Research Article

A Poisson Cluster Stochastic Rainfall Generator That Accounts for the Interannual Variability of Rainfall Statistics: Validation at Various Geographic Locations across the United States

Dongkyun Kim,[1] Jongho Kim,[2] and Yong-Sik Cho[3]

[1] Department of Civil Engineering, School of Urban and Civil Engineering, Hongik University, Wausan-ro 94,
Sangsu-dong, Mapogu, Seoul 121-791, Republic of Korea
[2] Department of Civil and Environmental Engineering, University of Michigan, 2350 Hayward Street, 2340 GG Brown Building,
Ann Arbor, MI 48109-2125, USA
[3] Department of Civil and Environmental Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong-gu,
Seoul 133-791, Republic of Korea

Correspondence should be addressed to Yong-Sik Cho; ysc59@hanyang.ac.kr

Received 13 June 2014; Accepted 6 August 2014; Published 27 August 2014

Academic Editor: Daoyi Xu

A novel approach for a Poisson cluster stochastic rainfall generator was validated in its ability to reproduce important rainfall and watershed response characteristics at 104 locations in the United States. The suggested novel approach, The Hybrid Model (THM), as compared to the traditional Poisson cluster rainfall modeling approaches, has an additional capability to account for the interannual variability of rainfall statistics. THM and a traditional approach of Poisson cluster rainfall model (modified Bartlett-Lewis rectangular pulse model) were compared in their ability to reproduce the characteristics of extreme rainfall and watershed response variables such as runoff and peak flow. The results of the comparison indicate that THM generally outperforms the traditional approach in reproducing the distributions of peak rainfall, peak flow, and runoff volume. In addition, THM significantly outperformed the traditional approach in reproducing extreme rainfall by 2.3% to 66% and extreme flow values by 32% to 71%.

1. Introduction

Stochastic rainfall generators provide synthetic rainfall input to hydrologic simulation models whenever the observed data with sufficient length are not available. Because they enable the Monte-Carlo simulation approach by providing an infinite length of rainfall time series to hydrologic simulation models, they are extensively utilized to assess the risks associated with hydrologic systems. Poisson cluster stochastic rainfall generation models [1, 2] are considered to be the most robust and practical stochastic rainfall generators because of their model structure that reflects well the seasonal and climatological features of rainfall generating mechanisms [3]. The performance of Poisson cluster rainfall models in reproducing conventional rainfall statistics such as mean, variance, autocorrelation, and probability of dry periods has

been well validated over various geographical locations across the world [4–11]. For this reason, Poisson cluster rainfall models have been applied in a wide range of practices for hydrological risk assessments dealing with flooding (e.g., [12]), drought (e.g., [13]), contaminant transport (e.g., [14]), and ecosystem behavior (e.g., [15]).

The Poisson cluster model has been constantly improved after Rodriguez-Iturbe et al. [1] have suggested the original model structure and parameter calibration scheme. For example, Rodriguez-Iturbe et al. [2] introduced an additional parameter that can account for the storm-to-storm variability of rain cell duration. Cowpertwait [17, 18] derived analytical expressions for the probability of a dry h-hr period and skewness of a synthetically generated rainfall time series, respectively, which can subsequently be used for the calibration of model parameters. Velghe et al. [19] and Onof et al. [20]
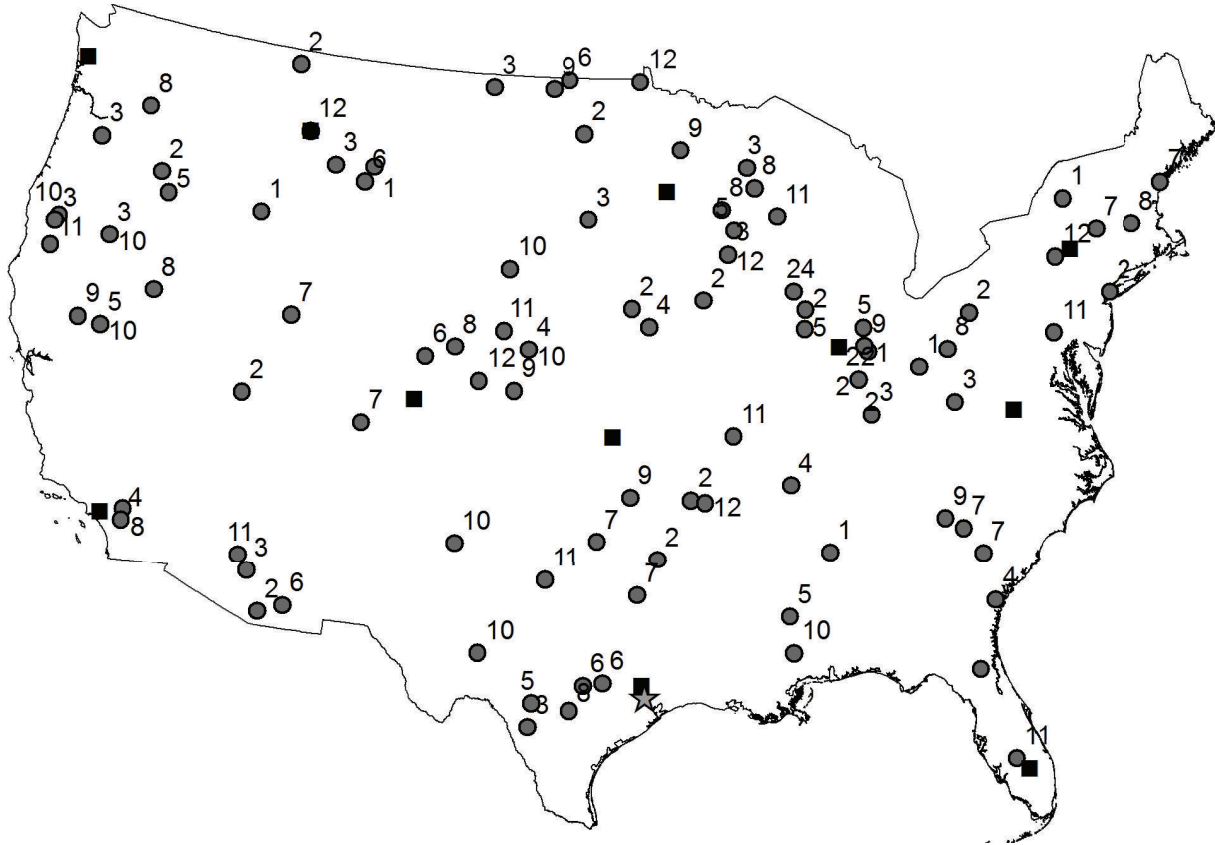
FIGURE 1: Location of the 104 US National Climate Data Center (NCDC) precipitation stations used for the validation of THM in the present study (gray circles). The numbers next to the dots represent the calendar month on which the validation was performed. The locations of the 11 gauges validated by Kim et al. [16] are shown as black squares. The location of the Texas rainfall gauge (NCDC ID 4300) used to generate Figure 3 is also shown (gray star).

replaced the one-parameter exponential distribution with the two-parameter gamma distribution to represent the distribution of rain cell intensity more accurately. Cowpertwait [17] proposed a model that can account for the rainfall process in which both convective and stratiform events simultaneously exist.

While most of these previous studies regarding Poisson cluster rainfall generators tried to enhance the performance of the model by modifying the fundamental assumptions of the model structures, Kim et al. [16, 21] indicated that the performance of the model can be enhanced not only by modifying the model structure, but also by providing more information about the rainfall process. Particularly, they suggested that the interannual variability of rainfall is highly associated with extreme rainfall events and that the conventional calibration scheme of a Poisson cluster rainfall model cannot account for this interannual variability. They suggested an approach that can take this variability into account, which was termed The Hybrid Model (THM). THM successfully reduced the systematic bias of extreme rainfall and flow values that exist in a synthetic rainfall time series generated by the traditional approach of Poisson cluster rainfall modeling. However, the performance of the model was tested only at 11 rain gauges across the United States

(black squares in **Figure 1**) for 12 calendar months (total of 132 months), so a general conclusion regarding the advantage of using THM over the traditional approach of Poisson cluster modeling can be constituted only after it is verified over various rainfall characteristics. For this reason, the present study tested THM for an additional 104 geographic locations across the United States. The results of this study are expected to expand the applicability of THM, especially in climatic regions that cannot be represented by the 11 gauges that were analyzed by Kim et al. [16, 21].

2. Methodology

2.1. Data Description. A total of 104 months of precipitation data observed at 104 US National Data Climate Center [22] precipitation gauges (one month per one gauge) across the contiguous United States (gray circles in **Figure 1**) were used in the analysis. The gauges were randomly drawn from the pool of all NCDC rain gauges that contain at least 50 years of records to make sure that the rainfall statistics represent well the rainfall characteristics at the gauges.

2.2. Modified Bartlett Lewis Rectangular Pulse (MBLRP) Model. THM shares its fundamental model structure with
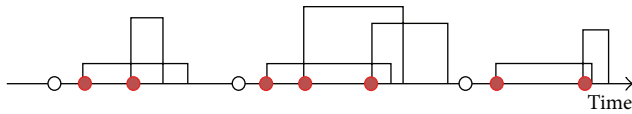
FIGURE 2: Schematic of the MBLRP model. White and gray circles represent the arrival time of storms and rain cells, respectively. Each rain cell is represented by a rectangle whose width and height represent rainfall duration and rainfall intensity.

the Modified Bartlett-Lewis rectangular pulse (MBLRP) model [2]. In the MBLRP model, the rainfall time series are represented as sequences of storms comprised of rain cells (see Figure 2). In the model, X1 [T] is a random variable that represents the storm arrival time, which is governed by a Poisson process with parameter λ; X2 [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 γ; X3 [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 β; X4 [T] is a random variable that represents the duration of the rain cells, which varies according to an exponential distribution with parameter η that, in turn, is a random variable represented by a gamma distribution with parameters ν and α; and X5 [L/T] is a random variable that represents the rain cell intensity, which varies according to an exponential distribution with parameter $1/\mu$. From a physical viewpoint, $\lambda[1/T]$ is the expected number of storms that arrives in a given period, $\gamma[1/T]$ is the inverse of the expected duration of storm activity, $\beta[1/T]$ is the expected number of rain cells that arrives within the duration of storm activity, $\eta[1/T]$ is the inverse of the expected duration of the rain cells, and $\mu[L/T]$ is the average rain cell intensity. Parameters $\nu[T]$ and α [dimensionless] do not have a clear physical meaning, but the expected value and variance of η can be expressed as α/ν and α/ν^2. Therefore, the model has six parameters: λ, γ, β, ν, α, and μ; however, it is customary to use the dimensionless ratios $\varphi = \gamma/\eta$ and $\kappa = \beta/\eta$ as parameters instead of γ and β. The estimation of the model parameters is accomplished by matching statistics of the simulated and observed rainfall time series. Some commonly used statistics are the precipitation depth mean, variance, probability of zero rainfall, and lag-s covariance at various time scales [8, 23].

2.3. The Hybrid Model (THM). The fundamental idea of THM arose from the fact that the rainfall statistics that are considered in the conventional approach of Poisson cluster rainfall modeling do not contain enough information about the variable characteristics of the rainfall process. Especially, it is noteworthy that the rainfall statistics that are used in the calibration process of the conventional approach are typically calculated for the entire period of a rainfall time series, overlooking the fact that they vary from year to year. See, for example, Figure 3. Each plot in the figure shows the monthly variation hourly rainfall statistics (mean, variance, lag-1 autocorrelation, and probability of zero rainfall) at the gauge

TX-4300 (star in Figure 1). Thirty-one years of continuous hourly rainfall data (1976–2005) was used to generate the plots. Small dots in each plot of Figure 3 represent the rainfall statistics of a given month of a given year (e.g., hourly mean rainfall of June, 1977). In contrast, hollow circles connected with a solid line represent the rainfall statistics of a given month over the entire recording period (e.g., hourly mean rainfall of June between the years of 1976 and 2005). It can be noted that the interannual variability of rainfall statistics (vertical spread of dots in each plot of Figure 3) is significant. The existing framework of Poisson cluster rainfall models only uses the average long-term statistics (hollow circles in Figure 3) that ignore the interannual variability. For this reason, the rainfall time series that is generated by the conventional approach cannot reflect the interannual variability of the rainfall. Considering that hydrologically important events such as floods and droughts are associated with marginal statistics that are far from its long-term mean, it is critical that stochastic rainfall models include an algorithm that accounts for the interannual statistical variability. Kim et al. [16, 21] proposed an algorithm to resolve this issue, which introduces an additional process of modeling the interannual variability of rainfall statistics to the conventional approach of Poisson cluster rainfall simulation.

The modeling framework of the traditional approach and THM approach is shown and compared in Figure 4. While the traditional approach generates the rainfall time series for a long period of time (e.g., 1000 months) based on the long-term rainfall statistics (namely, rainfall statistics corresponding to the entire period of the observed record), THM generates the rainfall time series with a short period of time (e.g., 1 month) based on the short-term rainfall statistics (namely, rainfall statistics corresponding to one calendar month).

THM firstly simulates the short-term rainfall statistics. Here, short-term statistics refer to the statistics for the time period of one calendar month. For example, if a rainfall time series is to be generated for the month of January for 100 times, THM first generates a 100-set of short-term rainfall statistics. The simulated short-term rainfall statistics are mean rainfall at 1 hourly accumulation level, variance, lag-1 autocorrelation coefficient, and probability of zero rainfall at 1, 3, 12, and 24 hourly accumulation levels (total of 13 statistics).

The sequence of generation of the 13 rainfall statistics is as follows, which is also depicted in Figure 5.

(1) Randomly draw mean (MEAN1) and autocorrelation coefficient (AC1) at hourly accumulation levels from the predetermined normal and gamma distributions, respectively. Figure 6 shows a sample histogram and the fitted distributions at the gauge TX-4300 for the month of June. From these distributions, hourly mean and lag-1 autocorrelation coefficients are drawn.

(2) Based on the randomly drawn mean rainfall at hourly accumulation levels (MEAN1), generate variance (VAR1 = STDEV1^2) and probability of zero rainfall (PROB1) at hourly accumulation levels using the correlation between the variables identified through
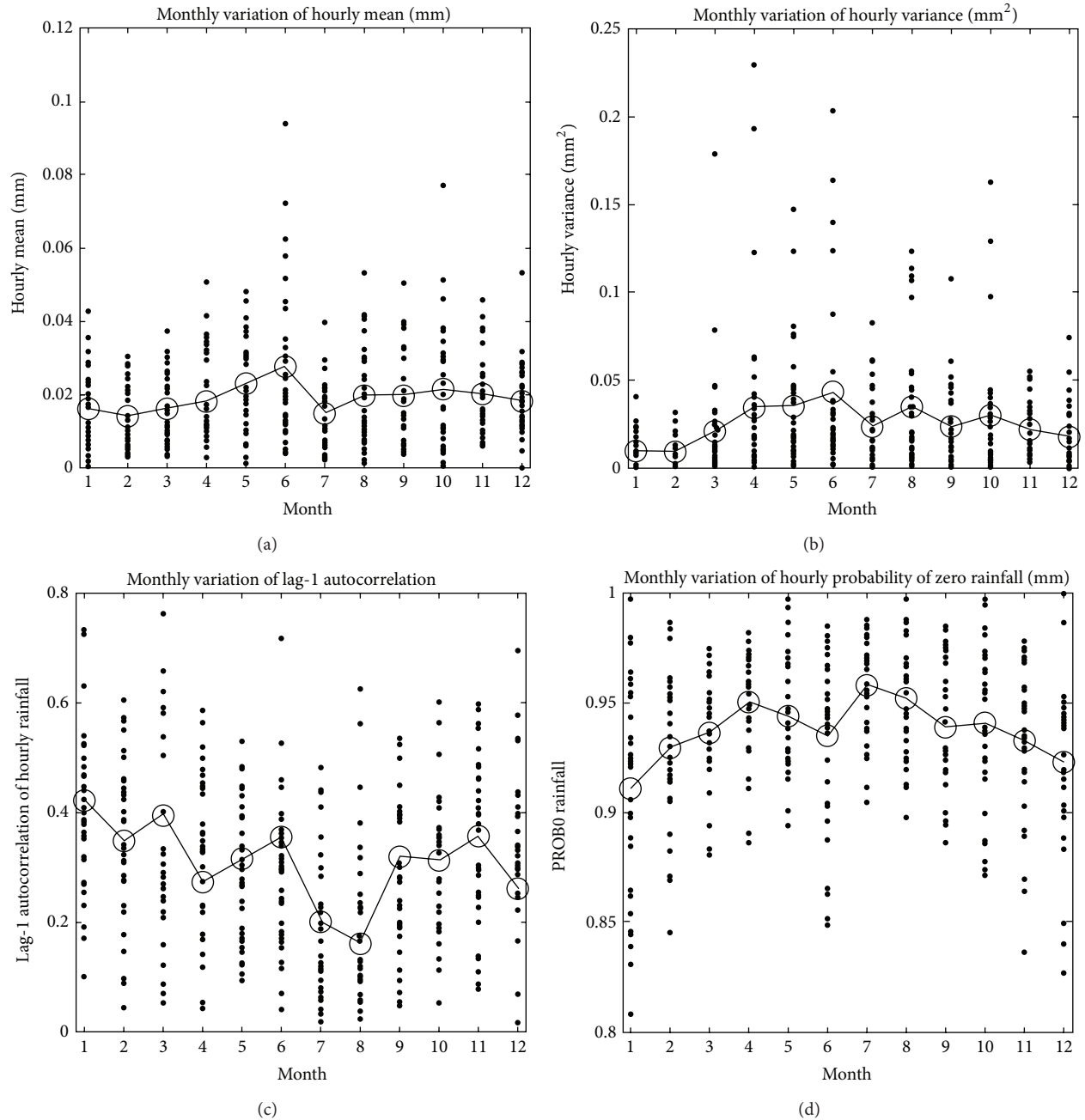
(a)

(b)

(c)

(d)

FIGURE 3: Monthly variation in hourly rainfall statistics at the gauge TX-4300. Small dots in each plot represent rainfall statistics corresponding to a given month of a given year. Hollow circles connected with a solid line represent the rainfall statistics for a given month over the entire recording period.

linear regression analysis. Figure 7 shows this correlation and the least square regression lines.

(3) Based on the generated VAR1 and PROB1, generate the variance and probability of zero rainfall at 3, 12, and 24 hours of accumulation levels (VAR3, VAR12, VAR24, PROB3, PROB12, and PROB24, resp.) using the correlation between the variables identified through linear regression analysis. Figure 8 shows this correlation and the least square regression lines.

(4) Based on the generated AC1 in step (1), generate the autocorrelation coefficients at 3, 12, and 24 hours of accumulation levels (AC3, AC12, and AC24, resp.) using the correlation between the variables identified through linear regression analysis. Figure 6 shows this correlation and the least square regression lines.

The second part of THM simulates the rainfall time series using the Modified Bartlett-Lewis rectangular pulse model [2] based on the simulated short-term rainfall statistics.
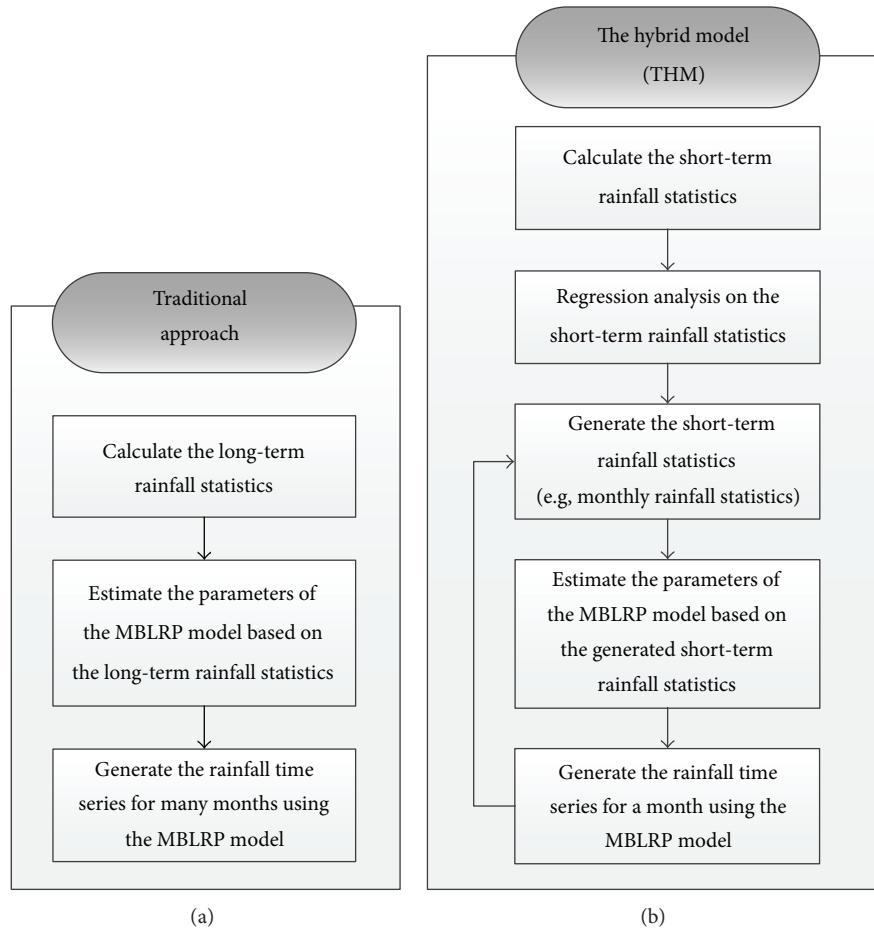
(a)   (b)

FIGURE 4: Schematic of the traditional approach of Poisson cluster rainfall modeling (a) and the approach of this study (b).
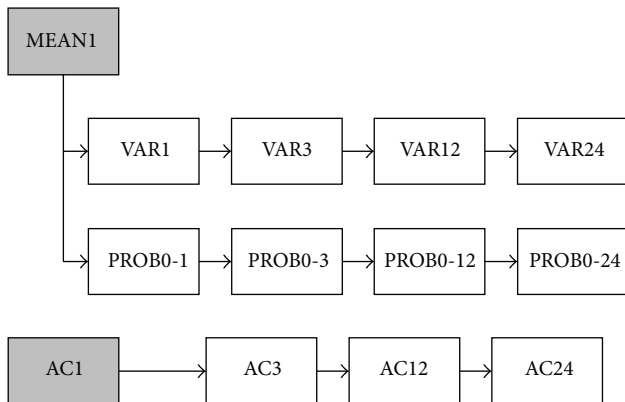

FIGURE 5: The sequence of a short-term rainfall statistics generation.

Firstly, the 6 parameters of the MBLRPM corresponding to each simulated month are estimated using isolated particle swarm optimization (ISPSO, [24]). Then, 20 rainfall time series with the length of one month are generated based on the estimated parameters. Lastly, the rainfall time series of which statistics is the closest to the short-term rainfall statistics that were used in parameter estimation is selected.

For a more detailed description on the methodology of THM, readers can refer to Kim et al. [16].

2.4. Comparison of the Two Models in Reproducing the Distribution of the Observed Rainfall and the Corresponding Watershed Responses. The performance of THM was tested in its ability to reproduce the distribution of monthly maximum rainfall depths, monthly peak flows, and monthly runoff volumes. For each of the chosen months of the 104 gauges, 100 months of synthetic rainfall time series were generated using both THM and the traditional approach of the MBLRP model. Accordingly, each of the gauges has three different types of rainfall time series including the observed ones. Then, the following values were calculated using all three types of rainfall time series: monthly maximum rainfall depths with the duration of 1, 3, 6, 12, and 24 hours; monthly runoff depth; and monthly peak flow. The SCS curve number method and SCS curvilinear unit hydrograph method [25] were used to calculate the last two values. The watershed characteristics assumed were a lag time of 2 hours, drainage area of 7.5 km^2, and a curve number of 50, 60, 70, 80, and 90.

As a result, each gauge was associated with 500 monthly maximum rainfall depths based on THM (i.e., 5 rainfall durations × 100 years of simulation—P_{THM}^1, P_{THM}^3, P_{THM}^6,
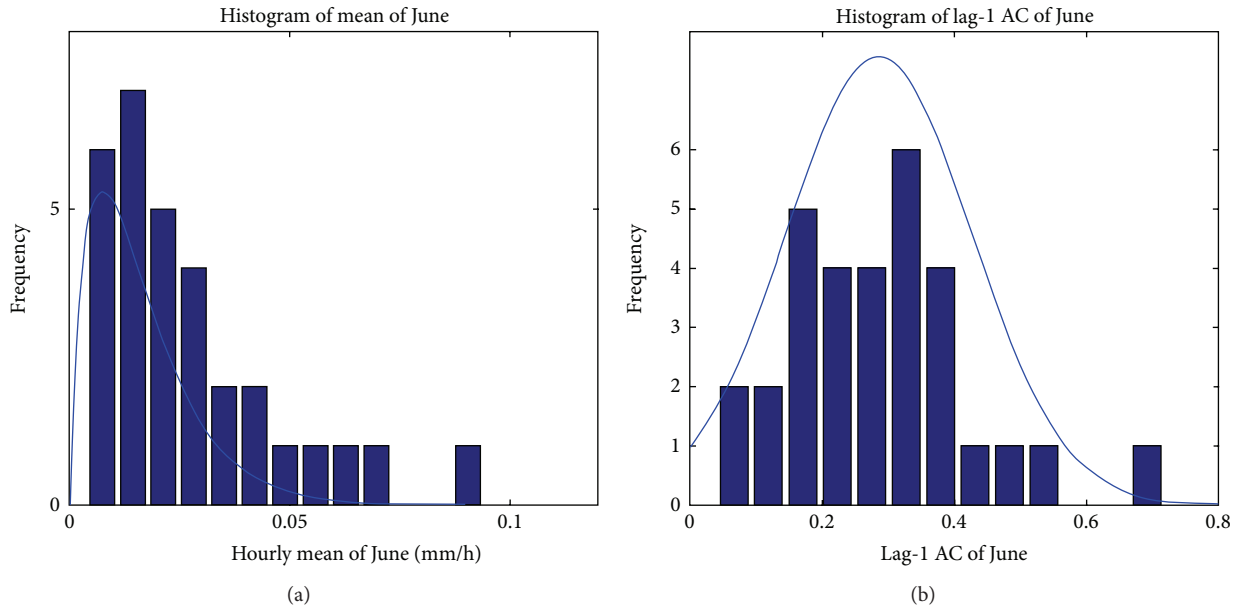
FIGURE 6: Histograms and the fitted gamma and normal distribution of hourly mean (a) and lag-1 autocorrelation coefficients (b) at the gauge TX-4300 for the month of June. From these distributions, MEAN1 and AC1 are randomly drawn.
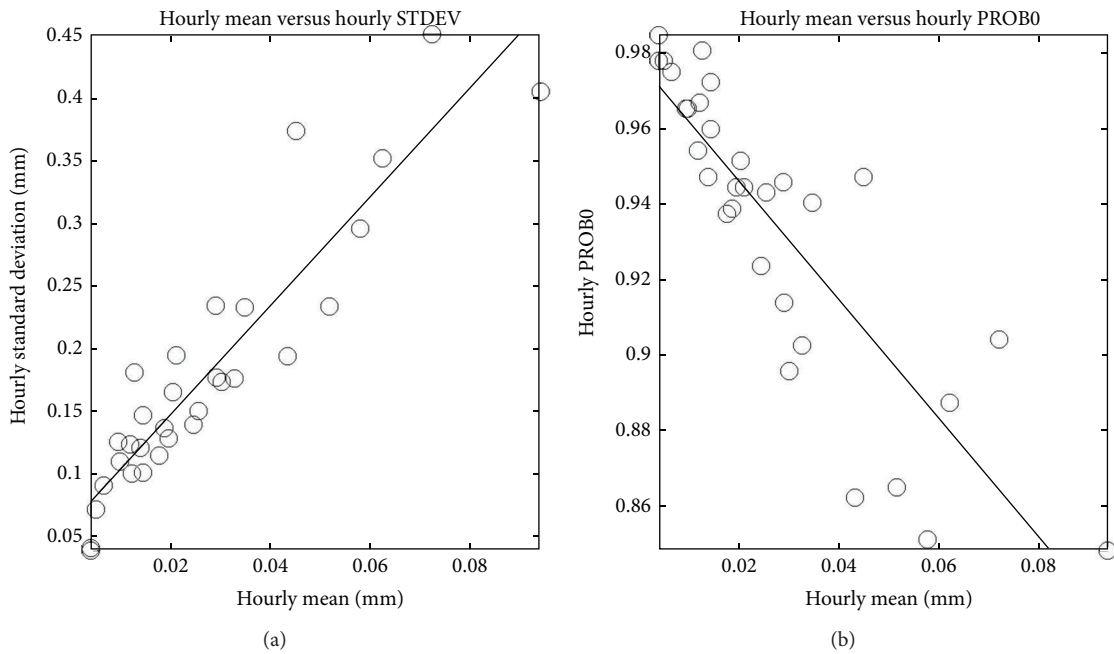

FIGURE 7: Relationship of MEAN1 to VAR1 (a) and PROB1 (b) and the least square regression lines at the gauge TX-4300 for the month of June.

P_{THM}^{12}, P_{THM}^{24}); 500 monthly maximum rainfall depths based on the traditional MBLRP model approach (i.e., 5 rainfall durations × 100 years of simulation—P_{MBLRP}^{1}, P_{MBLRP}^{3}, P_{MBLRP}^{6}, P_{MBLRP}^{12}, P_{MBLRP}^{24}); a number of monthly maximum observed rainfall depths (i.e., 5 rainfall durations × years of record—P_{Obs}^{1}, P_{Obs}^{3}, P_{Obs}^{6}, P_{Obs}^{12}, P_{Obs}^{24}); 5 (different curve numbers) sets of 100 monthly runoff depths based on THM (R_{THM}^{50},

R_{THM}^{60}, R_{THM}^{70}, R_{THM}^{80}, R_{THM}^{90}); 5 sets of 100 monthly runoff depths based on the traditional approach (R_{MBLRP}^{50}, R_{MBLRP}^{60}, R_{MBLRP}^{70}, R_{MBLRP}^{80}, R_{MBLRP}^{90}); 5 sets of +50 monthly runoff depths based on observed rainfall time series (R_{Obs}^{50}, R_{Obs}^{60}, R_{Obs}^{70}, R_{Obs}^{80}, R_{Obs}^{90}); 5 (different curve numbers) sets of 100 monthly peak flows on THM (F_{THM}^{50}, F_{THM}^{60}, F_{THM}^{70}, F_{THM}^{80}, F_{THM}^{90}); 5 sets of 100 monthly peak flows based on the traditional approach
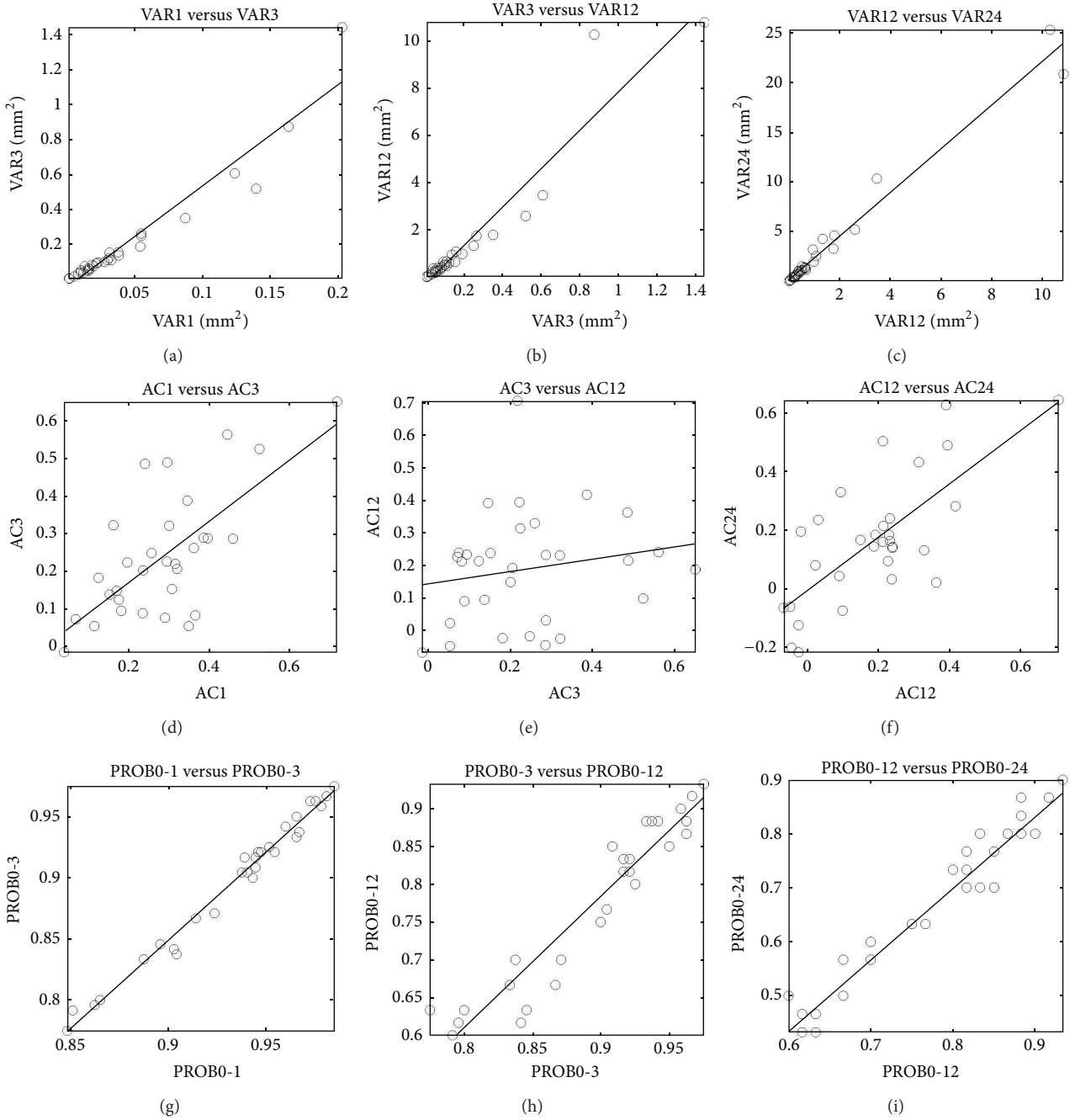
FIGURE 8: Relationship between the rainfall statistics at 1, 3, and 12 hourly accumulation levels and the least square regression lines at the gauge TX-4300 for the month of June.

(F_{MBLRP}^{50}, F_{MBLRP}^{60}, F_{MBLRP}^{70}, F_{MBLRP}^{80}, F_{MBLRP}^{90}); and 5 sets of +50 monthly peak flows based on observed rainfall time series (F_{Obs}^{50}, F_{Obs}^{60}, F_{Obs}^{70}, F_{Obs}^{80}, F_{Obs}^{90}).

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 the ones calculated from the synthetic rainfall time series. The test statistic of the two-sample Kolmogorov-Smirnov test, which compares

the distributions of the data set X_1 and the data set X_2, is as follows:

$$\max \left| F_1(X) - F_2(X) \right|, \tag{1}$$

where $F(X)$ is the empirical cumulative density function of the value X. The null hypothesis of the test is that the sets x_1 and x_2 are from the same continuous distribution. Therefore, if the results of the test indicate that the null hypothesis is not

rejected, one can say that set x_1 and set 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.

In this study, a set of two tests should be performed to tell if THM outperforms the traditional approach. For example, if the test comparing P^1_{Obs} and P^1_{THM} indicates that both variables are from the same continuous distributions and the test comparing P^1_{Obs} and P^1_{Trad} indicates that they are from the different distributions, the advantage of using THM over the traditional approach to predict the maximum precipitation depth at hourly duration is proved. This set of tests was repeated for the 15 variables (P^1, P^3, P^6, P^{12}, P^{24}, R^{50}, R^{60}, R^{70}, R^{80}, R^{90}, F^{50}, F^{60}, F^{70}, F^{80}, and F^{90}) to see how the performance of THM compares to the one of the traditional approach based on long-term statistics.

2.5. Comparison of the Two Models in Reproducing the Extreme Rainfall and the Peak Flow Values. The result of the K-S test only tells the overall similarity or difference between the two distributions. However, extreme events are more highly associated with the upper tail of the distribution than they are to the overall shape of the distribution. In other words, there can be a case in which the simulated rainfall time series cannot reproduce the extreme events even if the result of the K-S test indicates the similarity of the distributions. For this reason, the design rainfall and the corresponding peak flow values at the virtual watershed with some given recurrence intervals were calculated for each of the rainfall time series and were compared with each other. A generalized extreme value distribution was used to model the distribution of the monthly peak rainfall and the monthly peak flow, and the method of L-moment [26] was used to estimate the parameters of the distribution. Then, the residual of each model's design precipitation was normalized as follows:

$$RP^{RI}_{model} = \frac{DP^{RI}_{model} - DP^{RI}_{observed}}{RP^{RI}_{observed}}, \qquad (2)$$

$$RQ^{RI}_{model} = \frac{DQ^{RI}_{model} - DQ^{RI}_{observed}}{RQ^{RI}_{observed}}, \qquad (3)$$

where RP and RQ represent the normalized residual of the design precipitation and the corresponding design flow at the virtual watershed, respectively, DP and DQ represent the estimated design precipitation and design flow, respectively, and superscript and subscript attached to the letter DP and DQ represent recurrence interval and the type of time series on which the calculation is based (either the MBLRPM, THM, or observed), respectively. A value of RP or RQ that is close to 0 means that the extreme precipitation or the extreme flow with a given recurrence interval produced by the rainfall model is close to their observed counterpart. In addition, a positive (or negative) value of RP or RQ means that the extreme precipitation or extreme flow with a given recurrence interval produced by the rainfall model is greater (or smaller) than their observed counterpart, and vice versa. For example, a RQ^{50}_{MBLRPM} value of -0.1 means that the 50-year flow value
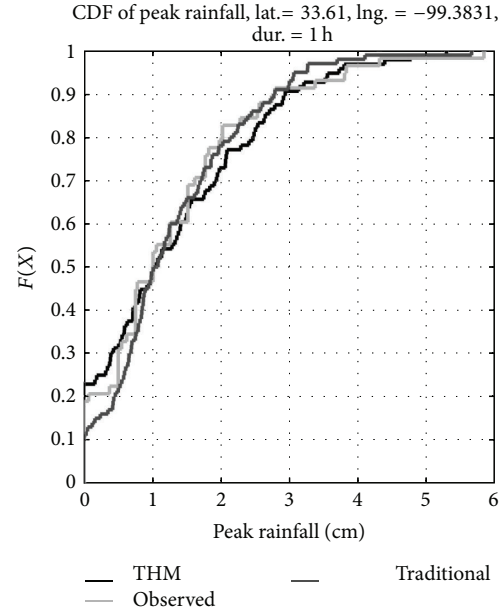

FIGURE 9: The empirical cumulative density function of the monthly peak rainfall depth made from the rainfall time series generated by THM, the traditional approach of MBLRPM, and observation.

that is reproduced by the traditional MBLRPM model is 10% smaller than its observed counterpart. In this study, this residual analysis was performed for all 104 gauges and a histogram of the residuals was prepared. The centeredness and peakedness near the value of 0 in the shape of this histogram can be a good measure of the overall performance of the model in terms of reproducing extreme rainfall and flow values.

3. Results

3.1. Reproduction of Annual Maximum Rainfall Distribution and Extreme Rainfall Depth. Figure 9 compares the cumulative density function (CDF) of the monthly peak rainfall depth with a duration of 1 hour derived from the simulated and the observed rainfall time series at the NCDC rainfall gauge located in northern Texas (Latitude = 33.61, Longitude = −99.38) for the month of November. It can be noted that both the traditional method and THM reproduce well the distribution of the monthly peak rainfall depth of the observed rainfall time series.

To acquire more general conclusions, the K-S test was performed for all 104 geographical locations and for all 5 rainfall durations. Figure 10 summarizes the results of the K-S test analysis. The figure shows the proportions among the 104 gauges that THM and the traditional approach succeeded in reproducing the distribution of the observed monthly maximum rainfall depths with 1, 3, 6, 12, and 24 hours of duration.

It is notable that THM and the traditional approach succeeded in reproducing hourly maximum rainfall depth only for 47 and 63% of the entire 104 stations, respectively.

Success ratio of reproducing monthly peak rainfall depth distribution

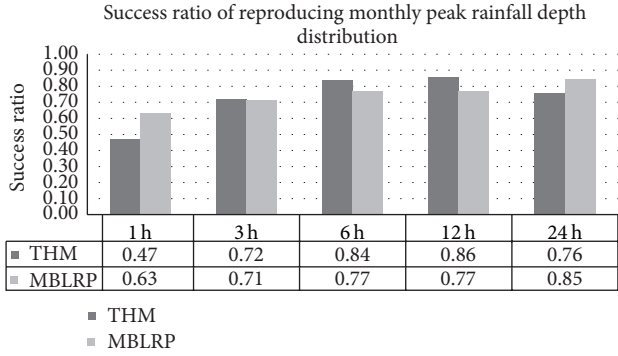|	1 h	3 h	6 h	12 h	24 h
■ THM	0.47	0.72	0.84	0.86	0.76
■ MBLRP	0.63	0.71	0.77	0.77	0.85

■ THM
■ MBLRP

Figure 10: The proportions among 104 gauges that THM and the traditional approach succeeded in reproducing the distribution of the observed monthly maximum rainfall depths with 1, 3, 6, 12, and 24 hours of duration.

This seems to be particularly because the monthly maximum rainfall with 1-hour duration is only a small portion of the time series and the statistics used in the model calibration (mean, variance, lag-1 autocorrelation, and probability of zero rainfall) does not directly reflect this event. For this reason, the reproduction of extreme values using Poisson cluster models has been the primary subject for many studies. For example, Cowpertwait [18] described the limitation of Poisson cluster rainfall models in which only a single set of parameters can be applied to represent the characteristics of both extreme and frequent rainfall events. In the meantime, the effort to identify the patterns of the statistical properties, in which both approaches have a high probability of failure in reproducing 1-hour duration maximum rainfall depth, was successful; see Figure 11.

Each plot in the figure shows the relationship between the hourly statistical values (e.g., mean and variance) and the success index in reproducing the monthly maximum rainfall depth with 1-hour duration. The density of circles in the first column of the first plot increases with the decrease in mean hourly rainfall. This suggests that both approaches are more likely to fail in a relatively dry region. In a similar manner, the second plot suggests that both approaches are more likely to fail when the hourly variance of the rainfall time series decreases. Considering the high correlation between rainfall mean and variance (meaning that a dry region is mostly associated with low rainfall depth variability), it can be generalized that Poisson cluster models should be used with caution when used to model extreme events in relatively dry regions (e.g., mean rainfall less than 0.1 mm/hr).

Secondly, THM outperformed the traditional approach in reproducing the distribution of the monthly maximum rainfall depths with a duration starting from 3 to 12 hours, while it did not for the 1-hour duration rainfall. This is particularly because the extreme values were better represented by the short-term rainfall statistics that were additionally introduced by THM as the duration of the extreme rainfall increased. In other words, THM, which incorporates more statistical information about the interannual variability than the traditional approach does, has more information on

the maximum rainfall depths, resulting in better performance. An analysis that is similar to the previous one was performed to identify the pattern of statistical properties in which THM particularly outperforms the traditional approach, but no notable pattern was observed.

Figure 12 shows the histograms of the normalized residuals of the 1-hour design precipitation (2) for THM (a) and the MBLRPM (b). It can be noted that the center of the histogram corresponding to THM is nearer to the value of 0 for all three recurrence intervals of design precipitation, while the one corresponding to the traditional approach is biased toward the left. This means that the extreme rainfall values generated by THM are closer to the observed ones, while the traditional approach consistently underestimates the observed extreme rainfall values. Table 1 shows the mean and standard deviation of the RP values for 1-hour, 3-hour, and 6-hour design precipitation values.

The mean of RP_{THM} values was consistently closer to 0 compared to RP_{MBLRPM} for all three rainfall durations, which means that the overall performance of THM in reproducing the extreme rainfall values (namely, the upper tail part of the distribution) of the observed rainfall time series. The standard deviation of RP_{THM} values was consistently greater than that of RP_{MBLRPM}. This indicates that the traditional MBLRP model has greater consistency, but it also means that the traditional approach underestimates the extreme rainfall values more consistently compared to THM. The last column of Table 1 shows the proportion among all 104 gauge locations that the model design precipitation value had less than 20% discrepancy from the observed design precipitation. The values are consistently greater for THM. This means that THM has a greater probability of successfully reproducing the extreme rainfall values than does the traditional MBLRP model. It has been also noted that the performance of both THM and the traditional approach was better for the rainfall with the greater accumulation interval. This seems to be because the conceptualization of the rainfall process of MBLRPM is more appropriate for the greater accumulation interval. In other words, the gradual increase of the rainfall rate at a given point at hourly accumulation level which occurs in reality cannot be modeled well using the current MPLRPM and THM framework because they conceptualize the arrival of the rainfall as an abrupt process using time-intensity rectangles in time axis.

3.2. Reproduction of Annual Maximum Peak Flow Distribution and Extreme Flow.

Figure 13 shows the proportions, among 104 gauges, that THM and the traditional approach succeeded in reproducing the distribution of the peak flow at the virtual watershed with varying values of curve numbers between 50 and 90. For all of the cases, THM slightly outperformed the traditional approach. This result is encouraging in that the tested hypothetical watershed has 2 hours of lag time and THM did not outperform the traditional approach in reproducing a 2-hour duration maximum rainfall. The traditional approach notably outperformed the THM for the 1-hour duration and THM slightly outperformed
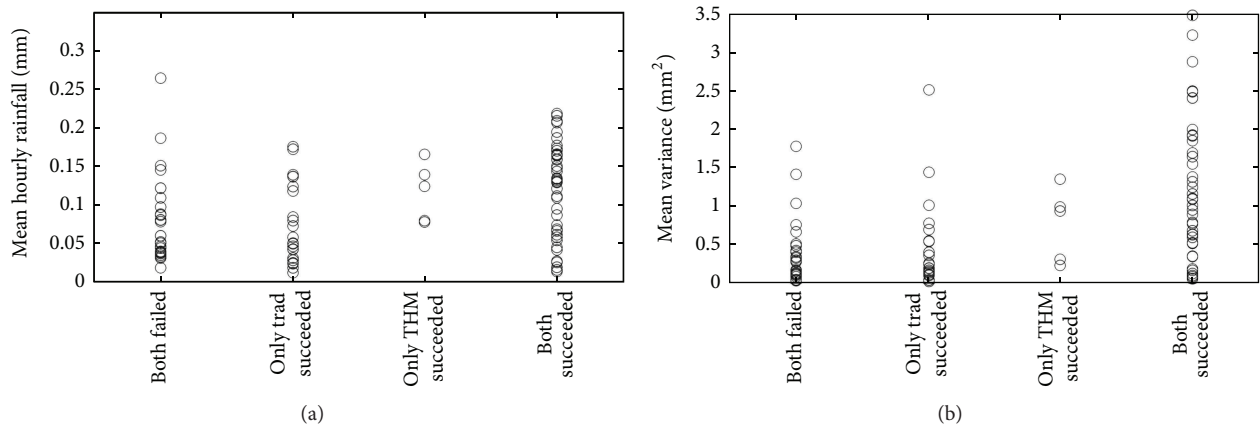
(a)

(b)

FIGURE 11: Relationship between the hourly rainfall statistics and whether THM and traditional approaches succeeded in reproducing maximum monthly rainfall depth with 1-hour duration.
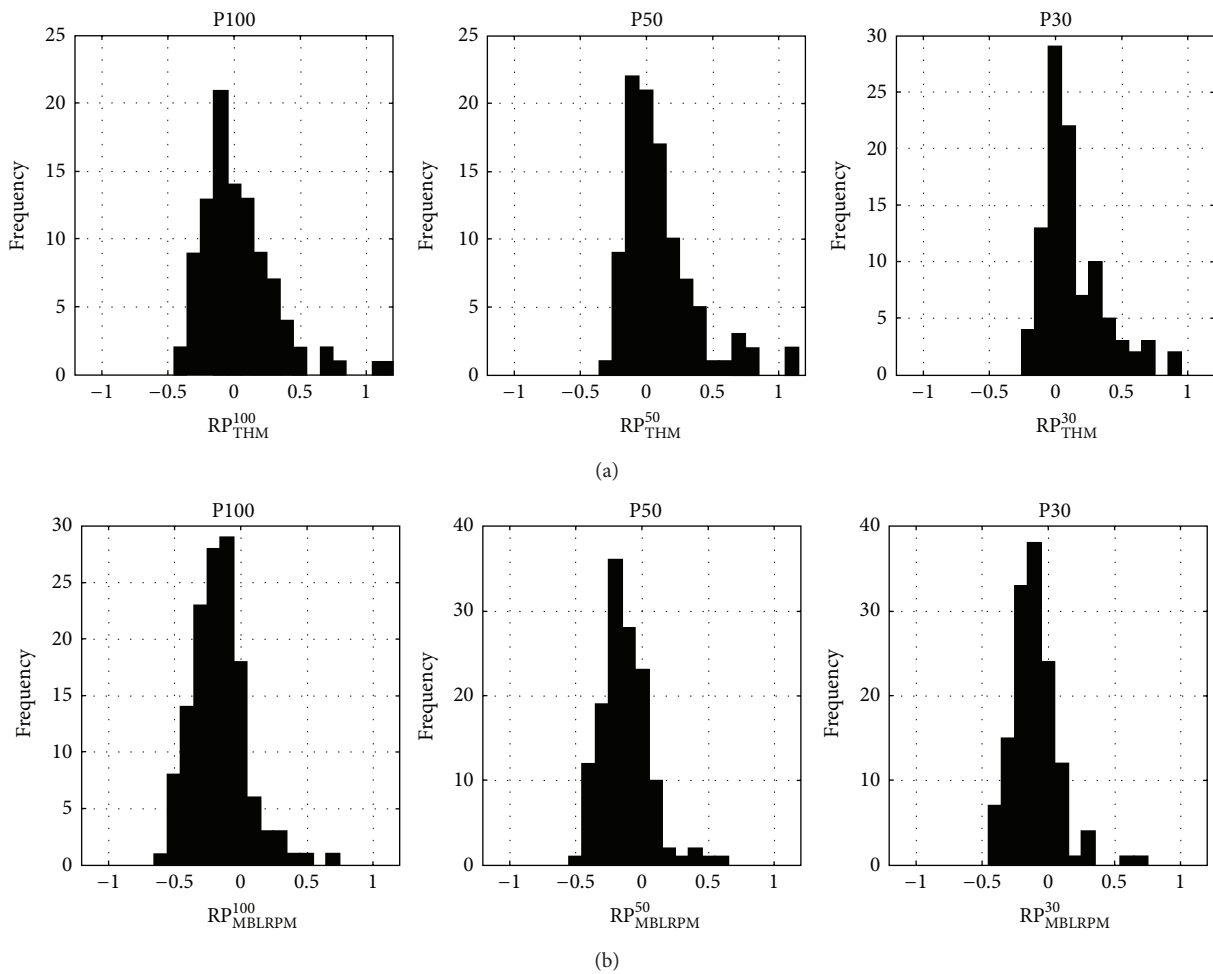

(a)

(b)

FIGURE 12: Histogram of the residual of the design precipitation with duration of 1 hour (RP^{100}, RP^{50}, and RP^{30}) for THM (a) and the MBLRPM (b).

TABLE 1: Performance comparison between THM and the traditional approach in reproducing design rainfall values.

| Recurrence interval	Type of statistic	THM			Traditional approach		
		1 hours	3 hours	6 hours	1 hour	3 hours	6 hours
100 years	Mean	0.058	0.022	−0.035	−0.157	−0.162	−0.189
	Standard deviation	0.310	0.296	0.234	0.210	0.193	0.157
	Within ±20%	0.640	0.710	0.730	0.522	0.522	0.441
50 years	Mean	0.097	0.046	−0.019	−0.127	−0.143	−0.168
	Standard deviation	0.261	0.240	0.191	0.182	0.164	0.140
	Within ±20%	0.670	0.790	0.800	0.610	0.596	0.515
30 years	Mean	0.129	0.065	−0.006	−0.101	−0.126	−0.150
	Standard deviation	0.233	0.210	0.168	0.172	0.152	0.136
	Within ±20%	0.700	0.810	0.860	0.684	0.684	0.610
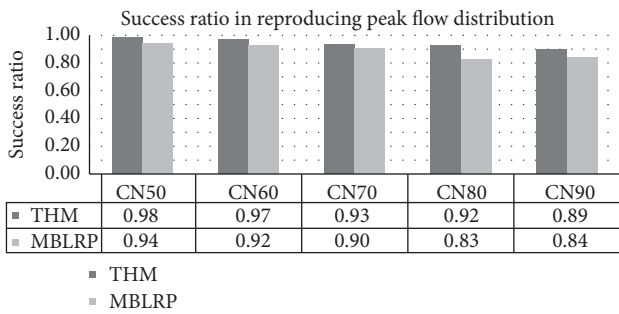
FIGURE 13: The proportions, among 104 gauges, that THM and the traditional approach succeeded in reproducing the distribution of the observed peak flows of the hypothetical watershed with curve numbers of 50, 60, 70, 80, and 90.
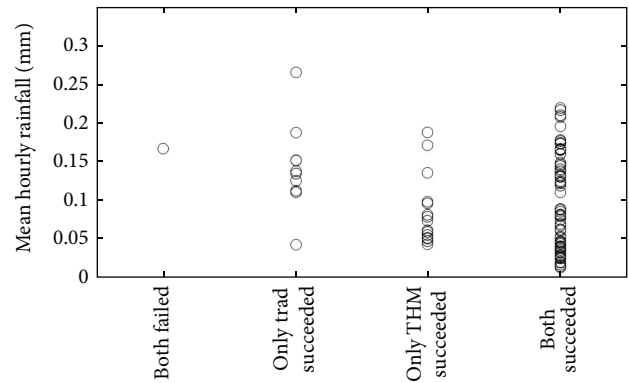

FIGURE 14: Relationship between the hourly rainfall mean and whether THM and traditional approaches succeeded in reproducing peak flow at the virtual watershed with curve number 90.

the traditional approach for the three hour duration. See Figure 10.

It is also noteworthy that the success ratio of peak flow distribution reproduction was significantly greater than that of the extreme rainfall depth distribution reproduction for both models. The reason may be because the watershed response is not only a function of peak flow depth but also a function of the rainfall depth that occurs around the time of a peak rainfall event, and Poisson cluster models are good at reproducing the latter. Also, the amount of discrepancy between the observed extreme rainfall and the model rainfall decreased as the rainfall converted into runoff due to the infiltration process (the so-called the "damping effect").

An analysis was performed to identify the statistical pattern at which THM outperforms the traditional approach. Figure 14 shows the relationship between the hourly rainfall mean and the success index of reproducing the monthly peak flows at the virtual watershed with curve number 90. It can be noted that the most gauge locations at which only THM succeeded in reproducing peak flow values had a rainfall mean of less than ~0.1 mm/hr. A similar pattern was observed for the analyses based on virtual watersheds with different curve numbers.

Figure 15 shows the histograms of the RF values for the virtual watershed with a curve number of 90. The plots in the upper row show the ones corresponding to THM, and

the plots in the lower row show the ones corresponding to the traditional MBLRP model. It can be seen that the histograms corresponding to THM are more centered near the value of 0, meaning that the overall performance of reproducing extreme peak flow values is better for THM compared to the traditional MBLRP model. Table 2 summarizes the results of this histogram analysis. It can be noted that the degree of underestimation of extreme flow values for the traditional MBLRP approach is significant, with the RQ value varying between −0.34 and −0.47, which means that the extreme flow reproduced by the traditional MBLRP model can be smaller than the observed flow by 34% to 47%. This degree of underestimation was significantly reduced by employing THM, which varied between −0.12 and −0.23. The last column of Table 2 shows the proportions among all 104 gauge locations where the model design flow value had less than 30% residual from the observed design flow, and the values were consistently greater for THM. This means that THM has a greater probability of successfully reproducing extreme flow values than the traditional MBLRP model does.

3.3. Reproduction of Runoff Depth Distribution. Figure 16 shows the proportions, among 104 gauges, that THM and the traditional approach succeeded in reproducing the distribution of the monthly runoff depth at the hypothetical

TABLE 2: Performance comparison between THM and the traditional approach in reproducing design flow values.

| Recurrence interval	Type of statistic	THM			Traditional approach		
		CN = 70	CN = 80	CN = 90	CN = 70	CN = 80	CN = 90
100 years	Mean	−0.184	−0.178	−0.117	−0.468	−0.420	−0.344
	Standard deviation	0.606	0.521	0.334	0.388	0.373	0.312
	Within ±20%	0.370	0.460	0.680	0.257	0.324	0.397
50 years	Mean	−0.225	−0.187	−0.123	−0.477	−0.429	−0.348
	Standard deviation	0.518	0.530	0.335	0.388	0.368	0.314
	Within ±20%	0.370	0.450	0.680	0.265	0.324	0.471
30 years	Mean	−0.226	−0.191	−0.126	−0.481	−0.432	−0.349
	Standard deviation	0.529	0.542	0.342	0.392	0.370	0.322
	Within ±20%	0.340	0.450	0.660	0.257	0.324	0.463
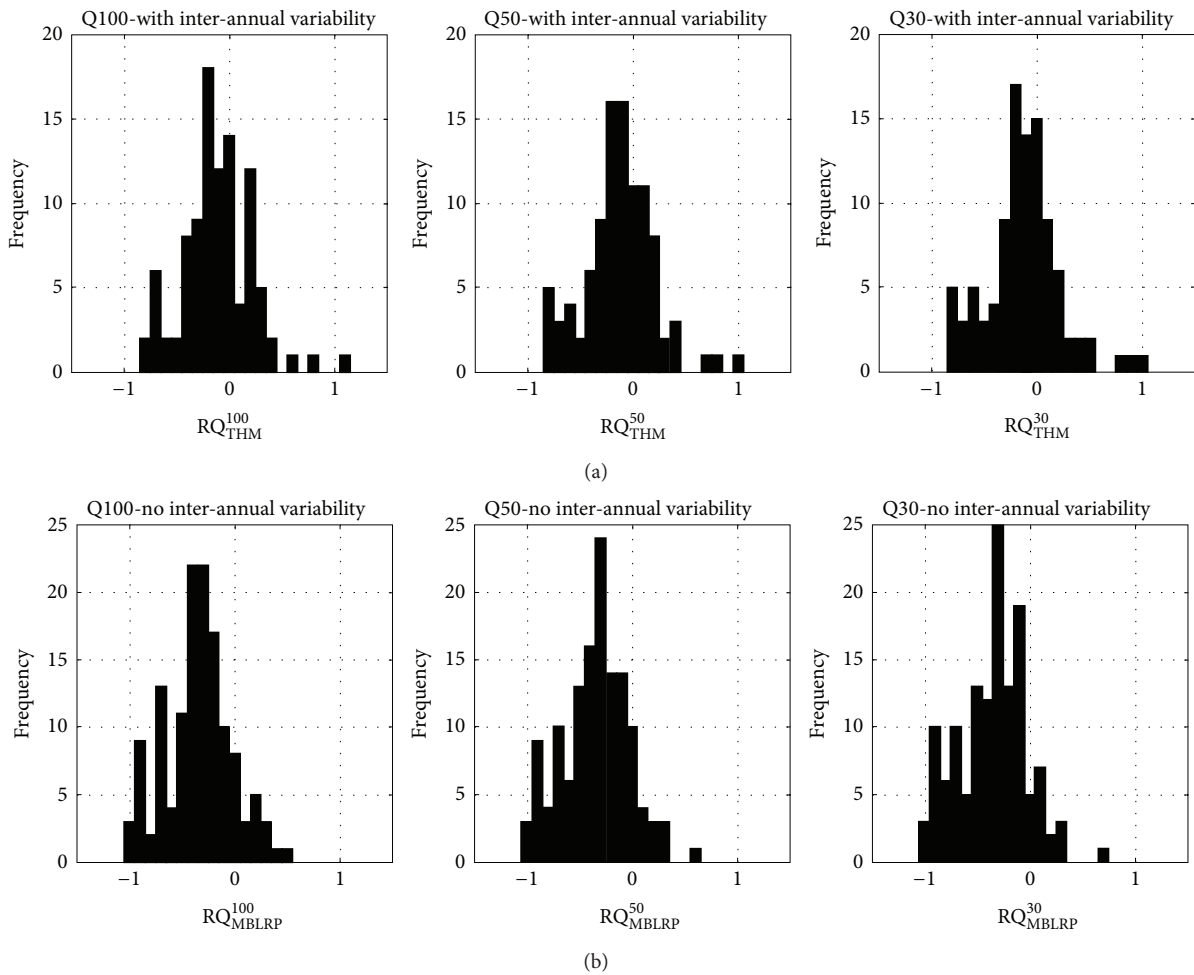
FIGURE 15: Histogram of the residual of design flow at the virtual watershed with curve number 90 (RF^{100}, RF^{50} and RF^{30}) for THM (a) and the MBLRPM (b).

watershed with varying values of curve numbers between 50 and 90. For all cases, THM notably outperformed the traditional approach. As opposed to maximum rainfall depths and peak flow values, runoff volume is more closely related to the rainfall statistics, and THM, which incorporates more rainfall statistics than the traditional approach, had

an improved performance in reproducing the distribution of runoff volumes.

An analysis was performed to identify the statistical pattern at which THM outperforms the traditional approach in reproducing runoff volume. Similar to the analysis of peak flow reproduction, it was observed that the most gauge

Success ratio in reproducing monthly runoff depth distribution

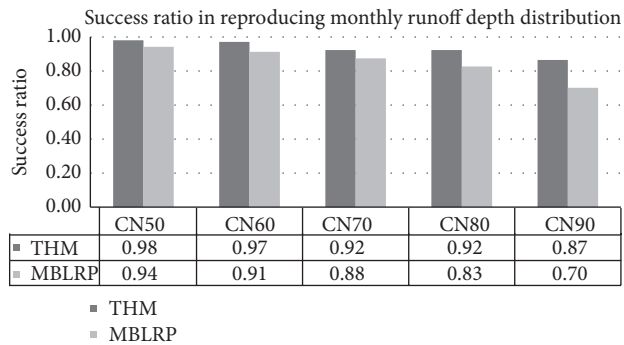|	CN50	CN60	CN70	CN80	CN90
■ THM	0.98	0.97	0.92	0.92	0.87
■ MBLRP	0.94	0.91	0.88	0.83	0.70

■ THM
■ MBLRP

FIGURE 16: The proportions, among 104 gauges, that THM and the traditional approach succeeded in reproducing the distribution of the observed runoff depths of the hypothetical watershed with curve numbers of 50, 60, 70, 80, and 90.

locations at which only THM succeeded in reproducing peak flow values had a rainfall mean of less than ~0.1 mm/hr.

Overall, this result indicates the advantage of using THM for continuous rainfall simulation in which not only the extreme values, but also the overall runoff value is considered important. In the meantime, it should be noted that the results regarding watershed response analysis were obtained using the SCS curve number method, which may yield inaccurate result in continuous watershed modeling. The analysis based on fully distributed hydrologic models such as the ones used in the Distributed Model Intercomparison Project—phase 1 (DMIP1) (http://www.nws.noaa.gov/oh/hrl/dmip/) might be able to provide more physically-based results provided with sufficient information on meteorological forcings, soil types, land use, topography, and initial conditions of any interested domain. However, the distributed continuous models were not adopted in this study because it is not only infeasible to obtain all relevant information but also computationally intensive to run the models for many locations.

4. Conclusion

In this study, the performance of a Poisson cluster stochastic rainfall generator that is capable of accounting for the interannual variability of rainfall statistics was validated over the various geographic locations across the contiguous United States. The results of the present study confirmed that the traditional approach using only a few number of "long-term" rainfall statistics calculated for the entire period of record cannot sufficiently represent the rainfall characteristics of a given calendar month varying from one year to another. Such an identified problem was resolved by using the newly suggested approach. The latter additionally included the process of simulating "short-term" rainfall statistics that varies from one year to another and combined the process into the traditional approach of Poisson cluster rainfall modeling. The suggested novel approach (The Hybrid Model, THM) generally outperformed the conventional approach of the Poisson cluster rainfall simulation when modeling the characteristics of extreme precipitation, extreme flood, and

runoff volume, while it did not outperform the traditional modeling approach in reproducing extreme rainfall depths on a fine time scale (e.g., 1 hour). The present study is especially meaningful in that it extended the applicability of THM by validating it over various geographic locations across the contiguous United States, in addition to what was investigated by Kim et al. [16, 21].

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

References

[1] I. Rodriguez-Iturbe, D. R. Cox, and V. Isham, "Some models for rainfall based on stochastic point processes," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 410, no. 1839, pp. 269–288, 1987.

[2] I. Rodriguez-Iturbe, D. R. Cox, and V. Isham, "A point process model for rainfall: further developments," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 417, no. 1853, pp. 283–298, 1988.

[3] J. Olsson and P. Burlando, "Reproduction of temporal scaling by a rectangular pulses rainfall model," *Hydrological Processes*, vol. 16, no. 3, pp. 611–630, 2002.

[4] Z. Bo, S. Islam, and E. A. B. Eltahir, "Aggregation-disaggregation properties of a stochastic rainfall model," *Water Resources Research*, vol. 30, no. 12, pp. 3423–3435, 1994.

[5] P. S. P. Cowpertwait, P. E. O'Connell, A. V. Metcalfe, and J. A. Mawdsley, "Stochastic point process modelling of rainfall. I. Single-site fitting and validation," *Journal of Hydrology*, vol. 175, no. 1–4, pp. 17–46, 1996.

[6] C. A. Glasbey, G. Cooper, and M. B. McGechan, "Disaggregation of daily rainfall by conditional simulation from a point-process model," *Journal of Hydrology*, vol. 165, no. 1–4, pp. 1–9, 1995.

[7] S. Islam, D. Entekhabi, R. L. Bras, and I. Rodriguez-Iturbe, "Parameter estimation and sensitivity analysis for the modified Bartlett-Lewis rectangular pulses model of rainfall," *Journal of Geophysical Research*, vol. 95, no. D3, pp. 2093–2100, 1990.

[8] M. N. Khaliq and C. Cunnane, "Modelling point rainfall occrrences with the modified Bartlett-Lewis rectangular pulses model," *Journal of Hydrology*, vol. 180, no. 1–4, pp. 109–138, 1996.

[9] C. Onof, P. Northrop, H. S. Wheater, and V. Isham, "Spatiotemporal storm structure and scaling property analysis for modeling," *Journal of Geophysical Research D: Atmospheres*, vol. 101, no. 21, pp. 26415–26425, 1996.

[10] C. Onof and H. S. Wheater, "Improvements to the modelling of British rainfall using a modified Random Parameter Bartlett-Lewis Rectangular Pulse Model," *Journal of Hydrology*, vol. 157, no. 1–4, pp. 177–195, 1994.

[11] N. Verhoest, P. A. Troch, and F. P. De Troch, "On the applicability of Bartlett-Lewis rectangular pulses models in the modeling

of design storms at a point," *Journal of Hydrology*, vol. 202, no. 1–4, pp. 108–120, 1997.

[12] H. S. Wheater, R. E. Chandler, C. J. Onof et al., "Spatial-temporal rainfall modelling for flood risk estimation," *Stochastic Environmental Research and Risk Assessment*, vol. 19, no. 6, pp. 403–416, 2005.

[13] C. Yoo, D. Kim, T. W. Kim, and K. N. Hwang, "Quantification of drought using a rectangular pulses Poisson process model," *Journal of Hydrology*, vol. 355, no. 1'4, pp. 34–48, 2008.

[14] G. Botter, T. Settin, M. Marani, and A. Rinaldo, "A stochastic model of nitrate transport and cycling at basin scale," *Water Resources Research*, vol. 42, no. 4, Article ID W04415, 2006.

[15] F. Laio, S. Tamea, L. Ridolfi, P. D'Odorico, and I. Rodriguez-Iturbe, "Ecohydrology of groundwater-dependent ecosystems: 1. Stochastic water table dynamics," *Water Resources Research*, vol. 45, no. 5, Article ID W05419, 2009.

[16] D. Kim, F. Olivera, and H. Cho, "Effect of the inter-annual variability of rainfall statistics on stochastically generated rainfall time series: part 1. Impact on peak and extreme rainfall values," *Stochastic Environmental Research and Risk Assessment*, vol. 27, no. 7, pp. 1601–1610, 2013.

[17] P. S. P. Cowpertwait, "A generalized point process model for rainfall," *Proceedings of the Royal Society of London A—Mathematical Physical and Engineering Sciences*, vol. 447, no. 1929, pp. 23–37, 1994.

[18] P. S. P. Cowpertwait, "A Poisson-cluster model of rainfall: high-order moments and extreme values," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 885–898, 1998.

[19] T. Velghe, P. A. Troch, F. P. de Troch, and J. van de Velde, "Evaluation of cluster-based rectangular pulses point process models for rainfall," *Water Resources Research*, vol. 30, no. 10, pp. 2847–2857, 1994.

[20] C. Onof, R. E. Chandler, A. Kakou, P. Northrop, H. S. Wheater, and V. Isham, "Rainfall modelling using poisson-cluster processes: a review of developments," *Stochastic Environmental Research and Risk Assessment*, vol. 14, no. 6, pp. 384–411, 2000.

[21] D. Kim, F. Olivera, H. Cho, and S. O. Lee, "Effect of the inter-annual variability of rainfall statistics on stochastically generated rainfall time series. Part 2. Impact on watershed response variables," *Stochastic Environmental Research and Risk Assessment*, vol. 27, no. 7, pp. 1611–1619, 2013.

[22] NCDC, NCDC: National Climate Data Center, Climate Data Online, 2014, http://www.ncdc.noaa.gov/cdo-web/.

[23] D. Kim and F. Olivera, "Relative importance of the different rainfall statistics in the calibration of stochastic rainfall generation models," *Journal of Hydrologic Engineering*, vol. 17, no. 3, pp. 368–376, 2012.

[24] H. Cho, D. Kim, F. Olivera, and S. D. Guikema, "Enhanced speciation in particle swarm optimization for multi-modal problems," *European Journal of Operational Research*, vol. 213, no. 1, pp. 15–23, 2011.

[25] SCS, *Soil Conservation Service (SCS) National Engineering Handbook, Section 4: Hydrology*, Department of Agriculture, Washington, Wash, USA, 1972.

[26] J. R. M. Hosking, "L-moments: analysis and estimation of distributions using linear combinations of order statistics," *Journal of the Royal Statistical Society B: Methodological*, vol. 52, no. 1, pp. 105–124, 1990.