordan. *Arabian Journal of Geosciences*, 12(27):1-19, 2019. DOI: https://doi.org/10.1007/s12517-018-4205-z.Swain, S., Nandi,

[10] Eymen, A. and Köylü, Ü. Seasonal trend analysis and ARIMA modeling of relative humidity and wind speed time series around Yamula Dam. *Meteorology and Atmospheric Physics*, 131:601–612, 2019. DOI: https://doi.org/10.1007/s00703-018-0591-8.

[11] Lai, Y. and Dzombak, D.A. Use of Autoregressive Integrated Moving Average (ARIMA) model to forecast near-term regional temperature and precipitation. *Weather and Forecasting*, 2020. DOI: 10.1175/ WAF-D-19-0158.1.

[12] Shiri, G., Salahi, B., Samadzadeh, R. and Shiri, M. The investigation and forecasting of relative humidity variations of Pars Abad-e-Moghan, North-West of Iran, by ARIMA model. *Research Journal of Applied Sciences*, 6(2):81-87, 2011.

[13] Han, P., Wang, P., Tian, M., Zhang, S., Liu, J. and Zhu, D. Application of the ARIMA Models in Drought Forecasting Using the Standardized Precipitation Index. *In CCTA, 2012, Part I, IFIP AICT*, volume 392, pages 352–358, 2013.

[14] Wang, H.R., Wang, C., Lin, X. and Kang, J. An improved ARIMA model for precipitation simulations. *Nonlinear Processes in Geophysics*, 21:1159-1168, 2014. DOI: 10.5194/npg-21-1159-2014.

[15] Li, Z.Q., Zou, H.X. and Qi, B. Application of ARIMA and LSTM in Relative Humidity Prediction. In *IEEE 19th International Conference on Communication Technology*, pages 1544-1548, 2019.

[16] Bari, S.H., Rahman, M.T., Hussain, M.M. and Ray, S. Forecasting Monthly Precipitation in Sylhet City Using Arima Model. *Civil and Environmental Research*, 7: 69-77, ISSN 2224-5790, 2015.

[17] Shahriar, S.A., Kayes, I., Hasan, K., Hasan, M., Islam, R., Awang, N.R., Hamzah, Z., Rak, A.E and Salam, M.A. Potential of ARIMA-ANN, ARIMA-SVM, DT and CatBoost for Atmospheric $PM_{2.5}$ Forecasting in Bangladesh. *Atmosphere*, 12:1-21, 2021. DOI: https://doi.org/10.3390/atmos12010100.

[18] Dawood, M., Rahman, A., Ullah, S., Mahmood, S., Rahman, G. and Azam, K. Spatio-statistical analysis of rainfall fluctuation, anomaly and trend in the Hindu Kush region using ARIMA approach. *Natural Hazards*, 101:449–464, 2020. DOI: https://doi.org/10.1007/s11069-020-03881-5.

[19] Somvanshi, V.K., Pandey, O.P., Agrawal, P.K., Kalankar, N.V., Prakash, M.R. and Chand, R. Modelling and prediction of rainfall using artificial neural network and ARIMA techniques. *The Journal of Indian Geophysical Union*, 10(2):141-151, 2006.

[20] Chattopadhyay, S. and Chattopadhyay, G. Univariate modelling of summer-monsoon rainfall time series: Comparison between ARIMA and ARNN. *C. R. Geoscience*, 342:100-107, 2010. DOI: 10.1016/j.crte.2009.10.016.

[21] Chaudhuri, S. and Dutta, D., Mann–Kendall trend of pollutants, temperature, and humidity over an urban station of India with forecast verification using different ARIMA models. *Environmental Monitoring and Assessment*, 186: 4719-4724, 2014. DOI: 10.1007/s10661-014-3733-6.

[22] Hosamane, S.N., Prasanth, K.S. and Virupakshi, A.S. Assessment and prediction of PM_{10} concentration using ARIMA. In *Journal of Physics: Conference Series, 1st International Conference on Advances in Physical Science and Materials*, volume 1706, pages 1-9, 2020. DOI: 10.1088/1742-6596/1706/1/012132.

[23] Narayanan, P., Basistha, A., Sarkar, S. and Sachdeva, K. Trend analysis and ARIMA modelling of pre-monsoon rainfall data for western India. *Elsevier*, 345:22-27, 2013. DOI: http://dx.doi.org/10.1016/j.crte.2012.12.001.

[24] Narayanan, P., Sarkar, S., Basistha, A. and Sachdeva, K. Trend analysis and forecast of pre-monsoon rainfall over India. *Weather*, 71(4):94-99, 2016. DOI: 10.1002/wea.2699.

[25] Dimri, T., Ahmad, S. and Sharif, M. Time series analysis of climate variables using seasonal ARIMA approach. *Journal of Earth System Science*, 129(149):1-16, 2020. DOI: https://doi.org/10.1007/s12040-020-01408-x.

[26] Nirmala, M. and Sundaram, S.M. A seasonal ARIMA model for forecasting monthly rainfall in Tamilnadu. *National Journal on Advances in Building Sciences and Mechanics*, 1(2):43-47, 2010.

[27] Swain, S. and Patel, P. Development of an ARIMA Model for Monthly Rainfall Forecasting over Khordha District, Odisha, India. *Recent Findings in Intelligent Computing Techniques*, Advances in Intelligent Systems and Computing, 708: 325-331, 2018. DOI: 10.1007/978-981-10-8636-6_34.

[28] Shad, M., Sharma, Y.D. and Singh, A. Forecasting of monthly relative humidity in Delhi, India, using SARIMA and ANN models. *Modeling Earth Systems and Environment*, 2022. DOI: https://doi.org/10.1007/s40808-022-01385-8.

[29] Goswami, K., Hazarika, J. and Patowary, A.N. Monthly temperature prediction based on ARIMA model: a case study in Dibrugarh station of Assam, India. *International Journal of Advanced Research in Computer Science*, 8(8):292-298, 2017. DOI: http://dx.doi.org/10.26483/ijarcs.v8i8.4590.

[30] Murthy, K.V.N., Saravana, R. and Kumar, K.V. Modeling and forecasting rainfall patterns of southwest monsoons in North–East India as a SARIMA process.

Meteorological Atmospheric Physics, 130:99-106, 2018. DOI: https://doi.org/10.1007/s00703-017-0504-2.

[31] Das, U.D., Singh, B.P. and Roy, T.D. Temporal Variation of Temperature in Guwahati, Assam: An Application of Seasonal ARIMA Model. *Journal of Statistics Applications and Probability*, 9(1):169-180, 2020.

[32] Kumar, T.S., Das, H.S., Choudhury, U., Dutta, P.E., Guha, D. and Laskar, Y. Analysis and Prediction of Air Pollution in Assam Using ARIMA/SARIMA and Machine Learning. In *Innovations in Sustainable Energy and Technology. Advances in Sustainability Science and Technology*, pages 317-330, 2021. DOI:10.1007/978-981-16-1119-3.

[33] Barman, U., Hussain, A.E., Dahal, M.J., Barman, P. and Hazarika, M. Time Series Analysis of Assam Rainfall Using SARIMA and ARIMA. In *Smart Computing Techniques and Applications, Smart Innovation, Systems and Technologies*, pages 357-364, 2021. DOI: https://doi.org/10.1007/978-981-16-0878-0_35.

[34] Dickey, D.A. and Fuller, W.A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74, 427-431, 1979.

[35] Fuller, W.A. *Introduction to Statistical Time Series*, 2nd edition, New York: John Wiley, 1996.

[36] Said, S.E. and D.A. Dickey. Testing for Unit Roots in Autoregressive Moving Average of Unknown Order. *Biometrika*, 71, 599-607, 1984.

[37] Said, S.E. Unit Root Test for Time Series Data with a Linear Time Trend. *Journal of Econometrics*, 47:285-303, 1991.

[38] de Carvalho, J.R.P., Assad, E.D., de Oliveira, A.F. and Pinto, H.S. Annual maximum daily rainfall trends in the Midwest, southeast and southern Brazil in the last 71 years. *Weather and Climate Extremes*, 5-6:7–15, 2014.

[39] de Carvalho, J.R.P., Assad, E.D., Evangelista, S.R.M. and da Silveira Pinto, H. Estimation of dry spells in three Brazilian regions- Analysis of extremes. *Atmospheric Research*, 132-133:12-21, 2013. DOI: https://doi.org/10.1016/j.atmosres.2013.04.003.

[40] Adu, D.T. and Denkyirah, E.K. Economic growth and environmental pollution in West Africa: Testing the Environmental Kuznets Curve hypothesis. *Kasetsart Journal of Social Sciences*, 1-8, 2018. DOI: https://doi.org/10.1016/j.kjss.2017.12.008.

[41] Zhang, L., Lin, J., Qiu, R., Hu, X., Zhang, H., Chen, Q., Tan, H, Lin, D. and Wang, J. Trend analysis and forecast of $PM_{2.5}$ in Fuzhou, China using the ARIMA model. *Ecological Indicators*, 95(1), 702-710, 2018. DOI: https://doi.org/10.1016/j.ecolind.2018.08.032.

[42] Hipel, K.W., McLoed, A.I. and Lennox, W.C. Advances in Box Jenkins Modeling. 1. Model Construction. *Water Resources Research*, 13 (3), 567–575, 1977.

[43] Chattopadhyay, S., Jhajharia, D. and Chattopadhyay, G. Univariate modelling of monthly maximum temperature time series over northeast India: neural network versus Yule-Walker equation based appproach. *Meteorological Applications*, 18:70–8, 2011.

[44] Judge, G.G., Griffiths, W.E., Hill, R.C., Lütkepohl, H. and Lee, Tsoung-Chao. The theory and practice of econometrics. In *Wiley Series in Probability and Mathematical Statistics*, 2nd edition, John Wiley and Sons, pages 232-233, ISBN 0-471-84572-8, 1980.

[45] Kotu, V. and Deshpande, B. Time Series Forecasting. *Data Science, 395–445*, 2019. DOI:10.1016/b978-0-12-814761-0.00012-5.

[46] Bowerman, B.L. and O'connel, R.T. *Forecasting and time series: An applied approach*. 3rd edition, 1993.

[47] Profillidis, V.A. and Botzoris, G.N. Trend Projection and Time Series Methods. In *Modeling of Transport Demand*, pages 225-270, 2019. DOI: https://doi.org/10.1016/B978-0-12-811513-8.00006-6.

[48] Arora, S. and Keshari, A.K. ANFIS-ARIMA modelling for scheming re-aeration of hydrologically altered rivers. *Journal of Hydrology*, 601(126635):1-12, 2021. DOI: https://doi.org/10.1016/j.jhydrol.2021.126635.

[49] Khan, M.M.H., Muhammad, N.S. and El-Shafie, A. Wavelet based hybrid ANN-ARIMA models for meteorological drought forecasting. *Journal of Hydrology*, 590(125380):1-9, 2020. DOI: https://doi.org/10.1016/j.jhydrol.2020.125380.