Environ Earth Sci (2014) 71:3627–3640
DOI 10.1007/s12665-013-2755-7

ORIGINAL ARTICLE

Disaggregation of daily rainfall data using Bartlett Lewis Rectangular Pulse model: a case study in central Peninsular Malaysia

Zulkifli Yusop · Harisaweni Nasir · Fadhilah Yusof

Received: 2 August 2012 / Accepted: 21 August 2013 / Published online: 8 September 2013

Abstract Short duration rainfall data are required for certain hydrological risk assessments. However, short timescale rainfall intensity records are still scarce due to the high cost and low reliability of the monitoring systems. One way to solve this problem is by disaggregating rainfall data using stochastic methods. This study used the Bartlett Lewis Rectangular Pulse model to disaggregate daily rainfall into hourly rainfall for ten stations in the central region of Peninsular Malaysia. The performance of the model was evaluated on its ability to reproduce statistical properties, namely the mean and standard deviation, derived from the historical records over the disaggregated rainfall. The disaggregation of daily to hourly rainfall produced daily and hourly means that closely matched the historical records. However, the standard deviations of the disaggregated daily rainfall were lower than the historical values. Despite the significant differences in the standard deviation, both data series exhibit similar patterns and the model adequately preserved the trends of all the statistical properties used in evaluating its performance.

Keywords Disaggregation · Daily rainfall · Hourly rainfall · Bartlett Lewis Rectangular Pulse

Introduction

Most hydrological risk assessment techniques require information on rainfall data over a short timescale. However, such data are still limited worldwide due to the high cost of monitoring and low reliability of monitoring systems. Alternatively, an indirect technique to obtain short timescale rainfall data is possible using stochastic rainfall disaggregation model (Koutsoyiannis et al. 2003).

Stochastic models can generate rainfall data from a long timescale into short timescale using a disaggregation technique (Koutsoyiannis et al. 2003). Disaggregation techniques can also be used to improve the resolution and quality of other data that is needed for hydrological analysis, such as in weather data generation (Debele et al. 2007), stream flow data generation (Kumar et al. 2000; Nowak et al. 2010) and disaggregation of land types to develop a conceptual hydrologic hill slope response (Van Zijl 2010).

The stochastic disaggregation method was originally developed by Valencia and Schaake (1973) and was later modified and extended by Mejia and Rouselle (1976), (Lane 1982) and Stedinger et al. (1985). In the initial model by Valencia and Schaake (1973), the model structure was designed to preserve the variance and covariance among the seasonal values. However, this model requires a large number of parameters, yet it is not able to consistently preserve the moments. In the extended model by Mejia and Rouselle (1976), an additional term was included in the model to preserve the seasonal covariance between years, but the problem of inconsistent causal structure and excessive number of parameters still remained unresolved.

Recently, short timescale rainfall data is highly sought after for use in hydrological risks assessments because finer input data would generate a better modelling output. Therefore, indirect techniques to obtain short timescale rainfall data such as stochastic rainfall disaggregation techniques cannot be avoided. There are several methods available to generate short timescale rainfall data and

Z. Yusop · H. Nasir (✉) · F. Yusof
Water Research Alliance c/o Institute Environmental and Water Resource Management (IPASA), Faculty of Civil Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
e-mail: harisaweni@yahoo.com

researches along this line are still on-going in many parts of the world. One of the most proven methods is the Bartlett Lewis Rectangular Pulse (BLRP) model (Koutsoyiannis and Onof 2001; Koutsoyiannis et al. 2003; Gyasi-Agyei 2005; Gyasi-Agyei and Mahbub 2007). The BLRP model is a clustered point process model in time, and physically it is a realistic rainfall model (Koutsoyiannis and Onof 2001; Koutsoyiannis et al. 2003; Gyasi-Agyei 2005; Gyasi-Agyei and Mahbub 2007).

Bo and Islam (1994) used the modified BLRP model developed by Rodriguez-Iturbe et al. (1988) to capture the statistics of short timescale rainfall from observed daily rainfall in Arno basin central Italy. The model parameters were derived from the mean, standard deviation, autocorrelation and probability of dry days for simulating the sequence of rainfall events for any desired timescale. Glasbey et al. (1995) applied the modified BLRP model to disaggregate daily rainfall data by conditional simulation in Edinburgh, United Kingdom. They found that the generated hourly rainfalls are consistent with the historical daily totals. The statistical parameters for model calibration were derived from hourly data without considering seasonality.

In the early 2000s, the BLRP model was further improved by adding a proportional adjusting procedure to correct the time series simulation; so that the errors between the sums of the generated lower-level series and the corresponding higher-level variable are reduced (Koutsoyiannis and Onof 2001). Gyasi-Agyei (2005) regionalised the BLRP based on a hybrid model for daily rainfall disaggregation in Central Queensland, Australia. Koutsoyiannis et al. (2003) introduced multivariate approach by incorporating the spatial and temporal nonstationary rainfall data for multisite disaggregation of daily rainfall in Brue catchment in South-Western England. Gyasi-Agyei and Mahbub (2007) improved the earlier model and applied it for a very large region in Australian continent. They also enhanced the capping procedure and generalise it for any short timescale at each point. As a result, the model was successfully applied for disaggregating daily rainfall data throughout Australia into a finer timescale of 6 min.

Thus far, most of the rainfall disaggregation analyses using BLRP model were carried out in temperate or relatively dry regions. Only recently, Hanaish et al. (2011) has tested the model in Peninsular Malaysia using hourly rainfall data from a single station, in Petaling Jaya. Further work is still necessary to validate the performance of BLRP model in the humid tropic sites in view of discernible differences in their rainfall characteristics. Tropical rainfall is characterised by heavy burst over short period of time and higher number of rainy day in a year (Noguchi et al. 2003). This study evaluates the performance of BLRP model for disaggregating daily into hourly rainfall from ten stations in the central region of Peninsular Malaysia.

Methodology

Source of data

This study was carried out in Damansara, in the central region of Peninsular Malaysia (Fig. 1). The area receives heavy rainfall during the inter-monsoon periods in March/April and September/October with an average annual rainfall of 2,266 mm and monthly rainfall ranging from 120 to 280 mm. The monthly rainfall seldom dips below 100 mm. The rainfall pattern in Peninsular Malaysia is mainly influenced by two monsoons, namely the southwest monsoon (May–August) and the northeast monsoon (November–February). Rainfall data from ten stations were selected because their long-term rainfall record was the most complete (Table 1).

The selected rainfall stations are being maintained by the Department of Irrigation and Drainage, Malaysia. Daily and hourly rainfall records from 1998 to 2007 were obtained. The hourly data were recorded by automatic rain gauges (Model HOBO ONSET RG2-M) with 0.5 mm resolution whereas the daily rainfall was recorded by storage or manual rain gauges. Missing data from the selected stations are <20 % over the 10-year period. In this study, a wet day is defined as a day when the rainfall amount exceeds or equals a threshold value of 0.5 mm following the definition by Dale (1974).

Data distribution

Fitting of data distribution is a pre-requisite when analysing hydrological data in order to develop valid models of random processes and ensure the reliability of the analysis. This study used five distribution models, namely Gamma, Weibull, Beta, Log Pearson Type 3 and Generalised Pareto. These five distributions were selected because they are the most preferred for fitting the distributions of daily data (Duan et al. 1995; Burgueño et al. 2005; Deidda and Puliga 2006; Olofintoye et al. 2009; Suhaila and Jemain 2007, 2008).

Three goodness of fit tests were used to check the best distribution that could represent the daily rainfall data. They are the Kolmogorov–Smirnov (K–S), Anderson–Darling (A–D), and χ^2 tests. The K–S test is commonly used to decide whether a sample comes from a hypothesised continuous distribution, as follows:

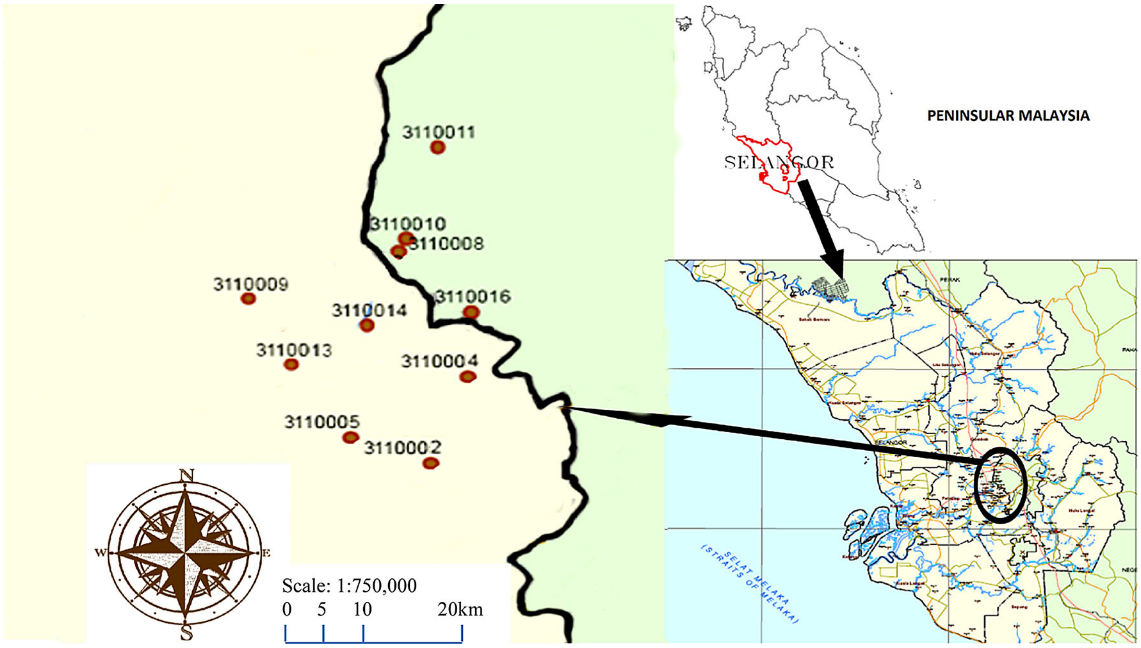


Fig. 1 Locations of the ten rainfall stations in Damansara, Peninsular Malaysia

Table 1 Selected rainfall stations and their coordinates

No	Station name	Station code	Latitude	Longitude
1.	Balai Polis Sea Park	3110004	03 07′19.9″	101 37′45.7″
2.	Balai Polis TTDI	3110010	03 09′8.6″	101 38′7.7″
3.	Bukit Kiara Golf Resort	3110016	03 08′1.5″	101 38′7.7″
4.	Kg Sg Penchala	3110011	03 10′32.4″	101 38′7.7″
5.	SM Sri Permata	3110002	03 06′7.0″	101 36′50.7″
6.	SMK Damansara Jaya	3110014	03 07′49.2″	101 38′7.7″
7.	SR Damansara Utama	3110008	03 08′56.4″	101 38′7.7″
8.	Surau Assyakirin	3110013	03 08′41.8″	101 38′7.7″
9.	Tropicana Golf	3110009	03 08′13″	101 38′7.7″
10	SMK Taman Sea	3110005	03 06′38″	101 38′7.7″

$$D = \max_{1 \le i \le N} \left(F(x_i) - \frac{i-1}{N}, \frac{i}{N} - F(x_i) \right). \quad (1)$$

The K–S test is based on an empirical distribution function with a given N as ordered data points at $x_1, x_2, \ldots, x_N$. F is the theoretical cumulative distribution of the tested distribution, and $x(i)$ is ordered from the smallest to the largest value. The hypothesis that the data fit the theoretical distributional is rejected if the test statistics, D, is greater than the critical value obtained from the K–S table at a 95 % confidence level.

The A–D procedure is a general test to compare the fit of an observed cumulative distribution function to an expected cumulative distribution function, as defined in Eq. (2):

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i-1).[\ln F(X_i) + \ln(1 - F(X_n - i + 1))] \quad (2)$$

where F is the cumulative distribution function of the specified distribution and X_i is the ordered data. The A–D test is a one-sided test and the hypothesis that the distribution is of a specific form is rejected if the test statistics, A, is greater than the critical value.

Meanwhile, the χ^2 test was used to determine whether a sample comes from a population of a similar distribution. This test is applied to binned data, so the value of the test statistic depends on how the data is binned, and this test is only available for continuous sample data. The χ^2 statistic is defined as follows:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \quad (3)$$

where O_i is the observed frequency for bin i, and E_i is the expected frequency for bin i calculated by:

$$E_i = N(F(x_2) - F(x_1)) \quad (4)$$

where F is the cumulative distribution function of the probability distribution being tested, x_1 is the upper limit for class i, x_2 is the lower limit for class i, and N is the sample size.

The idea behind the goodness of fit tests is to measure the distance between the data and the theoretical distribution, and to compare that distance to certain threshold values. If the distance (called the test statistic) is less than the threshold value (the critical value), that distribution is considered to be good. Thus, the distribution with the lowest statistic value is chosen as the best fit.

Bartlett Lewis Rectangular Pulse (BLRP)

The ability of the BLRP model to reproduce important features of rainfall data from an hourly to daily scale and higher has been discussed by Rodriguez-Iturbe et al. (1987) and Onof and Wheater (1994). In addition, BLRP has important features for representing rainfall in continuous time series.

According to Rodriguez-Iturbe et al. (1987), there are five general assumptions when applying a BLRP model: (1) Storm origins t_i occur following a Poisson process with a rate λ; (2) Each storm is characterised by a random number of cells, C ($C \geq 1$), and each storm origin is followed by a Poisson arrival at rate β of the cell's origin; (3) The intervals between successive cells are independent and identically distributed random variables, starting at each t_i, the cell arrival process terminates after an exponentially distributed time v_i with parameter γ; (4) Each cell has an exponentially distributed duration of parameter η; and (5) a uniform intensity X_{ij} with a specified distribution and mean μ_x. This is typically assumed to be exponential (parameter $1/\mu x$), or alternatively a two-parameter gamma with mean μ_x and expected mean square of cell intensity μ_x^2. The number of cells per storm, C, thus has a geometric distribution of mean $\mu_c = 1 + \kappa/\varphi$ where $\kappa = \beta/\eta$ and $\varphi = \gamma/\eta$ are dimensionless parameters (Rodriguez-Iturbe et al. 1987). Therefore, for this BLRP model there are five parameters governing the process: λ, κ, μ_x, φ and η, and all parameters are assumed to be constant (Rodriguez-Iturbe et al. 1987).

The inability to preserve the properties of dry and wet periods is one of weaknesses of cluster-based rectangular pulse models, including the BLRP model (Rodriguez-Iturbe et al. 1987). The model generates fewer wet periods than required, and underestimates the proportion of dry periods (Rodriguez-Iturbe et al. 1987). In order to overcome this problem, Rodriguez-Iturbe et al. (1988) introduced an extra parameter into the model to give a better fit to the statistics. In the extra parameter Rodriguez-Iturbe et al. (1988) placed the mean cell duration, $1/\eta$, which is allowed to vary randomly from storm to storm. The parameter η now follows two-parameter gamma distributions with a shape parameter α and scale parameter v. Subsequently, parameters β and γ also vary such that the ratios $\kappa = \beta/\eta$ and $\varphi = \gamma/\eta$ become constant. Each cell depth is a random constant that is exponentially distributed with mean $E[x]$.This results in a six-parameter model (λ, κ, φ, μ_x, α, and v).

Hyetos

This analysis used *Hyetos* rainfall disaggregation software developed by Koutsoyiannis and Onof (2001). *Hyetos* uses the BLRP model as a background stochastic model for rainfall disaggregation. It uses iteration to derive a synthetic rainfall series, which resembles the given series on a daily scale.

There are three essential steps to run *Hyetos*. The first step is to generate a time series using the BLRP model. In this step, *Hyetos* generates the short timescale by simulating rainfall based on the entered parameters produced from the higher-level time series that is daily rainfall data. The parameters for second-order properties were calculated using four statistical properties, namely mean, variance, lag-1 auto-covariance and probability dry at 1, 24 and 48 h of the historical daily rainfall. Then, the *Hyetos* program requires entering those statistical properties and also the historical daily rainfall time series to generate the hourly rainfall data (Koutsoyiannis and Onof 2000).

In the second step, *Hyetos* computes adjusting procedure to correct the generated short time series so that its terms add up to the corresponding long time series. This computation is to reduce error between the sum of the generated short time series and the corresponding higher-level variable. These procedures are sound because they help preserve certain statistics or in special cases, the entire distribution of the lower-level series even after the adjustment is performed. In the last step, the processes in steps 1 and 2 are repeated to ensure that the generated results are close to the historical values, thereby allowing accurate higher-order statistics to be obtained. The three processes are then repeated until the best values are obtained.

In addition to the four statistical properties, *Hyetos* requires historical daily rainfall data series to simulate and disaggregate hourly rainfall. The performance of the BLRP model in *Hyetos* was examined by plotting hourly simulated result against the hourly observed data. Similarly, the statistical properties (mean and standard deviation) of the disaggregated hourly data were plotted against the corresponding statistics of the hourly historical data.

Parameter estimation

The derivation of parameters for the rectangular pulse stochastic rainfall model from basic statistical properties of rainfall data is not practical, and is probably not the best procedure to use (Rodriguez-Iturbe et al. 1988). Rodriguez-Iturbe et al. (1988) suggested equalizing the characteristic features computed from the observed data with the corresponding model values. In this analysis, the method of moments (Valdes et al. 1977; Isham et al. 1990) was chosen to estimate the parameters by minimising the sum of squares, where the squared terms are the differences between the selected expressions of the model and their equivalent historical sampled values (Cowpertwait et al. 1996). The method of moments was chosen because it has been successfully used to fit historical data into time series models (Bo and Islam 1994; Entekhabi et al. 1989; Onof and Wheater 1993; Rodriguez-Iturbe et al. 1987; Verhoest et al. 1997).

Four statistical properties are required to estimate the parameters of the BLRP model, i.e., the mean, variance, lag-1 auto-covariance coefficient, and proportion of dry periods at 1, 24 and 48 h levels of aggregation. Letting k be the number of parameters to be fitted, p the statistics to be chosen from the historical data to fit the parameters, and these are denoted by the set T, where $T = (t_1, t_2, \ldots, t_p)$, which can include the mean, variance, etc. of various timescales. Then the functions to calculate the various statistics from the parameter values in the BLRP are using the following equations:

$$\text{Mean} = \lambda\mu \times \mu c \frac{\upsilon}{\alpha - 1} T \tag{5}$$

$$\text{Variance} = \frac{2\upsilon^{2-\alpha}T}{\alpha - 2}\left(k_1 - \frac{k_2}{\varphi}\right) - \frac{2\upsilon^{3-\alpha}}{(\alpha-2)(\alpha-3)} \cdot \left(k_1 - \frac{k_2}{\varphi^2}\right) + \frac{2}{(\alpha-2)(\alpha-3)} \cdot \left[k_1(T+\upsilon)^{3-\alpha} - \frac{k_2}{\varphi}(\varphi T + \upsilon)^{3-\alpha}\right] \tag{6}$$

where

$$k_1 = \left(2\lambda\mu C E^2[x] + \frac{\lambda\mu C^k \varphi E^2[x]}{\varphi^2 - 1}\right)\left(\frac{\upsilon^\alpha}{\alpha - 1}\right) \tag{7}$$

$$k_2 = \left(\frac{\lambda\mu C^k E^2[x]}{\varphi^2 - 1}\right)\frac{\upsilon^\alpha}{\alpha - 1}. \tag{8}$$

The cell depth X is assumed to be exponentially distributed, therefore, $E(X^2) = 2\,\mu^2 x$. The probability that a period of length h is dry is:

$$\text{prob[zero rainfall]} = \exp\left\{-\lambda T - \left[\frac{\lambda\upsilon}{\varphi(\alpha-1)}\left(1 + \varphi\left(\kappa + \frac{\varphi}{2}\right) - \frac{1}{4}\varphi(\kappa+\varphi)(\kappa+2\varphi)\right).(\kappa+4\varphi) + \frac{\varphi(\kappa+\varphi)(4\kappa^2+27\kappa\varphi+36)}{72}\right] + \frac{\lambda\upsilon}{(\alpha-1)(\kappa+\varphi)}\left(1-\kappa-\varphi+\frac{3}{2}\kappa\varphi+\varphi^2+\frac{\kappa^2}{2}\right) + \frac{\lambda\upsilon}{(\alpha-1)(\kappa+\varphi)}\left(\frac{\upsilon}{\upsilon+(\kappa+\varphi)T}\right)^{\alpha-1} \cdot \frac{\kappa}{\varphi}\left(1-\kappa-\varphi+\frac{3}{2}\kappa\varphi+\varphi^2+\frac{\kappa^2}{2}\right)\right\} \tag{9}$$

$$\text{Autocovariance(lags)} = \frac{k_1}{(\alpha-2)(\alpha-3)}\left\{[T(s-1)+\upsilon]^{3-\alpha} + [T(s+1)+\upsilon]^{3-\alpha} - 2(Ts+\upsilon)^{3-\alpha}\right\} + \frac{k_2}{\varphi^2(\alpha-2)(\alpha-3)}\left\{2(\varphi Ts+\upsilon)^{3-\alpha} - [\varphi T(s-1)+\upsilon]^{3-\alpha} - [\varphi T(s+1)+\upsilon]^{3-\alpha}\right\}. \tag{10}$$

Table 2 Distribution of daily rainfall by month for various stations

No.	Station	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	SMK Sri Permata	W	G	G	G	GP	W	G	W	LP3	G	B	B
2	Balai Polis Sea Park	W	G	G	GP	W	LP3	B	W	LP3	LP3	W	W
3	SMK Taman Sea	G	G	W	GP	LP3	LP3	G	LP3	W	LP3	G	G
4	SR Damansara Utama	W	G	LP3	W	LP3	LP3	LP3	W	G	G	LP3	LP3
5	Tropicana Golf	W	GP	W	LP3	W	LP3	G	G	B	LP3	G	W
6	Balai Polis TTDI	W	W	G	W	W	W	GP	W	B	W	G	W
7	Kg. Sg. Penchala	W	G	G	W	W	LP3	GP	W	LP3	W	W	W
8	Surau Assyakirin	W	G	W	LP3	W	LP3	W	W	LP3	LP3	G	W
9	SMK Damansara Jaya	W	G	W	W	LP3	W	G	W	LP3	GP	G	G
10	Bukit Kiara Golf Resort	W	W	W	W	W	W	GP	G	W	GP	B	W

W weibull, *G* gamma, *GP* Gen.Pareto, *LP3* log Pearson 3, *B* beta

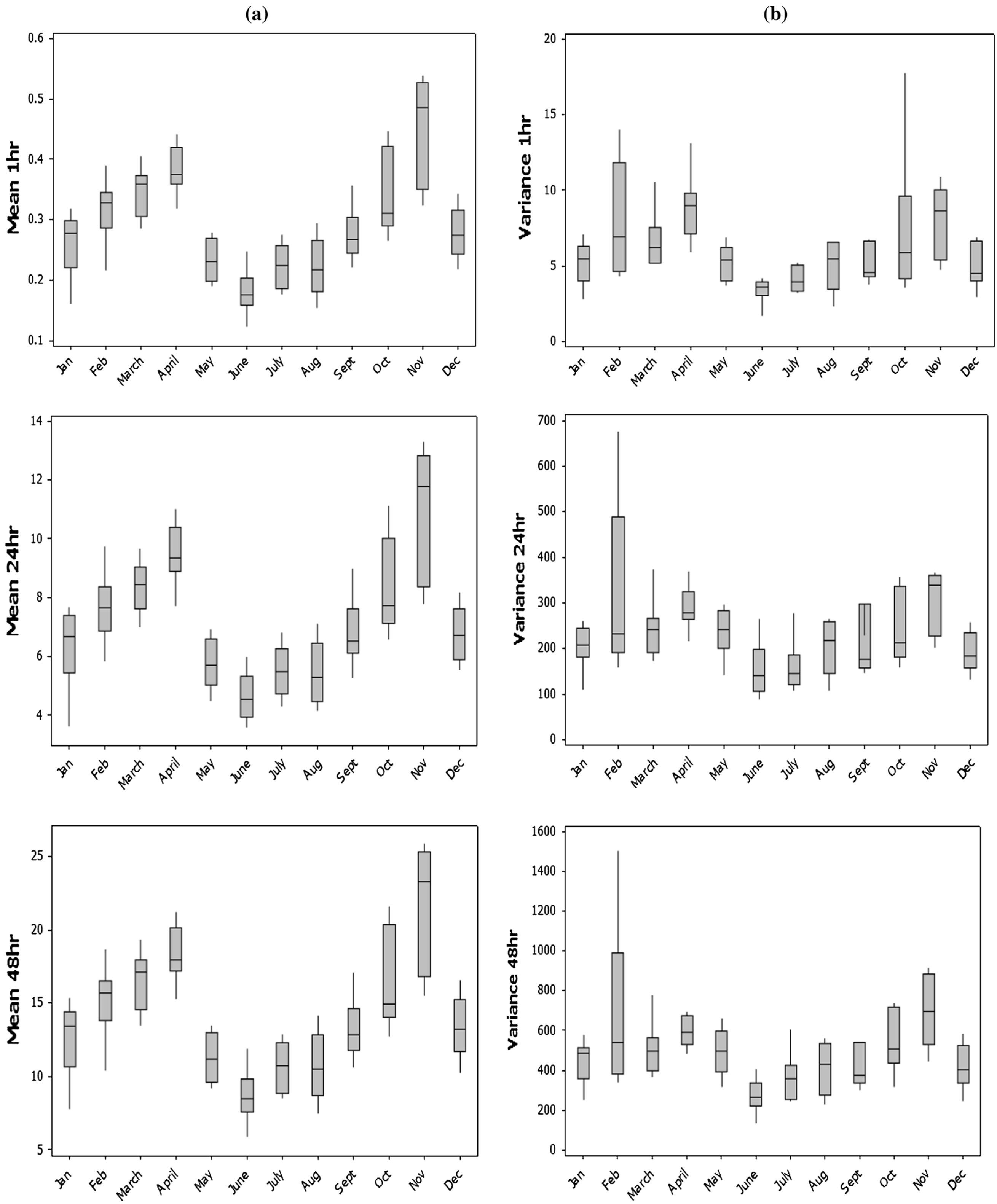


Fig. 2 *Box plots* show the statistical properties for 1, 24 and 48 h of historical daily rainfall for the ten rainfall stations. **a** Mean value, **b** variance value, **c** lag 1 auto-covariance and **d** probability of dry days

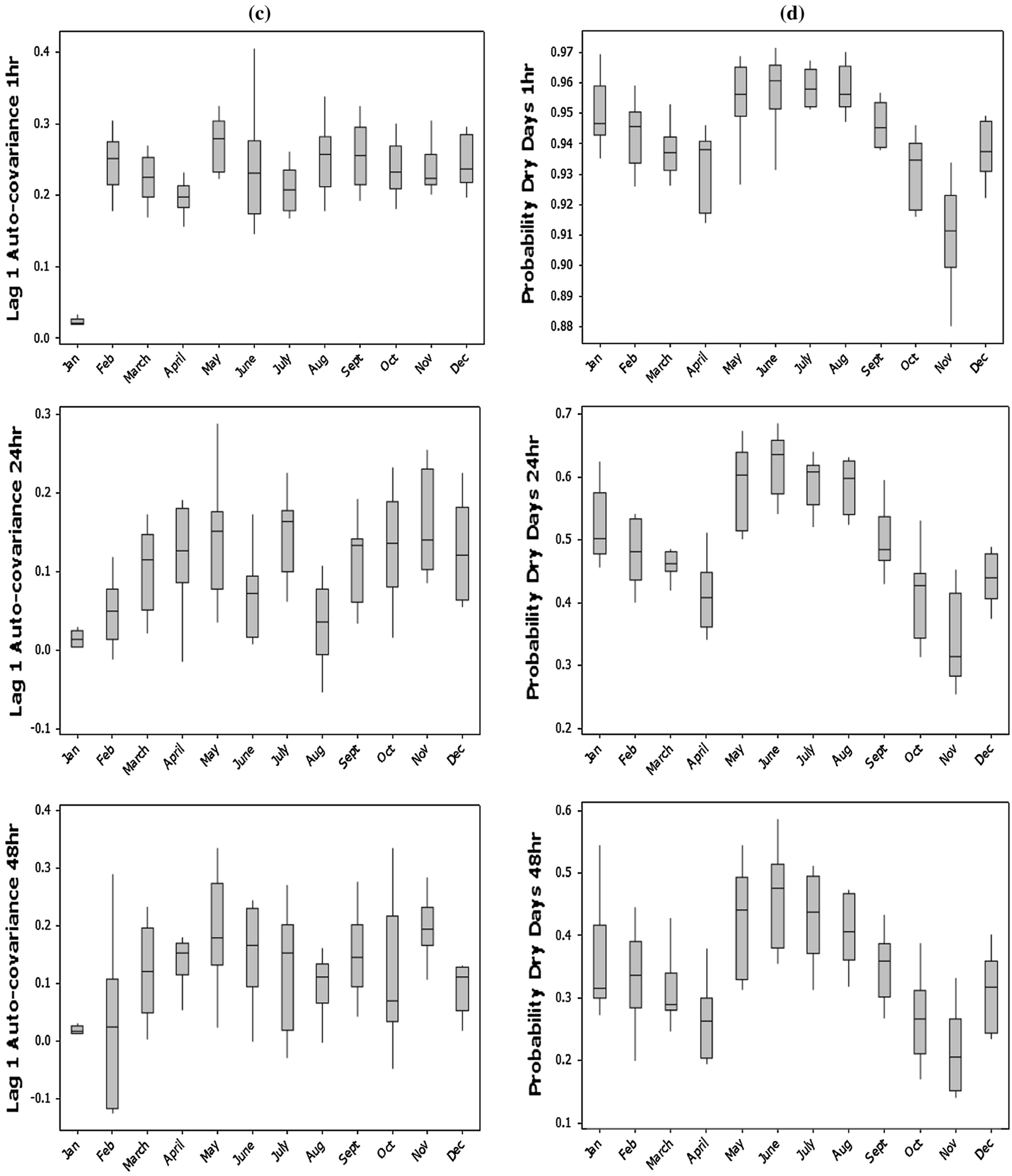


Fig. 2 continued

Table 3 Monthly parameter estimates for SMK Taman Sea station

Parameter/Month	λ	$\kappa = \beta/\eta$	$\varphi = \gamma/\eta$	α	v	μ_X	σ_X	Sum of weighted squared errors
Jan	3.81	1E−07	99	99	1.35	75.25	75.25	2.83
Feb	0.35	99	9.45	77.31	1.35	75.25	75.25	1.98
Mac	0.54	99	32.08	32.22	1.35	75.25	75.25	2.03
Apr	0.73	99	8.33	99	1.35	75.25	75.25	2.47
May	0.19	89.96	43.41	12.98	1.35	75.25	75.25	2.07
Jun	0.58	1E−07	99	14.14	1.35	75.25	75.25	2.14
Jul	0.56	19.93	99	16.26	1.35	75.25	75.25	1.91
Aug	0.24	99	6.83	99	1.35	75.25	75.25	2.31
Sep	0.46	99	11.24	76.28	1.35	75.25	75.25	2.22
Oct	0.74	99	12.99	99	1.35	75.25	75.25	2.37
Nov	0.95	99	10.80	92.68	1.35	75.25	75.25	2.21
Dec	0.53	99	14.87	78.82	1.35	75.25	75.25	2.19

Since the components have different orders of magnitude, they are first normalized to bring the adjusted values close to the historical rainfall. This was achieved by calculating the error-residual term, Z using Eq. 11,

$$Z = \min\left[\sum_{i=1}^{N} W_i\left(\frac{F_i(x)}{F_i'} - 1\right)^2\right]. \qquad (11)$$

The above technique is used to calibrate the models to the historical rainfall, where W_i is the weight attributed to all statistical values; the sum of the weighted mean, variance, dry probability, and lag 1 auto-covariance of 1, 24 and 48 h levels of aggregation. The objective function is minimised to reduce the error between the simulated and historical values. $F_i(x)$ is the corresponding analytical expression for statistic i as a function of the parameter vector x. F_i' is statistic i, estimated from historical data at various levels of aggregation. N is the number of statistics used in determining the parameter. For the original BLRP model, let $F_i = F_i\,(\lambda, \kappa, \phi, \gamma, v$ and $\mu x)$ be a function of the BLRP model, and F_i' be its historical sampled value. Velghe et al. (1994) and Verhoest et al. (1997) used $W_i = 1$ for all statistics, while Cowpertwait (1991) set $W_i = 100$ for the mean and $W_i = 1$ for all other moments. These weighted values were used in this analysis.

Results and discussion

Data distribution

Statistically, the behaviour of rainfall data can be described by the type of distribution to which the data belongs. The distributions that can best fit daily rainfall data for various stations by month are presented in Table 2.

About 42.5 % of the data follow the Weibull distribution, 24.7 % the Gamma, 20.8 % the Log Pearson 3, 7.5 % follows the Generalised Pareto, and 5 % follows a Beta distribution. According to Koutsoyiannis (1994) and Hidayah et al. (2010), the BLRP model only gives a satisfactory performance for data that can fit two-parameter gamma and/or exponential distributions (Hidayah et al. 2010). The performance of this model is discussed further in the next section.

Parameter estimation

The mean, variance, lag-1 auto-covariance and probability of dry days of the historical daily rainfall were calculated as input parameters for BLRP model. Each of these parameters is needed for 1, 24 and 48 h rainfall for every station. The box plots showing monthly statistical properties for all the ten stations are presented in Fig. 2a–d.

Most of the box plots show skewed distributions of the statistical properties. A longer upper hinge indicates a negatively skewed (to the left) distribution while a longer lower hinge indicates positive skewness. The mean and variance generally share quite similar patterns with two maxima in April and November for the mean, and in February and November for the variance. June recorded the least rainfall whereas November the highest. The probabilities of dry days for 24 and 48 h tend to show opposite trends with the mean and variance, being lowest in April and November. However, the pattern for lag-1 auto-covariance is not clear. The length of the box plot indicates the degree of peakedness.

Using all those statistical properties, the monthly parameters of the BLRP model were estimated and used as inputs into *Hyetos*. The values were derived on a monthly basis for all stations. An example of the calculated parameter estimates for SMK Taman Sea station is shown in Table 3.

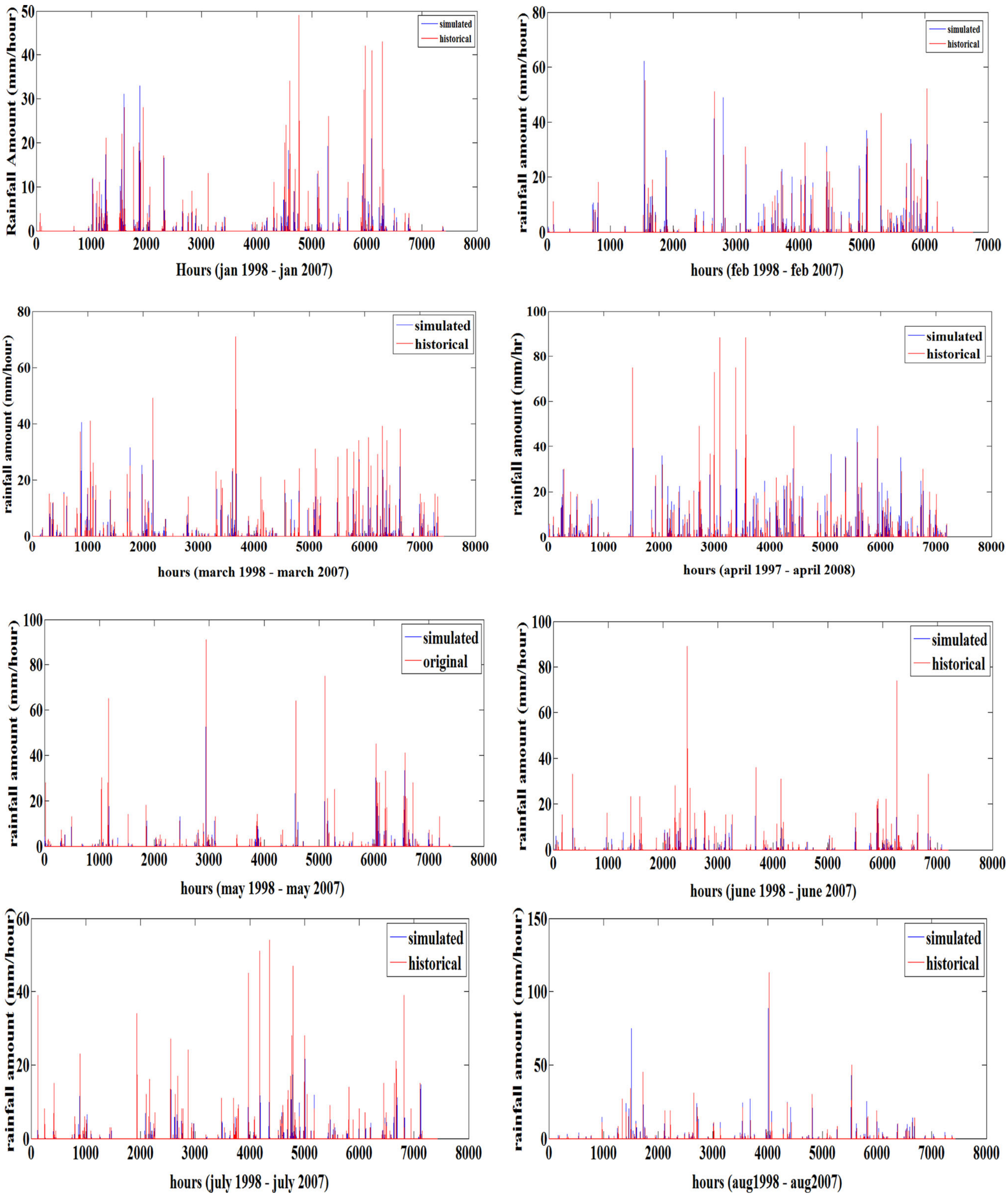


Fig. 3 Comparison of historical and simulated hourly rainfall by month based on data from Jan 1988 to Dec 2007. Note: Each plot consists of hourly rainfall of the same month (e.g., January) for ten different years

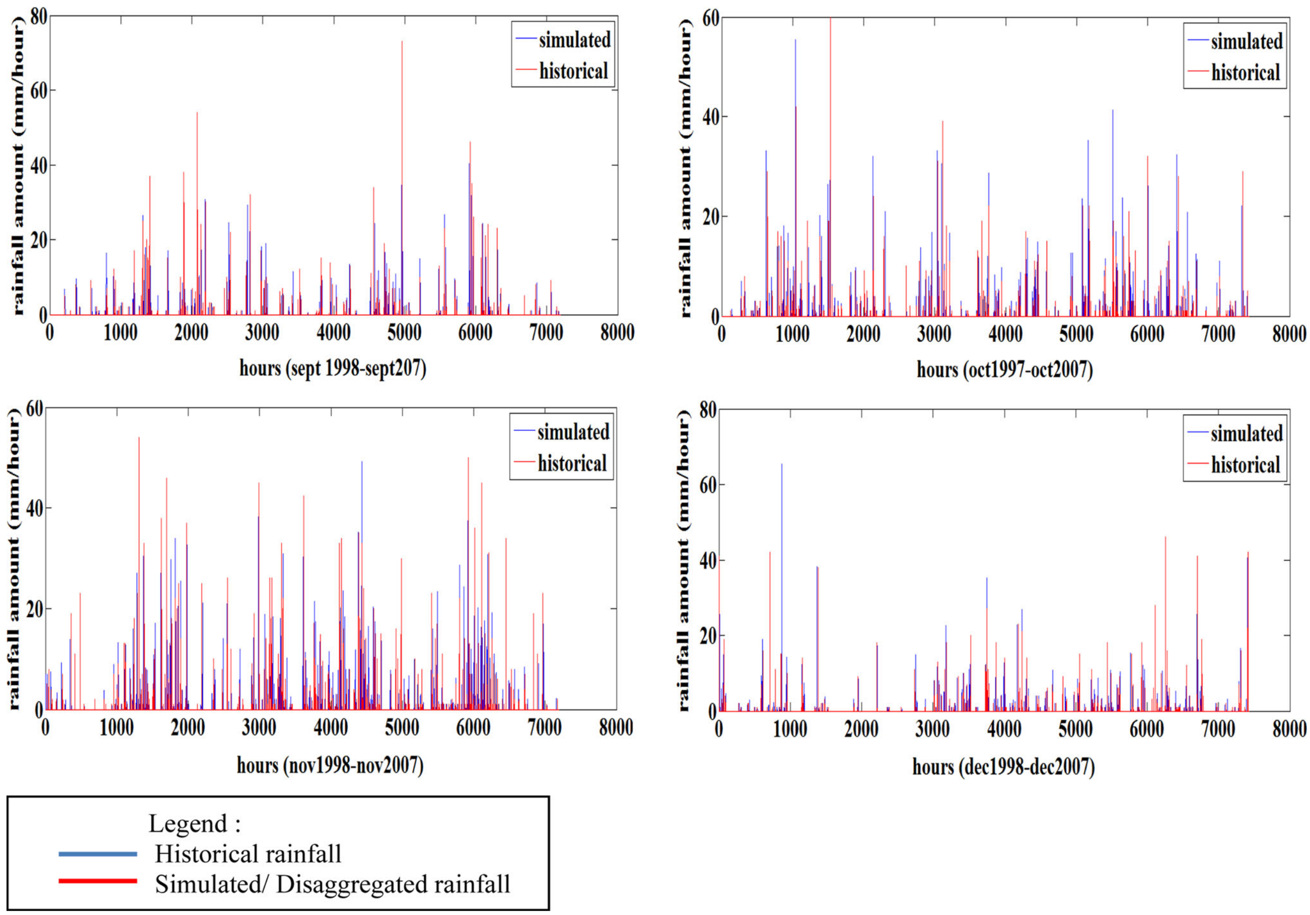


Fig. 3 continued

Performance of the model

The *Hyetos* simulation results for SMK Taman Sea station were also chosen as an example to compare the simulated hourly rainfall with the observed values Fig. 3. The results were plotted on a monthly basis (January–December) using 10-year rainfall data (1998–2007).

It is observed that most of the times the simulated hourly rainfall (red bar) fail to catch up with the extreme values of the historical data (blue bar). However, for small and moderate storms there are many events where the simulated hourly rainfall exceeded the historical rainfall especially in February, July and August which are usually quite dry.

The failure of *Hyetos* model to match the extreme rainfall might be a real limitation of the model when applied in tropical region. Rainfalls in the tropics are mainly convective that are characterised by sudden burst, very high intensities and short duration. The intensities also fluctuate but decrease rapidly as the storm progress.

Another method of checking the model's performance in disaggregating daily data into hourly data is by comparing the statistical properties (mean and standard deviation) of the historical hourly with the disaggregated (simulated) hourly as shown in Fig. 4. Figure 4a (left side) shows the monthly mean values whereas Fig. 4b (right side) shows the monthly standard deviation derived from ten stations. The mean and standard deviation are useful to check the consistency of the simulated rainfall against the historical rainfall.

The simulated result of daily to hourly mean rainfall for various stations closely matched the historical hourly means, except for the SR Damansara Utama station in January, March, July, August, November and December, and the Tropicana Golf station in January, February, March, April, September and October. For the SR Damansara Utama station, the BLRP model resulted in higher disaggregated daily–hourly means than those historical, and the opposite was found for the Tropicana Golf station.

The level of error in disaggregating rainfall data depends on the accuracy of the parameter estimates for the BLRP model. The sum weighted errors of the monthly rainfall were high, ranging from 1.91 to 2.83 (see Table 3). According to Gupta et al. (1999), in order to achieve good simulation result the weighted of error

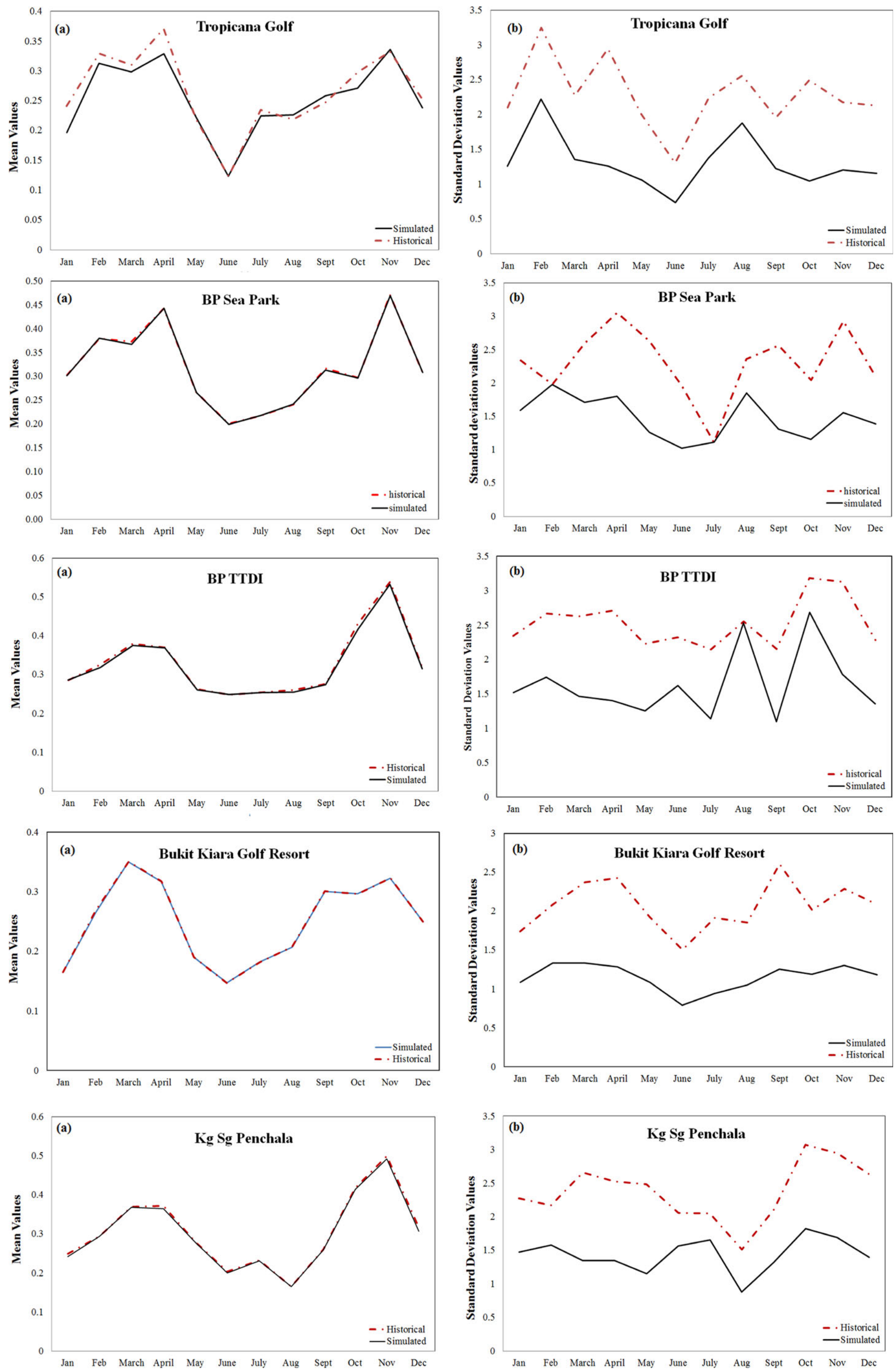
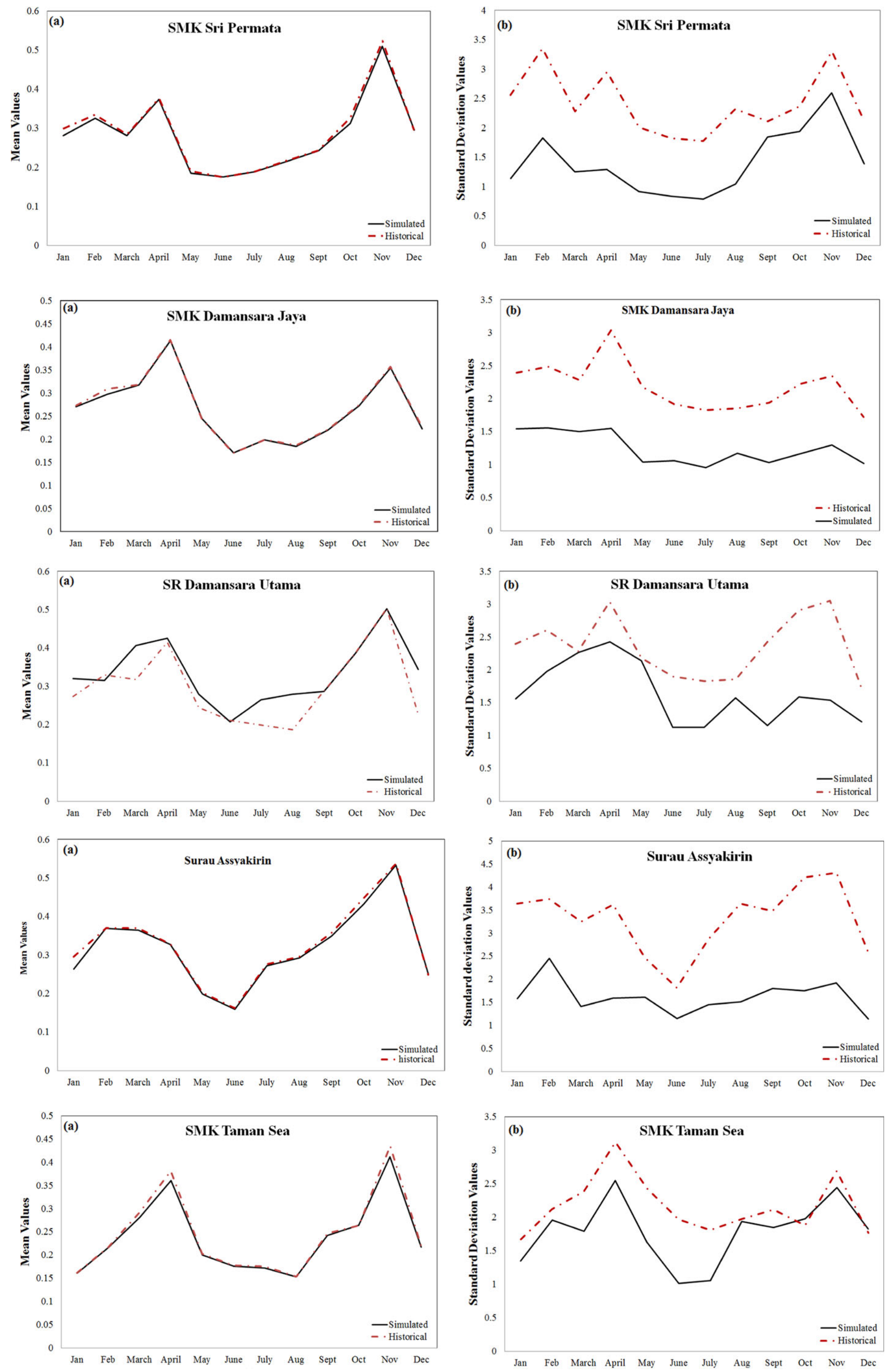


Fig. 4 Comparison of mean (*left side*) and standard deviation (*right side*) of the disaggregated and historical hourly rainfall for the ten stations in Damansara

should be close to 0 (zero) because values of 0 (zero) indicated that as a perfect fit or as accurate model simulation. Lower sum weighted errors, were obtained in dry months (January, May, June, August and December) rather than in wet months (March, April, September and October). This is expected because of more fluctuation in the rainfall amount during wet months.

Unlike the mean, the standard deviation of the daily disaggregated rainfall (hourly rainfall) was generally lower than the historical series, but both present similar trends (see Fig. 4b). The standard deviation provides measure of the degree of dispersion of the data from the mean. A larger standard deviation than the mean indicates that the data are spread out over a large range of values. On the other hand a small standard deviation indicates clustering of data around the mean. For example, large standard deviations obtained for the SR Damansara Utama station have resulted in overestimation

Fig. 4 continued



of hourly mean, and the pattern was different from the historical data.

Another factor that could affect the *Hyetos* performance is rainfall distribution. It was found that the daily rainfall distribution varied between stations (see Table 2). Thus, direct application of *Hyetos* without first checking the data distribution of each rainfall station would result in serious disaggregation errors (Hidayah et al. 2010). The BLRP model in *Hyetos* is computed based on the assumption that the rainfall data follow a gamma or exponential distribution (Koutsoyiannis and Onof 2001). The most dominant distributions found in this study are Weibull, Log Pearson 3 and Gamma. Therefore, the large standard deviations of the disaggregated rainfall in January, May, June, August and December could be attributed to the historical data which follow Log Pearson 3 or Weibull distributions. This was rather unexpected because Log Pearson type 3 or Weibull distributions have an almost similar structure to Gamma. However, results of this study show that the BLRP model can fulfil the objective of disaggregating data from daily into hourly rainfall. In addition, the BLRP model is able to preserve the statistical properties of the rainfall data.

Conclusion and recommendation

This paper discusses the use of the BLRP model to disaggregate daily rainfall data in Damansara, which is located in the central region of Peninsular Malaysia. The model was used to disaggregate daily rainfall into hourly rainfall using 10 years of data from ten selected stations. The model parameters were estimated from four statistical properties, namely the mean, standard deviation, lag-1 auto-covariance and probability of hourly, 24 and 48 h dry periods from the observed daily data.

The mean of disaggregated rainfall shows perfect match with the historical mean. Although the standard deviations of the disaggregated daily rainfall and the observed daily rainfall generally followed similar patterns, the former was significantly lower. In general, the BLRP model shows good performance in preserving the statistical properties, and has a good ability to disaggregate the daily rainfall into hourly rainfall, especially for stations that have daily rainfall data resembling gamma or exponential distributions. It is necessary to test the BLRP model for different regions in Malaysia to ensure consistency of results with different rainfall regimes. It is also worthwhile to improve the statistical model using such techniques as parameter optimisation, data clustering or using another time series model. The latest work along this line is to evaluate the ability of BLRP model for disaggregating rainfall at multiple sites.

Acknowledgments This project was sponsored by the Ministry of Higher Education (MOHE) through the Fundamental Research Grant Scheme (FRGS) with Grant Number Q.J130000.7822.3F601. 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 Higher Education (MOHE) Malaysia.

References

Bo ZPD, Islam S (1994) Aggregation–disaggregation properties of a stochastic rainfall model. Water Resour Res 30(12):3423–3435. doi:10.1029/94WR02026

Burgueño A, Martínez MD, Lana X, Serra C (2005) Statistical distributions if the daily rainfall regime in Catalonia (Northeastern Spain) for the years 1950–2000. Int J Climatol 25:1381–1403. doi:10.1002/joc.1197

Cowpertwait PSP (1991) Further developments of the Neyman–Scott clustered point process for modelling rainfall. Water Resour Res 27(7):1431–1438. doi:10.1029/91WR00479

Dale WL (1974) The rainfall of Malaya, Part II. In: Bee OJ, Sien CL (eds) The climate of West Malaysia and Singapore. Oxford University Press, Oxford, pp 144–161

Debele B, Srinivasan R, Yves Parlange J (2007) Accuracy evaluation of weather data generation and disaggregation methods at finer timescales. Adv Water Resour 30(5):1286–1300

Deidda R, Puliga M (2006) Sensitivity of goodness-of-fit statistics to rainfall data rounding off. Phys Chem Earth 31:1240–1251. doi:10.1016/j.pce.2006.04.041

Duan J, Sikka AK, Grant GE (1995) A comparison of stochastic models for generating daily precipitation at the H.J. Andrews experimental forest. Northwest Sci 69(4):318–329was cited in the page 5 line 156

Entekhabi D, Rodriguez-Iturbe I, Eagleson PS (1989) Probabilistic representation of the temporal rainfall process by a modified Neyman–Scott rectangular pulse model: parameter estimation and validation. Water Resour Res 25(2):295–302. doi:10.1029/WR025i002p00295

Glasbey CA, Cooper G, McGechan MB (1995) Disaggregation of daily rainfall by conditional simulation from a point-process model. J Hydrol 165(1–4):1–9

Gupta H, Sorooshian S, Yapo P (1999) Status of automatic calibration for hydrologic models: comparison with multilevel expert calibration. J Hydrol Eng 4(2):135–143. doi:10.1061/(ASCE)1084-0699(1999)4:2(135)

Gyasi-Agyei Y (2005) Stochastic disaggregation of daily rainfall into one-hour time scale. J Hydrol 309(1–4):178–190

Gyasi-Agyei Y, Mahbub SMPB (2007) A stochastic model for daily rainfall disaggregation into fine time scale for a large region. J Hydrol 347(3–4):358–370

Hanaish IS, Ibrahim K, Jemain AA (2011) Stochastic modelling of rainfall in peninsular Malaysia using Bartlett Lewis rectangular pulses models. Model Simul Eng 2011:37. doi:10.1155/2011/253735

Hidayah E, Iriawan N, Nadjadji A, Edijatno (2010) Generating Hourly Rainfall Model Using Bayesian Time Series Model (A case study at Sentral Station, Bondowoso). Majalah IPTEK. Institut Teknologi Sepuluh November, Surabaya, pp 21–25

Isham S, Entekhabi D, Bras RL, Rodriguez-Iturbe I (1990) Parameter estimation and sensitivity analysis for the modified Bartlett-Lewis rectangular pulses model of rainfall. J Geophys Res 95(D3):2093–2100. doi:10.1029/JD095iD03p02093

Koutsoyiannis D (1994) A stochastic disaggregation method for design storm and flood synthesis. J Hydrol 156:193–225. doi:10.1016/0022-1694(94)90078-7

Koutsoyiannis D, Onof C (2000) A computer program for temporal rainfall disaggregation using adjusting procedures. In: 25th General Assembly of the European Geophysical Society, Geophysical Research Abstracts, vol 2. European Geophysical Society, Nice

Koutsoyiannis D, Onof C (2001) Rainfall disaggregation using adjusting procedures on a Poisson cluster model. J Hydrol 246(1–4):109–122

Koutsoyiannis D, Onof C, Wheater HS (2003) Multivariate rainfall disaggregation at a fine timescale. Water Resour Res 39(7):1173. doi:10.1029/2002wr001600

Kumar DN, Lall U, Petersen MR (2000) Multi-site disaggregation of monthly to daily stream flow. Water Resour Res 36(7):1823–1833

Lane WL (1982) Corrected parameter estimates for disaggregation schemes. In: Singh VP (ed) Statistical analysis of rainfall and runoff. Water Resour Publ, Highlands RanchColo

Mejia JM, Rouselle J (1976) Disaggregation models in hydrology revisited. Water Resour Res 12:185–186. doi:10.1029/WR012i002p00185

Noguchi S, Abdul Rahim N, Tani M (2003) Rainfall characteristics of tropical rainforest at Pasoh Forest Reserve, Negeri Sembilan, Peninsular Malaysia. In: Okuda T, Manokaran N, Matsumoto Y, Niiyama K, Thomas S, Ashton P (eds) Pasoh: ecology and natural history of a Southeast Asian lowland tropical rain forest. Springer, Japan, pp 51–58. doi:10.1007/978-4-431-67008-7_4

Nowak K, Prairie J, Rajagopalan B, Lall U (2010) A nonparametric stochastic approach for multisite disaggregation of annual to daily stream flow. Water Resour Res 46(8):W08529. doi:10.1029/2009wr008530

Olofintoye OO, Sule BF, Salami AW (2009) Best-fit probability distribution model for peak daily rainfall of selected Cities in Nigeria. N Y Sci J 2(3):1–2

Onof C, Wheater HS (1993) Modelling of British rainfall using a random parameter Bartlett-Lewis Rectangular Pulse Model. J Hydrol 149(1–4):67–95

Onof C, Wheater HS (1994) Improvements to the modelling of British rainfall using a modified random parameter Bartlett-Lewis rectangular pulses model. J Hydrol 157:177–195. doi:10.1016/0022-1694(94)90104-x

Rodriguez-Iturbe I, Cox DR, Isham V (1987) Some models for rainfall based on stochastic point process. Proc Royal Soc A Math Phys Eng Sci 410:269–288. doi:10.1098/rspa.1987.0039

Rodriguez-Iturbe I, Cox DR, Isham V (1988) A point process model for rainfall: further developments. Proc Royal Soc A Math Phys Eng Sci 417:283–298. doi:10.1098/rspa.1988.0061

Stedinger JR, Pel D, Cohn TA (1985) A condensed disaggregation model for incorporating parameter uncertainty into monthly reservoir simulations. Water Resour Res 21:665–675. doi:10.1029/WR021i005p00665

Suhaila J, Jemain AA (2007) Fitting daily rainfall amount in Peninsular Malaysia using several types of exponential distributions. J Appl Sci Res 3(10):1027–1036

Suhaila J, Jemain AA (2008) Fitting the statistical distribution for daily rainfall in Peninsular Malaysia based on AIC criterion. J Appl Sci Res 4(12):1846–1857

Valdes JB, Rodriguez-Iturbe I, Vicens GJ (1977) Bayesian generation of synthetic stream flows 2. The multivariate case. Water Resour Res 13:291–295

Valencia D, Schaake JC Jr (1973) Disaggregation processes in stochastic hydrology. Water Resour Res 9:580–585. doi:10.1029/WR009i003p00580

Van Zijl G, Le Roux, PAL and Turner, DP (2010) Disaggregation of land types, using terrain analysis, expert knowledge and GIS methods. World Congress of Soil Science, Soil Solutions for a Changing World, Brisbane Australia

Velghe T, Troch PA, De Troch FP, Van de Velde J (1994) Evaluation of cluster-based rectangular pulses point process models for rainfall. Water Resour Res 30(10):1431–1438. doi:10.1029/94wr01496

Verhoest N, Troch PA, De Troch FP (1997) On the applicability of Bartlett-lewis rectangular pulses models in the modelling of design storms at a point. J Hydrol 202:108–120. doi:10.1016/s0022-1694(97)00060-7