

# Effect of the inter-annual variability of rainfall statistics on stochastically generated rainfall time series: part 1. Impact on peak and extreme rainfall values

Dongkyun Kim · Francisco Olivera ·  
Huidae Cho

Published online: 2 February 2013  
© Springer-Verlag Berlin Heidelberg 2013

**Abstract** A noble approach of stochastic rainfall generation that can account for inter-annual variability of the observed rainfall is proposed. Firstly, we show that the monthly rainfall statistics that is typically used as the basis of the calibration of the parameters of the Poisson cluster rainfall generators has significant inter-annual variability and that lumping them into a single value could be an oversimplification. Then, we propose a noble approach that incorporates the inter-annual variability to the traditional approach of Poisson cluster rainfall modeling by adding the process of simulating rainfall statistics of individual months. Among 132 gage-months used for the model verification, the proportion that the suggested approach successfully reproduces the observed design rainfall values within 20 % error varied between 0.67 and 0.83 while the same value corresponding to the traditional approach varied between 0.21 and 0.60. This result suggests that the performance of the rainfall generation models can be largely improved not only by refining the model structure but also by incorporating more information about the observed rainfall, especially the inter-annual variability of the rainfall statistics.

## 1 Introduction

Stochastic rainfall generators are widely used in hydrologic analysis because they can provide precipitation input to models in situations where data are not available. In general, stochastic rainfall generators are classified into the three following categories: (1) the multi-scaling models, which are based on the observation that rainfall patterns have “self-similarity” at a given range of timescales (Lovejoy and Schertzer 1990 among many), (2) the non-parametric resampling models, which forms the new rainfall time series by borrowing the fragments from the instrumental data with similar statistical properties (Lall and Sharma 1996; Tarboton et al. 1998; Westra et al. 2012), (3) the Poisson cluster rainfall models, which is being considered in this study. The Poisson cluster rainfall models (Rodriguez-Iturbe et al. 1987, 1988), a type of stochastic rainfall models, represent rainfall as a sequence of storms composed of rain cell clusters (Kavvas and Delleur 1975). According to Olsson and Burlando (2002), “the representation of storm occurrence as a point process and the internal storm intensity structure as a cluster of rectangular pulses mimicking rain cells has proven to be (...) a physically realistic way to describe temporal rainfall”. The applicability of the Poisson cluster rainfall models has been validated over various geographic locations with different rainfall characteristics (Isham et al. 1990; Bo et al. 1994; Onof and Wheater 1994; Glasbey et al. 1995; Khaliq and Cunnane 1996; Onof et al. 1996; Cowpertwait et al. 1996; Verhoest et al. 1997) and these models have been used in a wide range of studies dealing with flooding (e.g. Wheater et al. 2005), drought (e.g. Yoo et al. 2008), contaminant transport (e.g. Botter et al. 2006), and ecosystem behavior (e.g. Laio et al. 2009), among others. A more thorough explanation about the

---

D. Kim (✉)  
School of Urban and Civil Engineering, Hongik University,  
Seoul, Republic of Korea  
e-mail: deokaykim@gmail.com

F. Olivera · H. Cho  
Zachry Department of Civil Engineering, Texas A&M  
University, College Station, TX, USA

*Present Address:*  
H. Cho  
Water Resources Engineer, Dewberry, 8401 Arlington Blvd.,  
Fairfax, VA, USA

developments, applications, and limitations to overcome of the Poisson cluster rainfall models can be found in Onof et al. (2000).

In Poisson cluster models, model parameters are typically calculated for each month of the year based on the multi-annual rainfall statistics of the corresponding months to account for the seasonal variability of rainfall (i.e., month-to-month variability), resulting in twelve parameter sets instead of one (Burton et al. 2008 among others). However, to our knowledge, the inter-annual variability of the precipitation statistics has not been incorporated in the estimation of the model parameters, and its effect on stochastically generated rainfall time series has not been studied.

In this study, it is hypothesized that overlooking the inter-annual variability of the statistics when generating stochastic rainfall causes errors and biases that might affect the overall rainfall depth variability and frequency of extreme values. To validate the proposed approach, a rainfall generator that can account for inter-annual variability is developed. Then, the rainfall time series generated based on this model was compared to the observed ones and also to the time series generated without accounting for inter-annual variability.

Section 2 describes the modified Bartlett-Lewis rectangular pulse (MBLRP) model, the stochastic rainfall model used in this study. Section 3 describes inter-annual variability of precipitation in detail. Section 4 describes how the new method of rainfall generation called “the hybrid model” (THM) is developed. Section 5 presents its results and discussion. Finally, Sect. 6 concludes this study.

## 2 The modified Bartlett-Lewis rectangular pulse (MBLRP) model

The MBLRP model, rainfall time series are represented as sequences of storms comprised of cluster of rain cells (see Fig. 1). In the model,  $X_1$  [T] is a random variable that represents the storm arrival time, which is governed by a Poisson process with parameter  $\lambda$  [1/T];  $X_2$ [T] is a random variable that represents the duration of storm activity (i.e., the time window after the beginning of the storm within which rain cells can arrive), which varies according to an exponential distribution with parameter  $\gamma$  [1/T];  $X_3$  [T] is a random variable that represents the rain cell arrival time within the duration of storm activity, which is governed by a Poisson process with parameter  $\beta$  [1/T];  $X_4$  [T] is a random variable that represents the duration of the rain cells, which varies according to an exponential distribution with parameter  $\eta$  [1/T] that, in turn, has a gamma distribution with parameters  $\nu$  [T] and  $\alpha$  (dimensionless); and  $X_5$  [L/T] is a random variable that represents the rain cell intensity, which varies according to an exponential

distribution with parameter  $1/\mu$  [T/L]. From the physical viewpoint,  $\lambda$  is the expected number of storms that arrive in a given period,  $1/\gamma$  is the expected duration of storm activity,  $\beta$  is the expected number of rain cells that arrive within the duration of storm activity,  $1/\eta$  is the expected duration of the rain cells, and  $\mu$  is the expected rain cell intensity. Parameters  $\nu$  and  $\alpha$  do not have a clear physical meaning, but the expected value and variance of  $\eta$  can be expressed as  $\alpha/\nu$  and  $\alpha/\nu^2$ . Therefore, the model has six parameters:  $\lambda$ ,  $\gamma$ ,  $\beta$ ,  $\nu$ ,  $\alpha$  and  $\mu$ ; however, it is customary to use dimensionless ratios  $\varphi = \gamma/\eta$  and  $\kappa = \beta/\eta$  as parameters instead of  $\gamma$  and  $\beta$ . Based on these model assumptions, Rodriguez-Iturbe et al. (1988) derived the equations for the statistics of the simulated rainfall time series at an accumulation interval T as follow:

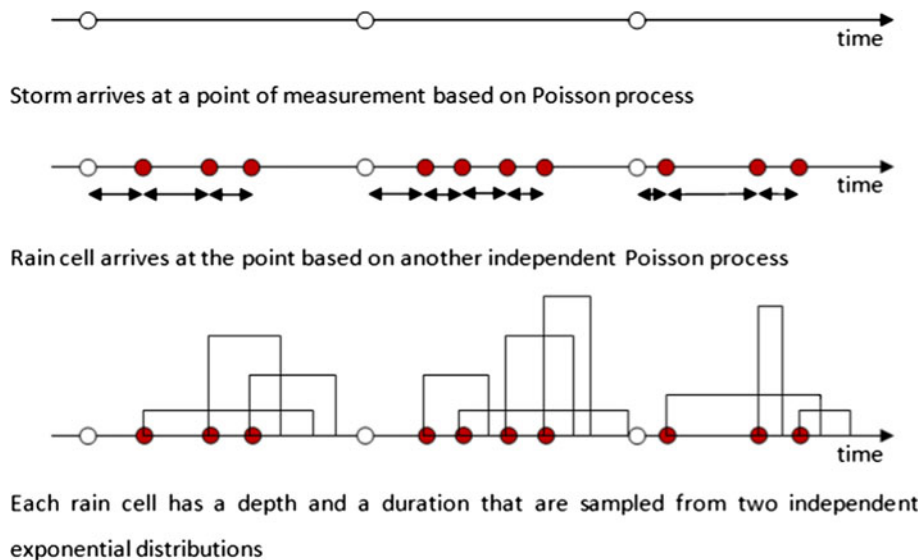
$$E[Y_t^{(t)}] = \lambda \mu \mu_c \frac{\nu}{\alpha - 1} T \quad (1)$$

$$\begin{aligned} \text{Var}[Y_t^{(T)}] = & \frac{2 \nu^{2-\alpha} T}{\alpha - 2} \left( k_1 - \frac{k_2}{\phi} \right) \\ & - \frac{2 \nu^{3-\alpha}}{(\alpha - 2)(\alpha - 3)} \left( k_1 - \frac{k_2}{\phi^2} \right) \\ & + \frac{2}{(\alpha - 2)(\alpha - 3)} \\ & \times \left[ k_1 (T + \nu)^{3-\alpha} - \frac{k_2}{\phi^2} (\phi T + \nu)^{3-\alpha} \right] \end{aligned} \quad (2)$$

$$\begin{aligned} \text{Cov}[Y_t^{(T)}, Y_{t+s}^{(T)}] = & \frac{k_1}{(\alpha - 2)(\alpha - 3)} \left\{ [T(s - 1) + \nu]^{3-\alpha} \right. \\ & \left. + [T(s + 1) + \nu]^{3-\alpha} - 2(Ts + \nu)^{3-\alpha} \right\} \\ & + \frac{k_2}{\phi^2 (\alpha - 2)(\alpha - 3)} \\ & \times \left\{ 2(\phi Ts + \nu)^{3-\alpha} - [\phi T(s - 1) + \nu]^{3-\alpha} \right. \\ & \left. - [\phi T(s + 1) + \nu]^{3-\alpha} \right\} \end{aligned} \quad (3)$$

$$\begin{aligned} P(\text{Zero rainfall}) = & \exp \left\{ -\lambda T - \frac{\lambda \nu}{\phi(\alpha - 1)} \left[ 1 + \phi(\kappa + \phi) \right. \right. \\ & \left. \left. - \frac{1}{4} \phi(\kappa + \phi)(\kappa + 4\phi) \right. \right. \\ & \left. \left. + \frac{\phi(\kappa + \phi)(4\kappa^2 + 27\kappa\phi + 72\phi^2)}{72} \right] \right\} \\ & + \frac{\lambda \nu}{(\alpha - 1)(\kappa + 4\phi)} \\ & \times \left( 1 - \kappa - \phi + \frac{3}{2} \kappa\phi + \phi^2 + \frac{\kappa^2}{2} \right) \\ & + \frac{\lambda \nu}{(\alpha - 1)(\kappa + \phi)} \left[ \frac{\nu}{\nu + (\kappa + \phi)T} \right]^{\alpha-1} \frac{\kappa}{\phi} \\ & \times \left( 1 - \kappa - \phi + \frac{3}{2} \kappa\phi + \phi^2 + \frac{\kappa^2}{2} \right) \end{aligned} \quad (4)$$

**Fig. 1** Schematic of the MBLRP model. The *white* and *gray circles* represent the arrival time of storms and rain cells, respectively. Each rain cell is represented by a *rectangle* whose width and height represent its duration and rainfall intensity, respectively



where

$$k_1 = \left( 2\lambda\mu_c\mu^2 + \frac{\lambda\mu_c\kappa\phi\mu^2}{\phi^2 - 1} \right) \left( \frac{v^\alpha}{\alpha - 1} \right) \tag{5}$$

$$k_2 = \left( \frac{\lambda\mu_c\kappa\phi\mu^2}{\phi^2 - 1} \right) \left( \frac{v^\alpha}{\alpha - 1} \right) \tag{6}$$

$$\mu_c = 1 + \frac{\kappa}{\phi}, \tag{7}$$

where  $s$  is the lag time in number of accumulation intervals, and  $Yt^{(T)}$  is the rainfall time series at an accumulation interval  $T$ .

Estimation of MBLRP model parameters is accomplished by minimizing the discrepancy between the statistics of simulated and observed rainfall time series. Some statistics commonly used are the mean, variance and lag- $s$  covariance of the precipitation depth, and the probability of zero rainfall at various accumulation levels (Khaliq and Cunnane 1996). Bo et al. (1994) suggested the following equation as an objective function in calibration:

$$\sum_{k=1}^n w_k \left[ 1 - \frac{F_k(\vec{\theta})}{f_k} \right]^2 \tag{8}$$

where  $\vec{\theta}$  is a vector of the model parameters,  $n$  is the number of rainfall statistics being matched at various temporal accumulation level,  $F_k$  is the  $k$ th statistic of the synthetic rainfall time series (Eqs. 1 to 4),  $f_k$  is the  $k$ th statistic of the observed rainfall time series, and  $w_k$  is the weight factor for the  $k$ th statistic (Kim and Olivera 2012).

### 3 Rainfall inter-annual variability

Figure 2 shows the seasonality and inter-annual variability of the statistics of the rainfall time series observed in the

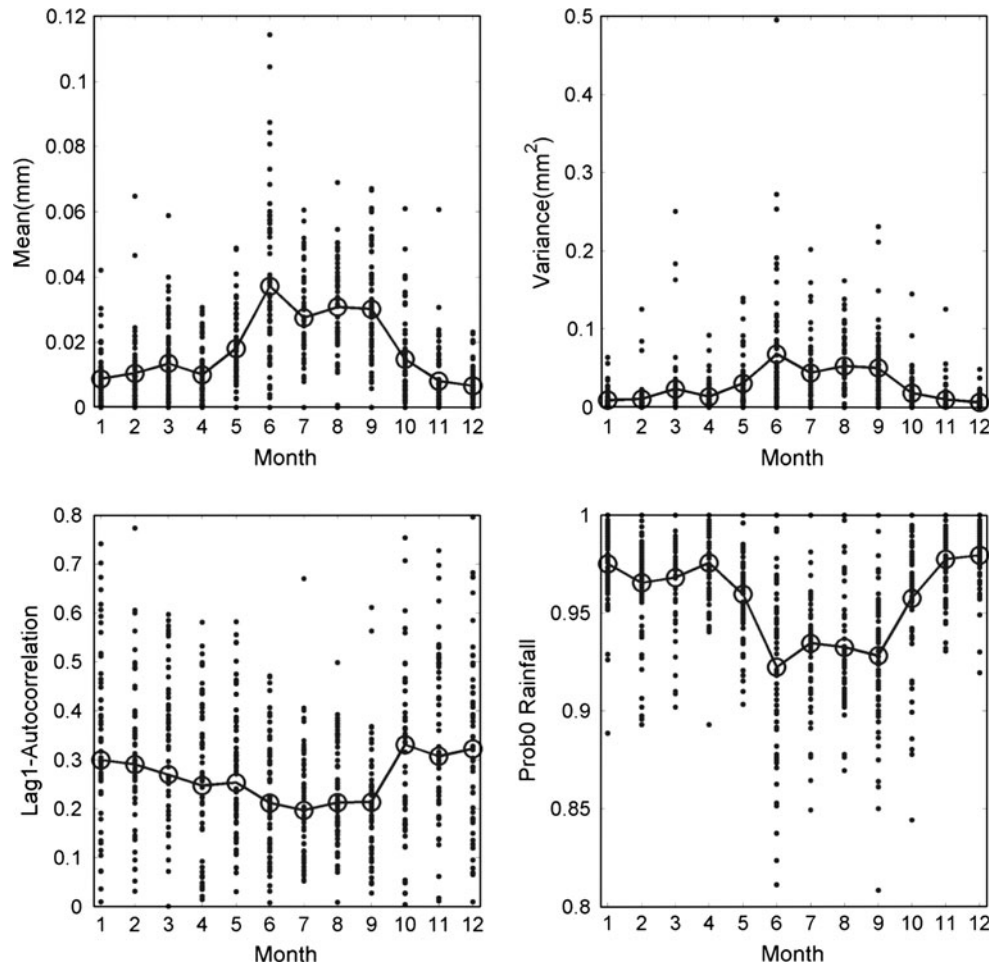
NCDC (2011) gage FL-9148 (star mark in Fig. 3). In the figure, the hollow circles connected by solid lines represent the statistics calculated for the entire period of record; while the solid dots show the statistics of individual months that vary with years (e.g. January of 1985, January of 1986). The wide range of the vertical distribution of the solid dots in each plot suggest that the monthly statistics vary significantly from year to year and that lumping them into a single value could be an oversimplification. Here, it is noteworthy that the conventional approaches of Poisson cluster rainfall modeling (Burton et al. 2008 among others) generates the rainfall time series based on this “lumped” rainfall statistics. This study proposes a noble approach that can incorporate the inter-annual variability of rainfall statistics and shows how it enhances the properties of the generated rainfall time series. We name this approach THM and explain it in the following section.

## 4 The hybrid model (THM)

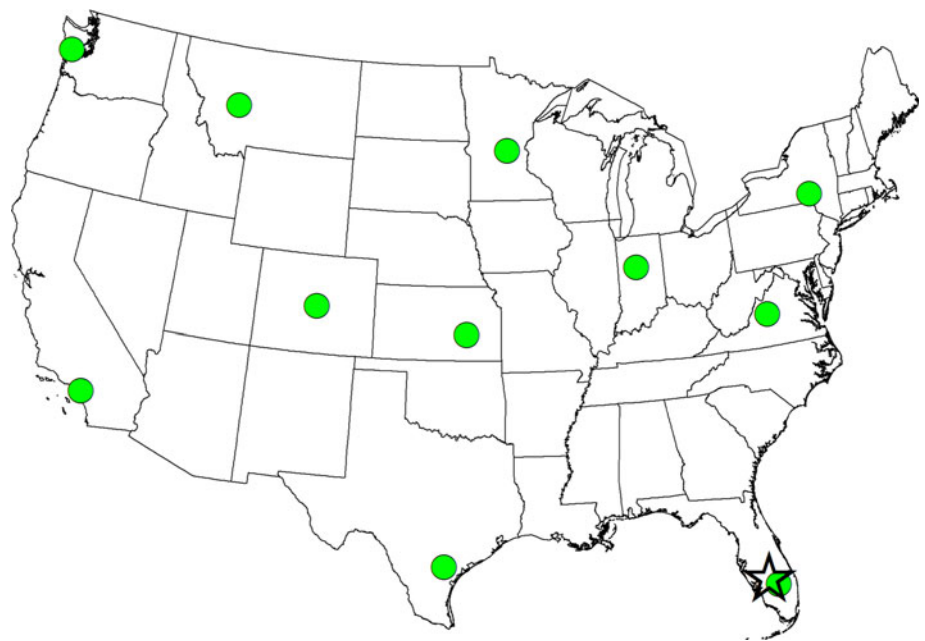
### 4.1 Overview

The model presented in this study incorporates the inter-annual variability of rainfall statistics by simulating short-term rainfall statistics (that is, rainfall statistics of individual months) based on the correlation between observed rainfall statistics. Then, the model generates rainfall time series using the MBLRP model based on the simulated short-term rainfall statistics. This model is named “THM” because it combines the process of generating rainfall statistics and the process of generating rainfall time series. The development of the model is explained in detail using the precipitation data observed at gage FL-9148 (star mark in Fig. 3). Finally, the model is validated using the rainfall data observed at 11

**Fig. 2** Monthly variations of the rainfall statistics observed at gage FL-9148. The statistics referring to the month of the entire length of rainfall time series are shown as the *hollow circles* along with the *lines* and the ones referring to the month of a specific year are shown as the *solid dots*



**Fig. 3** Gage FL-9148 (*star*) and the 11 gages used for model validation (*circles*)



rain gages (total of  $11 \times 12$  months = 132 gage-months) located across the United States (Fig. 3).

#### 4.2 Generation of rainfall statistics

The first step of THM is to stochastically generate rainfall statistics of individual months that is to be used for parameter calibration of the MBLRP model. The model generates the following monthly rainfall statistics based on the correlation between them; mean at an hourly accumulation level (MEAN1), variances at hourly, 3-, 12-, and 24-hourly accumulation levels (VAR1, VAR3, VAR12, and VAR24); probability of zero rainfall at hourly, 3-, 12-, and 24-hourly accumulation levels (PROB0-1, PROB0-3, PROB0-12, and PROB0-24); lag-1 autocorrelation at hourly, 3-, 12-, and 24-hourly accumulation levels (AC1, AC3, AC12, and AC24).

Figure 4 shows the scatter plots of the STD1 ( $=\sqrt{\text{VAR1}}$ ), PROB0-1 and AC1 versus MEAN1 of the rainfall time series observed at gage FL-9148 for the month of June. Each point in the plots represents the monthly statistic of the rainfall time series (e.g. June 1981, June 1995, etc.). Consequently, there are 61 points in each plot because the gage has 61 years of records. Linear regression analysis was performed to identify the relationship between the variables. Table 1 shows the regression coefficients between the variables. VAR1 and PROB0-1 showed a clear correlation with MEAN1 while AC1 did not. THM randomly generates MEAN1 first, and then VAR1 ( $=\text{STD1}^2$ ) and PROB0-1 based on the generated MEAN1 according to the relationship identified through these regression analyses. AC1 was generated independently since it did not show a considerable correlation with the other statistics.

The histogram of MEAN1 for the month of June for gage FL-9148 and its corresponding fitted gamma distribution are shown in Fig. 5. The gamma distribution was

used to model MEAN1 as suggested by Ozturk (1981). The two parameters of the distribution were estimated using the method of maximum likelihood. Then, MEAN1 was randomly drawn from the gamma distribution with the estimated parameters. The histogram of AC1 observed at gage FL-9148 during the month of June is also shown in Fig. 5. Because of its near-symmetrical shape, a normal distribution was adopted to fit the distribution, from which AC1 is randomly drawn.

Because  $\text{STD1} (= \sqrt{\text{VAR1}})$  and PROB0-1 are highly correlated with MEAN1 (Fig. 4), VAR1 and PROB0-1 are generated based on MEAN1 using the result of the regression analysis in Table 1. For example, the following two equations specifically address the relationships between these variables for gage FL-9148 for the month of June:

$$\text{STD1} = 4.7203 \text{ Mean1} + 0.04973 + \varepsilon_1 \tag{9}$$

$$\text{PROB0} - 1 = -0.9718 \text{ Mean1} + 0.9583 + \varepsilon_2 \tag{10}$$

where  $\varepsilon_1$  and  $\varepsilon_2$  represent residual errors expected when using the regression equation.

$\varepsilon_1$  and  $\varepsilon_2$  in Eqs. (9) and (10) were randomly drawn from a normal distribution with mean 0 and variance equal to the error variance of the regression equation (Column “Var(-Error)” in Table 1).

A strong linear correlation was identified between the rainfall statistics at different temporal accumulation levels with the correlation coefficient ranging between 0.41 and 0.99. The result of the linear regression analysis is summarized in Table 1. The same methodology that was used for the generation of VAR1 and PROB0-1 from MEAN1 was used to generate the rainfall VAR, AC, and PROB0 at 3-, 12-, and 24-hourly accumulation levels: VAR1 is used for the generation of VAR3, which is, in turn, used for the generation of VAR12. VAR12, then again, is used as the basis of the generation of VAR24. The same principle was applied for the generation of PROB0-3, PROB0-12, and PROB0-24 from

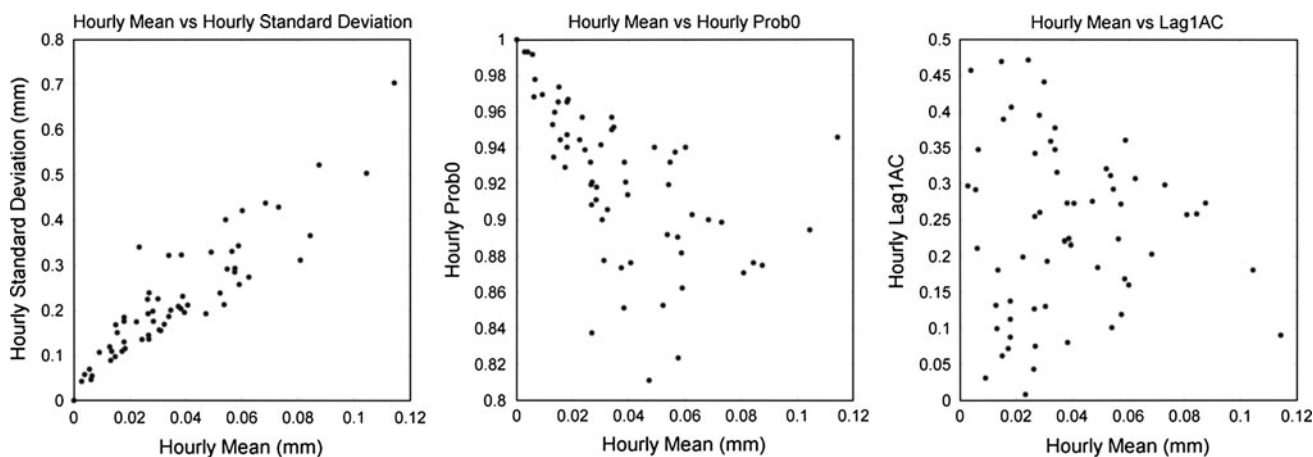


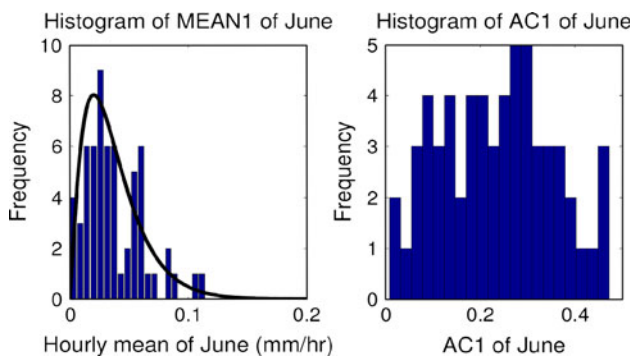
Fig. 4 Correlations between MEAN1 and STD1, MEAN1 and PROB0-1, and MEAN1 and AC1 at rainfall gage FL-9148 for the month of June



**Table 1** Results of the regression analysis between the rainfall statistics of the month of June at gage NCDC-FL9148

Independent variable (x)	Dependent (y)	a	b	Var (error)	R <sup>2</sup>
MEAN1	STD1(= $\sqrt{\text{VAR1}}$ )	4.7203	0.04973	0.003095	0.8222
MEAN1	PROB0-1	-0.9718	0.9583	0.0013	0.3236
MEAN1	AC1	-0.2059	0.2422	0.0142	0.0019
VAR1	VAR3	3.4488	0.0287	0.0017	0.9799
VAR3	VAR12	5.8397	0.0011	0.2463	0.9202
VAR12	VAR24	1.9743	0.0664	0.5789	0.9541
PROB0-1	PROB0-3	1.3929	-0.4056	0.0002	0.9592
PROB0-3	PROB0-12	1.6505	-0.7018	0.0013	0.8896
PROB0-12	PROB0-24	1.4899	-0.5256	0.0017	0.9368
AC1	AC3	0.4605	0.0748	0.0149	0.1692
AC3	AC12	0.5399	-0.0547	0.0255	0.17
AC12	AC24	0.5698	0.1757	0.0405	0.1975

The linear regression equation  $y = ax + b$  was assumed. Var(Error) represents the variance of the residuals or  $\left[ \frac{\sum (y_{\text{observed}} - y_{\text{regressed}})^2}{n-1} \right]$ .



**Fig. 5** Histogram of MEAN1 and AC1. The fitted curve of a gamma distribution for MEAN1 is shown along with the histogram (left side)

PROB0-1; and AC3, AC6, AC12, and AC24 from AC1. The generated statistics values were perturbed by the amount specified by the variance of error according to the regression analysis (Column “Var(Error)” in Table 1)

The described process is repeated for the number of months for which rainfall time series is generated. For example, if the model is to generate 50 years of rainfall time series, THM generates rainfall statistics sets for the 600 months (50 years  $\times$  12 months/year). This is the key concept of THM to integrate inter-annual variability of rainfall statistics, and it contrasts to the traditional approach which uses 12 sets (1 per each month regardless of the number of years being simulated) of observed rainfall statistics.

#### 4.3 Generation of rainfall time series using the MBLRP model

Once the rainfall statistics is generated as described in Sect. 4.2, the next step is to obtain the 6 parameters ( $\vec{\theta}$ ) of the

MBLRP model, which were calibrated using ISPSO (Cho et al. 2011) based on the rainfall statistics generated for the individual months using Eq. 8 as objective function. This procedure is repeated for the number of individual months to be used for simulation. For example, if 50 years of rainfall time series is to be simulated, the parameter calibration is repeated for 600 (=50  $\times$  12) times, which yields 600 model parameter set. Once the parameters are obtained, THM generates synthetic rainfall time series of each individual month using the MBLRP model based on the estimated parameter sets.

Because the generated rainfall time series for each of the months has the length of only one month, the statistics of it does not usually match the targeting rainfall statistics. To reduce these residuals in statistics, the process of rainfall generation is repeated for 20 times using a given parameter set, which yields 20 rainfall time series with the length of one month. Then, the rainfall time series with the lowest statistics residual is chosen as the finalist. The difference between the statistics of the simulated rainfall time series of the  $j$ th month and the target statistics is calculated as follow:

$$R_j = \sum_{i=1}^n \left[ (Z_i^{\text{simul}} - Z_i^{\text{target}})^2 \right] \quad (11)$$

,where  $Z_i^{\text{simul}}$  is the Z-score (departure of a given statistics from its global mean normalized by global standard deviation) of the  $i$ th statistics of the simulated rainfall time series and  $Z_i^{\text{target}}$  is the Z-score of the  $i$ th target statistics. Z-score was calculated based on the mean and standard deviation of the rainfall statistics that were estimated using the rainfall time series observed at 1,099 rain gages across the United States (Table 1 in Kim and Olivera 2011). This

study used MEAN1, VAR1, AC1, and PROB0-1 to calculate  $R_j$  ( $n = 4$ ). Then, the synthetic rainfall time series for the  $j$ th month with the lowest value of  $R_j$  is chosen as the finalist among the 20 months of generated rainfall time series.

#### 4.4 Model validation

The performance of THM was tested based on its ability to reproduce the distribution of the monthly maximum rainfall depths and extreme precipitation depth. A total of 132 months of precipitation data observed at 11 NCDC precipitation gages (12 months per one gage) across the coterminous United States (Fig. 3) were used for this validation procedure. All chosen gages have at least 50 years of records. For each of the 132 gage-months, 100 months of synthetic rainfall time series were generated using both THM and the conventional MBLRP modeling approach. Subsequently, each of the gage-months has three different types of rainfall time series including observed one. Then, monthly maximum rainfall depths with the duration of 1, 3, and 6 h were calculated. As a result, each gage-month is associated with 300 simulated monthly maximum rainfall depths accounting for statistic variability (i.e., 3 rainfall durations  $\times$  100 years of simulation) ( $P_{THM}^1, P_{THM}^3, P_{THM}^6$ ), 300 simulated monthly maximum rainfall depths not accounting for statistic variability (i.e., 3 rainfall durations  $\times$  100 years of simulation) ( $P_{Trad}^1, P_{Trad}^3, P_{Trad}^6$ ); and a number of monthly maximum observed rainfall depths (i.e., 3 rainfall durations  $\times$  years of record) ( $P_{Obs}^1, P_{Obs}^3, P_{Obs}^6$ ).

The two-sample Kolmogorov–Smirnov test (K–S test) was used to compare the distributions of the variables calculated from the observed rainfall time series and those calculated from the synthetic rainfall time series. The test statistic of the two-sample Kolmogorov–Smirnov test, which compares the distributions of data sets  $x_1$  and  $x_2$ , is as follows:

$$\max|F_1(x) - F_2(x)| \tag{12}$$

where  $F_1(x)$  is the proportion of data set  $x_1$  less than or equal to  $x$ . The null hypothesis of the test is that data sets  $x_1$  and  $x_2$  are from the same continuous distribution. Therefore, if the result of the test indicates that the null hypothesis is not rejected, one can say that data sets  $x_1$  and  $x_2$  are from the same continuous distribution with a given significance level that is specified in the test. In this study, a significance level of 5 % was used.

A set of two tests should be performed to tell if THM outperforms the traditional approach. For example, if a test comparing  $P_{Obs}^1$  and  $P_{THM}^1$  indicates that both variables are from the same continuous distributions and another test comparing  $P_{Obs}^1$  and  $P_{Trad}^1$  indicates that they are from different

distributions, the advantage of using THM over the traditional approach to predict the maximum precipitation depth at an hourly duration is justified. This set of tests was repeated for the three testing variables ( $P^1, P^3, P^6$ ) to see how the performance of THM compares to that of the traditional approach.

In addition to this rather qualitative approach to compare the models’ performance, each model’s performance to reproduce the extreme rainfall depth was quantitatively estimated. Frequency analysis was performed on  $P^1, P^3,$  and  $P^6$  to estimate extreme rainfall depth with 100-, 50-, and 30-months of recurrence interval. Here, it is noteworthy that the extreme rainfall depth has the frequency unit of month instead of year. For example, 100-month rainfall represents the rainfall depth that has the exceedance probability of 1 % of a given month. Another approach of estimating the rainfall depth with the frequency unit of year was considered, but such approach was not adopted because it disturbs the direct comparison between the variables by filtering out the maximum rainfall depths of the remaining 11 months.

Generalized logistic distribution (Asquith 1998) was used to model the distribution of the monthly maximum rainfall, and the method of L-moment (Hosking 1990) was used to estimate the parameters of the distribution. Then, the normalized residual of each model’s design precipitation estimate was calculated as follow:

$$RP_{trad}^{RI} = \frac{DP_{trad}^{RI} - DP_{obs}^{RI}}{DP_{obs}^{RI}} \tag{13}$$

$$RP_{THM}^{RI} = \frac{DP_{THM}^{RI} - DP_{obs}^{RI}}{DP_{obs}^{RI}} \tag{14}$$

,where RP represents the normalized residual of the extreme precipitation; DP represents the depth of the extreme precipitation; and superscript and subscript represent recurrence interval and the type of time series on which the calculation is based, respectively. If the distribution of these residuals is concentrated to 0, the advantage of using one model over the other can be justified.

## 5 Results and discussion

### 5.1 Reproduction of monthly maximum rainfall depth

Table 2 shows the performance of each model in reproducing the distribution of the observed monthly maximum rainfall depths. THM outperformed in reproducing the distributions of 3- and 6-hour duration monthly maximum precipitations while the traditional approach outperformed THM in reproducing the one with 1-hour duration. The success ratio of both approaches ranged between 45 and 81 %. Even though the values presented in Table 2 do not

**Table 2** Performance of THM and the traditional MBLRP approach in reproducing the distribution of extreme rainfall depths

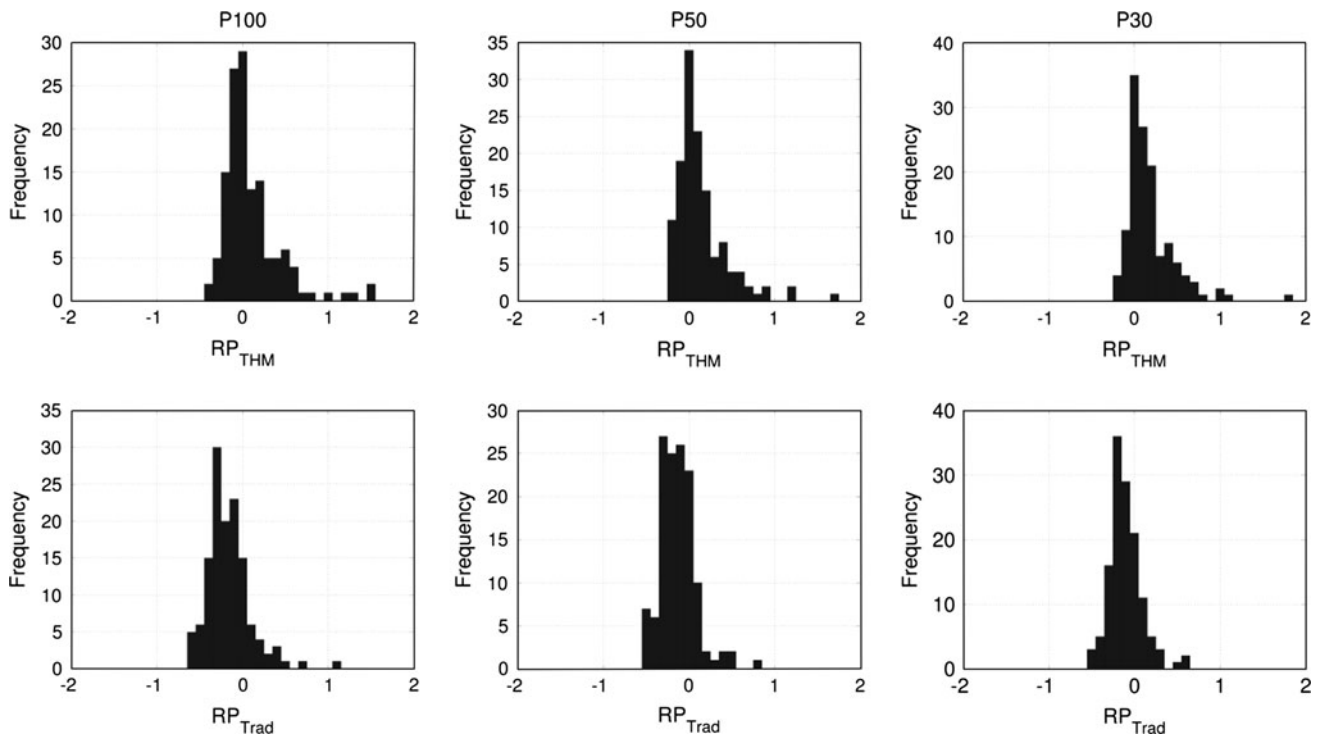
Rainfall duration	1 h	3 h	6 h
Both approaches succeeded	47	70	70
Only THM succeeded	12	21	37
Only traditional approach succeeded	29	19	17
Both failed	44	22	8
Overall THM success ratio	44.7 %	68.9 %	81.1 %
Overall traditional approach success ratio	57.6 %	67.4 %	65.9 %

explicitly prove the superiority of THM over the traditional modeling approach, they provide a general idea on the performance of both models in reproducing the distribution of the maximum rainfall depth. It also has to be noted that the result of the K–S test tends to be more sensitive near the center of the distribution than it is at the tails. This result suggests that considering inter-annual variability does not necessarily enhance the rainfall model’s performance to reproduce the overall distribution of the maximum rainfall depth.

### 5.2 Reproduction of extreme rainfall depth

The result of the analysis on extreme precipitation depth reproduction enlightens the perspective of THM that could

not be caught by the K–S test. Figure 6 shows the histograms of the normalized residual (Eqs 13 and 14) of 1-hour duration design precipitation with 100-, 50-, and 30-month recurrence interval for THM (upper row) and traditional MBLRP model (lower row). In the plots of Fig. 6, sharper shape of the distribution means higher model’s consistency. Also, model’s unbiasedness increases as the center of the distribution gets closer to 0. In general, it is observed that the distributions of  $RP_{Trad}$  for all three recurrence intervals are systematically biased to the left side of 0. This means that the traditional MBLRP approach systematically and significantly underestimates the extreme rainfall of the observed data. On the contrary, the level of biasedness of the distribution of  $RP_{THM}$  is significantly less than that of  $RP_{Trad}$ . The consistency of the model’s performance was similar for both of the approaches showing similar peakedness of the distribution. For some cases, THM had slightly less performance consistency compared to the traditional approach. However, the fact that the traditional approach has lower standard deviation than the suggested approach does not mean that it outperforms THM because it means the traditional approach underestimates the extreme rainfall values more consistently compared to THM. Another way to compare the model’s performance is to see the proportion that the estimated extreme rainfall depth (RP) falls within a given percentage (20 %) of the observed counterpart, which is shown in the 4th, 7th, and



**Fig. 6** Histograms of the normalized residual of the extreme precipitation depths reproduced by THM and the traditional approach of MBLRP rainfall modeling. From top to bottom, then left to right are shown the histograms of  $RP_{THM}^{100}$ ,  $RP_{THM}^{50}$ ,  $RP_{THM}^{30}$ ,  $RP_{Trad}^{100}$ ,  $RP_{Trad}^{50}$ , and  $RP_{Trad}^{30}$



**Table 3** Performance comparison between THM and the traditional approach on reproduction of design rainfall values

Recurrence interval	Type of statistics	THM			Traditional Approach		
		1 h	3 h	6 h	1 h	3 h	6 h
100 year	Mean	0.0963	0.0439	−0.0183	−0.1663	−0.1729	−0.1928
	Standard deviation	0.3350	0.2530	0.2064	0.2567	0.2337	0.1982
	Within $\pm 20\%$	0.6667	0.7424	0.7273	0.4091	0.4167	0.3939
50 year	Mean	0.1417	0.0730	0.0019	−0.1322	−0.1509	−0.1714
	Standard deviation	0.2992	0.2274	0.1905	0.2129	0.1931	0.1681
	Within $\pm 20\%$	0.6894	0.7500	0.7955	0.5076	0.5076	0.4394
30 year	Mean	0.1765	0.0945	0.0165	−0.1022	−0.1309	−0.1526
	Standard deviation	0.2841	0.2169	0.1860	0.1916	0.1710	0.1520
	Within $\pm 20\%$	0.6742	0.7576	0.8333	0.5985	0.5985	0.5227

10th row of Table 3. These values are significantly greater for THM. Table 3 also summarizes the mean, standard deviation of the distribution of RP. Similar result was obtained for all other comparison cases (extreme rainfall depth with 3 h and 6 h duration). The mean of  $RP_{THM}$  is notably closer to 0 compared to  $RP_{Trad}$  while the standard deviation is slightly greater for THM (again, this does not lead to the conclusion that the traditional approach outperforms THM). Also, the probability that the estimated RP values falls within  $\pm 20\%$  of the observed counterpart was significantly greater for  $RP_{THM}$ .

In Fig. 6, the histograms of  $RP_{THM}$  have stronger positive skew than those corresponding to  $RP_{Trad}$ , which also causes the greater variability of the  $RP_{THM}$  values. This means that the extreme precipitation simulated by THM is more likely to be greater than the observed one compared to the one estimated using the traditional method. The most probable cause of this problem is that the Gamma distribution adopted to model the mean rainfall for all gages and months. The distribution of the mean rainfall can follow other distributions especially because the gages used for the model validation are physically distant thus can have different climatic properties. Considering that mean monthly rainfall is used as the basis of the generation of the other statistics and the MBLRP model parameter calibration, choosing the right distribution type of the mean rainfall can greatly affect the accuracy of the model. This problem may be resolved if the correct type of distribution is customized for each of the gage-months based on the result of the statistical tests such as Kolmogorov–Smirnov test.

It also has to be noted that some recent developments in non-parametric stochastic rainfall generation has been successful in reproducing the extreme rainfall depths more closely (e.g., Westra et al. 2012). In this broader context of the stochastic rainfall generation, the contribution of this study in improving the ability of stochastic rainfall generation model in reproducing extreme rainfall values should be considered limited.

## 6 Conclusions

We proposed a novel approach of stochastic rainfall generation based on MBLRP model that can account for the inter-annual variability of rainfall statistics. While the newly proposed approach did not show apparent advantage over the traditional approach in reproducing the distribution of the maximum rainfall depth, it clearly outperformed the conventional one in reproducing extreme precipitation depths by reducing the systematic bias. This result indicates that the inter-annual variability of rainfall contains the important information about extreme values of precipitations that the lumped long-term monthly statistics could easily miss.

Numerous previous studies regarding Poisson cluster rainfall models have tried to enhance the model's ability to capture extreme rainfall. Cowpertwait (1998) modified the model such that it can explicitly count for the 3rd order moment of the observed rainfall time series. Evin and Favre (2012) refined the model structure by introducing the concept of transient storm arrival rate. These studies tried to refine the model structure to enhance the performance. In this context, the result of this study suggests that the performance of the rainfall generation model can be improved not only by altering or refining model structure but also by incorporating more information about the observed rainfall time series, especially its internal variability. We expect the same principle be applicable in other types of stochastic rainfall generators to enhance the performance.

**Acknowledgments** This work was supported by the Hongik University new faculty research support fund.

## References

- Asquith WH (1998) Depth-duration frequency of precipitation for Texas. US Geological Survey, Water-Resources Investigations Report 98-4044 (<http://pubs.usgs.gov/wri/wri98-4044>)

- Bo Z, Islam S, Eltahir EAB (1994) Aggregation-disaggregation properties of a stochastic rainfall model. *Water Resour Res* 30(12):3423–3435
- Botter G, Settin T, Marani M, Rinaldo A (2006) A stochastic model of nitrate transport and cycling at basin scale. *Water Resour Res* 42(4):1–5
- Burton A, Kilsby CG, Fowler HJ, Cowpertwait PSP, O’Connell PE (2008) RainSim: a spatial-temporal stochastic rainfall modeling system. *Environ Model Softw* 23:1356–1369
- Cho H, Kim D, Olivera F, Guikema SD (2011) Enhanced speciation in particle swarm optimization for multi-modal problems. *Eur J Oper Res* 213(1):15–23
- Cowpertwait PSP (1998) A Poisson-cluster model of rainfall: high-order moments and extreme values. *Proc R Soc Lond Ser A* 454:885–898. doi:10.1098/rspa.19980191
- Cowpertwait PSP, O’Connell PE, Metcalfe AV, Mawdsley JA (1996) Stochastic point process modelling of rainfall. I Single-site fitting and validation. *J Hydrol* 175(1–4):17–46
- Evin G, Favre AC (2012) Further developments of transient Poisson-cluster model for rainfall. *Stoch Environ Res Risk Assess* 2012:8. doi:10.1007/s00477-012-0612-y
- Glasbey CA, Cooper G, McGeahan MB (1995) Disaggregation of daily rainfall by conditional simulation from a point-process model. *J Hydrol* 165:1–9
- Hosking JRM (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. *J R Stat Soc Ser B* 52:105–124 JSTOR 2345653
- Isham S, Entekhabi D, Bras RL (1990) Parameter estimation and sensitivity analysis for the modified Bartlett-Lewis rectangular pulses model of rainfall. *J Geophys Res* 95(D3):2093–2100
- Kavvas ML and Delleur JW (1975) The stochastic and chronologic structure of rainfall sequences—application to Indiana, Technical Report 57. Water Resources Research Center, Purdue University, West Lafayette
- Khaliq M, Cunnane C (1996) Modelling point rainfall occurrences with the modified Bartlett-Lewis rectangular pulses model. *J Hydrol* 180:109–138
- Kim D, Olivera F (2012) On the relative importance of the different rainfall statistics in the calibration of stochastic rainfall generation models. *J Hydrol Eng* 17:368
- Laio F, Tamea S, Ridolfi L, D’Odorico P, Rodriguez-Iturbe I (2009) Ecohydrology of groundwater-dependent ecosystems: 1. Stochastic water table dynamics. *Water Resour Res* 45:W05419. doi:10.1029/2008WR007292
- Lall U, Sharma A (1996) A nearest neighbor bootstrap for resampling hydrological time series. *Water Resour Res* 32:679–693
- Lovejoy S, Schertzer D (1990) Multifractals, universality classes, and satellite and radar measurements of cloud and rain fields. *J Geophys Res* 95:2021–2031
- NCDC (2011) Precipitation Data, National Climatic Data Center (NCDC)—National Oceanic and Atmospheric Administration (NOAA). Available at <http://gis.ncdc.noaa.gov/map/precip/as> August 19, 2011
- Olsson J, Burlando P (2002) Reproduction of temporal scaling by a rectangular pulses rainfall model. *Hydrol Process* 16:611–630
- Onof C, Wheater HS (1994) Improvements to the modeling of British rainfall using a modified random parameter Bartlett-Lewis rectangular pulse model. *J Hydrol* 157(1–4):177–195
- Onof C, Northrop P, Wheater HS, Isham V (1996) Spatiotemporal storm structure and scaling property analysis for modeling. *J Geophys Res Atmospheres* 101:26415–26425
- Onof C, Chandler RE, Kakou A, Northrop P, Wheater HS, Isham V (2000) Rainfall modelling using Poisson-cluster processes: a review of developments. *Stoch Environ Res Risk Assess* 14(6):384–411
- Ozturk A (1981) On the study of a probability-distribution for precipitation totals. *J Appl Meteorol* 20(12):1499–1505
- Rodriguez-Iturbe I, Cox DR, Isham V (1987) Some models for rainfall based on stochastic point processes. *Proc R Soc Lond Ser A* 410(1839):269–288
- Rodriguez-Iturbe I, Cox DR, Isham V (1988) A point process model for rainfall: further developments. *Proc R Soc Lond Ser A* 417(1853):283–298
- Tarboton DG, Sharma A, Lall A (1998) Disaggregation procedures for stochastic hydrology based on nonparametric density estimation. *Water Resour Res* 34(1):107–119
- Verhoest N, Troch PA, De Troch FP (1997) On the applicability of Bartlett-Lewis rectangular pulses models in the modeling of design storms at a point. *J Hydrol* 202(1–4):108–120
- Westra SP, Mehrotra R, Sharma A, Srikanthan R (2012) Continuous rainfall simulation: 1. A regionalized subdaily disaggregation approach. *Water Resour Res* 48:W01535-1-W01535-16
- Wheater HS et al (2005) Spatial-temporal rainfall modeling for flood risk estimation. *Stoch Environ Res Risk Assess* 19(6):403–416
- Yoo C, Kim D, Kim TW, Hwang KN (2008) Quantification of drought using a rectangular pulses Poisson process model. *J Hydrol* 355(1–4):34–48