Stochastic Modeling of Rainfall Series in Kelantan Using an Advanced Weather Generator

A. H. Syafrina
Department of Mathematics
Faculty of Science
Universiti Putra Malaysia
Serdang, Selangor, Malaysia
syafrina@upm.edu.my

A. Norzaida
UTM Razak School of Engineering
and Advanced Technology
Universiti Teknologi Malaysia
Kuala Lumpur, Malaysia
zaida.kl@utm.my

O. Noor Shazwani
UTM Razak School of Engineering
and Advanced Technology
Universiti Teknologi Malaysia
Kuala Lumpur, Malaysia
noorshazwaniosman@gmail.com

Abstract—**Weather generator is a numerical tool that uses existing meteorological records to generate series of synthetic weather data. The AWE-GEN (Advanced Weather Generator) model has been successful in producing a broad range of temporal scale weather variables, ranging from the high-frequency hourly values to the low-frequency inter-annual variability. In Malaysia, AWE-GEN has produced reliable projections of extreme rainfall events for some parts of Peninsular Malaysia. This study focuses on the use of AWE-GEN model to assess rainfall distribution in Kelantan. Kelantan is situated on the north east of the Peninsular, a region which is highly susceptible to flood. Embedded within the AWE-GEN model is the Neyman Scott process which employs parameters to represent physical rainfall characteristics. The use of correct probability distributions to represent the parameters is imperative to allow reliable results to be produced. This study compares the performance of two probability distributions, Weibull and Gamma to represent rainfall intensity and the better distribution found was used subsequently to simulate hourly scaled rainfall series. Thirty years of hourly scaled meteorological data from two stations in Kelantan were used in model construction. Results indicate that both probability distributions are capable of replicating the rainfall series at both stations very well, however numerical evaluations suggested that Gamma performs better. Despite Gamma not being a heavy tailed distribution, it is able to replicate the key characteristics of rainfall series and particularly extreme values. The overall simulation results showed that the AWE-GEN model is capable of generating tropical rainfall series which could be beneficial in flood preparedness studies in areas vulnerable to flood.**

Keywords-weather generator; flood; rainfall intensity; probability distribution; northeastern monsoon

I. INTRODUCTION

Weather generator is a numerical tool that uses existing meteorological records to produce a series of long daily synthetic weather data that is often short and non-continuous [1-2]. Various weather generators have been built throughout the years, such as the Weather Generator (WGEN) [3-4], Climate Generator (CLIGEN) [5], Long Ashton Research Station-Weather Generator (LARS-WG) [6]. Weather generators are often based on empirical statistical relationships that maintain the autocorrelation and correlation properties of the various variables. According to [7], the simulated time scales range from daily to annual periods. However, one of the drawbacks of daily weather generators is that they underestimate monthly and inter-annual variances due to the lack of consideration in estimating low-frequency component of climate variability. Authors in [8] used a monthly generator (based on first-order autoregressive model) to adjust the low frequency capability based on the daily WGEN model. Even though the results are well simulated, this model is not able to capture the inter-annual variability. In [9], authors introduced a method for pairing of two different time scales modeled stochastic hydrological time series model. Two resembling time series were produced where one preserves important statistical properties on a finer time scale while the other is on a coarser scale of time. The adjustment is then made on a series of finer time scales so that the series is consistent with a series of coarser time scales. The results show that the coupling method can produce a series of daily rainy days which preserves some important statistical properties on daily, monthly and yearly scales. Other studies of weather generator (e.g. [10-13]) were conducted to address daily weather generator related problems. Authors in [14] compared the performance of different stochastic weather generators for long term climate data simulation. In particular, CLImate GENerator (CLIGEN), Long Ashton Research Station Weather Generator (LARS-WG), and Weather Generators (WeaGETS) were compared in terms of their ability to capture vital statistics features. The observed daily monitoring statistical features and minimum as well as maximum daily air temperatures are well simulated using both CLIGEN and LARS-WG models. These generators can also simulate maximum growth periods and increasing degree days, making them ideal for plant growth simulation. However, WeaGETS model is not quite capable of capturing the descriptive statistics, output value distributions, and evaluate extreme variables.

On the other hand, the Advanced Weather Generator (AWE-GEN) model, developed by authors in [15], has been successful in producing a broad range of temporal scale weather variables from the high-frequency hourly values to the low-frequency inter-annual variability. In Malaysia, the model

had demonstrated its ability in producing projections of extreme rainfall events for Peninsular Malaysia [16]. In this study, the AWE-GEN model is used to analyze rainfall activities in Kelantan using meteorological data from selected stations. Embedded within the AWE-GEN model is the Neyman Scott Rectangular Pulse (NSRP) model, developed in [17]. NSRP model uses parameters to denote the physical rainfall characteristics, thus it is highly crucial to appoint the correct probability distributions representing the parameters to allow reliable results to be produced. In Malaysia, several types of distributions have been tested for rainfall intensity and the results varied according to the models used. For instance, Generalized Pareto [18], Mixed Exponential [19, 20] were used. This study compares the performance of two probability distributions, Weibull and Gamma to represent rainfall intensity. The distribution with the best fit will be used to simulate hourly scaled rainfall series. Weibull is a heavy-tailed distribution which could capture the extreme rainfall events, frequently experienced in the studied region, while Gamma distribution has been successfully employed in [15] in AWE-GEN to assess rainfall in Tucson Airport, Arizona, USA.

II. DATA

Malaysia lies within the equatorial belt with high temperatures and rains all year long. Its rainfall distribution is influenced by the monsoon regime and lately, is inconsistent from year to year. There are two main monsoon winds seasons, the Southwest Monsoon (between May to September) and the Northeast Monsoon (from November to March). The east coast of Peninsular continues to be affected by the northeast monsoon which brings more rain. The studied region, Kelantan is located in the north east and is highly susceptible to monsoon flood almost on a yearly basis. In this study, the AWE-GEN model is constructed based on 30 years (1975-2005) of historical meteorological data. Data on hourly scale such as rainfall amount, temperature, relative humidity and wind speed are taken from two selected rainfall stations (see Table I). Figure 1 shows the location of the rainfall stations in this study.

III. MODEL DEVELOPMENT

Within the AWE-GEN model is the NSRP model to assess the intra-annual variability. Work by authors in [20] and [21] indicated that the Neyman Scott methodology is suitable to be used in Malaysia. Parameters describing the physical characteristics of rainfall in the NSRP model are given in Table II. Further description of NSRP model and theoretical formations could be found in [19]. As mentioned before, appointing the correct probability distributions in representing the parameters of NSRP model is critical. Hence in this study, the performance of Gamma and Weibull in representing cell intensity is compared. The Gamma distribution that is associated in NSRP is as follows,

$$P(x) = \begin{cases} \dfrac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1} e^{\left(\frac{-x}{\theta}\right)} & , \quad x > 0 \\ 0 & , \quad x \leq 0 \end{cases} \qquad (1)$$

where θ is the scale parameter $(\theta > 0)$, α is the shape parameter $(\alpha > 0)$ and x is the hourly rainfall amount.
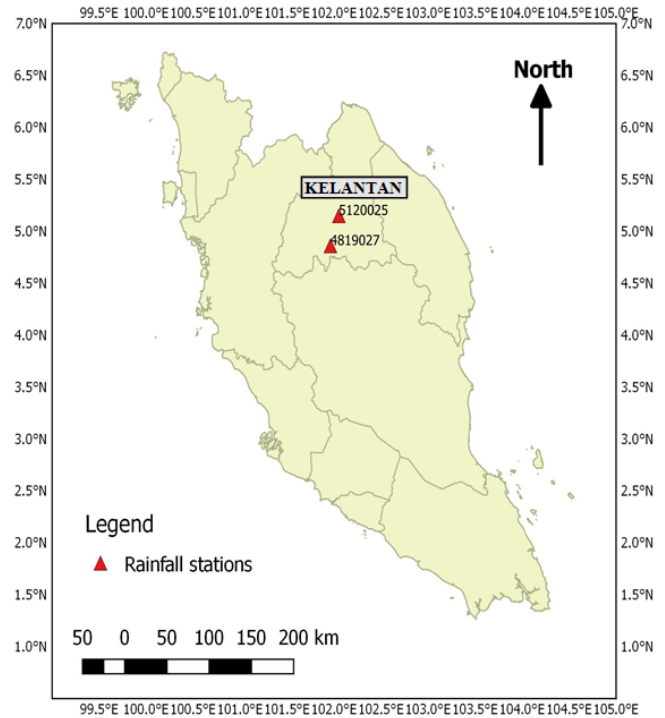

Fig. 1.   Map of Peninsular Malaysia with location of rainfall stations

TABLE I.   STATION DATA

Station ID	Name of station	Lat (°)	Lon (°)
4819027	Gua Musang, Kelantan	2.92	101.37
5120025	Balai Polis Bertam, Kelantan	3.43	101.18

Meanwhile, the Weibull distribution is as follows,

$$P(x) = \left(\frac{\alpha}{\gamma}\right)\left(\frac{x}{\gamma}\right)^{\alpha-1} e^{\left(-\left(\frac{x}{\gamma}\right)^{\alpha}\right)} \qquad (2)$$

where γ and α are the scale and shape parameters, respectively.

The first phase of this study individually fits Gamma and Weibull distributions to the data of the selected stations. Generated rainfall amount of each station are then compared with the observed data. The performance of both distributions are judged using Root Mean Square Error (RMSE),

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (3)$$

where n is the total number of data, y_i is the i-th actual rainfall amount and $\hat{y}_i$ is the simulated rainfall amount. Lower value of RMSE denotes the more efficient distribution at a particular station. The second phase of the study involves generating rainfall series using the identified distribution representing rainfall intensity at each rainfall station. In order

to validate the model, the simulated hourly rainfall was divided into two non-overlapping periods. These periods are (i) 1975 to 1989 and (ii) 1990 to 2005. The first period is the reference period where a multiplicative factor is computed based on the simulation output and the high resolution observational data. Next, the changing factors were then used to amend the biases of the simulation output from 1990 to 2005. The revised hourly rainfall is then compared to the observation from the identical period of 1975-1989.

TABLE II.   RAINFALL PARAMETERS OF THE NSRP MODEL

Param.	Meaning
λ	Mean storm origin arrivals *(h)*
β	Mean waiting time for cell origins after the origin of the storm *(h)*
η	Mean duration of the cell *(h)*
μ_c	Mean number of cell per storm
α, θ	Shape and scale parameters of rainfall intensity using Gamma
γ, θ	Shape and scale parameters of rainfall intensity using Weibull

IV.   RESULTS AND DISCUSSION

Table III shows the RMSE values for both distributions at each rainfall station. There is not much difference in the values between Gamma and Weibull, but overall, Gamma has the better fit for rainfall intensity at both stations with a lower value of RMSE. The NSRP model parameters with Gamma representing intensity are displayed in Table IV.

TABLE III.   RMSE VALUES OF GAMMA AND WEIBULL DISTRIBUTIONS

Station ID	Name of station	RMSE	
		Gamma	Weibull
4819027	Gua Musang	216*	224.7
5120025	Balai Polis Bertam	238.7*	240.0

* indicates lowest RMSE value

TABLE IV.   ESTIMATED NSRP PARAMETERS OF THE AWE-GEN MODEL WITH GAMMA

Station 4819027						
Month	λ	β	μ_c	η	α	θ
Jan	0.005	0.02	22.25	2.39	0.43	8.10
Feb	0.005	0.02	8.17	2.40	0.62	10.07
Mar	0.007	0.03	2.67	2.22	3.26	5.70
Apr	0.008	0.01	3.36	2.65	1.80	10.60
May	0.011	0.01	3.41	2.63	1.85	10.62
Jun	0.009	0.01	3.27	2.46	2.41	8.23
Jul	0.013	0.01	1.82	3.12	4.98	5.86
Aug	0.014	0.01	2.28	2.83	2.86	8.09
Sep	0.022	0.01	1.29	2.80	20.00	1.84
Oct	0.023	0.02	1.74	2.23	3.21	6.54
Nov	0.021	0.08	7.86	2.07	0.28	16.22
Dec	0.009	0.03	13.64	1.63	0.66	6.18
Station 5120025						
Jan	0.005	0.02	6.69	1.94	1.51	4.73
Feb	0.004	0.02	5.53	2.11	2.18	4.46
Mar	0.006	0.01	3.02	2.68	3.73	4.31
Apr	0.008	0.03	2.57	2.21	1.59	12.04
May	0.016	0.09	1.16	2.48	10.82	3.14
Jun	0.010	0.01	3.01	2.77	1.80	10.54
Jul	0.011	0.01	1.88	3.79	3.58	9.91
Aug	0.010	0.01	2.90	2.68	2.90	7.93
Sep	0.014	0.02	1.83	2.32	5.37	5.17
Oct	0.018	0.01	1.69	2.16	2.40	8.91
Nov	0.007	0.03	17.87	1.70	0.18	19.36
Dec	0.005	0.03	27.31	1.75	0.34	9.03

Parameter estimates for λ and β denote the estimated storm origin arrival rate and waiting time for cell origin after the storm origin, respectively. There are no significant differences in λ and β at both stations. Meanwhile, μ_c represents the mean number of cell per storm and η represents the mean duration of the cell. From the table, mean number of cells per storm, recorded higher values in November to February for both stations, indicating that rainfall activity is heavier during these months. It should be noted that this period is referred to as the northeast monsoon season which causes heavier rainfall in the eastern coast of the peninsular. Kelantan which is located at the north eastern coast is exposed to the northeast monsoon wind during this period. However, the mean duration of the cell, η recorded relatively higher values in July for both stations. It should be noted that July corresponds to the southwest monsoon which takes place from May to August. Hence, η does not give high contribution to the rainfall activity in Kelantan. Increase in shape parameter α implies higher probability occurrence of extreme and heavy rainfall [22]. The value of α for stations 4819027 and 5120025 is highest in September and May, respectively. In contrast, both stations have smaller value of α in November, implying the probability of the occurrence of rainfall extremes is small.

However, the highest value for scale parameter θ is observed in November, with 16.22 mm/h for station 4819027 and 19.36 mm/h for station 5120025, implying higher rainfall extreme intensities during this month. According to [22], inter-annual variability in the extreme rainfall is driven by the leading mode of variability in the scale parameter. Second order mode of the variability of rainfall extremes is associated with the leading mode of the variability of shape parameter. Such results are consistent with the findings of this study where the scale parameter contributes more significant to the extreme rainfall activity in Kelantan as compared to the shape parameter. AWE-GEN with Gamma is used to generate 30 years of rainfall series at hourly scale in each station. The results in the form of graphical comparison are shown in Figure 2. The mean, variance, lag-1 autocorrelation and skewness are well simulated for both stations.

Results also indicate that the frequency of non-precipitation as well as the transition probability are quite challenging for both stations to capture. Due to its stations' location at the eastern region of Peninsular Malaysia that receives higher amount of rainfall every year, the mean of dry interval is underestimated at both stations. For stations 4819027 and 5120025, it is underestimated by 0.8 and 0.7 days respectively as confirmed in Figure 3. Similarly, the mean of wet interval is also underestimated but to a lesser degree by 0.2 and 0.3 days, respectively. In tropical climate, the mean is usually preserved but the shape of the distribution can deviate from the observed. The simulated and observed extremes rainfall for time aggregation periods of 1 hour and 24 hours are illustrated in Figure 4. For both stations, there is good match between simulated and observed values, especially for return periods up to 20-30 years. Extreme values of dry intervals are poorly captured by the model. In contrast, extreme values of wet intervals are well captured by the model.
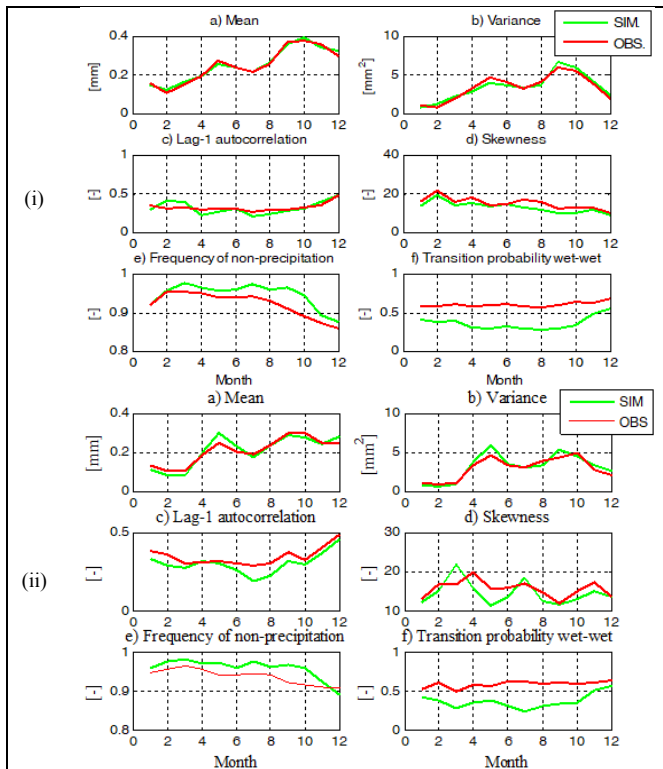
Fig. 2.   Comparison between observed (red) and simulated (green) monthly statistics of rainfall of 1 hour scale for (i) Station 2913001 and (ii) Station 3411017
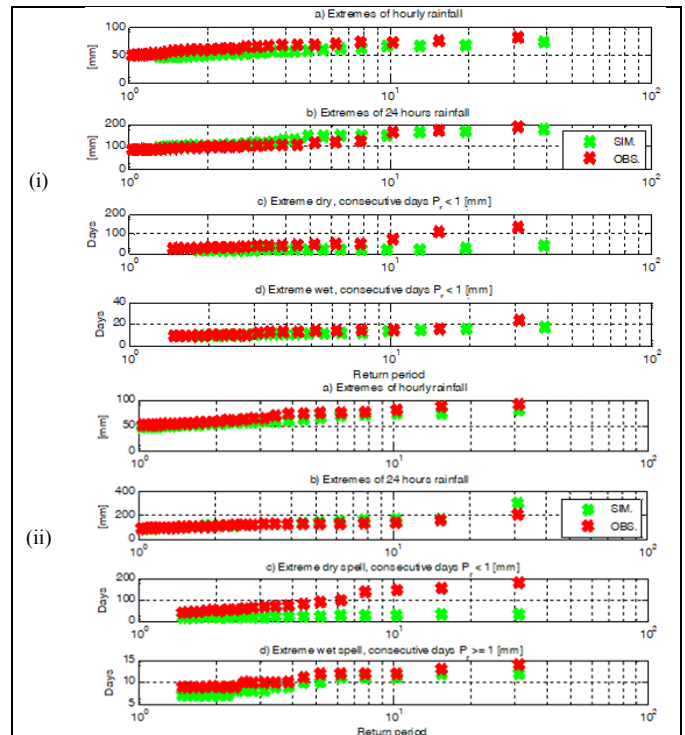

Fig. 4.   Comparison between observed (red) and simulated (green) values of extreme precipitation (green crosses) at (a) 1-hour and (b) 24-hour aggregation periods, (c) extremes of dry and (d) wet interval durations for (i) Station 4819027 and (ii) Station 5120025
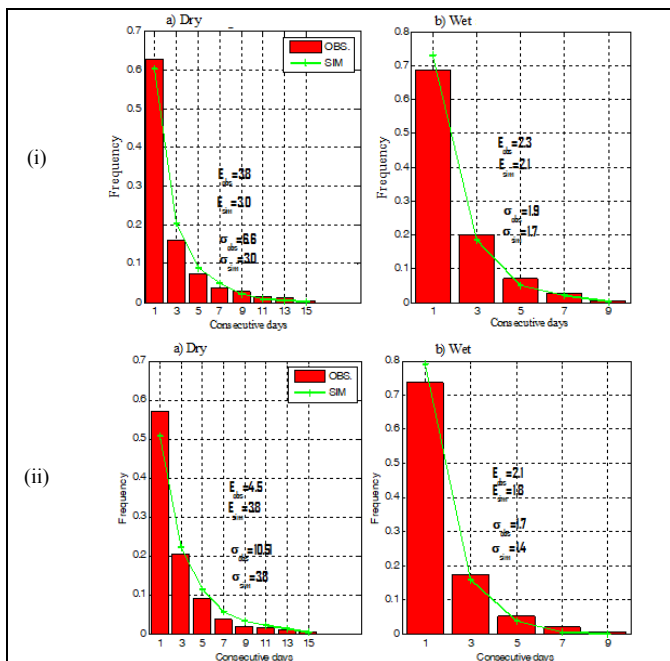
V.   CONCLUSIONS

Overall, the AWE-GEN model is capable of replicating the monthly rainfall series in Kelantan through Gamma distribution. Although Gamma is not a heavy-tailed distribution, it was found to be better in representing the rainfall intensity compared to the Weibull distribution. The estimated parameter results indicate that there are no significant differences in λ and β at both stations. Higher values of μ_c and θ suggest that there are heavy rainfall activities in Kelantan, especially during the northeast monsoon season whereas η and α do not contribute significantly to the rainfall activity. Results of this study are valuable, particularly to agricultural and storm water management planning. Specifically, this model could be beneficial in terms of handling issues of insufficiency of hydrological data especially rainfall, at remote stations.

References

[1]   D. S. Wilks, R. L. Wilby, "The Weather Generation Game: A Review of Stochastic Weather Models", Progress in Physical Geography, Vol. 23, No. 3, pp. 329-357, 1999

Fig. 3.   Comparison between observed (red) and simulated (green) of fraction of time, for dry interval length (a) consecutive days with precipitation depth lower than 1 [mm] and for wet interval length distribution (b) consecutive days with precipitation depth larger than 1 [mm] for (i) Station 4819027 and (ii) Station 5120025. E_{obs} and σ_{obs} are the observed mean and standard deviation and E_{sim} and σ_{sim} are the simulated ones.

[2] M. Kevin, R. Ramesh, O. John, A. Imran, G. Bahram, "Evaluation of weather generator ClimGen for southern Ontario", Canadian Water Resources Journal, Vol. 30, No. 4, pp. 315-330, 2005

[3] C. W. Richardson, "Stochastic simulation of daily precipitation, temperature, and solar radiation", Water Resources Research, Vol. 17, No. 1, pp. 182‑190, 1981

[4] C. W. Richardson, D. A. Wright, WGEN: A model for generating daily weather variables, US Department of Agriculture, Agricultural Research Service, ARS-8, 1984

[5] A. D. Nicks, L. J. Lane, G. A. Gander, "Weather generator", in: USDA_Water Erosion Prediction Project: Hillslope Profile and Watershed Model Documentation, NSERL Report No. 10, Chapter 2, West Lafayette, 1995

[6] M. A. Semenov, E. M. Barrow, "LARS-WG, A Stochastic Weather Generator for Use in Climate Impact Studies, User Manual", Available at: http://resources.rothamsted.ac.uk/sites/default/files/groups/mas-models/download/LARS-WG-Manual.pdf, 2002

[7] C. G. Kilsby, P. Jones, A. Burton, A. Ford, H. Fowler, C. Harpham, P. James, A. Smith, R. L. Wilby, "A Daily Weather Generator for Use in Climate Change Studies", Environmental Modelling and Software, Vol. 22, No. 12, pp. 1705–1719, 2007

[8] M. Dubrovsky, J. Buchteke, Z. Zalud, "High-frequency and low-frequency variability in stochastic daily weather generator and its effect on agricultural and hydrologic modeling", Climatic Change, Vol. 63, No. 145-179, pp. 1123-1130, 2004

[9] Q. J. Wang, R. J. Nathan, "A method for coupling daily and monthly time scales in stochastic generation of rainfall series", Journal of Hydrology, Vol. 346, No. 3-4, pp. 122‑130, 2007

[10] D. E. Keller, A weather generator for current and future climate conditions, PhD Thesis, ETH-Zurich 2015

[11] E. M. Furrer, R. W. Katz, "Improving the simulation of extreme precipitation events by stochastic weather generators", Water Resources Research, Vol. 44, No. 12, 2008

[12] J. Chen, F. P. Brissette, R. Leconte, "Uncertainty of Downscaling Method in Quantifying the Impact of Climate Change on Hydrology", Journal of Hydrology, Vol. 401, No. 3-4, pp. 190-202, 2011

[13] M. A. Semenov, "Simulation of Extreme Weather Events by a Stochastic Weather Generator", Climate Research, Vol. 35, pp. 203-212, 2008

[14] S. Mehan, T. Guo, M. W. Gitau, D. C. Flanagan, "Comparative study of different stochastic weather generators for long-term climate data simulation", Climate, Vol. 5, No. 26, 2017

[15] S. Fatichi, V. Y. Ivanov, E. Caporali, "Simulation of Future Climate Scenarios with a Weather Generator", Advances in Water Resources, Vol. 34, No. 4, pp. 448-467, 2011

[16] A. H. Syafrina, M. D. Zalina L. Juneng, "Historical trend of hourly extreme rainfall in Peninsular Malaysia", Theoretical & Applied Climatology, Vol. 120, No. 1, pp. 259–285, 2015

[17] I. Rodriguez-Iturbe, P. S. Eagleson, "Mathematical Models of Rainstorm Events in Space and Time", Water Resources Research, Vol. 23 No.1, pp. 181-190, 1987

[18] S. Dan'azumi, S. Shamsudin, A. Aris, "Modeling the distribution of rainfall intensity using hourly data", American Journal of Environmental Sciences, Vol. 6, No. 3, pp. 238-243, 2010

[19] Y. Fadhilah, M. D. Zalina, V. T. V. Nguyen, Y. Zulkifli, "Performance of Mixed Exponential and Exponential Distribution Representing Rain Cell Intensity in Neyman-Scott Rectangular Pulse (NSRP) Model", Malaysian Journal of Civil Engineering, Vol. 10, No. 1, pp. 55-72, 2007

[20] A. Norzaida, M. D. Zalina, Y. Fadhilah, "A comparative study of Mixed Exponential and Weibull distributions in a stochastic model replicating a tropical rainfall process", Theoretical & Applied Climatology, Vol. 118, pp. 597-607, 2014

[21] A. Norzaida, M. D. Zalina, Y. Fadhilah, "Application of Fourier series in managing the seasonality of convective and monsoon rainfall", Hydrological Sciences Journal, Vol. 61, No. 10, pp. 1967-1980, 2016

[22] O. Zolina, A. Kapala, C. Simmer, S. K. Gulev, "Analysis of extreme precipitation over Europe from different reanalyses: A comparative assessment", Global and Planetary Change, Vol. 44, No. 1-4, pp. 129–161, 2004