Jurnal Kejuruteraan Awam 16(2): 56-65 (2004)

SYNTHETIC SIMULATION OF STREAMFLOW AND RAINFALL DATA USING DISAGGREGATION MODELS

Nurul Adzura Ismail[1], Sobri Harun[1], Zulkifli Yusop[2]
[1]*Department of Hydraulics and Hydrology, Faculty of Civil Engineering, Universiti Teknologi Malaysia, Skudai 81310, Johor, Malaysia.*
[2]*Institute of Environmental and Water Resources Management, Faculty of Civil Engineering, Universiti Teknologi Malaysia, Skudai 81310, Johor, Malaysia.*

Abstract: Synthetic hydrological series is useful for evaluating water supply management decision and reservoir design. This paper examines stochastic disaggregation models that are capable of reproducing statistical characteristics especially mean and standard deviation of historical data series. Simulation was carried out on both transformed and untransformed streamflow and rainfall series of Sungai Muar. The Synthetic Streamflow Generation Software Package (SPIGOT) model was found to be the most robust for streamflow simulation. On the other hand, the Valencia-Schaake (VLSH) model is more superior for generating rainfall series.

Keywords: *Streamflow; Rainfall; Simulation; Disaggregation.*

Abstrak: Siri hidrologi sintetik ialah satu kaedah yang berguna untuk menilai keputusan dalam pengurusan bekalan air dan rekabentuk empangan. Kertas kerja ini bertujuan menguji model disaggregasi stokastik yang berupaya menghasilkan semula ciri-ciri statistik terutama min dan sisihan piawai ke atas siri bulanan aliran sungai dan hujan. Simulasi telah dilakukan dengan kaedah transformasi dan tanpa transformasi menggunakan data aliran sungai dan hujan tadahan Sungai Muar. Model *Synthetic Streamflow Generation Software Package* (SPIGOT) dipilih sebagai model yang paling baik untuk simulasi aliran sungai manakala model Valencia-Schaake (VLSH) memberi keputusan yang terbaik dalam menjana siri hujan.

Katakunci: *Aliran Sungai; Hujan; Simulasi; Disaggregasi.*

1. Introduction

The streamflow and rainfall sequences to be analyzed may be thought of as one particular realisation, produced by underlying probability mechanism of the phenomenon. In other words, in analysing streamflow and rainfall sequences many hydrologists regard it as a realisation of a stochastic process. The generated data sequences, particularly monthly time series such as streamflow or rainfall are widely used in water resources planning and management to understand the variability of future system performance. Stochastic data generation aimed at generating synthetic data sequences that are statistically similar to the observed data sequences. Therefore, the generated data is important for more accurate solution of various complex planning, design and operational problems in water resources development (Yevjevich, 1989). Typically, stochastic simulation of hydrologic time series such as streamflow and rainfall are based on mathematical models (Salas, 1993; Hipel and McLeod, 1994). Simulation methods for the hydrologic time series can be classified into disaggregation and aggregation methods (Harun, 1999). Valencia and Schaake (1973) proposed the so-called disaggregation model that subsequently becomes a major technique for modeling hydrologic series. Further modification and applications of disaggregation model have led to the development of other models such as Mejia - Rouselle, SPIGOT and Lane.

The main objective of this study was to develop and test various stochastic models capable of reproducing the historical statistical characteristics especially mean and standard deviation of monthly streamflow and rainfall series. The streamflow and rainfall records of Sungai Muar were obtained from the Department of Irrigation and Drainage (DID), Malaysia. Ten data sets were simulated from the historical records and each set consists of one hundred years of flow and rainfall sequences. The model performance for untransformed and transformed data series was compared.

2. Methodology

Several disaggregation models namely Valencia-Schaake model (Valencia and Schaake, 1973), Mejia-Rouselle model (Mejia and Rouselle**,** 1976), Lane (Lane, 1979) and Synthetic Streamflow Generation Software Package (Grygier and Stedinger, 1991) were tested. The basic form of Valencia-Schaake model (VLSH) can be written in a matrix form as:

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

for the case of disaggregating annual flows into seasonal flows, $\mathbf{Y_t}$**,** $\mathbf{Y_t^i = \left(y_{t,1}, y_{t,2}, \ldots . y_{t,m}\right)}$ is a column matrix containing the seasonal flow values which sum to $\mathbf{Q_t}$, and for the same time **Q** and **Y** are referred as flow values. $\mathbf{Q_t}$ is a 1 x 1 matrix of the annual streamflow value of year t, m is taken as 12 month, $\boldsymbol{\varepsilon_t}$ is the m x 1 matrix of independent standard normal deviates, and **A** and **B** are the parameters matrices with dimensions of m x 1 and m x m, respectively. The Mejia-Rouselle (MJRS) model takes the following form:

$$\mathbf{Y_t = A\,Q_t + B\,\varepsilon_t + C\,Y_{t-1}} \tag{2}$$

in which $\mathbf{Y_t}$, $\mathbf{Q_t}$, $\boldsymbol{\varepsilon_t}$, **A** and **B** are similar to the Valencia-Schaake model and **C** is the additional (m x m) parameter matrix. The Lane model (LANE) for a single site can be written as:

$$\mathbf{Y_{v,t} = A_t\,Q_v + B_t\,\varepsilon_{v,t} + C_\tau\,Y_{v,t-1}} \tag{3}$$

in which $\mathbf{Y}_{v,t}$ is the seasonal streamflow vector; $\mathbf{Q}_v$ is the annual streamflow vector; $\boldsymbol{\varepsilon}_{v,t}$ is the vector of normally distributed noise term with mean zero and the identity matrix as its variance – covariance matrix. The noises $\boldsymbol{\varepsilon}_{v,t}$ are independent in time and space; *v* denotes the number of year and *t* is for season (month). The condensed disaggregation procedures are based on Synthetic Streamflow Generation Software Package (SPIGOT). The model takes the following form:

$$\mathbf{Y_{v,t} = A_t + B_t\,Q_{v,t} + E_{v,t}} \tag{4}$$

where *t* is for month; v is for year; $\mathbf{Y_{v,t}}$ is the seasonal streamflow vector; $\mathbf{E_{v,t} = C_t\,E_{v,t-1} + \varepsilon_{v,t}}$ is the normally distributed innovations of $\boldsymbol{\varepsilon}_{v,t}$, the independent zero-mean normal random vectors; and $\mathbf{Q}_{v,t}$ is the generated annual streamflow vector.

For the case of rainfall simulation, the VLSH model can be written as:

$$\mathbf{X_t = A\,Z_t + B\,\varepsilon_t} \tag{5}$$

where $\mathbf{X_t}$, $X_t^i = \left(x_{t,1}, x_{t,2}, \ldots . x_{t,m}\right)$ is a column matrix containing the seasonal rainfall values which sum to $\mathbf{Z_t}$, and for the same time Z and X represent rainfall values. $\mathbf{Z_t}$ is a 1 x 1 matrix of the annual rainfall value of year *t*, *m* is taken as 12 month, $\boldsymbol{\varepsilon_t}$ is the m x 1 matrix of independent standard normal deviates, and **A** and **B** are the parameters matrices with dimensions m x 1 and m x m, respectively. For the case of MJRS model, the rainfall series simulation reads:

$$\mathbf{X_t = A\,Z_t + B\,\varepsilon_t + C\,X_{t-1}} \tag{6}$$

where the $\mathbf{X_t}$, $\mathbf{Z_t}$, **A**, **B** and $\boldsymbol{\varepsilon_t}$ are similar with equation (5) and **C** is the additional (m x m) parameter matrix. The LANE model for rainfall series can be written as:

$$\mathbf{X_{v,t} = A_t\, Z_v + B_t\, \varepsilon_{v,t} + C_t\, X_{v,t-1}} \tag{7}$$

The development of streamflow and rainfall simulation model involves three procedures; statistical analysis of data, fitting a stochastic model, and generating a synthetic series. The simulation models used in this study are based on work by Toa and Delleur (1976) and Grygier and Stedinger (1991). The quality of the data was examined by performing time series plotting, distribution plot and descriptive analyses. The statistical characteristics of the observed data series are important factors in selecting the type of model. Basically, a streamflow and rainfall series can be characterised by their mean, standard deviation, skewness coefficient and season-to-season correlation coefficient.

In the case of non-normally distributed monthly streamflow and rainfall series the data was first transformed. Logarithmic, Box-Cox and power transformation techniques were generally sufficient to obtain normal distribution. The use of normally distributed series (after transformation) is preferred because their statistical properties are well established compared to the original series (non normal distribution). The generated series were then compared with the historical records for the untransformed and transformed cases. The normality of the data was tested using the skewness of normality (Valencia and Schaake, 1973). The descriptive analysis consists of parameter estimation and model testing to ensure that the model can fit the data well. The parameter estimation step was done after the type of model(s) was selected. The model parameters were estimated either by the method of moments (MOM) or by the least squares methods (LSM).

The parameter estimations of the stochastic model were tested to ensure that the model comply with the model requirement. The goodness of fit test involved checking the residuals and comparing the model with the historical properties. The basic assumptions about the residual are that they are normal and independent. For testing whether the residuals are independent, two approaches were used. The first compute the correlation coefficient and check whether the values are statistically equal to zero. The second approach used the Porte Manteau lack of fit test as described by Valencia and Schaake (1973) and Grygier and Stedinger (1991).

3. Results and Discussion

This analysis used 26 years of streamflow and 59 years of rainfall data of Sungai Muar. The statistics of the historical and generated streamflow and rainfall sequences were computed and compared. Figures 1 and 2 show that all disaggregation candidate models can adequately preserve the historical mean for transformed and untransformed streamflow. However, the SPIGOT model preserves the historical standard deviation better than the other candidate models (Figures 3 and 4). VLSH is obviously the best model for flow simulation as it consistently preserve the statistical properties of most of the mean monthly values for both untransformed and transformed flow series. Nevertheless, all the tested models failed to preserve the skewness coefficient for both untransformed and transformed streamflow and rainfall series.

In addidition, Box and Whisker plots of the annual flow for each model are shown in Figures 5 and 6. The candidate models consistently demonstrate a good reproduction of the historical properties (annual mean) for the transformed flow sequences (Figure 5). The tested models, however, are less promising in preserving the historical annual standard deviation except for the VLSH model (Figure 6).

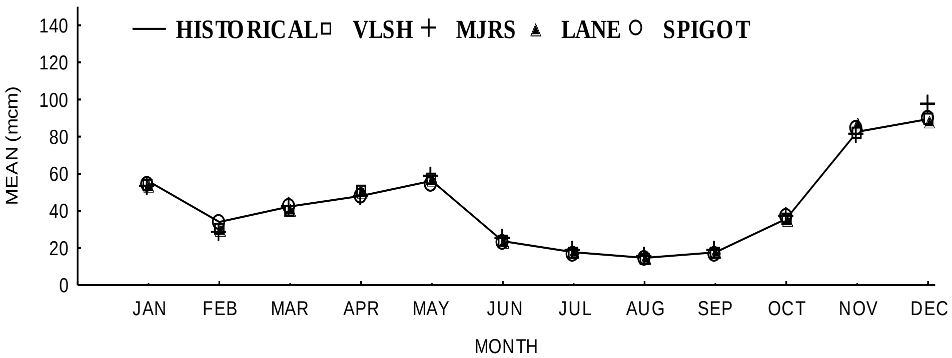

Figure 1: Mean monthly streamflow obtained by various disaggregation models (transformed flow)

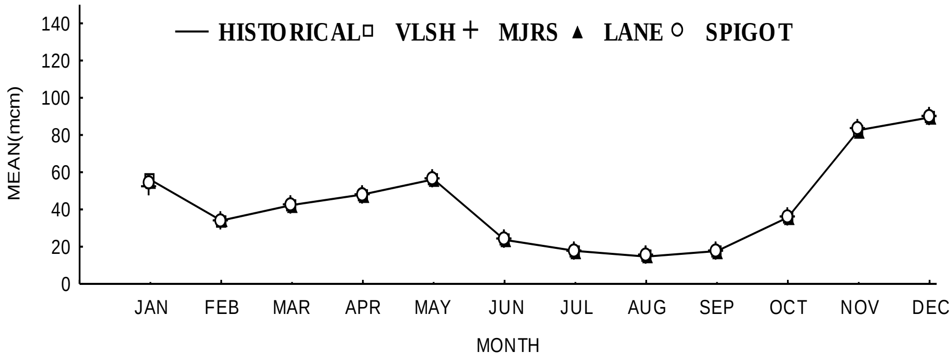

Figure 2: Mean monthly streamflow obtained by various disaggregation models (untransformed flow)

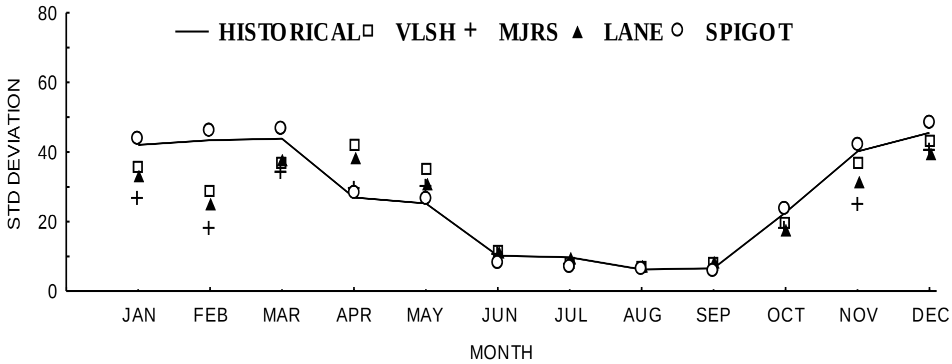

Figure 3: Mean monthly standard deviation of streamflow obtained by various disaggregation models (transformed flow)

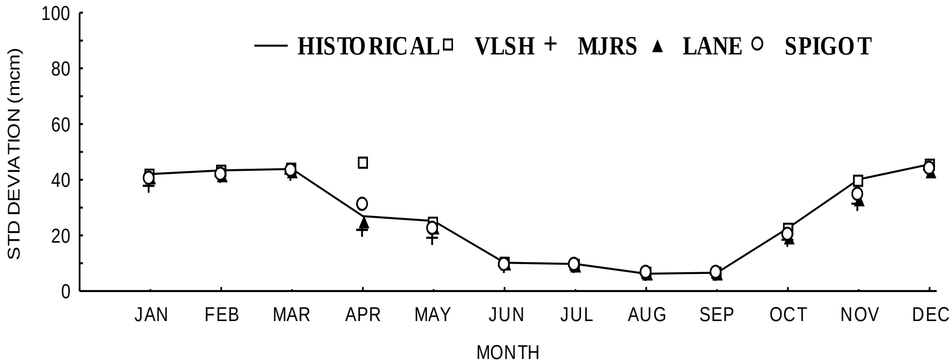

Figure 4: Mean monthly standard deviation of streamflow obtained by various disaggregation models (untransformed flow)

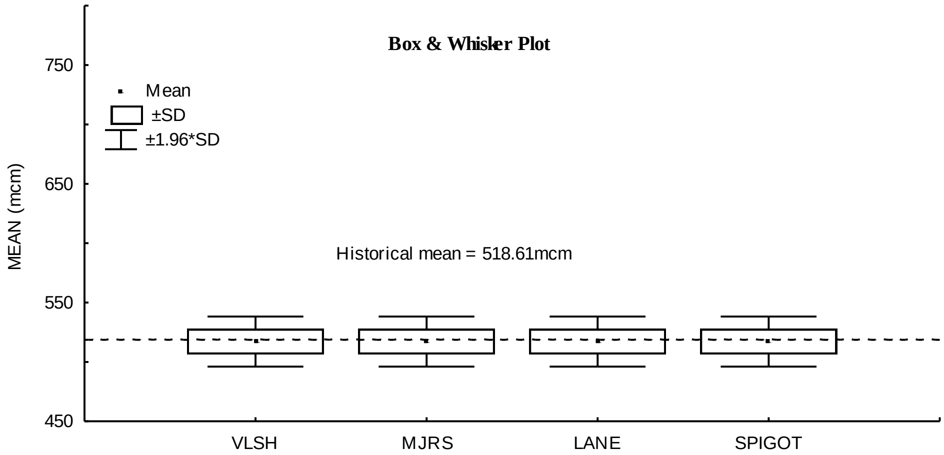

Figure 5: Annual mean streamflow obtained by various disaggregation models (transformed flow)

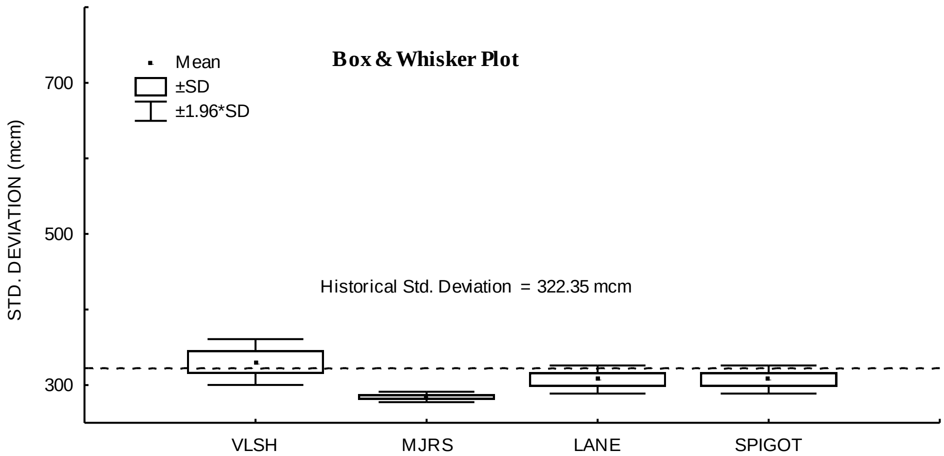

Figure 6: Annual standard deviation of streamflow obtained by various disaggregation models (transformed flow)

Similar comparisons were made for the rainfall simulation. In Figures 7 and 8, all the disaggregation models are able to produce very similar mean monthly with the historical mean for both transformed and untransformed rainfall series. Similar results were obtained for the preservation of mean monthly standard deviation (Figure 9). As for the annual data, the candidate models are successful in preserving the historical annual mean for transformed rainfall series (Figure 10). However, the MJRS and LANE models slightly underestimated the historical standard deviation (Figure 11). In overall, the VLSH model is the most robust for preserving the historical annual standard deviation for the transformed rainfall sequences.

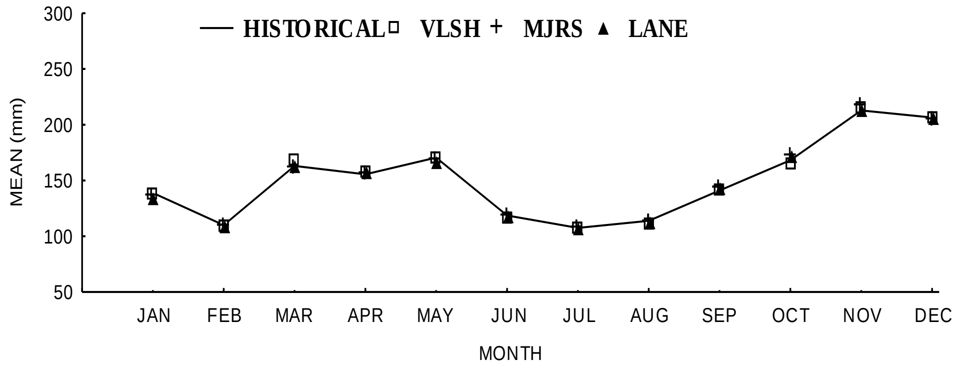

Figure 7: Mean monthly rainfall generated by various disaggregation models (transformed series)

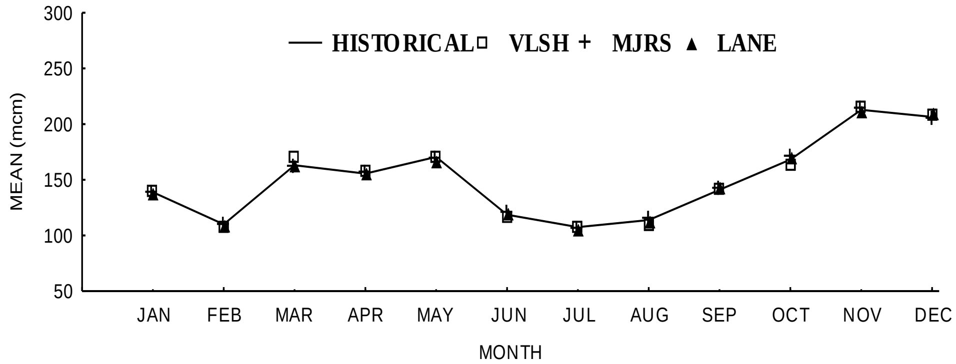

Figure 8: Mean monthly of rainfall obtained by various disaggregation models (untransformed series)

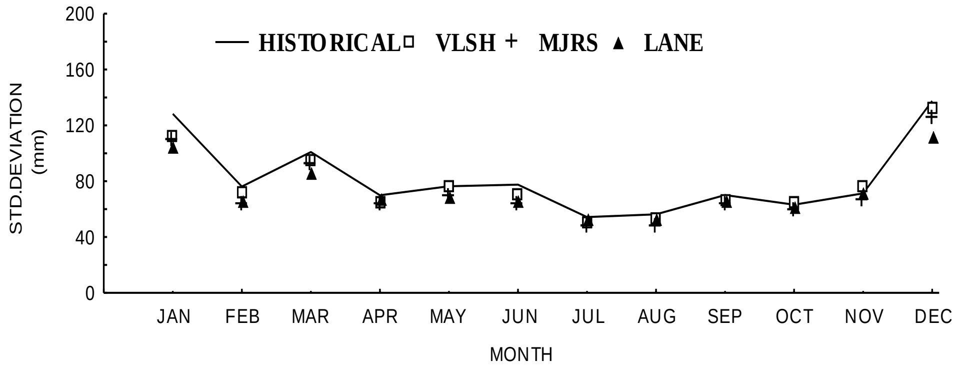

Figure 9: Mean monthly standard deviation of rainfall obtained by various disaggregation models (transformed series)

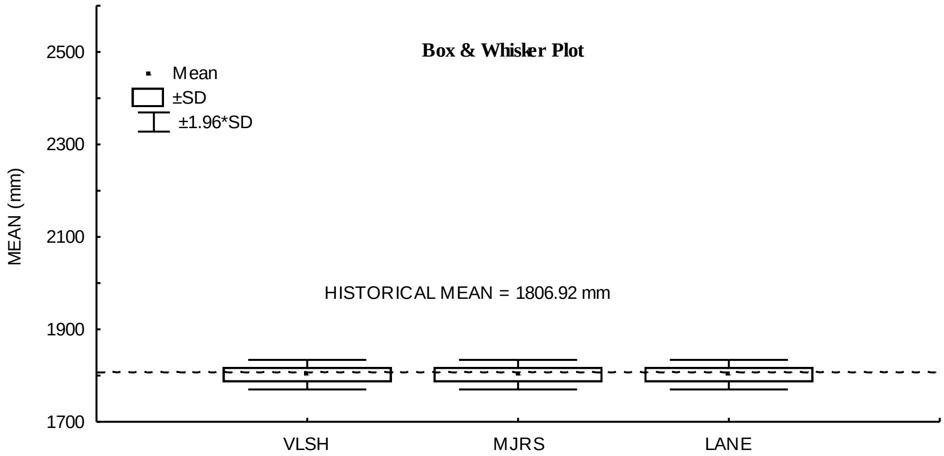

Figure 10: Annual mean of rainfall obtained by various disaggregation models (transformed series)

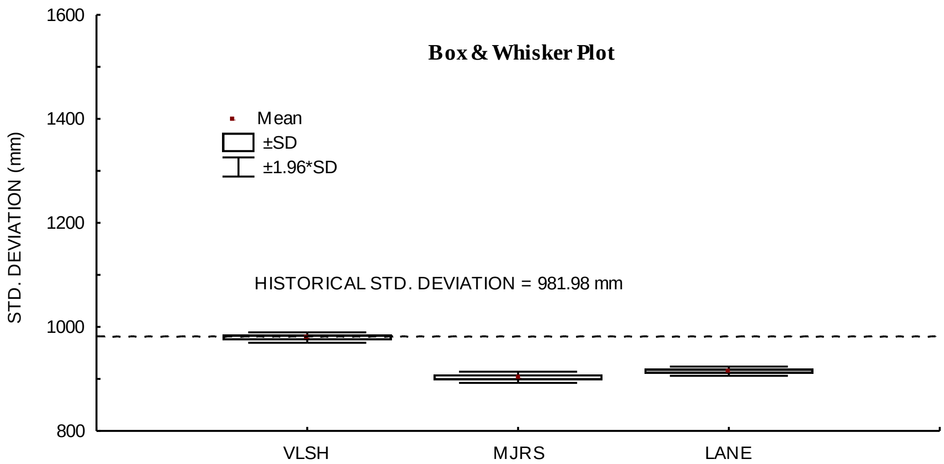

Figure 11: Annual standard deviation of rainfall obtained by various disaggregation models (transformed series)

4. Conclusion

The tested disaggregation models have successfully preserved the historical mean and standard deviation of streamflow and rainfall series of Sungai Muar. Nevertheless, the models failed to preserve the skewness coefficient. VLSH model was found to be the best stochastic disaggregation technique as it produces very similar properties to the historical streamflow and rainfall series. These findings, however, must be considered preliminary as the models were only tested on one site. Due to variability of hydrological data, this study must be replicated to a number of rivers in Malaysia before any generalization can be made.

Acknowledgements

This research has been supported by RMC-UTM Vote No 75034 and financial support provided by NSF Ministry of Science and Technology. The Department of Irrigation and Drainage (DID) is gratefully acknowledged for providing the data.

References

Grygier, J.C.and Stedinger, J.R. (1991) *SPIGOT: A Synthetic Streamflow Generation Software Package User's Manual V2.6*, School of Civ. and Envir. Eng., Cornell Unversity, Ithaca, N.Y.

Harun, S. (1999) *Forecasting and Simulation of Net Inflows for Reservoir Operation and Management*. Ph.D. Thesis, University Teknologi Malaysia, 220 pp.

Hipel, K.W., and McLeod, A.I. (1994) *Time Series Modeling of Water Resources and Environmental Systems*. Elsevier, Amsterdam.

Lane, W.L. (1979). *Applied Stochastic Techniques (LAST computer package), User Manual*. Division of Planning Technical services, Bureau of Reclamation, Denver, Colarado, USA.

Mejia, J.M., and Rouselle, J. (1976). Disaggregation models in hydrology revisited. *Water Resources Research*, 12(2): 185-186.

Salas, J.D. (1993) Analysis and modeling of hydrologic time series. *In* Maidment, D. (Ed.), Handbook of Hydrology, McGraw Hill, New York., p. 19.1-19.72.

Toa, P.C., and Delleur, J.W. (1976) Seasonal and nonseasonal ARMA models. *Journal of the Hydraulics Division*, 102(HY10): 1541-1559.

Valencia, R.D. and Schaake, J.C. (1973) Disaggregation process in stochastic hydrology. *Water Resources Research*, 9(3): 1295-1306.

Yevjevich, V. (1989) General introduction to application of stochastic hydrology in water resources. Proceedings of the NATO Advanced Study Institute on Stochastic Hydrology and its use in Water Resources Systems Simulation and Optimization, Peñíscola, Spain. 18-29 September 1989