INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING VOL. 13 NO. 1 (2021) 303-310

IJIE

Journal homepage: http://penerbit.uthm.edu.my/ojs/index.php/ijie
ISSN : 2229-838X e-ISSN : 2600-7916

The International Journal of Integrated Engineering

Evaluation the Performances of Stochastic Streamflow Models for the Multi Reservoirs

Nurul Nadrah Aqilah Tukimat[1*], Sobri Harun[2]

[1]Faculty of Civil Engineering Technology,
Universiti Malaysia Pahang, 26300 Kuantan, Pahang, MALAYSIA

[2]School of Civil Engineering,
Universiti Teknologi Malaysia, 81310 Skudai, Johor Bahru, MALAYSIA

*Corresponding Author

DOI: https://doi.org/10.30880/ijie.2021.13.01.027
Received 07 July 2021; Accepted 15 December 2020; Available online 25 February 2021

Abstract: Pedu-Muda reservoirs responsible to supply sufficient water capacity during paddy cultivation period twice a year. Thus, improper management and operation of the reservoirs creating the scarcity issue of water availability especially during dry season. Synthetic streamflow being as a main role in predicting the capability and sustainability of these reservoirs to cope the demand. This study evaluated the performances of stochastic streamflow model to produce the synthetic streamflow generation. Two comparable models; Valencia Schaake (VS) and Thomas Fiering (TF) represented for disaggregation and aggregation models, respectively. Each model was analyzed for 100 times of simulation to generate the long-term synthetic streamflow. There were 3 basis of the statistical analyses consist of lag one correlation, mean, mean absolute error (MAE), and standard deviation (St.D) of annual and monthly levels for both models were evaluated to compare the model performances. The results revealed the generated streamflow series by VS models had better performances to the historical streamflow record than the TF model in term of annual and monthly excepted on Feb, Aug, Sept, and Oct with less correlation values. The errors of these months between historical and generated correlation values are in the range of 0.14 to 0.20. However, both models can preserve a good agreement to the mean even the range of monthly streamflow were overestimated/underestimated by VS and TF models respectively. The average annual generated streamflow is predicted to reduce 0.7% (by VS) and 2.4% (by TF) from the historical record.

Keywords: Thomas fiering, Valencia schaake, streamflow, reservoir management, stochastic model

1. Introduction

Stochastic streamflow models have been widely applied in several decades to forecast the hydrological activities [1]-[3]. The synthesizing of streamflow series becomes significant information in preparing a proper plan of flood mitigation, irrigation, reservoirs management, and hydropower generation [4]. The models were derived using discrete time scale by sampling the continuous process *y(t)* at discrete points in time or by integrating the continuous time series over successive time interval [5]. The stochastic streamflow models were classified into 2 groups known as disaggregation and aggregation models based on the extraction data input. The disaggregation model refers to the ability model to disaggregate the inflow series in term of annual to season, month to daily, daily to hour, and station to sub-stations. Meanwhile the aggregation model is the direct model where the generated inflow series depends on the historical record such as annual to annual and season to season to avoid the application of coefficient into the model. The first stochastic streamflow model was introduced by Thomas and Fiering in year 1962 known as Thomas-Fiering model or PAR(1). It uses Markovian nature with periodic parameters in term of mean, standard deviations and the lag-

*Corresponding author: nadrah@ump.edu.my

one backward to form a regression model [6]. The application of Thomas-Fiering model is extensively used because the model requires limited samples and potentially to simplify the parameter estimation [7]. The model preserves the non-stationary of generated streamflow in term of mean, variance, and covariance at lag one [8]. Then, several types of classical model were developed to solve the hydrological problems in term of accuracy such as autoregression moving average (ARMA), periodic (PARMA) or contemporaneous (CARMA), autoregressive integrated moving average (ARIMA), shifting mean (SM), and gamma autoregression model (GAR). However, the aggregation models only can preserve the relevant statistic characteristic at the same level of aggregation. For example, the monthly synthetic streamflow could preserve the mean, standard deviation, skewness, and correlation to the historical monthly record but not in the annual synthetic streamflow.

The disaggregation models were developed to solve the complexity of the hydrological process in the spatial and temporal space. There are Valencia Schaake [9], Mejia-Rousselle [10] and Lane [11]. These models were applied in several reasons and purposes. [12] in their study were disaggregated the daily rainfall into hourly rainfall data due to insufficient data record. The stochastic results are expected to accurate and reliable if the region was dominated by consistence of the hydro climatic processes. [13] were practiced disaggregation model to disaggregate the daily rainfall series into regional hourly rainfall to trace the storm pattern. [3] revealed the Valencia Schaake model preserved high correlation coefficient and small error about 0.98 and 15% respectively.

Therefore, the main aim of this study is to evaluate the performance between Valencia Schaake model with the Thomas-Fiering model in generating the synthetic streamflow over 100 times simulations. The Valencia Schaake and Thomas-Fiering models were selected as represented the disaggregation and aggregation techniques respectively. The comparison of synthetic streamflow generation is very important in preparing the reliable reservoir management for the long-term planning.

1.1 Study Area

The study focused on the Pedu and Muda river, Kedah which contributed to the Pedu-Muda reservoirs, northwestern part of peninsular Malaysia. The reservoirs responsible to irrigate sufficient water demand for Muda Irrigation Scheme area, the largest paddy cultivation under Muda Agriculture Development Authority (MADA, Kedah). The combination systems between Pedu reservoir and Muda reservoir had been built because the Muda Reservoir has large catchment area with low storage capacity which requires water supply from Pedu Reservoir that having large storage capacity with small catchment. To enhance the water ability, Muda reservoir storage reacts as a back-up to the Pedu reservoir storage especially during dry season through to the Saiong Tunnel.

These reservoirs were built at the upstream of Pedu River and Muda River in 80km to the east of paddy field. In general, the monthly pattern of streamflow exhibits the curve where the quantity starts decrease in February, then escalates in November and December. At Pedu-Muda reservoir, the monthly streamflow data is based on the Pedu-Muda net inflow in year 1972 to 2000. The flow entering Pedu-Muda reservoir during these years are in the range of 190 to 400cumecs/year. The net inflow is practicing in this study because the flow of Muda reservoir will enter automatically to the Pedu reservoir through to the Saiong Tunnel without any controlling/management. All the data were provided by Muda Agriculture Development Authority (MADA, Malaysia). Fig. 1 illustrates the location of Pedu-Muda reservoir.

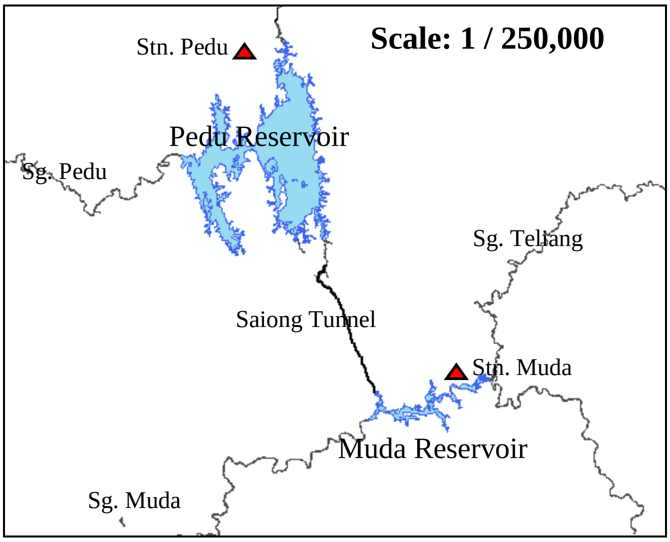

Fig. 1 - Location of Pedu-Muda reservoirs

2. Synthetic Streamflow Series using Valencia Shaake

A sequence flow of reservoir was generated by using spatial disaggregation technique known as Valencia Schaake model (VS). It was selected because the model has an ability to divide the annual flow into finer scale time series (in month) within year based on cross-correlation concept. The undemanding data series, easier understand and simple application make VS model becomes well-liked method in recent years ago. Besides, the model can generate unlimited stochastic flow by using limited historical flow time series. The potentials of VS model with other available disaggregation techniques were proved by [13], [14]. The basic form of VS is:

$$Y_t = AQ_t + B\varepsilon_t \tag{1}$$

where:

Y_t = a seasonal flow value which sum to Q_t

Q_t = annual of flow value of year t_{th}

Meanwhile, ε_t refers to the m x 1 matrix of independent standard normal deviates, *A* and *B* are referred to the parameter matrices with dimensions of m x 1 and m x m respectively by methods of moment (MOM) with m refers to the 12 months in a year. The parameter matrices of *A* and *B* were estimated by the following equations:

$$A = M_0(YQ)M_0^{-1}(Q) \tag{2}$$

$$BB^T = M_0(Y) - M_0(YQ)M_0^{-1}(Q)M_0(QY) \tag{3}$$

Three steps involve in VS model; analyze, fit and generate. First step, the flow time series will be analyzed using basic statistical parameters to observe the properties of historical flow in general. These are mean, standard deviation, coefficient of variation, and skewness. The streamflow data series is tested for normality using normal probability paper and normality test. Four types of transformations agent were engaged in this analysis there are logarithmic, gamma, power, and box cox to normalize the data series. The normality result is presented in skewness and filliben test at 10% of significance level.

In the second step is a fit stage. Fitting refers to the model selection where in this analysis the water flow time series will use VS model to disaggregate the inflow data series from annual to monthly. The model uses coefficient value to disaggregate the data using temporal disaggregation. It classifies as indirect modeling because the annual data of site study can depend on annual data at another site and the seasonal data is correspond from the annual data. The method of moment (MOM) is used to estimate the model parameters. In the fitting stage, the performances of the method are tested based on the mean, standard deviation and correlation of the generated results with the observed data. For the final step, the historical flow series are generated to unlimited time series without require calibration and validation process.

3. Synthetic Streamflow Series using Thomas Fiering

Thomas-Fiering (TF) model is a classical method that uses to generate synthetic flows time series based on the linear regression equation. It is also known as periodic lag-1 autoregressive model or noted as PAR(1). The equation for the next month seems like depending on the backward monthly flow volume. For example, flow in February is relates with the flow in January and continue for the next month. The equation of TF model was showed as follows:

$$Q_{t+1} = \overline{Q}_{t+1} + b_t(Q_t - \overline{Q}_t) + s_{t+1}I_{t+1}(1 - r_t^2)^{\frac{1}{2}} \tag{4}$$

where:

Q_{t+1} and Q_t = synthetic mean flows for month $t+1$ and t

$\overline{Q}_{t+1}$ and $\overline{Q}_t$ = observed mean monthly flows for month $t+1$ and t

b_t = observed regression coefficient of Q_t on Q_{t+1}

s_{t+1} = observed standard deviation of flow for month $t+1$

I_{t+1} = value of random deviate at $t+1$ (I_{t+1} has zero mean and unit variance)

r_t = observed correlation coefficient between Q_{t+1} and Q_t

It was classified as aggregation model because the formation of annual flow obtained from the summation of monthly or seasonal flow. By applying this model, the generated data flow is direct formation like month to month, season to season, and annual to annual. Moreover, the generated data is only focusing on that particular station without able to subdivide or disaggregate.

3.1 Performance Evaluations

To evaluate the performances of the model, the correlation at lag-1, mean absolute error (MAE) and standard deviation (St.D) were computed based on historical and generated streamflow record over 100 times simulations. The function of correlation at lag-1 is to measure the association between the monthly coefficient and the seasonal characteristics. MAE is to measure the accuracy of continuous variables through the average of errors between the two sets of data. It is relatively simple mathematic equation that widely use in forecasting analysis. The consideration of average error makes the reading is representing the whole disparity of two data sets. Meanwhile, St.D is to measure the widespread and distribution of a data estimation. It can show the range of variation from the observed mean value for the entire data. Smaller value of St.D indicates the majority of estimation data are very close to the mean of the estimation data set. The formulas of these mathematical statistics were presented in Table 1.

Table 1 - List of statistical analyses

Name	Formula	Description	Eq.
Correlation at lag-k	$r_{k,t} = \frac{m_{k,t}}{\left(m_{0,t} \times m_{0,t} - k\right)^2}$	$m_{k,t}$ = sample variance at month t	(5)
MAE	$\frac{1}{n}\sum(X_{esti} - X_{histi})$	X_{esti} = estimated of streamflow at i month X_{histi} = historical streamflow at i month	(6)
St.D	$\sqrt{\frac{1}{n}\sum(X_{est} - \overline{X_{est}})^2}$	X_{esti} = estimated of streamflow at i month $\overline{X_{est}}$ = estimated of mean streamflow at i month	(7)

4. Results and Discussions

Pedu-Muda reservoir inflow series were generated using VS and TF models. These models belong to the family of disaggregation and aggregation model respectively. During the analysis, the data set was transformed for normality by using transformation agent; power agent during month of Sept, Oct, and Nov and logarithmic agent during month of Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, and Dec. Table 2 shows the performance of transformation agent for each month. The purpose is to normalize the distribution of data set to minimize problem in the parameter estimation. To analyse the generated Pedu-Muda net inflow series, the statistical characteristics of transformed and untransformed types were evaluated graphically.

Table 2 - The transformation agent in disaggregation model

Month	Type of transformation	Skewness	Filliben value	Test result
Jan	Log	-0.05	0.99	Accept
Feb	Log	-0.08	0.99	Accept
Mar	Log	-0.17	0.98	Accept
Apr	Log	0.01	0.99	Accept
May	Log	-0.07	0.99	Accept
Jun	Log	-0.03	0.98	Accept
Jul	Log	-0.29	0.99	Accept
Aug	Log	0.05	0.99	Accept
Sep	Power	0.35	0.99	Accept
Oct	Power	0.05	0.99	Accept
Nov	Power	0.22	0.99	Accept
Dec	Log	-0.11	0.99	Accept

Fig. 2 reveals the stochastic average inflow produced by untransformed data was in good agreement to the historical average inflow in term of mean and standard deviation analyses compared to the transformed data. Obviously, the mean monthly inflow produced by untransformed data gave better accuracy than the transformed data. The inflow volume estimated by the transformed data is slightly lower than historical inflow especially in May to Oct. The error might occur due to the coefficient adjustment that applied in the transformation process. The annual net inflow produced by untransformed and transformed results are 683.5 MCM/year (-10% from historical) and 669.0 MCM/year (-12% from historical) respectively. The highest and lowest average inflow were recorded on Nov and Feb respectively. Based on the standard deviation performance, the result reveals the widespread of the simulated result by untransformed data consistent to the historical trend in a good agreement. However, the transformed data was slightly underestimated the standard deviation value through a year except in Jan, Feb, Mar, and Oct.

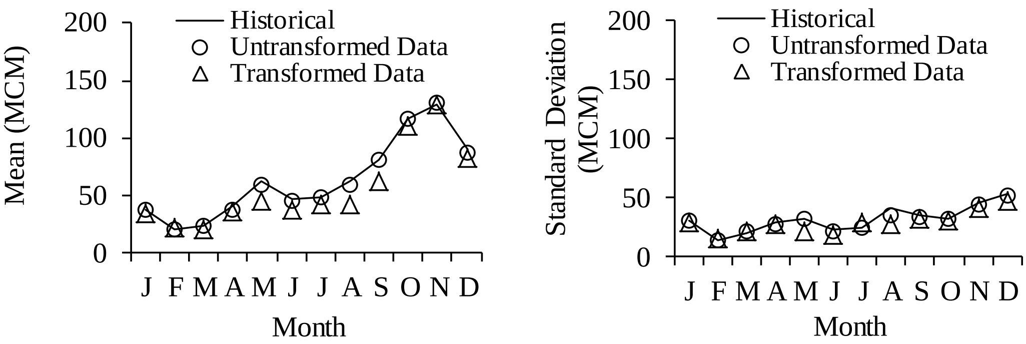

Fig. 2 - Comparison of monthly mean and standard deviation of Pedu-Muda inflow

Fig. 3 shows the spreading error in the form of box plot for every month produced by untransformed and transform simulation data. Obviously, the untransformed data produced smaller error in every month compared to the transformed data. Applying the untransformed data, the error of simulated value is estimated high in Aug to Jan. But, the error produced by transformed data was estimated higher than untransformed data according to the month. Thus, the untransformed data is expected to produce better simulated result compare to the transformed data due to the smaller wide range of monthly error.

These statistical characteristics results confirmed that the untransformed data leading in accuracy and reliability of flow generation. Further, the monthly error could be minimizing with set aside the monthly coefficient factors in the analysis. Therefore, the untransformed data will be used to generate the synthetic streamflow of Pedu-Muda based on VS and TF models.

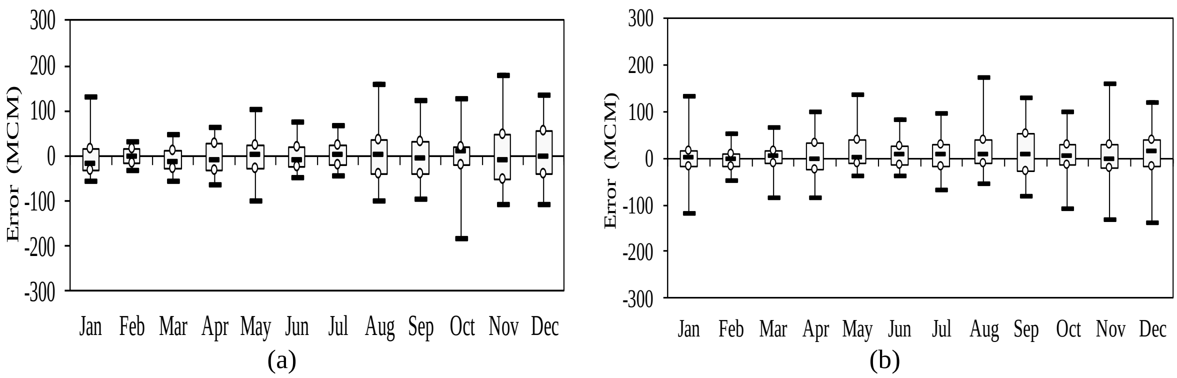

Fig. 3 - Error dispersion of (a) untransformed data (b) transformed data with the historical inflow record

4.1 Streamflow Generation

With applying untransformed data series, the synthetic streamflow of Pedu-Muda reservoir was generated to 100 samples of seasonal (12month) data for each 30 years length record. To simulate the VS model, two autoregressive (AR) models were considered in the fitting stage of untransformed data there are AR(1) and AR(0). The Akaike information criterion coefficient (AICC) was used as an agent to measure the goodness of fit of the model. The smallest value of AICC is considered as good analysis to generate the synthetic inflow series. Therefore, the AR(0) model displayed a best performer that produced lower AICC value (341.0) than AR(1) is 342.1. Hence, the AR(0) model was selected to generate the VS model. Meanwhile, the TF model was used the periodic lag-1 autoregressive model or noted as PAR(1). The AICC by this analysis is 342.1.

To ensure the accuracy of the simulation, the generated synthetic inflow series produced by VS model and TF model were compared to the historical record by correlation of lag-one season to season to season shown in Fig. 4. As overall, the results reveal the generated monthly inflow time series using VS model was successfully preserved a good agreement to the historical record for the whole months excepted on Feb, Aug, Sept, and Oct with less correlation values. The errors of these months between historical and generated correlation values are in the range of 0.14 to 0.20. The correlation value produced by TF model is predicted consistent to the VS model but with the lower correlation value especially in Aug and Nov. The highest errors between the historical and generated correlation value produced by TF model is 0.4 during Nov. Even though, the generated synthetic inflow series produced by VS and TF models are considered acceptable to be used in this study.

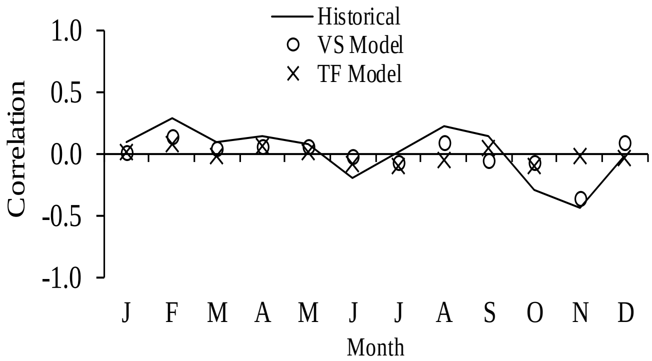

Fig. 4 - Comparison of monthly lag one season to season correlation using VS model and TF model with the historical value

Table 3 shows the statistical comparison between VS and TF models in term of annual and seasonal of generated synthetic streamflow series. The mean absolute error (MAE) and standard deviation (St.D) were used to compare the model performance. The results claim the VS model preserved better generated streamflow volume for annual and seasonal levels than the TF model. Similarly, the MAE and St.D of seasonal streamflow is predicted lower than annual generated flow may affect from the historical monthly streamflow that was practiced in this study.

Table 3: Statistical comparison between VS and TF models

Statistical Analysis		VS model	TF Model
Annual (MCM/year)	MAE	143	146
	St.D	142	144
Seasonal (MCM/year)	MAE	2	3
	St.D	33	34

Fig. 5 shows the generated 100 series of synthetic inflow based on 30years length record for both models. The synthetic inflow is estimated to high/low inconsistently during 100 of generation results based on the historical inflow pattern. The synthetic inflow series produced by VS model and TF model are predicted to produce similar trend. Generally, the annual flow is predicted to reduce at the end of 100years. The minimum and maximum flows are estimated as 400MCM/year and 1100MCM/year respectively. The highest annual inflow is estimated to experience in upcoming year period of 38 and 83 achieve more than 1050MCM. The generated inflow expected to drop less 400MCM/year at upcoming year period of 73 and 75. Even thought, the overall generated inflow series was reflected the historical record. This is because the formulation of synthetic inflow time series was completely depends on the historical inflow record as input without considers any climatic factors. Therefore, the average annual generated synthetic inflow is 761MCM/year (reduced 0.7% from the historical record) and 748MCM/year (reduced 2.4% from the historical record) for VS model and TF model respectively.

Fig. 6 presents the range of monthly streamflow for VS and TF models compared to the historical mean streamflow. Generally, the streamflow pattern of Nov and Dec are experienced to receive higher average inflow volume meanwhile month of Feb and Mar as the lesser average inflow volume. From this figure, the average inflow volume that estimated by VS and TF models have good agreement to the historical record after 100times simulations. Month of Nov is predicted to overestimate and underestimate by VS and TF models respectively with smaller error. The synthetic inflow produced by VS model is also expected higher than TF model in month Jan, Feb, May, Jul, Aug, Nov and Dec. The higher different inflow volumes between VS and TF models are estimated in Jan, Feb, May, Aug and Dec with the range of 6 - 46MCM/month.

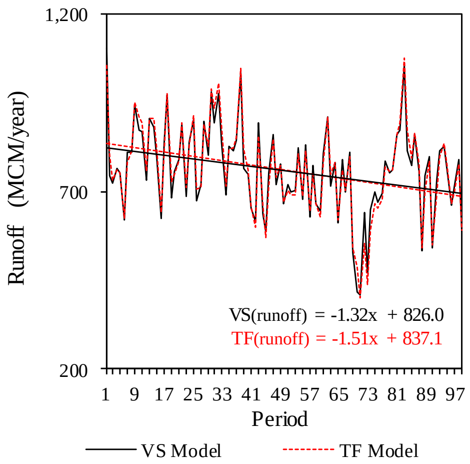

Fig. 5 - Generated 100 series of annual inflow using VS model and TF model

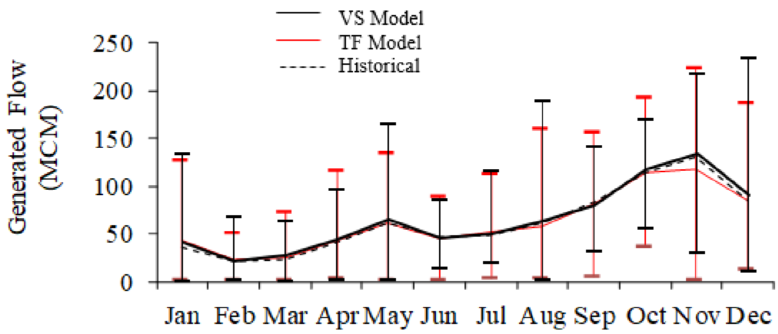

Fig. 6 - Box plot for average generated monthly inflow using VS model and TF model

5. Conclusion

The performances of the VS and TF models in generating the synthetic streamflow time series were investigated and presented in this study. The comparison among these stochastic models is very necessary to ensure the reliability of the reservoir management planning. There were 3 statistical analyses were considered to measure the performances among models; monthly lag one season to season correlation, MAE, and St.D in the annual and seasonal comparison. The results proved the untransformed data yield the most accuracy to estimate the synthetic streamflow without require any transformation agent. The use of original data can control the reliability and accuracy in the analyses. The application of monthly coefficient factors in the analysis could raise the monthly error even the normalized data set produced better skewness and Filliben test value than the actual data set.

In this study, 100 series of synthetic streamflow for Pedu-Muda reservoirs were generated by both models to obtain the streamflow series in the long-term. By this, it can preserve the accuracy of the findings based on average 100 times simulation. The statistical analyses proved the VS model produced better generated synthetic results in the annual and seasonal levels than the TF model with lower value of MAE and St.D. Even the generated mean monthly streamflow by the TF model is well to the periodical mean of historical data, but the generated synthetic streamflow was generally underestimated compared to the historical annual streamflow.

Acknowledgement

This research is supported by Universiti Malaysia Pahang (grant vot RDU1803156), Ministry of Higher Education (MHE) (FRGS/1/2019/TK01/UMP/02/1 grant vot RDU1901141), Malaysian Meteorological Department (MMD), and Drainage and Irrigation Department (DID).

References

[1] Brunner, M. I. (2019). Technical Note: Stochastic Simulation of Streamflow Time Series using Phase Randomization. Hydrology and Earth System Sciences, 23(8), 3175-3187

[2] Pender, D., Patidar, S., Pender, G. & Haynes, H. (2016). Stochastic Simulation of Daily Streamflow Sequences using a Hidden Markov Model. Hydrology Research, 47(1), 75-88

[3] Farmer, W. H. & Vogel, R. M. (2016). On the Deterministic and Stochastic Use of Hydrologic Models. Water Resources Research, 52 (7), 5619-5633

[4] Vogel, R. M. (2017). Stochastic Watershed Models for Hydrologic Risk Management. Water Security, 1, 28-35

[5] Salas, J. D., (1993). Analysis and Modeling of Hydrologic Time Series. Handbook of Hydrology, McGraw-Hill.

[6] Yohannes, H. S., (2009). Stochastic Simulation of Streamflow and Hydrologic Drought Analysis (Case Study: Upper Blue Nile Basin). Master Thesis, Addis Ababa University

[7] Rajagopalan, B., Salas, J. & Lall, U. (2010). Stochastic methods for modeling precipitation and streamflow, in Advances in Data-based Approaches for Hydrologic Modeling and Forecasting, Advances in Data-Based Approaches for Hydrologic Modeling and Forecasting. World Scientific Report, pp 17-52

[8] Yousif, A. A., Aswad, F. K. & Ibrahim, S. A. (2016). Performance of ARIMA model and modified Thomas Fiering model for predicting the monthly rainfall data for Tallafar Station. Journal of Polytechnic, 6(1), 17-27.

[9] Valencia, R. D. & Schaake Jr, J. C. (1973). Disaggregation processes in stochastic hydrology. Water Resources Research, 9(3), 315-326

[10] Mejia, J. M. & Rousselle, J. (1976). Disaggregation models in hydrology revisited. Water Resources Research, 12(2), 185–186

[11] Lane, W. L. (1998). Applied Stochastic Techniques: User Manual. Technical Service Center, Water Resources Services

[12] Park, H. & Chung, G. (2020). A nonparametric stochastic approach for disaggregation of daily to hourly rainfall using 3-day rainfall pattern. Water. 12 (8), 2306

[13] Socolofsky, S. A., Adams, E. & Entekhabi, D. (2001). Disaggregation of daily rainfall for continuous watershed modeling. Journal of Hydrologic Engineering, 6(4), 300-309

[14] Ismail, N. A., Harun, S. & Yusop, Z. (2004) Synthetic simulation of streamflow and rainfall data using disaggregation models. Jurnal Kejuruteraan Awam, 16(2), 56-65