European Journal of Statistics and Probability
Vol.11, No.1, pp.,20-38, 2022
Print ISSN: 2055-0154(Print),
Online ISSN 2055-0162(Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

Forecasting Climatic Variables using Vector Autoregression (VAR) Model

[1]Md. Nuruzzaman, [2]Most Sifat Muntaha Soni, [2]Dr. Md. Shohel Rana, [2]Dr Md. Sazzad Hossain, [*]Md. Matiur Rahman Molla

[1]Bangladesh Food Safety Authority
[2]Department of Statistics, Islamic University, Kushtia, Bangladesh

doi: https://doi.org/10.37745/ejsp.2013/vol11n1119 Published August 2 2023

Citation: Nuruzzaman M., Soni M.S.M., Rana M.S., Hossain M.S., Molla M.M.R. (2023) Forecasting Climatic Variables using Vector Autoregression (VAR) Model, *European Journal of Statistics and Probability*, Vol.11, No.1, pp.,20-38

ABSTRACT: *Although it is a topic of global concern, climate change implementation is typically regional. Finding an adequate Vector Autoregression (VAR) model to predict temperature, rainfall, and cloud coverage for the Jessore region of Bangladesh was the goal of this research project. The stationarity of variables was determined by ADF, PP, and KPSS unit root tests. Granger causality test was used to verify the endogenity among the variables. Employing AIC, VAR (11) model found best. The parameters associated with the model were estimated using the ordinary least square approach. Forecast error variance decomposition and impulse response function were utilized to reveal structural analysis, and the outcome revealed endogenous in the future. The predicted value showed a trend toward increasing temperature and a trend toward decreasing rainfall and cloud coverage.*

KEYWORDS: vector autoregression (VAR), impulse response function, granger causality, forecast error variance (FEV) decomposition, white noise.

INTRODUCTION

In fact, Bangladesh is one of the countries with the most experience with climate change. Most of the parts of her are not so high above sea levels and almost all the part of our country is flooded over year to year. Natural calamities like flood, tsunami, cyclones, storm, and drought are occurred regularly. People who reside in rural areas suffer severe consequences in significant numbers. That is why study of climate related variables badly in need for Bangladesh. Numerous factors influence climate, but those three factors—temperature, rainfall, and cloud coverage—have been chosen for forecasting, revealing relationships, demonstrating past and future behavior, and movement—all of which are crucial for Bangladesh, a country that is particularly vulnerable to climate change. The significance of this research lies in the fact that climate change is not just a local issue, but a national and global issue, and it is an integrated process.

In particular, time series modeling with specially vector Autoregression is a powerful way to extract past and future movements of multiple interrelated variables that are endogenous. In order to describe dynamic behavior in economic and financial time series, Sim (1980) used VAR models for forecasting. It is shown that, compared to the large scale Structural Economic Model, it has better predictive capacity and accuracy. Various research and studies on Climate change and its effects in Bangladesh at various times have been carried out by various governmental and non-governmental organizations and institutions. However, there has not yet been much attention paid to research on climate change behavior, estimation of movement and projections for future periods in southern Bangladesh. So, it is necessary to explore hidden pattern, movement, and their future behavior of restated climate variables.

LITERATURE REVIEW

Many researchers home and abroad have applied VAR model to predict multivariate time series data. A VAR model of democracy and trade balance was applied by Khan and Hossain (2010). Referable numbers of researchers have studied climate variables home and abroad. In order to monitor climate variability in Rajshahi and Rangpur Division, Ferdous and Baten (2011) applied a least square method for analysing the evolution of weather data temperature, rainfall, relative humidity and sunshine. In order to predict the temperature, humidity and cloud coverage in Rajshahi district of Bangladesh, Shahin et al. (2014) used a VAR model. The VAR model was also used by Liu et al., (2011) to study how climate problem predicts major international events, and how climate science feedback influences media and congressional attention on global warming, and climate change. **Adenoma** et al. (2013) used **a** VAR model **to analyze the** dynamic relationship between time series rainfall and temperature data in Niger **State,** Nigeria, and found that bidirectional **causality exists.** Moneta et al. (2011) applied structural VAR models **to search** for **causality.** Awokuse and Bessler (2003) **adopted the** VAR model to the US economy. Kleiber et al. (2013) developed a bivariate stochastic model applied to a daily temperature (minimum and maximum) **dataset** covering **a** complex **landscape in** Colorado, USA, **to investigate** climate **effects,** and successfully **accounted for significant time-varying** non-stationarity **of quarter as a direct covariate** and cross-covariance functions.

MATERIAL AND METHODS

Data

Temperature is an objective comparative measurement of hot or cold. The amount of water falling in rain within a given time and area usually expresses as a unit of measurement such as millimeter, inch or cm is considered as rainfall. Cloud coverage refers to the fraction of the sky obscured by clouds when observed from a particular location. The units of measurement of the considered variables are Celsius, mm, and octas respectively. This research is based on the monthly data of temperature, rain and cloud coverage in Jessore region from 1st January 1983 to 31st December 2015, which was collected by Bangladesh Meteorological Department. That is why sample size is

365. The data has been organized and no missing value has been found. We have arranged, furnished, and tabulated of the original raw data to pursue our objective of the study. The well-known software MS excel, MS word, and R package has been used to arrange this data set as a time series data and subsequent analysis has been conducted by various R packages.

Test of Stationarity:

To study VAR model, it is necessary to know stationarity for each of the variable under study. A time series is said to be stationary if it's mean, variance remain constant over time and autocovariance depends only one the lag period not actual time period (see Gujarity(2003)). The are several test are available for testing stationarity of time series both graphical approach (time series plot and correlogram) and quantitative approach that is well known unit root test ((i) The Augmented Dickey-Fuller test, (ii) Philips-Perron (PP) Test, and (iii) Kwiatkowski-Philips-Schmidt-Shin (KPSS) Test). The time series plot aid us to observe whether a time series is trending (upward or downward) behavior or random walk then it is said to be non-stationary. Autocorrelation function (ACF) and Partial Autocorrelation Function (PACF) against lag are also helpful to identify non-stationarity. Both ADF and PP tests consider the following hypothesis $H_0: \delta = 0$ (non-stationary) against $H_0: \delta < 0$ (stationary). The null hypothesis is rejected if ADF test statistic (tau statistic) is less than the critical value. The PP test is on the basis of same hypothesis like ADF, $H_0: \delta = 0$ (non-stationary) against $H_0: \delta < 0$ (stationary). Unlike the ADF test, which is parametric in nature, the PP test is non-parametric in nature to ensure the serial correlation without the augmented term.

In case of assuming hypothesis, Kwiatkowski-Philips-Schmidt-Shin Test (1992) test differs from the other tests. In this test, the series is assumed to be stationary under null hypothesis. The KPSS statistic is based on the residuals from the ordinary least square (OLS) regression of y_t on the exogenous variables of lagged y_t. For one exogenous variables using one lagged value of y_t, the regression model can be written as follows:

$$y_t = \delta y_{t-1} + u_t$$

Where $\delta = \rho - 1$ and u_t is the error term if $\delta = \rho - 1 = 0$ then the series is said to be nonstationary.

Vector Autoregressive Model

The model was made famous by Chris Sims's paper in 1980 for macro-economic forecasts. After then, VAR are well furnished through many text books and scientific articles. The renowned developers of VAR modeling are (Hamilton, 1994; Johansen, 1995; Hatanaka, 1996; Lu¨tkepohl and Kra¨tzig 2004; Lu¨tkepohl2005; Litterman1986; Canova1995; Sun and Ni 2004; Ni and Sun, 2005; Liu and Theodoridis, and so on).

Selected Variables under Study

In a simultaneous equation model, some variables are treated as endogenous and others as exogenous. Sim (1980) states that all variables in a Vector Autoregressive (VAR) system are endogenous. The idea behind creating a VAR model is that all variables under study are endogenous and usually none are exogenous (Gujarati 1993). Logically, the variables in our study: temperature, precipitation and cloud are endogenous. The endogeneity of variables can be tested using the Granger causality process proposed by Granger (1969) and later popularized by Sims (1972). Finally, variables with endogenous characteristics are selected for VAR analysis.

Making a Model of Order P (Arbitrary)

Let us denote the temperature, rainfall, and cloud coverage by T_t, R_t and C_t; t =1, 2,…, N (Sample size = N), respectively. The three variable VAR model of arbitrary order p can be denoted by VAR (p) and written as:

$$T_t = c_1 + a_{11}^1 T_{t-1} + \cdots + a_{1p}^1 T_{t-p} + a_{11}^2 R_{t-1} + \cdots + a_{1p}^2 R_{t-p} + a_{11}^3 C_{t-1} + \cdots + a_{1p}^3 C_{t-p} + \varepsilon_{1t}$$

$$R_t = c_2 + a_{21}^1 T_{t-1} + \cdots + a_{2p}^1 T_{t-p} + a_{21}^2 R_{t-1} + \cdots + a_{2p}^2 R_{t-p} + a_{21}^3 C_{t-1} + \cdots + a_{2p}^3 C_{t-p} + \varepsilon_{2t}$$

$$C_t = c_3 + a_{31}^1 T_{t-1} + \cdots + a_{3p}^1 T_{t-p} + a_{31}^2 R_{t-1} + \cdots + a_{3p}^2 R_{t-p} + a_{31}^3 C_{t-1} + \cdots + a_{3p}^3 C_{t-p} + \varepsilon_{3t}$$

In matrix notation, we can write

$$\begin{pmatrix} T_t \\ R_t \\ C_t \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} + \begin{pmatrix} a_{11}^1 & a_{11}^2 & a_{11}^3 \\ a_{21}^1 & a_{21}^2 & a_{21}^3 \\ a_{31}^1 & a_{31}^2 & a_{31}^3 \end{pmatrix} \begin{pmatrix} T_{t-1} \\ R_{t-1} \\ C_{t-1} \end{pmatrix} + \cdots + \begin{pmatrix} a_{1p}^1 & a_{1p}^2 & a_{1p}^3 \\ a_{2p}^1 & a_{2p}^2 & a_{2p}^3 \\ a_{3p}^1 & a_{3p}^2 & a_{3p}^3 \end{pmatrix} \begin{pmatrix} T_{t-p} \\ R_{t-p} \\ C_{t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}$$

Therefore, the reduced form of VAR process of order p can be written as follows

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \varepsilon_{t,} \qquad t = 1, 2, ..., N$$

Where, y_t is a 3×1 vector, c is a 3×1vector of constants (intercept), A_i is a 3×3 matrix (for each $i = 1, 2, ..., p$), and ε_t is distributed as $\varepsilon_t \sim NID(0,\Omega)$.

Lag order Selection:

The first step in building a VAR model is to choose the order of the VARs which can be determined by Akaike Information Criteria (AIC) (Akaike, 1974). The mathematical form of AIC is as follows:

$$AIC = \log\left|\hat{\Omega}(p)\right| + \frac{2m(p^2+1)}{N}$$

With $\hat{\Omega}(p) = N^{-1}\sum_{i=1}^{N} \varepsilon_t \hat{\varepsilon}_t'$ and $m(p^2+1)$ is the total number of the parameters in each equation, m is the number of equation or variables in VAR model, and p determines the lag order.

Parameter Estimation

The OLS estimate of the considered equation can be briefly expressed as follows:

$$Y = BZ + U$$

Where, $= (c, A_1, A_2, \cdots, A_p)$; $Z_t = [1\ y_t, \ldots, y_{t-p+1}]'$and $U = [u_1, u_2, \cdots, u_T)]'$

It may be noted that the multivariate LS estimator $\hat{\beta}$ is identical to the ordinary LS (*OLS)* estimator obtained by minimizing

$$\bar{S}(\beta) = u'u = [y - (Z' \otimes I_k)\beta]'[y - (Z' \otimes I_k)\beta]$$

Now, the LS estimate of the problem can be found after solving this matrix operation

$$\boldsymbol{\hat{\beta}} = ((\boldsymbol{ZZ}')^{-1}\boldsymbol{Z}\otimes\boldsymbol{I_k})[(\boldsymbol{Z}' \otimes \boldsymbol{I_k})\boldsymbol{\beta} + \boldsymbol{u}]$$

Diagnostic Checking

Several diagnostic tests criteria are required to declare a model to be a good fit, including: Q-Q plot to check normality of residuals, unit root to check stationary residuals, and Durbim-Watson (1951) d-test to check autocorrelation. Outliers were checked using standardized residual plots.

Forecasting:

Forecasting is predicting the value of a variable based on known values in the past. In VAR models, this prediction also depends on the lag values of other endogenous variables. The compacted form of the VAR model is suitable for forecasting because it represents the conditional mean of the stochastic process. The one-step ahead forecasts are represented as follows:

A VAR (P)-model is given as

$$\hat{y}_{t+1} = \hat{A}_1 y_t + \hat{A}_2 y_{t-1} + \cdots\cdots + \hat{A}_p y_{t-p+1}$$

And successively we will find $\hat{y}_{t+2}, \hat{y}_{t+3}, \ldots.., \hat{y}_{t+h}, \ldots..$

Forecast Error Variance Decompositions

The Forecast Error Variance Decomposition (FEVD) quantify the question, how much of the forecast error variance is caused by the structural shock? Using orthogonal shocks η_t, the h-step forward forecast error vector with well-known VAR determinants can be written as

$$Y_{T+h} - Y_{T+h|T} = \sum_{s=0}^{h-1} \Theta_s \eta_{T+h-s}$$

Where $Y_{T+h|T}$ is h-step forecasts based on information available at time T

For a particular variable $Y_{i,T+h}$, this forecast error has the form

$$Y_{i,T+h} - Y_{i,T+h|T} = \sum_{s=0}^{h-1} \theta_{i1}^{s} \eta_{1,T+h-s} + \cdots + \sum_{s=0}^{h-1} \theta_{in}^{s} \eta_{n,T+h-s}$$

Where $\sigma_{\eta j}^2 = \text{var}(\eta_{jt})$, is the fraction of var($Y_{i,T+h} - Y_{i,T+h|T}$) cause by the shock η_j,

$$FEVD_{i,j}(h) = \frac{\sigma^2_{\eta j}\sum_{s=0}^{h-1}(\theta^s_{ij})^2}{\sigma^2_{\eta 1}\sum_{s=0}^{h-1}(\theta^s_{i1})^2 + \cdots + \sigma^2_{\eta n}\sum_{s=0}^{h-1}(\theta^s_{in})^2}, i, j = 1, \ldots n$$

A VAR with n variables has n^2 $FEVD_{i,j}(h)$ values. It is important to note that FEVD is dependent on the causal order used to determine structural shocks (η_t), and is not unique. Alternatively, different sequences of causal produce different FEVD values.

Impulse Response Functions

Any covariance stationary VAR (p) process has a Wold representation of the form

$$Y_t = \mu + \varepsilon_t + \Psi_1\varepsilon_{t-1} + \Psi_2\varepsilon_{t-2} + \cdots\cdots$$

Where the (n × n) moving average matrices Ψ_s are determined recursively using

$$\Psi_s = \sum_{j=1}^{p}\Psi_{s-j}\Pi_j$$

It is tempting to interpret the (i, j)-th element, Ψ^s_{ij}, of the matrix Ψ_s as the dynamic multiplier or impulse response i.e. Ψ^s_{ij} represents the effect of unit shocks on system variables.

$$\frac{\delta yi,t+s}{\delta\varepsilon j,t} = \frac{\delta yi,t}{\delta\varepsilon j,t-s} = \Psi^s_{ij} \qquad i, j = 1, 2, \ldots\ldots\ldots.., n$$

However, this illustration is only possible if var (ε_t) = Σ is a diagonal matrix in which the elements of ε_t are uncorrelated.

RESULTS AND DISCUSSION

Selecting Variables under Study:

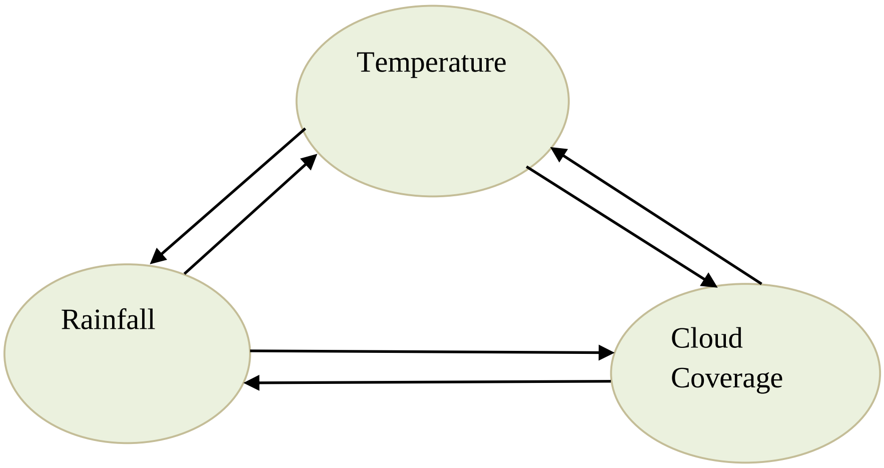

Figure 1. Graph for the course of causality

Table 1. F statistics for the pair wise Granger causality test on selected variables

Null Hypothesis	Lag								
	1	2	3	4	5	6	7	8	9
$R \nrightarrow T$	7.61*	9.41*	24.84*	32.29*	34.72*	23.20*	9.91*	7.71*	7.97*
$T \nrightarrow R$	76.15*	51.34*	40.22*	28.95*	20.82*	14.74*	12.26*	10.33*	9.41*
$C \nrightarrow R$	89.49*	60.04*	37.99*	25.77*	14.05*	11.47 *	8.44*	6.54*	6.68*
$R \nrightarrow C$	1.50	0.81	1.66	5.46*	5.95*	6.06*	5 .50*	4.83*	3.91*
$T \nrightarrow C$	85.86*	34.95*	16.81*	7.46*	9.46*	11.72*	16.59*	17.16*	14.55*
$C \nrightarrow T$	4 .56	49.70*	87.62*	82.65*	75.99*	49.87*	27.35*	20.83*	20.64*

***Remark**:* The symbol $\nrightarrow$ means "does not granger cause". The asterisk (*) indicates the statistical significance at the 1% level.

Lag Order Selection:

Table 2. Determination of lag length of VAR (P) model

Lag	Log L	AIC
1	-3578.45	11.296
2	-3436.266	10.611
3	-3320.624	10.112
4	-3254.022	9.863
5	-3206.635	9.684
6	-3175.795	9.613
7	-3149.915	9.557
8	-3118.766	9.457
9	-3096.662	9.427
10	-3071.906	9.377
11	-3031.258	9.231

Comment: From above table, AIC indicates that the adequate model is VAR (11) regarding our research variables.

Parameter Estimation

The parameter of VAR (11) model are estimated by Ordinary Least Square (OLS). Estimated value of the parameter of VAR (11) model including constant and trend are given as follows:

Table 3. The estimated coefficient of VAR (11) model

Variables	R_t	T_t	C_t
const	$\hat{c}_1 = -979.4^{*}$	$\hat{c}_2 = 31.0528^{***}$	$\hat{c}_3 = 2.42$
R_{t-1}	$\hat{a}^1_{11} = 0.0251$	$\hat{a}^1_{21} = -0.0007$	$\hat{a}^1_{31} = -0.0008$
R_{t-2}	$\hat{a}^1_{12} = 0.0377$	$\hat{a}^1_{22} = -0.0003$	$\hat{a}^1_{32} = -0.00040$
R_{t-3}	$\hat{a}^1_{13} = -0.0421$	$\hat{a}^1_{23} = 0.0006$	$\hat{a}^1_{33} = 0.0001$
R_{t-4}	$\hat{a}^1_{14} = 0.0323$	$\hat{a}^1_{24} = 0.0004$	$\hat{a}^1_{34} = -0.0007$
R_{t-5}	$\hat{a}^1_{15} = 0.0985$	$\hat{a}^1_{25} = -0.0002$	$\hat{a}^1_{35} = -0.0003$
R_{t-6}	$\hat{a}^1_{16} = 0.0516$	$\hat{a}^1_{26} = -0.0007$	$\hat{a}^1_{36} = 0.0006$
R_{t-7}	$\hat{a}^1_{17} = 0.0217$	$\hat{a}^1_{27} = 0.0002$	$\hat{a}^1_{37} = -0.0005$
R_{t-8}	$\hat{a}^1_{18} = 0.0488$	$\hat{a}^1_{28} = 0.0013*$	$\hat{a}^1_{38} = -0.0008$
R_{t-9}	$\hat{a}^1_{19} = 0.0423$	$\hat{a}^1_{29} = -0.0006$	$\hat{a}^1_{39} = -0.00031$
R_{t-10}	$\hat{a}^1_{110} = 0.0896$	$\hat{a}^1_{210} = -0.0009$	$\hat{a}^1_{310} = 0.00006$
R_{t-11}	$\hat{a}^1_{111} = 0.0463$	$\hat{a}^1_{211} = 0.0061$	$\hat{a}^1_{311} = -0.0001$
T_{t-1}	$\hat{a}^2_{11} = 4.6800$	$\hat{a}^2_{21} = 0.3486***$	$\hat{a}^2_{31} = -0.0248$
T_{t-2}	$\hat{a}^2_{12} = 0.0652$	$\hat{a}^2_{22} = -0.0674$	$\hat{a}^2_{32} = 0.0245$
T_{t-3}	$\hat{a}^2_{13} = 7.528$	$\hat{a}^2_{23} = -0.1414*$	$\hat{a}^2_{33} = 0.0276$
T_{t-4}	$\hat{a}^2_{14} = 0.1297**$	$\hat{a}^2_{24} = -0.1809**$	$\hat{a}^2_{34} = 0.0298$
T_{t-5}	$\hat{a}^2_{15} = -5.277$	$\hat{a}^2_{25} = 0.1078$	$\hat{a}^2_{35} = -0.0635*$
T_{t-6}	$\hat{a}^2_{16} = 0.0824$	$\hat{a}^2_{26} = 0.0482$	$\hat{a}^2_{36} = -0.0160$
T_{t-7}	$\hat{a}^2_{17} = -0.594$	$\hat{a}^2_{27} = 0.0331$	$\hat{a}^2_{37} = -0.0032$
T_{t-8}	$\hat{a}^2_{18} = 1.707$	$\hat{a}^2_{28} = -0.0719$	$\hat{a}^2_{38} = -0.0569$
T_{t-9}	$\hat{a}^2_{19} = -2.817$	$\hat{a}^2_{29} = -0.1561*$	$\hat{a}^2_{39} = -0.0692*$
T_{t-10}	$\hat{a}^2_{110} = 4.963$	$\hat{a}^2_{210} = -0.0813$	$\hat{a}^2_{310} = 0.0777*$
T_{t-11}	$\hat{a}^2_{111} = 5.070$	$\hat{a}^2_{211} = 0.2692***$	$\hat{a}^2_{311} = 0.0894*$
C_{t-1}	$\hat{a}^3_{11} = 11.32$	$\hat{a}^3_{21} = 0.1195$	$\hat{a}^3_{31} = 0.1848**$
C_{t-2}	$\hat{a}^3_{12} = -0.192$	$\hat{a}^3_{22} = 0.1844$	$\hat{a}^3_{32} = 0.1383*$
C_{t-3}	$\hat{a}^3_{13} = 10.66$	$\hat{a}^3_{23} = -0.3402$	$\hat{a}^3_{33} = 0.0854$
C_{t-4}	$\hat{a}^3_{14} = -2.56$	$\hat{a}^3_{24} = -0.3328$	$\hat{a}^3_{34} = 0.0550$
C_{t-5}	$\hat{a}^3_{15} = -5.277$	$\hat{a}^3_{25} = 0.1298$	$\hat{a}^3_{35} = -0.0911$
C_{t-6}	$\hat{a}^3_{16} = 0.824$	$\hat{a}^3_{26} = -0.0212$	$\hat{a}^3_{36} = -0.1744**$
C_{t-7}	$\hat{a}^3_{17} = 0.5943$	$\hat{a}^3_{27} = 0.0101$	$\hat{a}^3_{37} = 0.0883$
C_{t-8}	$\hat{a}^3_{18} = 2.147$	$\hat{a}^3_{28} = -0.1842$	$\hat{a}^3_{38} = 0.0183$

C_{t-9}	$\hat{a}_{19}^{3} = 17.55*$	$\hat{a}_{29}^{3} = 0.1154$	$\hat{a}_{39}^{3} = -0.0180$
C_{t-10}	$\hat{a}_{110}^{3} = -13.66$	$\hat{a}_{210}^{3} = -0.0266$	$\hat{a}_{310}^{3} = 0.1091$
C_{t-11}	$\hat{a}_{111}^{3} = 12.81$	$\hat{a}_{211}^{3} = 0.2877$	$\hat{a}_{311}^{3} = 0.291***$
Trend	-0.05788	0.00239*	-0.002047***

Note: The asterisk ***, **, and * demonstrates the statistical significance at the 1%, 5%, and 10% levels, respectively.

Diagnostic checking of VAR model

Diagnostic checking is an important step in the model building process. In this section, we measured the fit of the VAR model. After fitting the VAR model, we should check whether our assumptions have been met or not. If the default setting does not satisfy the selected model, it may be misleading. In most modeling cases, overall goodness of fit is assessed by the nature of the residuals. Thus, it is necessary to quickly assess the residual behavior before claiming model adequacy.

Stationary and autocorrelation test of residual:

The ADF test assumed the following hypothesis.

H_0: The series is nonstationary

H_1: The series is stationary

Table 4. Augmented Dickey-Fuller test for residual obtained from VAR(11) model

Series of Residuals	ADF value	P-Value	Level of significance	Decision
Rainfall	-6.233	0.01	0.05	Stationary
Temperature	-6.431	0.01	0.05	Stationary
Cloud Coverage	-5.943	0.01	0.05	Stationary

Durbin Watson test assumed the following hypothesis

H_0: $\rho = 0$ i.e. there is no autocorrelation.

H_1: $\rho \neq 0$ i.e. there is autocorrelation.

Table 4. Durbin-Watson test for checking autocorrelation

Series of Residuals	DW- Value	Comment
Rainfall	2.004	No Autocorrelation
Temperature	2.088	No Autocorrelation
Cloud Coverage	2.067	No Autocorrelation

Normal Q-Q plot:

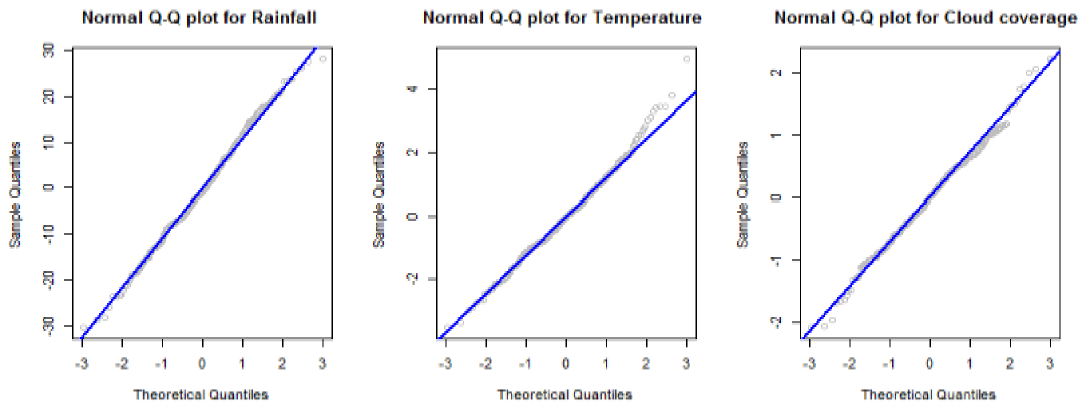

Figure 2. Normal Q-Q plot for the residuals of VAR (11) model

Comments: From the Q-Q plot, it is appeared to be normally distributed, because it is observed that the points of all residual for our study variable (Rainfall, temperature and cloud coverage) lie quite close to the straight line; close enough to say these data come from a normal distribution. It is also seen that a few number of random point rotate about the line; this does not incapacitate these data from being normal.

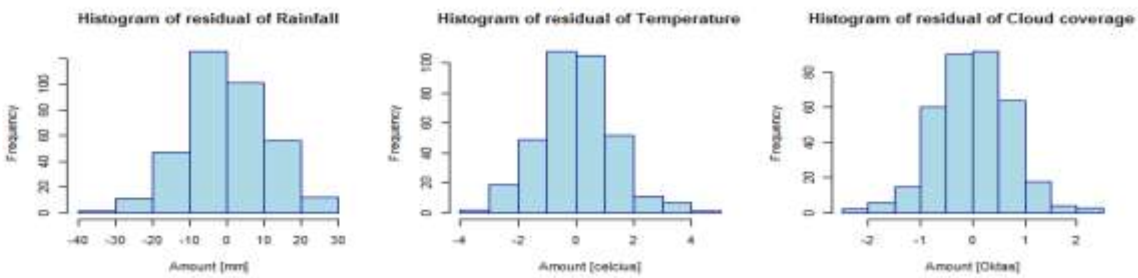

Figure 3. Histogram of the residual of VAR (11) model

Comment: The histogram of residual VAR (11) model's variables are approximately normally distributed.

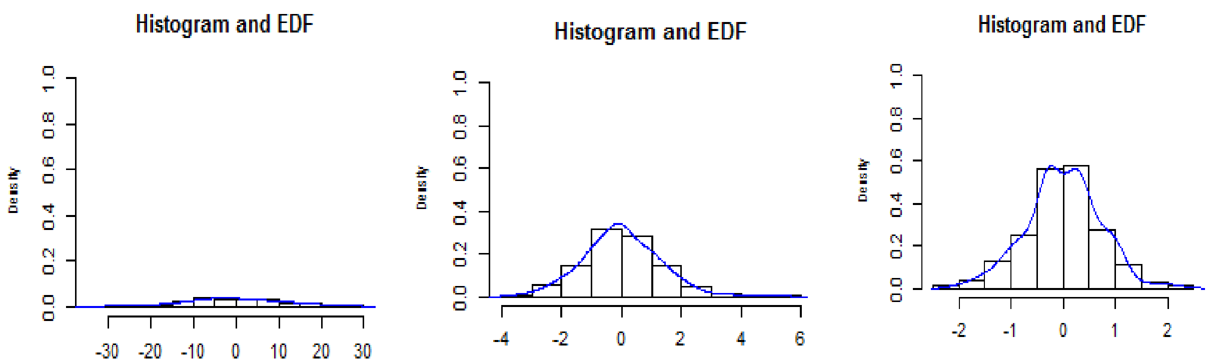

Figure 4.(a) Figure 4.(b) Figure 4.(c)

Figure 4. EDF of the residual of VAR (11) model of Rainfall, Temperature and, Cloud Coverage.

Comment: From the above histogram and empirical distribution function (EDF), it can be seen that the residual of rainfall, temperature and cloud coverage for our selected VAR model are approximately normally distributed. It also exhibit that by using VAR (11) model the histogram of residual for rainfall, temperature and cloud coverage are roughly platykurtic (flat), leptokurtic and leptokurtic (thin) respectively.

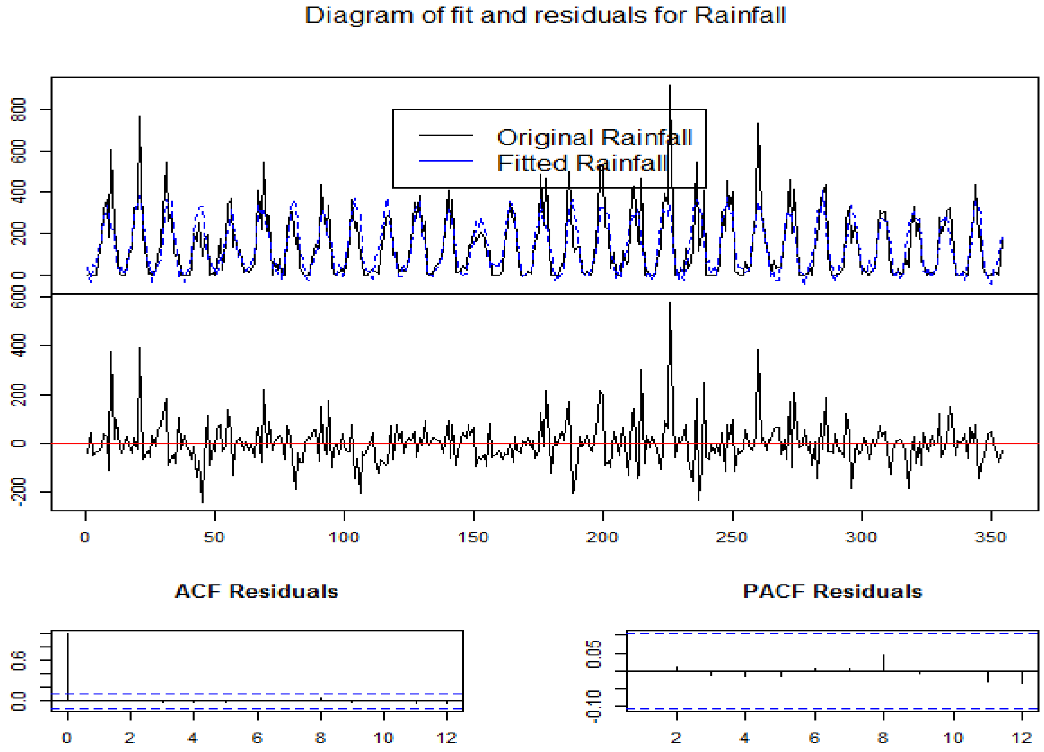

Figure 5. Comparison of original value and fitted value, residual and ACF and PACF of residual for rainfall

Comment: From the above graph, we see that the original rainfall and fitted rainfall by using VAR (11) model are approximately resemblance. The residuals are fluctuating positively and negatively around the line zero. The residual of the VAR (11) model's autocorrelation and partial autocorrelation graphs support the assumption of white noise.

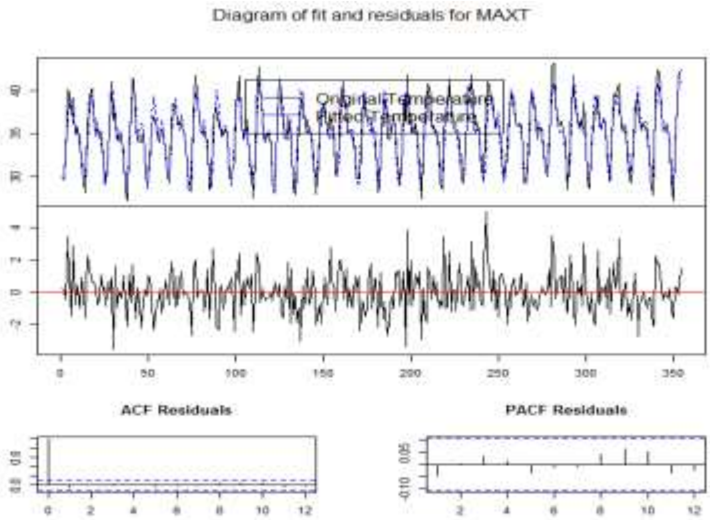

Figure 6. Comparison of original vs fitted value, residual and ACF and PACF of residual for maximum temperature

Comment: We notice from the preceding graph that the initial temperature and the estimated temperature by using VAR (11) model are identical. Around the line zero, the residuals are varying positively as well as negatively. The residual of the VAR (11) model's autocorrelation and partial autocorrelation graphs support the assumption of white noise.

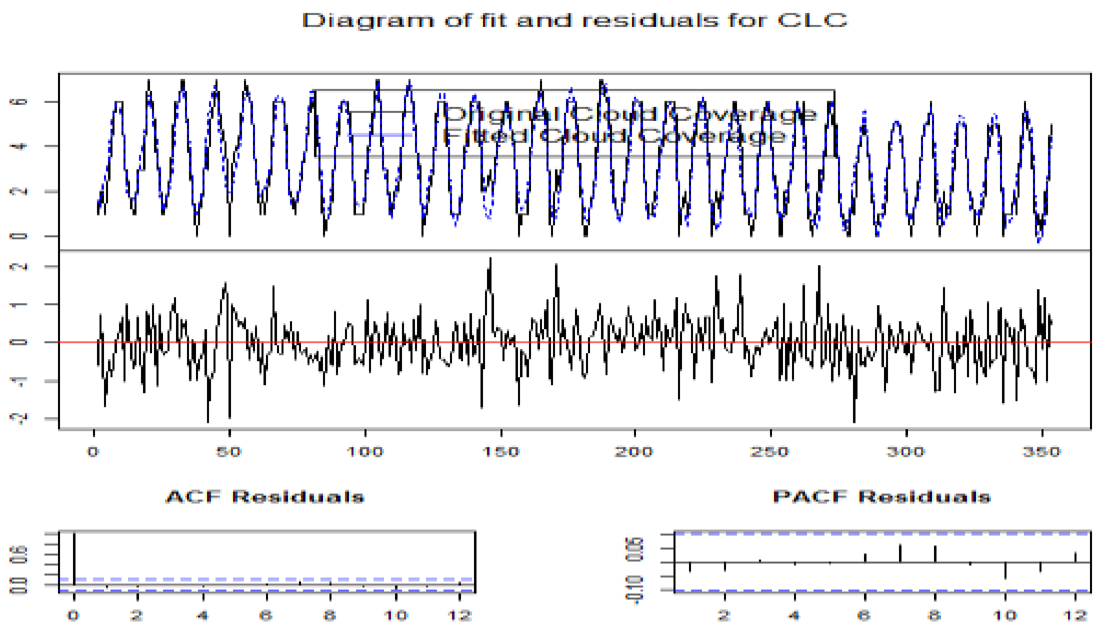

Figure 7. Comparison of original vs fitted value, residual and ACF and PACF of residual for cloud coverage

Comment: The aforementioned graph demonstrates that the initial amount of cloud coverage and the amount of cloud coverage that was fitted using the VAR (11) model are about the same. Around the line zero, the residuals oscillate in a positive and negative direction. White noise is supported by the residual of the Autocorrelation & Partial Autocorrelation Graphs from the VAR model (11).

Outlier Detection by Standardized Residual:

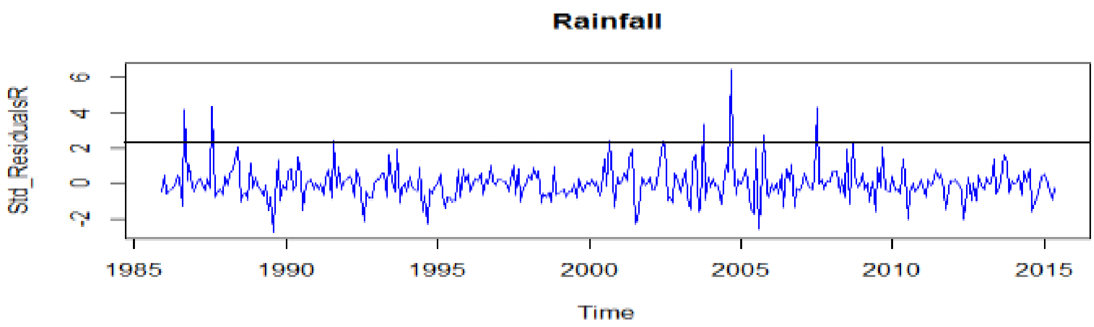

Figure 8. Outlier detection plot for Rainfall

Comment: We see that there are five residuals for the series of rainfall falls outside 3.0, with their absolute values being regarded as outliers, from the standardized residuals in addition to from the graphical depiction. September 1986 = 4.134, August 1987 = 4.318, October 2003 = 3.349, September 2004 = 6.424, and July 2007 = 4.253 are the values of the five residuals.

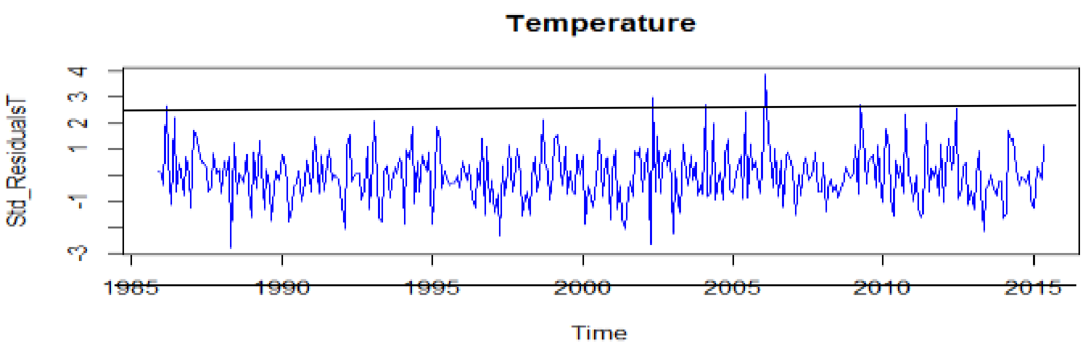

Figure 9. Outlier detection plot for Temperature

Comment: Similar to this, it appears that there is a residual for the series temperature that is greater than 3.0, measured in absolute terms. February 2006 = 3.85 is the value of the residuals.

European Journal of Statistics and Probability
Vol.11, No.1, pp.,20-38, 2022
Print ISSN: 2055-0154(Print),
Online ISSN 2055-0162(Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

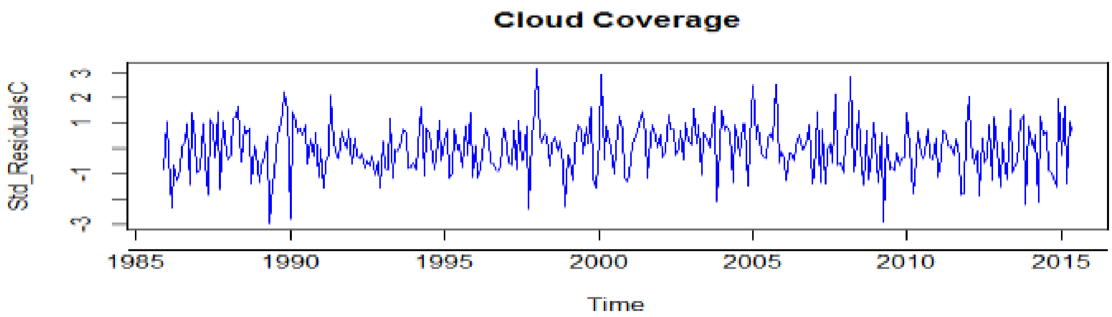

Figure 10. Outlier detection plot for Cloud Coverage

Comment: We can see from the aforementioned graphs that there is a residual with an absolute value for the series where cloud coverage is outside of 3.0. A tiny outlier can be identified in the residual value of January 1998, which is 3.163, but it has a very little effect overall the model's performance.

Forecast Error Variance Decomposition:
The time is referred to here as a month because the study data are collected on a monthly basis. The following table displays the FEV decomposition value for the predicted horizons of 1, 6, 12, 18, and 24 months. The decomposition separates the forecast variance into various components each with their own innovations to explain them.

Table 5. Forecast Error Variance Decomposition of VAR (11) Model.

FEV in	Period(month)	Rainfall	Temperature	Cloud Coverage
Rainfall	01	100.000	00.000	00.000
	06	93.053	05.267	01.678
	12	87.540	06.259	06.200
	18	84.771	07.151	05.077
	24	82.010	08.210	09.778
Temperature	01	02.253	97.746	00.000
	06	02.485	90.275	07.238
	12	05.366	86.406	08.226
	18	04.455	82.669	12.875
	24	05.107	81.823	13.069
Cloud Coverage	01	10.268	37.212	86.009
	06	10.784	05.065	84.151
	12	10.785	10.085	79.129
	18	10.471	09.632	79.897
	24	09.758	12.587	77.672

Comments: The table shows the impact of temperature, cloud coverage, and rainfall on each other. After a year (12 periods), the rainfall appeared to be the less exogenous factor in the system that explained more than 85% of its FEV. More than 6% of the fluctuation was accounted for by the

European Journal of Statistics and Probability
Vol.11, No.1, pp.,20-38, 2022
Print ISSN: 2055-0154(Print),
Online ISSN 2055-0162(Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

temperature and cloud coverage. Additionally, we notice that the erogeneity of rainfall is steadily waning while temperature and cloud cover accounted for more than 8% and 9% of the total for the upcoming year's (24) periods, respectively. After a year, the temperature alone accounted for more than 85% of the FEV. Rainfall and cloud cover accounted for more than 5% and 8% of the temperature fluctuation, respectively. Additionally, we observed that rainfall and cloud coverage accounted for more than 5% and 13% of the increase in the next year's (24) periods' rainfall and exogenity of temperature, respectively. After a year, the cloud coverage itself accounted for more than 79% of the FEV. Additionally, temperature and precipitation contributed more than 10% of the FEV to the variance in cloud cover. Therefore, there is definitely a relationship among climate-related variables.

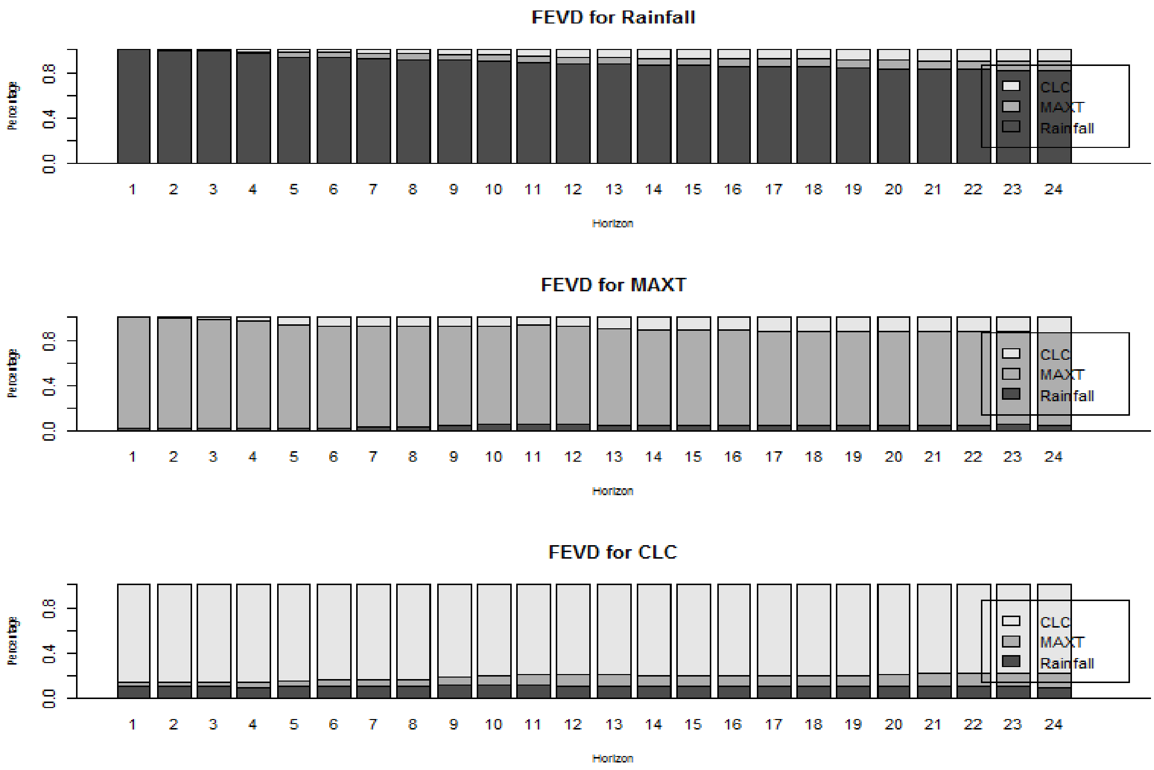

Figure 11. Graphical representations of Forecast Error Variance Decomposition (FEVD) of selected variables

Impulse Response Function for the VAR (11) Model

To find out the response of an endogenous variable to a change in one of the innovations in the VAR system is the main task of Impulse Response Function (IRF). Impulse response analysis is a statistical technique for investigating the relations between the variables in a VAR model. In this function X axis lies in the horizon on the horizontal and response on Y axis lies in the vertically. It traces the effect of a one standard deviation shock to one of the innovations on current and future values of the dependent variables through the dynamic structure of the VAR.

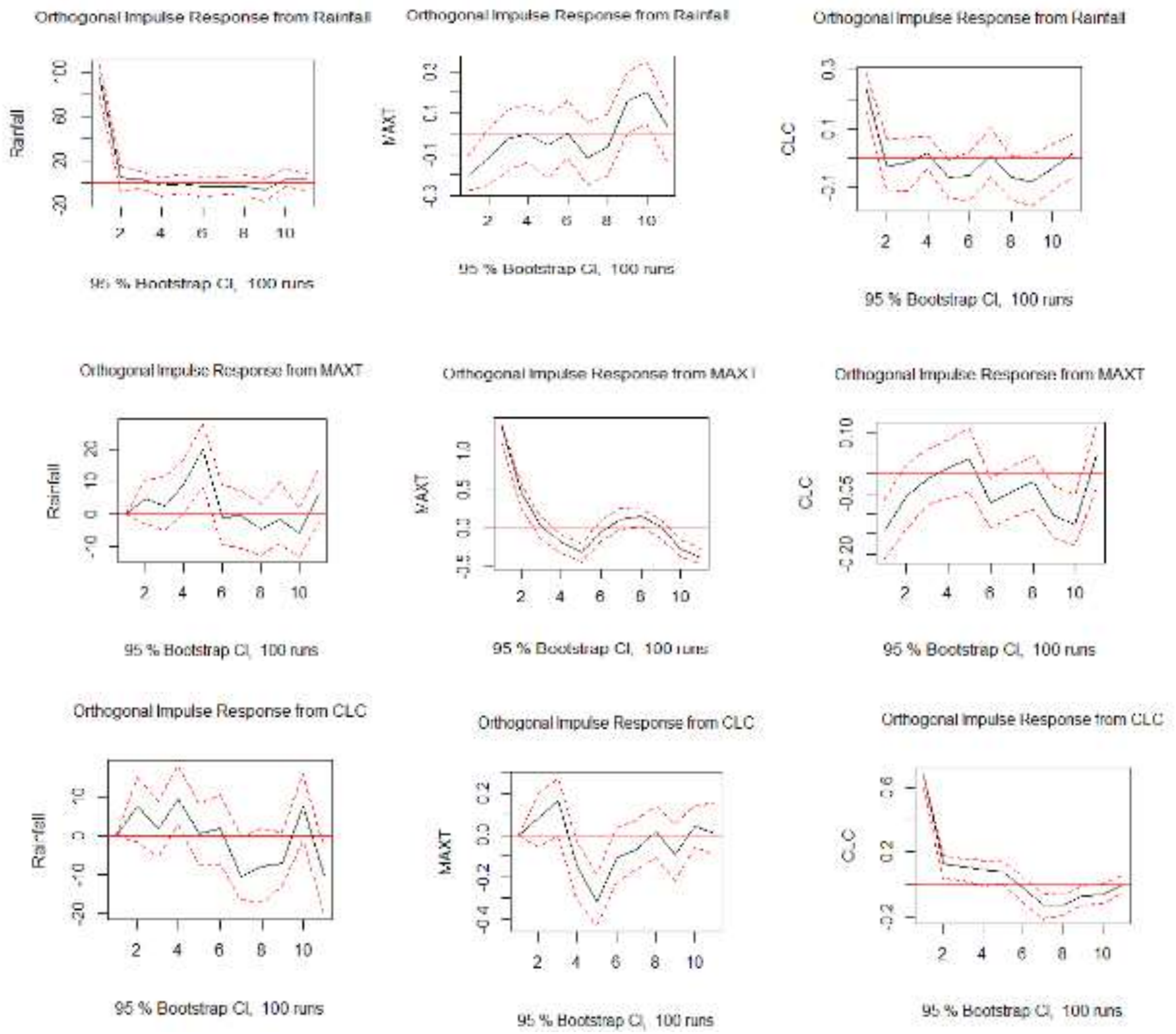

Figure 12. Impulse Response Function (IRF) from the cholesky decomposition for VAR (11) model. The scale in the X axis indicate the lag in month

Comment: From the graph we observe over the eleven month period considered a shock in rainfall has significant impact on rainfall up to three months into the future and then the impact of the effect dies out quickly and again showed significant after 9 month.. It also seen that the impulse response from rainfall to temperature is gradually increasing. Also response from rainfall to cloud coverage showed a significant impact on second and eleven month. Temperature has a significant impact on itself after 4 month. And orthogonal response from temperature to cloud coverage the response of the amount cloud coverage has an obvious fluctuation; there is lowest negative effect on the first month, highest effect on the eleven month. And at last the orthogonal impulse form cloud coverage to rainfall, temperature and itself, it is seen that about half response of rainfall is positive and rest of the half is negative. Cloud coverage has significant impact on itself up to three months into the future and then the impacts of the effect are diminishing after time to time.

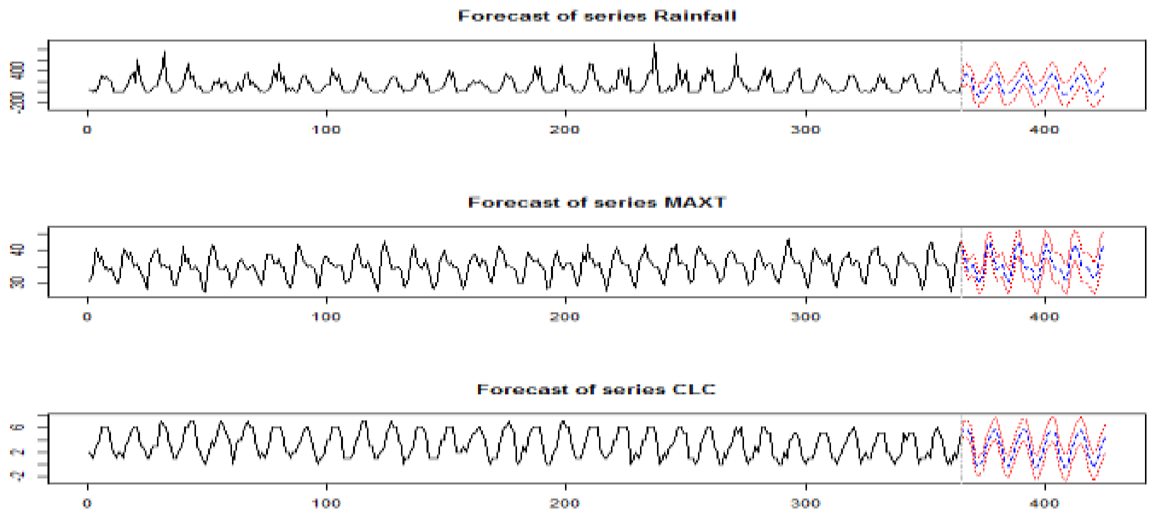

Fig.1.14: Forecasted plot from the fitted VAR (11) model of the climatic variables of rainfall, temperature, and cloud coverage.

CONCLUSION

Three significant climatic variables were examined in this study: rainfall, temperature, and cloud coverage. Also bidirectional causality detected among these variables under granger causality and it shows all the variables are interrelated which support use of VAR analysis. Appropriate VAR order was determined using AIC and found that order 11gives the best selection. At the diagnostic checking step, VAR (11) model reveals the residual are non-autocorrelated, stationary and approximately normally distributed. From the model, it is clear that more than 61%, 85%, and 87% of endoginoty of the climatic variable can be explained by rainfall, temperature and cloud coverage regression equation. Forecast error variance decomposition (FEVD) was utilized, and it shows that after a year (12 periods), rainfall proved to be the less exogenous factor in the system, explaining more than 85% of its FEV. More than 6% of the fluctuation was accounted for by the temperature and cloud coverage. After one year, temperature accounted for more than 85% of FEV. Rainfall and cloud coverage quantify for more than 5% and 8% FEV of the temperature fluctuation, respectively. After a year, the cloud coverage itself accounted for more than 79% of the FEV. Once more, in the variability of cloud coverage, temperature and rainfall contributed for more than 10% of the FEV. Therefore, there is a relationship between the climatic factors.

REFERENCES

Adenomon, M. O.& Ojehomon, V.E.T. Oyejola, B. A (2013). Modelling the dynamic relationship between rainfall and temperature time series data in Niger State, Nigeria, Mathematical Theory and Modeling Vol.3, No.4, 2013 pp: 53-71

Adenomon, M. O. & Oyejola, B. A. Adenomon, C. A. (2012): On the relationship between savings and investment in Nigeria: A Vector Autoregressive Analysis Approach, Journal of the Nigerian Statistical Association. 24:45-58.

Ahmed R, & Karmakar, S. (1993). Arrival and withdrawal date of the summer monsoon in Bangladesh, International Journal of Climatology 13: 727–740.

Ahmed, R, & Kim I. (2013). Patterns of daily rainfall in Bangladesh during the summer monsoon season: case studies at three stations, Physical Geography 24(4): 295–318.

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Contr

Akaike, H. (1969). Fitting autoregressive models for prediction. Ann Inst Stat Math 21:243–247

Altaf, M.Z.& Arshad, I.A. Ilys,M.R. (2012) Vector autoregression application on macroeconomic variables of Pakistan's economic growth ,Sindh Univ Res J 44(2): 267-272

Anderson, T.W. (1984). An introduction to Multivariate Statistical Analysis, 2nd ed., New York: John Wiley and Sons Inc.

Awokuse,T.O.& Bessler, D.A. (2003). Vector Autoregressions, Policy Analysis, And Directed Cyclic Graphs, An Application To The U.S. Economy, Journal of Applied Economics, Vol. VI, No. 1 (May 2003), 1-24.

Azad, A.S. & Hasan, M.K. Rahman M.A.I, Rahman M.M. (2014). Exploring the Behavior and Changing Trends of Rainfall and Temperature Using Statistical Computing Techniques . "Computational Intelligence Techniques in Earth and Environmental Sciences" pp. (79-81).

Bovas A.& May, A.C.(1989). Outlier Detection and Time Series Modeling, Taylor& Francis, Ltd. on behalf of American Statistical Association andAmerican Society for Quality " Vol. 31, No. 2 (May, 1989), pp. 241-248

Box, G.E.P. & Jankins, G.W. Reinsel G.(1994). Time Series Analysis, Forecasting and Control, 3rd edition, Englewood Cliffs, N.J.; Prentice-Hall.

Brillinger, D.R. (1989). Consistent detection of a monotonic trend superimposed on a stationary time series, Biometrika, Vol.76, pp.23-30

Brokwell,P. J.& Davis,R. A. (1996) .Introduction to Time Series and Forecasting, Springer, New York

Dickey,D.A. & Fuller,W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. J Am Stat Assoc 74:427–431

Durbin,J. & Watson, G.S. (1951). Testing for serial correlation in least square regression, Biometrica 38: 159-17.

Engle, R.F.& Granger C.W.J. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing." Econométrica 55:251-76.

Ferdous, M.G.& Baten, M.A. (2011). Climatic variables of 50 years and their trends over Rajshshi and Rangpur division". J Environ Sci Nat Resour 4(2): 147-150

Gebhard K. & Wolters,J.(2007). Introduction to Modern Time Series Analysis, Springer, New York.

Granger, C.W.J. (1969). Investigating Causal Relations by Econometric Models and Cross Spectral Methods, Econometrica, 37, 424-438.

Gujarati, D. (2002). Basic Econometrics, 4th edition, McGraw-Hill.

Hamilton, J. (1994). Time Series Analysis, Princeton University Press,NJ.

Hannan, E.J. & Quinn, B.G. (1979). The determination of the order of an autoregression. J R Stat Soc B 41:190–195

Hatanaka, M. (1996). Time series based econometrics: unit roots and co-integration. Oxford University Press, Oxford

Ivanov, I.& Killian, L. (2001). A practitioner guide to lag order selection for vector autoregressions. CEPR discussion paper no. 2685.Center for Economic Policy Research, London.

Janjua, P.Z., Khan, S.J. NU (2010) Impact of climate change on wheat production: A case study of Pakistan. Pak Dev Rev 49(4):799-821

Karmakar,S.& Shrestra, ML (2000). Recent climate change in Bangladesh, SMRC No, SMRC, Dhaka.

Khan, M.Z.S.& Hossain, M.I. (2010). Democracy and trade balance: a vector autoregressive analysis Bangladesh Dev Stud 33(4): 23-27.

Kilian, L.(2001). Impulse response analysis in vector autoregressions with unknown lag order. Journal of Forecastin. 20, 161-179.

Kleiber, W. &Katz R.W. Rajagopalan, B, (2013). DAILY MINIMUM AND MAXIMUM TEMPERAT URE SIMULATION OVER COMPLEX TERRAIN. *The Annals of Applied Statistics*Vol. 7, No. 1, 588–612

Kwiatkowski, D. & Philips, P.C.B. Schmidt, P. Shin, Y. (1992). Testing the Null hypothesis of stationary against the alternative of a unit root, Journal of Econometrical 54: 159-178.

Lütkepohl, H. & Poskitt, D.S.(1996). Specification of Echelon-Form VARMA Models, Biometrika.

Lütkepohl, H. (2005), New Introduction to Multiple Time Series Analysis", Springer, New York.

Lütkepohl, H. (2005). New Introduction to Multiple Time Series Analysis", Berlin: Springer.

Mann, H.B. (1945), Nonparametric tests against trend, Econometrica, 13, 245-259.

Moneta, A.& Chlab, N. Entner, D. Hoyer, P.(2011).Causal search in structural vector autoregressive models. Workshop ConfProc 12:95-118

Moseale TJ, Stephenson DB, Collins M, Mills TC (2006) Granger causality of coupled climate processes: Ocean feedback on the North Atlantic oscillation. Journal of Climate 19:1182-1194

Shahid,S.(2009).Rainfall variability and the trends of wet and dry periods in Bangladesh,International journal of climatology 30:2299–2313 (2010) Published online 20 November 2009 in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/joc.2053

Shahin, M.A.& Ali,A.Ali A.B.M.S. (2014). Vector Autoregression (VAR) Modeling and Forecasting of Temperature, Humidity, and Cloud Coverage, Computational Intelligence Techniques in Earth and Environmental Sciences, pp. (29-51)

Sims, C.A. (1972). Money, Income and Causality. Am Econ Rev 62:540–552 Sims CA (1980) Macroeconomics and reality. Econometrica 48(1):1–48

Sims, C.A. (1972). Money, income and causality. Am Econ Rev 62:540–552 Sims CA (1980) Macroeconomics and reality. Econometrica 48(1):1–48.

Tularam, G.A.& Illahee, M.(2012). Time Series Analysis of Rainfall and Temperature Interactions in Coastal Catchments, Journal of Mathematics and Statistics, 6 (3): 372-380, 2010, ISSN 1549- 3644.

Wang, W.& Niu,Z. (2009).VAR model of PM2.5 , weather and traffic in Los Angeles-long beach area. International conference on environmental science and information application technology, 4-5 July, 2009, Wuhan, 3:66-69, ISBN: 978-0-7695-3682-8.DOI: 10.1109/ESIAT.2009.226