

An Investigation towards the Suitability of Vector Autoregressive Approach on Modeling Meteorological Data

Siti Mariam Norrulashikin¹, Fadhilah Yusof¹ & Ibrahim Lawal Kane²

¹ Department of Mathematical Science, Faculty of Science, Universiti Teknologi Malaysia, Johor Bahru, Johor, Malaysia

² Department of Mathematics and Computer Science, Umaru Musa Yar'adua University, Katsina State, Nigeria

Correspondence: Siti Mariam Norrulashikin, Department of Mathematical Science, Faculty of Science, Universiti Teknologi Malaysia, 81300 Johor Bahru, Johor, Malaysia. Tel: 60-017-796-7464. E-mail: fadhilahy@utm.my

Received: May 6, 2015

Accepted: June 19, 2015

Online Published: September 30, 2015

doi:10.5539/mas.v9n11p89

URL: <http://dx.doi.org/10.5539/mas.v9n11p89>

Abstract

In most meteorological problems, two or more variables evolve over time. These variables not only have relationships with each other, but also depend on each other. Although in many situations the interest was on modelling single variable as a vector time series without considering the impact other variables have on it. The vector autoregression (VAR) approach to multiple time series analysis are potentially useful in many types of situations which involve the building of models for discrete multivariate time series. This approach has 4 important stages of the process that are data pre-processing, model identification, parameter estimation, and model adequacy checking. In this research, VAR modeling strategy was applied in modeling three variables of meteorological variables, which include temperature, wind speed and rainfall data. All data are monthly data, taken from the Kuala Krai station from January 1985 to December 2009. Two models were suggested by information criterion procedures, however VAR (3) model is the most suitable model for the data sets based on the model adequacy checking and accuracy testing.

Keywords: meteorology, vector autoregressive, multivariate time series

1. Introduction

The climate change has been a global issue and always one of the most imperative topics in water resources. Weather parameters such as Precipitation, Temperature, Wind speed and Relative Humidity modelling and forecasting could be practically useful in risk management, water resource management and making decisions on climate change. These variables have undeniable effects on the hydrological cycle, agriculture and the environments. Modelling these physical processes deterministically may become a very challenging task due to the complexity of natural systems. As an alternative, stochastic models are used. The application of stochastic models to these climatological variables has been undertaken for long along the globe. However, most of the literatures dealt with single variable; for example; studies on precipitation can be seen in (Gil-Alana, 2012; Valdez-Cepeda *et al.*, 2012; Ibrahim & Fadhilah, 2013; etc.). Studies on temperature can be seen in (Smith, 1993; Fraedrich & Blender, 2003; etc.). Studies on relative humidity can be seen in (Shiri *et al.*, 2011; Jäntschi, 2011; Jamiyansharav, 2011; etc.).

In various time series problems, two or more random variables evolve over time. These variables not only have relationships with each other, but also are dependent. Generally, if variables are empirically dependent, then multivariate models should be considered. Although in many situations we are only interested in modelling and predicting only one variable, however, there is need to also consider all of these variables as a vector time series (Li and Genton, 2009). For example, the fluctuating nature of precipitation as a result of anthropogenic climate change has been the potential subject of significant in recent modelling framework (e.g. Wong *et al.* 2009; Wan Zawiah, 2012 and many others), but, many uncertainties remain, it is generally agreed that as temperatures increase, the intensity of heavy precipitation events also will increase (Meehl *et al.*, 2007).

Vector autoregression (VAR) model, introduced by Sims (1980) is a technique that could be used to capture the linear interdependencies among multiple time series as well as to characterize the joint dynamic behavior of a

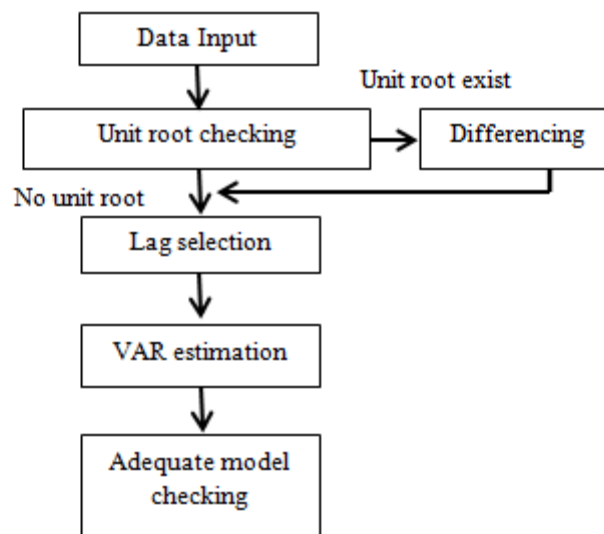
collection of variables without requiring strong restrictions of the kind needed to identify underlying structural parameters. It has become a prevalent method of time-series modeling. VAR models generalize the univariate autoregression (AR) models by allowing for more than one evolving variable. All variables in a VAR are treated symmetrically in a structural sense (although the estimated quantitative response coefficients will not in general be the same); each variable has an equation explaining its evolution based on its own lags and the lags of the other model variables. VAR modeling does not require as much knowledge about the forces influencing a variable, as do structural models with simultaneous equations: The only prior knowledge required is a list of variables which can be hypothesized to affect each other intertemporally.

The main advantage of the VAR is that there is no need to specify which variables are the endogenous variables and which are the explanatory variables because in the VAR, all selected variables are treated as endogenous variables. That is, each variable depends on the lagged values of all selected variables and helps in capturing the complex dynamic properties of the data (Brooks, 2002).

However, Engle and Granger (1987) suggest that if a time-series system under study includes integrated variables of order 1 and satisfied the conditions of cointegration relations, then this system will be more appropriately specified as a vector error-correction model (VECM) rather than a VAR. These types of modeling strategy receives less application to meteorological data sets more specifically in the Malaysian meteorological phenomena. Therefore, the aim of the present work is to investigate the suitability of using the vector autoregressive method in modeling meteorological data sets of Kuala Krai in the northeast of Malaysia.

2. Methodology

A description of the data to be used and a brief overview of the methodology to be implemented in this research work are presented. The theoretical model, which serves as the basic framework of our analysis, is the Vector Autoregressive model of order p (VAR (p)). The compact mode of our methodology is designed in the framework below.



2.1 Study Area and Data Collection

The data used were taken from the Kuala Krai station in the center of the State of Kelantan at northeast of Peninsular Malaysia. The Kuala Krai station is located at latitude $5^{\circ} 32' N$ and longitude $102^{\circ} 12' E$. The land is hilly and it is once an area of tropical rain forest. Kuala Krai is an area in which the meeting point of two main rivers and formed Sungai Kelantan flowing to the estuary in South China Sea near the State capital of Kota Bharu (Ababa, 2012). Kuala Krai is influenced by an extreme monsoon whereby the average annual temperature is $26.8^{\circ} C$ while the annual rainfall is at the average of 2713mm. The daily meteorological data were collected from Jabatan Meteorologi Malaysia, which contained 24 hours mean temperature ($^{\circ}C$), maximum wind speed (m/s) and rainfall (mm) data from year 1985 to year 2009.

2.2 Parameter Estimation

2.2.1 Lag Selection

The information criteria of Akaike, Schwartz and Hannan-Quinn models would determine the lag length for VAR

order, p (Miształ, 2010). The following criteria are;

- i. Akaike's Information Criteria,

$$AIC = \ln \frac{1}{T} \sum_{t=1}^T (\hat{u}_t^{(p)})^2 + m \frac{2}{T}$$

- ii. Schwartz Criteria

$$SC = \ln \frac{1}{T} \sum_{t=1}^T (\hat{u}_t^{(p)})^2 + m \frac{\ln T}{T}$$

- iii. Hannan-Quinn Criteria

$$HQC = \ln \frac{1}{T} \sum_{t=1}^T (\hat{u}_t^{(p)})^2 + m \frac{2 \ln(\ln T)}{T}$$

$\hat{u}_t^{(p)}$ the estimated residuals of the AR(p) process, while m is the number of estimated parameters.

2.2.2 Unit Root

In this study, we focused on three unit root test, Augmented Dickey-Fuller (ADF) test, Phillip Peron (PP) test and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. ADF and PP test testing on the null hypothesis of a unit root against the alternative of stationary, while KPSS test the null hypothesis of stationary against the alternative of a unit root.

- a. ADF test (Dickey & Fuller, 1979)

- i. Without constant and trend, $Y_t = \rho Y_{t-1} + e_t$, $t = 1, 2, \dots$
 ii. With constant, $Y_t = \mu + \rho Y_{t-1} + e_t$, $t = 1, 2, \dots$
 iii. With constant and trend, $Y_t = \mu + \beta t + \rho Y_{t-1} + e_t$, $t = 1, 2, \dots$

Where $Y_0 = 0$, ρ is a real number and $\{e_t\}$ is a sequence of independent normal random variables with mean zero and variance σ^2 [i.e., e_t NID (0, σ^2)].

The hypothesis is:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

Decision rule:

If $t^* >$ ADF critical value or p-value $>$ significance level 0.05, do not reject the null hypothesis, i.e., unit root exists.

If $t^* <$ ADF critical value or p-value $<$ significance level 0.05, reject the null hypothesis, i.e., unit root does not exist.

- b. PP test (Phillips & Perron, 1988)

Phillips and Perron (1988) have derived the following equation;

$$Y_t = \alpha + \rho Y_{t-1} + u_t, \quad t = 1, 2, \dots$$

where u_t is I(0) and may be heteroskedastic. The PP tests correct for any serial correlation and heteroskedasticity in the errors u_t of the test regression.

The hypothesis:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

Decision rule:

Do not reject the null hypothesis (unit root exists) if p-value $>$ significance level 0.05.

Reject null hypothesis (no unit root) if p-value $<$ significance level 0.05.

- c. KPSS test (Kwiatkowski & Phillips, 1992)

Kwiatkowski et al. (1992) have derived test for their hypothesis, given by the equation,

$$y_t = \xi t + r_t + \varepsilon_t, \quad r_t = r_{t-1} + u_t$$

where r_t is a random walk and u_t are iid (0, σ_u^2).

The hypothesis: H_0 : the series of data is stationary

H_1 : the series of data is not stationary

Decision rule:

If p - value $>$ significance level 0.05, then do not reject the null hypothesis.

If p - value $<$ significance level 0.05, then reject the null hypothesis.

2.2.3 Model Estimation

A VAR specification was used to model each variable as a function of all the lagged endogenous variables in the system. Johansen (1988) considered that the process y_t is defined by an unrestricted VAR system of order (p):

$$y_t = \delta + \Gamma_1 y_{t-1} + \Gamma_2 y_{t-2} + \dots + \Gamma_p y_{t-p} + u_t, \quad t = 1, 2, 3, \dots, T$$

where y_t is independent $I(1)$ variables, the Γ 's are estimable parameters and $u_t \sim \text{iid}(0, \Sigma)$ is vector of impulses which represent the unanticipated movements in y_t . However, such a model is only appropriate if each of the series in y_t is integrated to order zero, $I(0)$, meaning that each series is stationary. (Wong, Chan, & Chiang, 2007)

The hypothesis: H_0 : The data sets follows VAR model

H_1 : The data sets does not follows VAR model

3. Results and Discussion

3.1 Data Analysis

The data of this paper consist of the monthly observations of meteorological variables in northern Malaysia for the period of 1985-2009, which include mean temperature ($^{\circ}C$), wind speed (m/s), and rainfall (mm). The summary statistics were presented and the time series plot and correlogram were plotted, followed by a unit root test, parameter estimation, model checking, causality test and lastly impulse response function.

3.1.1 Summary Statistics

Table 3.1. Descriptive statistics of the variables

Variables	Mean	Standard Deviation	Skewness	Kurtosis	Min	Max
Temperature	26.22	0.9321862	-0.135978	3.014978	23.77	29.39
Wind Speed	7.702	1.16346	0.08087192	2.671254	5.190	11.057
Rainfall	6.87541	5.915796	2.477805	11.6369	0.02143	39.13000

Table 3.1 presents the descriptive statistics of the meteorological variables where all variables display a positive value of the mean. The standard deviation of temperature and wind speed were smaller as compared to rainfall, indicates that the variation of the data set were not far away from its mean, but the standard deviation for rainfall was a bit large, indicates the possibility of some outliers occurred in the rainfall data set. Wind speed and rainfall distribution showed a positively skewed while temperature exhibit a negative skewed, but it was still approaching zero. In terms of kurtosis, both temperature and wind speed values were approaching 3, indicates that they had a normal shape of the distribution, while rainfall data set was having a leptokurtic distribution.

3.1.2 Time Series Plot

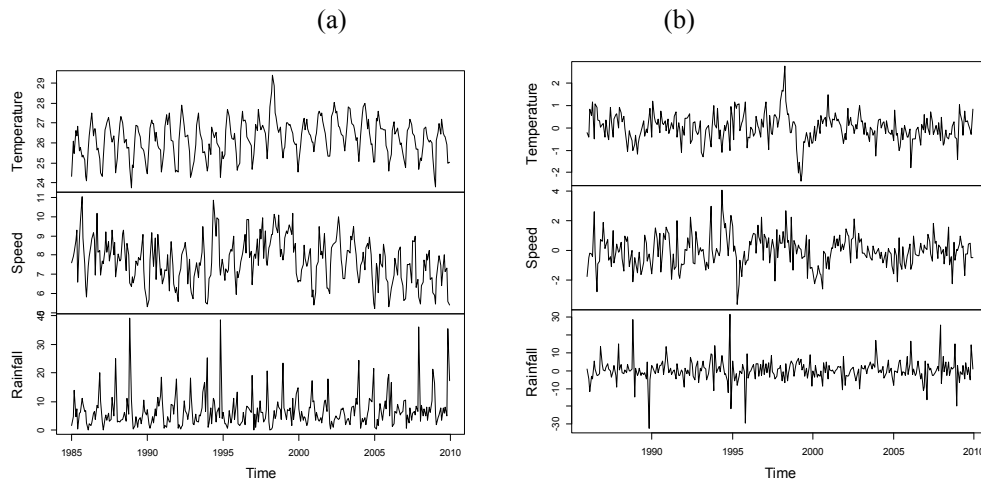


Figure 3.1. (a) time series plot for original data set (b) time series plot after seasonal differencing

Figure 3.1 displays the time series of all variables that were plotted against time. From the plot (Figure 3.1 (a)), we can see that the data follow seasonality pattern and all of the graphs fluctuated around its mean. Although the data seemed to be stationary, we might have to do seasonal differencing to remove the seasonal pattern. Figure 3.1 (b) shows the graph of all variables after seasonal differenced.

In order to confirm that the data of all three variables are having a seasonal pattern, the autocorrelation function was conducted. Figure 3.2 displayed the existence of seasonal pattern from the autocorrelation function test which repeated periodically every 12 months.

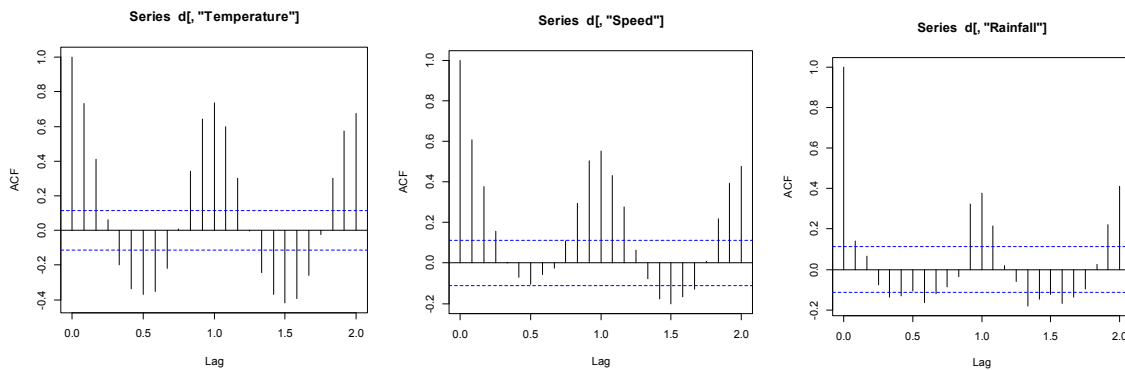


Figure 3.2. Autocorrelation function of variables

3.1.3 Testing for Unit Root

Three tests were used to test whether the time series data were in a stationary state, namely Augmented Dickey Fuller (ADF), Phillip Perron (PP) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests. For ADF and PP test, the null hypothesis is rejected (i.e., the data does not have unit roots) when the p-value is less than the significance level, 0.05. Meanwhile, for KPSS test, rejecting the null hypothesis, when p-value less than significance level 0.05, indicate that the data has unit root.

Table 3.2. Unit root testing for stationary

Variables	ADF test		KPSS test		PP test	
	Level	Seasonal Difference	Level	Seasonal Difference	Level	Seasonal Difference
Temperature	-8.936 (0.01)	-4.8349 (0.01)	0.5018 (0.04126)	0.1761 (0.1)	-91.0795 (0.01)	-190.9745 (0.01)
Wind Speed	-5.9299 (0.01)	-4.1996 (0.01)	0.5896 (0.02359)	0.0967 (0.1)	-119.1996 (0.01)	-237.2388 (0.01)
Rainfall	-9.2192 (0.01)	-5.7487 (0.01)	0.2433 (0.1)	0.0395 (0.1)	-241.3231 (0.01)	-387.7442 (0.01)

Table 3.2 presents the p-value for all tests that have been tested. At level, the p-value for ADF and PP test was 0.01, showing that the data for all three variables has no unit root. In Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests, the p-value for temperature and wind speed shows the existence of unit root while rainfall data is stationary, since the p-value was more than 0.05.

After seasonal difference, all p-values for ADF and PP test were less than significant level and the p-value for the KPSS test were more than the significance level, which indicate that the series were already stationary.

3.1.4 Lag Length Selection

Order of lag length is one of the most important aspects that should be included in VAR modeling because if we had chosen a different order of lag length, we would encounter with different result that could lead to misleading interpretation. After removing the seasonal pattern, the correlograms of the seasonally adjusted meteorology series (Figure 3.3) suggested that the first six orders of temperature variable, for instance might be auto-correlated.

Table 3.3. Lag selection for VAR

	1	2	3	4	5	6
AIC(n)	2.337299	2.295235	2.265127	2.315649	2.333903	2.376906
HQ(n)	2.399445	2.403991	2.420494	2.517626	2.582490	2.672102
SC(n)	2.492273	2.566441	2.652564	2.819317	2.953803	3.113036
FPE(n)	10.353290	9.927069	9.633212	10.133488	10.321917	10.778132

Four criterion procedure to measure the relative quality of a statistical model were estimated in order to identify the correct number of lag order, p . **Akaike's Information Criterion (AIC)**, Hannan–Quinn information criterion (**HQC**), Schwarz **Criterion (SC)**, and Final Prediction Error (FPE) were used to identify the number of lag order that would be used in the VAR modeling. According to the analysis below in Table 3.3, AIC and FPE suggested that an optimal lag length, $p=3$ is appropriate for the time series data while HQC and SC suggested that lag length, $p=1$ should be used for modeling the time series. In this case, we decided to model the VAR process using both lag order, $p=1$ and $p=3$ and yet to identify which model would give the best performance by comparing the mean square error for both models.

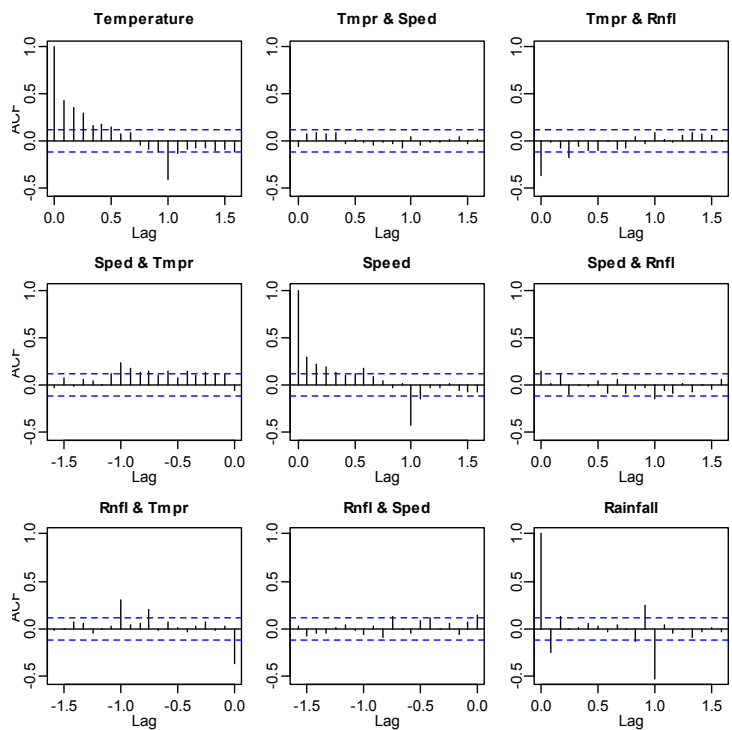


Figure 3.3. Correlograms

3.1.5 Estimation of Parameters

After identifying the lag order for VAR model, the estimation process of VAR modeling, including the constant and trend was performed.

The parameter estimation of VAR (1) and VAR (3) are summarized in equation 3.1 and 3.2. If we compare the coefficient and the p-value for both equations, there was not much difference between them. However, the p-value for the equation of all variables; temperature, wind speed and rainfall in both tables were significant in this study, where the null hypothesis is rejected, which indicate that the series of data might not be suitable for VAR modeling. Nevertheless, the model checking that is going to be done in the next process will determine further the suitability of VAR modeling.

VAR(1) model :

$$\begin{bmatrix} T_{(t)} \\ S_{(t)} \\ R_{(t)} \end{bmatrix} = \begin{bmatrix} 0.0694 \\ -0.0362 \\ -0.0829 \end{bmatrix} + \begin{bmatrix} -0.0004 \\ 0.0000 \\ 0.0018 \end{bmatrix} t + \begin{bmatrix} 0.4819 & 0.0412 & 0.0151 \\ 0.2226 & 0.2888 & 0.0032 \\ -0.7888 & 0.6902 & -0.2929 \end{bmatrix} \begin{bmatrix} T_{(t-1)} \\ S_{(t-1)} \\ R_{(t-1)} \end{bmatrix} + \begin{bmatrix} U_{(1t)} \\ U_{(2t)} \\ U_{(3t)} \end{bmatrix} \tag{3.1}$$

VAR(3) model :

$$\begin{bmatrix} T_{(t)} \\ S_{(t)} \\ R_{(t)} \end{bmatrix} = \begin{bmatrix} 0.0493 \\ -0.0378 \\ 0.0468 \end{bmatrix} + \begin{bmatrix} -0.0003 \\ 0.0000 \\ 0.0010 \end{bmatrix} t + \begin{bmatrix} 0.3910 & 0.0265 & 0.0136 \\ 0.1447 & 0.2372 & 0.0083 \\ -1.1330 & 0.6133 & -0.2854 \end{bmatrix} \begin{bmatrix} T_{(t-1)} \\ S_{(t-1)} \\ R_{(t-1)} \end{bmatrix} + \begin{bmatrix} 0.1523 & 0.0338 & -0.0027 \\ 0.1868 & 0.1102 & 0.0190 \\ -0.2211 & -0.6743 & 0.0762 \end{bmatrix} \begin{bmatrix} T_{(t-2)} \\ S_{(t-2)} \\ R_{(t-2)} \end{bmatrix} + \begin{bmatrix} 0.01283 & 0.0209 & -0.0170 \\ -0.0683 & 0.0957 & -0.0228 \\ 1.4226 & 0.2767 & 0.0805 \end{bmatrix} \begin{bmatrix} T_{(t-3)} \\ S_{(t-3)} \\ R_{(t-3)} \end{bmatrix} + \begin{bmatrix} U_{(1t)} \\ U_{(2t)} \\ U_{(3t)} \end{bmatrix} \tag{3.2}$$

3.1.6 Johansen Cointegration Rank Test

The Johansen cointegration rank test was applied to check whether or not cointegration exist among the variables and yet to determine the cointegration rank of the variables. VECM is applied in the model instead of VAR when there exists a cointegration relationship between the variables.

Table 3.6. Johansen test

Rank	Test statistics	Critical value		
		10%	5%	1%
0	312.09	39.06	42.44	48.45
1	190.41	22.76	25.32	30.45
2	73.20	10.49	12.25	16.26

The results for the test is presented in Table 3.6. Rejecting the null hypothesis when the test statistics are less than critical values and it indicates that there exists a cointegration relationship between the variables. From Table 3.6, all the test statistics are more than the critical values which indicate that from rank 0 to 2, there is no cointegration relationships exist.

3.1.7 Model Adequacy Checking

Table 3.7. Model checking

	Lag length order, $p=1$		Lag length order, $p=3$	
	Chi square test statistics	p-value	Chi square test statistics	p-value
Jarque-Bera test	468.3302	< 2.2e-16	527.1804	< 2.2e-16
Breusch-Godfrey LM test	74.3114	0.00387	47.2256	0.3817
ARCH LM test	221.8539	0.01833	193.3146	0.2357

Model checking was needed after fitting the model. Table 3.7 shows the result for VAR(1) and VAR(3); normality test (Jarque-Bera test), autocorrelation test (Breush-Godfrey LM test) and ARCH test (ARCH LM test). For VAR(1), the p-value for all tests were less than 0.05 indicate that this model was not normally distributed, correlated between variables and had heteroscedasticity effect. However, in VAR(3) model, the results shows that the model was not normally distributed, but the model had no auto-correlation between variables and it had no heteroscedasticity effect on the model. When the series of data is auto-correlated and had heteroscedasticity effect, then the estimated variances of the estimated coefficients will be biased and inconsistent, and therefore hypothesis testing is no longer valid.

3.1.8 Causality Test

Causality test is one of the advantages when dealing with multivariate time series model, or in other word, we could not perform causality test when analyzing the univariate time series model.

Table 3.8. Granger causality test

			Lag length order, $p=1$		Conclusion
Null hypothesis			F-test	P-value	
Temperature	do not	Granger-cause	3.6308	0.02691	Reject null hypothesis
Speed and Rainfall					
Speed	do not	Granger-cause	4.481	0.01159	Reject null hypothesis
and Rainfall					
Rainfall	do not	Granger-cause	3.875	0.02112	Reject null hypothesis

Temperature and Speed					
Lag length order, $p=3$					
Temperature do not Granger-cause Speed and Rainfall	2.0902	0.05214	Do not reject null hypothesis		
Speed do not Granger-cause Temperature and Rainfall	1.6869	0.1212	Do not reject null hypothesis		
Rainfall do not Granger-cause Temperature and Speed	4.5053	0.0001676	Reject null hypothesis		

The result for the Granger causality test are shown in Table 3.8. At 95% confidence level, temperature and wind speed Granger-cause on the other variables for VAR(1) but for VAR(3), temperature and wind speed did not Granger-cause the other variables. However, for rainfall variability, it showed that rainfall was Granger cause temperature and wind speed for both VAR (1) and VAR (3). Rainfall was said to Granger-cause temperature and wind speed, meaning that temperature and wind speed could be better predicted using all the three variables; temperature, wind speed and rainfall than it could by using only temperature and wind speed alone.

3.1.9 Impulse Response

Impulse response function (IRF) is a process to investigate the impacts of each variable within the system. Through dynamic structure, IRF demonstrate the effect of one standard deviation shocks in the error terms from a variable to another endogenous variables (Gujarati, 2004). From this study, one might be interested to understand how sudden and unexpected change of a variable would affect another variable over a period of time.

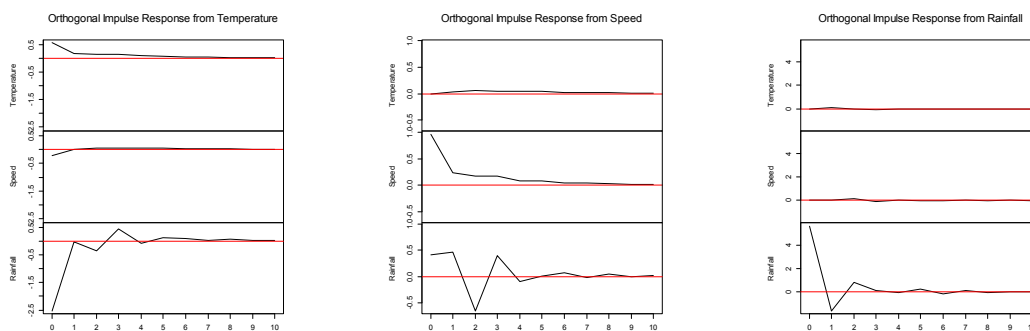


Figure 3.4. Impulse response function

Figure 3.4 shows the impact of the endogenous variables to one standard deviation shock of temperature, wind speed and rainfall. All of the endogenous variables showed positive impacts on temperature and wind speed at the initial time periods, but after that they became insignificant and slowly became zero. Temperature showed a negative impact on the rainfall in the first few period of time and slowly became insignificant and zero. Wind speed and rainfall demonstrated that initially their impact towards rainfall offer positive and significant. Thereafter the effect remained zero (Pervez Zamurad Janjuasamad et al., 2010).

3.1.10 Accuracy

Table 3.9. Accuracy checking

Order	ME	RMSE	MAE	MASE
VAR(1)	-3.17913e-18	0.5614094	0.4316827	0.9065021
VAR(3)	2.877891e-19	0.5402731	0.4191903	0.8823528

When comparing two models, we looked at the smallest values given by any one of those listed in Table 3.9, either mean error (ME), root mean square error (RMSE), mean absolute error (MAE) or mean absolute square error (MASE) to identify which model should be chosen that would give a better predicted model. From the

result, we clearly see that VAR (3) was the best model based on the smallest values obtained from any one of the four listed below.

3.2 Discussion and Conclusion

In this study, we presented VAR modeling based on meteorological data that consisted of 3 variables, namely temperature, wind speed and rainfall amount which were collected from the Kuala Krai station from the year 1985 till the year 2009. Data plotting and unit root testing showed that the data were stationary with a seasonal pattern. Seasonal differencing was required in order to remove the seasonal pattern. The lag length was chosen by information criteria to determine the order of VAR. For this set of data, VAR(1) and VAR (3) were chosen to be estimated.

The estimation of VAR (1) and VAR (3) showed that most of the parameters were significant, since they had smaller p-values where the smaller p-value indicate that the model is significant in rejecting the null hypothesis. The correlation of residuals between variables showed that the temperature had negative correlation between wind speed and rainfall. However, wind speed was positively correlated with rainfall.

Model checking was required to check whether the estimated model was an adequate model. VAR (3) model indicated that there were no autocorrelation and no heteroscedasticity effects between variables. However, VAR (1) showed that the variables were auto-correlated between them and there was a heteroscedasticity effect where the presence of heteroscedasticity can invalidate statistical tests of significance. Accuracy testing such as ME, RMSE, MAE and MASE were performed to check the smallest error given by the model. VAR (3) gave the smallest error from all of the tests.

As a conclusion, VAR modeling might not be a suitable model when dealing with these data sets since the p-value for all equations are significant in rejecting the null hypothesis. However, from the model checking we are able to verify that VAR is suitable to be applied in these data sets. VAR (3) is more suitable as compared to VAR (1) in terms of their model checking and accuracy testing. Furthermore, Johansen cointegration rank tests showed that there is no cointegration relationship among the variables that lead to conclude that VECM model is not suitable for the data set. Future work will focus on other multivariate time series methods to find a more suitable model for these data sets.

Acknowledgments

The authors would like to thank Universiti Teknologi Malaysia (UTM) and grant vote no. 08H21 for the financial funding.

References

- Ababa, A. (2012). A Time Series Analysis Of Climate Variability And Its Impacts On Food Production In North Shewa Zone In Ethiopia Climate change (CC) manifest in the form of temperature increases, changes in precipitation and sea level rise, and the intensification of. *African Crop Science Journal*, 20, 261–274.
- Brooks, C. (2002). *Introductory Econometrics for Finance*. Cambridge University Press, United Kingdom
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74(366), 427–431.
- Fraedrich, K., & Blender, R. (2003). Scaling of atmosphere and ocean temperature correlations in observations and climate models. *Phys. Rev. Lett.*, 90, 108-501.
- Gujarati. (2004). *Basic Econometrics* (4th ed). The McGraw–Hill Companies, 2004.
- Hayati, R., Meriam, N., Sulaiman, N., & Ali, B. S. (2014). Forecasting of Air Pollution Potential for a Selected Region in Malaysia. *International Journal of Engineering Development and Research*, 2(3), 3338–3358.
- 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>
- Shukla, J., & Misra, B. M. (1977). Relationships between Sea Surface Temperature and Wind Speed Over the Central Arabian Sea, and Monsoon Rainfall Over India. *Monthly Weather Review*, 105, 998–1002.
- Jamiyansharava, K., Ojima, D., Pielke, R. A., Parton, W., Morgan, J., Beltrán-Przekurat, A., LeCain, D., & Smith, D. (2011). Seasonal and interannual variability in surface energy partitioning and vegetation cover with grazing at shortgrass steppe. *Journal of Arid Environments*, 75(4), 360- 370.
- Jäntschi, dr. L. (2011). Relative outside Humidity - Analysis of Variance. *ES1002 : Workshop March 22nd-23rd 2011*.

- Kwiatkowski, D., & Phillips, P. C. B. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root How sure are we that economic time series have a unit root? *Journal of Econometrics*, *54*, 159–178.
- 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.
- Meehl, G., Stocker, T. L., & Solomon, S. (2007). Global climate projections, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., pp. 747–845, Cambridge Univ. Press, Cambridge, U. K.
- Misztal, P. (2010). Foreign Direct Investments as a factor for economic growth in Romania. *Review of Economic and Business Studies (REBS)*, *3*, 39–53. Retrieved from <http://www.cceol.com/asp/getdocument.aspx?logid=5&id=eeb8459a2e10498bb57c4e5d7092c961>
- Pervez, Z. J., Ghulam, S., & Nazakat, U. K. (2010). Impact of Climate Change on Wheat Production : A Case Study of Pakistan. *The Pakistan Development Review*, *49*(Winter), 799–820.
- Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, *75*(2), 335–346.
- Shiri, G., Salahi, B., Samadzadeh, R., & Shiri, M. (2011). The Investigation and Forecasting of Relative Humidity Variation of Pars Abad-e-Moghan, North-West of Iran, by ARIMA Model. *Research Journal of Applied Sciences*, *6*(2), 81-87.
- Smith, R. L. (1993). Long-range dependence and global warming. In Barnett, V., & Turkman, K. F. (Eds.), *pub. J. Wiley Statistics for the Environment* (pp. 141-161).
- Wan Zawiah, W. Z., Jemain, A., & Ibrahim, K. (2012). Bayesian Changepoint Analysis of the Extreme Rainfall Events. *Journal of Mathematics and Statistics*, *8*(1), 85-91.
- Wong, C. L., Venneker, R., Uhlenbrook, S., A. Jamil, A. B. M., & Zhou, Y. (2009). Variability of rainfall in Peninsular Malaysia. *Hydrol. Earth Syst. Sci. Discuss.*, *6*, 5471–5503.
- Wong, J. M. W., Chan, A. P. C., & Chiang, Y. H. (2007). Forecasting construction manpower demand: A vector error correction model. *Building and Environment*, *42*(8), 3030–3041. <http://dx.doi.org/10.1016/j.buildenv.2006.07.024>

Appendix A

Parameter Estimation

Table A1. Parameter estimation for lag 1

Vector Autoregressive Model VAR(1)			
Sample size : 287 Log Likelihood: -1546.78			
Roots of the characteristic polynomial: 0.5193 0.2755 0.2339			
Standard error in () and p-value in []			
	Temperature	Wind speed	Rainfall
Temperature (t-1)	0.4818530 (0.0573487) [2.18e-15]	0.2226 (0.1017) [0.0294]	-0.788755 (0.630098) [0.2117]
Wind speed (t-1)	0.0412267 (0.0320992) [0.20007]	0.2888 (0.05691) [7.06e-07]	0.690168 (0.352678) [0.0513]
Rainfall (t-1)	0.0151352 (0.0056463) [0.00778]	0.003206 (0.01001) [0.7490]	-0.292927 (0.062036) [3.69e-06]
Trend	-0.0003954 (0.0004059) [0.33086]	-4.652e-05 (7.196e-04) [0.9485]	0.001825 (0.004460) [0.6827]
Constant	0.0694101	-0.03618	-0.082895

	(0.0677904) [0.30676]	(0.1202) [0.7636]	(0.744822) [0.9115]
Residual standard error	0.5664	1.004	6.223
F-statistic	18.96	7.678	6.109
p-value	7.989e-14	6.942e-06	9.974e-05

Table A2. Parameter estimation for lag 3

Vector Autoregressive Model VAR(3)			
Sample size : 285 Log Likelihood: -1507.308			
Roots of the characteristic polynomial: 0.7397 0.6202 0.6202 0.5522 0.5522 0.4726 0.4726 0.3407 0.3407			
Standard error in () and p-value in []			
	Temperature	Wind speed	Rainfall
Temperature (t-1)	0.3909784 (0.0657558) [8.35e-09] ***	0.1447 (0.1175) [0.219177]	-1.132993 (0.736897) [0.1253]
Temperature (t-2)	0.1522831 (0.0688089) [0.02771] *	0.1868 (0.1230) [0.129736]	-0.221119 (0.771111) [0.7745]
Temperature (t-3)	0.0128383 (0.0657485) [0.84533]	-0.06826 (0.1175) [0.561691]	1.422592 (0.736815) [0.0545]
Wind speed (t-1)	0.0265331 (0.0340754) [0.43685]	0.2372 (0.06089) [0.000123]	0.613265 (0.381868) [0.1094]
Wind speed (t-2)	0.0338371 (0.0345165) [0.32779]	0.1102 (0.06168) [0.075162]	-0.674269 (0.386811) [0.0824]
Wind speed (t-3)	0.0209122 (0.0336650) [0.53499]	0.09568 (0.06015) [0.112855]	0.276723 (0.377269) [0.4639]
Rainfall (t-1)	0.0135976 (0.0058630) [0.02112]	0.008307 (0.01048) [0.428490]	-0.285407 (0.065704) [1.97e-05]
Rainfall (t-2)	-0.0027094 (0.0060442) [0.65432]	0.01899 (0.01080) [0.079824]	0.076195 (0.067735) [0.2616]
Rainfall (t-3)	-0.0169895 (0.0058714) [0.00411]	-0.02283 (0.01049) [0.030437]	0.080535 (0.065799) [0.2220]
Trend	-0.0002572 (0.0004007) [0.52159]	1.869e-05 (7.160e-04) [0.979197]	0.001004 (0.004491) [0.8233]
Constant	0.0492922 (0.0672244) [0.46403]	-0.03778 (0.1201) [0.753368]	0.046774 (0.753355) [0.9505]
Residual standard error	0.551	0.9846	6.175
F-statistic	10.11	5.101	3.3
p-value	1.83e-14	7.826e-07	0.0004648

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).