# A Spatial-temporal Rainfall Generator for Flood Response Analysis

Yejia Qiang[1,2]
[1] *HKUST Shenzhen Research Institute, Nanshan, Shenzhen 518057, China*
[2] *Dept. of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, China*

Limin Zhang[1,2]
[1] *HKUST Shenzhen Research Institute, Nanshan, Shenzhen 518057, China*
[2] *Dept. of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, China*

Te Xiao
*Dept. of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, China*

ABSTRACT: The Guangdong-Hong Kong-Macao Bay Area, located in the southeast of China, suffers typhoon-related storms and floods. This paper presents a spatial-temporal rainfall generation model for regional flood response analysis, with its parameters easily obtainable from historical point observations. The model generates point rainfall event series at different rainfall stations with variables describing the external structure and a predefined internal profile within an alternating renewal model framework. Spatial correlation of rainfall process between different sites within the study area is considered, and the areal rainfall distribution of each time slot is obtained from multi-point rainfall amounts. The model performs well in the reproduction of regional rainfall statistical characteristics.

## 1. INTRODUCTION

Due to global warming, extreme rainstorms will increase in terms of both frequency and intensity in some areas. The occurrence of floods is closely related to intense rainfall. The Guangdong-Hong Kong-Macao Bay Area located in the southeast of China consists of 11 cities including Hong Kong, Macao, Guangzhou, Shenzhen, etc. Figure 1 shows the range and elevation of the study area. It is one of China's most developed regions. However, the bay area frequently suffers damage from typhoon related storms, and was ranked as the first by Swiss Re (2013) in terms of population affected by storm and river flood.

The Guangdong-Hong Kong-Macao Bay Area has a humid subtropical monsoon climate characterized by clear rainy and dry seasons, and the average annual rainfall is over 1300mm/year.

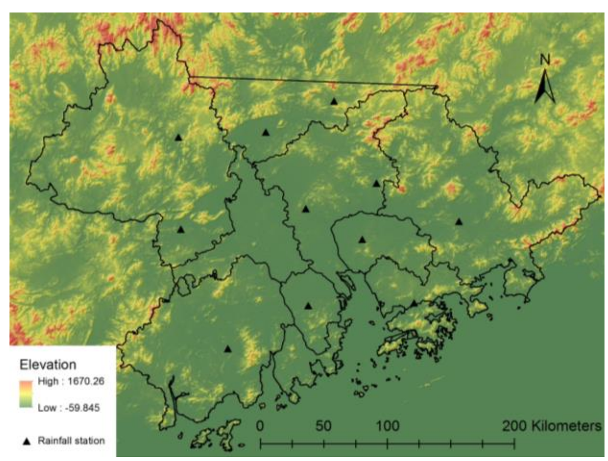


*Figure 1: Topography of the study area and rainfall stations*

Generally, the largest rainfall amounts occur in June, May and August. To study the risk related to rainfall-induced flood, spatially variable rainfall processes are needed as input for large-scale flood analysis. A stochastic rainfall generator may be a good alternative.

Stochastic rainfall simulation models can be used to identify continuous rainfall processes to provide input for flood routing modelling. Based on point rainfall observations, simple models for point synthetic rainfall sequence generation were developed, such as Markov renewal model (Haan et al., 1976; Foufoula-Georgiou and Lettenmaier, 1987) and alternating renewal model (Acreman, 1990; Haberlandt, 1998; Haberlandt et al., 2008). To capture the spatial and temporal characteristics of rainfall, multisite generation of time series is widely used when point observations are available. For example, Wilks (1998) established a chain dependent process model for simultaneous simulation of daily precipitation occurrences and amounts at multiple locations. Yang et al. (2005) made improvements in modelling spatial dependence for time series based on the spatial-temporal model using generalized linear models proposed by Chandler and Wheater (2002).

The objective of this paper is to set up a model for spatial-temporal rainfall synthesis for regional flood response analysis, with parameters easily obtainable from historical point observations.

## 2. METHODOLOGY

The proposed model consists of three components: single-site rainfall event series generation, multisite rainfall event series generation and spatial-temporal varying rainfall synthesis.

### *2.1. Temporal rainfall simulation for each single station*

An alternating renewal model developed by Haberlandt et al. (1998, 2008) is used to generate point rainfall time series of each station within the study area because the physical meaning of the model is easy to follow and the parameters can be obtained from historical observations.

A sequence of rainfall events can be approximated as an alternating renewal process which models a system alternating between two states over time, i.e. raining or not raining here. Time of raining or not raining are recognized as wet spell duration (*W*) and dry spell duration (*D*) respectively, and the basic assumption is that pairs of *W* and *D* form an independent, identically distributed sequence. Furthermore, *W* and *D* of the same pair are considered independent here. For each rainfall event (one wet spell), the average intensity (*I*) defined as cumulated rainfall amount divided by the wet spell duration is not independent of the wet spell duration. The relation between W and I is described using a copula here. The three variables characterizing the occurrence and rainfall amount of rainfall events are regarded as the external structure of rainfall time series (Figure 2). Besides, temporal variation of rainfall intensity within one event is regarded as the internal structure.

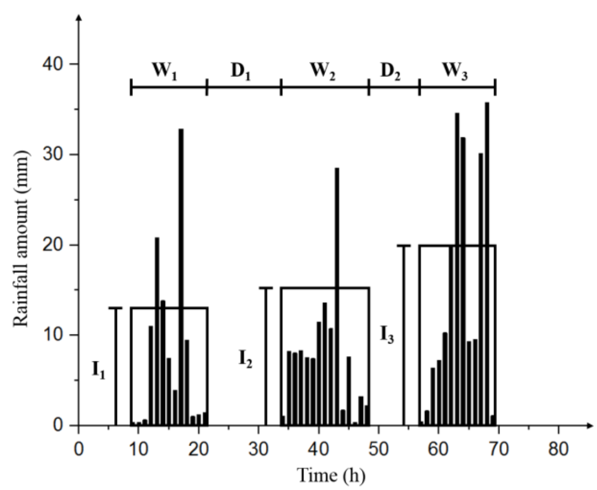


*Figure 2. Illustration of precipitation event process*

Different probability distributions of the three variables can be fitted to the rainfall dataset of the study area, which together constitute the external structure of the rainfall event. In the case of this paper, the Weibull distribution, the Exponential distribution and the Lognormal distribution are found to be suitable for modelling dry spell duration *D*, wet spell duration *W* and intensity *I*, respectively.

A two dimensional copula is used to describe the dependence between wet spell duration and rainfall intensity (Sklar, 1973; Nelsen, 2007):

$$F(w,i) = C[F_W(w), F_I(i)] = C(u_1, u_2) \qquad (1)$$

where $F(w,i)$ is the joint cumulative distribution function (CDF) of *W* and *I*; $F_W(w)$ and $F_I(i)$ are the marginal CDFs of *W* and *I*, respectively; C is the copula function that relates $F(w,i)$ and $F_W(w)$, $F_I(i)$; both $u_1$ and $u_2$ are uniformly distributed in the interval [0, 1]. According to previous research (Haberlandt et al., 2008), the Frank copula is used in this paper as the parameters for this model can be estimated easily.

The internal structure of the rainfall events is simplified to a predefined profile. It is assumed that the instantaneous rainfall intensity within a certain rainfall event first increases exponentially to the peak and then decreases exponentially:

$$i(t) = i_m \times e^{-\gamma|t-t_m|} \qquad (2)$$

where $i_m$ is the peak value during the event, $t_m$ is peak time, and $\gamma$ is an event specific parameter. The peak time is assumed to evenly occur during the event:

$$t_m = \alpha \times w \qquad (3)$$

where $\alpha$ is a random number uniformly distributed on [0, 1]. The peak rainfall intensity is assumed to be related to the average rainfall intensity, and parameters *a* and *b* are obtained through a simple regression:

$$i_m = a \times i^b, a \geq 1, b \geq 1 \qquad (4)$$

For each specific rainfall event, $\gamma$ can be estimated using the Newton-Raphson method after the determination of values of all the other variables:

$$\int_0^{t_m} i_m \times e^{\gamma(t-t_m)} dt + \int_{t_m}^{w} i_m \times e^{\gamma(t_m-t)} dt = wi \qquad (5)$$

With the external structure and the internal structure, single point rainfall event series can be generated by assembling sequences of *W*, *D* and *I* sampled from their corresponding probability distributions.

### *2.2. Generation of time series for multiple stations and areal rainfall distribution*

After obtaining the single point rainfall event series, the next step is to generate multi-point series within the study area considering the spatial dependence of rainfall processes between different points. A Gaussian spatial correlation structure is used, assuming the spatial dependences of the three variables are the same:

$$\rho_{pq} = \exp(-\frac{{\tau_{pq}}^2}{\delta^2}) \qquad (6)$$

where $\rho_{pq}$ is the correlation coefficient between variable values at two stations *p* and *q*; $\tau_{pq}$ is the separation distances between stations *p* and *q*; $\delta$ is a model parameter.

For a specific value of $\delta$, any number of rainfall events at multi-points can be generated. One rainfall event at $n_s$ stations is considered as an example. First, a $n_s \times n_s$ auto-correlation matrix $\boldsymbol{R} = [\rho_{pq}]$ is formed according to Eq. (6). Then, a $n_s \times 1$ correlated standard normal random vector $\boldsymbol{N}$ can be obtained by:

$$\boldsymbol{N} = \boldsymbol{LU} \qquad (7)$$

where $\boldsymbol{L}$ is a lower triangular matrix satisfying Cholesky decomposition $\boldsymbol{R} = \boldsymbol{LL}^{\mathrm{T}}$; $\boldsymbol{U}$ is a $n_s \times 1$ independent standard normal random vector. $\boldsymbol{N}$ can then be transformed into other types of random field according to variable distribution types (Low and Tang, 2007).

Note that *D* is independent of *W* and *I*, while *W* and *I* are cross-correlated, as mentioned in Section 2.1. For one rainfall event, the $n_s \times 1$ vectors of $\boldsymbol{D}$, $\boldsymbol{W}$ and $\boldsymbol{I}$ can be obtained as follows:

1. Randomly generate $n_s \times 3$ potential samples of a standard uniformly distributed variable at $n_s$ stations, with no dependence between all the samples.

2. Transform the first column of the potential samples into an independent standard normally distributed vector $\boldsymbol{U}_D$ firstly; then transform it into a standard normally distributed vector $\boldsymbol{N}_D$ with spatial correlation between stations according to Eq. (7). The $n_s$×1 vector of $\boldsymbol{D}$ for $n_s$ stations that follows Weibull distribution can be obtained by:

$$\boldsymbol{D} = \lambda\{-\ln[1-\Phi(\boldsymbol{N}_D)]\}^{1/k} \tag{8}$$

   where $k$ and $\lambda$ are the shape parameter and scale parameter for the Weibull distribution, respectively; $\Phi(\cdot)$ is the CDF of a standard normal distribution.
3. Transform the last two columns of the potential samples into a pair of standard uniformly distributed variables obeying a given copula function (e.g., Frank copula) with independence between stations.
4. Similar to Step 2, transform the two vectors firstly into two standard normally distributed vectors $\boldsymbol{U}_W$ and $\boldsymbol{U}_I$ with independence between stations, and then into two standard normally distributed vectors $\boldsymbol{N}_W$ and $\boldsymbol{N}_I$ with spatial correlation between stations using Eq. (7). The wet spell duration series $\boldsymbol{W}$ and average rainfall intensity series $\boldsymbol{I}$ that follow Exponential distribution and Lognormal distribution, respectively, can be obtained by:

$$\boldsymbol{W} = -\beta \ln[1-\Phi(\boldsymbol{N}_W)] \tag{9}$$

$$\boldsymbol{I} = \exp(\mu + \sigma \boldsymbol{N}_I) \tag{10}$$

   where $\beta$ is the scale parameter of exponential distribution; $\mu$ and $\sigma$ are the mean and standard deviation of logarithmized samples for the lognormal distribution, respectively.

Repeatedly performing the above steps, the matrices $\boldsymbol{D}$, $\boldsymbol{W}$ and $\boldsymbol{I}$ for $n_s$ stations with several rainfall events can be obtained eventually.

To evaluate the rationality of the generated rainfall process, the correlation coefficients of instantaneous rainfall intensity $\rho_I$ when simultaneous rainfall occurs at stations $p$ and $q$ are calculated, and the instantaneous rainfall intensity is approximated with rainfall depth within very short time slots (0.1 hours here):

$$\rho_{I,pq} = \frac{\mathrm{cov}(z_p, z_q)}{\sqrt{\mathrm{var}(z_p)\times \mathrm{var}(z_q)}} \tag{11}$$

where $z_p$ and $z_q$ are the cumulative rainfall depth in 0.1 hours. A correlation function between $\rho_I$ and separation distance $\tau$ of stations then can be obtained by regression:

$$\rho_I = \exp(-\frac{\tau}{\tau_0}) \tag{12}$$

where $\tau_0$ is a regression coefficient for the simultaneous rainfall intensity correlation function.

The value of parameter $\delta$ is estimated to find the optimal simulated results by minimizing the gap $\Delta\tau_0$ between repeated simulations and long-term observations:

After the point rainfall event series at multiple stations have been obtained, rainfall amounts within each time slot at several stations are known. The areal rainfall distribution in each time slot then can be obtained by interpolation using kriging (Oliver and Webster, 1990). To avoid negative values in interpolation, rainfall value is assumed to be lognormally distributed spatially. Repeated realizations of rainfall process generation can be carried out to obtain sufficient inputs for the probabilistic flood response analysis.

## 3. RAINFALL GENERATION CASE STUDY

The model is applied to the Guangdong-Hong Kong-Macau bay area. The distribution of 11 rainfall stations within this region is shown in Figure 1. Hourly rainfall data are recorded by these rainfall stations and daily rainfall data are accessible. At the same time, high time resolution rainfall data is obtained from dataset of NASA Tropical Rainfall Measuring Mission (TRMM) Rainfall Estimate (Huffman et al., 2007). The temporal and spatial resolution of rainfall data are 3 hour and 0.25º×0.25º, respectively. The observation period is 4 months (May to August) of each year from 1998 to 2017.

### *3.1. Determination of model parameters*

The parameters of probability distribution of *W*, *D* and *I*, as well as the copula parameter are estimated from historical rainfall data at the locations of the 11 stations using maximum likelihood estimation (DeGroot and Baecher, 1993; Fenton, 1999a, b) (Table 1), and parameter differences between stations are ignored. To determine the value of $\delta$, simultaneous rainfall intensity correlation function obtained from simulated results with a large scale of $\delta$ are compared with that from historical rainfall data, and a value of 350 is chosen to minimize $\Delta\tau_0$ (Figure 3). The simultaneous rainfall intensity correlation functions obtained from observation and simulation are shown in Figure 4. It is shown that the simulated results capture well the characteristics of the spatial correlation of simultaneous rainfall intensity.

*Table 1: Parameters used in the rainfall generator*

| *Variable* | *Description* | *Parameter* | *Value* |
|---|---|---|---|
| *D* | $F_D(d) = 1 - e^{-(d/\lambda)^k}$ | *k* | *0.836* |
| | | $\lambda$ | *28.786* |
| *W* | $F_W(w) = 1 - e^{-\frac{w}{\beta}}$ | $\beta$ | *5.507* |
| *I* | $F_I(i) = \frac{1}{2} + \frac{1}{2} erf[\frac{\ln i - \mu}{\sqrt{2}\sigma}]$ | $\mu$ | *0.405* |
| | | $\sigma$ | *0.920* |
| *C* | $C(u_1, u_2; \theta) = -\frac{1}{\theta}\ln[1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1}]$ | $\Theta$ | *2.205* |
| $i_m$ | $i_m = a \times i^b$ | *a* | *1.224* |
| | | *b* | *1.119* |

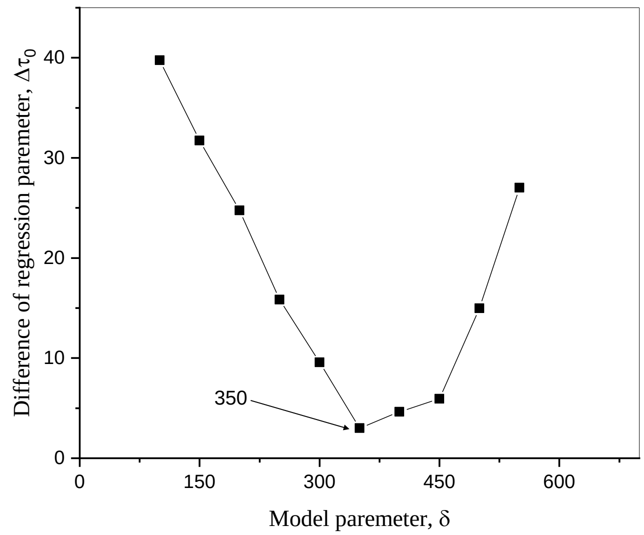


*Figure 3: Identification of parameter $\delta$*

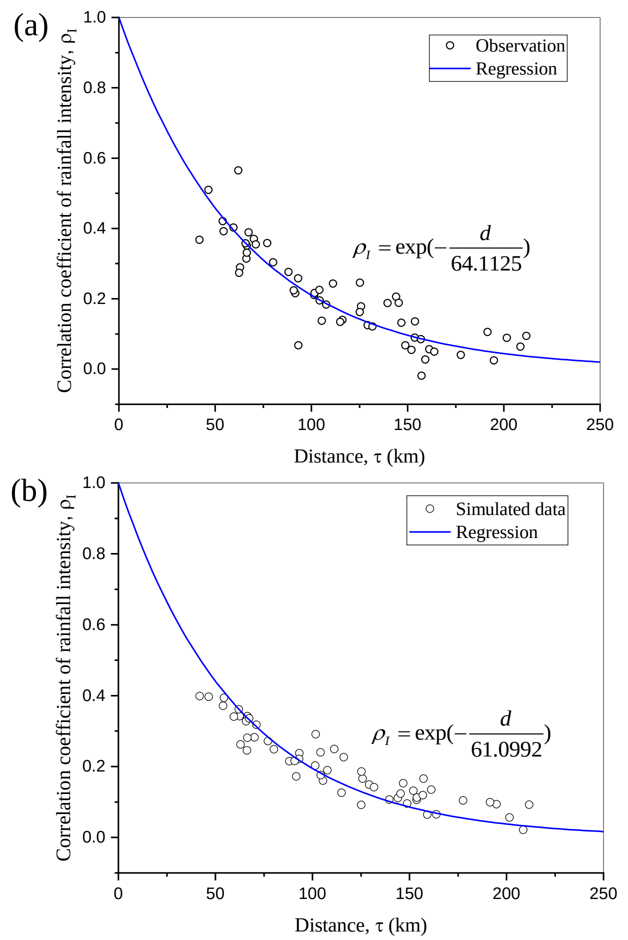


*Figure 4: Correlation functions of simultaneous rainfall intensity: (a) from observation data; (b) from simulated data)*

### 3.2. Results

The rainfall process of any duration can be generated within the study area. A realization of 4-month rainfall process at one station, as well as the spatial rainfall distribution at one time slot, are shown in Figure 5 as examples.

Ten realizations of 20-year rainfall generation (4-month for each year) are carried out and the characteristics of generated rainfall are compared with the observations. The results are shown in Table 2. It can be seen that the model captures the characteristics of the observed rainfall well. The simulated number of events and the total 4-month rainfall amount are slightly smaller than the observations, and the standard deviation of the simulated single event rainfall amounts is larger than the observed one. Moreover, 3-hour maximum rainfall and daily maximum rainfall are calculated.

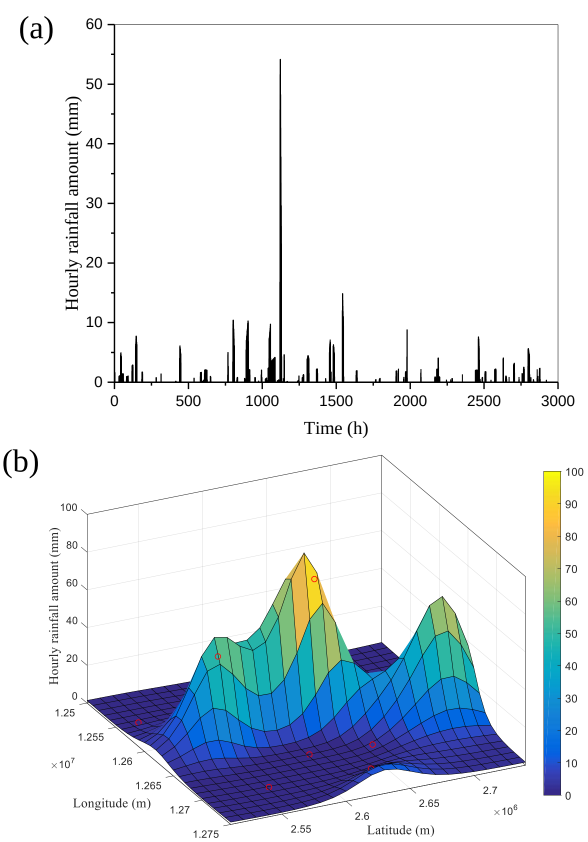


*Figure 5: Example of a generated 4-month rainfall process: (a) hourly rainfall time series at one station; (b) areal rainfall distribution at peak hour of the above station*

*Table 2: Comparison of characteristics of generated rainfall time series and observations*

| | *Observation* | *Simulation* | |
|---|---|---|---|
| | | *Range* | *Average* |
| *Average number of events* | *78.0* | *73.0-78.0* | *75.6* |
| *Average single event rainfall amount (mm)* | *15.8* | *14.6-16.7* | *15.9* |
| *Standard deviation of single event rainfall amount (mm)* | *23.9* | *26.2-32.8* | *29.5* |
| *Total rainfall amount within 4 months (mm)* | *1230.5* | *1087.4-1264.3* | *1196.8* |

## 4. CONCLUSIONS

A model for spatial-temporal rainfall synthesis reflecting regional rainfall characteristics is proposed for flood analysis. The model generates point rainfall time series based on an alternating renewal model, and considers spatial correlation of the multi-point rainfall time series and the areal rainfall distribution. Statistical characteristics of the regional rainfall process are captured. However, more extreme values exist in the generated rainfall than in the observations. So the model needs further improvement in terms of extreme value control.

## ACKNOWLEDGEMENTS

This research was supported by Science and Technology Plan of Shenzhen (20180428170335970) and the Research Grants Council of the Hong Kong SAR Government (Nos. C6012-15G and 16206217).

## REFERENCES

Acreman, M. C. (1990). “A simple stochastic model of hourly rainfall for Farnborough, England.” *Hydrological Sciences Journal*, 35(2), 119-148.

Chandler, R. E., and Wheater, H. S. (2002). “Analysis of rainfall variability using generalized linear models: a case study from the west of Ireland.” *Water Resources Research*, 38(10), 10-1-10-11.

DeGroot, D. J., and Baecher, G. B. (1993). “Estimating autocovariance of in-situ soil properties." *Journal of Geotechnical Engineering*, 119(1), 147-166.

Fenton, G. A. (1999a). "Estimation for stochastic soil models." *Journal of Geotechnical and Geoenvironmental Engineering*, 125(6), 470-485.

Fenton, G. A. (1999b). "Random field modeling of CPT data." *Journal of geotechnical and geoenvironmental engineering*, 125(6), 486-498.

Foufoula-Georgiou, E., and Lettenmaier, D. P. (1987). "A Markov renewal model for rainfall occurrences." *Water Resources Research*, 23(5), 875-884.

Haan, C. T., Allen, D. M., and Street, J. O. (1976). "A Markov chain model of daily rainfall." *Water Resources Research*, 12(3), 443-449.

Haberlandt, U. (1998). "Stochastic rainfall synthesis using regionalized model parameters." *Journal of Hydrologic Engineering*, 3(3), 160-168.

Haberlandt, U., Ebner von Eschenbach, A. D., and Buchwald, I. (2008). "A space-time hybrid hourly rainfall model for derived flood frequency analysis." *Hydrology and Earth System Sciences 12 (2008)*, Nr. 6, 12(6), 1353-1367.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., and Stocker, E. F. (2007). "The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales." *Journal of Hydrometeorology*, 8(1), 38-55.

Low, B. K., and Tang, W. H. (2007). "Efficient spreadsheet algorithm for first-order reliability method." *Journal of Engineering Mechanics*, 133(12), 1378-1387.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Oliver, M. A., and Webster, R. (1990). "Kriging: a method of interpolation for geographical information systems." *International Journal of Geographical Information System*, 4(3), 313-332.

Swiss Re. (2013). *Mind the risk: a global ranking of cities under threat from natural disasters*. Swiss Re.

Sklar, A. (1973). "Random variables, joint distribution functions, and copulas." *Kybernetika*, 9(6), 460.

Wilks, D. S. (1998). "Multisite generalization of a daily stochastic precipitation generation model." *Journal of Hydrology*, 210(1-4), 178-191.

Yang, C., Chandler, R. E., Isham, V. S., and Wheater, H. S. (2005). "Spatial - temporal rainfall simulation using generalized linear models." *Water Resources Research*, 41(11), W11415.