International Journal of Latest Engineering and Management Research (IJLEMR)
ISSN: 2455-4847
www.ijlemr.com || Volume 03 - Issue 11 || November 2018 || PP. 48-56

Bivariate vector autoregressive Approach for modelling Rainfall and Temperature of Katsina-Nigeria

Ibrahim Lawal Kane[1] & Umar Saminu[2]
[1]*Department of Mathematics and Computer Science,*
Umaru Musa Yar'adua University, 2218, Katsina – Nigeria
[2]*Department of Geography,*
Umaru Musa Yar'adua University, 2218, Katsina – Nigeria

Abstract: Climate change has been a worldwide issue and one of the most serious environmental threats of the 21^{st} centuries. Climate change caused severe weather which is devastating and can lead to natural disaster. Owing to the magnitude of current climate change, there is increasing researches on climate parameters such temperature, rainfall, wind speed and so on around the globe. However, most of the researches on climatic variables focus on modelling only one variable. A better understanding of the dynamics of the climatic variables can be attained by studying the variables together. Bivariate vector autoregressive (VAR) modelling technique was utilizedin this work to study the dynamic flows of rainfall and temperature of Katsina-Nigeriafor the period 1949–2014. The repercussions of this analysis are discussed. Finally, VAR forecast error variance decomposition reveals that 99.9% of forecast error variance of rainfall is accounted by the past rainfall and only 0.089% by temperature, 99.4% of the forecast error variance of temperature is accounted by its past values and 0.06% by rainfall. These indicate that neither temperature nor rainfall have a positive impact in predicting one another of Katsina which belong to steppe climates.

Keywords: Vector autoregressive, Granger causality, CUSUM, Impulse response analysis, Forecast error variance decomposition (FEVD)

Introduction

The climate change has been a global issue and one of the most imperioussubjects in water resources. Climate change caused severe weather which is devastating and can lead to natural disaster. A disaster typically can cause a serious disruption to human and environments. It also affects the ways individuals cope with natural resources. Since, 1950s, global warming has been unambiguous and many researchers have observed the fact that the changes will be unprecedented over decades (Norrulashikin et al. (2018). Due to the magnitude of nowadays climate change more work is now being done on climate parameters such temperature, rainfall, wind speed and so on around the globe. Even though most of researches on climatic variables focus on modelling only one variable, example, on rainfall (Ibrahim & Fadhilah 2013, Lovallo*et al.* 2013), temperature (Zheng et al. 2007; Gil-Alana LA. 2003); humidity (Sarraf et al. 2011, Jamaludin et al. 2015) and wind speed (Drobinski et al. 2015; Jamaludinet al. 2016)however, there is need alsoto consider all of these variables as a vector time series (Li and Genton, 2009). A better understanding can be obtained by studying several related variables together rather than just studying one variable. A Multivariate time-series approach is an essential statistical tool to study the comportment of time dependent variables based on the history of the data variation.Until now, only few researches are available on joint modelling of the climate variables,for example, Shahin (2014) developed an appropriate vector autoregression (VAR) model for forecasting monthly temperature, humidity, and cloud coverage of Rajshahi district in Bangladesh. Norrulashikin et al., (2015) observed the joint behaviour of temperature, rainfall and wind speed of Kuala Krai, Malaysia which belong to tropical rain forest zone. In a related study, Norrulashikin et al., (2018) modelled the dynamics of rainfall, humidity, wind speed and temperature of Alor Star station situated in the north-western of Peninsular Malaysia using Classical VAR-DCC Approach.Weather parameters such as Precipitation and Temperature, have irrefutable effects on hydrological cycle, agriculture and the environments. These variables not only have relationships with each other, but also are dependent. Modelling these meteorological variables could be practically useful in water resource management, risk management and making decisions on climate change. Researches concerning rainfall and temperature variations and trend in Semi-arid climates / steppe climates catchments received less attention. To investigate these stochasticmeteorological bi-variables and on different gauges should be of ultimate importance in the field climate change. However, developing an analytical time series models to explain the relationships among correlated meteorological variables has been difficult because of the complexity of the system, and an inadequate understanding of the physical mechanisms responsible for the interaction (Martınez-Gomez et al., 2009).

This paper tends to develop a time series model to model the rainfall and temperature of Katsina which belong to steppe climates and experiences a periodic rainfall between May and September with peak in August,for adequate understanding of the physical mechanisms responsible for their interaction using Vector Autoregressive (VAR) framework.Climate change is affecting Katsina particularly increases in rainfall variability and air temperature, which have resulted in more frequent hydrologic extremes, such as intense storms, flooding, and drought events.

Data and Methods

This paper examined 30 years data of rainfall and temperature of Katsina metropolis which belong to steppe climates under Köppen Climate Classification system for the period 1946 to 2014.The climate here is considered to be a local steppe climate. During the year, there is little rainfall in Katsina. The temperature here averages 26.0 °C and rainfall averages 600 mm.Rainfall is the key climatic variable, and there is a marked alternation of wet and dry seasons in most areas. Two air masses control rainfall--moist northward-moving maritime air coming from the Atlantic Ocean and dry continental air coming south from the African landmass.

Methodological Framework

A brief description of the methodology applied in this paper is presented in this section. The theoretical model, which serves as the basic framework of this analysis is the bivariate Vector Autoregressive (bVAR) modelling. The flowchart for the methodology adopted in this paper is given in figure 1.

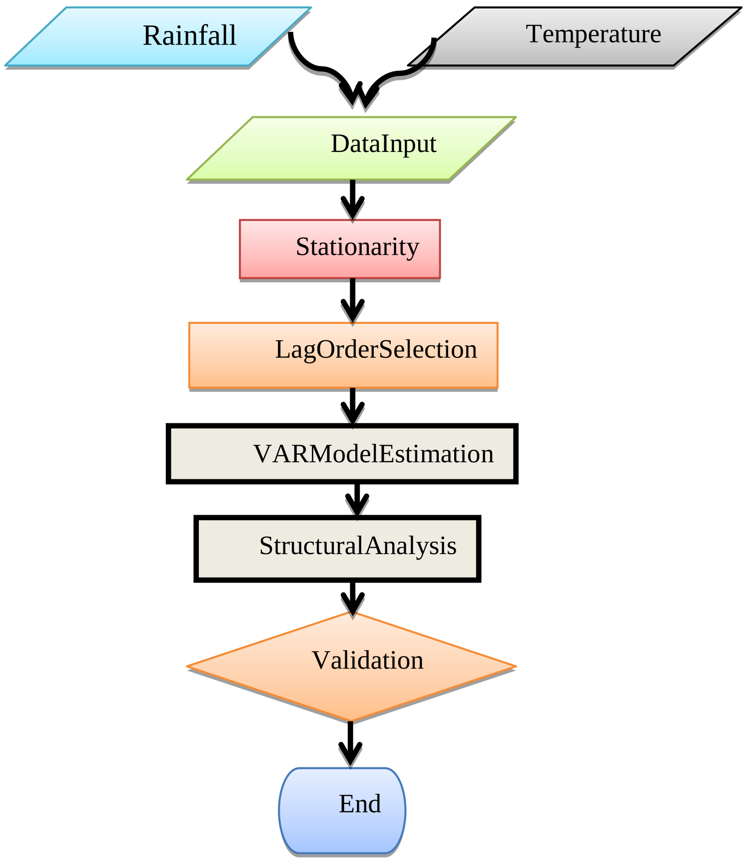


Figure 1: Flowchart of the methodology

Lag Length Selection

The lag length for the VAR modelling may be determined using model selection criteria. The general approach is to fit VAR(p) models with orders p= 0, ..., $p$max, and choose the value of $p$which minimizes some model selection criteria. Four most commonly used selection criteria adopted in this work are the Akaike (AIC), Schwarz-Bayesian (BIC), Hannan-Quinn (HQ) and Final Prediction Error (FPE) given as:

I. Akaike information criterion by Akaike (1974)

$$AIC(p) = \ln\left|\tilde{\Sigma}(p)\right| + \frac{2}{\hat{T}} \; p\mathrm{M}^2$$

II. Bayesian criterion by Gideon Schwarz (1978)

$$SBC(p) = \ln\left|\tilde{\Sigma}(p)\right| + \frac{\ln\left(\hat{T}\right)}{\hat{T}} \; p\mathrm{M}^2$$

III. Hannan-Quinn by Edward J. H and Barry G. Quinn (1979)

$$HQ(p) = \ln\left|\tilde{\Sigma}(p)\right| + \frac{2\ln\left(\ln\left(\hat{T}\right)\right)}{\hat{T}} \; p\mathrm{M}^2$$

IV. Final Prediction Error (FPE) by Akaike

$$\mathrm{FPE}(p) = \left|\tilde{\Sigma}(p)\right| + \left(\frac{\hat{T} + Mp + 1}{\hat{T} - Mp - 1}\right)^M$$

for moderate and large $\hat{T}$, FPE and AIC are essentially equivalent (Lutkepohl, 2006).

VAR Model Estimation

The theoretical model, which serves as the basic framework of our analysis, is the Vector Autoregressive model of order *p* (VAR(*p*)). VAR models were initially built base on the economic variables that are assumed to be stationary. Recently, the model starts to appreciate meteorological applications.

For a set of k variables let $Y_t = (y_{1t}, ..., y_{kt})'$ denote an (n×1) vector of time series variables. The basic VAR (*p*) model to captures their dynamic inter relationships which is given by:

$Y_t = c + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \varepsilon_t$, for (t = 1, …,N),

which can be rewritten as:

$\phi(L) Y_t = c + \varepsilon_t$

where c is the constant term, $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t} ..., \varepsilon_{kt})'$ is an unobserved zero means independent white noise process with time invariant positive definite covariance matrix $E(\varepsilon_t \varepsilon_t') = \sum_k$ and $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$ is a matrix of a lag polynomial with k x k coefficient matrices ϕ_j , for j = 1,…, p.

Model Diagnostic Checking

To test the null hypothesis that the residuals of the VAR model are uncorrelated, the Ljung-Box portmanteau test for autocorrelation can be applied. This test statistic in its multivariate form is defined as the following according to Hosking (1980):

$$LB(s) = T(T+2)\sum_{j=1}^{s} \frac{1}{T-j} tr\left\{C_{0j} C_{00}^{-1} C_{0j}' C_{00}^{-1}\right\} \sim \chi^2_{k(s-L)}$$

where

$$C_{0j} = T^{-1} \sum_{t=j+1}^{T} \varepsilon_t \varepsilon_{t-j}'$$

Under the null hypothesis (of independence) the Ljung-Box test statistic has approximately a χ^2 distribution with $k^2(s\text{-}L)$ degrees of freedom. Here $T$is the length of the series, and s denotes the order of autocorrelation. It is important to note that the test can be implemented only when the order of autocorrelation is higher than the lag length in the VAR model, i.e. $s > L$.

The CUSUM Test for Stability of Model Parameter

Parameters are constant through the sample period, is a key assumption in econometric models. In this study the Cumulative Sum of recursive residuals test (CUSUM) is used to investigate the stability of parameters.

The $$CUSUM_T = \sum_{t=M+1}^{T} \frac{\hat{u}_t^{(T)}}{\hat{\sigma}_u}, \ \mathrm{t} = M+1,\ldots,T$$

It is plotted for t = M+1,…,T together with the two lines $\pm C_r\left[\sqrt{T-M}+2\left(T-M\right)/\sqrt{T-M}\right]$ where C_r depends on the desired significance level of the resulting test. If the CUSUMs walk outside these lines; then it is evidence against structural stability of the underlying model, Brown et al., 1975).

Structural Analysis

When an adequacy of the model for the systems of equations of the specified variables has been realized, there are three interdependent approaches to the interpretation of VAR models (Lutkepohl, 1991);

1. Granger causality
2. Impulse response analysis
3. Forecast error variance decomposition (FEVD)

Data Description

This work examined 30 years data of rainfall and temperature from Katsina which belong to steppe climates under Köppen Climate Classification system from 1946 to 2014.This paper has focused on the relationship between temperature and rainfall of Katsina experiences rainfall between May and September with peak in August and is periodic in 12 month. The time series plots of all the variables considered are plotted against time. These plots are shown in figure.2.The three components of rainfall and temperature time series are visualized. The top panel shows the main data followed by a linear trend and the seasonal cycle that are obtained from the raw time series. The bottom panel envisions the residuals, which are obtained by subtracting the seasonal cycle and the linear trend from the raw data. Temperature series is trending upward since 1960 while rainfall was trending downward up to the year 2000 then start to trend upward and a seasonal influence were realised in both. The residuals display a sign of variation which shows that both rainfall and temperature fluctuates over time.

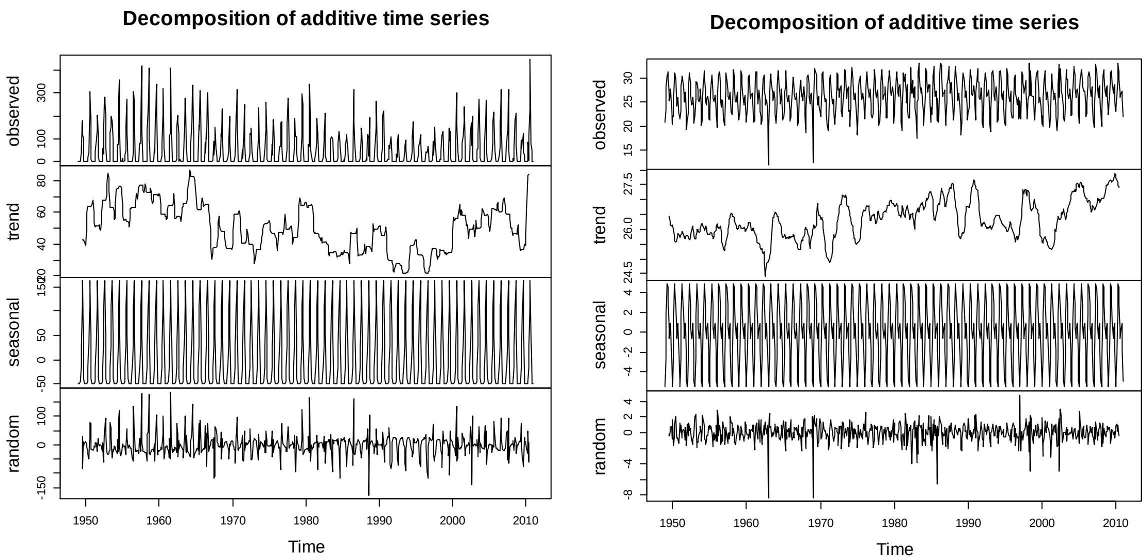


Figure 2: Components of the rainfall and temperature time series Katsina

Results and discussions

Numerous unit root test have been developed with a view to determining stationarity in time series analysis, in this work we consider Augmented Dickey Fuller(ADF) test.The ADF statistic tests the null hypothesis of presence of unit root against the alternative of stationary and the decision is to reject the null hypothesis when the value of test statistic is less than the critical value (Venus *et al.,* 2005).

Table 1: The ADF unit root test results

Test	ADF								
	None			Drift			Trend		
Critical Level	10%	5%	1%	10%	5%	1%	10%	5%	1%
	-1.62	-1.95	-2.58	-2.57	-2.86	-3.43	-3.12	-3.41	-3.96
Parameter	Statistic			Statistic			Statistic		
Rainfall	-13.2768			-17.2228			-17.2676		
Temperature	-19.2320			-20.0948			-20.3082		

Table 1 gives the results for the ADF tests for the rainfall and temperature series, for all series, the tests reject the null hypothesis of presence of unit root because the values of the tests statistic are less than the critical values, so there is evidence that the rainfall and temperature series does not behave as unit root.

VAR Model Identification

We estimate VAR model of rainfall and temperature series with number of lags order based on information criteria that is values of AIC, HQC, FPE and BIC given by the result in Table 2. The optimal routine is to select a lag that exhibits the smallest values of information criterion.

Table 2: Lag order selection

Lag	1	2	3	4	5	6
AIC(n)	8.877950	**8.874558**	8.876135	8.881247	8.879837	8.889631
HQ(n)	**8.887199**	8.893056	8.903883	8.918243	8.926083	8.945126
SC(n)	**8.901989**	8.922636	8.948253	8.977403	9.000033	9.033866
FPE(n)	7172.075	**7147.7861**	7159.073	7195.766	7185.640	7256.379

From Table 2, AIC and FPE suggested an optimal lag length of 2, p= 2 for the time series while HQC and SC suggested lag length of 1, p=1. In this case, we decided to model the VAR process using both lag order of p=1 and p=2 and identify which model would give the best performance by comparing the mean square error and the mean absolute error of the models.

VAR Models Estimation

After identifying the lag order for the VAR models, theestimated parameters of VAR (1) and VAR (2) are summarized in the following equations;

VAR (1) model :

$$\begin{bmatrix} R_{(t)} \\ T_{(t)} \end{bmatrix} = \begin{bmatrix} -0.0309 \\ 0.0134 \end{bmatrix} t + \begin{bmatrix} 0.0604 & -0.1612 \\ -0.0009 & 0.1467 \end{bmatrix} \begin{bmatrix} R_{(t-1)} \\ T_{(t-1)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{(1t)} \\ \varepsilon_{(2t)} \end{bmatrix}$$

VAR (2) model :

$$\begin{bmatrix} R_{(t)} \\ T_{(t)} \end{bmatrix} = \begin{bmatrix} 0.063 \\ 0.034 \end{bmatrix} t + \begin{bmatrix} 0.338 & -0.051 \\ -0.079 & 0.181 \end{bmatrix} \begin{bmatrix} R_{(t-1)} \\ T_{(t-1)} \end{bmatrix} + \begin{bmatrix} 0.003 & 0.156 \\ 0.296 & 0.386 \end{bmatrix} \begin{bmatrix} R_{(t-2)} \\ T_{(t-2)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{(1t)} \\ \varepsilon_{(2t)} \end{bmatrix}$$

VAR Model Checking

When a model for a time series data has been constructed, it is common practice to execute checks for the model adequacy. The Portmanteau (LB) test and Breusch-Godfrey (BG) testfor residual autocorrelation are prominent tools for this task. To judge whether ARCH effects and autocorrelation have been completely removed or not, a well-known example is the ARCH-LM test. These tests were applied to check for the adequacy of the developed VAR(1) and VAR(2) models;and the results are summarized in table 3.

Table 3: P-values ofSerial Correlation and ARCH Effect tests

Models	LB Test	BG	ARCH LM Test
VAR(1)	0.030.050.01		
VAR(2)	0.12	0.32	0.04

The existence of serial correlations in the models residuals is rejected based on the results of the LB test for VAR(2) model with p-value above 0.05 and accepted for VAR(1) model (Wang *et al.*, 2005). Similarly, BG test results indicate the acceptance of the null hypothesis of model adequacy at 95% confidence level for both models hence the models are specified correctly. ARCH LM tests with *p*-values lessthan 0.05 indicate presence of heteroskedasticity. Hence the linear aspect of the data sets are well captured.

CUSUM Test for Stability

Figures 3 give the graphical representation of CUSUM test for stability of VAR (1) and VAR (2) models parameters respectively. The CUSUM plots of the variables do not cross critical bounds which show no evidence of any significant instability.

(a) (b)

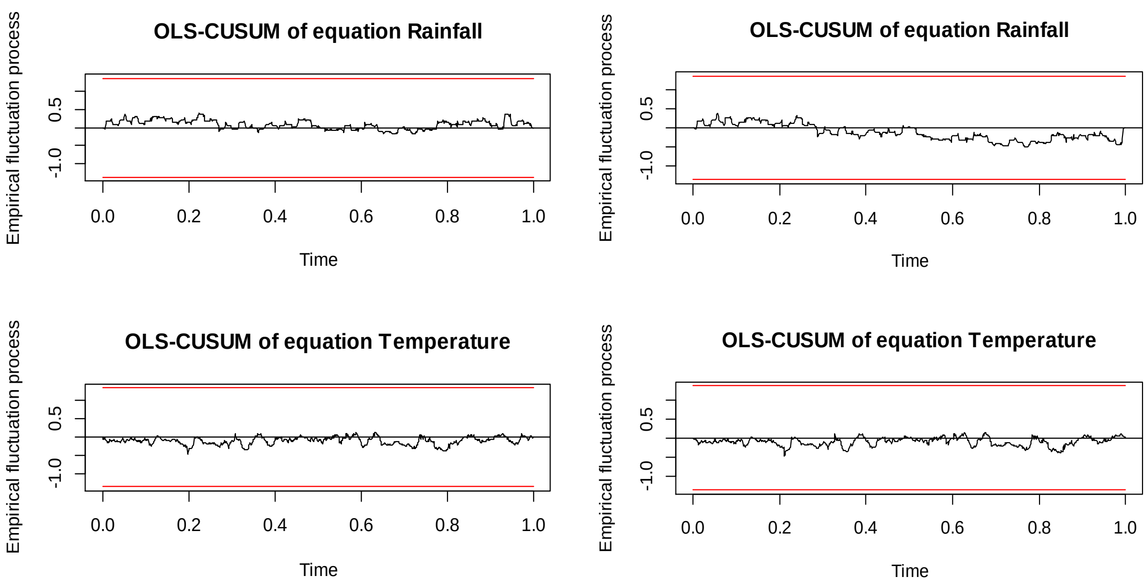


Figure 3: The Ordinary Least Square Cumulative Sum (OLS-CUSUM) test

Structural Analysis

Granger causality Test

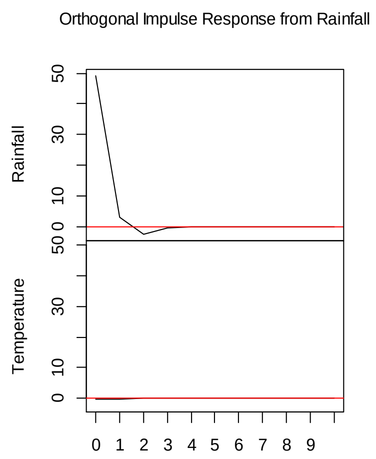
Table 4 presented the results of Granger Causality for the developed VAR(1) and VAR(2) models. The estimatedresults show that rainfall does not Granger-cause temperature and temperature does not Granger-cause rainfall. Therefore, neither temperature nor rainfall could have a positive impact in predicting one another.

Table 4: Result of Granger Causality Test

Model	Null Hypothesis	F-Statistic	P-Value	Decision
VAR(1)	Temperature do not Granger-cause Rainfall	0.0251	0.8742	Do not reject null
	Rainfall do not Granger-cause Temperature	0.6231	0.4307	Do not reject null
VAR(2)	Temperature do not Granger-cause Rainfall	0.3594	0.6982	Do not reject null
	Rainfall do not Granger-cause Temperature	0.3115	0.7324	Do not reject null

Impulse Response Analysis

The impulse response analysis is used to trace out responses of current and future values of each variable to one unit increase in the current value of one of the VAR errors. It might be interested to realisethat sudden and unexpected change in temperature have no effect on rainfall and vice versa over a period of time. The curved lines trace out the impulse response functions, it could be observed that rainfall showed positive impact on temperature at the initial periods, but later became insignificant and slowly became zero.



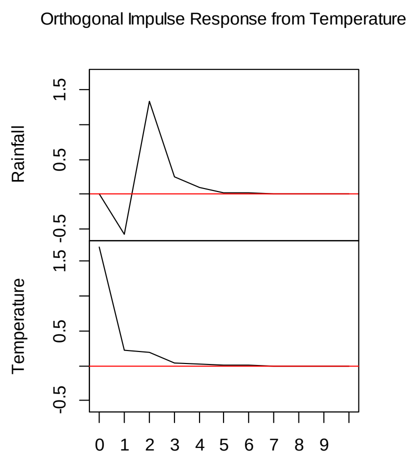


Figure 4: VARForecast Error Impulse Response

Forecast error variance decomposition (FEVD)

They are fractions of the forecast error variances of the variables attributed to their own innovations or the innovations of the other variables.

Table 5a: Proportions of Forecast Error in "Rainfall" Accounted for by:

Forecast horizon	rainfall	Temperature
[1,]	1.0000000	0.0000000000
[2,]	0.9998641	0.0001359146
[3,]	0.9991371	0.0008629420
[4,]	0.9991132	0.0008867661
[5,]	0.9991095	0.0008905331
[6,]	0.9991092	0.0008907844
[7,]	0.9991091	0.0008908755
[8,]	0.9991091	0.0008908858
[9,]	0.9991091	0.0008908875
[10,]	0.9991091	0.0008908877
[11,]	0.9991091	0.0008908877
[12,]	0.9991091	0.0008908877

Table 5b: Proportions of Forecast Error in "Temperature" Accounted for by:

Forecast horizon	rainfall	Temperature
[1,]	0.004352503	0.9956475
[2,]	0.005621432	0.9943786
[3,]	0.005733814	0.9942662
[4,]	0.005745558	0.9942544
[5,]	0.005746940	0.9942531
[6,]	0.005747234	0.9942528
[7,]	0.005747275	0.9942527
[8,]	0.005747280	0.9942527
[9,]	0.005747281	0.9942527
[10,]	0.005747281	0.9942527
[11,]	0.005747281	0.9942527
[12,]	0.005747281	0.9942527

From the result of tables 5a, the VAR forecast error variance decomposition shows that at the period of 12, 99.9% of the forecast error variance of rainfall is accounted by the past rainfall and 0.089% by temperature. For temperature, the result from table 5b reveals that at the period of 12, 99.4% of the forecast error variance of temperature is accounted by its past values and 0.06% by rainfall.

Models Comparison

The performance of VAR(1) and VAR(2) models are compared. The result is presented in Table 6.The VAR(2) model seems to give better result based on low values ofMAE and RMSE which are presented in order to detect the best time series model for the given data sets.

Table 6: Models comparison

Model	RMSE	MAE
VAR(1)	0.7667	0.6523
VAR(2)	0.3401	0.2782

Summary and Conclusion

The study of behaviour of the climatic variables becameessential for understanding the future changes among the climatic variables and implementing important policies. This paper presents the dynamic flow of rainfall and temperature of Katsina.Test for stationarity of the time series variables has been established with augmented Dickey–Fuller. AIC and FPE suggested VAR (2) model for the time series while HQC and SC suggested VAR (1) model. Both models were developed, and by comparing their performanceusing RMSE and MAE, VAR (2) became the appropriate model for the climate-variables considered.

Structural analyses were performed using impulse response function and forecast error variance decomposition. Unlike in the tropical rainforest where climate variables not only have relationships with each other, but also depend on each other, It might be interested to realise that sudden and unexpected change in temperature in steppe climate have no effect on rainfall and vice versa over a period of time and that neither rainfall nor temperature has a significantimpact on forecasting one another.These might be as a result of high seasonal variation which may be considered at future work.

Acknowledgement

The authors would like to thank Umaru Musa Yar'adua University, Katsina and TET Fund grant vote number UMYU/UBR/BC/002 for funding the research.

References

[1]. Akaike, H. (1974). A look at the statistical model identification, IEEE transactions on automatic control, AC-19, 716-723.

[2]. AmirpouyaSarraf, A., FarnoodVahdat S, Behbahaninia A. (2011). Relative Humidity and Mean Monthly Temperature Forecasts in Ahwaz Station with ARIMA Model in time Series Analysis, *International Conference on Environment and Industrial Innovation, Singapore IPCBEE vol.12, pp 149-153.*

[3]. Brown, R.L., Durbin, J. & Evans, J.M. 1975.Techniques for testing the constancy of regression relationships over time. Journal of the Royal Statistical Society 37(2): 149-192.

[4]. Drobinski, P., Coulais, C. &Jourdier, B. (2015) Boundary-Layer Meteorol 157: 97. https://doi.org/10.1007/s10546-015-0035-7

[5]. Edward, J.H and Quinn, B.G. (1979).The determination of the order of an autoregression. J. Royal Statis. Soc. Series B 41: 190-195.

[6]. Gideon, S. (1978). Estimating the dimension of a model.The Annals Statis.6: 461- 464.

[7]. Gil-Alana LA. 2003. An application of fractional integration to a long temperature time series. International Journal of Climatology23: 1699-1710.

[8]. Hosking, JRM. (1980). The Multivariate Portmanteau Statistic. Journal of the American Statistical Association 75(371):602-608.DOI: 10.1080/01621459.1980.10477520

[9]. Ibrahim, L. K., & Fadhilah, Y. (2013).Assessment of Risk of Rainfall Events with a Hybrid of ARFIMA-GARCH.*Modern Applied Science*, *7*(2), 78-89. http://dx.doi.org/10.5539/mas.v7n12p78

[10]. Jamaludin, AaishahRadziahBt, Fadhilah Yusof & Ibrahim Lawal Kane (2015). Temporal Dynamics of Trend in Relative Humidity with RH-SARIMA Model.Modern Applied Science; Vol. 9, No. 3.

[11]. Li, Y., &Genton, M. (2009).Single-Index Additive Vector Autoregressive Time Series Models. Scandinavian Journal of Statistics, 36, 369–388. http://dx.doi.org/10.1111/j.1467-9469.2009.00641.x.5314-5326.

[12]. Lovallo, M., Shaban, A., Darwich, T. and Telesca, L. (2013). Investigating the time dynamics of monthly rainfall time series observed in northern Lebanon by means of the detrended fluctuation analysis and the Fisher-Shannon method. ActaGeophysica, DOI 10.2478/s11600-012-0094-9.

[13]. Lutkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*, Berlin: Springer

[14]. Martınez-Gomez, E, Guerrero, V. M, Estrada F. (2009) A Multivariate Time Series Analysis for Climate Modeling based on the Solar-Terrestrial Relationship. Physical and Engineering Sciences – JSM 2009.p2035-2047.

[15]. Norrulashikin, S.M, Yusof, F & Kane, I.L. (2015).Modern Applied Science; Vol. 9, No. 11; pp89-100. ISSN 1913-1844 E-ISSN 1913-1852.

[16]. Norrulashikin, S.M, Yusof, F & Kane, I.L. (2018).Meteorological Multivariable Approximation and Prediction with Classical VAR-DCC Approach.SainsMalaysiana 47(2): 409-417. http://dx.doi.org/10.17576/jsm-2018-4702-24

[17]. Shahin M.A., Ali M.A., Ali A.B.M.S. (2014) Vector Autoregression (VAR) Modeling and Forecasting of Temperature, Humidity, and Cloud Coverage. In: Islam T., Srivastava P., Gupta M., Zhu X., Mukherjee S. (eds) Computational Intelligence Techniques in Earth and Environmental Sciences. Springer, Dordrecht.

[18]. Venus, K..L., Sie-Hoe, L. and Siew-Eng, L. (2005). A Complementary Test for ADF Test with An Application to the Exchange Rates Returns. Munich Personal RePEc Archive, 6.

[19]. Verlag.

[20]. Wang, W., van Gelder, P. H. A. J. M., and Vrijling, J. K. (2005) Long memory properties of streamflow processes of the Yellow River. Proceedings of the International Conference on Water Economics, Statistics and Finance, 1, Rethymno-Crete, Greece, 481–490, 8–10 July, 2005.

[21]. Yusof, F. & Kane, I.L. 2013.Volatility modeling of rainfall time series. Theoretical and Applied Climatology 113(1-2): 247-258. doi:10.1007/s00704-012-0778-8.

[22]. Zheng Z. F., Zhang Xiu-Li, Z., Hong-Xing, C., Zhuang, X. and Jia-Hua, P. (2007).Characteristics of Long-Term Climate Change in Beijing with Detrended Fluctuation Analysis.Chinese Journal of Geophysics. 50(2), 393-398.