MATEMATIKA, 2018, Volume 34, Number 1, 73–85

Performance Evaluation of a New Hybrid Multivariate Meteorological Model Analysis: A Simulation Study

[1]Siti Mariam Norrulashikin, [2]Fadhilah Yusof* and [3]Ibrahim Lawal Kane
[1,2]Department of Mathematical Sciences
Faculty of Science, Universiti Teknologi Malaysia
81310 UTM Johor Bahru, Johor, Malaysia
[3]Department of Mathematics and Computer Science, Faculty of Science
Umaru Musa Yar'adua University, Nigeria
*Corresponding author: fadhilahy@utm.my

Article history
Received: 26 July 2017
Received in revised form: 14 November 2017
Accepted: 10 December 2017
Published on line: 1 June 2018

Abstract Simulation is used to measure the robustness and the efficiency of the forecasting techniques performance over complex systems. A method for simulating multivariate time series was presented in this study using vector autoregressive base-process. By applying the methodology to the multivariable meteorological time series, a simulation study was carried out to check for the model performance. MAPE and MAE performance measurements were used and the results show that the proposed method that consider persistency in volatility gives better performance and the accuracy error is six time smaller than the normal hybrid model.

Keywords Simulation; meteorology; time series simulation

Mathematics Subject Classification 62H12, 62M10, 62P99, 91C20

1 Introduction

Simulation process is a cycle of enhancement to well match the prerequisites of the issues element which incorporates the theoretical model building, its execution to mechanized model, and experimentations to generate large data sets [1]. Realistic and accurate data collection is needed for any decision-making process studies. Setting up the simulation experiments enable the researchers to produce large data sets that may represent many situations and people's profiles by building a process that is easy to understand. In addition, by varying the simulation parameters, the researchers are able to test the efficiency and the robustness of developed algorithm [2].

Chai [3] had developed an algorithm for multivariate autoregressive time series simulation using autoregressive base-process method. Azimmohseni *et al* [4] had proposed a simulation methodology to generate real multivariate stationary process. Chai [5] provided a general method for simulating univariate and multivariate time series data of flood variables. Cario and Nelson [6] simulated random variables from fitted autoregressive-to-anything (ARTA) using the base-process method. They have specified that one of the limitations of this approach is that the estimated base-process may not be stationary, and for this reason, a long-simulated time series may not be available. Fiorentino *et al* [7] utilized continuous hydrological simulation to investigates complexities related to spatial and temporal dependencies relating to flood, rainfall and runoff process.

A good quality of simulated data will be produced if these three conditions are met, that are, appropriate observed data sets, successful model, and precise specifications of simulation conditions. In other words, by distorting the model or conveying excessive information may prompt poor simulation process [8]. Unfortunately, there are up to 100 types of simulations from the simulation taxonomy which presume that it is practically difficult to extract the general guidelines and proper methodology for simulation processes [9]. Reliable simulation methods for multivariate stochastic process are very challenging and demanding.

In order to overcome the general guideline issues, there is a critical step required in each of the simulation processes, by checking on the verification and validation of the model and their behavior whether the simulation procedures operates the way the researcher intended and whether it behaves the way the real system does [1]. In this paper, we implemented the base-process simulation analysis method from vector autoregressive (VAR) model to the multivariate meteorological data series.

In our context of research, the simulation procedure includes several original features. The method of simulation is within the scope of problem solving and at the stage of validating the developed hybrid model building, where the developed hybrid model building is required to deal with the heteroscedastic data and support the multivariable and correlated data sets. The purpose of this study is to evaluate and verify the performance of hybrid proposed model using simulated meteorological data sets. To this end, the normal hybrid of VAR – dynamic conditional correlation (DCC) and the proposed hybrid VAR – hidden Markov model (HMM) – DCC model have been developed and the simulated data will be generated based on base-process of VAR.

This paper is organized as the following: Section 2 provides the methodology used for model building and simulation techniques, while Section 3 presents the results and discuss the findings. Lastly, conclusion is discussed in Section 4.

2 Methodology

In this section, the multivariate meteorological data is applied to the model building procedures. It includes pre-processing analysis (e.g., stationarity), developing a model and diagnostic checking. Then, the simulation analysis was carried out to validate and verify the developed model accuracy.

2.1 Model Building

Model building analysis covers multivariate time series: vector autoregressive (VAR) model, for the base model; hidden Markov model (HMM), where it is used to capture the hidden state of volatility in residuals; and dynamic conditional correlation (DCC) model, from a multivariate generalized autoregressive conditional heteroscedastic (GARCH) family, to capture the fluctuation in the volatility.

2.1.1 VAR 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):

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

where y_t is independent and integrated to order one, $I(1)$ variables, the Γ's are estimable parameters and u_t˜iid $(0, \Sigma)$ is a 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 [10].

The hypothesis:

$$\begin{aligned} H_0 &: \text{The data sets follows VAR model} \\ H_1 &: \text{The data sets do not follows VAR model} \end{aligned}$$

2.1.2 HMM Model

The joint likelihood of observations $O_{1:T}$ and hidden states, $S = \{S_1, S_2, ..., S_5\}$, given model parameters θ and covariates $\mathbf{z}_{1:T} = (\mathbf{z}_1, ..., \mathbf{z}_T)$ in a dependent mixture model is written as follow:

$$P\left(O_{1:T}, S_{1:5} \left| \theta, \mathbf{z}_{1:T} \right.\right) = \pi_i\left(\mathbf{z}_1\right) \mathbf{b}_{S_1}\left(O_1 \left| \mathbf{z}_1 \right.\right) \prod_{t=1}^{T-1} a_{ij}\left(\mathbf{z}_t\right) \mathbf{b}_{S_t}\left(O_{t+1} \left| \mathbf{z}_{t+1} \right.\right)$$

where the model is described by the following elements:

1. $\pi_i\left(\mathbf{z}_1\right) = P\left[S_1 = i \left| \mathbf{z}_1 \right.\right] \ , \quad 1 \leq i \leq 5$, provides the initial probability of class or states i at time $t = 1$ with covariates $\mathbf{z}_1$.

2. $a_{ij}\left(\mathbf{z}_t\right) = P\left[S_{t+1} = j \left| S_t = i, \mathbf{z}_t \right.\right] \ , \quad 1 \leq i, j \leq 5$, gives the transition probability from state i to state j with covariates $\mathbf{z}_t$.

3. $\mathbf{b}_{S_t}$ is observation densities vector $b_j^k\left(\mathbf{z}_t\right) = P\left[O_t^k \left| S_t = j, \mathbf{z}_t \right.\right], 1 \leq j \leq 5$ that delivers the conditional densities of observations O_t^k associated with latent class or states j and covariates $\mathbf{z}_t$, $j = 1, ..., 5$ and $k = 1, ..., m$ where m is the number of time series variables.

4. $S = \{S_1, S_2, ..., S_5\}$ giving the hidden states.

The process of an HMM is described as following. For the first step, a hidden state distribution is labelled as π at time t. Next, a certain hidden state transfers from the initial state to the next state according to the state transition probability matrix a_{ij} which describes the probabilities of particular transitions. All elements in A are positive, less than one and the total sum of every row is one. Each state emits observations according to the emission probabilities $\mathbf{b}_{S_t}$ which describe probability density of observation in a certain hidden state. The observations end at a time t_T where T is the length of the observations [11].

2.1.3 DCC Model

This model split up the volatility modelling into two stages. The first stage is to acquire the volatility series $\{\sigma_{ii,t}\}$ for $i = 1, ..., k$. In the practical estimation of DCC models, we consider a k-dimensional innovation a_t to the residuals series z_t. Univariate GARCH models are used to acquire estimates of the volatility series $\{\sigma_{ii,t}\}$. Let $F_{t-1}^{(i)}$ denote the σ-field generated by the former information of a_{it}. That is, $F_{t-1}^{(i)} = \sigma\{a_{i,t-1}, a_{i,t-2}, ...\}$. Univariate GARCH models obtain $Var(a_{it}| F_{t-1}^{(i)})$. Then again, the multivariate volatility $\sigma_{ii,t}$ is $Var(a_{it}| F_{t-1})$.

The last stage is to model the dynamic dependence of the correlation matrices ρ_t. Let $\Sigma_t = [\sigma_{ij,t}]$ be the volatility matrix of a_t given F_{t-1}, which represents the information accessible at the time $t - 1$. Then, the conditional correlation matrix is

$$\rho_t = D_t^{-1}\Sigma_t D_t^{-1}$$

where $D_t = diag\left\{\sigma_{11,t}^{1/2}, ..., \sigma_{kk,t}^{1/2}\right\}$ is the diagonal matrix of the k volatilities at time t. Let $\eta_t = (\eta_{1t}, ..., \eta_{kt})'$ be the marginally standardized innovation vector, where $\eta_{it} = a_{it}\big/\sqrt{\sigma_{ii,t}}$. Then, ρ_t is the volatility matrix of η_{it}. The DCC models is projected by Engle [12] and is defined as

$$\begin{aligned} Q_t &= (1 - \theta_1 - \theta_2)\,\bar{Q} + \theta_1 Q_{t-1} + \theta_2 \eta_{t-1}\eta_{t-1}' \\ \rho_t &= J_t Q_t J_t \end{aligned} \tag{1}$$

where for η_t, $\bar{Q}$ is the unconditional covariance matrix, θ_i are non-negative real numbers fulfilling $0 < \theta_1 + \theta_2 < 1$, and $J_t = diag\left\{q_{11,t}^{-1/2}, ..., q_{kk,t}^{-1/2}\right\}$, with $q_{ii,t}$ denotes the (i, i)th component of Q_t. From the delineation, Q_t is a positive-definite matrix and J_t is just a normalisation matrix. The correlations dynamic dependence is administered by Equation (1) with parameters θ_1 and θ_2 [12].

2.2 Simulation Method

The simulation in this paper generates a gaussian distribution of meteorological time series data sets. Let $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{kt})'$ be an observed multivariate time series for k-dimensional vector time series and $t = 1, 2, ..., n$. Then the simulation method can be summarized as listed below.

Step 1 Identify the coefficient of the vector autoregressive model estimation including covariance matrices

The coefficient of estimated VAR model has been identified from the observed multivariate time series $\mathbf{y}_t$ of size n using the formulation below [13]

$$y_t = \delta + \psi D_t + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + u_t\ , \qquad t = 0, \pm1, \pm2, ...$$

Step 2 Define base process

The base process of time series is defined by a causal process [3]

$$\mathbf{z}_t = \phi_1\mathbf{z}_{t-1} + \phi_2\mathbf{z}_{t-2} + ... + \phi_p\mathbf{z}_{t-p} + \mathbf{u}_t + \theta_1\mathbf{u}_{t-1} + \theta_2\mathbf{u}_{t-2} + ... + \theta_q\mathbf{u}_{t-q}$$

where $\mathbf{z}_t = (z_{1t}, z_{2t}, ..., z_{kt})'$, ϕ_i and θ_j are fixed $k \times k$ coefficient matrices for $i = 1, ..., p$ and $j = 1, ..., q$. $\mathbf{u}_t = (u_{1t}, u_{2t}, ..., u_{kt})'$ is $k-$dimensional normally distributed random variables with $E(\mathbf{u}_t) = \mathbf{0}_k$ and covariance matrices Σ_u as such

$$E\left(\mathbf{u}_t \mathbf{u}'_{t-h}\right) = \begin{cases} \Sigma_u & \text{if } h = 0 \\ \mathbf{0}_{k\times k} & \text{otherwise} \end{cases}$$

Step 3 Simulate a long-time series data generating process from the base process and analyze it using the proposed developed model

In this study, there is only a few information, taken from the parameter estimation of the observed time series data. Hence, it is difficult to produce the generated time series data as similar to the observed data [3]. Mean absolute percentage error (MAPE) and mean absolute error (MAE) was used as the performance measure to evaluate the performance of the proposed model.

3 Results and Discussion

In this study, the observed data of rainfall, temperature, humidity and wind speed monthly data series were collected from Alor Star meteorological stations for 25 years from 1985 to 2009. The data were found to be non-stationary and having a seasonal pattern. Seasonal differencing was used to remove the seasonality in the data series. Akaike information criterion (AIC) has been used to identify the possible number of lag length and it proposed lag two (p =2) as the optimum lag length to fit the VAR model, as shown in Equation (2).

$$\begin{bmatrix} R \\ T \\ H \\ W \end{bmatrix}_t = \begin{bmatrix} 1.5738 \\ 0.0035 \\ -0.0131 \\ -0.0227 \end{bmatrix} + \begin{bmatrix} 0.0467 & 2.842 & 0.2932 & -0.6448 \\ 0.0008 & 0.396 & -0.0210 & 0.0178 \\ -0.0030 & -0.320 & 0.4116 & 0.0073 \\ -0.0004 & 0.132 & -0.0039 & 0.2091 \end{bmatrix} \begin{bmatrix} R \\ T \\ H \\ W \end{bmatrix}_{t-1}$$
$$+ \begin{bmatrix} -0.0192 & -0.879 & 0.7783 & 0.8063 \\ 0.0010 & 0.230 & -0.0093 & 0.0262 \\ -0.0019 & 0.253 & 0.1452 & -0.3281 \\ -0.0009 & -0.224 & -0.0210 & 0.1124 \end{bmatrix} \begin{bmatrix} R \\ T \\ H \\ W \end{bmatrix}_{t-2} + \begin{bmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \\ e_{4t} \end{bmatrix} \tag{2}$$

Diagnostic checking for the VAR(2) residual reveals that it contains heteroscedastic effect in the residual, or in another word, the sum of square of residual is not constant. DCC model was adopted to VAR(2) model, to capture the heteroscedasticity in the residual. Table 1 presented the model parameter estimation of VAR(2)-DCC model. As can be seen, the summation of α and β is close to one (0.9791), indicate that there exists high persistency in the volatility. Three hidden states of the volatility: high, medium and low volatility were classified from VAR(2) residual using probabilistic HMM model to overcome the problem of high persistence in the volatility. Next, each volatility state from VAR(2)-HMM was modelled again using DCC model to capture the heteroscedastic effect in the residual. Table 1 displayed the model parameter estimation of both models, VAR(2)-HMM-DCC and VAR(2)-DCC model with the standard error (in parentheses). The summation of two parameters in all levels of VAR(2)-HMM-DCC is now not as high as in VAR(2)-DCC model (0.8197, 0.8513, and 0.6523).

Table 1: Model Parameter Estimation

Parameter	VAR(2)-DCC (standard error)	VAR(2)-HMM-DCC (standard error)		
		High volatility	Medium volatility	Low volatility
α	**0.0451** (0.0204)	0.0925 (0.0623)	0.0235 (0.0135)	0.0492 (0.0248)
β	**0.9340** (0.8803)	0.7272 (0.3618)	0.8278 (0.4216)	0.6031 (0.3884)
$\alpha + \beta$	**0.9791**	0.8197	0.8513	0.6523
AIC	20.329	18.263	18.042	17.4432

Diagnostic checking has been tested to the developed model, as presented in Table 2. As can be seen, there exists no autocorrelation in the residual data. However, the result from autoregressive conditional heteroscedastic-Lagrange multiplier (ARCH-LM) test for heteroscedasticity shows that for VAR(2) model, the p-value is less than 0.05, indicating that the existence of heteroscedasticity occurs in VAR(2) residual. This issue has been discussed in details in previous paragraph, by adopting DCC model to VAR(2) and VAR(2)-HMM. Both developed hybrid models, VAR(2)-DCC and VAR(2)-HMM-DCC model are free from autocorrelation and heteroscedastic effect.

Table 2: Residual Analysis of All Models

		Test statistics (p-value)				
Adequacy checking	Test	VAR	VAR-DCC	VAR-HMM-DCC (high)	VAR-HMM-DCC (med)	VAR-HMM-DCC (low)
Autocorrelation	Breusch-Godfrey	0.1180 (0.9432)	0.1464 (0.9315)	1.3844 (0.5575)	3.0845 (0.2999)	2.2057 (0.3335)
Heteroscedasticity	ARCH LM test	19.0280 (**0.0093**)	0.9715 (0.8065)	1.5062 (0.5811)	2.2078 (0.6374)	7.4159 (0.1367)

A simulation study is then used to evaluate the efficiency of the developed two-time series models presented in this paper, that are, VAR(2)-DCC and VAR(2)-HMM-DCC model. The motivation for this is to establish whether the model that takes into consideration the volatility persistence, VAR(2)-HMM-DCC model performs better than the one ignoring the volatility persistence, VAR(2)-DCC model. In this simulation study, a multivariate time series of length 272 from the estimated VAR(2) base process is simulate for 30 replicates [14], [15] from multivariate meteorological time series that consists of rainfall, temperature, humidity and wind speed variables. The equation below presented the coefficients and covariance matrices of VAR(2) parameter estimation that will be used for data generating process.

$$\begin{bmatrix} R \\ T \\ H \\ W \end{bmatrix}_t = \begin{bmatrix} 1.5738 \\ 0.0035 \\ -0.0131 \\ -0.0227 \end{bmatrix} + \begin{bmatrix} 0.0467 & 2.842 & 0.2932 & -0.6448 \\ 0.0008 & 0.396 & -0.0210 & 0.0178 \\ -0.0030 & -0.320 & 0.4116 & 0.0073 \\ -0.0004 & 0.132 & -0.0039 & 0.2091 \end{bmatrix} \begin{bmatrix} R \\ T \\ H \\ W \end{bmatrix}_{t-1}$$
$$+ \begin{bmatrix} -0.0192 & -0.879 & 0.7783 & 0.8063 \\ 0.0010 & 0.230 & -0.0093 & 0.0262 \\ -0.0019 & 0.253 & 0.1452 & -0.3281 \\ -0.0009 & -0.224 & -0.0210 & 0.1124 \end{bmatrix} \begin{bmatrix} R \\ T \\ H \\ W \end{bmatrix}_{t-2} + \begin{bmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \\ e_{4t} \end{bmatrix}$$

where $\mathbf{e}_t$ is a white noise vector with $E(\mathbf{e}_t) = 0_{4\times 4}$ and covariance matrix Σ_e such that

$$E\left(\mathbf{e}_t \mathbf{e}'_{t-h}\right) = \begin{cases} \Sigma_e, & \text{if } h = 0 \\ 0_{k\times k}, & \text{otherwise} \end{cases}$$

and

$$\Sigma_e = \begin{bmatrix} 12099.5 & -25.031 & 176.76 & -3.1153 \\ -25.031 & 0.2188 & -0.8195 & 0.0216 \\ 176.76 & -0.8195 & 0.7251 & -0.7861 \\ -3.1153 & 0.0216 & -0.7861 & 0.7744 \end{bmatrix}$$

The descriptive statistics of the generated data sets of each variable are given in Table 3, 4, 5, and 6, respectively. It can be seen from the mean, minimum, and maximum values of all tables that the generated data does not follow the similar range of the observed data. It has been stated in Section 2.2 that it is not easy to generate as similar to the observed data since only a few information used from the estimated parameter of the base-model [3]. The skewness and kurtosis show all variables are having a normal distribution. All 30 replicates simulation data then went through the modelling process, using the same algorithm and model as the observed data. A comparison was carried out between simulated results of the two models. This is aimed at validating the claim that VAR-HMM-DCC model performs better than the VAR-DCC model. In general, the model that considers volatility persistence in the model improves the forecast accuracy of the normal hybrid model, VAR-DCC model in meteorological time series forecasting. There is no consensus on the appropriate performance measure to access forecasting techniques, hence, MAPE and MAE are used as a performance measure in this study. Table 7 and 8 displayed the accuracy results of simulation modelling using MAPE and MAE obtained from 30 replicates simulated data sets.

It can be observed from the tables that VAR(2)-HMM-DCC model gives the smaller error (in bold) compared to VAR(2)-DCC model, six times smaller in both MAPE and MAE. These results signify that VAR(2)-HMM-DCC model performs better than the normal hybrid model.

Table 3: Rainfall Simulation

Simulation	Mean	Median	Min.	Max.	St. Dev.	Skewness	Kurtosis
1	−64.5	−947.9	−37204.1	36352.7	11968.2	−0.0294	2.9893
2	927.4	1053.3	1734.8	39528.5	11933.6	0.0465	2.9414
3	−47.6	320.4	−35443.0	35470.1	11474.4	−0.2980	3.3156
4	−361.9	−142.1	−33015.1	32343.6	12587.7	0.0842	2.7403
5	−801.9	−1598.8	−31504.7	32971.4	12063.6	0.0139	2.6132
6	1053	1299	−31019	31429	12063.3	−0.0834	2.4890
7	278.8	54.29	−35813.9	34933.7	12153.6	−0.0637	2.9964
8	−1097.2	−707.1	−36679.5	34400.3	12117.4	−0.0137	2.9179
9	−1361	−1488	−32945	31092	11748.4	−0.0229	2.7259
10	−907	−1433	−39998	37697	11955.7	−0.0661	3.4793
11	423.3	1822.3	−32388.3	32747.5	13193.9	−0.1056	2.7805
12	626.7	1217.6	−37356.8	43061.8	11858.5	−0.0911	3.2913
13	−57.6	91.63	−32437.7	41119.5	12408.6	0.0764	2.9946
14	632.1	1039.5	−38506.6	38441.3	12121.0	−0.0234	3.1757
15	−791.5	−870.7	−31538.8	34526.0	11447.7	0.0713	2.7300
16	137.6	−723.1	−30148.2	38472.4	12012.9	0.1572	3.1694
17	−406.9	−149.0	−35165.6	34124.0	11746.7	−0.0023	2.9694
18	−682.3	−692.8	−32181.3	39514.7	11676.8	0.1479	3.2227
19	106.5	−402.2	−30620.5	31227.5	11858.8	0.0247	2.9012
20	−515.3	−224.5	−38331.8	36491.7	13066.3	−0.0557	2.8453
21	401.7	−1071.0	−40710.3	33680.5	12413.4	−0.0964	3.2450
22	−197.5	−1296.5	−39143.9	34694.2	12278.8	0.1222	3.0513
23	−61.7	99.52	−36581.1	34473.6	11552.9	−0.0959	3.4041
24	−111.4	−715.0	−31489.1	36348.9	12005.3	0.2827	3.2514
25	−842.5	−2179.2	−37785.5	31068.8	12470.7	0.0509	2.7620
26	−437.4	61.14	−38664.4	33610.1	11827.9	−0.1188	3.0584
27	−286	−1537	−40417	34186	12482.4	0.0115	3.0290
28	−494.1	−610.0	−38055.2	28500.5	11972.7	−0.0860	2.7350
29	−282.7	410.9	−35003.3	26474.0	11982.7	−0.1193	2.6072
30	−565	−1804	−28953	31627	11947.4	0.1497	2.4922
Range	[−1361, 1053]	[−2179.2, 1822.3]	[−40710.3, 31734.8]	[26474, 43061.8]	[11447.7, 13193.9]	[−0.298, 0.2827]	[2.489, 3.4793]

Table 4: Temperature Simulation

Simulation	Mean	Median	Min.	Max.	St.Dev.	Skewness	Kurtosis
1	0.438	1.559	−79.340	79.639	25.0530	−0.0018	3.1067
2	−2.269	−2.348	−75.094	62.820	24.9335	−0.0396	2.7942
3	0.354	−0.710	−81.041	69.307	24.2682	0.2876	3.4453
4	0.999	1.019	−71.978	74.352	26.6759	−0.0356	2.8831
5	2.235	2.098	−66.369	64.780	24.9887	0.0158	2.6402
6	−2.622	−2.906	−62.703	70.677	25.6246	0.1128	2.4161
7	−0.664	−0.905	−74.526	78.345	25.2560	0.0879	3.2685
8	2.878	1.741	−73.434	79.085	25.0831	0.0549	3.0156
9	3.115	2.339	−64.875	66.188	24.642	0.0135	2.7022
10	2.401	4.299	−79.374	86.426	24.8799	0.0640	3.5933
11	−0.939	−3.181	−67.374	69.159	27.354	0.1210	2.7908
12	−1.407	−2.330	−84.128	78.645	24.9015	0.1246	3.2459
13	0.240	−0.898	−82.875	63.967	25.8797	−0.0156	2.9161
14	−1.443	−1.106	−80.264	76.581	25.4055	−0.0119	3.1489
15	1.636	2.295	−72.465	65.705	23.7846	−0.0634	2.7157
16	−0.338	1.174	−88.113	69.775	25.3336	−0.1355	3.2468
17	1.089	1.700	−68.212	72.998	24.7755	−0.0306	2.8688
18	1.935	2.928	−82.382	65.993	24.8520	−0.1630	3.1110
19	−0.165	1.154	−65.110	62.433	24.9571	−0.0747	2.8039
20	1.534	1.836	−78.847	83.820	27.6707	0.0755	2.9913
21	0.856	0.799	−70.264	84.414	26.2021	0.0979	3.2400
22	0.694	3.243	−72.417	76.581	25.8798	−0.1206	2.8835
23	0.201	−0.724	−69.514	80.429	24.4212	0.1191	3.4845
24	0.087	1.347	−78.907	64.218	25.4774	−0.2787	3.1911
25	1.997	3.470	−69.386	77.002	25.7258	−0.0616	2.9010
26	1.154	0.232	−72.423	72.06	24.7794	0.1070	2.9749
27	0.806	3.108	−68.433	85.956	26.2086	0.0040	3.0333
28	1.328	1.708	−60.587	87.233	25.3637	0.1261	2.8853
29	0.802	−0.046	−56.977	74.559	25.2371	0.1277	2.6413
30	1.179	3.197	−62.855	58.711	24.7826	−0.1249	2.5066
Range	[−2.622, 3.115]	[−3.181, 4.299]	[−88.113, −56.977]	[58.711, 87.233]	[23.7846, 27.6707]	[−0.2787, 0.2876]	[2.4161, 3.5933]

Table 5: Humidity Simulation

Simulation	Mean	Median	Min.	Max.	St. Dev.	Skewness	Kurtosis
1	−3.09	−11.80	−567.34	515.01	183.23	−0.0471	2.9879
2	21.01	23.22	−488.52	513.69	175.71	−0.0425	2.7952
3	−1.80	3.92	−505.47	505.70	170.54	−0.2440	3.2192
4	−8.61	−4.92	−618.24	514.65	201.19	−0.0229	2.9580
5	−19.79	−24.77	−549.80	542.31	181.24	−0.0596	2.7973
6	24.37	23.25	−484.05	485.24	187.11	−0.0565	2.4831
7	6.50	4.48	−583.07	545.03	180.69	−0.1123	3.1875
8	−26.47	−12.23	−553.19	511.36	178.88	−0.0732	2.9137
9	−30.25	−34.53	−566.15	473.07	179.68	−0.0626	2.7625
10	−21.53	−39.63	−627.07	542.68	183.83	−0.0070	3.4130
11	9.23	24.30	−525.10	525.48	202.42	−0.1738	2.7545
12	13.56	21.32	−590.13	565.66	180.33	−0.1744	3.1954
13	−2.36	−5.94	−484.41	522.49	186.05	0.0445	2.9063
14	13.66	14.44	−653.65	570.56	186.10	−0.0973	3.3853
15	−16.39	−11.45	−482.76	483.59	175.19	−0.0324	2.7357
16	3.37	−1.99	−441.51	602.33	183.48	0.1190	3.0379
17	−10.08	−17.20	−543.10	457.43	180.86	0.0177	2.7916
18	−16.86	−9.05	−459.75	563.13	174.22	0.0866	3.1979
19	1.60	0.72	−520.54	476.78	184.51	−0.0256	2.9726
20	−13.64	−12.59	−597.42	518.78	201.57	−0.0976	2.8833
21	−8.77	−10.26	−688.18	566.20	193.75	−0.1268	3.2850
22	−5.67	−25.78	−639.92	495.50	186.95	0.0498	3.1860
23	−1.55	−9.75	−569.59	503.79	172.65	−0.0380	3.4548
24	−1.57	2.24	−495.04	546.33	183.20	0.1237	3.2824
25	−18.72	−36.39	−528.35	548.03	187.58	0.2626	2.8822
26	−10.25	−2.97	−543.69	510.92	183.15	−0.1631	3.0123
27	−6.72	−16.78	−645.24	479.31	190.80	−0.1588	3.0910
28	−12.24	−20.68	−554.22	450.20	176.67	0.0034	2.8334
29	−7.42	−10.74	−449.89	478.70	178.77	0.0867	2.6818
30	−11.97	−22.76	−433.53	456.72	180.70	0.0590	2.4781
Range	[−30.25, 24.37]	[−39.63, 24.3]	[−688.18, −433.53]	[450.2, 602.33]	[170.54, 202.42]	[−0.244, 0.2626]	[2.4781, 3.4548]

Table 6: Wind Speed Simulation

Simulation	Mean	Median	Min.	Max.	St.Dev	Skewness	Kurtosis
1	0.291	−0.307	−39.717	46.685	16.755	0.1963	2.6279
2	−2.464	−3.655	−39.021	35.548	14.721	0.2328	2.9400
3	−0.035	0.352	−37.134	43.501	14.499	0.0359	2.7964
4	0.883	−0.267	−49.765	66.776	19.601	0.1693	3.2914
5	2.263	0.153	−45.478	56.476	16.213	0.3050	3.2908
6	−3.026	−4.006	−44.235	33.464	17.113	0.0213	2.4350
7	−0.803	−0.589	−44.178	48.431	15.781	0.1321	2.9685
8	3.157	2.232	−34.293	45.417	15.391	0.2064	2.7005
9	3.670	3.733	−37.258	52.724	16.785	0.3052	3.0511
10	2.518	3.583	−44.288	50.383	17.480	−0.1899	2.9130
11	−1.261	−1.523	−45.519	49.250	18.666	0.3760	2.9206
12	−1.618	−2.923	−40.967	55.021	16.140	0.4138	3.4001
13	0.349	0.954	−51.330	44.388	16.447	−0.2906	3.0127
14	−1.615	−3.937	−45.735	54.388	17.124	0.3728	3.0936
15	2.056	0.493	−32.027	45.614	16.220	0.3002	2.7530
16	−0.500	−0.404	−46.180	43.405	16.606	−0.0185	2.6151
17	1.217	2.315	−45.138	45.355	16.403	−0.0386	3.0147
18	1.910	2.374	−44.945	48.205	15.247	−0.0631	3.3009
19	−0.025	−1.276	−49.128	62.232	17.438	0.2346	3.6162
20	1.721	2.191	−44.424	46.895	18.667	0.0934	2.5374
21	1.121	−0.624	−52.315	57.897	18.524	0.1597	2.9156
22	0.635	0.691	−48.848	61.454	16.926	0.0189	3.8629
23	0.167	0.412	−36.874	45.835	14.670	0.1089	3.4331
24	0.155	−0.043	−47.052	51.347	16.488	0.2611	3.3918
25	2.070	3.158	−55.746	41.896	17.050	−0.5960	3.5626
26	1.214	0.692	−43.990	48.985	17.386	0.2588	2.8853
27	0.699	−1.532	−43.156	52.859	17.182	0.4446	3.0851
28	1.384	1.836	−39.681	43.226	14.664	−0.2043	2.9038
29	0.915	2.480	−49.422	33.225	15.357	−0.4291	2.9239
30	1.478	1.570	−38.242	44.168	16.165	0.0773	2.5613
Range	[−3.026, 3.67]	[−4.006, 3.733]	[−55.746, −32.027]	[33.225, 66.776]	[14.499, 19.601]	[−0.596, 0.4446]	[2.435, 3.8629]

Table 7: MAPE Accuracy Checking

Model		Rainfall	Temperature	Humidity	Wind Speed	Average
VAR-DCC		1.8384	1.4659	2.8190	1.4725	1.8990
VAR-HMM-DCC	High	**0.0494**	**0.1010**	**0.1637**	0.1291	**0.1108**
	Medium	0.2637	0.5180	0.7543	1.0551	0.6478
	Low	0.1242	0.2182	0.1675	**0.0690**	0.1447

Table 8: MAE Accuracy Checking

Model		Rainfall	Temperature	Humidity	Wind Speed	Average
VAR-DCC		0.7821	0.7483	0.7304	0.1577	0.6046
VAR-HMM-DCC	High	0.1830	0.1861	0.1890	0.0454	0.1509
	Medium	0.5087	0.4933	0.4797	0.1134	0.3988
	Low	**0.0437**	**0.0429**	**0.0486**	**0.0098**	**0.0363**

4 Conclusion

This section is alienated into two major points; simulation of developed model building and the verification of the simulated multivariate data sets. This simulation study is based on the developed model building whereby VAR(2)-HMM-DCC model has undergone the step-by-step time series model procedures including data pre-processing analysis, thorough diagnostic checking, one year forecasting analysis and validating accuracy checks. This study was done to verify and prove that the developed hybrid multivariate meteorological time series model to be better than the normal hybrid multivariate time series model. Accuracy tests have revealed that the developed model was better that the normal model using the simulated time series data sets in meteorological application and probably can be used in many other areas. However, some proper simulation measures may need to be developed in future so that the simulated time series data can be as close as the observed data.

Acknowledgement

The author would like to thank Universiti Teknologi Malaysia university grant number 14H00 and Ministry of High Education (MOHE), STEM Grant with vote no. A. J091002.5600.07397 for research funding.

References

[1] Sargent, R. G. Validation and verification of simulation models. In *Proceedings of the 2004 Winter Simulation Conference*. 2004. 17–28.

[2] Duchêne, F., Garbay, C. and Rialle, V. A hybrid knowledge-based methodology for multivariate simulation in home health telecare. In *Proc. of the Joint Workshop Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP) of the 9th Artificial Intelligence in Medicine Europe conference (AIME)* 2003. 87–94.

[3] Cai, Y. Multi-variate time-series simulation. *J. Time Ser. Anal.* 2011. 32: 566–579.

[4] Azimmohseni, M., Soltani, A. R. and Khalafi, M. Simulation of real discrete time Gaussian multivariate stationary processes with given spectral densities. *J. Time Ser. Anal.* 2015. 36(6): 783–796.

[5] Cai, Y., Gouldby, B., Hawkes, P. and Dunning, P. Statistical simulation of flood variables: incorporating short-term sequencing, *J. Flood Risk Manag.* 1(1): 2008. 3–12.

[6] Cario, M. C. and Nelson, B. L. Numerical methods for fitting and simulating autoregressive-to-anything processes. *INFORMS J. Comput.* 1998. 10(1): 72–81.

[7] Fiorentino, M., Gioia, A., Iacobellis, V. and Manfreda, S. Analysis on flood generation processes by means of a continuous simulation model. *Adv. Geosci.* 2006. 7. 231–236,

[8] Shannon, R. E. Introduction to the art and science of simulation. In *Proc 30th Winter Simulation Conference*. 1998. 7–14.

[9] T. I. Oren. Simulation: taxonomy. In *Systems and Control Encyclopedia*. Pergammon Press. 1987. 4411–4414.

[10] Wong, J. M. W., Chan, A. P. C. and Chiang, Y. H. Forecasting construction manpower demand: A vector error correction model. *Build. Environ.* 2007. 42(8): 3030–3041.

[11] Sun, W., Zhang, H. and Palazoglu, A. Prediction of 8 h-average ozone concentration using a supervised hidden Markov model combined with generalized linear models. *Atmos. Environ.* 2013. 81. 199–208,

[12] Tsay, R. S. *Multivariate Time Series Analysis with R and Financial Applications*, vol. 1. John Wiley & Sons, 2014.

[13] Lütkepohl, H. *New Introduction to Multiple Time Series Analysis*. Springer. 2005.

[14] Zhang, G. P. Patuwo, B. E. and Hu, M. Y. A simulation study of artificial neural networks for nonlinear time-series forecasting. *Comput. Oper. Res.* 2000. 28(4): 381–396.

[15] Taylor, B. L. The reliability of using population viability analysis for risk classification of species. *Conserv. Biol.* 1995. 9(3): 551–558.