

# Uncertainty Estimation in Design Rainfalls: A Modelling Framework for Qatar Arid Region

A.A. Mamoon<sup>a,b</sup>, A. Rahman<sup>b,c</sup> and H. Qasem<sup>a</sup>

<sup>a</sup>Ministry of Municipality and Urban Planning, Doha, Qatar

<sup>b</sup>School of Computing, Engineering and Mathematics, Western Sydney University, Australia

<sup>c</sup>Institute for Infrastructure Engineering, Western Sydney University, Australia

Email: [A.Rahman@westernsydney.edu.au](mailto:A.Rahman@westernsydney.edu.au)

**Abstract:** In stormwater system design, flood insurance studies and flood protection works, hydrological models are adopted to estimate design flows. Design flows are referred to as a runoff discharge associated with a given average recurrence interval (ARI) or annual exceedance probability (AEP). These models require design rainfall as the most important input among other inputs such as catchment characteristics representing runoff routing behaviour and losses. The design rainfall, often known as intensity-duration-frequency (IDF) data, is generally derived using a regional frequency analysis approach based on a group of rainfall stations that form a homogeneous region. In Australia, design rainfall is known as intensity-frequency-duration (IFD) data. A large degree of uncertainty is associated with IDF data, which often is not quantified and considered in majority of hydrologic modelling applications.

This paper presents a modelling framework to quantify uncertainty in design rainfalls for Qatar due to uncertainties arising from limited data length and parameters of the adopted probability distribution model. Qatar is situated in arid region, which has limited rainfall data in terms of number of stations, resolution of data (e.g. only daily rainfall data is available for most of the stations) and record length of the available data. The proposed modelling framework accounts for the uncertainty in the rainfall data using a Monte Carlo simulation technique where a multivariate normal distribution is adopted in accounting for the uncertainty in the parameters of the log Pearson Type 3 (LP3) distribution. A bootstrapping method is adopted to estimate the mean and standard error values and the correlations among the three parameters of the LP3 distribution to define the multivariate normal distribution. A total of 10,000 simulations are carried out to develop the 90% confidence intervals for the 24-hour duration rainfall quantile. It has been found that uncertainty in IDF curves is quite high; to reduce the uncertainty band in estimated rainfall quantiles, a higher record length is needed, which however is not currently available in Qatar region.

The proposed modelling framework is in the developmental stage, which is applied in this paper to a single station and for one rainfall duration (24-hour). The proposed method is being enhanced by adding other sources of uncertainties in design rainfall estimation e.g. uncertainty due to data quality and climate change. Furthermore, other rainfall durations from a large number of stations will be considered, which will enable better quantification of the uncertainty in the design rainfalls in Qatar.

**Keywords:** *Arid regions, floods, IDF, Monte Carlo simulation, Qatar, rainfalls*

## 1. INTRODUCTION

Design rainfall is commonly expressed in the form of a 3-way relationship known as intensity, duration and frequency (IDF) curves, which in Australia is known as IFD curves. By using IDF curves, design rainfall intensity of a given duration and average recurrence interval (ARI) can be obtained. Design rainfall is an essential input to a hydrologic model, which is used to estimate design discharge that is needed in the planning and design of many hydraulic and drainage infrastructures. Design rainfall essentially provides a probabilistic estimation of rainfall intensity that is likely to occur at a particular location in future during the design life of an infrastructure project. The design rainfall estimation is generally made using a regionalization technique based on a regional database of observed rainfalls from a large number of stations within a region/country (Haddad *et al.*, 2011).

Numerous research studies have been carried out on design rainfall estimation around the globe, e.g. Australia (I. E. Aust., 1987; Haddad *et al.*, 2011; Haddad and Rahman, 2014; Johnson *et al.*, 2012), U. K. (NERC, 1975), USA (Bonnin *et al.*, 2006; Trefry *et al.*, 2005), Hong Kong (Yu and Cheng, 1998), Italy (Baldassarre *et al.*, 2006), Denmark (Madsen *et al.*, 2009) and Qatar (Mamoon *et al.*, 2013, Mamoon *et al.*, 2014a).

As compared to humid region, design rainfall estimation in the arid region is more challenging mainly due to significant spatial and temporal variability in the observed rainfall, and limited availability of recorded rainfall data and lack of adequate monitoring stations (Kwarteng *et al.*, 2009; Nasrallah and Balling, 1993). These factors introduce significant uncertainty in the derived IDF data, which often is not specified. Furthermore, climate change brings another dimension of uncertainty in the IDF derivation as in many cases the historical rainfall data may not satisfy the stationarity assumption (Ishak *et al.*, 2013; Seidou *et al.*, 2012; Laz *et al.*, 2014).

Many attempts have been made to examine uncertainties in design rainfall estimation and rainfall runoff modelling (Renard *et al.*, 2010; Wu *et al.*, 2011; Tung and Wong, 2014, and Soulis *et al.*, 2015). In general, the uncertainty in hydrological modelling can be divided into two main categories: (i) data and sampling errors and (ii) modelling or structural errors (Haddad and Rahman, 2014). The data uncertainty is originated from measurements errors resulting from instrumental and human errors and also due to inadequate representativeness of a data sample due to temporal and spatial variability of the data. The use of a limited quantity of rainfall data (such as data of short record length) in the frequency analysis introduces sampling uncertainty, which is transmitted to the model coefficients and, eventually to the design rainfall amount and adopted hyetograph (Tung and Wong, 2014).

Based on the twenty rain gauges in Ontario, Wang and McBean (2014) found a linear relationship between the uncertainties (in the form of the proportion of the 95% confidence intervals, compared with the expected values) and the historical rainfall record length. Using this linear relationship, it is possible to quantify the record length needed to achieve a specified level of uncertainty. Based on data from twenty rain gauges in Ontario, Canada Wang and McBean (2014) found that at least 49, 62 and 73 years of records are needed to achieve a 95% confidence interval as small as 10% of the predictions, for 5, 10 and 25 years ARI, respectively.

The uncertainty in the model parameters is attributed to inability in accurately quantifying the input parameters for a model. The parameters obtained from the calibration process are not free from uncertainty due to various reasons including data uncertainty (e.g. data used to calibrate the model may contain error), insufficient amount of data from which the parameters in an assumed model are being estimated, and model uncertainty (e.g. the model structure used to calibrate the model is not adequate). Rupa *et al.* (2015) found that major sources of uncertainty in the IDF relationships are due to insufficient quantity and quality of data leading to parameter uncertainty associated with the probability distribution being fitted to the data.

Deletic *et al.* (2012) found that uncertainties are highly interlinked, suggesting that assessing the impact of a single source is not adequate and that simultaneous propagation of key sources of uncertainties is required. A summary of various sources of uncertainties associated with design rainfall estimations are given in Table 1.

It should be noted that uncertainty is intrinsic in any modelling process. Uncertainties cannot be eliminated, but their amplitude should be estimated and, if possible, reduced (Deletic *et al.*, 2012). There have been many studies on design rainfall estimation; however, the uncertainty in design rainfall estimation

has not been incorporated in the modelling framework for derivation of IDF curves in most of the previous applications. This paper presents a modelling framework for identifying and incorporating uncertainty due to limited data length and parameters of the adopted probability model in derivation of IDF curves.

**Table 1.** Uncertainties in design rainfall estimation (Mamoon and Rahman, 2014b).

Uncertainty type	Sources of uncertainty
Data and sampling uncertainty	<ul style="list-style-type: none"> <li>• Short record length</li> <li>• Instrumental error, failure in gauges</li> <li>• Gaps in the data</li> <li>• Manual error in entering the data</li> <li>• Insufficient stations</li> <li>• Selection of AMS and PDS events</li> <li>• Non-concurrence of the data</li> <li>• Uncertainty due to correlation</li> </ul>
Uncertainty in regional methods	<ul style="list-style-type: none"> <li>• Delineation of regions</li> <li>• Lack of homogeneity within the region</li> <li>• Choice of distributions</li> <li>• Predictive variables for regression equation (e.g. longitude, latitude, elevation etc.)</li> <li>• Quantile estimation</li> <li>• Ungauged location with no measured data points</li> <li>• Parameter and quantile estimation</li> <li>• Design hyetograph</li> </ul>
Uncertainty due to climate change	<ul style="list-style-type: none"> <li>• Trend analysis</li> <li>• Selection of emissions scenario</li> <li>• Selection of global climate models (GCM)</li> <li>• Scale and resolution of GCM</li> <li>• Parameterization of GCMs</li> <li>• Downscaling of data to local scale</li> <li>• Extrapolation of areal reduction factor</li> <li>• Natural variability of climate patterns</li> </ul>

## 2. METHOD

In this study a Monte Carlo simulation (Rahman *et al.*, 2002) is adopted to assess the uncertainty in design rainfall estimation by incorporating uncertainties due to sample size and model parameters. The procedure is described below.

Assume that a three-parameter distribution, log-Pearson Type 3 (LP3) is suitable to estimate rainfall quantile, which has the following form:

$$\ln(I_Y) = M + K_Y S \tag{1}$$

where  $I_Y$  = rainfall intensity having an ARI of  $Y$  years (design rainfall or rainfall quantile);

$M$  = mean of the natural logarithms of the annual maximum (AM) rainfall time series, which in this case is formed by taking the annual maximum value of 24-hour duration for each calendar year for the entire period of data availability;

$S$  = standard deviation of the natural logarithms of the AM series of the 24-hour duration; and

$K_Y$  = frequency factor for the LP3 distribution for ARI of  $Y$  years, which is a function of the ARI and the skewness ( $SK$ ) of the natural logarithms of the AM rainfall series, there are tables of  $K_Y$  values in statistical text book that can be applied in practice.

To specify the uncertainty in the parameters of the LP3 distribution, a multivariate normal (MVN) distribution is adopted where it is assumed that  $M$ ,  $S$  and  $SK$  at the site of interest have a MVN distribution.

The MVN distribution is specified by the mean and standard error of each of three parameters of the LP3 distribution (i.e.  $M$ ,  $S$  and  $SK$ ) and correlations among these parameters, as explained below.

$M$ :  $N \sim (y_1, s_{y1})$ :  $y_1$  is estimated mean for the site (obtained by bootstrapping) and  $s_{y1}$  = standard error of mean (obtained by bootstrapping). The adopted bootstrapping assumes that AM rainfall series has no serial correlation, which is a reasonable assumption.

$S$ :  $N \sim (y_2, s_{y2})$ :  $y_2$  is estimated  $S$  for the site and  $s_{y2}$  = standard error of  $S$  (obtained by bootstrapping).

$SK$ :  $N \sim (y_3, s_{y3})$ :  $y_3$  is estimated  $SK$  for the site and  $s_{y3}$  = standard error of  $SK$  (obtained by bootstrapping).

Correlations among ( $M$ ,  $S$  and  $SK$ ) are estimated by bootstrapping where many samples are generated from the observed AM rainfall data and  $M$ ,  $S$  and  $SK$  values are computed for each of the boot-strapped samples, which are then used to estimate the Pearson correlation coefficients among  $M$  and  $S$ ,  $S$  and  $SK$  and  $M$  and  $SK$ .

Monte Carlo simulation: 10, 000 sets of values of  $M$ ,  $S$  and  $SK$  at the site of interest are generated using the above MVN. Using equation 1, 10,000 values of  $I_Y$  are then estimated, which are then ranked to get the 5<sup>th</sup> and 95<sup>th</sup> percentiles of  $I_Y$ , which define the 5% and 95% CL of  $I_Y$ .

In the above approach, the uncertainty due to limited data length is accounted for by generating many different samples from the original data by boot-strapping. The uncertainty in the parameters of the LP3 distribution is accounted for by introducing a standard error value in the mean of the parameters of the LP3 distribution. The method needs further enhancement by adding other sources of uncertainty e.g. poor data quality and trends in the AM rainfall data.

### 3. RESULTS

To estimate the mean and standard error values of the of the  $M$ ,  $S$  and  $SK$  parameters of the LP3 distribution (equation 1), a bootstrapping is conducted. In this paper, the analysis is done for one station and single rainfall duration (24-hour). The results of this bootstrapping for Site S002 are presented here. Site S002 has 39 years of 24-hour duration AM rainfall data. A total of 5,000 samples are generated for the site by bootstrapping (with replacement), each with sample size of  $N-5 = 34$ . The results of the bootstrapping for Site 002 are summarized in Table 2. The average of the simulated  $M$  values is 1.298 mm/h, against 1.361 mm/h for the original full length data, representing a difference of 4.6%, which seems to be reasonable.

**Table 2.** Mean and standard error for  $M$ ,  $S$  and  $SK$  of  $\ln(I)$  data estimated by bootstrapping (Station S002).

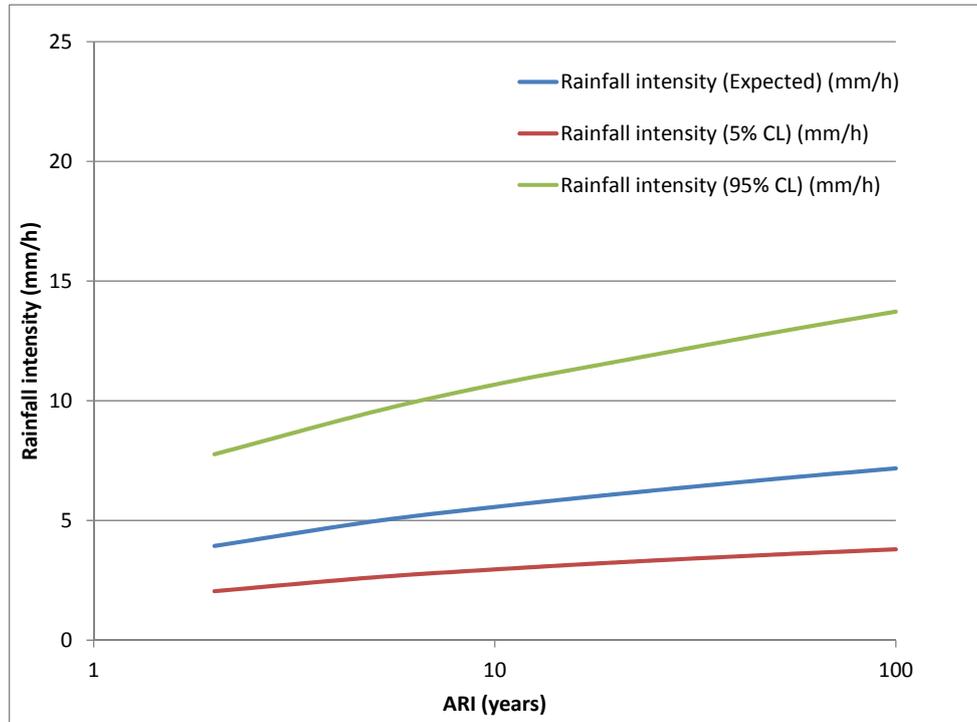
Parameter of LP3 distribution	Average value from bootstrapping	Standard error value from bootstrapping
$M$	$y_1 = 1.298$ (mm/h)	0.066 (mm/h)
$S$	$y_2 = 0.275$ (mm/h)	0.134 (mm/h)
$SK$	$y_3 = -1.103$	0.917

Based on 10,000 simulated values of  $I_Y$  from the multivariate normal distribution, the estimated expected values of the quantiles for the 24-hour duration rainfall are shown in column 2 of Table 3. The 5% and 95% confidence limits (CL) of the rainfall quantiles for ARIs of 2, 5, 10, 20, 50 and 100 years are shown in Table 3. The confidence limits presented in Table 3 and Figure 1 show that the uncertainty in rainfall estimates increases with increasing ARIs as expected.

It should be noted here that the confidence limit shown here is mainly governed by the standard error values (Table 2, column 3), which is inversely proportional to the sample size, i.e. to reduce the confidence limits, data length should be higher. However, in the arid regions like Qatar, rainfall record length is quite short. Other sources of uncertainty e.g. model type needs to be included in the Monte Carlo simulation and hence GEV distribution will be applied with method of  $L$ -moments as this is one of the most widely adopted distributions in IDF estimation. Furthermore, uncertainty due to climate change will be incorporated in the analysis. The climate change is expected to results in a trend in the AM rainfall series which makes the mean, standard deviation and skew of the AM rainfall data non-stationary. A non-stationary frequency analysis needs to be adopted to account for this type of uncertainty (e.g. Cunderlik and Ouarda 2006; Khaliq et al. 2006; El Adlouni et al. 2007; Ishak et al., 2013).

**Table 3.** Rainfall quantiles from Monte Carlo simulation (Site S002).

ARI (years)	Rainfall intensity (Expected) (mm/h)	Rainfall intensity (5% CL) (mm/h)	Rainfall intensity (95% CL) (mm/h)
2	3.94	2.05	7.77
5	4.97	2.63	9.55
10	5.57	2.96	10.68
20	6.1	3.25	11.63
50	6.74	3.58	12.86
100	7.18	3.81	13.73



**Figure 1.** Plot of 90% confidence interval for 24-hour rainfall quantile (Station 002 in Qatar).

#### 4. DISCUSSION

This exploratory analysis illustrates how uncertainty in design rainfalls can be estimated by Monte Carlo simulation. In this example, sampling uncertainty is focused, which can be extended by adding other sources of uncertainty e.g. serial correlation in the AM data and model uncertainty. The uncertainty due to record length is the dominant source of uncertainty as record lengths of rainfall data in many cases are quite limited. For example, Wang and McBean (2014) showed that to get 95% confidence interval as small as 10% of the predictions, about 62 years of record is needed for 10 years ARI in Ontario, Canada. They noted that since rainfall record lengths are generally smaller than 50 years (the typical data length), the associated uncertainty in design rainfall estimates is substantial, which needs to be considered in design applications such as sizing urban drainage systems. In this regard, Rupa et al. (2015) stated that the common approaches for design rainfall estimation often leads to systematic underestimation (overestimation in only few cases) mainly due to short data length. They highlighted that in most applications, parameter uncertainty is ignored in design rainfall estimation. Both of these uncertainties (i.e. uncertainties due to record length and model parameters) can be evaluated by the method proposed in this study.

The uncertainty analysis in design rainfall estimates is important in practical applications where hydrologic models are used to estimate design flows. For example, a single design rainfall value is generally used in estimating design discharge without considering the impacts of uncertainty in the design rainfall, which can undermine the design flow estimates. The uncertainty in design rainfall can be incorporated into the hydrologic model by applying a Monte Carlo simulation where the uncertainty of all the input variables including input rainfall and model parameters can be included. In Australian Rainfall and Runoff (ARR) 2015, a Monte Carlo simulation approach is proposed to estimate design flows where uncertainty in design

rainfall can be considered in design flow estimation. This presents a paradigm shift in hydrologic modelling in Australia where uncertainty will be a major focus in ARR 2015.

## 5. CONCLUSION

Uncertainty in design rainfall estimates can arise from different sources, such as poor quality data, limitation in record length, model type and parameter estimation procedure. A method is presented in this paper to quantify uncertainty in design rainfall estimates due to limited data length. The proposed method involves use of Monte Carlo simulation technique where 10,000 estimates of rainfall quantiles for a given average recurrence interval (ARI) are generated from the fitted log Pearson Type 3 (LP3) distribution by adopting a multivariate normal distribution. The multivariate normal distribution is defined by bootstrapping of the observed annual maximum (AM) series of 24-hour rainfall duration at the selected station. This bootstrapping method provides an estimate of the mean and standard error of the parameters of the LP3 distribution, and their correlation structure. The proposed Monte Carlo simulation technique is being enhanced by adding other sources of uncertainty (e.g. data quality and trends in the AM rainfall data), which will enable to define uncertainty in the final design rainfall data in Qatar.

## ACKNOWLEDGMENTS

The authors would like to acknowledge the Ministry of Municipality and Urban Planning, Qatar for providing the data for the study.

## REFERENCES

- Baldassarre, D.G., Brath, A. and Montanari, A. (2006). Reliability of different depth-duration-frequency equations for estimating short-duration storms, *Water Resources Research*, 42, W12501. doi:10.1029/2006WR004911.
- Bonnin, G.M., Martin, D., Lin, B., Parzybokt, T., Yekta, M. and Riley, D. (2006). Precipitation-Frequency Atlas of the United States, Vol. 1, NOAA Atlas 14, NOAA, USA.
- Cunderlik, J.M. and Ouarda, T. (2006). Regional flood-duration-frequency modeling in the changing environment, *Journal of Hydrology*, 318, 1-4, 276-291.
- Deletic, A., Dotto, C.B.S. et al. (2012). Assessing uncertainties in urban drainage models, *Journal of Physics and Chemistry of Earth*, 42-44, 3-10.
- El Adlouni S, Ouarda T, Zhang X, Roy R, Bobee B 2007 Generalized maximum likelihood estimators for the non-stationary generalized extreme value model”, *Water Resources Research*, 43, W03410, doi:10.1029/2005WR004545.
- Haddad, K. and Rahman, A. (2014). Design rainfall estimation and changes, In Handbook of Engineering Hydrology: Modeling, Climate Change and Variability, Edited by SaeidEslamian, CRC Press, 173-190.
- Haddad, K., Rahman, A. and Green, J. (2011). Design rainfall estimation in Australia: a case study using L-moments and generalized least squares regression, *Stochastic Environmental Research and Risk Assessment*, 25, 6, 815-825.
- Haddad, K. and Rahman, A. (2014). Derivation of short duration design rainfalls using daily rainfall statistics, *Natural Hazards*, 74, 1391-1401.
- Institution of Engineers Australia (I. E. Aust.) (1987). Australian Rainfall and Runoff: A Guide to Flood Estimation. Edited by D. H. Pilgrim, Vol. 1, I. E. Aust., Canberra.
- Ishak, E.H., Rahman, A., Westra, S., Sharma, A. and Kuczera, G. (2013). Evaluating the non-stationarity of Australian annual maximum flood, *Journal of Hydrology*, 494, 134-145.
- Johnson, F., Haddad, K., Rahman, A. and Green, J. (2012). Application of Bayesian GLSR to estimate sub-daily rainfall parameters for the IFD Revision Project. Hydrology and Water Resources Symposium, 19-22 Nov 2012, Sydney, Australia.
- Khaliq, M.N., Ouarda, T.B.M.J., Ondo, J.C., Gachon, P. and Bobee, B. (2006). Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: a review, *Journal of Hydrology*, 329, 3-4, 534-552.

- Kwarteng, A.Y., Dorvlob, A.S. and Gupta, V.K. (2009). Analysis of a 27-year rainfall data (1977–2003) in the Sultanate of Oman, *International Journal of Climatology*, 29, 605–17.
- Laz, O.U., Rahman, A., Yilmaz, A. and Haddad, K. (2014). Trends in sub hourly, sub daily and daily extreme rainfall events in south-east Australia, *Journal of Water and Climate Change*, 5, 4, 667-675.
- Madsen, H., Arnbjerg-Neilsen, K., Mikkelsen, P.S. (2009). Update of regional intensity-duration-frequency curves in Denmark: tendency towards increased storm intensities, *Atmospheric Research*, 92, 343-349.
- Mamoon, A.A., Joergensen, N.E., Rahman, A. and Qasem, H. (2013). Estimation of Design Rainfall in Arid Region: A Case Study for Qatar Using L Moments. Proceedings of the 35th IAHR World Congress, September 8 to 13, 2013 Chengdu, China, 1-9.
- Mamoon, A.A., Jeorgensen, N.E., Rahman, A. and Qasem, H. (2014a). Derivation of new design rainfall in Qatar using L-moments based index frequency approach, *International Journal of Sustainable Built Environment*, 3, 111-118.
- Mamoon, A.A. and Rahman, A. (2014b). Identification of Rainfall Trends in Qatar, International Conference on Environmental Systems Science and Engineering, 15 -16 Dec 2014, Sydney, Australia, 8, 12, Part VI, 735-739.
- Mamoon, A.A. and Rahman, A. (2014c). Uncertainty in Design Rainfall Estimation: A Review, *Journal of Hydrology and Environment Research*, 2, 1, 65-75.
- Nasrallah, H.A. and Balling, Jr, R.C. (1993). Analysis of recent climatic changes in the Arabian Gulf region, *Environmental Conservation*, 20(3), 223–226.
- Natural Environmental Research Council (NERC) (1975). Flood Studies Report, Vol. II Meteorological studies. NERC, London.
- Rahman, A., Weinmann, P.E., Hoang, T.M.T. and Laurenson, E.M. (2002). Monte Carlo Simulation of flood frequency curves from rainfall, *Journal of Hydrology*, 256, 196-210.
- Renard, B., Kavetski, D., Kuczera, G., Thyer, M. and Franks, S.W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resources Research*, 46, W05521, doi:10.1029/2009WR008328.
- Rupa, C., Saha, U. and Majumdar, P.P. (2015). Model and parameter uncertainty in IDF relationships under climate change. *Advances in Water Resources*, 79, 127–139.
- Seidou, O., Ramsay, A. and Nistor, I. (2012). Climate change impacts on extreme floods I: combining imperfect deterministic simulations and non-stationary frequency analysis, *Natural Hazards*, 61, 647–659.
- Soulis, E.D., Princz, D. and Wong, J. (2015). Renewal and Update of MTO IDF Curves: Defining the Uncertainty. *Journal of Water Management Modeling*, C386.
- Trefry, C.M., Watkins Jr, D.W. and Johnson, D. (2005). Regional Rainfall Frequency Analysis for the State of Michigan, *Journal of Hydrologic Engineering*, 10, 437-449.
- Tung, Y. and Wong, C. (2014). Assessment of design rainfall uncertainty for hydrologic engineering applications in Hong Kong, *Stochastic Environmental Research and Risk Assessment*, 28,583–592.
- Yu, P. and Cheng, C. (1998). Incorporating uncertainty analysis into a regional IDF formula, *Hydrological Processes*, 12, 713-726.
- Wang, Y. and McBean, E.A. (2014). Uncertainty Characterization of Rainfall Inputs used in the Design of Storm Sewer Infrastructure. *Journal of Water Management Modeling*, C367, doi: 10.14796/JWMM.
- Wu, S.J., Yang, J.C. and Tung, Y.K. (2011). Risk analysis for flood control structure under consideration of uncertainties in design flood, *Natural Hazards*, 58(1), 117–140.