Jurnal
Teknologi

Full Paper

DEVELOPING AND CALIBRATING A STOCHASTIC RAINFALL GENERATOR MODEL FOR SIMULATING DAILY RAINFALL BY MARKOV CHAIN APPROACH

Article history
Received
15 July 2015
Received in revised form
2 August 2015
Accepted
26 August 2015

N. S. Dlamini[a*], M. K. Rowshon[a], Ujjwal Saha[b], A. Fikri[a], S. H. Lai[c], M. S. F. Mohd[d]

[a]Department of Biological and Agricultural Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
[b]Department of Civil Engineering, IIEST, Shibpure, Howrah, West Bengal, India
[c]Department of Civil Engineering, University of Malaya, 50603 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia
[d]Department of Water Resources Research Centre and Climate Change, NAHRIM, 43300 Seri Kembangan, Selangor, Malaysia

*Corresponding author
dlaminileko@gmail.com

Graphical abstract

Abstract

Rainfall is an important parameter in tropical humid regions for which paddy production systems depend. A significant portion of paddy water requirements is supplied by natural rainfall. Several studies have predicted changes in rainfall patterns and in the amount of rain that may be obtainable in future owing to climate change. There is increased concern about future water availability for an important crop such as rice. Need to develop new water management tools for sustainable production is inevitable, but such tools require long-term climate data that is credible and consistent with the time. This study concerns itself with evaluating a stochastic weather generator (WGEN) model for simulating daily rainfall series. The model is assessed using long-term historical rainfall data obtained from a rice growing irrigation schemes in Malaysia. The model is based on a first-order two-state Markov chain approach which uses two transition probabilities and random number to generate rainfall series. Selected statistical properties were computed for each station and compared against those retrieved from the model after model training and testing. The results obtained from these comparisons are quite satisfactory giving confidence about the performance and future outputs from the model. The model has shown good skill in describing the rainfall occurrence process and rainfall amounts for the area. The model will be adapted in a subsequent study for downscaling and simulating effective daily rainfall series corresponding to future climate scenarios.

Keywords: Rainfall, climate change, paddy, Markov chain, weather generator (WGEN), transition probabilities, Malaysia

1.0 INTRODUCTION

Rainfall is an important component of agricultural production that drives both rain-fed and irrigated agriculture. Long series of daily rainfall are increasingly required not only for hydrologic modeling of river catchments but also to serve as inputs in irrigation scheduling and crop growth models in the agriculture sector. It plays a significant role in designing irrigation systems and fixing crop water requirements. However, because water resources have direct link with climate, changes in climate have the possibility of altering the severity and prevalence of extreme hydrological and meteorological events including heavy storms, flooding and sustained droughts. Rainfall patterns and magnitudes are reported to be changing across the globe [1] which puts pressure on agricultural water availability for the farming sector.

Rice production is one of the important sectors in Malaysia with approximately 394,000 ha under irrigation confined on 8 granary areas designated for rice production. Rainfall is the primary source of irrigation water with more than 50% of paddy water requirements being satisfied by natural rainfall in these schemes. The natural pattern is characterized by poor distribution and small amounts of rainfall [2] with the wet season practicing supplemental irrigation while full irrigation is practiced during the dry season. And therefore, the irrigation infrastructure is not designed for full supply, but rather, a significant amount of the irrigation water is expected to be supplied by rainfall especially during the wet season [3]. Estimation of paddy water demand therefore relies mainly on the balance of many factors for which evapotranspiration and effective rainfall are key [4], as shown in equation 1. Effective rainfall is important as it can determine the size of conveyance infrastructure and can also help reduce annual pumping costs in areas where gravity-fed irrigation is not possible.

$$\mathrm{SW_j = SW_{j-1} + IR_j + RF_j - ET_j - SP_j - DR_j} \qquad (1)$$

Where, SW_j is ponding water depth in the paddy field (mm), IR_j is irrigation water supplied during the period (mm), RF_j is rainfall (mm), ET_j is crop evapotranspiration during the period (mm), SP_j is seepage during the period (mm), DR_j is drainage in the paddy field outlet (mm).

There is growing concern about future water availability in this country for an important crop such as rice where rainfall is a limiting factor in production, considering the severe strain that climate change will place on paddy and other rain-fed commercial crops. However, it is not yet clear how the likely future changes in rainfall will impact on paddy irrigation requirements. There is obviously a need to develop new tools for managing water resources for sustainable production.

One step towards evaluating this challenge is through quantitative studies on future rainfall projections for different scenarios for robust long term planning and mitigation strategies. Climate models are currently the most widely used tools for simulating the present and future climate variables. Because of their coarse resolution, climate models often require downscaling to adjust their information to be more realistic with local scale scenarios. Several downscaling techniques are available in literature. Weather generator models are increasingly receiving a lot of attention in research nowadays due to their stochastic nature and also because of their inherent ability to generate data on a (daily) time scale that is often required by many impact models [5]. But more importantly, they are cheaper and easy to implement and do not require lot of data input to calibrate [6-9].

With the increasing strain in water demand, paddy farming is under pressure to continue providing rice food despite the looming changes in climate. Therefore to sustain itself, there is need to develop new and robust water management tools, but such tools rely on long term hydrological and climate data that is credible and consistent with the time. Previous rainfall studies in the country have only focused on a range of other important rainfall issues including; rainfall spatial trends [10], assessment of extreme rainfall events [11], characterizing rainfall using drought indices [12], wet and dry spells spatial distribution [13]. These studies need to be extended in light of projected rainfall scenarios.

Thus, the purpose of this study is to apply a stochastic rainfall model in one of the rice granary schemes for simulation of daily rainfall for the rice scheme. Once validated successfully, the model will be used as a component in a subsequent study of developing a climate-smart water management tool for modeling water allocation and irrigation schedules in paddy fields. Section 2 of this paper describes the study area and the methods used in the study. Results are presented and discussed in Section 3 and lastly, Section 4 makes a conclusion of the study and draws recommendations.

2.0 EXPERIMENTAL

2.1 Study Area

Tanjung Karang Irrigation Scheme (TAKRIS) (shown on the graphic abstract) is located in Northwest of the State of Selangor occupying about 20,000 ha of irrigable paddy area. Main source of irrigation water supply is conveyed from the Bernam River through a 15 km Feeder Canal into the scheme's tertiary network canal system. The scheme experiences a tropical climate with long dry spells occurring during the Dry season, and heavy rains during the Wet season. According to the rainfall stations records, the scheme receives in the range of 1450 to 1800 mm of rainfall annually. Water rationing cases are more prevalent during the dry season which has led the scheme suffer frequent water shortages in the past. Therefore, knowledge of how rainfall is likely to evolve in future is undoubtedly essential to assist in minimizing drought impacts.

2.2 Model Description

The model used in this study is the modified version of Weather Generator (WGEN-type) stochastic rainfall generator developed by Richardson [14]. The model generates long series of daily rainfall *by* analyzing statistical parameters of the observed data from stations and then by using a uniform random number, to simulate rainfall series for current and future period. In the WGEN model, rainfall occurrence and rainfall amount is modeled separately.

2.2.1 Rainfall Occurrence

Daily rainfall occurrence is modeled using the Markov chain approach. Markov chain can be viewed as a way of describing the chance of an event whose occurrence is dependent only on the state attained in the previous event. This study has adopted a two-state first-order Markov chain to simulate rainfall occurrence. This model is characterized by the two transition probabilities; P_{dw} the probability of a rainy day preceded by a dry day and P_{ww} the probability of a rainy day preceded by a rainy day, expressed by equations 2 and 3 respectively.

$$P_{dw} = \mathrm{P}\{\text{wet on day t} \mid \text{dry on day t} - 1\} \qquad (2)$$

$$P_{ww} = \mathrm{P}\{\text{wet on day t} \mid \text{wet on day t} - 1\} \qquad (3)$$

To simulate rainfall occurrence $P_s(t)$ on day t, a random number U_t is generated using MATLAB program and is compared with the critical transition probability (equation 4) which depends on rainfall state of the previous day t-*1* where rain day = *w* and dry day = *d*. A flowchart describing this process is shown in Figure 1.

$$P_c = \begin{cases} P_{dw} \ if \ P_s(t-1) = d \\ P_{ww} \ if \ P_s(t-1) = w \end{cases} \qquad (4)$$

A rainy day is simulated when the random number is less than the critical probability (equation 5), otherwise it is simulated as a dry day.

$$P_s(t) = \begin{cases} w \ if \ U_t \ \le P_c \\ d \ if \ U_t \ > \ P_c \end{cases} \qquad (5)$$

The other two complementary probabilities for dry day following a dry day and dry day following a wet day are related by equations 6 and 7 respectively.

$$\mathrm{P_{dd}} = 1 - \mathrm{P_{dw}} \qquad (6)$$

$$\mathrm{P_{wd}} = 1 - \mathrm{P_{ww}} \qquad (7)$$

The transition probabilities can be estimated from the series of daily observed rainfall throughout the period as in the table below. Transitional probabilities are assumed to be constant for each month although vary from month to month [14].

		Current day		Total
Previous day		Dry (d)	Rain (w)	
	Dry (d	N_{dd}	N_{dw}	N_d
	Wet (w)	N_{wd}	N_{ww}	N_w

Where;
N_{dd} = No. dry days followed by dry days
N_{dw} = Number of rain days followed by dry days
N_{wd} = Number of dry days followed by rain days
N_{ww} = Number of rain days followed by rain days
N_d = Number of total dry days *(N_{dd} + N_{dw})*
N_w = Number of total rain days *(N_{wd}* + N_{ww})

The maximum likelihood estimation is used to estimate the transition probabilities for all the 12 months using the following equation;

$$\mathrm{P_{ij}} = \frac{\mathrm{N_{ij}}}{\mathrm{N_t}} \text{ for i, j} = 1{,}2 \qquad (6)$$

Where
$N_t = N_d$ when *i* = d
= N_r when *i* = w

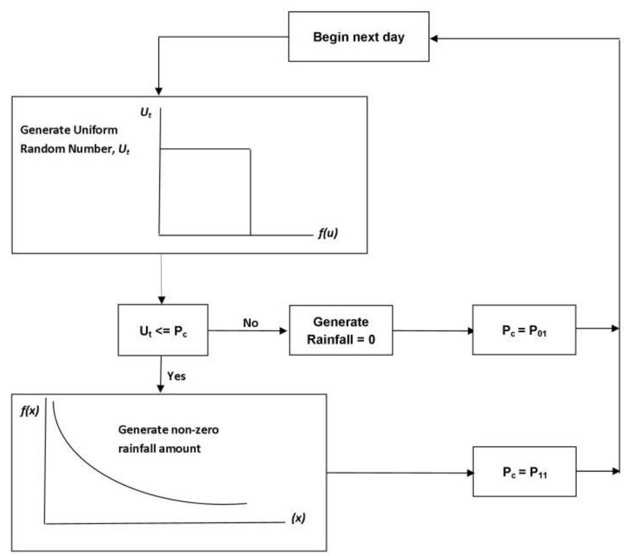

Figure 1 Flowchart for daily rainfall generation using Markov chain generators

2.2.2 Modeling Rainfall Amount

After generating the rainfall states in terms of wet or dry, the model generates rainfall amounts for each day that has been modeled as wet by fitting a probability distribution to all days with rainfall amounts. A threshold of 1 mm was assumed as wet day in this study in order to remove trace rainfall amounts. Selection of threshold value is not fixed but depends largely on the climate and the need of each particular study area [15], most studies select between 0.1 up to 5 mm depending on the purpose. In Malaysia a value of 1 mm has been widely used by previous studies due to the high humidity condition in the country [11, 16].

Rainfall amount is characterized by fitting a probability density function that best describes rainfall amounts. Several distributions have been proposed by other researchers. In Malaysia, Jamaludin and Jemain [17] found that the Mixed Exponential is appropriate for describing the daily rainfall amount. In this study, the gamma distribution (equation 7) was adopted because it is most popular choice in rainfall studies, while the former has been used comparatively rarely.

$$f(x) = \frac{\left({}^{x}/_{\beta}\right)^{\alpha-1} exp\left({}^{-x}/_{\beta}\right)}{\beta\Gamma(\alpha)}; \alpha, \beta > 0; x > 0 \qquad (7)$$

where α is a shape parameter, β is a scale parameters, and $\Gamma(a)$ represents the gamma function. The maximum likelihood estimators are used to estimate the gamma parameters [18].

2.3 Model Evaluation Criteria

Before applying the model for downscaling and future simulation, its performance in simulating rainfall series at the study area is evaluated. WGEN, being a site-specific model requires calibration and verification at every station. In this study, daily rainfall data set from 3 stations (shown on Table 1) obtained from the Department of Irrigation and Drainage, in Ampang for the 30 year period (1976 to 2006) was used for training and testing the model. The data was provided to the model to compute model parameters from generated daily series by running the model 100 times. Generated rainfall statistics describing rainfall occurrence, quantity and distribution (including monthly mean rainfall, rainfall standard deviation, rainy days, wet and dry spells and annual maximum rainfall) were computed from all 3 stations and compared against those retrieved from the model. The changes were consistent to all 3 stations.

The results (presented in Section 3) obtained from these comparisons are quite acceptable giving reasonable confidence about the performance and future outputs from the model.

Table 1 List of rainfall stations for the study area

Station	Latitude	Longitude	% missing value
3610014	03° 37' 16''	101° 02' 28''	12.6
3609012	03° 40' 54''	100° 59' 31''	4.4
3710006	03° 43' 43''	100° 04' 59''	9.8

2.4 Data Analysis

In this study analysis were done using MATLAB program. MATLAB script or program was developed for loading data, simulation of dry and wet days, calculation of gamma parameters, generating rainfall amounts based on single gamma distribution, and simulation of monthly statistical properties. All graphs were produced from the MATLAB program.

3.0 RESULTS AND DISCUSSION

This section presents results of model performance for simulation of observed rainfall series. For illustration, only selected results from station no.3609012 are presented in this paper. Results are presented in terms of the rice growing seasons, that is, Wet and Dry seasons.

3.1 Transitional Probabilities and Gamma Parameters

Estimated transition probabilities for the observed period are shown in Table 1. It can be observed that during the Dry season (February to August), there is a high probability of no rain in a day if the previous day is dry. Similarly, the probability of a wet day increases if the previous day is wet during the Wet season (September to January). Estimated parameters of Gamma distribution fitted for the wet days are shown in Table 2. The derived model parameters were used to generate new rainfall series.

3.2 Statistical Parameters (Mean and Standard Deviation)

Derived rainfall parameters were provided in the model to generate synthetic data and compared with observed. The results obtained in these comparisons are presented in Figure 2(a). As it can be observed, the daily mean rainfall from month to month is accurately duplicated by the model with R-squared value of 0.99. In terms of standard deviation comparison, the model shows good performance skill, although there is a slight underestimate throughout all the months. The month of October and November were more underestimated compared to the other months, but overall it gives acceptable results with R-squared value of 0.96. It is generally acknowledged that most models are poor in modeling rainfall variance (Wilby *et al*. (2004). The results show close resemblance with the rainfall pattern in Malaysia where most of the rain fall during the northeast monsoon season also known as the main (wet) season in the irrigation sector. Lower rainfall are during the dry season (commonly referred to as off-season) which occurs during the southwest monsoon season accompanied by low river flows which usually trigger water shortages and rationing in some cases.

3.3 Mean Number of Wet and Dry Spells Lengths

The duration of wet and dry spell series for the site was reasonably simulated with R-squared values of 0.70 and 0.68 respectively as shown by Figure 2(b). The model failed to simulate the peak wet spell for October and that of January month. However, the pattern of wet spell during both seasons was reproduced more precisely for almost all months than the wet spell values. The model showed that higher wet days occur during the main cropping season (August through to

December sometimes overlapping to January), and that the dry cropping season (February July) is occupied by less wet days. Modeling of dry spell length also followed a similar trend as shown by Figure 2(b). The model consistently overestimated dry spell length throughout both seasons except in April and May. However, it underestimated slightly the month (June) which has the highest dry spell length. The model has shown skill in simulating the pattern of dry spell length during both seasons. As can be expected, there are few dry days during the main (wet) rice growing season characterized by the northeast monsoon rainfall and more dry days are experienced during the off-season rice growing period.

Table 2 Summary of estimated transition probabilities for different months for the observed period

January			**February**			**March**		
	Dry day	Rain day		Dry day	Rain day		Dry day	Rain day
Dry day	0.80	0.20	Dry day	0.80	0.20	Dry day	0.80	0.20
Rain day	0.45	0.55	Rain day	0.67	0.33	Rain day	0.60	0.40
April			May			June		
	Dry day	Rain day		Dry day	Rain day		Dry day	Rain day
Dry day	0.75	0.25	Dry day	0.81	0.19	Dry day	0.86	0.14
Rain day	0.62	0.38	Rain day	0.69	0.31	Rain day	0.75	0.25
July			August			September		
	Dry day	Rain day		Dry day	Rain day		Dry day	Rain day
Dry day	0.78	0.22	Dry day	0.82	0.18	Dry day	0.72	0.28
Rain day	0.66	0.34	Rain day	0.73	0.27	Rain day	0.58	0.42
October			November			December		
	Dry day	Rain day		Dry day	Rain day		Dry day	Rain day
Dry day	0.67	0.33	Dry day	0.67	0.33	Dry day	0.76	0.24
Rain day	0.54	0.46	Rain day	0.52	0.48	Rain day	0.57	0.43

Table 3 Estimated parameters of the gamma distributions for the observed period

Gamma parameters	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
α	0.96	0.83	0.85	0.91	0.86	0.90	0.73	0.84	0.91	0.97	0.97	1.11
β	13.09	19.93	16.94	16.75	17.42	16.34	20.6	19.28	16.54	17.46	17.04	14.27

3.4 Annual Maximum Rainfall

Figure 2(c) shows a comparison of annual maximum rainfall for the 30-year period. The model was unable to simulate well the peak annual maximum rainfall as it underestimated rainfall values throughout the study period. However, was able to show the rainfall pattern well. The maximum rainfall was recorded in 2005 while the previous year recorded the least rainfall during the period.

3.5 Rainy Days

Figure 2(d) shows comparison of observed rainy days with simulated rainy days for the study period. The model was able to reproduce the rainy days throughout the months of 30 year period with an R-squared value of 0.99. The analysis indicates that during the study period, October and November have received rain for more days (350 days) than any other month, while June received the least rain (150 days) throughout the period.

4.0 CONCLUSION

In this study, a stochastic rainfall generator model was evaluated for its skill in simulating observed rainfall using historical long-term rainfall data at three stations for Tanjung Karang Irrigation Scheme in Malaysia. The evaluation was based on the capability of the model in reproducing selected statistical parameters of the observed data from each station. These were compared against those derived after model run. The analysis indicates that the first-order two-state Markov chain model is suitable in describing the rainfall occurrence process in the study area, although there was no comparison with other similar stochastic models. Also, the fitting of the gamma distribution for rainfall amounts produced well matching rainfall amounts for the wet days, implying that, although this type of distribution has not been used in Malaysia for stochastic rainfall generation, it is well suited for the study area.

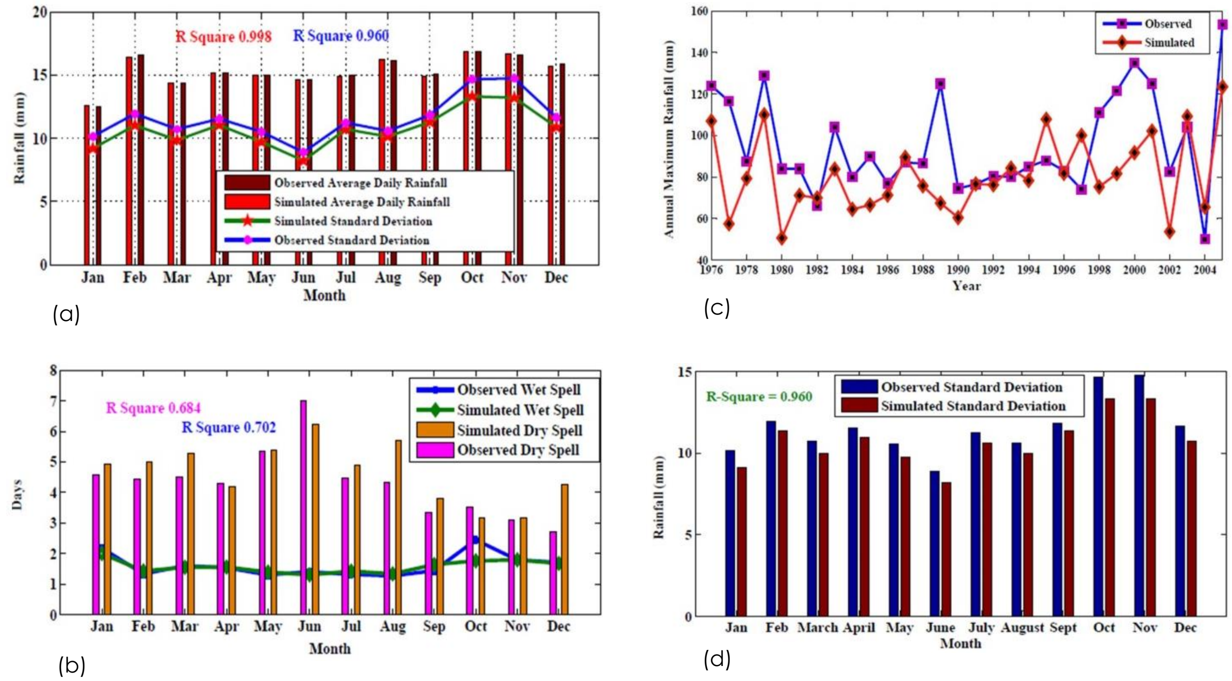

Figure 2 Comparison of rainfall statistics at station no.3609012: (a) mean daily rainfall and standard deviation, (b) wet and dry spell lengths, (c) annual maximum rainfall, and (d) rainy days

Based on the results obtained, it can be concluded that the model has good predictive skill and can thus be used for downscaling and simulating future rainfall state by pertubing model parameters for future period. However, its limitation is the inability to model extreme events since it is limited to 2 states and one order. Further study could be carried out by considering 3 states to capture future extreme rainfall and drought events.

Acknowledgement

The authors wish to thank the Ministry of Higher Education and Universiti Putra Malaysia for sponsoring this work. This research is funded under grant number FRGS2013-2. Sincere appreciation is also extended to the reviewers for their valuable critics and suggestions in improving this manuscript.

References

[1] Intergovernmental Panel on Climate Change (IPCC). 2007: *The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report Of The Intergovernmental Panel on Climate Change*. *Cambridge*: Cambridge University Press.

[2] Lee, T. S., Haque, M. A. and Najim, M. 2005. Scheduling the Cropping Calendar in Wet-Seeded Rice Schemes in Malaysia. *Agricultural water Management*. 71(1): 71-84.

[3] DID. 1996. Detailed Study of Water Resources Availability for Northwest Selangor Integrated Agricultural Development Project. Department of Irrigation and Drainage, Kuala Lumpur Office, Malaysia.

[4] Jang, Min-Won, Choi, Jin-Yong, and Lee, Jeong-Jae. 2007. A Spatial Reasoning Approach to Estimating Paddy Rice Water Demand in Hwanghaenam-Do, North Korea. *Agricultural Water Management.* 89(3): 185-198.

[5] Caron, Annie, Leconte, Robert, and Brissette, François. 2008. An Improved Stochastic Weather Generator for Hydrological Impact Studies. *Canadian Water Resources Journal. 33*(3): 233-256.

[6] Furrer, Eva M, and Katz, Richard W. 2008. Improving the Simulation of Extreme Precipitation Events by Stochastic Weather Generators. *Water Resources Research.* 44(12).

[7] Dian, B, Gameda, Samuel, and Hayhoe, Henry. 2008. Performance of Stochastic Weather Generators LARS-WG and AAFC-WG for Reproducing Daily Extremes of Diverse Canadian Climates. *Climate Research.*

[8] Semenov, Mikhail A. 2007. Simulation of Extreme Weather Events by a Stochastic Weather Generator. *Climate Research.* 35(3): 203.

[9] Wilks, Daniel S, and Wilby, Robert L. 1999. The Weather Generation Game: A Review of Stochastic Weather Models. *Progress in Physical Geography.* 23(3): 329-357.

[10] Deni, Sayang Mohd, Suhaila, Jamaludin, Zin, Wan Zawiah Wan, and Jemain, Abdul Aziz. 2010. Spatial Trends of Dry Spells Over Peninsular Malaysia During Monsoon Seasons. *Theoretical and Applied Climatology.* 99(3-4): 357-371.

[11] Zin, W. Z. W., *et al*. 2010. Recent Changes in Extreme Rainfall Events in Peninsular Malaysia: 1971–2005. *Theoretical and Applied Climatology*. 99(3-4): 303-314.

[12] Yusof, Fadhilah, Hui-Mean, Foo, Suhaila, Jamaludin, Yusop, Zulkifli, and Ching-Yee, Kong. 2014. Rainfall Characterisation by Application of Standardised Precipitation Index (SPI) in Peninsular Malaysia. *Theoretical and Applied Climatology.* 115(3-4): 503-516.

[13] Deni, Sayang Mohd, and Jemain, Abdul Aziz. 2009. Fitting the Distribution of Dry and Wet Spells with Alternative Probability Models. *Meteorology and Atmospheric Physics.* 104(1-2): 13-27.

[14] Richardson, Clarence W. 1981. Stochastic Simulation of Daily Precipitation, Temperature, and Solar Radiation. *Water Resources Research.* 17(1): 182-190.

[15] Tolika K, Maheras P. 2005. Spatial and Temporal Characteristics of Wet Spells in Greece. *Theor Appl Climatol* 81(1-2): 71-85.

[16] Deni, Sayang Mohd, Jemain, Abdul Aziz, and Ibrahim, Kamarulzaman. 2010. The Best Probability Models for Dry and Wet Spells in Peninsular Malaysia During Monsoon Seasons. *International Journal of Climatology*. 30(8): 1194-1205.

[17] Jamaludin, Suhaila, and Jemain, Abdul Aziz. 2007. Fitting the Statistical Distributions to the Daily Rainfall Amount in Peninsular Malaysia. *Jurnal Teknologi.* 46(C): 33-48.

[18] Haan, Charles Thomas. 1977. *Statistical Methods in Hydrology.*