



A Hybrid Stochastic Rainfall Model That Reproduces Rainfall Characteristics at Hourly through Yearly Time Scale

Jeongha Park¹, Christian Onof², and Dongkyun Kim¹

¹Department of Civil Engineering, Hongik University, Seoul, 04066, Republic of Korea

²Department of Civil and Environmental Engineering, Imperial College, London, SW7 2AZ, UK

Correspondence to: Dongkyun Kim (kim.dongkyun@hongik.ac.kr)

An Subsequent

Abstract. A novel approach ^{to} of stochastic rainfall generation that can reproduce various statistical characteristics of observed rainfall at hourly through yearly time scale is presented. The model uses the Seasonal Auto-Regressive Integrated Moving Average (SARIMA) model to generate monthly rainfall. Then, it downscales the generated monthly rainfall to the hourly aggregation level using the Modified Bartlett-Lewis Rectangular Pulse (MBLRP) model, a type of Poisson cluster rainfall model. Here, the MBLRP model is fine-tuned such that it can reproduce the fine-scale properties of observed rainfall. This was achieved by first generating a set of fine scale rainfall statistics reflecting the complex correlation structure between rainfall mean, variance, auto-covariance, and proportion of dry periods, and then coupling ^{these} it to the generated monthly rainfall, ^{which were used as the basis of the MBLRP parameters to downscale monthly rainfall.} The approach was tested ^{on 29 gauges} at the 29 gauges located in the Midwest to the East Coast of the Continental United States with a variety of rainfall characteristics. The results of the test suggest that our hybrid model accurately reproduces the first through the third order statistics as well as the intermittency properties from the hourly to the annual time scale; and the statistical behaviour of monthly maxima and extreme values of the observed rainfall ^{were} reproduced as well.

1 Introduction and Background

Most human and natural systems affected by rainfall react sensitively to temporal variability of rainfall across small (e.g. quarter-hourly) through large (e.g. monthly, yearly) time scale. Small scale rainfall temporal variability influences short-term watershed responses such as flash flood (Reed et al., 2007) and subsequent transport of sediments (Ogston et al., 2000) and contaminants (Zonta et al., 2005). Large scale rainfall temporal variability influences long-term resilience of human-flood systems (Yu et al., 2017), human health (Patz et al., 2005), food production (Shisanya et al., 2011), and evolution of human society (Warner and Afifi, 2014) and ecosystems (Borgogno et al., 2007; Fernandez-Illescas and Rodriguez-Iturbe, 2004).

While the risk exerted by these impacts needs to be precisely assessed for the management of ^{Such} the systems, the observed rainfall record is oftentimes not long enough (Koutsoyiannis and Onof, 2001). Furthermore, the rainfall records do not exist when the risks need to be assessed for the future. For this reason, stochastic rainfall generators, which can ^{create} generate

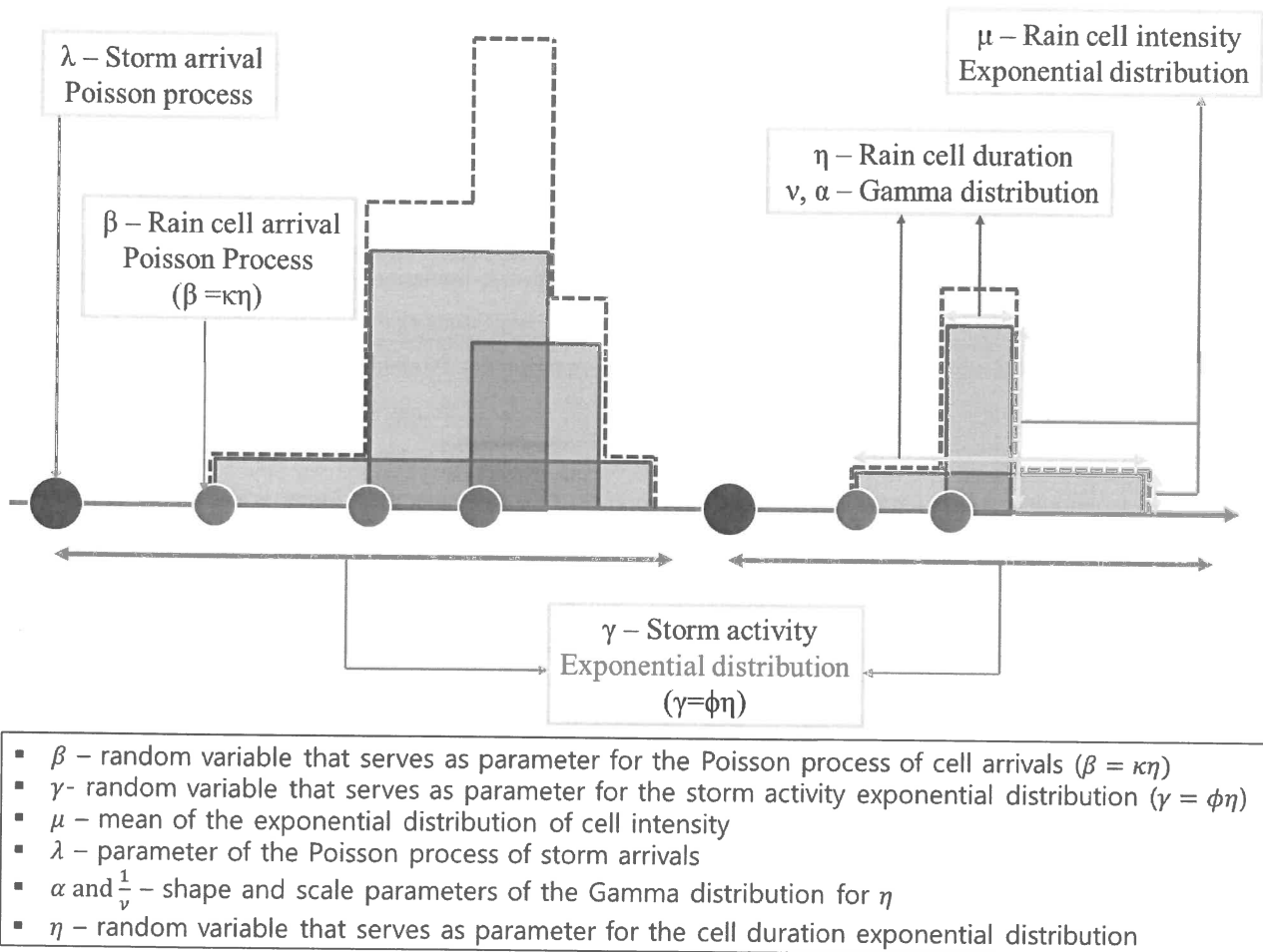


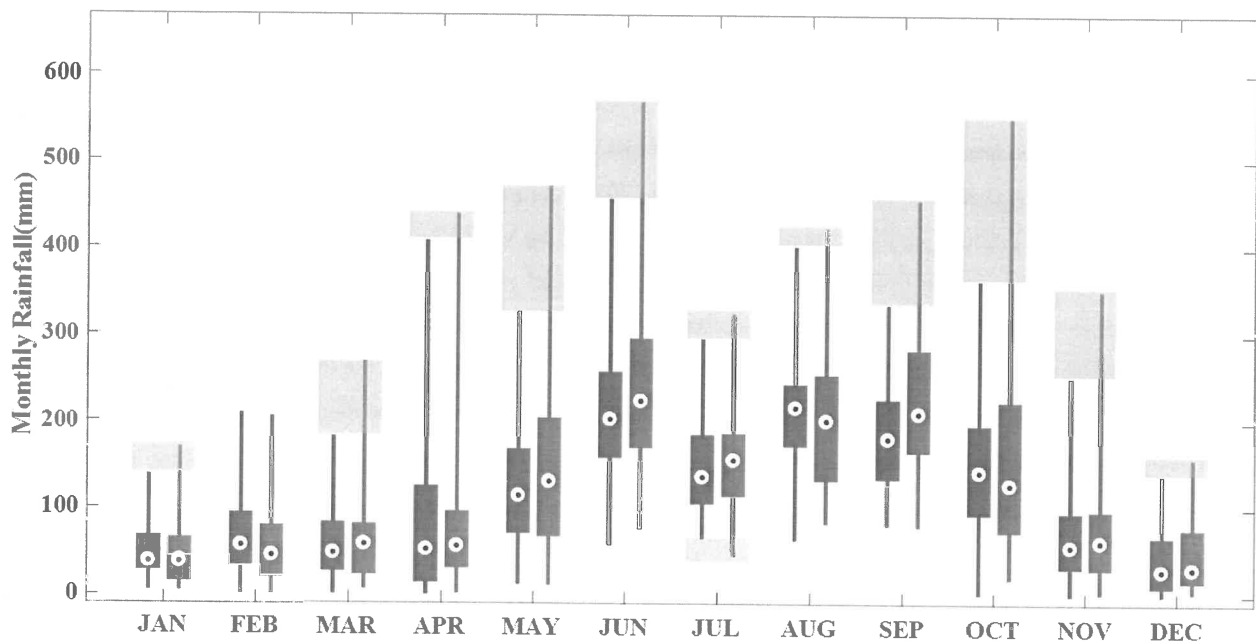
Figure 1: Schematic of the Modified Bartlett-Lewis Rectangular Pulse Model.

As suggested by the figure, Poisson cluster rainfall models are designed to reflect the original spatial structure of rain storms containing multiple rain cells (Austin and Houze Jr., 1972; Olsson and Burlando, 2002), so they are good at reproducing the first through the third order statistics of the observed rainfall at quarter-hourly through daily accumulation levels, as well as other hydrologically important statistics such as proportion of non-rainy period (Olsson and Burlando, 2002). The performance of the Poisson cluster rainfall models in reproducing the statistical properties of observed rainfall has been validated for various climates at numerous locations across the globe (Bo et al., 1994; Cameron et al., 2000; Cowpertwait, 1991; Cowpertwait et al., 2007; Derzekos et al., 2005; Entekhabi et al., 1989; Glasbey et al., 1995; Gyasi-Agyei and Willgoose; 1997, Gyasi-Agyei; 1999, Islam et al.; 1990, Kaczmarska et al., 2014; Khaliq and Cunnane, 1996; Kim et al., 2016; Kim et al., 2013b; Kim et al., 2014; Kossieris et al., 2015; Kossieris et al., 2016; Onof and Wheater, 1994a; Onof and Wheater, 1994b;



greater than that of the blue box plots in general, which implies that the variability of the observed rainfall is systematically greater than that of the synthetic rainfall. The discrepancy between the two are shown as the gray shading in the figure. In addition, the monthly extreme values shown as star marks are also underestimated by synthetic rainfall. This is, in particular, caused by the aforementioned limitations of the Poisson cluster rainfall models.

5 Considering that the management strategies of the water-prone human and natural systems may be governed by the few extreme rainfall values observed in the shaded domain of Figure 2, the risk analysis based on the rainfall data generated by Poisson cluster rainfall models may miss the system behaviour that is crucial for development of the management plans. As a matter of fact, other rainfall models have the similar issue: they cannot reproduce the temporal variability of observed rainfall across all time scales (Paschalis et al, 2014). For example, Markov chains, alternating renewal processes, and
 10 generalized linear models can reproduce the variability only at time scales coarser than one day. Models based on autoregressive properties of rainfall are typically good at reproducing the observed rainfall variability only for a limited range of scales, for instance from one month to a year or two (Mishra and Desai, 2005; Modarres and Ouarda, 2014; Yoo et al., 2016).



15 **Figure 2:** Box plots of the observed monthly rainfall at the NCDC Gauge 85663 in Florida, US. (red). The box plots of the synthetic monthly rainfall generated by the Modified Bartlett-Lewis Rectangular Pulse model at the same gauge are shown for reference (blue).

Several studies discussed the need to use composite rainfall models to resolve this scale problem of rainfall models. Koutsoyiannis (2001) used two seasonal autoregressive models with different temporal resolution to generate two different
 20 time series referring to the same hydrologic process. Then, they adjusted the fine scale time-series using their novel coupling

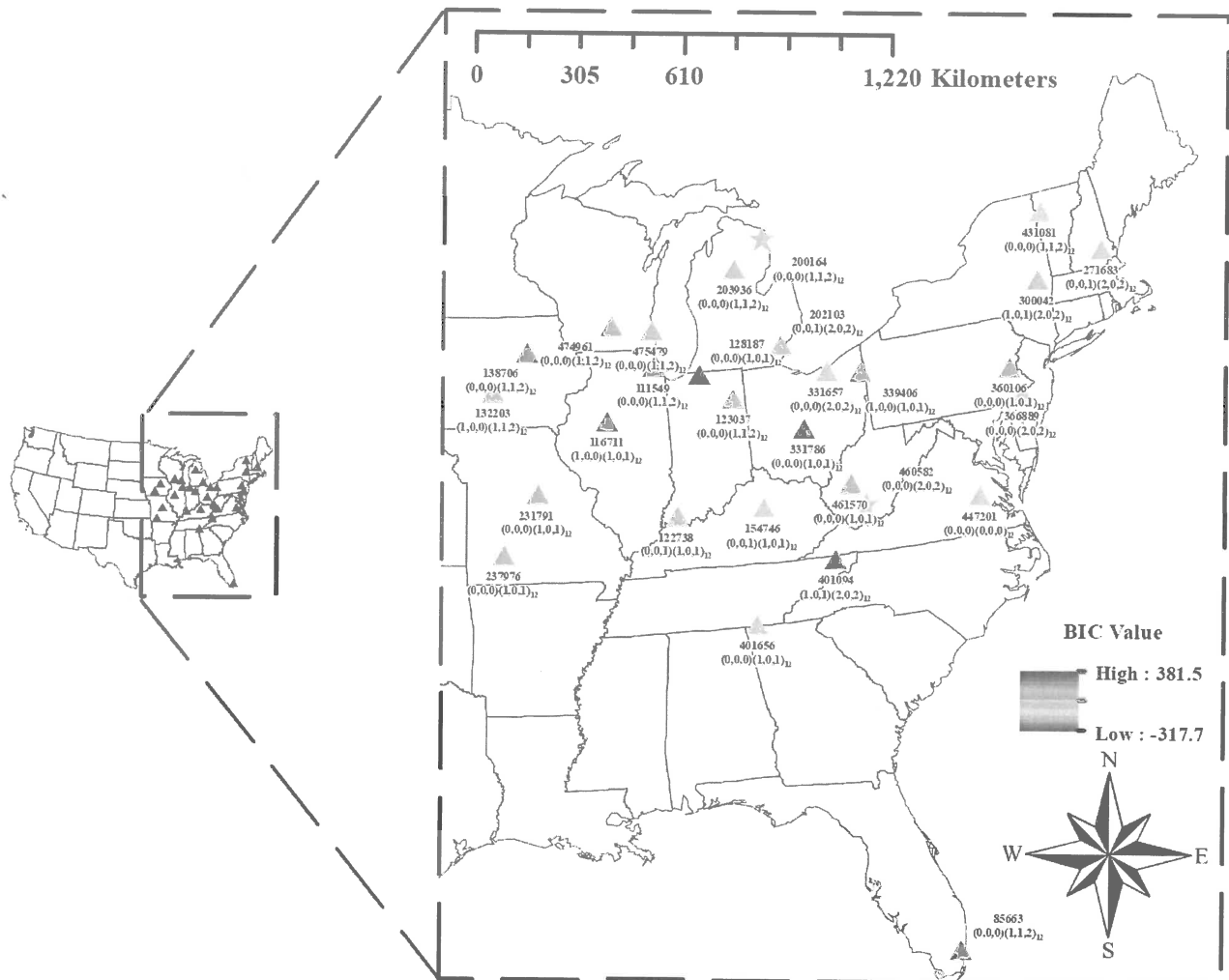


Figure 3: Study area and 29 NCDC hourly rainfall gauges.

3 Methodology

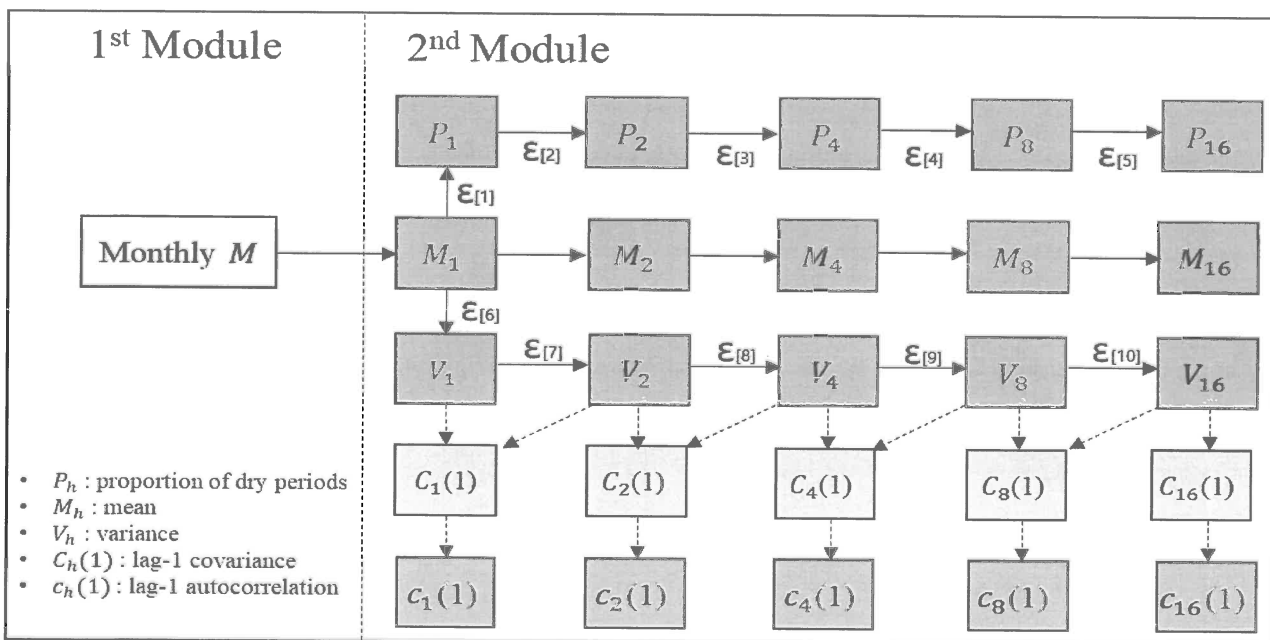
Figure 4 describes the model structure of this study. The model is composed of four distinct modules. The first module generates the monthly rainfall. The second module generates the fine-scale (1 hour through 16 hours) rainfall statistics corresponding to each of the generated monthly rainfall values in the first module. The third module estimates the parameters of the MBLRP model based on the fine-scale rainfall statistics generated by the second module. As a result of this process, each of the generated monthly rainfalls is coupled with the MBLRP parameter set reflecting its fine-scale statistical characteristics. The fourth module downscales each of the monthly rainfalls using the MBLRP model based on the parameters obtained in the third module.



(BIC) are calculated for the fitted SARIMA model. Lastly, the first to third steps are repeated for a combination of different values of p ($0 \leq p \leq 2$), d ($0 \leq d \leq 2$), q ($0 \leq q \leq 2$), P ($0 \leq P \leq 2$), D ($0 \leq D \leq 2$), and Q ($0 \leq Q \leq 2$), and the model structure with the lowest BIC is selected for the station. Therefore, a total of 729 ($=3^6$) SARIMA model structures were tested to obtain the optimal model for a station. The optimal model structure and the BIC values were shown in Figure 3. Through this process, we generated 200 years of monthly rainfall for the 29 gauges.

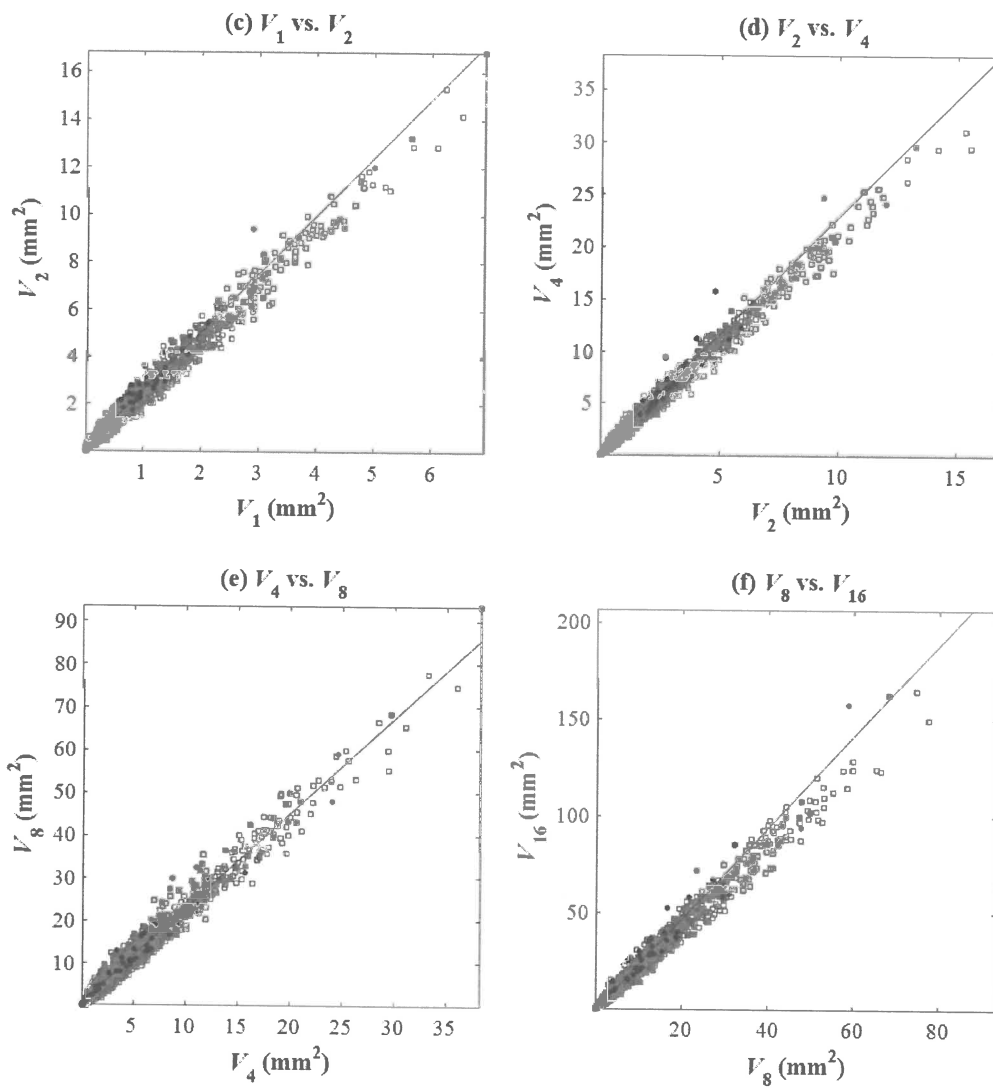
3.2 Generation of fine time scale rainfall statistics

The second module generates the fine time scale (1 hour through 16 hours) statistics corresponding to each monthly rainfall value generated through the SARIMA model. In so doing we are now considering the monthly rainfall, when divided by the number of days in the month times 24, as providing us with an estimate of the mean rainfall for that particular month. The second module consists of univariate regressions and functional relations linking the mean hourly rainfall to the other statistics that are required to fit the MBLRP model. With these statistics MLBRP model parameters are obtained and these will be used to disaggregate the generated monthly rainfall. Figure 5 describes the process of the fine time scale rainfall generation.



15 **Figure 5: Schematic of the algorithm to generate fine time-scale rainfall statistics.**

In the figure, M_h , V_h , $c_h(1) = C_h(1)/V_h$ and P_h in each rectangle represent the rainfall mean, variance, lag-1 autocorrelation, and proportion of dry periods at time-scale h hours, respectively. The statistic connected to each solid arrow head is stochastically generated based on its linear relationship to the one connected to the tail of the same arrow. The statistic





The autocorrelation lag- k is $c_h(k) = C_h(k)/V_h$, so, for $k = 1$ and $h = 1$ hour, we obtain the relation:

$$c(1) = \frac{V_2}{2V_1} - 1 \quad (4)$$

If we estimate the lag-one autocorrelation using standard estimators of the terms in the right-hand side of this relation, i.e. by using $\frac{\hat{V}_2}{2\hat{V}_1} - 1$, how good is the estimation likely to be? The figure below compares this estimator with the standard estimator

5 $\widehat{c(1)}$ of the autocorrelation.

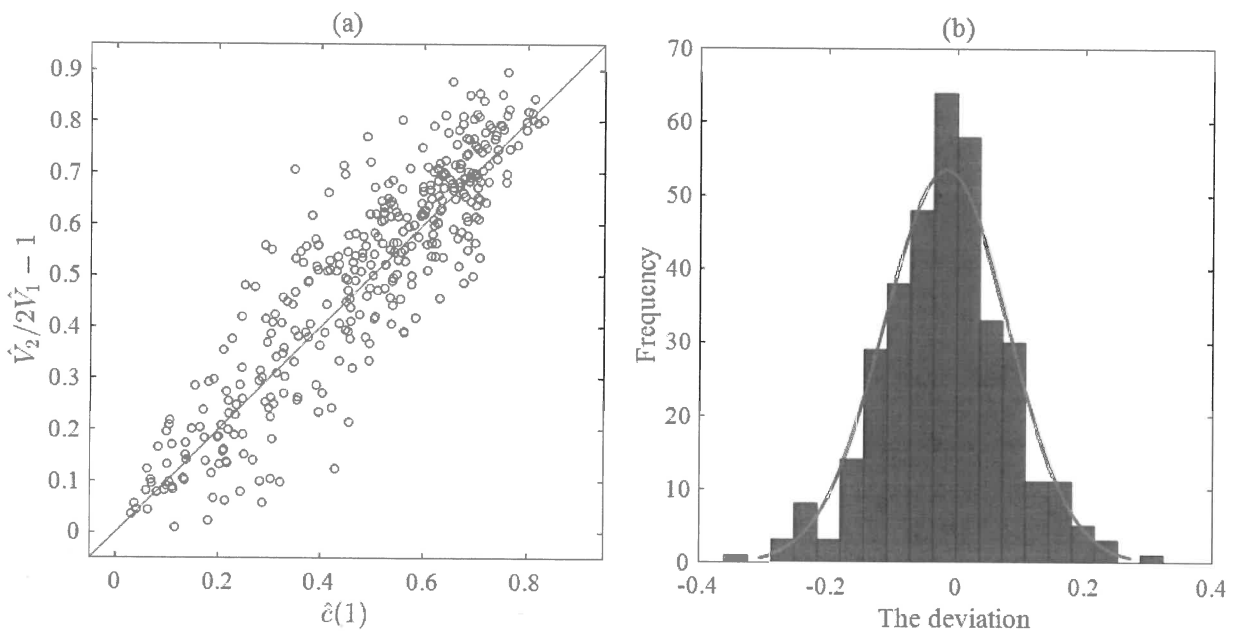


Figure 7: (a) Comparison of estimator $\widehat{c(1)}$ (horizontal axis) with estimator $\frac{\hat{V}_2}{2\hat{V}_1} - 1$ (vertical axis) of the autocorrelation lag-1 of hourly rainfall, (b) The histogram of the discrepancies between these two estimators.

Using the discrepancies ε between these two estimators which are approximately normally distributed as shown in Figure 7(b),

10 i.e. $\varepsilon \sim N(0, \sigma^2)$ we therefore estimate the autocorrelation lag-1 of hourly rainfalls using $\frac{\hat{V}_2}{2\hat{V}_1} - 1 + \varepsilon$.

Looking now at the solid arrows in Figure 5, we see that the residual terms (denoted $\varepsilon_{[i]}$) are likely to be correlated.

For example, consider the following equations relating V_1 to V_2 and V_2 to V_4 :

$$V_2 = a_{[7]}V_1 + \varepsilon_{[7]} \quad (5)$$

$$V_4 = a_{[8]}V_2 + \varepsilon_{[8]} \quad (6)$$

15 From equation (4), it is clear that term $\varepsilon_{[7]}$ is dependent upon the hourly autocorrelation (lag-1) coefficient, and similarly therefore that $\varepsilon_{[8]}$ in equation (6) is dependent upon the two-hourly (lag-1) autocorrelation coefficient.



3.3 MBLRP Model Parameter Estimation

In this process, each of the monthly rainfall values generated by the SARIMA model is coupled with one set of six MBLRP model parameters that define the random nature of rain storm and rain cell arrival frequency, and the intensity and duration of rain cells (Figure 1).

5 In this study, the parameters of the MBLRP model were determined such that the rainfall statistics of the generated rainfall resemble the 20 fine-scale rainfall statistics that were coupled with the monthly rainfall generated by the SARIMA model. The Isolated-Speciation Particle Swarm Optimization (ISPSO, Cho et al., 2011) algorithm was employed to identify a set of parameters that minimizes the following objective function:

$$OF = \sum_{i=1}^{20} w_i \cdot \left[1 - \frac{F_i(\lambda, \nu, \alpha, \mu, \phi, \kappa)}{f_i} \right]^2 \quad (8)$$

10 F_i is the i^{th} statistic of the synthetic rainfall time series (e.g. mean of hourly rainfall, standard deviation of 4-hourly rainfall, etc.). The mathematical formulae for the F_i s were derived by Rodriguez-Iturbe et al. (1988) as a function of the six parameters ($\lambda, \nu, \alpha, \mu, \phi, \kappa$); f_i is the i^{th} generated statistic, and w_i the weighting factor given to the i^{th} rainfall statistic depending on the use of the synthetic rainfall time series (Kim and Olivera, 2011). Here, it should be noted that a time step with rainfall less than 0.5mm was considered dry when the proportion of non-rainy period was calculated because small rainfall values are known to distort the “true” proportion of non-rainy period exerting an adverse effect on calibration process (Kim et al, 2016, Cross et al., 2018).

15 It is noteworthy that Module 2 may fail to generate a realistic set of fine scale rainfall statistics due to the complex interdependencies between them. The unrealistic fine scale rainfall statistics cannot be represented by the MBLRP model that reflects the original spatial structure of rainfall in reality, which entails poorly calibrated model parameters with high objective function value of Equation 8. To exclude the poorly calibrated parameter sets caused by the unrealistic fine scale rainfall statistics generated by Module 2, we repeated the process of Module 2 and Module 3 until the objective function value of Equation 8 becomes lower than a given threshold value (0.8 in this study). If the algorithm fails to find the parameter set after 50 repetitions, the parameter set with the lowest objective function value is chosen. Figure 4 describes this filtering process, and the red squares in Figure 6 shows the chosen parameter sets.

25

3.4 Downscaling of Monthly Rainfall Using the MBLRP Model

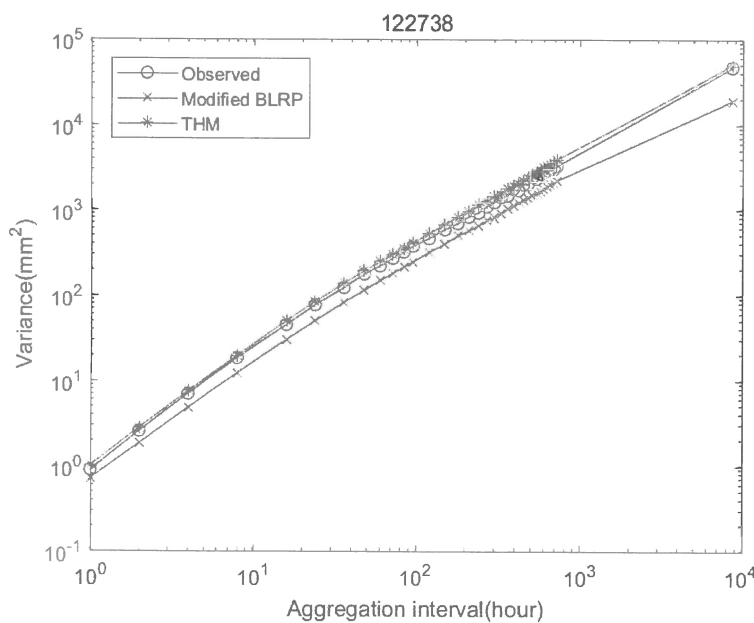
The MBLRP model was used to downscale the monthly rainfall to the hourly aggregation level. First, the MBLRP model generates the hourly rainfall time series using the parameter set for the monthly rainfall being downscaled. Second, the discrepancy between the generated fine time scale statistics and the statistics of the generated synthetic hourly rainfall time series is calculated using the following formula:

$$D^j = \sum_{i=1}^{20} \left[\frac{S_i^j - f_i}{R_i} \right]^2 \quad (9)$$



4.2 Reproduction of Large Scale Rainfall Variability

Figure 10 shows the behaviour the rainfall variance varying with temporal aggregation interval between 1 hour and 1 year at the NCDC gauge 122738. The behaviour corresponding to the observed rainfall (black) and the 200 years of synthetic rainfall generated by the MBLRP model (blue) and by our hybrid model (red) are shown together. While our model successfully reproduces the rainfall variance across the time scale, the MBLRP model is successful in reproducing the rainfall variance only at the hourly accumulation level. This reflects the fact that Poisson cluster rainfall models are not designed to preserve the rainfall persistence at the aggregation interval that is greater than the typical model storm duration, i.e. a few hours. See Figure 1 for example. Within the duration of one storm, rainfall at different time steps may be similar insofar as a portion of it is from the same rain cell. However, the rainfall within one storm is independent of the rainfall within another storm. Therefore, it is natural that Poisson cluster rainfall models tend to underestimate the observed rainfall variance (which reflects the covariance structure - see Equation 1) at time scales exceeding the rain storm duration. Kim et al. (2013b), when mapping the average model storm duration across the continental United States using Equation 11, showed that the model storm duration of the MBLRP model approximately ranges from 2 to 100 hours, so it is not only at the annual scale, but already at the scale of several hours (depending upon the location) that the variability may be underestimated by the MBLRP model.



15 **Figure 10: Behaviour of the rainfall variance with regard to aggregation interval of the observed rainfall time series (black) and the 200 years of synthetic rainfall generated by the MBLRP model (blue) and our hybrid model (red).**

A similar trend as exhibited in Figure 10 was observed at all of the 29 gauges. Figure 11(a) compares the variance of the observed (x) and synthetic (y) rainfall time series at yearly (purple), monthly (red), 15-daily (yellow), weekly (blue), and 32-hourly (green) aggregation levels. The comparison of the variance at the finer time scale is carried out in the following

20

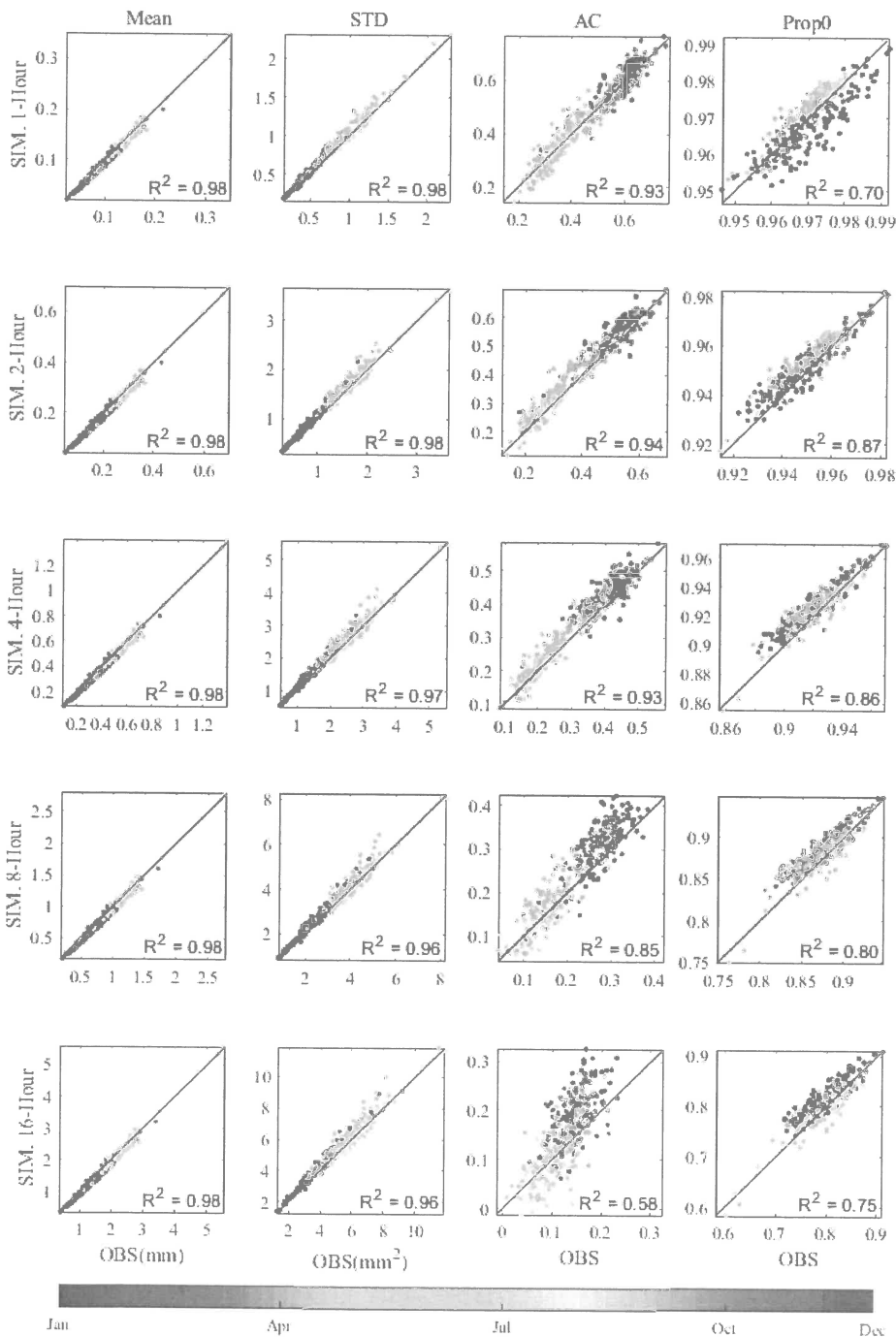
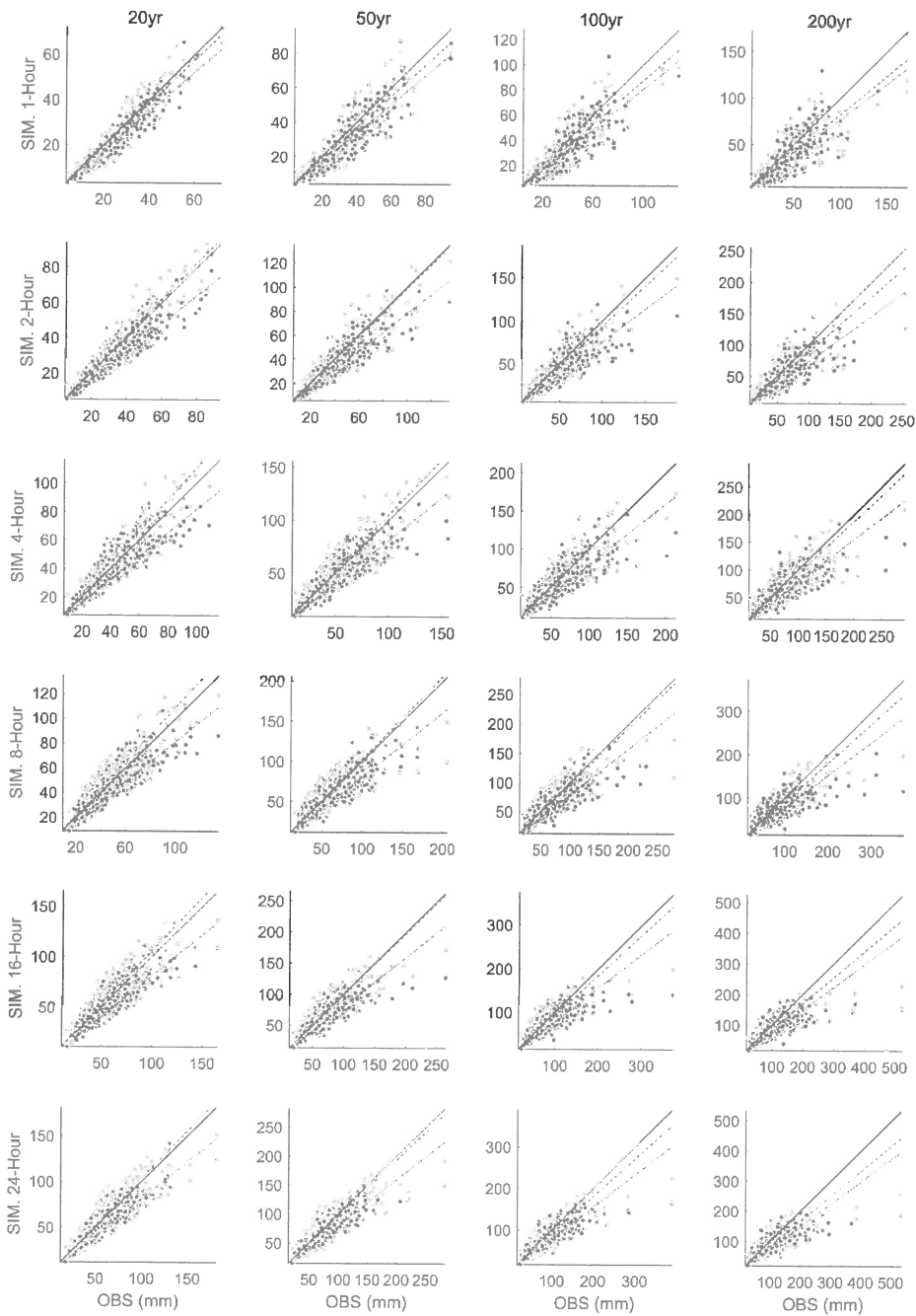


Figure 12: Comparison of the statistics of the observed (x) and synthetic (y) rainfall time series at sub-daily time scale. The colour of the dots represents the statistics of each calendar month.



5 **Figure 13: Comparison of the extreme rainfall values estimated from the observed (x) and synthetic rainfall generated by the model of this study (y). The plots comparing (a) 20-, (b) 50-, (c) 100-, and (d) 200-year rainfall are shown. Different colours of the scatter represent the rainfall duration. A circle represents the result according the hybrid model, and a triangle represents the result according to the traditional MBLRP model**



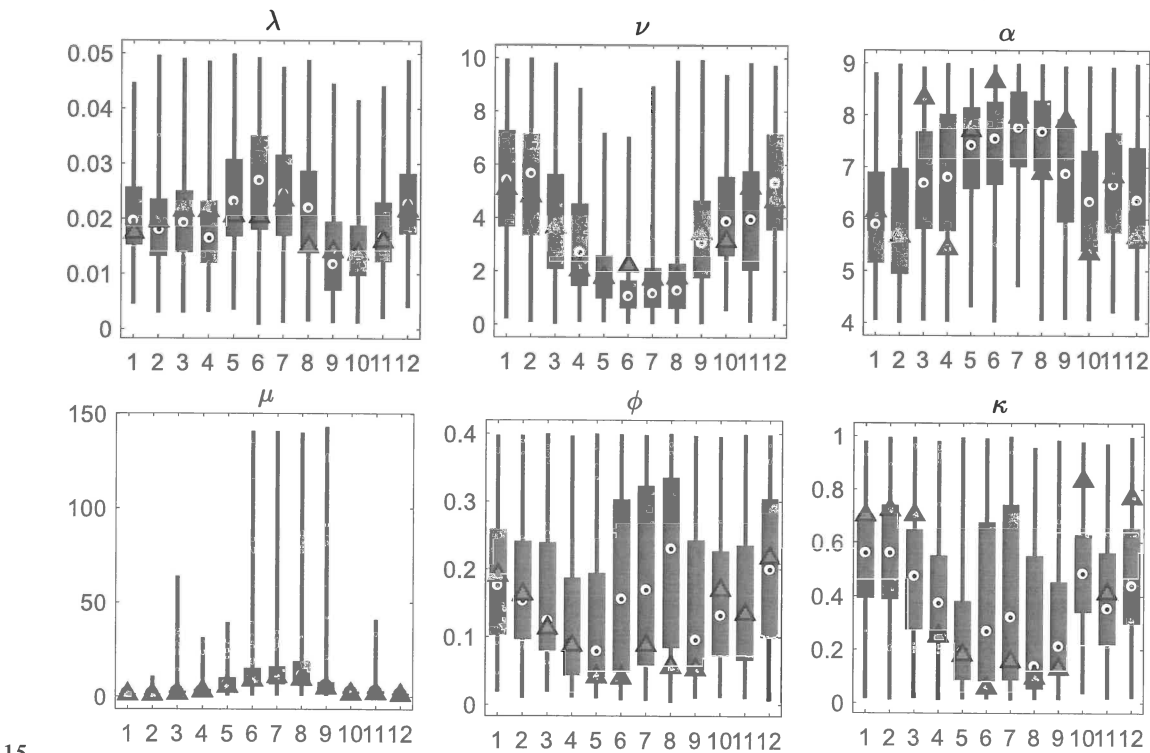
5 Discussion

5.1 Variability of the Parameters of the MBLRP model and Extreme Values

Our model uses different parameter sets of the MBLRP model to disaggregate different monthly rainfalls. This means that one given calendar month can have many different parameter sets. By contrast, the traditional MBLRP model uses one parameter set for each calendar month. Therefore, if we look at the variability of each month's parameters, we can see how the model of this study explains the variability of rainfall unlike the MBLRP model. Figure 15 shows a box plot of the parameters for each month at the NCDC Gauge 460582. The parameters of the traditional MBLRP model are shown together for reference (triangles).

While significant variability is observed for all six parameters, the parameter μ , which represents the average rain cell intensity, showed the greatest variability, ranging over two orders of magnitudes. This explains why our model is good at both reproducing large scale rainfall variability and small scale extreme values: the variability of the rain cell intensity parameter has the effect of stretching out the distribution of rainfall depths at a range of levels of aggregation, thereby increasing the probability of very large values. And it is course the variability of this cell intensity parameter that is also the most important factor responsible for the increase in the large scale rainfall variance.

Not clear



15 **Figure 15: Variability of the six parameters of the MBLRP model of this study (box plot) at the NCDC Gauge 460582 (star mark in Figure 3). The parameters of the traditional MBLRP model are shown together for reference (triangle).**

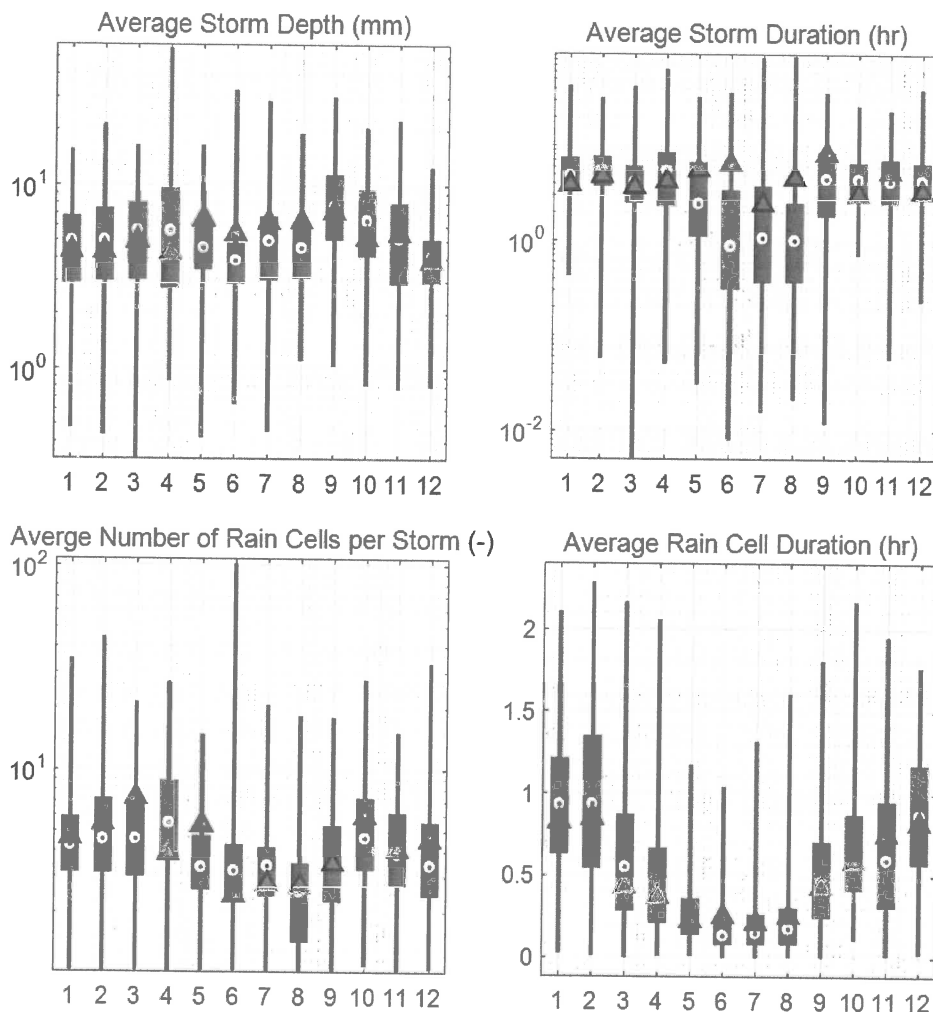


Figure 16: Variability of the rainfall characteristics of the MBLRP model of this study (box plot) at the NCDC Gauge 460582 (star mark in Figure 3). The rainfall characteristics of the traditional MBLRP model are shown together for reference (triangle).

Figure 16 shows box plots of the various rainfall characteristics for each month at the NCDC Gauge 460582. The values were calculated using Equations 10 through 14. The rainfall characteristics of the traditional MBLRP model are shown together for reference (triangles). The variability of the average storm depth, the average storm duration, and the average number of rain cells per storm was significant, so the y-axes of the box plots were drawn in log-scale. This result suggests that the parameter variability that is incorporated in our model's distinct algorithm contributes to the highly variable external (average storm depth, average storm duration) and internal (average number of rain cells per storm, average rain cell duration) properties of the generated rainfall.



- Borgogno, F., D'Odorico, P., Laio, F. and Ridolfi, L.: Effect of rainfall interannual variability on the stability and resilience of dryland plant ecosystems, *Water Resour. Res.*, 43, 2007.
- Burton, A., Fowler, H., Blenkinsop, S. and Kilsby, C.: Downscaling transient climate change using a Neyman–Scott Rectangular Pulses stochastic rainfall model, *J. Hydrol.*, 381, 18-32, 2010.
- 5 Cameron, D., Beven, K. and Naden, P.: Flood frequency estimation by continuous simulation under climate change (with uncertainty), *Hydrol. Earth Syst. Sci. Discuss.*, 4, 393-405, 2000.
- Cho, H., Kim, D., Olivera, F., and Guikema, S. D.: Enhanced speciation in particle swarm optimization for multi-modal problems, *Eur. J. Oper. Res.*, 213(1), 15-23, 2011.
- Cowpertwait, P. S.: A Poisson-cluster model of rainfall: some high-order moments and extreme values, *P. Roy. Soc. A-Math. Phys.*, 1998.
- 10 Cowpertwait, P. S.: Further developments of the Neyman-Scott clustered point process for modeling rainfall, *Water Resour. Res.*, 27, 1431-1438, 1991.
- Cowpertwait, P., Isham, V. and Onof, C.: Point process models of rainfall: developments for fine-scale structure, *P. Roy. Soc. A-Math. Phys.*, 2007.
- 15 Cross, D., Onof, C., Winter, H. and Bernardara, P.: Censored rainfall modelling for estimation of fine-scale extremes, *Hydrol. Earth Syst. Sc.*, 22, 727, 2018.
- Delleur, J. W. and Kavvas, M. L.: Stochastic models for monthly rainfall forecasting and synthetic generation, *J. Appl. Meteorol.*, 17, 1528-1536, 1978.
- Derzekos, C., Koutsoyiannis, D. and Onof, C.: A new randomised Poisson cluster model for rainfall in time, *Enrgy. Proced.*, 2005.
- 20 Entekhabi, D., Rodriguez-Iturbe, I. and Eagleson, P. S.: Probabilistic representation of the temporal rainfall process by a modified Neyman-Scott Rectangular Pulses Model: Parameter estimation and validation, *Water Resour. Res.*, 25, 295-302, 1989.
- Faramarzi, M., Abbaspour, K. C., Schulin, R. and Yang, H.: Modelling blue and green water resources availability in Iran, *Hydrol. Process.*, 23, 486-501, 2009.
- Fatichi, S., Ivanov, V. Y. and Caporali, E.: Simulation of future climate scenarios with a weather generator, *Adv. Water Resour.*, 34, 448-467, 2011.
- Fernandez-Illescas, C. P. and Rodriguez-Iturbe, I.: The impact of interannual rainfall variability on the spatial and temporal patterns of vegetation in a water-limited ecosystem, *Adv. Water Resour.*, 27, 83-95, 2004.
- 30 Furrer, E. M. and Katz, R. W.: Improving the simulation of extreme precipitation events by stochastic weather generators, *Water Resour. Res.*, 44, 2008.
- Glasbey, C., Cooper, G. and McGechan, M.: Disaggregation of daily rainfall by conditional simulation from a point-process model, *J. Hydrol.*, 165, 1-9, 1995.



- Kossieris, P., Makropoulos, C., Onof, C. and Koutsoyiannis, D.: A rainfall disaggregation scheme for sub-hourly time scales: Coupling a Bartlett-Lewis based model with adjusting procedures, *J. Hydrol.*, 2016.
- Kottek, M., Grieser, J., Beck, C., Rudolf, B. and Rubel, F.: World map of the Köppen-Geiger climate classification updated, *Meteorol. Z.*, 15, 259-263, 2006.
- 5 Koutsoyiannis, D.: Coupling stochastic models of different timescales, *Water Resour. Res.*, 37, 379-391, 2001.
- Koutsoyiannis, D. and Onof, C.: Rainfall disaggregation using adjusting procedures on a Poisson cluster model, *J. Hydrol.*, 246, 109-122, 2001.
- Marani, M.: On the correlation structure of continuous and discrete point rainfall, *Water Resour. Res.*, 39, 2003.
- Menabde, M. and Sivapalan, M.: Modeling of rainfall time series and extremes using bounded random cascades and levy-stable distributions, *Water Resour. Res.*, 36, 3293-3300, 2000.
- 10 Mishra, A. and Desai, V.: Drought forecasting using stochastic models, *Stoch. Env. Res. Risk A.*, 19, 326-339, 2005.
- Modarres, R. and Ouarda, T. B.: Modeling the relationship between climate oscillations and drought by a multivariate GARCH model, *Water Resour. Res.*, 50, 601-618, 2014.
- Molnar, P. and Burlando, P.: Preservation of rainfall properties in stochastic disaggregation by a simple random cascade model, 15 *Atmos. Res.*, 77, 137-151, 2005.
- Müller, H. and Haberlandt, U.: Temporal rainfall disaggregation using a multiplicative cascade model for spatial application in urban hydrology, *J. Hydrol.*, 2016.
- Ogston, A., Cacchione, D., Sternberg, R. and Kineke, G.: Observations of storm and river flood-driven sediment transport on the northern California continental shelf, *Cont. Shelf Res.*, 20, 2141-2162, 2000.
- 20 Olsson, J. and Burlando, P.: Reproduction of temporal scaling by a rectangular pulses rainfall model, *Hydrol. Process.*, 16, 611-630, 2002.
- Onof, C., Meca-Figueras, T., Kaczmarska, J., Chandler, R. and Hege, L.: , Modelling rainfall with a Bartlett-Lewis process: thirdorder moments, proportion dry, and a truncated random parameter version, 2013.
- Onof, C. and Wheater, H. S.: Improved fitting of the Bartlett-Lewis Rectangular Pulse Model for hourly rainfall, *Hydrolog. Sci. J.*, 39, 663-680, 1994a.
- 25 Onof, C. and Wheater, H. S.: Improvements to the modelling of British rainfall using a modified random parameter Bartlett-Lewis rectangular pulse model, *J. Hydrol.*, 157, 177-195, 1994b.
- Onof, C. and Wheater, H. S.: Modelling of British rainfall using a random parameter Bartlett-Lewis rectangular pulse model, *J. Hydrol.*, 149, 67-95, 1993.
- 30 Paschalis, A., Molnar, P., Fatichi, S. and Burlando, P.: On temporal stochastic modeling of precipitation, nesting models across scales, *Adv. Water Resour.*, 63, 152-166, 2014.
- Patz, J. A., Campbell-Lendrum, D., Holloway, T. and Foley, J. A.: Impact of regional climate change on human health, *Nature*, 438, 310, 2005.



Yu, D. J., Sangwan, N., Sung, K., Chen, X. and Merwade, V.: Incorporating institutions and collective action into a sociohydrological model of flood resilience, *Water Resour. Res.*, 53, 1336-1353, 2017.

Zonta, R., Collavini, F., Zaggia, L., and Zuliani, A.: The effect of floods on the transport of suspended sediments and contaminants: a case study from the estuary of the Dese River (Venice Lagoon, Italy), *Environ. Int.*, 31(7), 948-958, 2005.