Stoch Environ Res Risk Assess
DOI 10.1007/s00477-016-1373-9

ORIGINAL PAPER

Stochastic modelling of seasonal and yearly rainfalls with low-frequency variability

Jing Lin Ng[1] · Samsuzana Abd Aziz[1] · Yuk Feng Huang[2] · Aimrun Wayayok[1] · M. K. Rowshon[1]

Abstract Stochastic rainfall models are important for many hydrological applications due to their appealing ability to simulate synthetic series that resemble the statistical characteristics of the observed series for a location of interest. However, an important limitation of stochastic rainfall models is their inability to preserve the low-frequency variability of rainfall. Accordingly, this study presents a simple yet efficient stochastic rainfall model for a tropical area that attempts to incorporate seasonal and inter-annual variabilities in simulations. The performance of the proposed stochastic rainfall model, the tropical climate rainfall generator (TCRG), was compared with a stochastic multivariable weather generator (MV-WG) in various aspects. Both models were applied on 17 rainfall stations at the Kelantan River Basin, Malaysia, with tropical climate. The validations were carried out on seasonal (monsoon and inter-monsoon) and annual basis. The third-order Markov chain of the TCRG was found to perform better in simulating the rainfall occurrence and preserving the low-frequency variability of the wet spells. The log-normal distribution of the TCRG was consistently better in modelling the rainfall amounts. Both models tend to underestimate the skewness and kurtosis coefficient of the rainfall. The spectral correction approach adopted in the TCRG successfully preserved the seasonal and inter-annual variabilities of rainfall amounts, whereas the MV-WG tends to underestimate the variability bias of rainfall amounts. Overall, the TCRG performed reasonably well in the Kelantan River Basin, as it can represent the key statistics of rainfall occurrence and amounts successfully, as well as the low-frequency variability.

Keywords Stochastic rainfall model · Low-frequency variability · TCRG · MV-WG · Tropical climate

1 Introduction

Stochastic rainfall generators can generate long synthetic rainfall series which are statistically similar to the observed weather sequences. The time series of complete and good quality of rainfall data which can represent proper persistence is very important in hydrological simulations and ensures the reliability of the hydrological results. However, long time series of rainfall data are often unavailable due to missing data, random errors and measurement errors. Stochastic rainfall generators are particularly useful to provide long and complete synthetic time series with high computational efficiency under limited availability of the observed data. Therefore, they are ideal for risk analysis, agricultural decision making and climate change studies (Wheater et al. 2005; Cantet et al. 2011; Hauser and Demirov 2013; Langousis and Kaleris 2014; Huang et al. 2015; Hong et al. 2016).

The stochastic generation of rainfall usually involves the generation of rainfall occurrence and rainfall amount. The Markov chain model is used to simulate the rainfall occurrence where the next sequences of wet and dry days are conditioned on the basis of the current state of a model system (Srikanthan et al. 2005; Kim et al. 2008; Sonnadara

✉ Jing Lin Ng
jinglin.ng787@gmail.com

[1] Department of Biological and Agricultural Engineering, Faculty of Engineering, Universiti Putra Malaysia (UPM), 43400 Serdang, Selangor, Malaysia

[2] Department of Civil Engineering, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Kuala Lumpur, Malaysia

Published online: 30 December 2016

and Jayewardene 2014; Mandal et al. 2014). Numerous studies have reported that the Markov chain model performed reasonably well in representing the occurrence characteristics of rainfall especially in temperate climates (Harrison and Waylen 2000; Chen et al. 2012b). This is due to the straightforward structure of the Markov chain which eases the process of building the probability models for fitting the dry and wet spells. However, the short memory of the Markov chain may limit the reproduction of persistence events, such as droughts and floods. Alternatively, rainfall occurrence can be modelled using alternating renewal model which treats rainfall as a sequence of alternating wet and dry states (Racsko et al. 1991; Semenov et al. 1998; Bernardara et al. 2007; Paschalis et al. 2014). This model can provide direct consideration of the length of wet and dry series, but it will create an issue if the independence between wet and dry spells is not assumed. Next, the rainfall amount is generated by fitting various distribution models, such as gamma (Hasan and Dunn 2010), log-normal (Sharma and Singh 2010), Weibull (Mandal et al. 2014) and mixed exponential(Liu et al. 2011) distributions, to the historical data. Following these works, various stochastic weather generators exhibiting different features had been proposed over the years, such as the LARS-WG (Semenov and Brooks 1999); ClimGen (Castellvi et al. 2002) and WeaGETS (Chen et al. 2012a).

As noted earlier, while stochastic weather generators are well known for their ability in modelling rainfall, they are often found to underestimate the low-frequency components (variance) of rainfall (Katz and Parlange 1998; Wilks 1999; Dubrovský et al. 2004; Chen et al. 2009; Serinaldi and Kilsby 2014). Several researchers have viewed this drawback as an overdispersion phenomenon due to the inadequacy of high-frequency variation rainfall models and the failure of taking into consideration the aspect of low-frequency components (Katz and Parlange 1998; Mehrotra and Sharma 2007; Breinl et al. 2013; So et al. 2015). The concept of stationarity of the stochastic model often assumes that the rainfall events are not to change over time. As a matter of fact, in some cases, there are some non-stationarities present due to climate change that alters the characteristics of rainfall over time. Besides, the inter-annual climate fluctuation, known as the El-Nino–Southern Oscillation (ENSO), also affect the rainfall anomalies by inducing low-frequency climatic variations (Harrold 2003; Tangang et al. 2012). The inability of the stochastic rainfall models to represent the low-frequency components (variance) of rainfall can cause the results of the simulation process to be inaccurate and may lead to a potentially large source of model uncertainty. For this reason, it is imperative to investigate the formulations of a stochastic rainfall model to incorporate the low-frequency components of rainfall.

Various approaches have been devoted to investigate this discrepancy to correct the low-frequency variability bias (Chen et al. 2010; Khazaei et al. 2013). The conceptualization and formulation of the approaches may be comparatively complex due to the high rainfall fluctuations in terms of space and time. To preserve the inter-annual variability characteristics of rainfall, Hansen and Mavromatis (2001) proposed a disaggregated variance weather generator to perturb the monthly parameter values. A multivariate first-order Gaussian distribution was used to resample the monthly mean for the purpose of correcting the negative bias of inter-annual variability. The low-frequency correction scheme was tested at 25 locations in the continental USA. It was found that the approach efficiently discarded the negative bias and improved the inter-annual variability of monthly rainfall. However, the improvements came at the expense of overestimating the rainfall frequency which led to positive bias of rainfall. Dubrovský et al. (2004) utilized a stochastic monthly weather generator exhibiting realistic inter-monthly correlations to simulate the monthly averages of rainfall. The results showed that the synthetic series of rainfall was better simulated, with the improvement on hydrological output reliability. However, the standardization of monthly intensity was independent of the daily intensity. This may lead to the incorrect simulation of large monthly rainfall with short length of wet days. Chen et al. (2010) presented a power spectrum method for coupling timescales of rainfall series models. They compared the performance of their method with the coupled models of Wang and Nathan (2007). The study indicated that the spectral method could estimate the standard deviation of monthly and annual rainfall very well, and it performed better than the Wang and Nathan method which did not consider the annual timescale correction. Khazaei et al. (2013) developed a new weather generator (IWG) which generated rainfall by using the Neyman–Scott Rectangular Pulses (NSRP) model. The rainfall output was compared with those of the LARS-WG weather generator. They concluded that their model outperformed LARS-WG in reproducing the low-frequency variability and rainfall extremes. However, the linkage of daily and monthly weather generator of IWG did not consider the annual scale specifically and thus may have limited effects on correcting the inter-annual variability of rainfall.

Malaysia has a tropical climate and experiences seasonal variations which are dominated by monsoon seasons. The monsoon seasons of Malaysia consist of the south-west monsoon (SWM), the north-east monsoon (NEM) and the two inter-monsoon seasons (ITM). The SWM occurs from May to August, and it brings rain to the lowlands and the west coast regions of the peninsula. It usually implies drier and warmer climate conditions as most of the regions, particularly the south-western areas, are observed to have lower frequency of wet days and amount of rainfall (Suhaila et al. 2010; Varikoden et al. 2011). There are two inter-monsoon periods, where ITM1 occurs from September to October and ITM2 occurs from March to April. Heavy rainfall in the form of convective rain is very common during these inter-monsoon periods. The NEM is the major rainy season in Malaysia and prevails between November and February. During the NEM, the east coast regions of the peninsula and western part of Sarawak received high amounts of daily rainfall (Varikoden et al. 2010; Pour et al. 2014; Mayowa et al. 2015). Generally, the intensity of rainfall varies from year to year and exhibits inter-annual variability (Sundaresan et al. 2013). The heavy and long duration rainfall associated with the NEM usually lead to seasonal floods, causing serious damage to livelihood and infrastructures (Pradhan 2010; Tehrany et al. 2015). This issue has raised considerable concern to understand and analyse the characteristics of monsoon seasonal rainfall to cope with the flood hazards and other natural disasters. The monsoon seasonal rainfall amounts and duration are the crucial factors of the flood events as they affect the infiltration, flow accumulation and rate of run-off directly. Although a considerable amount of literature has been published on stochastic rainfall studies in a tropical region (Jones and Thornton 1993, 2000, 2013; Tingem et al. 2007; Cowden et al. 2008), those studies were carried out based on the overall basis where the seasonal components are not included in the formulation and calibration process. Specifically, most of the stochastic rainfall studies in Malaysia are conducted on daily, monthly and annual basis (Shui and Haque 2004; Hassan and Harun 2013; Dlamini et al. 2015). Shui and Haque (2004) applied a daily stochastic rainfall model for irrigation projects in Terengganu, Malaysia. The model was validated for the monthly timescale, and it was found that the model gave satisfactory results. Hassan and Harun (2013) utilized a stochastic weather generator, namely the Long Ashton Research Station Weather Generator (LARS-WG), to simulate synthetic daily rainfall for the Kerian catchment, Malaysia, where the LARS-WG was only calibrated and validated for daily, monthly and yearly timescales. Dlamini et al. (2015) evaluated the capability of a stochastic rainfall generator before applying it to the Tanjung Karang Irrigation Scheme in Malaysia. The basic statistics of the rainfall output, such as the mean, standard deviation, wet and dry spells and annual maximum rainfall, were only evaluated for the monthly and yearly timescales.

Accordingly, this study aims to develop an efficient stochastic rainfall model, tropical climate rainfall generator (TCRG), which can preserve the seasonal and inter-annual variabilities of rainfall for the Kelantan River Basin, Malaysia. The two important features affecting the seasonal and inter-annual variabilities of rainfall are the variance of the number of wet days and the variance of wet day rainfall amounts (Wilks and Wilby 1999; Schoof et al. 2005; Li et al. 2013). The spectral correction method proposed by Chen et al. (2010) was applied to the rainfall data to correct the low-frequency variability bias on the seasonal and annual basis. The output of stochastic rainfall model was validated and compared against the synthetic rainfall series generated by the new multivariable weather generator (MV-WG) (Fodor et al. 2010). The rainfall generation approaches of MV-WG had shown to perform very well in reproducing the characteristics of rainfall (annual, seasonal and monthly scales) in a non-tropical region (Fodor et al. 2010, 2013). MV-WG reproduced the variance of annual, seasonal and monthly values very well although the MV-WG does not utilized any low-frequency correction approach. It would be interesting to compare the overall performance of both stochastic models over a tropical region.

In the remainder of the article, Sect. 2 describes the characteristics of the study area and the information of historical datasets. The framework of calibrating and validating the TCRG and MV-WG to obtain the simulated rainfall series is described in detail in Sect. 3. The results from the simulations and the discussions are presented in Sect. 4. Finally, the concluding remarks and suggestions for further studies are summarized in Sect. 5.

2 Study area and data description

The Kelantan River Basin, Malaysia, is situated at the north-eastern region of Peninsular Malaysia, and it comprises an area of 13,000 km^2. The Kelantan River is the main stream of the Kelantan State and flows northwards into the South China Sea. The major economic activities along the Kelantan River Basin are agricultural plantation, fishing and sand mining (Peck Yen and Rohasliney 2013). The Kelantan River Basin has uniformly high temperatures throughout the year, abundant rainfall, high relative humidity and light wind speeds. Generally, the Kelantan River Basin is characterized with a tropical climate with extensive monsoonal influence. It receives heavy monsoonal rainfall during the NEM which has caused the increment of the rainfall intensity and frequency of extreme rainfall events. This contributes to the seasonal flood that occurs almost every year. Therefore, it would be interesting to conduct a stochastic rainfall study by taking into account the variability of rainfall in this study area. In addition, the availability of long daily rainfall data is another reason for the selection of this study area. Considering the limited rainfall stations with long observed data in Malaysia, the length of 26–60 years of observed data in this study area is considered as long series.

The observed daily rainfall data of 17 rainfall stations were extracted from the Malaysian Meteorological Department (MMD) for the period 1953–2012 on the basis of data availability. All the rainfall stations chosen for this study are scattered over the Kelantan state, as illustrated in Fig. 1. Table 1 shows the general information of all the rainfall stations used. The length of the observed rainfall data utilized in this study covers a period of at least 26 years to ensure robust assessments and reliable estimations of the model characteristics. Besides, observed data with more than 10% missing values were eliminated to prevent incorporation of any possible source of error that might result in biased estimations of the long-term rainfall data.

3 Methodology

3.1 Description of rainfall models used in the comparison

The tropical climate rainfall generator (TCRG) and multivariable weather generator (MV-WG) being compared here consist of two main stochastic components, which are the rainfall occurrence and rainfall amount models. They operated similarly by transforming pseudorandom number streams into long sequences of synthetic data which were statistically similar to the observed data. Although they are stochastic-based rainfall models, it is important to highlight the main difference of the simulation methods between them. For the rainfall occurrence generation procedure, the TCRG incorporates the Markov chain whereas the MV-WG adopts the serial approach. Following that, the TCRG and the MV-WG used the log-normal and the Weibull distributions, respectively, to generate the rainfall amounts. A remarkable note is that the TCRG applies the spectral correction approach (Chen et al. 2010) to correct the underestimation of the low-frequency variability of rainfall, which is a common limitation of the stochastic rainfall model. The following paragraphs describe how both models were fitted to the stochastic process to generate the rainfall occurrences and amounts at each station.

3.2 Tropical climate rainfall generator (TCRG)

3.2.1 Step 1: Generation of the rainfall occurrence

The first step to develop a stochastic rainfall model was to simulate the occurrence of daily rainfall events in the study area. A two-state, third-order Markov chain model was selected to generate the consecutive dry and wet spells based on the observed rainfall data. The third-order of Markov model was chosen after comparing with the first and second order by employing the Akaike information

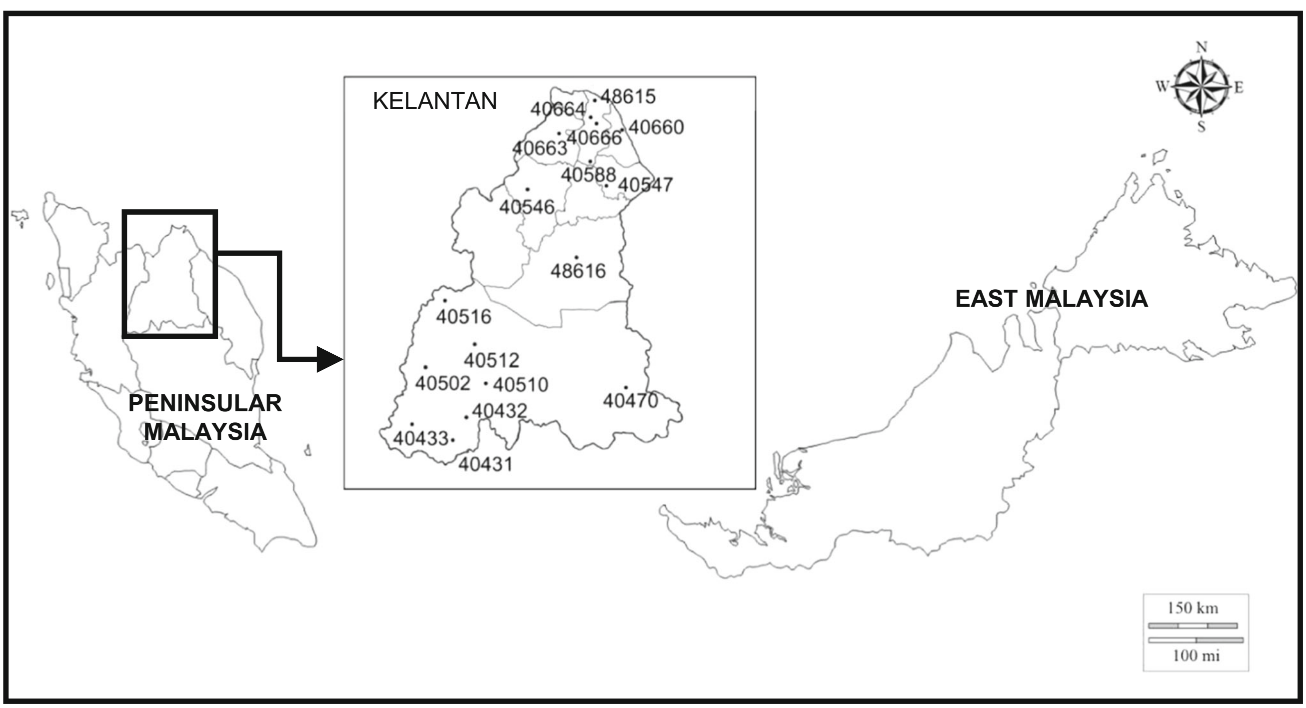

Fig. 1 Location of 17 rainfall stations in Kelantan River Basin, Malaysia

Table 1 List of rainfall stations used in this study

Station code	Station name	Record period	Duration	Latitude	Longitude
40431	Pos Blau	1979–2012	34	04°39′N	101°41′E
40432	RPS Kuala Betis	1975–2012	38	04°42′N	101°45′E
40433	Pos Hau	1979–2012	34	04°42′N	101°32′E
40470	Pos Lebir	1979–2012	34	04°56′N	102°23′E
40502	Pos Bihai	1979–2012	34	05°00′N	101°341′E
40510	Pos Tehoi	1979–2012	34	05°03′N	101°45′E
40512	Pos Wias	1979–2012	34	05°07′N	101°49′E
40516	Pos Gob	1979–2012	34	05°17′N	101°38′E
40547	Mardi Jeram Pasu	1985–2012	28	05°48′46″N	102°20′40″E
40546	Pusat Ternakan Haiwan Tanah Merah	1981–2012	32	5°48′40″N	102°00′33″E
40588	Pusat Pertanian Melor	1969–2012	44	05°58′N	102°18′E
40660	Pusat Pertanian Bachok	1975–2012	38	6°03′N	102°24′E
40663	Pusat Pertanian Pasir Mas	1977–2012	36	06°02′N	102°07′E
40664	Pusat Pertanian Lundang	1975–2012	38	06°06′N	02°14′E
40666	Mardi Kubang Keranji	1983–2012	30	06°05′N	102°17′E
48615	Kota Bharu	1953–2012	60	06°10′N	102°18′E
48616	Kuala Krai	1987–2012	26	05°32′N	102°12′E

Table 2 Number of times each Markov order is chosen as the best model for all stations by using AIC and BIC criteria

Markov order	Inter-monsoon 1		South-west monsoon		Inter-monsoon 2		North-east monsoon		Yearly	
	AE	BE	AE	BE	AE	BE	AE	BE	AE	BE
First	0	0	0	0	2	5	0	0	0	0
Second	6	**11**	2	5	7	**9**	0	7	0	0
Third	**11**	6	**15**	**12**	**8**	3	**17**	**10**	**17**	**17**

Bold values indicate the highest frequencies of the Markov order get chosen by AIC and BIC criteria

AE AIC estimates, *BE* BIC estimates

criteria (AIC) (Akaike 1974) and Bayesian information criteria (BIC) (Schwarz 1978), as shown in Table 2. The third order was found to be the optimum order for this study which gave the minimum estimates of both criteria most often and reached a balance between the least number of parameters and the maximum amount of information. For this study, the third-order Markov model was formulated based on the assumption that the probability of rain on a certain day is dependent on the rainfall state at lags of one, two and three time periods. The transitional probability can be obtained as

$$P_{hijk} = \Pr\left\{X_q = k \middle| X_{q-1} = j \middle| X_{q-2} = i \middle| X_{q-3} = h \middle|\right\} \tag{1}$$

where the h, i, and j and k are values of 0 or 1 implying dry and wet day, respectively, and X_q is the rainfall state on day q. The estimation of model parameters involved the calculation of conditional relative frequencies to generate maximum likelihood estimators (MLEs), which is defined as

$$\hat{P}_{hijk} = \frac{n_{hijk}}{\sum_{k=1}^{n} n_{hijk}} \tag{2}$$

where n_{hijk} is the observed frequency of the transition counts. The sequences of wet and dry days were computed by using a random number generator, which produced uniform random numbers that range between 0 and 1. The random variates generated were compared with the critical probability, and a wet day was predicted to occur if the random number was smaller than the corresponding transitional probability.

3.2.2 *Step 2: Generation of the rainfall amounts*

After computing the rainfall occurrence, the rainfall amounts on wet days were modelled by using the log-normal distribution. This distribution was selected, in preference to exponential, gamma, skewed normal, mixed exponential, generalized Pareto or Weibull distribution, for its adequacy based on the AIC and BIC (Table 3). The probability density function of the log-normal distribution is given by

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), \quad \sigma > 0,\ \mu > 0 \tag{3}$$

where x is the daily rainfall amounts, μ and σ denote the mean and standard deviation of the natural logarithms of rainfall amounts, respectively. The log-normal distribution was fitted to the observed rainfall data, and the method of maximum likelihood (ML) was adopted to determine the parameters. The maximum likelihood estimators of μ and σ are expressed as

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \ln x_i \tag{4}$$

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\ln x_i - \hat{\mu})^2}. \tag{5}$$

where n is the sample size. The Kolmogorov–Smirnov (K–S) tests were used to test the fit of log-normal distribution used in this study at significance level of 0.05. The null hypothesis stated that the rainfall series are drawn from the log-normal distribution. As shown in Table 4, all the computed p values were greater than the significance level, which indicated that the null hypothesis cannot be rejected. The histograms with a log-normal distribution fit of the observed seasonal and yearly data were plotted to inspect the goodness of fit of the log-normal distribution at the Pusat Pertanian Bachok station (Fig. 2). The histograms of the observed rainfall showed a good agreement with the probability density functions of the fitted log-normal distribution. The model parameters were estimated to simulate the synthetic daily rainfall series.

3.2.3 Step 3: Correction of the seasonal variability

Recognizing the fact that the synthetic rainfall series often are subjected to the underestimation of low-frequency variability, the spectral correction approach suggested by Chen et al. (2010) was used for the purpose of low-frequency correction. Previously, Chen et al. (2010) applied this approach in correcting the monthly and inter-annual variabilities. However, herein the correction procedure applied to the Kelantan River Basin was based on seasonal and annual timescales. The synthetic seasonal rainfall series consist of the south-west monsoon (SWM), north-east monsoon (NEM) and two inter-monsoon seasons (ITM) rainfall series. The low-frequency variability of synthetic seasonal rainfall data was computed using the fast Fourier transform (FFT). The main feature of the FFT is that the rainfall data are the discrete data that can be described as the sum of a number of sine functions, which is defined as

$$c = a + ib \tag{6}$$

where c is a general complex number, a and b are real numbers and i^2 equal to -1. The phase ϕ and magnitude ρ being determined as

$$\phi = \tan^{-1}\left(\frac{b}{a}\right) \tag{7}$$

$$\rho = \sqrt{a^2 + b^2} \tag{8}$$

For each seasonal rainfall series, the power spectrum was produced through FFT where the time domain components were transformed into the frequency domain components. Random phases were drawn from the output of a uniform distribution, with the range of 0–2π. The

Table 3 Number of times each probability distribution is chosen as the best model for all stations by using AIC and BIC criteria

Distribution	Inter-monsoon 1		South-west monsoon		Inter-monsoon 2		North-east monsoon		Yearly	
	AE	BE	AE	BE	AE	BE	AE	BE	AE	BE
Exponential	0	0	0	0	0	0	0	0	0	0
Gamma	4	4	0	0	0	0	5	5	1	1
Log-normal	**12**	**12**	**17**	**17**	**17**	**17**	**12**	**12**	**13**	**13**
Skewed normal	1	1	0	0	0	0	0	0	0	0
Mixed exponential	0	0	0	0	0	0	0	0	1	1
Generalized Pareto	0	0	0	0	0	0	0	0	2	2
Weibull	0	0	0	0	0	0	0	0	0	0

Bold values indicate the highest frequencies of the distribution get chosen by AIC and BIC criteria

AE AIC estimates, *BE* BIC estimates

Table 4 Results of K–S tests for the log-normal distribution in all rainfall stations

Station name	*p* values of K–S tests				
	Inter-monsoon 1	South-west monsoon	Inter-monsoon 2	North-east monsoon	Yearly
Pos Blau	0.3333	0.7430	0.8594	0.9848	0.8115
RPS Kuala Betis	0.4187	0.8511	0.7502	0.6606	0.8893
Pos Hau	0.1509	0.8178	0.8457	0.6663	0.6464
Pos Lebir	0.3918	0.6251	0.2479	0.5438	0.4326
Pos Bihai	0.3662	0.3241	0.9831	0.9963	0.9423
Pos Tehoi	0.122	0.5269	0.9838	0.4571	0.9650
Pos Wias	0.1353	0.8262	0.9685	0.8020	0.9927
Pos Gob	0.4703	0.2177	0.5425	0.9584	0.9918
Mardi Jeram Pasu	0.8277	0.9877	0.8187	0.6588	0.7724
Pusat Ternakan Haiwan Tanah Merah	0.7422	0.3505	0.6786	0.9705	0.8995
Pusat Pertanian Melor	0.8228	0.8697	0.9999	0.7078	0.6741
Pusat Pertanian Bachok	0.834	0.968	0.930	0.623	0.881
Pusat Pertanian Pasir Mas	0.7566	0.9940	0.7418	0.7244	0.8421
Pusat Pertanian Lundang	0.411	0.861	0.978	0.680	0.9586
Mardi Kubang keranji	0.9504	0.9207	0.4649	0.8393	0.6993
Kota Bahru	0.7457	0.5432	0.5148	0.5406	0.7262
Kuala Krai	0.3491	0.2902	0.5844	0.8906	0.9353

generated random phases were then assigned to each of the spectral components where the modifications were made on the phase and variance of each component. Subsequently, the new seasonal rainfall series with identical frequency spectrum can be produced and the frequency domain function was reverted back to its time domain components using inverse FFT, given in its complex form

$$c = \rho e^{i\phi} \tag{9}$$

A linear function was applied to adjust the synthetic seasonal rainfall series (before seasonal correction) to the new seasonal rainfall series. Throughout the whole correction process, the underestimation of the seasonal variances can be corrected and the seasonal variability can be preserved.

3.2.4 *Step 4: Correction of the inter-annual variability*

Inter-annual variability correction was applied to the new seasonal series generated in step 3. This step was similar to step 3 where the FFT was applied to the synthetic annual rainfall series. A new signal was produced by modifying the variance and phase of every component. An inverse FFT was then used to revert the annual times series back to its associated frequency components. As a result, the inter-annual variability of the rainfall can be preserved. However, the inter-annual spectral correction may lead to the overestimation of the seasonal variance, which was corrected in step 3. This is due to the dependence between the seasonal data and the annual data. Therefore, the seasonal correction and inter-annual correction were carried out iteratively to determine the best results for the seasonal and inter-annual variances.

3.3 Stochastic weather generator MV-WG

3.3.1 *Step 1: Generation of the rainfall occurrence*

MV-WG incorporates the serial approach in simulating the sequences of wet and dry days. The implicit assumption of the serial approach is that the occurrence of successive wet and dry spell lengths is independent of each other. Firstly, a weighted random number generator was used to determine the occurrence (wet or dry) of the first day. The weighting factor was calculated by considering the ratio of the number of the wet and dry days within the historical rainfall series. Once the wet or dry state of the first day was confirmed, the length of the given spell was generated and followed by the generation of the opposite type of a new spell. The transition to the same state is not possible; therefore, each wet spell is followed by a dry one and vice versa. Subsequently, the varying lengths of the alternating wet and dry intervals were then fitted separately by using the three-parameter Weibull distribution. The Weibull distribution with three parameters, β, η and γ, are representing the shape, scale and location parameters, respectively.

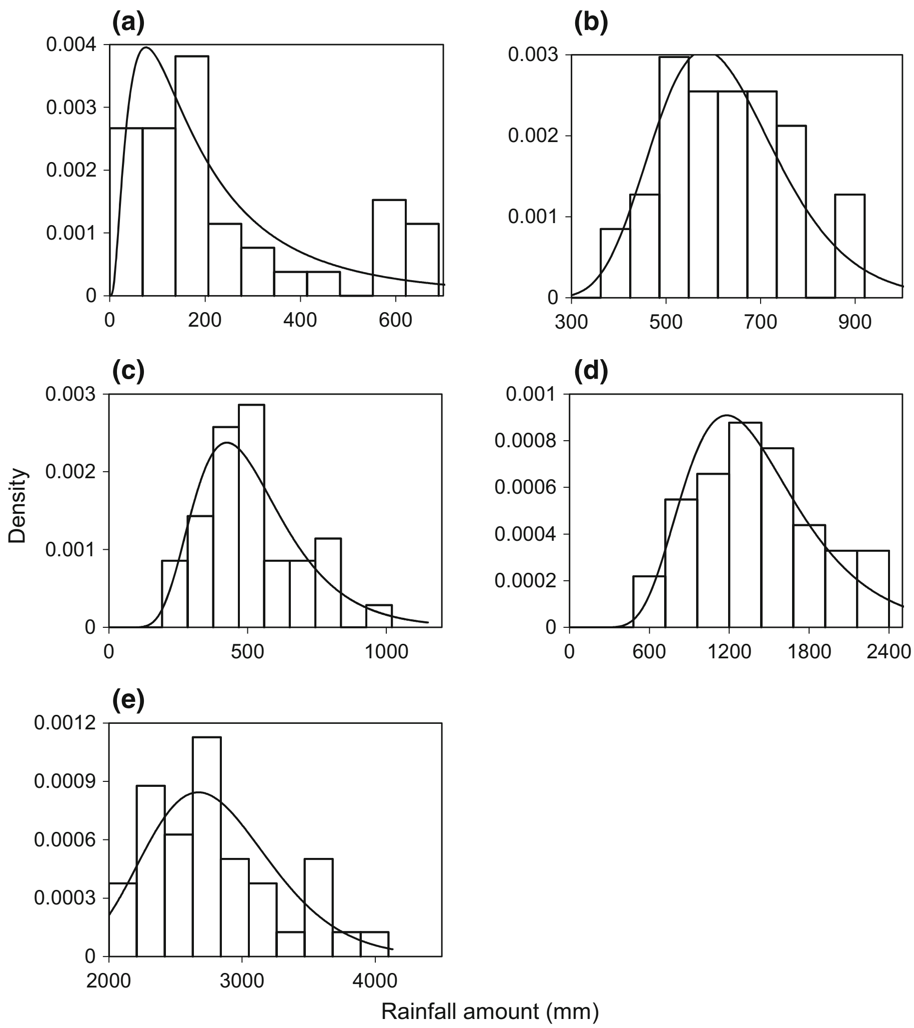

Fig. 2 Histogram and the fitted log-normal distribution of **a** inter-monsoon 1, **b** south-west monsoon, **c** inter-monsoon 2, **d** north-east monsoon and **e** yearly rainfall amounts at Pusat Pertanian Bachok station

$$f(x) = \frac{\beta}{\eta}\left(\frac{x-\gamma}{\eta}\right)^{\beta-1} \exp -\left(\frac{x-\gamma}{\eta}\right)^{\beta}, \quad \beta > 0,\ \eta > 0,\ x \geq \gamma \geq 0 \tag{10}$$

3.3.2 *Step 2: Generation of the rainfall amounts*

Given the simulated occurrence of wet days, the rainfall amounts were also generated by fitting the Weibull distribution (Eq. 10) to the observed rainfall series. As the Weibull distribution was applied here, the parameters were not estimated by nonlinear regression, but by maximum likelihood, as previous studies reported that the maximum likelihood method is able to generate smallest sampling variance of the estimated parameters (Rao and Hamed 2000). The K–S tests were used to evaluate the goodness of fit of Weibull distribution, and the results indicated that all the stations passed the tests by producing p values greater than the significance level of 0.05 (Table 5). The histograms with a Weibull distribution fit were also constructed for the Pusat Pertanian Bachok station, and it showed that the shape of Weibull distribution is suitable to model the rainfall data (Fig. 3).

3.4 The evaluation of model performance

Five hundred realizations of the synthetic rainfall of equivalent length to the historical rainfall were generated for each station in order to study the statistical properties of rainfall. The generation of such long synthetic rainfall series was needed to capture more climate events characteristics and avoid biased simulation of the true climate. The performance of both models at seasonal and yearly timescales were compared and evaluated using basic statistical metrics, such as the mean, standard deviation, skewness, kurtosis and extreme values. The synthetic

Table 5 Results of K–S tests for the Weibull distribution in all rainfall stations

Station name	*p* values of K–S tests				
	Inter-monsoon 1	South-west monsoon	Inter-monsoon 2	North-east monsoon	Yearly
Pos Blau	0.9194	0.8479	0.7770	0.9869	0.9792
RPS Kuala Betis	0.8716	0.8013	0.5467	0.4313	0.6951
Pos Hau	0.3356	0.7776	0.8191	0.7434	0.8327
Pos Lebir	0.9231	0.6330	0.2343	0.3468	0.2422
Pos Bihai	0.8462	0.6613	0.8705	0.9498	0.9701
Pos Tehoi	0.6062	0.4967	0.9716	0.5925	0.8670
Pos Wias	0.8463	0.9542	0.9988	0.8350	0.9148
Pos Gob	0.8293	0.1683	0.4270	0.7390	0.9812
Mardi Jeram Pasu	0.9756	0.9577	0.9716	0.4691	0.5262
Pusat Ternakan Haiwan Tanah Merah	0.7882	0.8905	0.7514	0.5775	0.7256
Pusat Pertanian Melor	0.9571	0.9974	0.9999	0.7801	0.9039
Pusat Pertanian Bachok	0.888	0.998	0.668	0.647	0.902
Pusat Pertanian Pasir Mas	0.9911	0.9994	0.5949	0.7182	0.8591
Pusat Pertanian Lundang	0.477	0.709	0.882	0.295	0.7440
Mardi Kubang keranji	0.9839	0.7981	0.2403	0.9568	0.6186
Kota Bahru	0.7434	0.9997	0.9209	0.9424	0.9389
Kuala Krai	0.8143	0.6366	0.2593	0.9469	0.9286

seasonal and yearly rainfall series were obtained by summing up the daily rainfall generated by the TCRG and MC-WG. The seasonal rainfall series consists of the south-west monsoon (SWM), north-east monsoon (NEM) and two inter-monsoon seasons (ITM) rainfall series. The observed and synthetic probability distributions were compared by computing the 10th, 30th, 50th, 70th and 90th percentiles of rainfall. The mean absolute percentage error (MAPE) (Swanson et al. 2011; Lu et al. 2015), computed as average of the absolute values of the percentage errors, was calculated. A scale to judge the accuracy of the model using MAPE developed by Lewis (1982) is shown in Table 6. The model that produced lower MAPE provides a better representation of the rainfall characteristics. The ability of each model in representing the inter-annual variability was assessed by variance overdispersion (Wilks 1999)

$$\text{Variance overdispersion} = \left(\frac{\text{Observed variance}}{\text{Generated variance}} - 1\right) \times 100\,\% \tag{11}$$

The MAPE and variance overdispersion were averaged over all stations and months. Since the frequency distribution of rainfall is usually skewed and not to follow a normal distribution, instead of parametric tests, nonparametric statistical tests were utilized. The mean, population distribution and standard deviation of the observed and generated series were compared by using the nonparametric Wilcoxon rank sum, Kolmogorov–Smirnov (K–S) and squared ranks tests, respectively. All the tests were computed at the significance level of 0.05. Furthermore, an acceptability index (AI) was used to indicate the relative efficacy of both models (Fodor et al. 2010), expressed as

$$\mathrm{AI} = 100\left(1 - \frac{\mathrm{TS}}{\mathrm{MS}}\right) \tag{12}$$

where TS refers to the total score calculated from the statistical tests and the MS refers to the maximum score. The maximum score of each test for the seasonal and yearly rainfalls are 68 (17 stations $\times$ 4 monsoon seasons) and 17 (17 stations $\times$ 1 year), respectively. The higher the AI, the better the model performance. A simple flow chart describing the methodology of this study is shown in Fig. 4.

4 Results and discussions

4.1 Rainfall occurrence (wet and dry spells characteristics)

Again, the TCRG simulated wet and dry spells using a third-order Markov chain while the MV-WG modelled them using a serial approach. Scatter plots of the mean and the extreme wet spells are shown in Fig. 5 to describe the characteristics of the rainfall occurrence. Figure 5a shows

Fig. 3 Histogram and the fitted Weibull distribution of **a** inter-monsoon 1, **b** south-west monsoon, **c** inter-monsoon 2, **d** north-east monsoon and **e** yearly rainfall amounts at Pusat Pertanian Bachok station

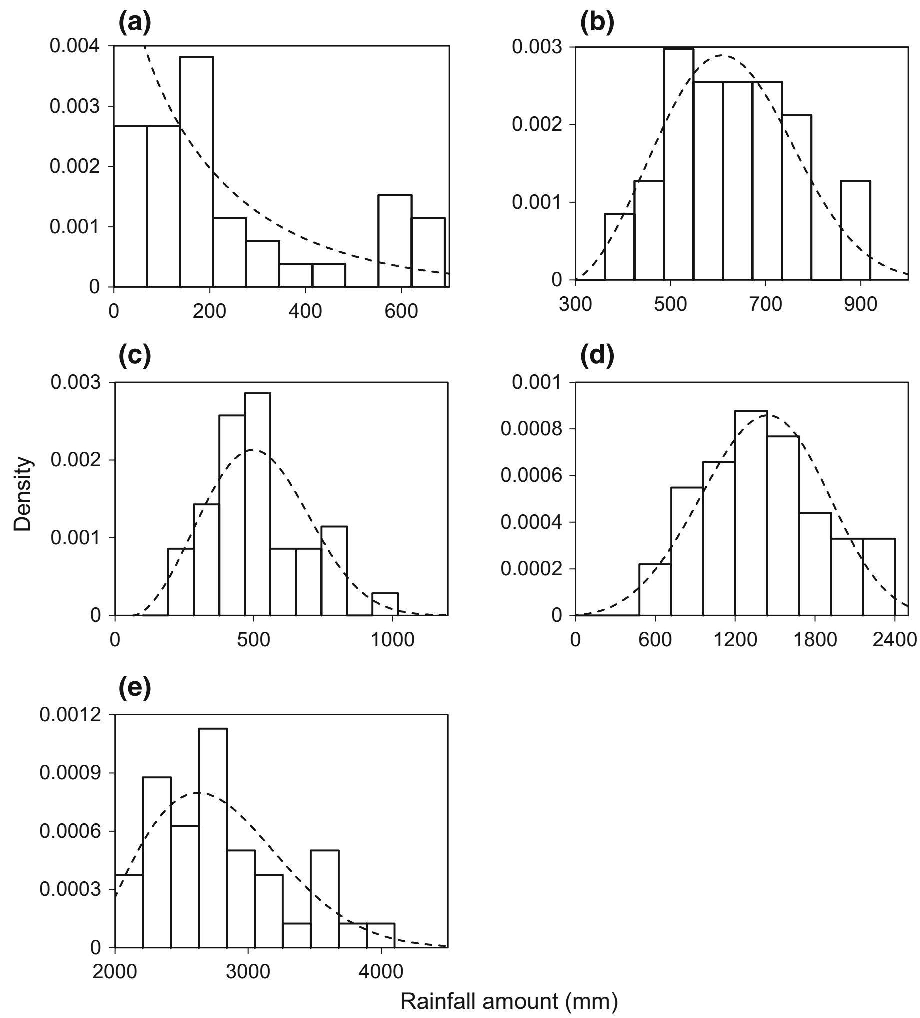

Table 6 A scale of judgement of accuracy (Lewis 1982)

MAPE	Judgement of forecast accuracy
Less than 10%	Highly accurate
11–20%	Good forecast
21–50%	Reasonable forecast
51% or more	Inaccurate forecast

that TCRG reproduced the mean of wet spells reasonably well as the data points are very near to the 1:1 line of perfect fit, whereas an underestimation can be observed from the MV-WG. The Wilcoxon rank sum tests were further used to evaluate the equality between the mean of observed and generated wet and dry spells. As illustrated in Table 7, the tests of TCRG did not show any significant differences for all stations, thus resulted in 100% acceptability index (AI) for both seasonal and yearly wet and dry spells. In comparison, the MV-WG produced 82.35 and 80.88% AI for the mean of seasonal wet and dry spells, respectively. The TCRG produced a lower mean absolute percentage error (MAPE), which were 1.04, 0.17 and 2.55 for the mean of wet spells, dry spells and wet days per month, respectively. The values were within the range of highly accurate forecast of the judgement of forecast accuracy (Tables 6, 8).

Next, the Kolmogorov–Smirnov (K–S) test was applied to test the significance difference between the observed and generated frequency distribution of the wet and dry spells. The third-order Markov chain of the TCRG obtained better results by passing most of the K–S tests, thus reproducing 98.53% (AI) for the distribution of seasonal wet and dry spells (Table 7). These values were higher than those of the serial approach, which were 67.64% for both seasonal wet and dry spells. For the yearly wet and dry spells, TCRG also achieved higher AI than the MV-WG. All the results mentioned above highlighted that the third-order Markov chain of TCRG performed better than the serial approach of MV-WG in generating the mean and distribution of the rainfall occurrence. This is consistent with previous studies where the Markov chain performed reasonably well in

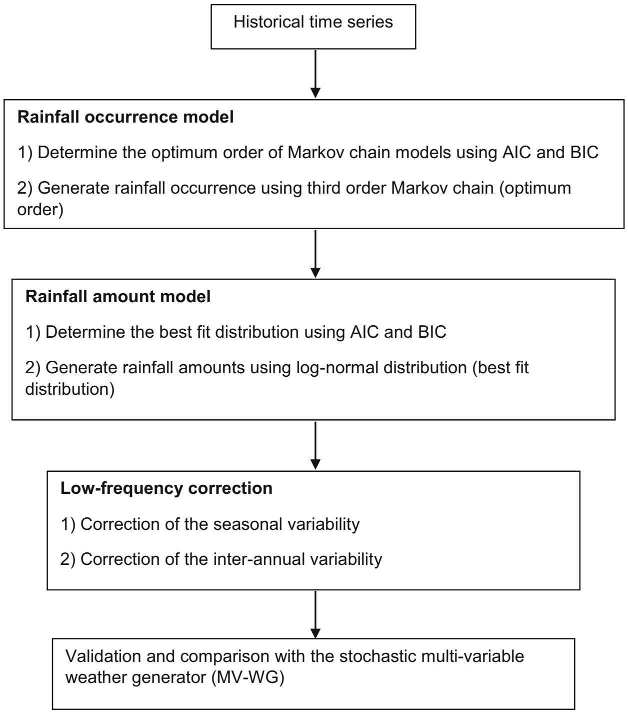

Fig. 4 Flow chart of the research methodology

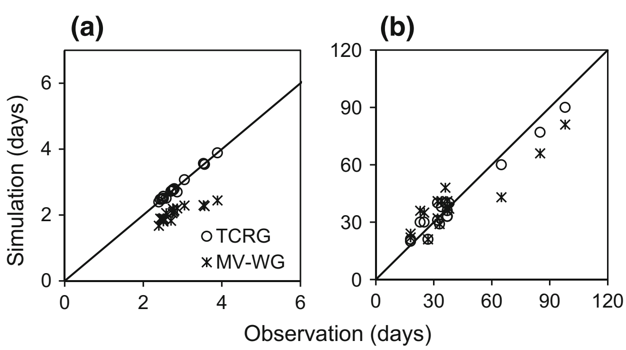

Fig. 5 Scatter plots of the **a** mean and **b** extreme values of the observed and simulated wet spells from TCRG and MV-WG for all stations

modelling the statistical characteristics of the rainfall occurrence (Sonnadara and Jayewardene 2014; Chen and Brissette 2014; Kannan and Farook 2015). The main reason for this is that the third-order Markov chain takes into consideration the weather condition of three preceding days and leads to the improvement of long-term dependency of the wet and dry spells. The MV-WG assumes independence of rainfall events; therefore, one or several dry days that occurred between the wet sequences may still correspond to the same rainfall event and subsequently lead to unsatisfactory simulation results.

Table 8 Mean absolute percentage error (MAPE) of the wet and dry spells and the wet days per month of the TCRG and MV-WG for all stations

	Mean	Std	Maximum
Wet spells			
TCRG	1.04	6.76	19.67
MV-WG	22.90	18.98	27.89
Dry spells			
TCRG	0.17	7.10	8.25
MV-WG	28.50	50.53	24.72
Wet days per month			
TCRG	2.55	15.89	
MV-WG	20.14	55.86	

Std standard deviation

Regarding the extreme values of the rainfall occurrence, the 1:1 plot of the observed versus generated series of TCRG and MV-WG were scattered (Fig. 5b). It appears that neither of them can exactly simulate the extreme rainfall characteristics. In general, both methods tend to underestimate the length of wet spells, especially when the extreme spells are large. This deficiency is hard to be eliminated as both the occurrence models are unable to fit more exactly to the observed data. However, the results of MAPE showed that the performance of TCRG was better than the MV-WG (Table 8). The corresponding MAPE of extreme wet and dry spells for TCRG were 19.67 and 8.25, respectively, which were still under the good forecast and highly accurate category. In comparison, the MAPE of extreme wet and dry spells for MV-WG were unsatisfactory with the values 27.89 and 24.72, respectively. The underestimation by TCRG and MV-WG should be

Table 7 Acceptability indices (%) for the wet and dry spells in passing the Wilcoxon rank sum, Kolmogorov–Smirnov (K–S) and squared ranks tests for all stations

	Wilcoxon rank sum tests		K–S tests		Squared ranks tests	
	TCRG	MV-WG	TCRG	MV-WG	TCRG	MV-WG
Wet spells						
Seasonal	100	82.35	98.53	67.64	98.53	77.94
Yearly	100	100	88.24	64.71	94.12	76.47
Dry spells						
Seasonal	100	80.88	98.53	67.64	98.53	77.94
Yearly	100	94.12	88.24	70.59	88.24	82.35

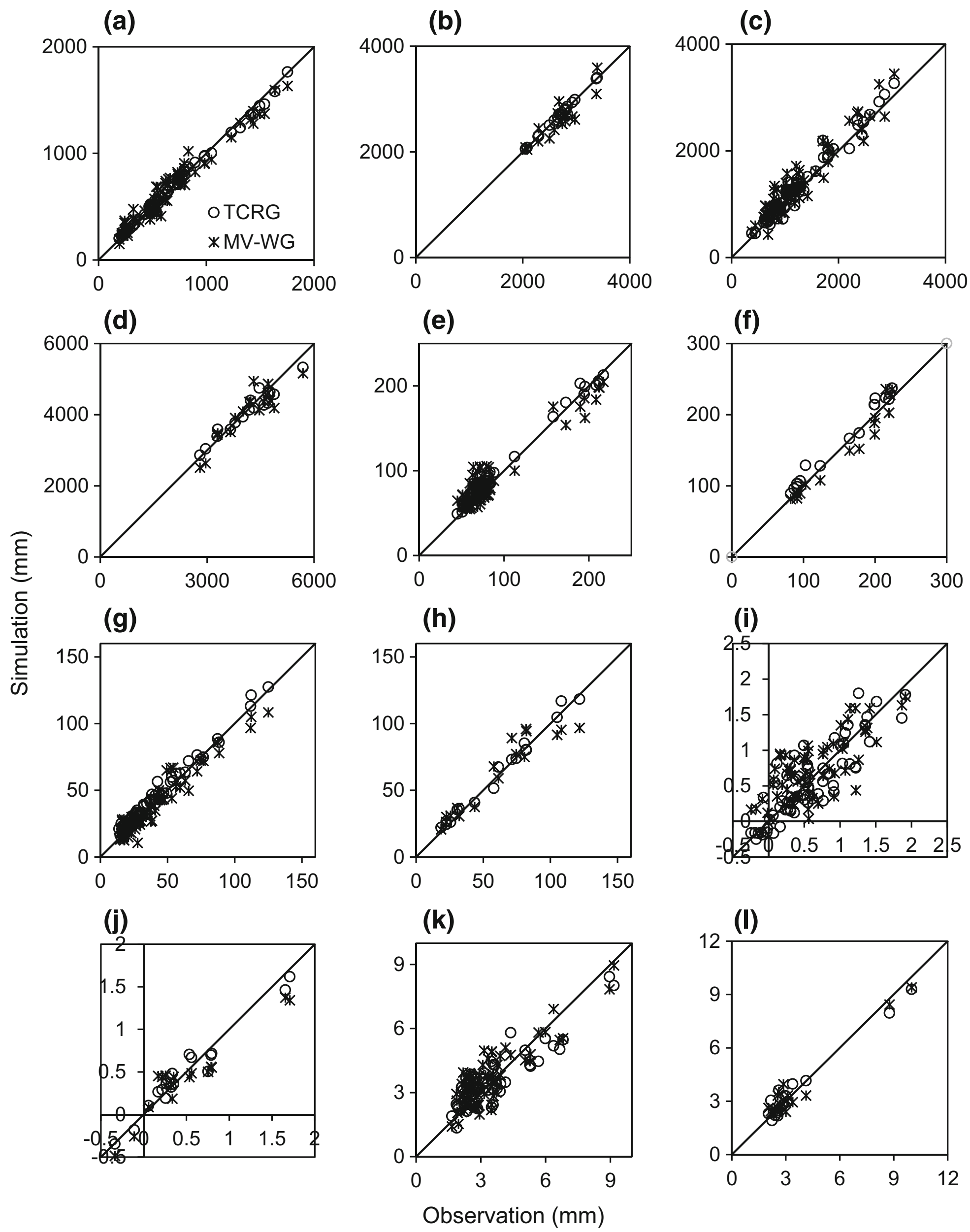

Fig. 6 Scatter plots of simulated (TCRG and MV-WG) versus observed mean, maximum, skewness and kurtosis coefficients of rainfall for all stations. **a** mean of seasonal total rainfall, **b** mean of yearly total rainfall, **c** maximum of seasonal total rainfall, **d** maximum of yearly total rainfall, **e** mean of seasonal maximum daily rainfall, **f** mean of yearly maximum daily rainfall, **g** standard deviation of seasonal maximum daily rainfall, **h** standard deviation of yearly maximum daily rainfall, **i** skewness coefficient of seasonal total rainfall, **j** skewness coefficient of yearly total rainfall, **k** kurtosis coefficient of seasonal total rainfall and **l** kurtosis coefficient of yearly total rainfall

attributed to the fact that both rainfall occurrence models are inadequate to describe the behaviour of the extreme events.

As a matter of fact, location and season are two main factors contributing to the characteristics of rainfall. For this study area which experiences tropical climate, the comparisons above showed that the Markov chain of the TCRG is more consistent in generating the statistical characteristics of the rainfall occurrence. Contrary to reports in earlier literature which claimed that the serial approach is more flexible than the Markov chain (Racsko et al. 1991), it was observed that this is not always true for rainfall occurrence generation as indicated in this study area, Kelantan River Basin, Malaysia. The third-order Markov chain is a higher-order model which considers three previous weather states that implicitly preserve the

Table 9 Acceptability index for the rainfall amounts in passing the Wilcoxon rank sum, Kolmogorov–Smirnov (K–S) and squared ranks tests for all stations

	Wilcoxon rank sum tests		K–S tests		Squared ranks tests	
	TCRG	MV-WG	TCRG	MV-WG	TCRG	MV-WG
Seasonal	100	89.71	98.53	75	86.76	73.53
Yearly	100	100	94.12	88.23	100	76.47

chain dependency between the sequences of rainfall and consistently models the series of wet and dry spells in the Kelantan River Basin. Also, Malaysia is a tropical country which experiences copious amounts of rainfall throughout the year. In particular, the temperature range of Malaysia is very small and the occurrence of drought is very rare. Thus, the third-order Markov chain which exhibits longer memory appears to be sufficient to model the persistent nature of the wet and dry sequences particularly to this study area.

4.2 Rainfall amounts (rainfall characteristics)

While the log-normal and the Weibull distributions of TCRG and MV-WG, respectively, have been applied to model the rainfall amounts on a basis of seasonal and yearly timescales, it is of interest to provide a metric interpretation of the model performance. As shown in Fig. 6a and b, the observed and generated mean of seasonal and yearly rainfall amounts of the TCRG and MV-WG were in a good agreement. The graph revealed that both methods are able to imitate the rainfall mean accurately, which is one of the most important statistics for a stochastic rainfall generator. The similarly good performance of both methods were consistent with the results of AI (Table 9), where all the station–year combinations (17 stations $\times$ 1 year) of both methods passed the Wilcoxon rank sum tests at 5% significance level. The station–monsoon season combinations (17 stations $\times$ 4 monsoons) also gave promising results by yielding 100 and 89.71% for the TCRG and MV-WG, respectively. This is further justified by the results of MAPE (Table 10) with values less than 20% that indicated that the log-normal and Weibull distributions preserved the mean of rainfall amounts accurately. This confirms previous findings reported in the literature that often highlighted the accurate estimation of the mean by stochastic models (Wan et al. 2005; Chen et al. 2012a; Jaafar et al. 2016). The possible explanation for this is that both distributions can fit to the shape of rainfall pattern and describe the average behaviour of the rainfall, thus giving good simulation of the rainfall mean.

Table 10 Mean absolute percentage error (MAPE) of the rainfall generated by TCRG and MV-WG for all stations

	Mean	Std	Maximum	Skewness	Kurtosis
TCRG	3.01	7.59	9.77	81.67	20.38
MV-WG	12.88	20.28	18.55	190.49	22.93

Std standard deviation

For the case of seasonal and yearly extreme values, the observed and generated series of both methods agreed reasonably well (Fig. 6c, d). The simulated rainfall extremes were close to the corresponding observed data which indicated that the true persistence of the real world can be reproduced adequately. The MAPE of log-normal distribution with value of 9.77 was lower than the value of Weibull distribution, which was 18.55, implying a slightly better performance of log-normal distribution (Table 10). The capabilities of both methods in generating the properties (mean, standard deviation and distribution) of seasonal maximum daily rainfall and yearly maximum daily rainfall were evaluated using statistical tests (Table 11). It was found that the log-normal distribution reproduced higher AI values than the Weibull distribution. Nevertheless, both methods showed promising results, especially the reproduction of seasonal and yearly mean of the extreme rainfall where the Wilcoxon rank sum tests of both methods showed non-significant differences between the observed and simulated series for all 17 stations. The results were further supported since the mean and standard deviation data points of seasonal maximum daily rainfall and yearly maximum daily rainfall were close to the 1:1

Table 11 Acceptability index for the seasonal maximum daily rainfall and yearly maximum daily rainfall in passing the Wilcoxon rank sum, Kolmogorov–Smirnov (K–S) and squared ranks tests for all stations

	Wilcoxon rank sum tests		K–S tests		Squared ranks tests	
	TCRG	MV-WG	TCRG	MV-WG	TCRG	MV-WG
Seasonal	100	100	89.71	88.24	83.82	80.88
Yearly	100	100	94.12	70.59	76.47	70.59

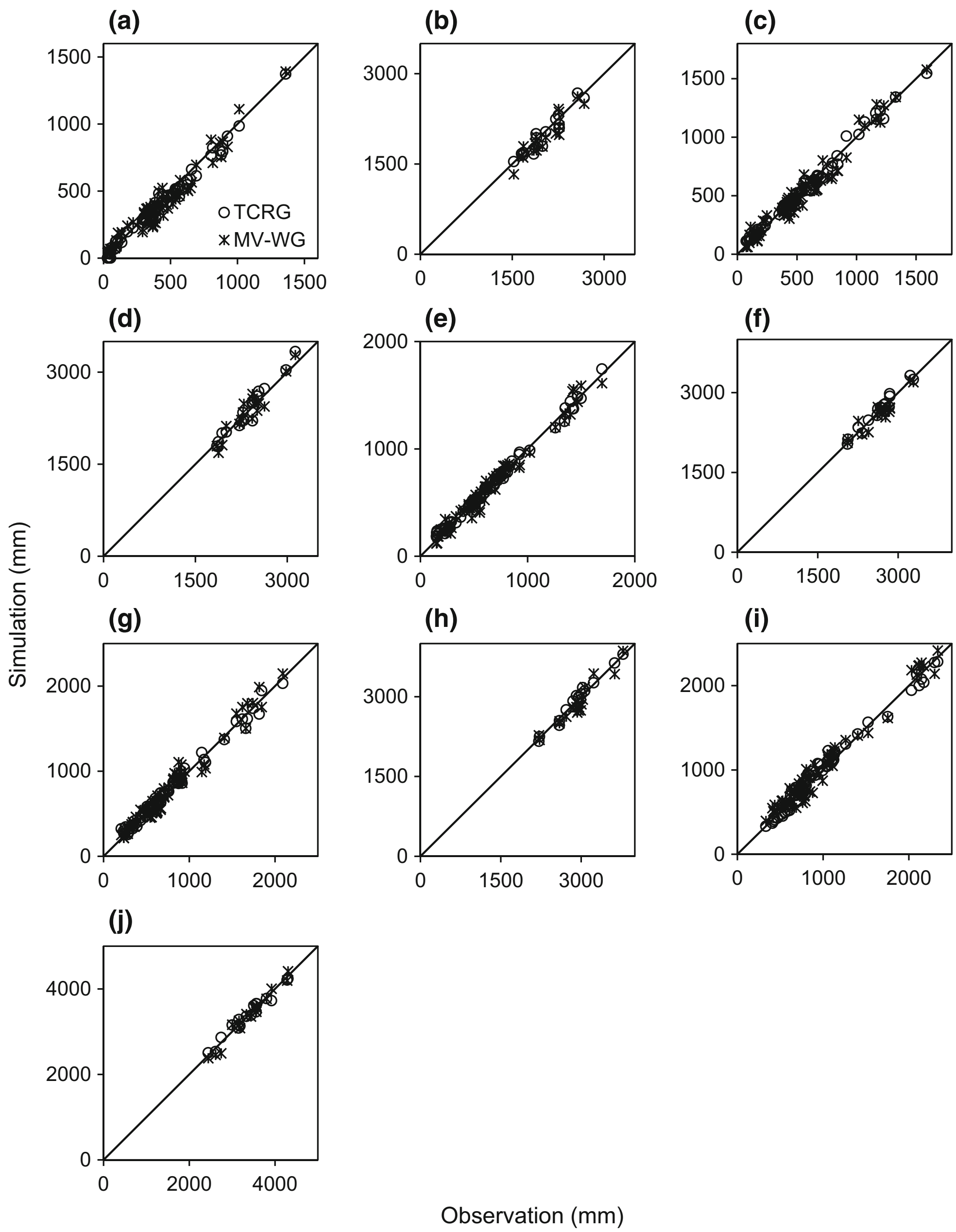

Fig. 7 Scatter plots of simulated (TCRG and MV-WG) versus observed percentiles of rainfall for all stations. **a** 10th percentile of seasonal rainfall, **b** 10th percentile of yearly rainfall, **c** 30th percentile of seasonal rainfall, **d** 30th percentile of yearly rainfall, **e** 50th percentile of seasonal rainfall, **f** 50th percentile of yearly rainfall, **g** 70th percentile of seasonal rainfall, **h** 70th percentile of yearly rainfall, **i** 90th percentile of seasonal rainfall and **j** 90th percentile of yearly rainfall

line (Fig. 6e–h). There is a general consensus that stochastic rainfall models often underestimate the extreme values of rainfall (Kyselý and Dubrovský 2005; Furrer and Katz 2008; Kaczmarska et al. 2014). However, both distributions performed moderately well in modelling the extreme events for the Kelantan River Basin. This emphasizes the idea that different study areas are characterized by different rainfall characteristics. Therefore, the best distribution selected for each area can be different. In this case, the log-normal and Weibull distributions seem to give satisfactory distribution although they are not tailed to model the extremes.

Next, the TCRG and MV-WG models performed less well in generating the seasonal and yearly skewness of rainfall amounts as there are some overestimations and underestimations which can be observed in Fig. 6i and j.

The simulated skewness is not consistent with the observed counterpart, and this may relate partly to the limited ability of both methods in generating the skewed rainfall. The MAPE values for the skewness were 81.67 and 190.49 for TCRG and MV-WG, respectively, indicating an inaccurate forecast (Table 10). The seasonal and yearly data of the kurtosis coefficients were moderately simulated by TCRG and MV-WG, as illustrated in Fig. 6k and l. TCRG and MV-WG reproduced a reasonable range of MAPE values for the kurtosis coefficients, which were 20.38 and 22.93, respectively (Table 10). These results are similar to the previous studies that highlighted the less well production of skewness and kurtosis coefficients by some stochastic weather generators (Chen and Brissette 2014). The existence of discrepancies between the observed and simulated data is understandable because both distributions are not sufficiently flexible to preserve all the aspects of rainfall.

Figure 7 shows a comparison between the observed and generated percentiles (10th, 30th, 50th, 70th and 90th) by TCRG and MV-WG at all stations at two timescales (seasonal and yearly). Both of them reproduced the seasonal and yearly results satisfactorily although there were slight underestimations or overestimations compared with the observed data. Nevertheless, the AI results suggested that the TCRG showed a slightly better performance than MV-WG in view of the higher AI values obtained by the TCRG (Table 9). This indicated that for most of the K–S tests of TCRG, the null hypothesis that stated the observed and generated data drawn from the same distribution are accepted.

In short, the results obtained revealed that the two-parameter log-normal distribution of TCRG was slightly better in modelling the statistical properties of rainfall amounts. In general, the degree of complexities needed are dependent on the characteristics of the climate that is to be modelled. Unlike other countries which experience spring, summer, autumn and winter, the Kelantan River Basin is located near to the equator which receives abundant rainfall throughout the year. There are no extreme conditions like snow or extended drought, which make the estimation of the rainfall properties become more straightforward and simple. This may be the reason why the simpler distribution is preferable than the complex distribution in this study area. The simpler distribution needs less computing procedure and easier to apply in practice. Nevertheless, neither of them can reproduce all the aspects of rainfall completely. For example, they tend to perform less well in generating the skewnesses and kurtosis coefficients of the rainfall. Therefore, more choices of rainfall distribution should be tested and examined in accordance to the study objectives, and this is beyond the scope of this paper.

4.3 Low-frequency variability of seasonal and yearly rainfall occurrences and amounts

The capability of the TCRG and the MV-WG in preserving the low-frequency variability of rainfall was evaluated and examined from the aspects of rainfall occurrence and rainfall amounts. In general, the inability of a stochastic rainfall model in representing the seasonal and inter-annual variabilities of rainfall is often associated with a phenomenon known as overdispersion (Katz and Parlange 1998; Kim et al. 2012). Figure 8 compares the relative performance of TCRG and MV-WG in representing the variability of rainfall occurrence and amounts. It is apparent that the TCRG simulated the standard deviation of the wet spells, seasonal and yearly rainfall considerably well as the data points of observed data were in a good correspondence with the generated counterparts. In contrast, the MV-WG appears to notably underestimate the standard deviation of the wet spells. The overestimation and underestimation of the standard deviation of seasonal and yearly rainfall by MV-WG also can be observed in Fig. 8b and c. The overestimation of the standard deviation of the rainfall could be the result of the sampling error in the observed and simulated standard deviation of the data

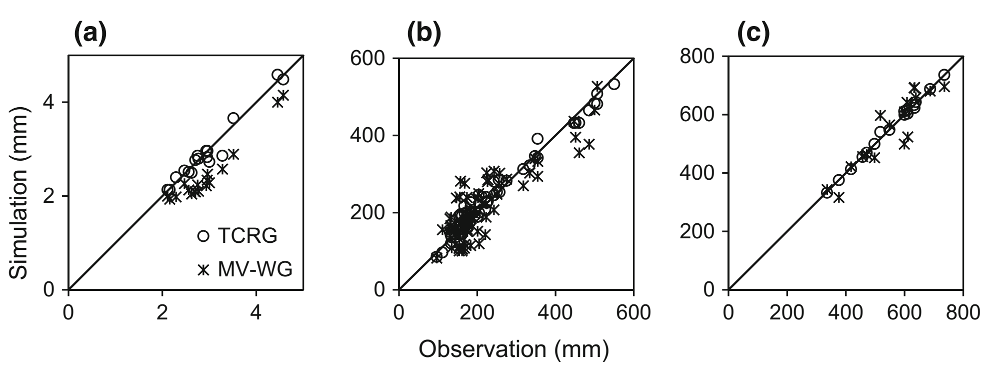

Fig. 8 Scatter plots of simulated versus observed standard deviation from TCRG and MV-WG for all stations. **a** standard deviation of wet spells, **b** standard deviation of seasonal rainfall and **c** standard deviation of yearly rainfall

Table 12 Variance overdispersion of wet spells and rainfall amounts

	Wet spells	Rainfall amounts
TCRG	1.68	−3.44
MV-WG	47.30	14.67

(Katz and Parlange 1998). As mentioned earlier, the underestimation of the standard deviation of the rainfall is one of the common shortcomings of the stochastic rainfall generator (Katz and Parlange 1998; Wilks 1999; Dubrovský et al. 2004; Chen et al. 2009; Serinaldi and Kilsby 2014). A possible explanation for this is that the stochastic models do not consider the low-frequency aspects of climate variability.

To further check whether both methods were able to capture the variability of rainfall, variance overdispersion were averaged over all stations and presented in Table 12. However, the results showed that the Markov chain of TCRG performed better in mitigating the overdispersion by producing substantially less value of overdispersion in wet spells variance, which was 1.68%. In comparison, the serial approach of MV-WG was found to poorly generate the wet spells variance by giving large average overdispersion, which was 47.30%. Next, TCRG and MV-WG reproduced −3.44 and 14.67%, respectively, for the overdispersion in wet day rainfall amounts. It seems that the spectral correction method TCRG contributed to the reduction of overdispersion, but retained some of the underdispersion, thus reproducing a negative average overdispersion.

The results were further shown by using AI as a criterion, as illustrated in Tables 7 and 9. The results highlighted that the TCRG achieved consistently higher AI with respect to the standard deviation of wet spells, seasonal and yearly rainfall amounts than the MV-WG. This is quantitatively supported as TCRG passed most of the squared ranks tests, reflecting that the generated standard deviations showed nonsignificant differences with the observed counterparts. This is also consistent with the results of MAPE (Tables 8, 10) where the TCRG achieved relatively lower MAPE values than the MV-WG in terms of the standard deviation of wet spells and rainfall amounts.

To sum up, the finding of this study suggested that the higher-order (third-order) Markov chain of the TCRG was clearly better in capturing the seasonal and inter-annual variabilities of rainfall occurrence. The serial approach of the MV-WG being less appropriate in this case could be a consequence of the assumption of implied independence within the wet spells lengths and thus was not suitable for the rainfall characteristics in this study area. On the other hand, the third-order Markov chain of the TCRG utilized in this study conditioned the probability of getting rain based on the rainfall state of the immediate three previous days, thus introducing more information to the model, and a lower overdispersion is to be expected. The finding is similar to some previous studies that reported the higher-order Markov chain can work better in mitigating the overdispersion while the low-order (first-order) Markov chain was inadequate in preserving the low-frequency variability of rainfall occurrence (Mason 2004; Dubrovský et al. 2004; Chen et al. 2012b).

In terms of the low-frequency variability of rainfall amounts, the overall results indicated that the TCRG is more accurate and robust in correcting the low-frequency variability of rainfall amounts in this study area. Although TCRG does not eliminate the overdispersion of rainfall completely, there was a significant improvement from the spectral correction approach of TCRG in preserving the seasonal and inter-annual variabilities of rainfall amounts. As explained in Sect. 3, the spectral correction approach modelled the low-frequency properties of the seasonal and yearly rainfall by using the fast Fourier transform. The variances of the seasonal and yearly rainfall were preserved very well, and the variability bias was controlled successfully. The underlying stationary assumption of MV-WG in general does not incorporate the low-frequency aspects of rainfall and tend to produce a smaller variance of the generated process.

5 Conclusions

This study proposed a simple yet effective stochastic rainfall model for the generation of rainfall occurrences and amounts in the tropical area, specifically for the Kelantan River Basin, Malaysia. It also demonstrated its applicability in preserving the seasonal and inter-annual variabilities of rainfall. The performance of this tropical climate rainfall generator (TCRG) was compared with the well-known MV-WG model. Both models have a similar stochastic rainfall generation procedure, but differ considerably in the choice of rainfall occurrence and amounts simulation methods.

In summary, it was highlighted that the third-order Markov chain of the TCRG was more accurate in generating the basic statistics of rainfall occurrence and the log-normal distribution of TCRG performed better in simulating the key features of rainfall amounts. Besides, the spectral correction approach (Chen et al. 2010) of the TCRG preserved the seasonal and inter-annual variabilities of the rainfall successfully. There was an appreciable improvement in mitigating the overdispersion problem, although it was not eliminated completely. Nevertheless, it was found that both the TCRG and MV-WG models performed less well in generating the skewnesses and kurtosis coefficients of rainfall. This has been attributed to that the distributions of both stochastic models may not be sufficiently flexible to model the inherent skewnesses and kurtosis coefficients. Future investigations should address this issue by considering more types of the distributions to produce better simulation results for the rainfall.

It can be concluded that the TCRG performed more efficiently in generating the essential properties of the rainfall occurrences and amounts, as well as the seasonal and inter-annual variabilities of rainfall in the Kelantan River Basin. The main strengths of the TCRG lie in its simplicity and straightforwardness. It acts as an easy and efficient tool for generating rainfall in the tropical climate for hydrological modelling purposes and may potentially be applied for the climate change assessments and to other tropical climate regions. The MV-WG may still be considered as a stochastic tool which can produce acceptable simulation of the synthetic rainfall, particularly the mean of the rainfall amounts. As with any model, it is essential to examine the model structure and performance before applying the generated series in any application because the climate regime of any given watershed is different. This study only evaluated and compared the performance of two stochastic rainfall models in this particular study area. Ultimately, it would be interesting to further investigate whether the tropical climate rainfall generator would work as well with other rainfall models and other types of frequency distributions under different climatic conditions. Further improvements to the stochastic rainfall model can be performed by simulating the rainfall characteristics based on climate model results in order to assess the climate change impacts and risks (Deidda et al. 2013; Langousis et al. 2015). A more exhaustive validation should also be carried out to evaluate stochastic rainfall models from different aspects, such as the presence of uncertainty and sensitivity (Gronewold et al. 2013; Chandra et al. 2015).

Acknowledgements The authors wish to express their appreciation to the Malaysian Meteorological Department (MMD) for providing the rainfall data and to the Ministry of Education Malaysia (MOE) for the financial support. The authors also would like to acknowledge the sincere appreciation to the reviewers for their valuable comments.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Contr 19:716–723. doi:10.1109/TAC.1974.1100705

Bernardara P, De Michele C, Rosso R (2007) A simple model of rain in time: an alternating renewal process of wet and dry states with a fractional (non-Gaussian) rain intensity. Atmos Res 84:291–301. doi:10.1016/j.atmosres.2006.09.001

Breinl K, Turkington T, Stowasser M (2013) Stochastic generation of multi-site daily precipitation for applications in risk management. J Hydrol 498:23–35. doi:10.1016/j.jhydrol.2013.06.015

Cantet P, Bacro J-N, Arnaud P (2011) Using a rainfall stochastic generator to detect trends in extreme rainfall. Stoch Environ Res Risk Assess 25:429–441. doi:10.1007/s00477-010-0440-x

Castellvi F, Stockle CO, Mormeneo I, Villar JM (2002) Testing the performance of different processes to generate temperature and solar radiation: a case study at Lleida (northeast Spain). Trans ASAE 45:571–580. doi:10.13031/2013.8936

Chandra R, Saha U, Mujumdar PP (2015) Model and parameter uncertainty in IDF relationships under climate change. Adv Water Resour 79:127–139. doi:10.1016/j.advwatres.2015.02.011

Chen J, Brissette FP (2014) Comparison of five stochastic weather generators in simulating daily precipitation and temperature for the Loess Plateau of China. Int J Climatol 34:3089–3105. doi:10.1002/joc.3896

Chen J, Zhang XC, Liu WZ, Li Z (2009) Evaluating and extending cligen precipitation generation for the Loess Plateau of China. J Am Water Resour Assoc 45:378–396. doi:10.1111/j.1752-1688.2008.00296.x

Chen J, Brissette FP, Leconte R (2010) A daily stochastic weather generator for preserving low-frequency of climate variability. J Hydrol 388:480–490. doi:10.1016/j.jhydrol.2010.05.032

Chen J, Brissette FP, Leconte R (2012a) WeaGETS—a Matlab-based daily scale weather generator for generating precipitation and temperature. Proc Environ Sci 13:2222–2235. doi:10.1016/j.proenv.2012.01.211

Chen J, Brissette FP, Leconte R, Caron A (2012b) A versatile weather generator for daily precipitation and temperature. Trans ASABE 55:895–906

Cowden JR, Watkins DW, Mihelcic JR (2008) Stochastic rainfall modeling in West Africa: parsimonious approaches for domestic rainwater harvesting assessment. J Hydrol 361:64–77. doi:10.1016/j.jhydrol.2008.07.025

Deidda R, Marrocu M, Caroletti G, Pusceddu G, Langousis A, Lucarini V, Puliga M, Speranza A (2013) Regional climate models' performance in representing precipitation and temperature over selected Mediterranean areas. Hydrol Earth Syst Sci 17:5041–5059. doi:10.5194/hess-17-5041-2013

Dlamini NS, Rowshon MK, Sahab U, Fikri A, Lai SH, Mohd MSF (2015) Developing and calibrating a stochastic rainfall generator model for simulating daily rainfall by Markov chain approach. J Teknol 76:13–19. doi:10.11113/jt.v76.5946

Dubrovský M, Buchtele J, Žalud Z (2004) High-frequency and low-frequency variability in stochastic daily weather generator and its effect on agricultural and hydrologic modelling. Clim Change 63:145–179. doi:10.1023/B:CLIM.0000018504.99914.60

Fodor N, Dobi I, Mika J, Szeidl L (2010) MV-WG: a new multi-variable weather generator. Meteorol Atmos Phys 107:91–101. doi:10.1007/s00703-010-0074-z

Fodor N, Dobi I, Mika J, Szeidl L (2013) Applications of the MVWG multivariable stochastic weather generator. Sci World J 2013:1–6. doi:10.1155/2013/571367

Furrer EM, Katz RW (2008) Improving the simulation of extreme precipitation events by stochastic weather generators. Water Resour Res 44:1–13. doi:10.1029/2008WR007316

Gronewold AD, Stow CA, Crooks JL, Hunter TS (2013) Quantifying parameter uncertainty and assessing the skill of exponential dispersion rainfall simulation models. Int J Climatol 33:746–757. doi:10.1002/joc.3469

Hansen JW, Mavromatis T (2001) Correcting low-frequency variability bias in stochastic weather generators. Agric For Meteorol 109:297–310. doi:10.1016/S0168-1923(01)00271-4

Harrison M, Waylen P (2000) A note concerning the proper choice for Markov model order for daily precipitation in the humid tropics: a case study in Costa Rica. Int J Climatol 20:1861–1872. doi:10.1002/1097-0088(20001130)20:14<1861:AID-JOC577>3.0.CO;2-9

Harrold TI (2003) A nonparametric model for stochastic generation of daily rainfall amounts. Water Resour Res 39:1343. doi:10.1029/2003WR002570

Hasan MM, Dunn PK (2010) A simple Poisson–gamma model for modelling rainfall occurrence and amount simultaneously. Agric For Meteorol 150:1319–1330. doi:10.1016/j.agrformet.2010.06.002

Hassan Z, Harun S (2013) Impact of climate change on rainfall over Kerian, Malaysia with Long Ashton Research Station Weather Generator (LARS-WG). Malays J Civ Eng 25:33–44

Hauser T, Demirov E (2013) Development of a stochastic weather generator for the sub-polar North Atlantic. Stoch Environ Res Risk Assess 27:1533–1551. doi:10.1007/s00477-013-0688-z

Hong N-M, Lee T-Y, Chen Y-J (2016) Daily weather generator with drought properties by copulas and standardized precipitation indices. Environ Monit Assess 188:383. doi:10.1007/s10661-016-5395-z

Huang YF, Puah YJ, Chua KC, Lee TS (2015) Analysis of monthly and seasonal rainfall trends using the Holt's test. Int J Climatol 35:1500–1509. doi:10.1002/joc.4071

Jaafar J, Baki A, Abu Bakar IA, Tahir W, Awang H, Ismail F (2016) Evaluation of stochastic daily rainfall data generation models. In: Tahir W, Abu Bakar PIDSH, Wahid MA, Mohd Nasir SR, Lee WK (eds) ISFRAM 2015. Springer, Singapore, pp 203–220

Jones PG, Thornton PK (1993) A rainfall generator for agricultural applications in the tropics. Agric For Meteorol 63:1–19. doi:10.1016/0168-1923(93)90019-E

Jones PG, Thornton PK (2000) MarkSim: software to generate daily weather data for Latin America and Africa. Agron J 92:445–453. doi:10.2134/agronj2000.923445x

Jones PG, Thornton PK (2013) Generating downscaled weather data from a suite of climate models for agricultural modelling applications. Agric Syst 114:1–5. doi:10.1016/j.agsy.2012.08.002

Kaczmarska J, Isham V, Onof C (2014) Point process models for fine-resolution rainfall. Hydrol Sci J 59:1972–1991. doi:10.1080/02626667.2014.925558

Kannan SK, Farook JA (2015) Stochastic simulation of precipitation using markov chain—mixed exponential model methodology precipitation occurrence. Appl Math Sci 9:3205–3212. doi:10.12988/ams.2015.54308

Katz RW, Parlange MB (1998) Overdispersion phenomenon in stochastic modeling of precipitation. J Clim 11:591–601. doi:10.1175/1520-0442(1998)011<0591:OPISMO>2.0.CO;2

Khazaei MR, Ahmadi S, Saghafian B, Zahabiyoun B (2013) A new daily weather generator to preserve extremes and low-frequency variability. Clim Change 119:631–645. doi:10.1007/s10584-013-0740-5

Kim T-W, Ahn H, Chung G, Yoo C (2008) Stochastic multi-site generation of daily rainfall occurrence in south Florida. Stoch Environ Res Risk Assess 22:705–717. doi:10.1007/s00477-007-0180-8

Kim Y, Katz RW, Rajagopalan B, Podesta GP, Furrer EM (2012) Reducing overdispersion in stochastic weather generators using a generalized linear modeling approach. Clim Res 53:13–24. doi:10.3354/Cr01071

Kyselý J, Dubrovský M (2005) Simulation of extreme temperature events by a stochastic weather generator: effects of interdiurnal and interannual variability reproduction. Int J Climatol 25:251–269. doi:10.1002/joc.1120

Langousis A, Kaleris V (2014) Statistical framework to simulate daily rainfall series conditional on upper-air predictor variables. Water Resour Res 50:3907–3932. doi:10.1002/2013WR014936

Langousis A, Mamalakis A, Deidda R, Marrocu M (2015) Assessing the relative effectiveness of statistical downscaling and distribution mapping in reproducing rainfall statistics based on climate model results. Water Resour Res. doi:10.1002/2015WR017556

Lewis CD (1982) Industrial and business forecasting methods: a practical guide to exponential smoothing and curve fitting. Butterworth Scientific, London

Li C, Singh VP, Mishra AK (2013) A bivariate mixed distribution with a heavy-tailed component and its application to single-site daily rainfall simulation. Water Resour Res 49:767–789. doi:10.1002/wrcr.20063

Liu Y, Zhang W, Shao Y, Zhang K (2011) A comparison of four precipitation distribution models used in daily stochastic models. Adv Atmos Sci 28:809–820. doi:10.1007/s00376-010-9180-6

Lu Y, Qin XS, Mandapaka PV (2015) A combined weather generator and K-nearest-neighbour approach for assessing climate change impact on regional rainfall extremes. Int J Climatol 35:4493–4508. doi:10.1002/joc.4301

Mandal KG, Padhi J, Kumar A, Ghosh S, Panda DK, Mohanty RK, Raychaudhuri M (2014) Analyses of rainfall using probability distribution and Markov chain models for crop planning in Daspalla region in Odisha, India. Theor Appl Climatol 121:517–528. doi:10.1007/s00704-014-1259-z

Mason SJ (2004) Simulating climate over western North America using stochastic weather generators. Clim Change 62:155–187. doi:10.1023/B:CLIM.0000013700.12591.ca

Mayowa OO, Pour SH, Shahid S, Mohsenipour M, Harun SB, Heryansyah A, Ismail T (2015) Trends in rainfall and rainfall-related extremes in the east coast of peninsular Malaysia. J Earth Syst Sci 124:1609–1622. doi:10.1007/s12040-015-0639-9

Mehrotra R, Sharma A (2007) Preserving low-frequency variability in generated daily rainfall sequences. J Hydrol 345:102–120. doi:10.1016/j.jhydrol.2007.08.003

Paschalis A, Molnar P, Fatichi S, Burlando P (2014) On temporal stochastic modeling of precipitation, nesting models across scales. Adv Water Resour 63:152–166. doi:10.1016/j.advwatres.2013.11.006

Peck Yen T, Rohasliney H (2013) Status of water quality subject to sand mining in the Kelantan River, Kelantan. Trop Life Sci Res 24:19–34

Pour S, Harun S, Shahid S (2014) Genetic programming for the downscaling of extreme rainfall events on the East Coast of Peninsular Malaysia. Atmosphere (Basel) 5:914–936. doi:10.3390/atmos5040914

Pradhan B (2010) Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. J Spat Hydrol 9:1–18

Racsko P, Szeidl L, Semenov M (1991) A serial approach to local stochastic weather models. Ecol Modell 57:27–41. doi:10.1016/0304-3800(91)90053-4

Rao AR, Hamed KH (2000) Flood frequency analysis. CRC Press, Boca Raton

Schoof JT, Arguez A, Brolley J, O'Brien JJ (2005) A new weather generator based on spectral properties of surface air temperatures. Agric For Meteorol 135:241–251. doi:10.1016/j.agrformet.2005.12.004

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6:461–464

Semenov MA, Brooks RJ (1999) Spatial interpolation of the LARS-WG stochastic weather generator in Great Britain. Clim Res 11:137–148. doi:10.3354/cr011137

Semenov MA, Brooks RJ, Barrow EM, Richardson CW (1998) Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates. Clim Res 10:95–107. doi:10.3354/cr010095

Serinaldi F, Kilsby CG (2014) Simulating daily rainfall fields over large areas for collective risk estimation. J Hydrol 512:285–302. doi:10.1016/j.jhydrol.2014.02.043

Sharma MA, Singh JB (2010) Use of probability distribution in rainfall analysis. New York Sci J 3:40–49

Shui LT, Haque A (2004) Stochastic rainfall model for irrigation projects. Pertanika J Sci Technol 12:137–147

So BJ, Kwon HH, Kim D, Lee SO (2015) Modeling of daily rainfall sequence and extremes based on a semiparametric Pareto tail approach at multiple locations. J Hydrol 529:1442–1450. doi:10.1016/j.jhydrol.2015.08.037

Sonnadara DUJ, Jayewardene DR (2014) A Markov chain probability model to describe wet and dry patterns of weather at Colombo. Theor Appl Climatol 119:333–340. doi:10.1007/s00704-014-1117-z

Srikanthan R, Harrold TI, Sharma A, McMahon TA (2005) Comparison of two approaches for generation of daily rainfall data. Stoch Environ Res Risk Assess 19:215–226. doi:10.1007/s00477-004-0226-0

Suhaila J, Deni SM, Zin WZW, Jemain AA (2010) Trends in Peninsular Malaysia rainfall data during the southwest monsoon and northeast monsoon seasons: 1975–2004. Sains Malays 39:533–542. doi:10.1007/s00703-010-0108-6

Sundaresan J, Sreekesh S, Ramanathan AL, Sonnenschein L, Boojh R (eds) (2013) Climate change and island and coastal vulnerability. Springer, Netherlands

Swanson DA, Tayman J, Bryan TM (2011) MAPE-R: a rescaled measure of accuracy for cross-sectional subnational population forecasts. J Popul Res 28:225–243. doi:10.1007/s12546-011-9054-5

Tangang FT, Juneng L, Salimun E, Sei KM, Le LJ (2012) Climate change and variability over Malaysia: gaps in science and research information. Sains Malaysiana 41:1355–1366

Tehrany MS, Pradhan B, Jebur MN (2015) Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. Stoch Environ Res Risk Assess 29:1149–1165. doi:10.1007/s00477-015-1021-9

Tingem M, Rivington M, Azam-Ali S, Colls J (2007) Assessment of the ClimGen stochastic weather generator at Cameroon sites. Afr J Environ Sci Technol 1:86–92

Varikoden H, Samah AA, Babu CA (2010) Spatial and temporal characteristics of rain intensity in the peninsular Malaysia using TRMM rain rate. J Hydrol 387:312–319. doi:10.1016/j.jhydrol.2010.04.023

Varikoden H, Preethi B, Samah AA, Babu CA (2011) Seasonal variation of rainfall characteristics in different intensity classes over Peninsular Malaysia. J Hydrol 404:99–108. doi:10.1016/j.jhydrol.2011.04.021

Wan H, Zhang X, Barrow EM (2005) Stochastic modelling of daily precipitation for Canada. Atmos Ocean 43:23–32. doi:10.3137/ao.430102

Wang QJ, Nathan RJ (2007) A method for coupling daily and monthly time scales in stochastic generation of rainfall series. J Hydrol 346:122–130. doi:10.1016/j.jhydrol.2007.09.003

Wheater HS, Chandler RE, Onof CJ, Isham VS, Bellone E, Yang C, Lekkas D, Lourmas G, Segond M-L (2005) Spatial-temporal rainfall modelling for flood risk estimation. Stoch Environ Res Risk Assess 19:403–416. doi:10.1007/s00477-005-0011-8

Wilks DS (1999) Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. Agric For Meteorol 93:153–169. doi:10.1016/S0168-1923(98)00125-7

Wilks DS, Wilby RL (1999) The weather generation game: a review of stochastic weather models. Prog Phys Geogr 23:329–357. doi:10.1177/030913339902300302