

TEMPORAL RAINFALL DISAGGREGATION BY A SIMPLE RANDOM CASCADE MODEL

Siti Mariam Saad^{a*}, Abdul Aziz Jemain^b, Noriszura Ismail^b

^aDepartment of Computer and Mathematical Sciences, UiTM, 13500, Permatang Pauh, Penang, Malaysia

^bScience Mathematics Studies Center, Faculty of Science and Technology, UKM, 43600, Bangi, Selangor, Malaysia

Article history

Received

3 May 2017

Received in revised form

25 June 2018

Accepted

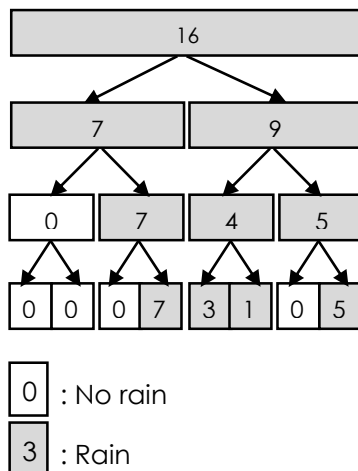
15 July 2018

Published online

5 October 2018

*Corresponding author
smariam.saad@
ppinang.uitm.edu.my

Graphical abstract



Abstract

This study evaluates the utility and suitability of a simple discrete multiplicative random cascade model for temporal rainfall disaggregation. Two of a simple random cascade model, namely log-Poisson and log-Normal β models are applied to simulate hourly rainfall from daily rainfall at seven rain gauge stations in Peninsular Malaysia. The cascade models are evaluated based on the capability to simulate data that preserve three important properties of observed rainfall: rainfall variability, intermittency and extreme events. The results show that both cascade models are able to simulate reasonably well the commonly used statistical measures for rainfall variability (e.g. mean and standard deviation) of hourly rainfall. With respect to rainfall intermittency, even though both models are underestimated, the observed dry proportion, log-Normal β model is likely to simulate number of dry spells better than log-Poisson model. In terms of rainfall extremes, it is demonstrated that log-Poisson and log-Normal β models gave a satisfactory performance for most of the studied stations herein, except for Dungun and Kuala Krai stations, which both located in the east part of Peninsula.

Keywords: Rainfall disaggregation, random cascade models, statistical moment scaling, rainfall variability, intermittency, rainfall extremes

Abstrak

Kajian ini dijalankan untuk menilai kebergunaan dan kesesuaian model lara rawak mudah secara pendaraban diskrit untuk mendisagregasi siri masa hujan. Dua daripada model lara rawak mudah iaitu model log-Poisson dan log-Normal β digunakan untuk mendisagregasi data hujan skala jam-jaman daripada data hujan skala harian di tujuh stesen hujan di Semenanjung Malaysia. Model lara tersebut dinilai berdasarkan keupayaan untuk mensimulasi data hujan yang dapat mengekalkan tiga ciri penting yang terdapat pada cerapan data hujan: kevariabelan data hujan, selang masa hujan terputus-putus dan peristiwa hujan ekstrem. Hasil analisis data menunjukkan kedua-dua model berupaya untuk mensimulasi data hujan per jam dengan agak baik di mana data simulasi mengekalkan ukuran statistik yang sering digunakan (min dan sisihan piawai). Berkenaan selang masa hujan terputus-putus pula, walaupun kedua-dua model menganggarkan nilai perkadaran jam kering kurang daripada perkadaran jam kering yang dicerap, model log-Normal β menghasilkan bilangan jam kering lebih baik daripada model log-Poisson. Berkenaan analisis ekstrem hujan, hasil kajian menunjukkan model log-Poisson dan log-Normal β memberikan keputusan yang memberangsangkan di hampir kesemua stesen hujan, kecuali di stesen Dungun dan Kuala Krai yang mana kedua-dua stesen ini terletak di bahagian timur semenanjung.

Kata kunci: Disagregasi hujan, model lara rawak, penskalaan momen statistic, kevariabelan data hujan, selang masa hujan terputus-putus, hujan ekstrem

© 2018 Penerbit UTM Press. All rights reserved

1.0 INTRODUCTION

Long and continuous rainfall data at hourly or finer resolutions are often needed in a wide range of real-world water engineering problems, related to diverse fields such as urban storm drainage, soil erosion and response studies of small watersheds to name a few. However, fine resolution rainfall data are not often available, especially in developing and under-developed countries, mainly because of the high cost and low reliability of monitoring such data. Rainfall series of considerable length are usually available at coarser resolution (e.g. daily) due to low cost and ease of collection. This situation has made it beneficial as one practical solution to overcome the paucity of fine resolution rainfall data. Fine resolution rainfall distributions can be estimated from those observed data at a lower resolution by using an appropriated stochastic rainfall disaggregation model.

The purpose of such stochastic rainfall disaggregation model is to extend historical records or generate new ones and it's essential that the model preserves important statistical properties of the observed rainfall records, such as the scaling of distributions, intermittency, autocorrelation characteristics and behaviour of extreme events. To this end, a vast number of rainfall disaggregation models have been proposed in the literature on which models are based on many different concepts (reader may refer to Sivakumar and Sharma [1] for some review on the existed disaggregation models). One of the most studied models is multiplicative random cascade models which is based on scale invariance theory (e.g. [2, 3, 4]). The interest in these models is related to their ability to describe the complex rainfall process over a wide range of scales with few parameters, as well as their link to the multifractal theory.

The multiplicative random cascade model, in the beginning, has been introduced into practice for describing geophysical fields [2, 3, 5, 6], and later, due to the encouraging results from modelling the multifractal structure of rainfall fields, the cascade model has been applied to the temporal rainfall disaggregation [7]. Successful tests of a simple random cascade model in temporal rainfall disaggregation have been reported in numerous studies, of which the cascade model showed the ability to simulate realistic fine resolution rainfall series. Some of the studies included Onof *et al.* [8] who has followed Deidda *et al.* [9] in using a discrete random cascade algorithm based on the log-Poisson generator to generate synthetic 5 min rainfall series from the 13 years of hourly rainfall for stations at Heathrow, near London. The cascade model did not only do well in reproducing the temporal structure, but it was also able to preserve the extreme behaviour of observed rainfall series. Molnar and Burlando [10] have adopted the intermittent log-Normal β model proposed by Gupta and Waymire [3] to disaggregate quasi daily rainfall down to 10 min intervals for 20-year record period in Zurich, Switzerland. They compared the performance of the model with a

cascade model with a bounded (i.e. time scale dependent) generator distributed as Beta probability density and found that the intermittent log-Normal β model is better at preserving the distribution of rainfall and its extreme behaviour at the 10-min scale.

Considering the encouraging results of a simple random cascade model to simulate rainfall series while preserving important statistical properties of the observed rainfall records, the present study thus intends to assess utility and suitability of the cascade model as temporal rainfall disaggregation for maritime continent conditions. By far, no study has been conducted in the present study areas on stochastic rainfall disaggregation using random cascade models. Two earlier studies only focused on the investigation of scaling behaviour for rainfall series in Singapore [11, 12]. A simple random cascade model will be employed to disaggregate daily rainfall data down to hourly rainfall data for seven rain gauge stations located sparsely in Peninsular Malaysia. The prime focus of this study is in the generation of synthetic rainfall time series and comparisons of those with the real observations in the context of rainfall variability and intermittency as well as rainfall extremes behaviour.

2.0 METHODOLOGY

2.1 Rainfall Data

The rainfall data set from seven selected rain gauge stations located in Peninsular Malaysia were obtained from the Department of Irrigation and Drainage Malaysia rainfall archive. The length of recorded years for these stations ranged from 22 to 42 years. The geographical coordinates of the stations along with the period of records are listed in Table 1 and Figure 1 show the locations of those stations.

Table 1 Geographic coordinates and period of records for seven rain gauge stations in Peninsular Malaysia

Stations	Latitude	Longitude	Period of records
Kuala Nerang	6.25	100.61	1985-2012
Tanjung Malim	3.68	101.52	1979-2012
Kajang	3.00	101.79	1976-2012
Raub	3.81	101.85	1979-2012
Pekan Nenas	1.52	103.49	1979-2012
Dungun	4.76	103.42	1971-2012
Kuala Krai	5.31	102.28	1990-2012

2.2 Multiplicative Random Cascade Disaggregation Model

The rationale of the discrete multiplicative random cascade framework is that a measurement (e.g. rainfall intensity, R) defined on a specific support at a given coarser scale can be recursively redistributed across finer scales through a multiplication process. The theory of random cascade arose in the statistical theory of turbulence, where its development was motivated by a

desire to model the fields and series that have statistically scale-invariant properties and it was put in the basic discrete form used here by Mandelbrot [13]. The ability of random cascade models to reproduce the structures in observed rainfall and its statistical properties such as the scaling of distributions, intermittency, and characteristics of extreme events, has justified their rapid implementation in numerous studies on generating fractal or multifractal rainfall fields or time series under different climatic conditions.

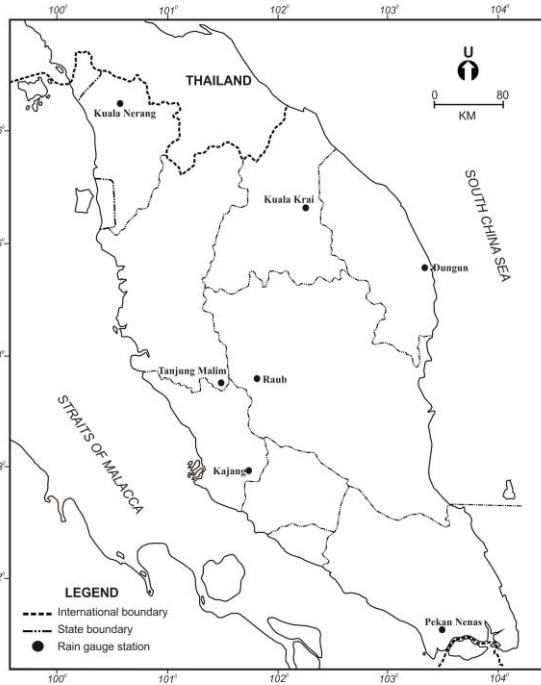


Figure 1 Map of Peninsular Malaysia showing the geographical regions and the location of the selected seven rain gauge stations

For temporal rainfall, the random cascade model starts by distributing the initial mass (in the present study, we used rainfall intensity, R) occurring at the coarse time scale, L among several subintervals of fixed size. The number of subintervals is defined by the branching number b , which often for simplicity (as done in this study), is set equal to 2. Figure 2 represents the basic structure of the random cascade model. At each cascade level, each segment is divided into b equal parts and the mass in each part is the product of the previous mass on the corresponding time scale and a non-negative weighted value derived from a specified distribution, which is known as the cascade generator W . The rainfall at the n -th cascade level at position j in time series can be expressed as:

$$R_{j,n} = R_0 \prod_{i=1}^n W_{f(i,j),i} \tag{1}$$

where for branching $b = 2$, $j = 1, 2, \dots, 2^n$. The function $f(i, j)$ indexes the position of the time interval at the i -th cascade level and is given by rounding up $j/2^{n-1}$ to the nearest integer.

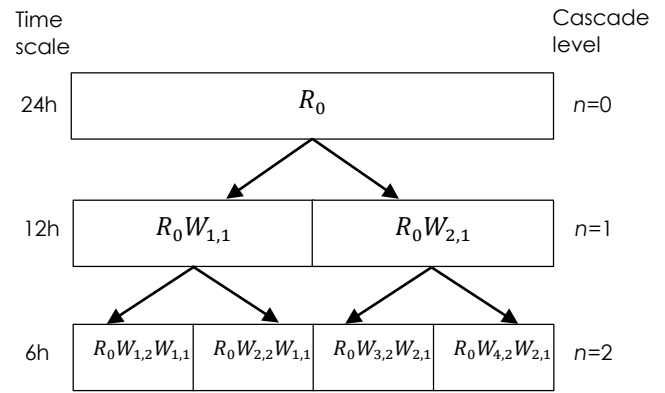


Figure 2 Example of the discrete MRC with branching number $b=2$ and cascade generator W for scales $n=0, 1$, and 2

2.3 Statistical Moment Scaling

The properties of the cascade generator W for cascade models used in the present study can be estimated using the statistical moment which characterized the fractal or scale-invariant behaviour of the multiplicative cascade process across scales. The statistical moment analysis of rainfall time series is considered in a non-dimensional framework as done by de Lima [14]. The time scale is nondimensionalized by a scale ratio $\lambda = L/l$ where it is defined as a ratio of the largest scale of interest, L (e.g. 32h) to the smallest scale, l (e.g. 1h). l is varying from 1h to L h by a factor of 2, e.g. $l = 1, 2, 4, 8, 16, 32$. While the dimensionless rainfall intensity ε_λ is obtained by dividing the intensity on a time scale of resolution λ , $R_{\lambda,i}$ and the ensemble average intensity of the process, $E(R_\lambda) = \sum_{i=1}^\lambda R_{\lambda,i} / \lambda$:

$$\varepsilon_\lambda = \frac{R_{\lambda,i}}{E(R_\lambda)} ; i = 1, \dots, \lambda \tag{2}$$

The statistical moments $M_q(\lambda)$ of (arbitrary) order q were estimated for each time resolution λ by:

$$M_q(\lambda) = E(\varepsilon_\lambda^q) \approx \lambda^{K(q)} \tag{3}$$

where $K(q)$ is the scaling exponent function, which can be obtained by estimating the slope of $M_q(\lambda)$ versus λ in the logarithmic domain for various moment orders q . The shape of $K(q)$ is linked to the q -moments of the random number generator W by:

$$K(q) = \log_b E(W^q) \tag{4}$$

The scaling exponent function $K(q)$ can be given a theoretical expression depending upon the nature of the random cascade. In the present study, we choose two theoretical distributions of cascade generator W in the light of the encouraging results by earlier studies as well as their appeal from a practical viewpoint since they are relatively parameter parsimonious. The two distributions are log-Poisson (e.g. [8, 15, 16]) and log-Normal (e.g. [7, 10, 17]). The distribution of W is assumed to be independent and identically distributed with $E(W) = 1$, so that mass, on the average, conserved from level to level in the cascade development. These

two cascade models are differed in the way they deal with the intermittency in rainfall.

2.3.1 log-Poisson Model

The cascade generator W for log-Poisson model is given by:

$$W = A\beta^N \quad (5)$$

where $P(N = m) = \frac{c^m e^{-c}}{m!}$. Since the condition of the mean intensity requires that $E(W) = 1$, therefore this leads to $A = e^{c(1-\beta)}$ which leaves only two independent parameters β and c that are needed to identify. The theoretical expression of $K(q)$ for log-Poisson cascade is given as:

$$K(q) = c \frac{q(1-\beta) + \beta^q - 1}{\ln(2)} \quad (6)$$

A feature of log-Poisson cascade is it is always equal to positive values. This means that this cascade generator does not generate any zeros. Therefore, this study needs to provide a method for generating zeros rainfall. Since the data were recorded from a tipping bucket measurement, thus we adopted approach by Onof and Arnbjerg-Nielsen [18] wherein a tipping bucket is used as a filter for the generated hourly rainfall series. The simulated hourly rainfall data is set to 0 if the total depth in that interval is below a given threshold (i.e. threshold is set to 0.1 mm), and the amount being subtracted is then added to the next hourly interval. Implementing this approach will avoid the problems of systematic errors from the tipping bucket rain gauge influencing the results, such as bias in the mean rainfall intensity [18].

2.3.2 log-Normal β Model

The cascade generator W in log-Normal β model is written as a product of two independent random variables, $W = BY$ where B is a random variable that controls the intermittency in rainfall and Y is a strictly positive random variable that represents the weight for assigning a rainfall amount in an interval. The rain and non-rain interval in a rainfall process is determined based on the following probabilities that derived according to the Bernoulli distribution:

$$P(B = 0) = 1 - b^{-c} \text{ and } P(B = b^c) = b^{-c} \quad (7)$$

where c is a parameter and $E(B) = 1$. Variability in the positive part of the generator is obtained from the log-Normal distribution, and with the condition $E(Y) = 1$, the expression is given as:

$$Y = b^{\frac{-\sigma^2 \ln b}{2} + \sigma X} \quad (8)$$

where X is a normal $N(0,1)$ random variable and σ^2 is a parameter defining the variance of Y . The generator W is then distributed as:

$$P(W = 0) = 1 - b^{-c} \text{ and } P(W = BY = b^c \cdot b^{\frac{-\sigma^2 \ln b}{2} + \sigma X}) = b^{-c} \quad (9)$$

and the scaling exponent function of W is:

$$K(q) = c(q-1) + \frac{\sigma^2 \ln b}{2}(q^2 - q) \quad (10)$$

2.4 Goodness-of-Fit Tests (GOF)

Three different goodness-of-fit tests (GOF) have been used as a measure of the relative quality of log-Poisson and log-normal β models in disaggregating hourly rainfall time series. The three tests, namely root mean squared error (RMSE), Nash-Sutcliffe efficiency (NSE) and percent bias (pbias) are among common GOF tests used to evaluate the model performance in hydrological modelling. The tests are based on the degree of discrepancy between observed and simulated values at hourly resolution. The formula of the GOF tests are given as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (S_i - O_i)^2}{n}} \quad (11)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (S_i - O_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (12)$$

$$\text{where } \bar{O} = \frac{\sum_{i=1}^n O_i}{n}$$

$$pbias = \frac{\sum_{i=1}^n (S_i - O_i)}{\sum_{i=1}^n O_i} \times 100 \quad (13)$$

where S_i is the i^{th} hourly simulated series, O_i is the i^{th} hourly observed series and n is the length of the hourly rainfall series. The preferred distribution is chose based on the minimum value of RMSE and the maximum value of NSE ($-\infty < NSE \leq 1$). Meanwhile, the optimum value of pbias is 0, with low-magnitude values indicate accurate model simulation. The positive values of pbias indicate the overestimation while the negative values indicate the underestimation of the model.

2.5 Analysis Setup

Our analysis setup for testing both of log-Poisson and log-Normal β models in the context of their ability for the disaggregation of coarse resolution rainfall data into synthetic fine resolution was as follows:

1. The observed rainfall series at the 1h resolution, defined as the scale $\lambda = 32$, was aggregated with $b = 2$ up to 32h resolution, where the scale is $\lambda = 1$. At the 1h resolution, the standard statistics that considered as important and necessary to be fulfilled for an adequate model in water resource studies were computed. These included mean, standard deviation, and dry proportions. We also computed the annual maxima for extreme value analysis which later will be analysed to examine the ability of the models used here to reproduce the distribution of annual extremes.
2. The statistical moment analysis was done for each rain gauge stations in a non-dimensional framework. The parameters of each cascade model will be estimated by fitting the scaling exponent function derived from the data to a theoretical form. The fitting is carried out by using the least squares method. If we denote $K_d(q)$ as the empirical scaling exponent function and $K_T(q)$

as the theoretical expression of each model, then the parameters are found with the following minimization program:

$$\min \sum_{j=1}^n [K_T(q_j) - K_d(q_j)]^2 \quad (14)$$

where n is the number of values of q -moments used in the fitting.

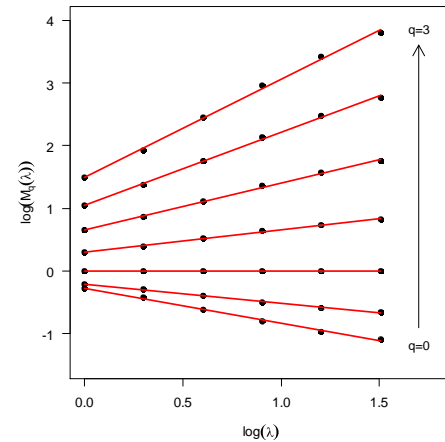
3. With the parameters for each cascade models, historical daily rainfall (aggregated from hourly data) were then disaggregated to hourly rainfall for 100 realizations. In this study, the daily data at 1440 min resolution (i.e. daily) was subdivided to 45 min after five successive disaggregation levels. The hourly rainfall series was obtained by performing linear interpolation from 45 min to the nearest 1h time step [19].
4. The performances of the cascade models in rainfall disaggregation were first evaluated based on the values of the three GOF criteria. To further evaluate the models' performance, the standard statistics and the behaviour of rainfall extremes in the simulated hourly data were analysed. The results of both cascade models were compared and statement regarding the utility and suitability of the models for rainfall disaggregation at the studied stations were made.

3.0 RESULTS AND DISCUSSION

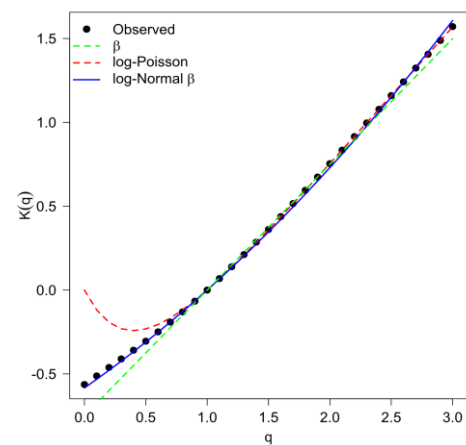
3.1 Parameter Estimation

Statistical moments for each scale λ were calculated per Eq. (3) and owing to uncertainties in the estimation of high order moments [14], the moment analysis was limited to orders q in the range $0 \leq q \leq 3$. A relationship between $M_q(\lambda)$ versus λ in the logarithmic domain for each q -moment was then developed. The slope of each fitted line was determined by linear regression as the function of scaling exponent $K(q)$. Figure 3 displays an example of the statistical q -moments with scale ratio λ on log-log plot and the empirical relation of $K(q)$ versus q , which presents the results obtained for Tanjung Malim rainfall data. The shape of the empirical $K(q)$ in Figure 3(b) was slightly convex, rather than a straight line. This is an indication of the presence of multifractal behaviour in the rainfall series in the studied station. In addition, the departure from monofractal behaviour in the data was evident (i.e. the theoretical $K(q) = c(q - 1)$ of monofractal β model would be shown as straight line [10]).

The parameters of log-Poisson and log-Normal β models were estimated from the $K(q)$ function given in Eq. (6) and Eq. (10), respectively, using the least square method in Eq. (11). For log-Poisson model, the estimation of $K(q)$ for values of q smaller than 1 is not of high quality, so these values were not considered in the fit. The fitted $K(q)$ for each model was then plotted along with the empirical $K(q)$ as shown in Figure 3(b). Table 3 presents the parameter values of each



(a) Moment scaling relationships



(b) The scaling exponent function, $K(q)$

Figure 3 An example for the observed data at the Tanjung Malim station

cascade models for rainfall series from all the stations studied herein. These parameters were computed from the entire historical records that were constant.

Table 3 Parameter values for the log-Poisson and log-Normal β model for all studied stations

Stations	log-Poisson		log-Normal β	
	c	β	c	σ^2
Kuala Nerang	0.580	0.083	0.586	0.151
Tanjung Malim	0.614	0.077	0.583	0.212
Kajang	0.578	0.046	0.643	0.140
Raub	0.598	0.074	0.619	0.154
Pekan Nenas	0.526	0.039	0.612	0.108
Dungun	3.264	0.649	0.452	0.271
Kuala Krai	0.462	0.016	0.586	0.070

3.2 Assessment of the Random Cascade Models

The performance of log-Poisson and log-Normal β models, for all the studied stations, were first evaluated using RMSE, NSE and pbias GOF tests. Results of the tests

are presented in Table 4. Based on the values of RMSE and NSE criteria, the log-Poisson model seems to be the preferred distribution for rainfall disaggregation as compared to the log-normal β model, except for Dungun station. The former model gave lower RMSE values and higher NSE values than the latter model. Results of the pbias test indicate that both models are either overestimated or underestimated the hourly rainfall data. The result of the pbias test, however, inconsistent with the RMSE and NSE tests. Thus, in this study, we chose the best suited distribution based on the RMSE and NSE values, while the pbias test is used to measure the degree of overestimation or underestimation by the models.

Table 4 The results of the goodness-of-fit tests (GOF) for log-Poisson and log-Normal β models for all studied stations

Stations	GOF tests	log-Poisson	log-Normal β
Kuala Nerang	RMSE	0.122	0.178
	NSE	0.994	0.986
	pbias	0.271	0.103
Tanjung Malim	RMSE	0.146	0.375
	NSE	0.995	0.968
	pbias	0.160	0.063
Kajang	RMSE	0.145	0.291
	NSE	0.994	0.977
	pbias	-0.129	0.031
Raub	RMSE	0.133	0.192
	NSE	0.993	0.983
	pbias	-0.073	0.124
Pekan Nenas	RMSE	0.125	0.216
	NSE	0.995	0.985
	pbias	-0.103	-0.291
Dungun	RMSE	1.089	1.058
	NSE	0.762	0.776
	pbias	0.129	-0.210
Kuala Krai	RMSE	0.182	0.258
	NSE	0.990	0.980
	pbias	0.448	-0.460

The performance of the cascade models was further evaluated to assess the models' ability to reproduce a number of important hourly rainfall statistics estimated from the observed records. This study will assess and compare how realistically each model simulates hourly rainfall variability, intermittency and rainfall extremes.

3.2.1 Rainfall Variability

Both log-Poisson and log-Normal β models were able to reproduce the mean of hourly rainfall amount at all rain gauge stations (see Table 5 and Figure 4). These results were expected since the mean is known to be well reproduced on theoretical grounds of cascade models. The log-Normal β model was slightly underestimated the mean at the hourly resolution for Dungun station. For Kuala Krai, the simulated mean of both cascade models was obviously differed from the observed mean, such that the log-Poisson model overestimated the mean, while the log-Normal β model underestimated the mean of hourly rainfall amount.

Table 5 Observed and model generated rainfall variability and dry proportion for hourly data across all studied rain gauge stations. Rainfall variability and dry proportion of cascade models are computed from average of 100 realizations, with variations between model runs shown in Figure 4-6

	Observed	log-Poisson	log-Normal β
<i>Mean</i>			
Kuala Nerang	0.205	0.206	0.206
Tanjung Malim	0.286	0.286	0.286
Kajang	0.244	0.244	0.244
Raub	0.203	0.203	0.203
Pekan Nenas	0.259	0.259	0.258
Dungun	0.281	0.281	0.280
Kuala Krai	0.258	0.259	0.257
<i>Standard deviation</i>			
Kuala Nerang	1.639	1.643	1.605
Tanjung Malim	2.149	2.142	2.054
Kajang	1.977	1.989	1.936
Raub	1.613	1.625	1.587
Pekan Nenas	1.864	1.875	1.828
Dungun	2.259	2.219	2.467
Kuala Krai	1.895	1.877	1.827
<i>Dry proportion</i>			
Kuala Nerang	0.936	0.884	0.924
Tanjung Malim	0.920	0.865	0.909
Kajang	0.939	0.897	0.932
Raub	0.937	0.886	0.925
Pekan Nenas	0.930	0.885	0.921
Dungun	0.923	0.775	0.886
Kuala Krai	0.927	0.895	0.913

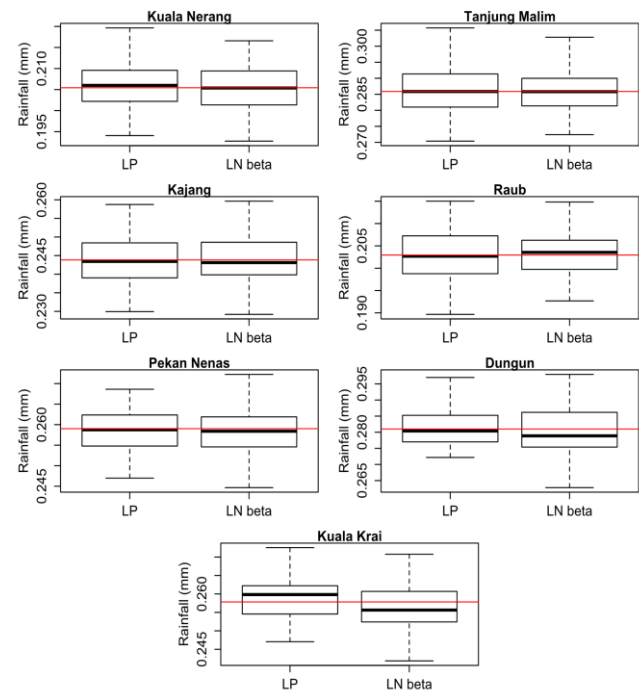


Figure 4 Mean of hourly rainfall amount across all studied stations, red solid line denoted observed mean

Figure 5 shows that log-Poisson model was better at preserving the standard deviation of rainfall series at hourly time scale than the log-Normal β model. The log-Poisson cascade model able to reproduce the observed standard deviation for most of the stations with only a slightly overestimated for Kajang and Pekan Nenas stations and underestimated for Dungun and Kuala Krai stations.

Meanwhile, the log-Normal β model unable to perform well at all the studied stations by consistently underestimating the hourly standard deviation, except at Dungun station where the model overestimated it. The poor simulation of the standard deviation by the log-Normal β model could be reasoned by the possibility of simulating two zero weights for a rainy period, and it could also be the inability of the variability parameter, σ^2 to adequately estimate the rainfall variability.

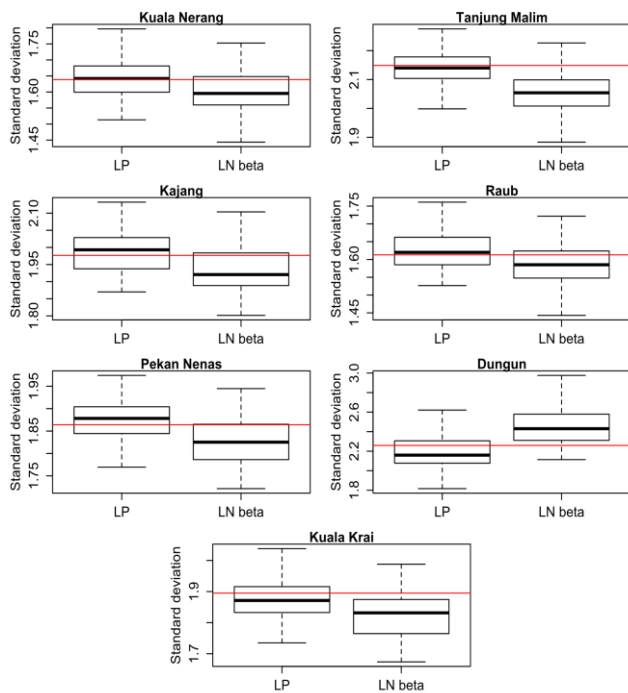


Figure 5 Standard deviation for hourly data across all studied stations, red solid line denoted observed standard deviation

3.2.2 Dry Proportions

The log-Poisson and log-Normal β models in general did not produce rainfall series with proper rainfall intermittency at the hourly resolution with underestimation for all the stations (see Figure 6). However, log-Normal β model able to generate the number of dry spells closer to the observed than log-Poisson model since the former model considered rainfall intermittency in its framework. The log-Poisson model, on the other hand, only generates positive values and the zeroes rainfall are obtained from the threshold mechanism described in Methodology section.

3.2.3 Extreme Value Analysis

In the present study, we analysed the behaviour of rainfall extremes using the intensity-frequency curves. It was formed based on an empirical cumulative distribution of the annual maxima which was extracted from observed and simulated rainfall series [17]. The annual maxima were ranked from the highest to the lowest and the annual exceedance probability (AEP) are estimated as:

$$AEP(m) = \frac{m}{N + 1} \quad (15)$$

where m is the rank and N is the length of record. The estimated AEPs were then plotted against the corresponding log-transformed rainfall intensities. Confidence bands for the simulated intensity-frequency curves for each model were constructed using results from 100 realizations and compared with those derived from observed curves.

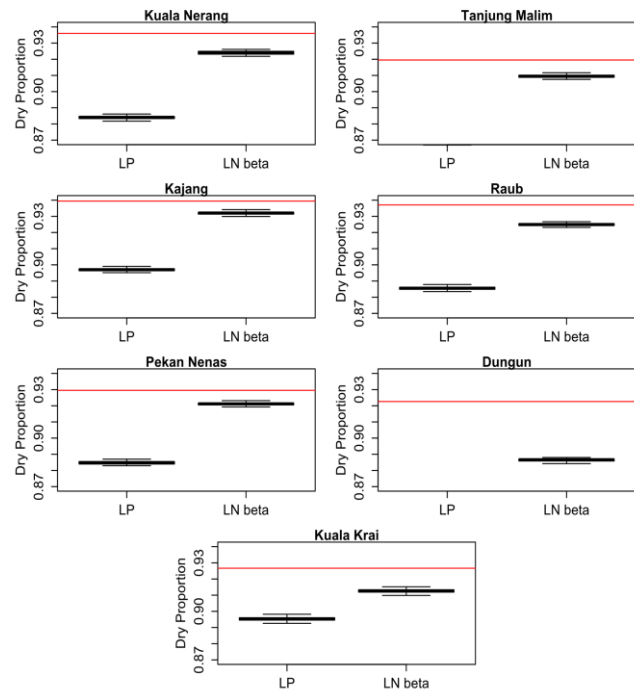


Figure 6 Dry proportion for hourly data across all studied stations, red solid line denoted observed dry proportion

Figure 7 shows the intensity-frequency curves for each model versus observed values at each studied stations. Interesting results were observed, such that the observed rainfall maxima for most of the stations fall within the simulated bounds of both log-Poisson and log-Normal β models even at low AEPs, except for Dungun and Kuala Krai stations. For Kuala Nerang station, the log-Normal β model reproduced the observed intensity-frequency curves well compared to the log-Poisson model which slightly underestimated the curves at middle up to low return periods. Meanwhile, for Tanjung Malim station, the performance of the log-Poisson model was better than the log-

Normal β model, particularly, at high return periods. Both models were slightly underestimated the hourly intensity-frequency at low return period. Nevertheless, the observed values still fall within the simulated bounds of these two models.

For Kajang and Raub stations, both cascade models performed reasonably well, except for a slight overestimation of high return periods by log-Normal β model. For Pekan Nenas station, the log-Poisson model was able to reproduce the observed intensity-frequency curves well at high return periods, but done poorly at low return period with an underestimation of the curves. Conversely, the log-Normal β model performed well at low return periods and a slight overestimation at high return periods. In Dungun station, both models gave about similar simulated intensity-frequency curves, with the observed values fall within the simulated bounds only at low return periods, while at high up to middle return periods, both models overestimated the curves.

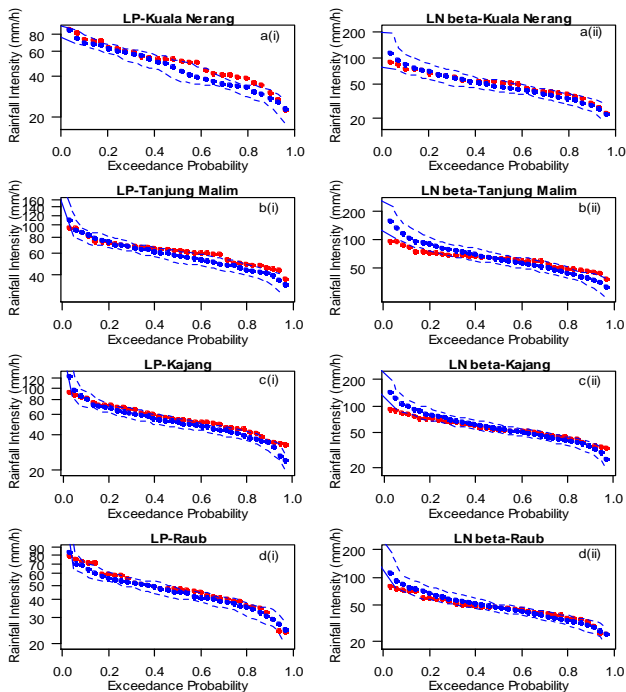


Figure 7 Empirical intensity-frequency curves derived from observed and cascade models for hourly data, with columns denoting cascade models (LP-log-Poisson and LN beta-log-Normal β) and rows denoting rain gauge stations. Red dots denoting observed intensity-frequency behaviour and blue dots and blue dashed lines denoting the median and 5th and 95th percentiles of 100 realizations, respectively

For Kuala Krai station, the log-Normal β model reproduced the curves rather well, apart from a slight underestimation of low return periods. In contrast, the log-Poisson model performed poorly with an underestimation of the hourly intensity-frequency from the middle to lower tails of the extreme value distribution. In view of these results, the log-Poisson model appeared to be fitted as a disaggregator tool

for Tanjung Malim, Kajang and Raub rainfall data, while the log-Normal β model was good in reproducing the rainfall series and extremes of the observed for Kuala Nerang and Pekan Nenas. However, both log-Poisson and log-Normal β models failed to simulate the extremes well for Dungun and Kuala Krai stations.

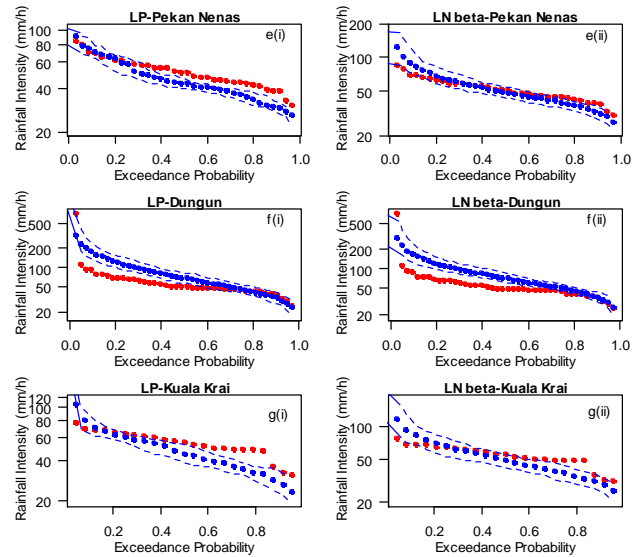


Figure 7 continued

4.0 CONCLUSION

The present study investigated the utility and suitability of a simple discrete multiplicative random cascade models for disaggregation of rainfall series from daily time scale to the hourly time scale in Peninsular Malaysia climate conditions. Two different distributions for cascade generator were used here, namely log-Poisson and log-Normal β .

Based on the RMSE and NSE goodness-of-fit tests, the log-Poisson model seems to be the best-suited model for all the stations studied herein, except for Dungun station which preferred log-Normal β model for rainfall disaggregation. The performances of the cascade models were further evaluated based on their ability to obtain realistic hourly rainfall series. In general, both log-Poisson and log-Normal β models are good at preserving the mean of rainfall at the hourly time scale. In terms of standard deviation, the log-Poisson model performs better at reproducing the variability of observed rainfall than the log-Normal β model. However, in terms of dry proportions, both models apparently failed to reproduce the rainfall intermittency. Nevertheless, the log-Normal β model does better than the log-Poisson model.

In terms of rainfall extremes, a fairly good performance by log-Poisson and log-Normal β models were shown for most of the studied stations, except for Dungun and Kuala Krai stations which both located in the east part of Peninsular Malaysia. Both models unable to reproduce the behaviour of observed rainfall extremes in those stations. This could be due to the

different climate experiencing in both stations where the rainfall in the east regions is very much influenced by the seasons. Perhaps by taking into consideration the seasonal factors in parameterization may improve the multiplicative random cascade models' performance, particularly in the aspect of extreme value simulation.

Overall, the performance of log-Poisson and log-Normal β models on the reproduction of some important rainfall or hydrological statistics can be judged as satisfactory, given their parsimonious nature. Although the present study only involved seven rain gauge stations in Peninsular Malaysia, the information gained from the study can at least guide engineers in water resource fields in making decisions on the suitable disaggregation model for the simulation of rainfall series at fine resolution using data from coarse resolution.

For the future research, the study may apply datasets from other rain gauge stations in Peninsular Malaysia to verify our findings on the suitability of cascade models, as well as employ microcanonical version of multiplicative random cascade model where the distribution of cascade weights is assumed to be dependent on time scale [7, 10, 20], on rainfall intensity and time scale [21, 22], and on interval class, i.e. intervals at the beginning, middle, or end of a rainfall event [4, 19].

Acknowledgement

The authors express gratitude to the staff of Malaysian Meteorology Department for providing rainfall data and Universiti Kebangsaan Malaysia for providing the facilities for this study. This study would not have been possible without the sponsorship from the Ministry of Higher Education (MOHE) and Universiti Teknologi Mara.

References

- [1] Sivakumar, B. and A. Sharma. 2008. A Cascade Approach to Continuous Rainfall Data Generation at Point Locations. *Stochastic Environmental Research and Risk Assessment*. 22(4): 451-459.
- [2] Schertzer, D. and S. Lovejoy. 1987. Physical Modeling and Analysis of Rain and Clouds by Anisotropic Scaling Multiplicative Processes. *Journal of Geophysical Research*. 92: 9692-9714.
- [3] Gupta, V. K. and E. C. Waymire. 1993. A Statistical Analysis of Mesoscale Rainfall as a Random Cascade. *Journal of Applied Meteorology*. 32(2): 145-164.
- [4] Olsson, J. 1998: Evaluation of a Scaling Cascade Model for Temporal Rainfall Disaggregation. *Hydrology and Earth System Sciences*. 2(1): 19-30.
- [5] Over, T. M. and V. K. Gupta. 1994. Statistical Analysis of Mesoscale Rainfall: Dependence of a Random Cascade Generator on Large-Scale Forcing. *Journal of Applied Meteorology*. 33: 1526-1542.
- [6] Over, T. M. and V. K. Gupta. 1996. A Space-Time Theory of Mesoscale Rainfall Using Random Cascades. *Journal of Geophysical Research*. 101(D21): 26319-26331.
- [7] Licznar, P., J. Łomotowski, and D. E. Rupp. 2011. Random Cascade Driven Rainfall Disaggregation for Urban Hydrology: An Evaluation of Six Models and a New Generator. *Atmospheric Research*. 99: 563-578.
- [8] Onof, C., J. Townend, and R. Kee. 2005. Comparison of Two Hourly to 5-min Rainfall Disaggregators. *Atmospheric Research*. 77: 176-187.
- [9] Deidda, R., R. Benzi, and F. Siccaldi. 1999. Multifractal Modeling of Anomalous Scaling Laws in Rainfall. *Water Resources Research*. 35(6): 1853-1867.
- [10] Molnar, P. and P. Burlando. 2005. Preservation of Rainfall Properties in Stochastic Disaggregation by a Simple Random Cascade Model. *Atmospheric Research*. 77: 137-151.
- [11] Sivakumar, B. 2000. Fractal Analysis of Rainfall Observed in Two Different Climatic Regions. *Hydrological Sciences Journal*. 45(5): 727-738.
- [12] Mandapaka, P. V. and X. Qin. 2015. A Large Sample Investigation of Temporal Scale-Invariance in Rainfall Over the Tropical Urban Island of Singapore. *Theoretical and Applied Climatology*. 122(3): 685-697.
- [13] Mandelbrot, B. B. 1974. Intermittent Turbulence in Self-Similar Cascades – Divergence of High Moments and Dimensions of Carrier. *Journal of Fluid Mechanics*. 62(2): 331-358.
- [14] de Lima, M. I. P. 1998. Multifractals and the Temporal Structure of Rainfall. Doctoral Dissertation. Wageningen Agricultural University, Wageningen.
- [15] Gaume, E., N. Mouhous, and H. Andrieu. 2007. Rainfall Stochastic Disaggregation Models: Calibration and Validation of a Multiplicative Cascade Model. *Advances in Water Resources*. 30: 1301-1319.
- [16] Mascaro, G., E.R. Vivoni, D.J. Gochis, C.J. Watts, and J.C. Rodriguez. 2014. Temporal Downscaling and Statistical Analysis of Rainfall Across a Topographic Transect in Northwest Mexico. *Journal of Applied Meteorology and Climatology*. 53: 910-927.
- [17] Pui, A., A. Sharma, R. Mehrotra, B. Sivakumar, and E. Jeremiah. 2012. *Journal of Hydrology*. 470-471: 138-157.
- [18] Onof, C. and K. Arnbjerg-Nielsen. 2009. Quantification of Anticipated Future Changes in High Resolution Design Rainfall for Urban Areas. *Atmospheric Research*. 92: 350-363.
- [19] Guntner, A., J. Olsson, A. Calver, and B. Gannon. 2001. Cascade-Based Disaggregation of Continuous Rainfall Time Series: The Influence of Climate. *Hydrology and Earth System Sciences*. 5(2): 145-164.
- [20] Menabde, M. and M. Sivapalan. 2000. Modeling of Rainfall Time Series and Extremes Using Bounded Random Cascades and Levy-Stable Distributions. *Water Resources Research*. 36(11): 3293-3300.
- [21] Rupp., D. E., R. F. Keim, M. Ossiander, M. Brugnach and J. S. Selker. 2009. Time Scale and Intensity Dependency in Multiplicative Cascades for Temporal Rainfall Disaggregation. *Water Resources Research*. 45(7): W07409.
- [22] Serinadi, F. 2010. Multifractality, Imperfect Scaling and Hydrological Properties of Rainfall Time Series Simulated by Continuous Universal Multifractal and Discrete Random Cascade Models. *Nonlinear Processes in Geophysics*. 17: 697-714.