Jurnal
Teknologi

Full Paper

TROPICAL DAILY RAINFALL AMOUNT MODELLING USING MARKOV CHAIN-MIXED EXPONENTIAL (MCME)

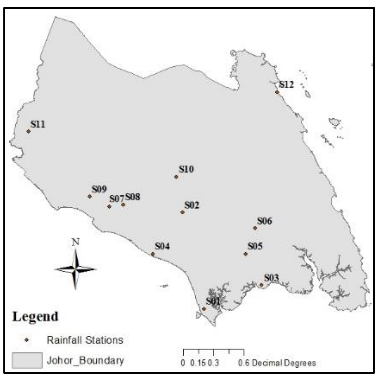
Fadhilah Yusof[a]*, Lee Mee Yung[a], Zulkifli Yusop[b]

[a]Department of Mathematical Sciences, Faculty of Science, University Teknology Malaysia, 81310 UTM Johor Bahru, Johor
[b]Institute of Environmental and Water Resource Management (IPASA), Faculty of Civil Engineering, University Teknologi Malaysia, 81310, UTM Johor Bahru, Johor

Article history
Received
24 April 2015
Received in revised form
4 May 2015
Accepted
9 May 2015

*Corresponding author
fadhilahy@utm.my

Graphical Abstract



Abstract

This study is concerned with the development of a stochastic rainfall model that can generate many sequences of synthetic daily rainfall series with the similar properties as those of the observed. The proposed model is Markov chain-mixed exponential (MCME). This model is based on a combination of rainfall occurrence (represented by the first-order two-state Markov chain) and the distribution of rainfall amounts on wet days (described by the mixed exponential distribution). The feasibility of the MCME model is assessed using daily rainfall data from four rainfall stations (station S02, S05, S07 and S11) in Johor, Malaysia. For all the rainfall stations, it was found that the proposed MCME model was able to describe adequately rainfall occurrences and amounts. Various statistical and physical properties of the daily rainfall processes also considered. However, the validation results show that the models' predictive ability was not as accurate as their descriptive ability. The model was found to have fairly well ability in predicting the daily rainfall process at station S02, S05 and S07. Nonetheless, it was able to predict the daily rainfall process at station S11 accurately.

Keywords: MCME, Markov chain, mixed exponential distribution, daily rainfall, rainfall station

Abstrak

Kajian ini adalah berkaitan dnegan pembentukkan model hujan stokastik yang boleh menjana banyak siri hujan harian sintetik dengan sifat yang sama dengan data yang dicerap. Model yang dicadangkan ialah rantai Markov-bergabung eksponen. Model ini adalah berdasarkan gabungan antara kejadian hujan (diwakili dengan rantai Markov) dengan taburan jumlah hujan harian (digambarkan oleh taburan bergabung eksponen). Keupayaan model ini dinilai menggunakan data hujan harian dari empat stesen hujan (stesen S02, S05, S07 dan S11) di Johor. Keputusan kajian mendapati bahawa model yang dicadangkan berupaya untuk menerangkan secukupnya sifat statistic dan fizikal proses hujan harian yang diambil kira untuk keempat-empat stesen hujan. Walau bagaimanapun, keputusan pengesahan menunjukkan bahawa keupayaan model untuk membuat ramallan tidaklah setepat keupayaan deskriptif. Model ini didapati mempunyai keupayaan sederhana untuk meramal proses hujan harian di stesen S11 dengan tepat. Secara keseluruhan, model ini boleh menerangkan pola bermusim terhadap sifat cerapan hujan untuk semua stesen hujan.

Kata kunci: MCME, rantai Markov, taburan bergabung eksponen, hujan harian, stesen hujan

1.0 INTRODUCTION

Rainfall is a continuous process where various sizes and shapes of isolated raindrops fall at different rates. A significant period of dry weather with no rainfall can have major consequences on water supply affecting plants and crop production, while excessive rainfall may cause flood which brings a great cost to human, economic and environmental systems. Therefore, knowledge of the frequency of occurrence and intensity of rainfall events is essential for water resources management.

A more recent climate issue of concern is the effects of rainfall variable on agriculture. Many studies have been conducted to study the impacts of rainfall variables, especially rainfall occurrence and intensity on crop production (Zhang *et al.* 2004, Zhang and Liu 2005, Yu *et al*. 2010, Yang *et al*. 2012). Such studies always require daily rainfall data as input. However, even when the rainfall records are available, they contain only limited and finite information regarding the historical rainfall data. With this limitation, stochastic simulations of rainfall have been widely used to generate many sequences of synthetic rainfall series that could accurately preserve the properties of the observed rainfall at a given location.

Generally, there are two stochastic models that are commonly being used in describing the rainfall process, namely cluster model (Kavvas and Delleur 1981) and occurrence-amount model (Woolhiser and Roldan 1982). Compared to cluster model, Han (2001) showed that the occurrence-amount model provides a better fit to rainfall amounts. Occurrence-amount model consists of two components: rainfall occurrence and rainfall amount. The rainfall occurrence is based on the sequence of wet and dry days while the rainfall amount is based on wet day amount. Two models to represent each component are combined to form an overall rainfall model. Markov chain-mixed exponential (MCME) is an example of occurrence-amount model that has successfully been employed to model daily rainfall series (Woolhiser and Pengram 1979, Woolhiser and Roldan 1986, Han 2001, Hussain 2008, Detzel and Mine 2011).

Fadhilah *et al.* (2007b) used MCME model for simulating hourly rainfall series in Peninsula Malaysia. It was found that the MCME model was able to preserve the statistical and physical properties of the rainfall process. The capability of this model need to be further assessed using different data sets.

Therefore, this study utilized the capability of MCME model using the daily rainfall series. The objectives are to generate synthetic daily rainfall series using MCME model and assess the performance of the model by comparing the synthetic daily rainfall series with the observed daily rainfall data for some areas in Johor. The model's ability to preserve accurately the statistical and physical properties of the observed data will be evaluated. This is because this model will offer significant help to the water resource and planning authorities in generating synthetic data at stations where data quality and records are inadequate.

2.0 EXPERIMENTAL

2.1 Study Area and Data

Johor is the largest state in the southern part Peninsular Malaysia and is located between the 1°20"N and 2°35"N latitudes. It covers a total land area of about 19210 km² and has an equatorial climate with northeast monsoon rain from November until February blowing from the South China Sea.

The daily rainfall dataset of 12 rainfall stations in Johor, which covers the period from January 1975 to December 2007 were used in this study. The data sets obtained from Malaysia Metrological Department were of good quality with no missing values throughout the 33 year period. Location of the rainfall stations can be seen in Figure 1 and Table 1.

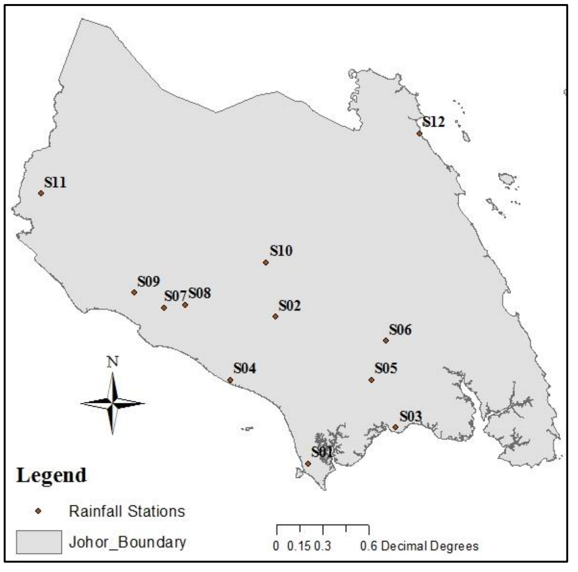


Figure 1 Location of rainfall stations in Johor

Table 1 List of rainfall stations considered with their geographical coordinates

Station	Station Name	Latitude	Longitude
S01	Ladang Getah Kukup Pontian	1°21'00"N	103°27'36"E
S02	Ladang Benut Rengam	1°50'24"N	103°21'00"E
S03	Stor JPS JB	1°28'12"N	103°45'00"E
S04	Pintu Kawalan Tampok Batu Pahat	1°37'48"N	103°12'00"E
S05	Senai	1°37'48"N	103°40'12"E
S06	Sek Men Bkt Besar	1°45'36"N	103°43'12"E
S07	Sek Men Inggeris Batu Pahat	1°52'12"N	102°58'48"E
S08	Pintu Kawalan Sembrong	1°52'48"N	103°03'00"E
S09	Pintu Kawalan Separap	1°55'12"N	102°52'48"E
S10	Kluang	2°01'12"N	103°19'12"E
S11	Tangkak	2°15'00"N	102°34'12"E
S12	Mersing	2°27'00"N	103°49'48"E

2.2 Markov Chain-Mixed Exponential (MCME) model

Markov chain-mixed exponential (MCME) model is a type of occurrence-amount model. This model can be expressed mathematically by assuming the amount of rainfall falling on t^{th} day and n^{th} year is a random variable. The MCME model $\{Z_n(t): t = 1, 2, \ldots;\ n = 1,2,\ldots\}$ is defined as:

$$Z_n(t) = X_n(t)Y_n(t) \qquad (1)$$

where $X_n(t)$ represents the occurrence process and $Y_n(t)$ represents the amount of rainfall when $X_n(t)$ is wet. The process of daily rainfall occurrences is represented by a first-order two-state Markov chain while the mixed exponential distribution is used to describe the distribution of daily rainfall amounts on wet days.

2.2.1 The Occurrence Process

A first-order Markov chain model is used to simulate daily rainfall occurrences due to its simplicity and the relative ease in estimating the model's two parameters. The rainfall data is treated as a series of two states, namely dry or wet; modelled as either a 0 or 1 respectively. The random variable represents the occurrence or non-occurrence of precipitation on day *t* of year *n* can be expressed as:

$$X_n = \begin{cases} 0 & \text{if day } t \text{ is dry} \\ 1 & \text{if day } t \text{ is wet} \end{cases} \qquad (2)$$

Thus, the transition probabilities of the first-order Markov chain are defined as follows:

$$p_{ij}(t) = P\{X_n(t) = j \mid X_n(t-1) = i\} \qquad (3)$$

where *i* and *j* can be 0 or 1, *t* = 1, 2,… and *n* = 1,2,3…

The maximum likelihood estimation is used to estimate transition probabilities by computing the observed number of transitions $a_{ij,k}(t)$ from state *i* on day *t* to state *j* on day *t*+1 in period *k* across the entire length of record (Woolhiser and Pegram 1979). By taking the year into *k* = 12 monthly periods, two transition probabilities to be estimated are formulated as follows:

$$p_{00,k}(t) = a_{00,k}(t)/[a_{00,k}(t) + a_{01,k}(t)] \qquad (4)$$
$$p_{10,k}(t) = a_{10,k}(t)/[a_{10,k}(t) + a_{11,k}(t)] \qquad (5)$$

where p_{00} is the probability of a day to be dry given that the previous day was dry and p_{10} is the probability of a day to be dry given that the previous day was wet.

2.2.2 The Amount Process

Motivated by Fadhilah *et al.* (2007a) and Suhaila *et al*. (2007) whom have proven that the mixed exponential distribution model is suitable in describing rainfall data in Peninsular Malaysia, the mixed exponential distribution is used in this work to model the daily rainfall amounts on wet days.

Let $Y_n(t)$ denotes the rainfall amount on the t^{th} day of the n^{th} year. If $X_n(t) = 1$, then $Y_n(t)$ is greater than or equal to a threshold value. In this study, the threshold value is equal to 1mm where rainfall amount less than 1mm is considered as dry day. The distribution of daily rainfall amounts $Y_n(t)$ is described by the mixed exponential as follows:

$$f_{Y_n(t)}(x) = (p/\beta_1)\exp(-\frac{x}{\beta_1} + \left(1 - \frac{p}{\beta_2}\right)\exp\left(-\frac{x}{\beta_2}\right) \qquad (6)$$

for x ≥ 1, $0 \le p \le 1$ and $0 < \beta_1 < \beta_2$ where p is the mixing probability, β_1 and β_2 explain a small mean and large mean respectively of two exponential distributions and x represents the daily rainfall amount on wet day.

The parameters of the mixed exponential distribution are estimated through the method of maximum likelihood (MLE) with the log-likelihood function defined as follows:

$$l = \ln L = \sum_{i=1}^{n} \ln \begin{bmatrix} \left(\frac{p}{\beta_1}\right)\exp\left(-\frac{x}{\beta_1}\right) + \\ (1 - p/\beta_2)\exp(-x/\beta_2) \end{bmatrix} \qquad (7)$$

where n is the sample size.

The iterative optimization technique is used to maximize the log-likelihood function which is in implicit form. Everitt and Hand (1981) suggested the following solutions for estimating the parameters to the log-likelihood equation.

$$\hat{p} = (1/n)\sum_{i=1}^{n} \hat{P}(1|x_i) \qquad (8)$$

$$\hat{\beta}_1 = (1/n\hat{p})\sum_{i=1}^{n} \hat{P}(1|x_i)x_i \qquad (9)$$

$$\hat{\beta}_2 = [1/n(1-\hat{p})]\sum_{i=1}^{n}\hat{P}(2|x_i)x_i \qquad (10)$$

Where

$$\hat{P}(1|x_i) = \{(p/\beta_1)[\exp(-x_i/\beta_1)]\}/\{(p/\beta_1)[\exp(-x_i/\beta_1)] + (1-p/\beta_2)[\exp(-x_i/\beta_2)]\}$$

$$\hat{P}(2|x_i) = \{(1-p/\beta_2)[\exp(-x_i/\beta_2)]\}/\{(p/\beta_1)[\exp(-x_i/\beta_1)] + (1-p/\beta_2)[\exp(-x_i/\beta_2)]\}$$

These iterative equations are used for getting the optimal solution using a method suggested by *Nguyen* and *Mayabi* (1990). Through this method, seven initial estimates for p and β_1 are formed by ranging p from 0.2 to 0.8 at intervals of 0.1 and β_1 from $0.2\hat{x}$ to $0.8\hat{x}$ at intervals of $0.1\hat{x}$ where $\hat{x}$ is the mean rainfall amounts of all wet days. The corresponding β_2 is calculated using the given p and β_1:

$$\beta_2 = [\hat{x} - (p/\beta_1)]/(1-p) \qquad (11)$$

The iteration provides the highest value out of the seven likelihood functions is taken to be the optimal solution to estimate the parameters.

2.3 Assessment of the MCME model

Daily rainfall series from station S02, S05, S07 and S11 (chosen randomly from 12 rainfall stations in Johor) are used to assess the performance of the daily MCME model. The series are divided into two-subseries with different length based on the rule of thumb. The longer sub-series (approximately 2/3 length of series) is used to calibrate the model and the shorter sub-series (approximately 1/3 length of series) is used to validate the model. Based on the parameter set estimated from calibration period (1975-1997), 30 simulations of synthetic daily rainfall series for 33 year period (1975-2007) were generated using random number generation process.

For the evaluation of MCME model's performance, simulated MCME parameters will be compared with the observed values for calibration and validation period (1998-2007). In addition, a set of statistical and physical properties will also be used for the evaluation of model's ability in preserving the observed properties of rainfalls. In this study, these comparisons are analysed monthly. For the month of February, the analysis is conducted separately for the non-leap years and leap years where February 1 and February 2 represent the February for the non-leap years and leap years respectively.

Model evaluation is based on the graphical comparison between the simulated and observed characteristics. Graphically, the simulated rainfall characteristics are represented by the boxplots and the observed characteristics are represented by the dots connected by the dashed lines. The proposed model is said to have an "excellent" or "very well" ability in conserving the characteristics of historical data if the observed value is comparable to the median value (the middle 50% value) of the boxplot. If the observed value falls on the whiskers and within the range defined by the simulated minimum and maximum, then the proposed model is said to have a "fairly well" ability. Otherwise, the model either underestimates or overestimates the observed characteristics.

Performance of MCME model in calibration period and validation period would be compared. If MCME model can perform well for calibration period, this means that the model has the ability to describe the daily rainfall process. On the other hand, if the MCME model can preserve observed characterize of rainfalls for the validation period, the model is said to have predictive ability.

3.0 RESULTS AND DISCUSSION

3.1 Performance of the Daily MCME Model in Calibration Period (1975-1997)

In assessing the descriptive ability of MCME model, daily rainfall series from year 1975 to 1997 is used in the calibration process. At each station, monthly MCME parameters for calibration period are estimated using MLE and the results are summarized in Table 2. These calibrated parameters are used to generate 30 simulations of synthetic time series for 33 year period. From the 30 sets of monthly simulated data, simulated characteristics (represented by boxplots) computed for all stations are compared with the empirical characteristics (represented by the dots connected by the dashed line). In this section, however only the graphical comparison for station S02 is depicted (see Figures 2 to 6).

3.1.1 Transition Probabilities for Calibration Period

Graphical comparisons (Figure 2) have shown that the simulated transition probabilities are well preserved and comparable to the empirical values for four rainfall stations. The median value of each boxplot excellently matched the empirical value. Besides, the seasonal trends of the probabilities are also well preserved. Thus, it can be concluded that the daily rainfall occurrence characteristics for all stations are well preserved by the MCME simulations.

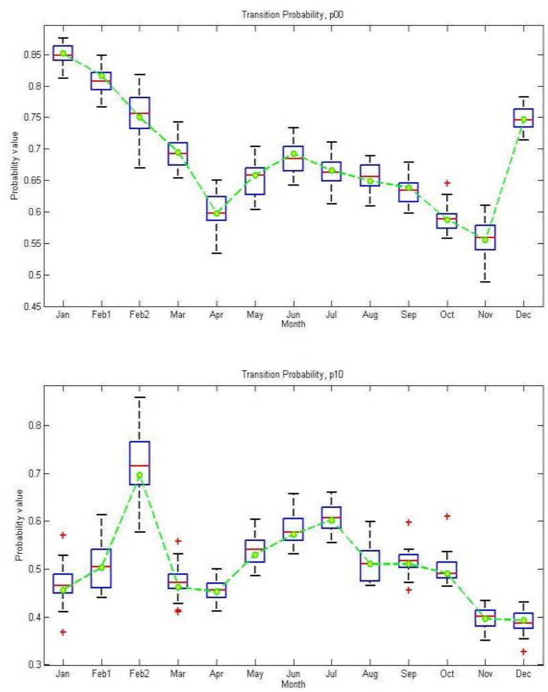


Figure 2 Comparison between simulated and empirical Markov chain parameters at station S02

At station S02, the probability of a dry day following another dry day decreases rapidly from the month of January till April followed by a slight increase till June and finally declines to its lowest value in November before rising again in December. Besides, the relatively low value of p_{10} in the months of November and December shows that a given day in these two months are more unlikely to be dry if the previous day was wet.

Transition probabilities at station S05 show that a day in the months of April and November tends to be dry if the previous day was dry. Similarly, there is a higher probability of rain on a given day if the previous day was also rainy in the months of January, November and December. Besides, comparison results also have shown that the month of November as the wettest month in both station S07 and S11. This is because the probability of a day being wet is the highest during this month, regardless of the status of the previous day.

In general, all stations show a low probability of having a given day is dry if the previous day was wet in the months of November, December and January.

3.1.2 Mixed Exponential Parameters for Calibration Period

From the comparisons, the middle 50% of simulated mixing parameters, p in the boxplots (see Figure 3), in particular, do not match the empirical values. This happens for the month of October and December in station S02, November and December in station S07 with the month of April in station S11. This is also true for the simulation of the smaller mean, β_1 and larger mean, β_2 in the same month in station S02. The middle 50% of boxplots for the smaller mean at station S07 and larger mean at station S11 however contain the empirical values. Only empirical larger mean for the month of December in station S07 and empirical smaller mean for the month of April in station S11 are on the whisker of the boxplots.

In contrast, the median of the simulated boxplots and the empirical parameter values for station S05 show close agreement in value as well as the trend. Its rainfall distribution largely consists of a larger mean distribution. Overall, daily rainfall amounts characteristics for all rainfall stations are well preserved by the MCME model.

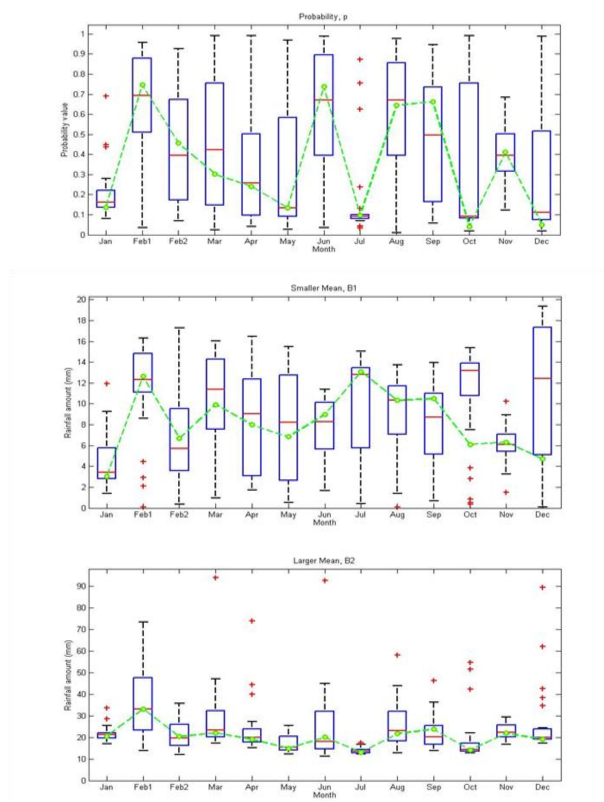


Figure 3 Comparison between simulated and empirical mixed exponential parameters at station S02

The wide differences in simulated and empirical values for the mixed exponential for some months in all rainfall stations can be explained mainly due to the fact that the mixed exponential parameters are estimated by the maximum likelihood method. This method does not aim to preserve the specific observed mixed exponential parameters but merely aims to find any parameters that maximize the likelihood of matching the mixed exponential function to the empirical distribution. Therefore, it is expected that the simulated rainfall series would provide a good match between the simulated mixed exponential function and the empirical rainfall distribution rather than preserve the exact values of mixed exponential parameters.

3.1.3 Statistical Properties for Calibration Period

Statistical properties (mean, standard deviation, kurtosis and skewness) of 30 simulations and observed rainfall series are compared (see Figure 4). At all stations, the means of the simulations have close agreement with the observed statistical properties. For the standard deviation, most of the empirical values are contained in the middle 50% of the boxplots except for the month of July and November in station S02 with the month of December in station S05. MCME model preserved the kurtosis of daily rainfall series accurately for the whole year except for the month of December in station S05 with the month of January and March in station S07. This is also true for the skewness of the observed series where the empirical values for the same months are underestimated by the model. Overall, statistical properties of the observed rainfall series are well preserved by the MCME model.

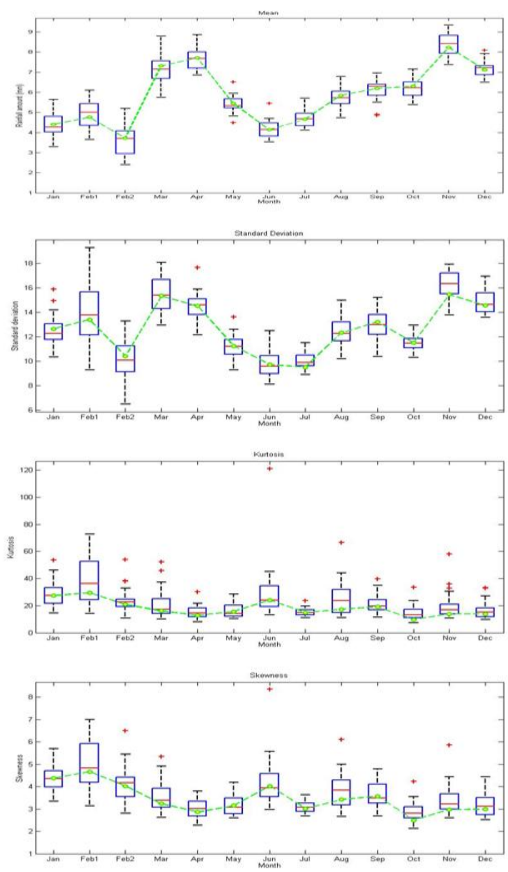


Figure 4 Comparison between simulated and empirical statistical properties at station S02

3.1.4 Physical Properies for Calibration Period

Physical properties include the daily maximum and number of dry or wet days is evaluated on monthly basis. For daily maximum rainfall, the observed values are contained in the middle 50% of the boxplots for most of the months in all stations, however only result of station S02 is displayed. Empirical values for the month of February1, April, October and November in station S02, May, October and November in station S05, January, April, June and August in station S07 and April, June and October in station S11, however fall on the whisker of the boxplots. Only empirical values for the month of December in station S05 and March in station S07 are underestimated by the MCME model.

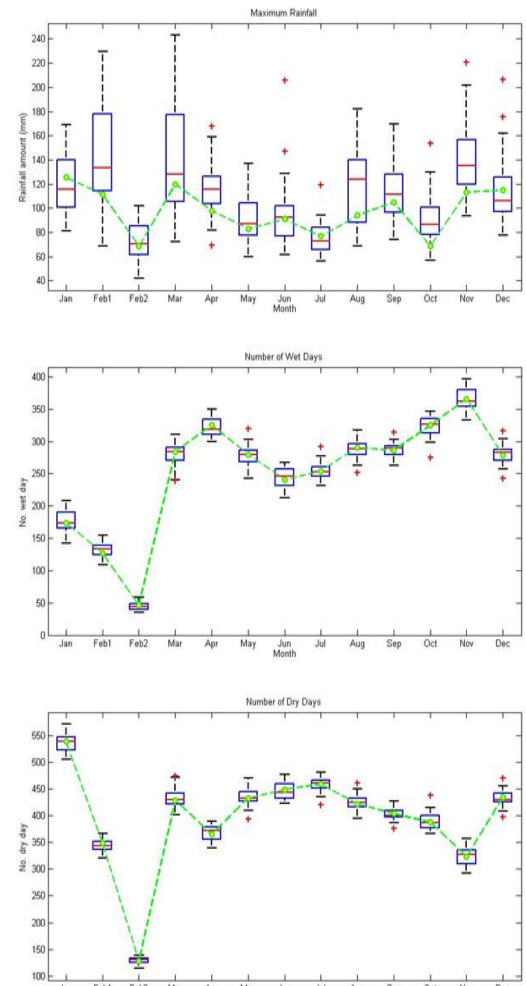


Figure 5 Comparison between simulated and empirical physical properties at station S02

In contrast, the number of wet days and dry days for all stations has shown excellent agreements between the observed values and the medians of simulated properties. Among the twelve months for all rainfall stations, January is the driest month while November is the month that gives most rainfall. In general, the simulations are able to preserve adequately various physical properties of the rainfall series considered.

Table 2 Summary of MCME parameters estimation for four rainfall stations

Stn	MCME parameters	Month												
		Jan	Feb1	Feb2	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
S02	p_{00}	0.8516	0.8166	0.7500	0.6946	0.5973	0.6582	0.6927	0.6659	0.6493	0.6386	0.5876	0.5556	0.7471
	p_{10}	0.4566	0.5039	0.6957	0.4629	0.4523	0.5305	0.5726	0.6024	0.5103	0.5105	0.4907	0.3962	0.3935
	p	0.1351	0.7463	0.4563	0.3022	0.2412	0.1337	0.7390	0.0965	0.6452	0.6636	0.0399	0.4108	0.0501
	β_1	3.0096	12.6368	6.6994	9.9421	7.9920	6.8397	8.9758	13.0811	10.3362	10.4892	6.1012	6.3224	4.7152
	β_2	20.4330	33.1432	20.2865	22.0159	19.0520	14.9723	20.0248	13.0846	21.5956	23.7107	14.1564	21.9325	18.9770
S05	p_{00}	0.8046	0.7781	0.7368	0.6841	0.5106	0.5618	0.6566	0.5966	0.6447	0.5876	0.5324	0.4586	0.6193
	p_{10}	0.4069	0.4966	0.5000	0.4570	0.4513	0.4824	0.5714	0.5446	0.5261	0.5331	0.4301	0.3950	0.3722
	p	0.4673	0.2949	0.4585	0.3298	0.2844	0.3411	0.4443	0.2765	0.4901	0.5824	0.4005	0.5094	0.7359
	β_1	5.4787	4.7769	5.0591	6.3675	6.6170	6.2977	6.6453	4.0306	6.5210	8.9870	5.5809	6.6787	9.0661
	β_2	23.7794	20.7607	21.9031	21.0554	19.2810	18.7958	18.7269	17.1076	22.5401	23.1960	19.8413	23.3577	36.5262
S07	p_{00}	0.7971	0.7411	0.7381	0.7303	0.6483	0.7149	0.7313	0.6979	0.7078	0.6722	0.6642	0.5714	0.6850
	p_{10}	0.4464	0.4790	0.6875	0.4457	0.5404	0.4704	0.5747	0.5909	0.5400	0.5167	0.4323	0.3934	0.4505
	p	0.6986	0.2831	0.3149	0.3511	0.8046	0.5193	0.3940	0.8980	0.4704	0.5039	0.7236	0.9088	0.9834
	β_1	7.0199	5.4717	6.2587	7.2774	14.0169	6.3468	6.5628	14.9127	9.3506	9.8564	10.3546	14.2634	15.6519
	β_2	27.2277	18.7184	29.1731	20.7365	32.7641	24.4345	20.2964	42.8463	22.5102	20.7982	23.7423	23.6356	50.5579
S11	p_{00}	0.9002	0.8408	0.8045	0.7105	0.6209	0.7134	0.7500	0.7111	0.7344	0.6636	0.6493	0.5620	0.7700
	p_{10}	0.4959	0.6061	0.6341	0.5156	0.6269	0.5897	0.7706	0.6590	0.5565	0.5960	0.5069	0.4461	0.4622
	p	0.3243	0.6765	0.2458	0.2256	0.0661	0.4041	0.4403	0.5255	0.1029	0.1151	0.5110	0.4507	0.9073
	β_1	7.9893	10.3615	7.1712	5.9123	6.4171	9.6409	8.9700	10.7701	4.9727	15.2235	9.2265	9.6229	12.3516
	β_2	16.1527	26.7960	25.3803	17.8925	19.1372	22.7398	22.3124	19.8585	16.7193	15.2336	18.7951	18.9417	30.8703

3.2 Performance of the Daily MCME Model in Validation Period (1998-2007)

In assessing the predictive ability of MCME model, daily rainfall series from same rainfall stations which covers the period from 1998 to 2007 is used in the validation process. Simulated MCME parameters and properties of 30 simulations for the same period are compared to the empirical values using graphical method. A further investigation for the simulated and empirical characteristics is conducted by summarizing the performance of simulated characteristics on the bar chart and the findings are discussed in the sections below.

3.2.1 Transition Probabilities for Validation Period

The performance of simulated transition probabilities at four rainfall stations is summarized in Figure 6. Among the stations, empirical transition probabilities, p_{00} from station S02 are most likely to be well preserved by the simulated values followed by station S11, S05 and finally S07. Besides, it is apparent that at each rainfall stations majority of the empirical transition probabilities, p_{00} are fairly well preserved. Empirical values from all stations have the possibility to be overestimated except for the station S07.

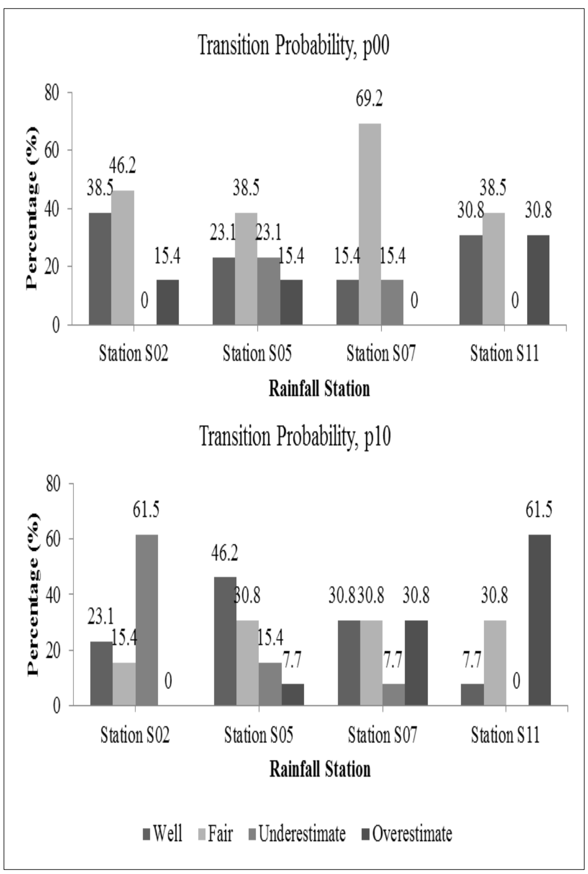


Figure 6 Performance of simulated transition probabilities at four rainfall stations

At station S02, over 50% of the empirical transition probabilities, p_{10} are underestimated by the MCME model. But, majority of the empirical transition probabilities at station S05 are excellently preserved by the simulated values. Furthermore, most of the empirical values at station S11 are overestimated.

Generally, MCME model is equally likely to well, fairly well preserve or underestimate the observed Markov chain parameters at station S02. However, the model tends to well or fairly well preserve the empirical values for the Markov chain parameters at station S05. Besides, MCME simulations are prone to fairly well preserve the empirical transitional probabilities at station S07 while overestimate at stationS11.

3.2.2 Mixed Exponential Parameters for Validation Period

Figure 7 summarizes the performance of simulated mixed exponential parameters at four rainfall stations. Among these stations, no empirical parameters are overestimated except for the mixing probability, p at station S07. Most of the simulated mixing probability at station S02, S05 and S11 are comparable to the empirical values. In contrast, majority of the empirical probabilities at station S07 are fairly well preserved.

At station S11, more than 50% of empirical smaller mean, β_1 are well preserved by MCME model. However, none of empirical smaller mean at station S02 is well preserved. They are either fairly good preserved or underestimated. Besides, majority of the empirical smaller mean at station S05 and S07 are fairly well preserved. For larger mean β_2, most of empirical values at station S02 are fairly well preserved and this is also true for other three stations.

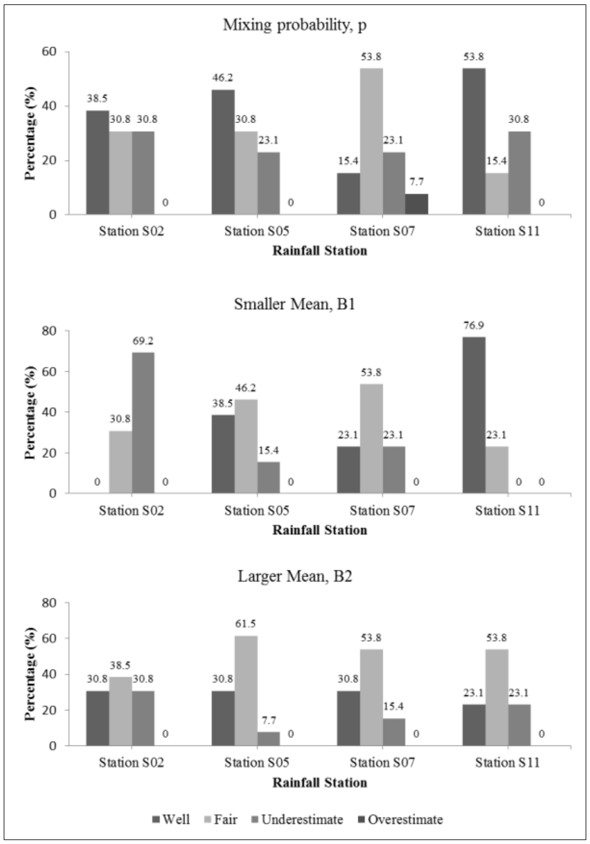


Figure 7 Performance of simulated mixed exponential parameters at four rainfall stations

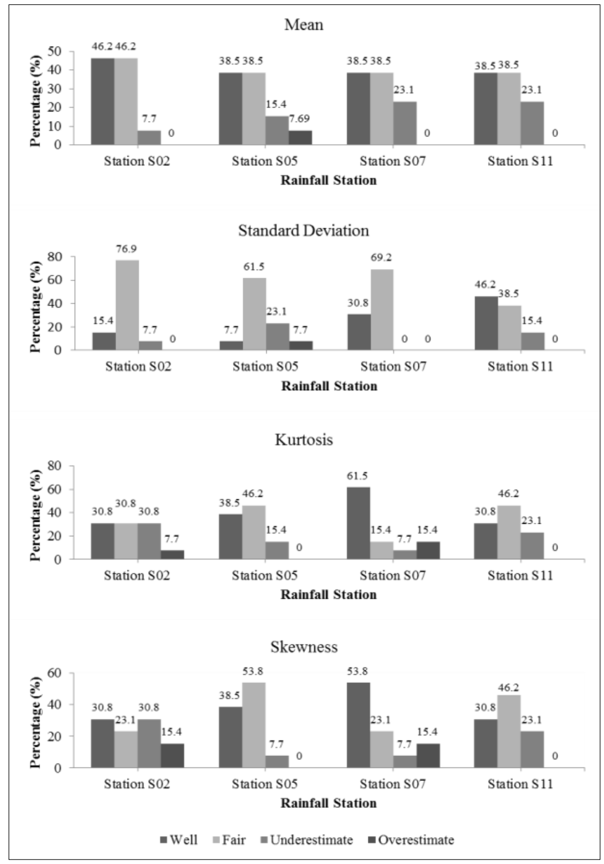


Figure 8 Performance of simulated statistical properties at four rainfall station

Overall, MCME model is likely to underestimate the empirical values for the mixed exponential at station S02. In contrast, simulations generated by the model tend to fairly well preserve the empirical mixed exponential parameters at station S05 and S07. At station S11, MCME model tends to well preserve the empirical values for mixed exponential parameters.

3.2.3 Statistical Properties for Validation Period

From Figure 8, it can be seen that at each station, the probability for the daily means of observed rainfall series to be well and fairly well preserved by MCME model is the same. Besides, there is no probability for the observed daily means to be overestimated except for station S05. For the standard deviations of observed rainfall series, most of the empirical values at station S02, S05 and S07 are fairly well preserved by the model. In contrast, over 40% of the empirical standard deviations at station S11 are well preserved.

Furthermore, the likelihood for the kurtosis of observed series to be underestimated, well and fairly well preserved, is the same at station S02. At station S05 and S11, most of the empirical kurtosis is fairly good preserved. Among the rainfall stations, the likelihood for the empirical skewness to be well preserved is highest at station S07. However, empirical skewness at station S05 has the highest probability to be fairly well preserved.

Overall, MCME simulations are prone to fairly well preserve the statistical properties of observed rainfall series at all stations except for the station S07.

3.2.4 Physical Properties for Validation Period

From Figure 9, it is apparent that there is a high percentage for the maximum rainfall of observed rainfall series at station S05, S07 and S11 to be well preserved. Most of the empirical maximum amount at station S02 however is fairly well preserved. Besides, there is no probability for the empirical maximum amount to be overestimated among the stations except for the station S05.

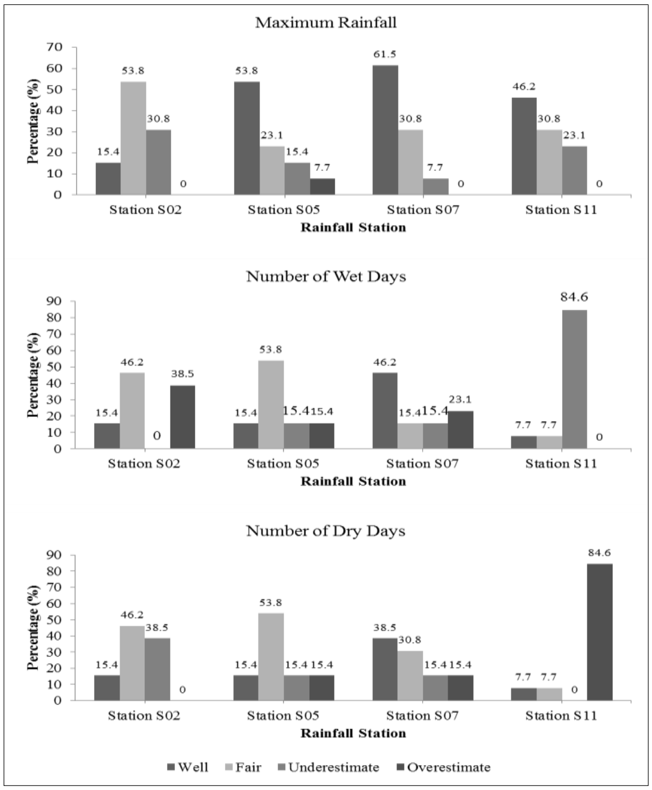


Figure 9 Performance of simulated physical properties at four rainfall stations

For the number of wet days in rainfall series, most of the empirical values at station S02 and S05 fall on the whisker of the boxplots. This is also true for the empirical values for the number of dry days at the same stations. But, at station S07, most of the empirical values for the number of wet and dry days in series are well preserved. At station S11, the percentage for the number of wet days to be underestimated is same as the percentage for the number of dry days to be overestimated.

In general, MCME simulations are prone to fairly well preserve the physical properties of observed rainfall series at station S02 and S05. On the other hand, MCME model is likely to well preserve the observed physical properties at station S07. At station S11, the model tends to underestimate the physical properties of observed rainfall series.

3.3 Comparison between the Performance of Daily MCME Model in Calibration and Validation Period

Performance of daily Markov chain-mixed exponential model is assessed based on the performance of simulated MCME parameters, statistical and physical properties of simulated rainfall series in calibration and validation period and the result is summarized (Figure 10). It is apparent that MCME simulations prone to well preserve the observed characteristics during calibration period for all rainfall stations. Therefore, at all rainfall stations, the model is able to describe the daily rainfall process accurately.

In contrast, the performance of daily MCME model in the validation period is not as good as in the calibration period. In validation period, MCME simulations tend to fairly well preserve the observed characteristics for all rainfall stations except for the station S11. Hence, the model has fair ability in predicting the daily rainfall process at station S02, S05 and S07 but tends to predict the daily rainfall process at station S11 accurately.

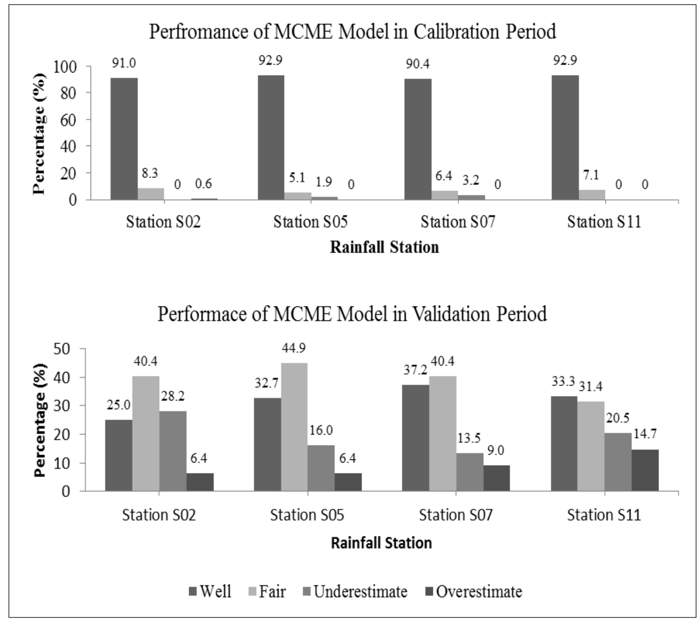


Figure 10 Performance of daily MCME model in both periods

4.0 CONCLUSION

The feasibility of the MCME model in simulating rainfall series is evaluated using daily rainfall data from four rainfall stations (station S02, S05, S07 and S11) in Johor for the 33 year period (1975-2007). At all stations, graphical comparisons have shown that the rainfall occurrence process can be well described by the first-order two-state Markov chain model. For the distribution of rainfall amounts on wet days, mixed exponential distribution is found to describe well. Besides, the statistical and physical properties of the underlying daily rainfall process are well described by the MCME daily model. The performance of MCME model in validation period is not as well as in the calibration period. The model is found to have a fairly well ability in predicting the daily rainfall process at station S02, S05 and S07 but tends to predict the rainfall process at station S11 accurately. In general, the model is able to preserve the seasonal trend of the observed rainfall properties for all stations.

Several recommendations may be suggested for improving the modelling of the MCME. The order of Markov chain should be investigated first before modelling rainfall occurrences using Markov chain. Distribution of daily rainfall amount should also fitted using various statistical distributions to find the most acceptable fit. Besides, parameters for the rainfall amount distribution should be estimated and compared using different methods to find the best method for estimating the parameters.

Acknowledgement

This project was sponsored by the Ministry of Education (MOE) through the Fundamental Research Grant Scheme with grant number 4F341. The authors would also like to thank the Research Management Centre (RMC) at Universiti Teknologi Malaysia (UTM) for managing the project. This study was also supported by the Asian Core Program of the Japanese Society for the Promotion of Science (JSPS) and the Ministry of Education (MOE) Malaysia.

References

[1] Detzel, D. H. M., and Mine, M. R. M. 2011. Generation of Daily Synthetic Precipitation Series: Analyses and Application in La Plata River Basin. *The Open Hydrology Journal*. 5: 69-77.

[2] Everitt, B. S., and Hand, D. J. 1981. *Finite Mixture Distributions. Monographs on Applied Probability and Statistics*. London: Chapman and Hall.

[3] Fadhilah, Y., Zalina, M. D., Nguyen, V. T. V., and Zulkifli, Y. 2007a. Fitting the Best-fit Distribution for the Hourly Rainfall Amount in the Wilayah Persekutuan. *Jurnal Teknologi*. 46(C): 49-58. doi: 10.11113/jt.v46.288.

[4] Fadhilah, Y., Zalina, M. D., Nguyen, V. T. V., and Zulkifli, Y. 2007b. Stochastic Modelling of Hourly Rainfall Series Using Markov Chain Mixed Exponential (MCME) Model. *Prosiding Simposium Kebangsaan Sains Matematik ke-15*, Concorde Hotel, Shah Alam, Selangor. 503-510.

[5] Han, S. Y. 2001. Stochastic Modelling of Rainfall Processes: A Comparative Study Using Data from Different Climatic Conditions. Thesis (Master). McGill University, Quebec, Canada.

[6] Hussain, A. 2008. Stochastic modelling of the Daily Rainfall Process: An Application of a MCME Model in Different Climatic Conditions. Thesis (Master). McGill University, Quebec, Canada.

[7] Nguyen, V. T. V., and Mayabi, A. 1990. Probabilistic Analysis of Summer Daily Rainfall for the Montreal region. *Canadian Water Resources Journal*. 15(3). doi: 10.4296/cwrj1601065.

[8] Suhaila, J., and Jemain, A. A. 2007. Fitting the Statistical Distributions to the Daily Rainfall Amount In Peninsular Malaysia. *Jurnal Teknologi*. 46(C): 33-48. doi: 10.11113/jt.v46.287.

[9] Woolhiser, D. A., and Pegram G. G. S. 1979. Maximum Likelihood Estimation of Fourier Coefficient to Describe Seasonal Variations of Parameters in Stochastic Daily Precipitation Models. *Journal of Applied Meteorology*. 18: 34-42. doi: 10.1175/1520 0450(1979)018 <0034:MLEOFC> 2.0. C O;2.

[10] Woolhiser, D. A., and Roldan, J. 1986. Seasonal and Regional Variability of Parameters for Stochastic Daily Precipitation Models: South Dakota, U.S.A. *Water Resources Research*. 22(6): 965-978. doi: 10.1029/WR022i006p00965.

[11] Yang, T. C., Chen, C., Kuo, C. M., Tseng, H. W., and Yu, P. S. 2012. Drought Risk Assessments of Water Resources Systems Under Climate Change: A Case Study in Southern Taiwan. *Hydrological Earth System Science Discussion*. 9: 12395-12433, doi:10.5194/hessd-9-12395-2012.

[12] Yu, P. S., Tseng H. W., and Chen S. T. 2010. The Impact of Climate Change on Agriculture Water Resources for Paddy Rice Over Southern Taiwan. *International Congress on Environmental Modeling and Software*, Ottawa, Ontario, Canada.

[13] Zhang, X. C., and Liu, W. Z. 2005. Simulating Potential Response of Hydrology, Soil Erosion, and Crop Productivity to Climate Change in Changwu Tableland Region on the Loess Plateau of China. *Agricultural and Forest Meteorology*. 131: 127-142. doi: 10.1016/j.agrformet.2005.05.005.

[14] Zhang, X. C., Nearing M. A., Garbrecht J. D., and Steiner J. L. 2004. Downscaling Monthly Forecasts to Simulate Impacts of Climate Change on Soil Erosion And Wheat Production. *Soil Science Society of America*. 68: 1376-1385,. doi: 10.2136/sssaj2004.1376.