Exploration of intensity change in daily precipitation using bias-corrected ERA-20c in South Korea for the 20th century (1900-2010)

Running Head: Daily rainfall intensity change in South Korea for the 20th century

By

Dong-Ik Kim[1], Hyun-Han Kwon[2*] and Dawei Han[1], Yong-Tak Kim[2]

[1:] Water and Environment Research Group, Department of Civil Engineering, University of Bristol, United Kingdom
[2:] Sejong University, Seoul, Korea
*: Corresponding Author: Hyun-Han Kwon, hkwon@sejong.ac.kr

Abstract

The rainfall frequency analysis has been routinely adopted for the estimation of design rainfall for a given specific return period. The annual maximum rainfall data are generally used for the frequency analysis in practice, but the parameters of the probability distribution are estimated from the limited data that are often available since the 1970s in many regions including South Korea. As an alternative, this study aims to utilize a century-long ERA-20c daily precipitation data, which have been provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). To reduce the systematic errors in the reanalysis data, we introduce a quantile delta mapping method using a composite Gamma-Pareto distribution (QDM-GP) that can better represent temporal trends and extreme events, compared with the stationary quantile mapping (SQM). We also evaluate the degree of uncertainty reduction in the estimation of design rainfall with the use of the bias corrected ERA-20c within a Bayesian modeling framework. Finally, the bias-corrected data are applied to explore the spatio-temporal change of design rainfall in South Korea for the 20^{th} century. To investigate changes in design rainfall under the nonstationary assumption, this study estimates the design rainfall using the data from three different periods (i.e. for the period of 1900-1936, 1937-1973, and 1974-2010). It is found that the QDM can substantially reduce the bias in annual maximum rainfall (AMR). The uncertainty ranges of design rainfall using the bias corrected ERA-20c are generally within those in the observed, suggesting the use of the bias corrected reanalysis data can reduce uncertainties in design rainfall by increasing the sample size. Furthermore, we have explored the role of the bias corrected rainfall for uncertainty reduction in design rainfall via three different experiments in the context of prior information within a Bayesian framework. In the experimental study, we conclude that the uncertainty reduction in the design rainfall can be mainly attributed to the use of prior distribution for the shape parameter, informed by the long-term reanalysis data. Moreover, a significant spatio-temporal change in design rainfall is observed over the entire South Korea. The significant change in design rainfall is mainly attributed to the recent increase in the rainfall intensity, leading to a potential increase in the flood risk in view of the present in most areas.

Keywords: Composite distribution, ERA-20c, Precipitation, MCMC, Quantile delta mapping, Uncertainty

1. Introduction

The rainfall frequency analysis is routinely adopted for the estimation of the design rainfall with a return period, which is an essential part in a water-related planning in a certain area. The extreme value distributions such as the generalized extreme value (GEV) or Gumbel distribution are commonly applied to the annual maximum rainfall series (AMRs). However, there exist significant uncertainties in the estimation of design rainfall due to sampling error, which is related to the limited AMRs and the use of an improper distribution (Huard *et al.*, 2010). Moreover, it has been acknowledged that the observed data are usually found to be insufficient for the spatial analysis of extreme value in ungauged catchment. In this context, geospatial approaches could be used for an interpolation to provide hydrologic variables (or parameters) over the entire watershed. For example, reliable long-term daily precipitation data over South Korea are not readily available, and a continuous daily precipitation records over the past 40 years are available from only about 50 stations. In this setting, the estimated design rainfall for a higher return period more than the length of AMRs might be exacerbated by the sampling error. Given these circumstances, with the reliable long-term precipitation data at a fine spatial resolution, reanalysis data could be an alternative in hydrologic modeling, especially for the design rainfall estimation.

Reanalysis datasets are produced by numerical models informed by advanced data assimilation techniques and they have been widely adopted in the climate change studies (Dee *et al.*, 2011; Zhang *et al.*, 2013; Hersbach *et al.*, 2015; Donat *et al.*, 2016; Gao *et al.*, 2016; Poli *et al.*, 2016). There are several types of reanalysis datasets available, but the three century-long daily reanalysis datasets are globally available, namely, the National Oceanic and Atmospheric Administration (NOAA) 20th century reanalysis (NOAA-20cR), the European Centre for Medium-Range Weather Forecasts (ECMWF) century atmospheric model ensemble (ERA-20cm) and ECMWF 20th century assimilation surface observations only (ERA-20c), covering from 1900 to 2010 (Compo *et al.*, 2011; Hersbach *et al.*, 2015; Poli *et al.*, 2016). The main differences among these datasets are the assimilation techniques

and spatial resolution. Unlike ERA-20c and NOAA-20cR, ERA-20cm does not consider observations in the assimilation process so that the synoptic patterns may not be well reproduced (Hersbach *et al.*, 2015; Donat *et al.*, 2016; Gao *et al.*, 2016; Poli *et al.*, 2016). It has been acknowledged that the mean climate is reasonably well reproduced in both NOAA-20cR and ERA-20c, whereas NOAA-20cR has coarser spatial resolution, 1.875°×1.9° compared to the ERA-20c which has the higher resolution, 0.125°×0.125. For these reasons, we mainly used ERA-20c reanalysis data to explore changes in design rainfall over the last century.

A primary concern of the use of reanalysis data for characterizing long-term climate trends is to reduce the systematic errors. Previous studies have shown that century-long reanalysis data may misrepresent long-term climatic trends or synoptic patterns, especially for the first half of twentieth century, and there exists the difference in temporal variability between century-long reanalyses and observations (Brands *et al.*, 2012; Krueger *et al.*, 2013; Poli *et al.*, 2013; Befort *et al.*, 2016; Donat *et al.*, 2016; Kim and Han, 2018). Thus, the reanalysis data without an attempt to adjust the bias can be problematic for many hydrologic applications. There are various approaches for the bias correction from simpler method (e.g. the delta method) to more complex procedures (e.g. the quantile mapping approach, multivariate approach and multiscale approach). Among these methods, the quantile mapping (QM) approach has been applied extensively to reduce the systematic biases in the climate model outputs (Themeßl *et al.*, 2011; Teutschbein and Seibert, 2012; Fang *et al.*, 2015; Maraun, 2016; Maraun and Widmann, 2018).

A fundamental assumption of the traditional QM approach is that the biases in the numerical modeling data are stationary for the reference period. However, recent studies have shown that climate variables including precipitation are often viewed as non-stationary. It has been well documented that there is a significant increasing trend over South Korea, especially in summer (Chang and Kwon, 2007; Choi *et al.*, 2009; Jung *et al.*, 2011; Cannon *et al.*, 2015; Miao *et al.*, 2016; Eum and Cannon, 2017; Nahar *et al.*, 2017). Recently, the bias correction method with the consideration of nonstationarity has

been proposed to better represent the nonstationarity in climate variables. Bürger et al. (2013) suggested the detrended quantile mapping (DQM) approach, removing the trends for future climate. Li et al. (2010) proposed an equidistant QM algorithm to reduce biases in the tails of the distribution of climate change scenarios. Cannon et al. (2015) proposed a quantile mapping approach for preserving trend, namely, the quantile delta mapping (QDM) approach. They confirmed that QDM is generally better in terms of reducing the uncertainty in the GCMs for the future period than DQM as well as the conventional QM. In this perspective, we adopt the QDM algorithm for the non-stationary bias correction of century-long reanalysis, especially for extreme rainfall events. The only difference from the QDM proposed by Cannon et al. (2015) is that we superimpose the delta change for the past period, not future period. For comparison purposes, the stationary QM (SQM) approach is also explored in this study.

On the other hand, the use of long-term data in the hydrologic frequency analysis can substantially reduce the uncertainty of design rainfall estimation (Coles *et al.*, 2003; Overeem *et al.*, 2008; Huard *et al.*, 2010; Tung and Wong, 2014; Van de Vyver, 2015). In this context, the use of the century-long reanalysis data can be used to reduce the uncertainty of design rainfall estimation in the rainfall frequency analysis. To explore reduction in the uncertainty of design rainfall, a set of parameters of probability distribution in the frequency analysis was estimated in a Bayesian framework (Reis and Stedinger, 2005; Kwon *et al.*, 2008, 2011; Huard *et al.*, 2010; Van de Vyver, 2015). There are several climate change studies for the historical rainfalls with the limited data to estimate the changes in rainfall intensity (Chung and Yoon, 2000; Ho *et al.*, 2003; Chang and Kwon, 2007; Choi *et al.*, 2009; Jung *et al.*, 2011). Here, the uncertainty of the parameters and their reduction with the use of the ERA-20c are studied for the first time, which are then evaluated by comparing the predictive posterior distribution of design rainfall.

From this background, we mainly focus on investigating the following questions:

(1) Can the QDM approach be more effective for reducing the errors compared with the SQM?

(2) Can the use of ERA-20c be effective in reduction of uncertainty in estimating design rainfall?

(3) How much has the design rainfall changed for the whole of 20th century in South Korea?

To address these questions, our aims are three-folds. First, we applied two bias correction schemes, SQM and QDM with a composite distribution of a gamma and GPD as for the transfer function. Here, their comparison is performed for the century-long ERA-20c with three different periods (1900-1936, 1937-1973, 1974-2010). Second, the degree of reduction in uncertainty of the design rainfall corresponded with the use of bias corrected AMRs was explored with that of the observed for the reference period 1974-2010, within the Bayesian model framework. Third, we explored the spatio-temporal change in design rainfall over the last century. The data used in this study are summarized in Section 2, and the theoretical background for the methodology is then introduced in Section 3. The results and discussion are summarized in Section 4 and concluding remarks are finally provided in Section 5.

2. Data

2.1 Weather Stations

In South Korea, there exist hundreds of local weather stations, but the historical records for more than 40 years are available from only several stations. Subsequently, we selected 48 rain gauges, covering from 1974 to 2010, over South Korea. Among these, for the evaluation, this study used 7 stations which provide a longer period of historical records: St 4. Gangneung (1912-2010), St. 5 Seoul (1910-2010), St. 6 Incheon (1910-2010), St. 17 Daegu (1912-2010), St. 18 Jeonju (1919-2010), St 21. Busan (1912-2010), and St 22. Mokpo (1910-2010). The daily precipitation data were collected and compiled from the Korea Meteorological Administration (KMA). The specific locations of weather stations used in this study are illustrated in Figure 1 and Table 1.

[Insert Figure 1 and Table 1]

2.2 ERA-20c daily precipitation

The ERA-20c reanalysis data is modelled by the ECWMF Integrated Forecasting System (IFS), covering the period 1900 to 2010. The ERA-20c global reanalysis data was modelled by data assimilation schemes with ocean-global atmosphere observing system, and using sea surface temperature and sea ice concentration as a set of boundary conditions. In the ERA-20c system, daily precipitation totals can be obtained as 24 hour accumulations (Poli *et al.*, 2016). In this study, we collected the daily precipitation totals over South Korea for the period 1900-2010, via the ECMWF web server on a fine grid, 0.125°×0.125°. It should be noted that the ERA-20c reanalysis data was finally obtained as a single simulation, without providing a large ensemble with uncertainties. The grid points of ERA-20c along with the locations of weather stations used in this study are illustrated in Figure 1. Here, the nearest grid centering at the target station was extracted for the subsequent analysis.

3. Methodology

In this section, two main approaches introduced in this study are demonstrated. First, the QDM is presented with a primary focus on the use of composite distribution as a transfer function. Second, a Bayesian parameter estimation approach to rainfall frequency analysis is briefly provided.

3.1 Quantile Delta Mapping with a Composite Distribution

QM is a commonly used method in bias correction studies. In the QM approach, the systematic biases can be efficiently removed by matching cumulative distribution function (CDF) of the modelled data into that of the observed (Teutschbein and Seibert, 2012; Rabiei and Haberlandt, 2015). The QMs for daily precipitation have involved transfer functions that are typically based on parametric and nonparametric distributions (Teutschbein and Seibert, 2012; Cannon *et al.*, 2015; Kim *et al.*, 2015a, 2015b; Eum and Cannon, 2017). Nonparametric QM can lead to an increase of bias in the upper quantile for the extreme values, moreover, parametric QM typically using a gamma distribution also

often fails to represent the extreme values (Maraun, 2016; Volosciuk *et al.*, 2017; Maraun and Widmann, 2018). In this context, one can consider a composite distribution of combining a gamma distribution for modeling the interior part of the distribution and the tails by generalized Pareto distribution (GPD) for heavy-tailed distribution, (Vrac and Naveau, 2007; Gutjahr and Heinemann, 2013; So *et al.*, 2015; Volosciuk *et al.*, 2017). More specifically, the composite distribution is composed by parametrically modeling the rainfall events over the upper threshold using a GPD and the events below the threshold using a gamma distribution, as follows:

$$x_{cor} = \begin{cases} F_{o,gam}^{-1}\left[F_{m,gam}(x_m)\right], & \text{if } x \leq \text{upper threshold} \\ F_{o,GPD}^{-1}\left[F_{m,GPD}(x_m)\right], & \text{if } x > \text{upper threshold} \end{cases} \quad (1)$$

Here, subscripts *o* and *m* represent the observed data and modelled data, respectively, and $F_{m,gam}$ and $F_{m,GPD}$ are the CDFs of the ERA-20c model for gamma and GPD. Similarly, $F_{o,gam}^{-1}$ and $F_{o,GPD}^{-1}$ are the inverse (or quantile) function of CDFs of observations for gamma and GPD, respectively. The cumulative distributions for gamma distribution and GPD are defined as follows (Coles, 2001; Gutjahr and Heinemann, 2013):

$$F(x|\alpha,\beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}\int_0^x t^{\alpha-1}e^{-t/\beta}dt\,;\ x \geq 0;\ \alpha,\beta > 0 \quad (2)$$

$$F(x) = P_r(X - u \leq x\,|X > u) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\theta}\right)^{-\frac{1}{\xi}} & for\ \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\theta}\right) & for\ \xi = 0 \end{cases} \quad (3)$$

where, α and β are the shape and scale parameters of the gamma distribution in Equation (2), and u, ξ, and $\theta = \sigma + \xi(u - \mu)$ represent the upper threshold, shape parameter and the reparametrized scale parameter of the GPD in Equation (3), respectively. Note that the parameters of gamma distribution are estimated on a monthly basis, whereas the parameters of GPD are estimated using the entire peak-over-thresholds for all months. For the upper threshold in Equation 3, we explored three thresholds, the 90th, 95th and 99th percentiles, for both the modelled and observed data, respectively. Among the thresholds, both 95th and 99th percentiles for the observation and ERA-20c give fairly good results, and two thresholds are used in the subsequent analysis. More detailed results on the threshold

selection were provided in Appendix A.

Basically, the conventional QM algorithm, SQM, assumes that the degree of bias in the climate model are stationary for the simulation period (Teutschbein and Seibert, 2012; Cannon *et al.*, 2015), and the QM algorithm is specifically designed to reduce bias in the climate model in the probability space. In this approach, CDFs of the historical records are constructed for the reference period, and CDFs of climate model outputs for the entire projection period are then mapped to that of observed as follows:

$$\hat{x}_{m,p} = F_{o,r}^{-1}[F_{m,r}\{x_{m,p}(t)\}] \quad (4)$$

Here, $F_{o,r}$ and $F_{m,\mathrm{r}}$ are the CDFs of the observed and modelled for the reference period, denoted by *r*, respectively, while $\hat{x}_{x,p}(t)$ and $x_{m.p}$ are the bias-corrected and uncorrected (or modelled) data at time *t* over the simulation period, denoted by *p*. In this study, historical records from 1974 to 2010 in 48 stations and their corresponding values in the climate model were used as reference data, while the period from 1900 to 2010 was considered as the simulation period. To begin with, we corrected the wet day frequency error, namely "drizzled effect", from the ERA-20c reanalysis data. It has been well acknowledged that the wet-day frequency of the modelled daily precipitation data from climate models is typically overestimated. For this reason, a cut-off threshold (TH) has been commonly employed in the bias correction scheme for modelled daily precipitation data (Schmidli *et al.*, 2006; Piani *et al.*, 2010; Themeßl *et al.*, 2011; Kim *et al.*, 2015a, 2015b; Rabiei and Haberlandt, 2015; Nyunt *et al.*, 2016; Volosciuk *et al.*, 2017). In this study, we set the wet-day frequency in the modelled precipitation equal to that of the observed. More specifically, we divided the whole period into three different periods with the same length (1900-1936, 1937-1973, 1974-2010) because climate is usually defined with 30 or more years, and the wet-day frequency of the modelled precipitation for each time period was set equal to that of the observed for the reference period (1974-2010). After adjusting for the wet-day frequency, a composite distribution is used to construct the CDFs (i.e. Equation 4). The SQM approaches with the 95$^{\mathrm{th}}$ and 99$^{\mathrm{th}}$ thresholds were named as SQM95 and SQM99, respectively. For the SQM, the bias corrected AMRs can be exceptionally higher than the range of the AMRs of the observed in the

reference period due to misrepresentation of the upper tail of the distribution. To reduce the over-estimation of the bias corrected AMRs, various extrapolation techniques have been proposed (Themeßl *et al.*, 2011; Eum and Cannon, 2017; Li *et al.*, 2017). Among these, in this study, we used a constant extrapolation scheme over the high quantiles suggested by Themeßl et al. (2011).

One major issue in the bias-correction is nonstationarity. As discussed in the introduction section, there may be significant nonstationarity in precipitation in many regions including South Korea, thus, quantile delta mapping (QDM) approach that is effective in preserving the long-term trend (Li *et al.*, 2010; Cannon *et al.*, 2015; Miao *et al.*, 2016; Eum and Cannon, 2017) was employed for correcting bias of the ERA-20c precipitation in terms of mean and extreme. We begin with adjusting the wet-day frequency with the same threshold used in SQM, and QDM algorithm is subseuqnetly applied for the bias correction of the ERA-20c daily precipitation ranging from 1900 to 2010. As noted in the previous section, ERA-20c daily precipitation was first divided into three period with the same data length and bias in precipitation in each period was then corrected by the QDM approach, as follows (Cannon *et al.*, 2015; Eum and Cannon, 2017):

$$\tau_{m,p}\,(t) = F_{m,p}^{(t)}\left[x_{m,p}(t)\right], \qquad \tau_{m,p}\,(t) \in \{0,1\} \tag{5}$$

$$\Delta_m(t) = \frac{F_{m,p}^{-1}\left(\tau_{m,p}\,(t)\right)}{F_{m,r}^{-1}\left(\tau_{m,p}\,(t)\right)} = \frac{x_{m,p}\,(t)}{F_{m,r}^{-1}[F_{m,p}\,(x_{m,p}(t))]} \tag{6}$$

$$\hat{x}_{m,p} = F_{o,r}^{-1}[\tau_{m,p}(t)] \times \Delta_m(t) = F_{o,r}^{-1}[F_{m,p}\{x_{m,p}(t)\}] \times \Delta_m(t) \tag{7}$$

Here, $\tau_{m,p}(t)$ is the nonexceedance probability associated with the value at time t, $\Delta_m(t)$ is the relative change in quantiles between the reference period (1974-2010) and the projected period, and $F_{m,r}$ and $F_{m,p}$ are the CDFs of the modelled for the reference period and simulation period, respectively. A composite distribution described in Equations 1 to 3 was adopted for estimating the CDFs in Equations 5 to 7. The QDMs with the 95th and 99th upper thresholds were abbrieviated as QDM95 and QDM99, respectively. More specific information on the QDM can be found from an earlier study (Cannon et al., 2015).

For the evaluation of the proposed models (SQM95, SQM99, QDM95 and QDM99), two efficiency measures such as the root mean square error (RMSE) and Nash-Sutcliffe efficiency (NSE) are considered, as shown in Equations 8 and 9:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i^{\mathrm{obs}} - Y_i^{sim}\right)^2}{n}} \tag{8}$$

$$NSE = 1 - \left[\frac{\sum_{i=1}^{n}\left(Y_i^{obs} - Y_i^{sim}\right)^2}{\sum_{i=1}^{n}\left(Y_i^{obs} - Y_i^{mean}\right)^2}\right] \tag{9}$$

Here, Y_i^{obs} is the i-th observation, Y_i^{mean} is the mean of the observation, while Y_i^{sim} is the modelled data, and n is the number of observations. For a favorable model performance, the NSE should be close to 1 while values close to 0 for the RMSE. In this analysis, we evaluated the bias corrected AMRs over a century long historical record (1910-2010) for 7 out of 48 stations (i.e. St.4 Gangneung, St.5 Seoul, St.6 Incheon, St.17 Daegu, St.18 Jeonju, St.21 Busan and St.22 Mokpo) and over the last four decades (1974-2010) for 48 stations. To find out how well the bias corrected AMRs can represent the observed values, the AMRs over three different periods, 1974-2010, 1937-1973 and 1910-1973, were additionally evaluated for the 7 stations.

3.2 Bayesian Parameter Estimation

Extreme precipitation is commonly characterized with the block maxima method of the extreme value theory. Specifically, AMRs were first obtained and fitted to a GEV (generalized extreme value) distribution. Suppose **R** indicates the AMRs in a given daily precipitation, and the CDF of the GEV distribution is then defined as bellows:

$$f(\mathbf{R};\mu,\alpha,\xi) = \begin{cases} \left(\frac{1}{\sigma}\right)\left(1+\frac{\xi(\mathbf{R}-\mu)}{\sigma}\right)^{-\frac{1}{\xi}-1} exp\left\{-\left[1+\frac{\xi(\mathbf{R}-\mu)}{\sigma}\right]^{-\frac{1}{\xi}}\right\}, & \xi \neq 0 \\ \left(\frac{1}{\sigma}\right) exp\left\{-exp\left[-\frac{(\mathbf{R}-\mu)}{\sigma}\right] - \frac{(\mathbf{R}-\mu)}{\sigma}\right\} & \xi = 0 \end{cases} \tag{10}$$

where μ, σ, ξ are the location, scale and shape parameter, respectively.

In this study, the parameters of distribution functions were estimated within a Bayesian modelling framework, and the derived posterior distributions of the parameters were further used to estimate design rainfalls and their uncertainties. Theoretically, the posterior distribution, $p(\boldsymbol{\theta}|\mathbf{R})$, of the parameter vector ($\underline{\boldsymbol{\theta}}$) is described as follow:

$$p(\boldsymbol{\theta}|\mathbf{R}) = \frac{p(\boldsymbol{\theta},\mathbf{R})}{p(\mathbf{R})} = \frac{p(\mathbf{R}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{R})} = \frac{p(\mathbf{R}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\boldsymbol{\theta})p(\mathbf{R}|\boldsymbol{\theta})d\theta} \propto p(\mathbf{R}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \qquad (11)$$

where, $\mathbf{R}$ is the vector of the AMRs, $p(\mathbf{R}|\boldsymbol{\theta})$ is the likelihood function, and $p(\mathbf{R})$ and $p(\boldsymbol{\theta})$ are the marginal distribution and prior distribution, respectively. The joint posterior distribution function $p(\boldsymbol{\theta}|\mathbf{R})$ for the rainfall frequency model can be formulated by combining the GEV likelihood function and prior distribution as follows:

$$p(\boldsymbol{\theta}|\mathbf{R}) \propto \prod_{i=1}^{n} GEV(\mathbf{R}|\mu,\sigma,\xi) \times N(\mu|0,10^3) \times N(\sigma|0,10^3) \times N(\xi|0,10^3) \qquad (12)$$

The posterior distribution for the parameters of GEV distribution was obtained by maximizing the joint posterior distribution as illustrated in Equation 12, via Markov Chain Monte Carlo (MCMC) algorithm, especially Metropolis-Hastings (MH) sampler. The MH method generates a sequence of random samples from a proposal density function, which subsequently approximate the desired distribution. Here, Gaussian distributions are used as prior distributions and the Markov chain eventually converges to the desired distribution through the rejection-acceptance process. The detailed information on the MCMC method can be found in Van de Vyver (2015). In this study, after 10,000 iterations, the chain is considered to be effectively converged and producing posterior distributions for the estimation of design rainfall over the entire weather stations and the corresponding grid points.

3.3 Spatio-temporal change in design rainfall

To explore the spatio-temporal changes in design rainfall, it is essential to obtain the spatio-temporal data for both the observed and modelled. For ERA-20c, the raw values should be first corrected against

the corresponding observed precipitation over the entire grid points, especially for the ungauged catchments. Conceptually, as described in Section 3.1, the QDM method builds a transfer function based on a one-to-one relationship between the observed and modelled data. However, there are limited observed data available for matching to the corresponding grids of ERA-20c precipitation data. Under such a condition, this study considered an interpolation technique to represent the pointwise values of parameters into the grid for the ungauged catchment, in turn, leading to the use of the interpolated parameters for building the transfer function.

One can consider the direct interpolation of daily precipitation based on the inverse distance weighting (IDW) method or the kriging method, however, there are many dry days in daily precipitation so that the interpolated daily precipitation data could be smoothed out, especially for the extreme. Such systematic errors associated with the spatial interpolation could be propagated through to the parameter estimation process in the QDM approach. Moreover, this direct interpolation cannot be applied without historical records in the early 20th century. For these reasons, we interpolated the estimated parameter over the entire grids, which can create the distribution parameters for the transfer function in parametric QDM approaches in ungauged basins. To implement the QDM approaches based on a composite distribution suggested in Section 3.1, the six parameters (TH, α, β, θ, ξ and u) should be estimated for a pair of the observed and modelled precipitation. Thus, we create the contour maps using the estimated parameters based on a scattered data interpolation method in matlab (Amidror, 2002), and then extract a set of distribution parameters covering the entire range. A flow chart for the proposed QDM procedure is illustrated in Figure 2. To evaluate the effectiveness of an interpolation approach, a leave-one-out cross validation framework is applied for the reference period (1974-2010). Specifically, one station is repeatedly excluded and validated using the estimated set of parameters from the remaining 47 stations. The bias-corrected AMRs were evaluated with regard to RMSE and NSE.

[Insert Figure 2]

With the spatially corrected reanalysis, this study explored the spatio-temporal changes in design rainfalls for a 100-year return period over South Korea. For this purpose, we analyzed the relative change (RC, %) in the observed and modelled design rainfalls for the three periods, 1900-1936, 1937-1973 and 1974-2010, as follows:

$$RC(\%) = \frac{D_r^{obs} - D_p^{sim}}{D_r^{obs}} \times 100 \quad (13)$$

Here, D_r^{obs} represents the design rainfall using the observed AMRs for the reference period (i.e. 1974-2010), while D_p^{sim} indicates the design rainfalls based on the bias corrected AMRs for the three periods. Note that we estimated the design rainfalls by fitting the bias corrected AMRs to GEV distribution, and those of the observed AMRs were obtained by an IDW method, which is commonly used in practice.

In addition, we conducted a retrospective analysis to explore the temporal changes in design rainfall for a given 100-year return period using the bias-corrected century long data. More specifically, the design rainfalls obtained from the bias corrected AMRs for the whole period (1900-2010) were compared with those by the observed for the reference period (1974-2010) over South Korea.

4. Results and Discussion

4.1 Evaluation for the bias corrected ERA-20c

To evaluate the performance of the proposed QDM approach, we collected the bias corrected AMRs and statistically compared them with those of the observed as illustrated in Figure 3. More specifically, Figures 3(a) describes the comparison between the raw ERA-20c and the bias corrected values over all stations for the reference period 1974-2010. The SQM and QDM approaches performed reasonably well, 0.914 and 0.891 for NSE and 18.65mm and 20.93mm for RMSE, respectively, while the raw ERA-20c showed -0.562 for NSE and 79.81 for RMSE. Here, QDM and SQM with the same upper threshold conceptually give the same error for the reference period. For a comparison with the century

long data for 7 stations, a significant reduction in the bias was identified by QDM and SQM, as shown in Figure 3(b). It can be shown that QDM approaches performed slightly better than the corresponding SQMs. For QDM99, the agreement to the observed were 27.11mm for RMSE and 0.824 for NSE, indicating the better performance than SQM99 with 28.11mm for RMSE and 0.810 for NSE. For QDM95, the model efficiency in terms of NSE was comparable with that of the SQM95, but RMSE was slightly smaller than that of the SQM95. These results suggest that QM approaches applied in this study can significantly reduce the bias in daily precipitation for the whole 20th century, and QDM approaches are more efficient than SQM schemes, especially for the AMRs, during the whole 20th century.

[Insert Figure 3]

However, the validation results in three different periods showed that the proposed bias correction scheme has a limitation in reproducing extreme values as shown in Figure 4. As illustrated in Figure 4(a), the AMRs over 7 stations during 1974-2010 are reasonably well reproduced and comparable to that of the observed in Figure 3(a). On the other hand, a relative increase in bias in AMRs is clearly seen in the period, 1937-1973, as shown in Figure 4(b) and 4(c).

[Insert Figure 4]

The bias corrected ERA-20c was significantly overestimated in the upper tail in QDM approaches, as illustrated in Figures 4(b) and 4(c). The large deviations in the top 5% of extremes between the observed (Δ_o) and modelled (Δ_m), which can be estimated from Equation 6, is most likely responsible for the overestimation. Conceptually, the QDM begins with the premise that the relative change in the modelled precipitation over the reference and simulation period is identical to these transformations of the observed. However, the relative changes for a few certain quantiles in the modelled are notably higher than the observed, especially for the high extremes, which can lead to the overestimation identified in Figure 4. More specifically, Figure 5 represents the relative change in a descending order of extreme rainfalls between the reference period (1974-2010) and the past period (1937-1973) for the

observed and raw ERA-20c. The relative changes generally showed a similar trend with a ratio around 1 in both the observed and modelled, but the large deviations are clearly identified for high extremes. For example, the relative change at St.17 Daegu station is about 1.3 for the modelled, while the value for the observed was less than 1. Under the assumption of QDM approaches, the bias corrected data for the simulation period is increased by 1.3, leading to the increased deviation in AMRs between the in-situ and modelled data. Apart from the misrepresentation of high extremes, other aspects could also influence the differences. The significant inconsistency in long-term trend, especially for the extreme in the first half of the 20th century, could also result in the bias (Befort *et al.*, 2016; Donat *et al.*, 2016).

[Insert Figure 5]

4.2 Uncertainty Reduction in Design Rainfall using ERA-20c

Although the suggested QM approaches still have the biases in the high extremes, the bias-corrected AMRs showed a significant reduction in the systematic bias and comparable results across three different periods. We explore changes in design rainfall and their uncertainties in the context of a century precipitation data. In many countries, the estimation of design rainfall is based on AMRs collected over a relatively short period of time that can lead to high uncertainty in estimating parameters for a given distribution. For this purpose, we evaluated the uncertainties of design rainfall with different return levels (i.e. 30-year, 50-year and 100-year return period) for both the observed and the bias corrected ERA-20c by QDM approaches (i.e. QDM95 and QDM99) over 48 stations. Note that the uncertainties derived from the bias corrected AMRs for the reference period (1974-2010) were named as QDM95v1 and QDM99v1, respectively, and the uncertainties using the values from 1900 to 2010 were named as QDM95v0 and QDM99v0, respectively. The uncertainty range of design rainfall for six stations (i.e. in St. 5, St. 13, St. 21. St. 29, St. 37 and St. 43) for a representative experiment is illustrated in Figure 6, and the results for the remaining stations can be found in Appendix B. For the reference period, the median values of design rainfalls obtained from the bias corrected ERA-20c are

comparable to those of the observed while their uncertainty range is largely extended, except for Busan and Gumi stations. As seen in Figure 6, design rainfalls by the observed also have large uncertainties for the reference period. On the other hand, the uncertainty range of design rainfall using a century precipitation data (i.e. QDM95v0 and QDM99v0) is much narrower than that for the reference period (i.e. QDM95v1 and QDM99v1). It is logical to assume that the uncertainty reduction in design rainfall is mainly attributed to the increase in sample size. Thus, the long-term bias corrected rainfall has its own advantage in terms of the increase of the sample size, leading to the uncertainty reduction in design rainfall.

[Insert Figure 6]

The increase in the uncertainty of design rainfall may be attributed to the GEV parameters, especially for the shape parameter, and this study is assuming that the associated uncertainty could be reduced by the use of long-term data. In this regard, we further explored the role of the bias corrected rainfall for uncertainty reduction in design rainfall in the context of prior information within a Bayesian framework. More specifically, the range of GEV parameters estimated from the bias corrected century long reanalysis data (i.e. QDM95v0 and QDM99v0) is considered as the prior distribution for the estimation of the distribution parameters within a Bayesian framework. We examined the role of the prior distribution informed by the bias corrected long-term reanalysis data in the uncertainty reduction in design rainfall. Three different cases with regard to the use of prior distributions were considered; (1) the sole use of prior distribution for shape parameter, (2) the use of prior distributions for both scale and location parameters, and (3) the combined use of prior distributions for all three parameters within the QDMs. The first experiments with QDM95v0 and QDM99v0 were named as Obs95a and Obs99a, respectively. The second experiments were named as Obs95b and Obs99b, and the final experiments were named as Obs95c and Obs99c, respectively. The comparison of three different experiments for the uncertainty reduction in design rainfall is illustrated in Figure 7. As shown in Figure 7, the median values are comparable over all cases presented here but a significant shrinkage of the uncertainty range

is seen in most cases where informative priors are considered. Among three approaches, the combined use of prior distribution for all the parameters (experiment 3) showed the greater reduction in the uncertainty than either the experiments 1 or 2. More specifically, for St.5 Seoul, St.21 Busan and St. 45 Gumi where the uncertainty range of design rainfall is exceptionally high, the degree of reduction in uncertainty for the experiment 1 (sole use of prior distribution for the shape parameter) is closely followed to that of the experiment 3, while experiment 2 (use of prior distributions for the location and shape parameters) still has large uncertainty. On the other hand, the reduction of uncertainty in other three stations, St.13, St.29 and St.37, for the experiment 1 is nearly similar to that of the experiment 2. In these contexts, we can conclude that the uncertainty reduction in the design rainfall can be mainly attributed to the use of prior distribution for the shape parameter, informed by the long-term reanalysis data.

[Insert Figure 7]

4.3 Spatio-temporal Change in Design Rainfall

To begin with, we evaluated the spatial interpolation approach for ERA-20c within a leave-one-out cross validation scheme. The bias-corrected AMRs for the reference period were compared with the corresponding observation over 48 stations in terms of RMSE and NSE. In Figure 8, the result indicated that the bias-corrected AMRs were generally comparable to those from the observed, although the slightly increased bias were observed, compared to the individual correction as illustrated in Figure 3(a). The result implies that the bias correction scheme based on a set of interpolated parameters informed by the observed parameters could reliably provide spatially interpolated long-term data, especially for exploring changes in design rainfall. It should be noted that QDM and SQM have the same results for the reference period, and QM approaches with the 95^{th} and 99^{th} percentiles are labelled as QM95 and QM99 in figure 8.

[Insert Figure 8]

For the exploration of spatio-temporal change in rainfall intensity over the 20th century, this study compared design rainfalls using the bias corrected ERA-20c over the entire areas for a given 100-year return period, with those of the observation for the reference period (1974-2010). The relative change (%) of design rainfalls between the observed and the modelled for three different periods (1900-1936, 1937-1973 and 1974-2010) is illustrated in Figure 9. For the reference period in Figure 9(a), the relative difference is generally limited within 10% in both QDM95 and QDM99 approaches, although a slight increase in bias of particular areas is observed. The QDM99 shows better performance although the difference between the two approaches is not significant. Thus, the results based on the QDM99 is mainly considered for subsequent analysis on the spatial-temporal change in rainfall intensity over South Korea.

For the period from 1937 to 1973 as shown in Figure 9(b), changes in the design rainfall vary spatially, and a noticeable change is observed in the southwest, suggesting a significant decrease in the rainfall intensity over the last three decades. On the other hand, an insignificant change in the rainfall intensity is identified in the western and eastern parts of South Korea for the period. Under the circumstances, the estimated design rainfall during that period may not be appropriate in a changing climate even if the bias in the ERA-20c is considered. For the period 1900-1936, changes in rainfall intensity relative to the current climate is illustrated in Figure 9(c), and they reveal a similar pattern over the entire areas, representing a noticeable increase in rainfall intensity over the last three decades. It can be concluded that the significant changes in rainfall intensity over different periods can lead to a misrepresentation of the design rainfall (or design flood) and are likely to misrepresent the flood risk, particularly at high return levels in the future.

Unlike the results presented in Figures 9(a) to 9(c), the AMRs over the entire period are used to further explore the changes in design rainfall, with that of the observed for the reference period (1974-2010), in Figure 9(d). As in the reference period in Figure 9(a), the relative changes in this period similarly limited within 10%. However, the spatial distribution of relative change is slightly different from the

changes based on the reference period from 1974 to 2010. More specifically, a positive change is more pronounced in the northern part of South Korea, confirming that its role of the recent increase in the rainfall intensity, while negative change is still remained in the south-western part of South Korea. This result implies that design rainfall estimated during that period can be significantly underestimated, leading to a potential increase in the flood risk in view of the present in most areas.

[Insert Figure 9]

Estimating design rainfall in a certain area plays an important role in managing risk associated with water-related hazards. In many countries including South Korea, design rainfall is routinely estimated with limited data for a given return period exceeding the length of the data record for planning structural or non-structural measures. In this context, the estimated design rainfall can be significantly influenced by sampling error, leading to an increase in uncertainty. Numerous studies have shown that temporal change in extreme rainfall has been observed, and especially an increasing trend has been reported in many parts of the world (Mason *et al.*, 1999; Jung *et al.*, 2011; Park *et al.*, 2011; Westra *et al.*, 2014; Yilmaz *et al.*, 2014). This non-stationarity in extreme rainfall is expected to be larger in the future and the current design practice may not be appropriate under the condition. For this reason, the guideline recommendations considering the potential impact of the nonstationarity on either design rainfall or design flood have been proposed by various studies (Lawrence and Hisdal, 2011; Madsen *et al.*, 2014; Environment Agency, 2017). These guidelines typically recommend to employ a correction factor, corresponding to the expected change, in the estimation of design rainfall and design flood (Madsen *et al.*, 2014). For instance, the Environment Agency (2017) in UK recommended to increase the rainfall intensity from 5 % to 40% for the future period (2015-2115) over the entire England and Wales. In Norway, a wider range of correction factors (i.e. 0-40%) were similarly recommended for design flood (Lawrence and Hisdal, 2011). This approach may help to reduce the flood risk for a given region in a changing climate. However, in consideration of the wider range of uncertainty in design rainfall with the limited data as illustrated in Figures 6 and A2, the use of longer

data plays a crucial role in the reliable estimation of design rainfall with the uncertainty reduction.

5. Concluding remarks

The objective of this study was to examine the ERA-20c reanalysis data and its use for assessing long term changes in design rainfall over South Korea, by extending the AMRs through the bias correction of the reanalysis data. In this context, we first applied the stationary and non-stationary QM approaches using a composite distribution, referred to as QDM and SQM, to reduce the biases. More specifically, this study evaluated not only the accuracy of the bias corrected AMRs but also their use for the uncertainty reduction in design rainfalls within the Bayesian framework. Finally, we explored the spatio-temporal changes in design rainfalls. The major findings obtained in this study are summarized as follows:

1. QM approaches (i.e. SQM and QDM) are significantly effective in reducing the bias of daily precipitation from the ERA-20c reanalysis data, and QDM approaches are more efficient than SQM schemes for the whole 20th century, especially for the AMRs. On the other hand, the validation results over different periods showed that the proposed bias correction scheme has a limitation in reproducing extreme values. To be more specific, the AMRs during 1974-2010 are reasonably well reproduced and comparable to that of the observed, however, a relative increase in bias in AMRs is clearly observed in the period, 1937-1973. It can be concluded that the increase in bias in that period is attributed to the large deviations for high extremes (i.e. top 5 events). Additionally, an inconsistency in long-term trend, especially for the extreme in the first half of the 20th century, could also result in the bias.
2. This study evaluated the uncertainties of design rainfalls with different return levels for both the observed and the bias corrected ERA-20c by QDM approaches. The uncertainty range of design rainfall using a century precipitation data is much narrower than that for the reference period due to the increase in sample size. Thus, the long-term bias corrected rainfall has their own

advantage in terms of the increase of the sample size, leading to the uncertainty reduction in design rainfall. We further explored the role of the bias corrected rainfall for uncertainty reduction in design rainfall via three different experiments in the context of prior information within a Bayesian framework. A significant shrinkage of the uncertainty range is seen in all the cases where informative priors are considered. In the experimental study, we can conclude that the uncertainty reduction in the design rainfall can be mainly attributed to the use of prior distribution for the shape parameter, informed by the long-term reanalysis data.

3. There were significant changes in design intensity according to the periods (1900-1936, 1937-1973 and 1974-2010), which can lead to a misrepresentation of the flood risk, particularly at high return levels in the future. Design rainfall change derived from the bias-corrected AMRs from 1900 to 2010 suggested that the recent increase in the rainfall intensity should be considered in managing risk associated with water-related hazards.
4. This study finally compared design rainfalls of using the bias corrected ERA-20c over the entire areas for a given 100-year return period, with those of the observation for the reference period (1974-2010). The spatial distribution of relative change using the AMRs over the entire period is different from the changes based on the reference period from 1974 to 2010. More specifically, a positive change is more pronounced in the northern part of South Korea, confirming that its role of the recent increase in the rainfall intensity. This result implies that design rainfall estimated during that period can be significantly underestimated, leading to a potential increase in the flood risk in view of the present in most areas.

The findings obtained in this study provide a meaningful perspective on the use of long-term reanalysis data for the uncertainty reduction in design rainfall. Further, this study helps to better understand the long-term changes in rainfall intensity over the past century in South Korea. Although the study has been performed only in South Korea, we hope this study will stimulate the hydro-meteorological community to explore the issues raised in the long-term reanalysis data in other countries under

different climate and geographical conditions.

Acknowledgements

The first author is funded by the Government of South Korea for performing his doctoral studies at the University of Bristol. We are grateful for the relevant data provided by KMA and ECMWF. The second author is supported by a grant (17AWMP-B121100-02) from Advanced Water Management Research Program (AWMP) funded by Ministry of Land, Infrastructure and Transport of Korean government.
Abbreviations and symbols used in this study are listed in Appendix C and D.

Appendix A. Selection of the Upper Thresholds

As described in Section 3.1, the QM approaches used in this study are based on a composite distribution of a gamma and GPD. Thus, determining the upper threshold is an essential part in this parametric approach. Previous papers have commonly applied high thresholds such as 95th or 99th percentile to the bias correction because the distribution over high values is asymptotically fitted to a GPD (Manton *et al.*, 2001; Wilson and Toumi, 2005; Acero *et al.*, 2011; Gutjahr and Heinemann, 2013; Chan *et al.*, 2015; Nyunt *et al.*, 2016; Kim *et al.*, 2018). This approach is based on the assumption that the upper thresholds for the observed and modelled are identical, although the distribution of each dataset can be different. In this context, we considered three different thresholds, the 90th, 95th and 99th percentile, and applied a mixture of the thresholds for the observed and modelled, respectively. For example, the CDF of the observed with the 90th percentile was matched to the modelled with the 90th, 95th and 99th percentile, respectively. Likewise, the nine different sets of thresholds were evaluated to determine the optimal threshold pair in 48 stations for the reference period (1974-2010) based on the QM schemes described in Section 3.1. The abbreviations for QM approaches with different thresholds are described in Table A1.

Table A1. Abbreviations for QM approaches depending on the thresholds

ID		ERA-20c		
		90th	95th	99th
OBS	90th	QM9090	QM9095	QM9099
	95th	QM9590	QM9595	QM9599
	99th	QM9990	QM9995	QM9999

Figure A1 illustrates the comparison of the AMRs between the observation in 48 stations and the bias-corrected ERA-20c in the corresponding grid points from 1974 to 2010. In Figure A1, the algorithm with a same pair of thresholds, especially QM9595 and QM9999, performed better than the other approaches. Interestingly, with the same threshold for the observed, the QM scheme with the lower modelled threshold like QM9099 underestimated the extremes, while the opposite case such as

QM9099 had relatively overestimated values. This result implies that how to set the upper threshold in the QM algorithm applied in this study may significantly affect the reliability of the bias-corrected value, especially for the extreme. Thus, we apply the 95th or 99th percentile pair as upper thresholds to the bias correction process.

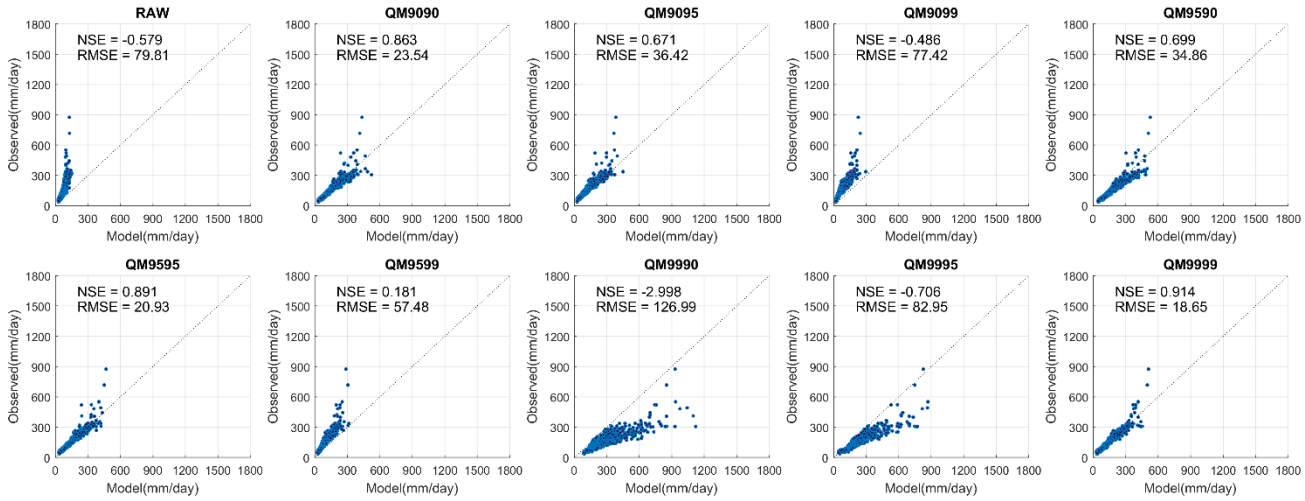
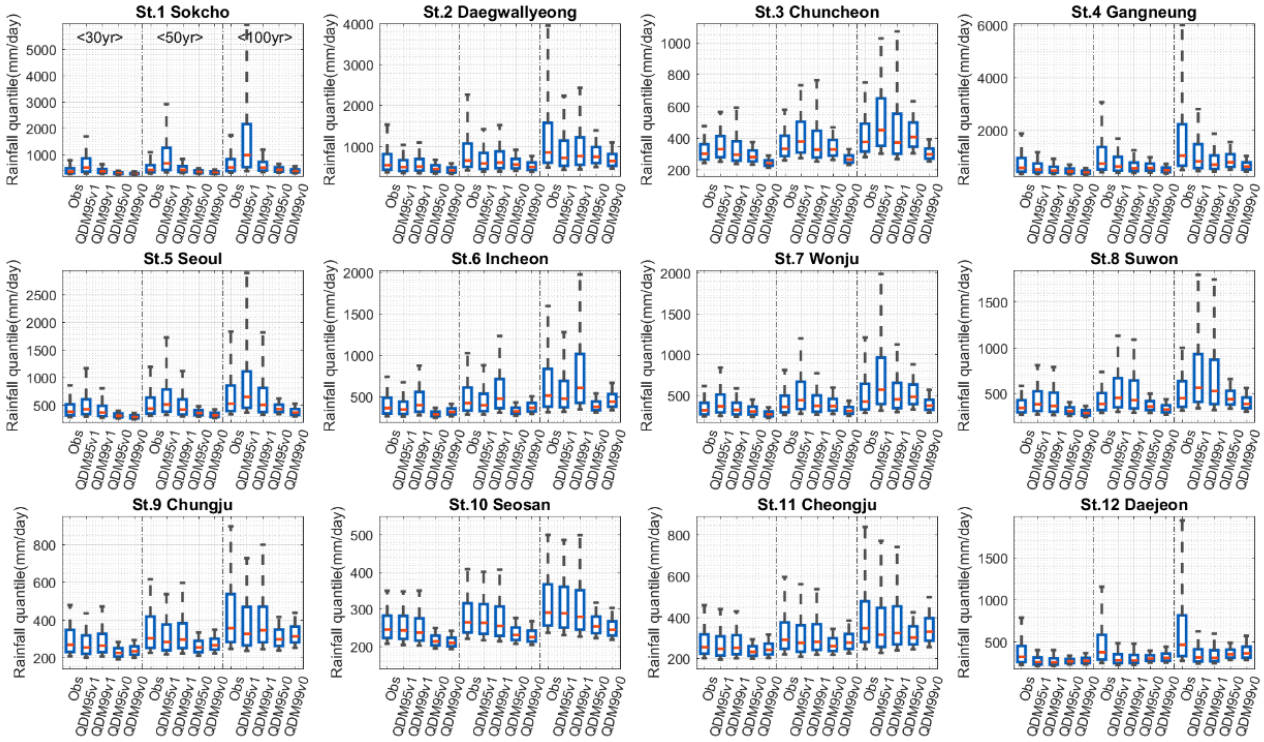
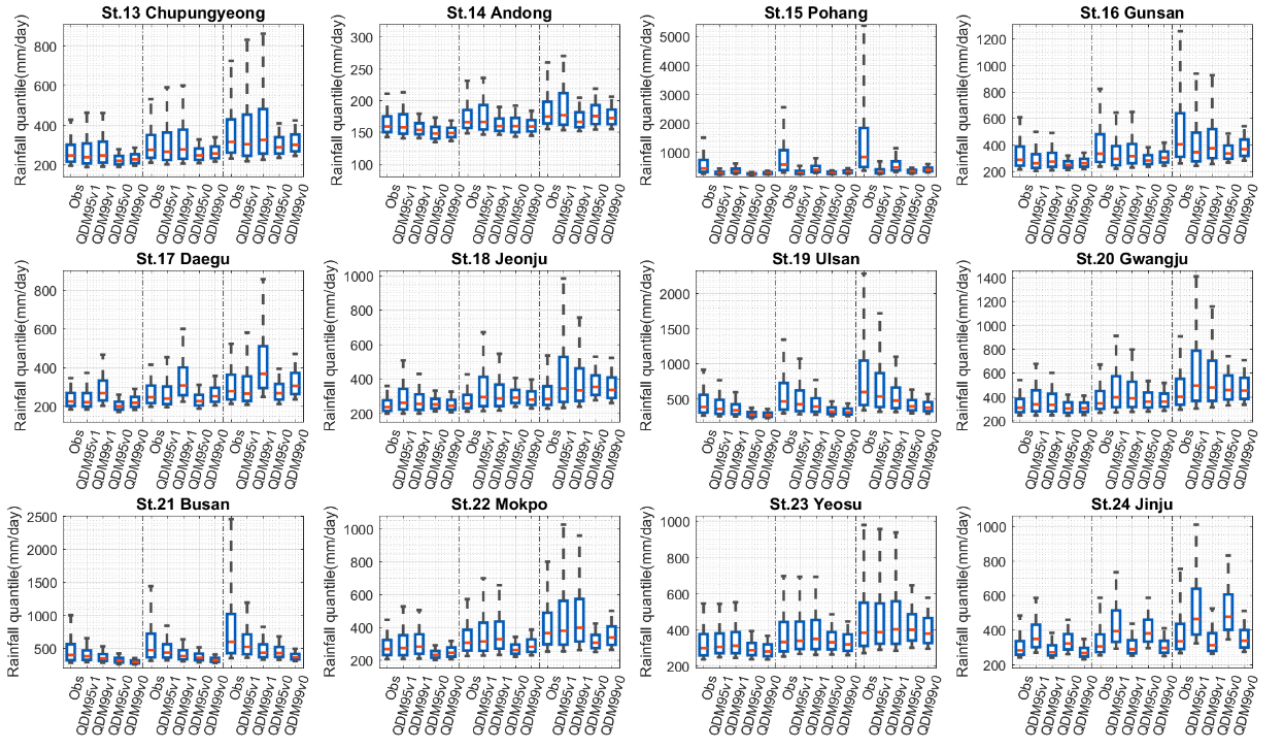
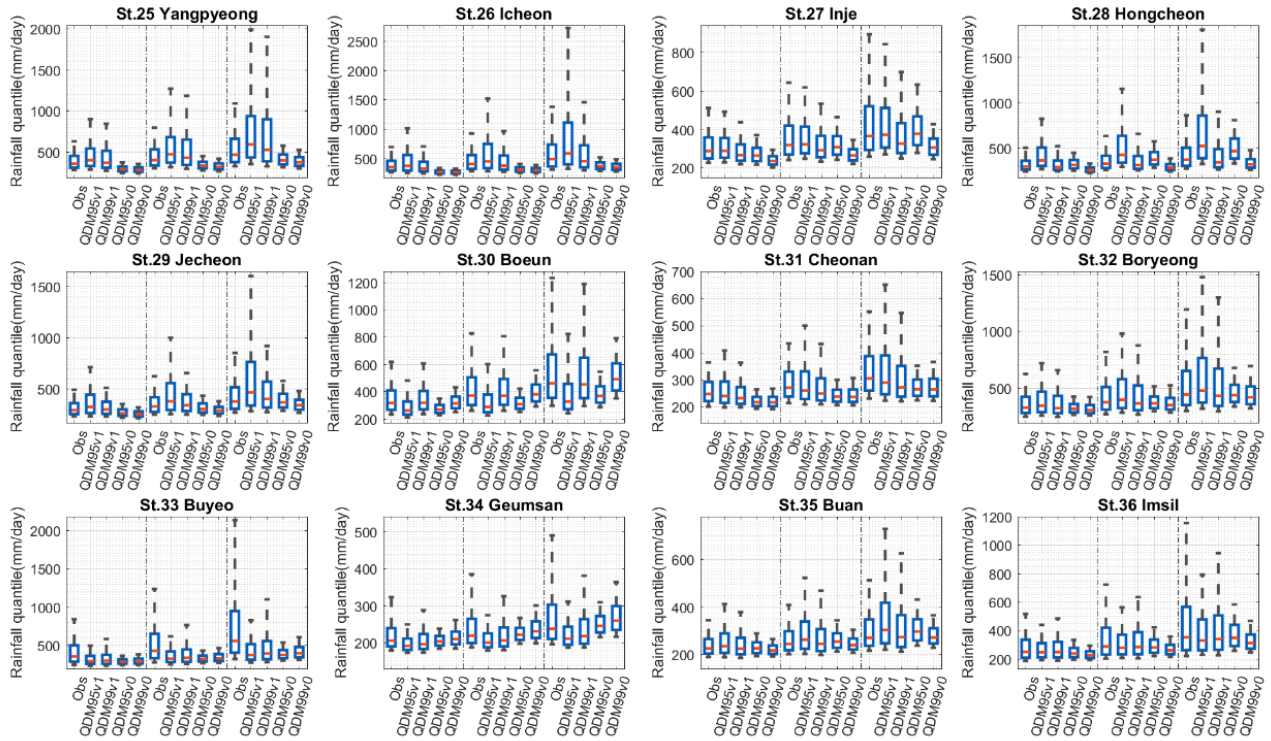


Fig A1. Scatter plot between the Annual maximum rainfalls of the observation and modelled (the raw ERA-20c(RAW) and the bias-corrected valued by QM approaches) for the reference period (1974-2010) in 48 stations

Appendix B. Uncertainty ranges of design rainfalls in 48 stations







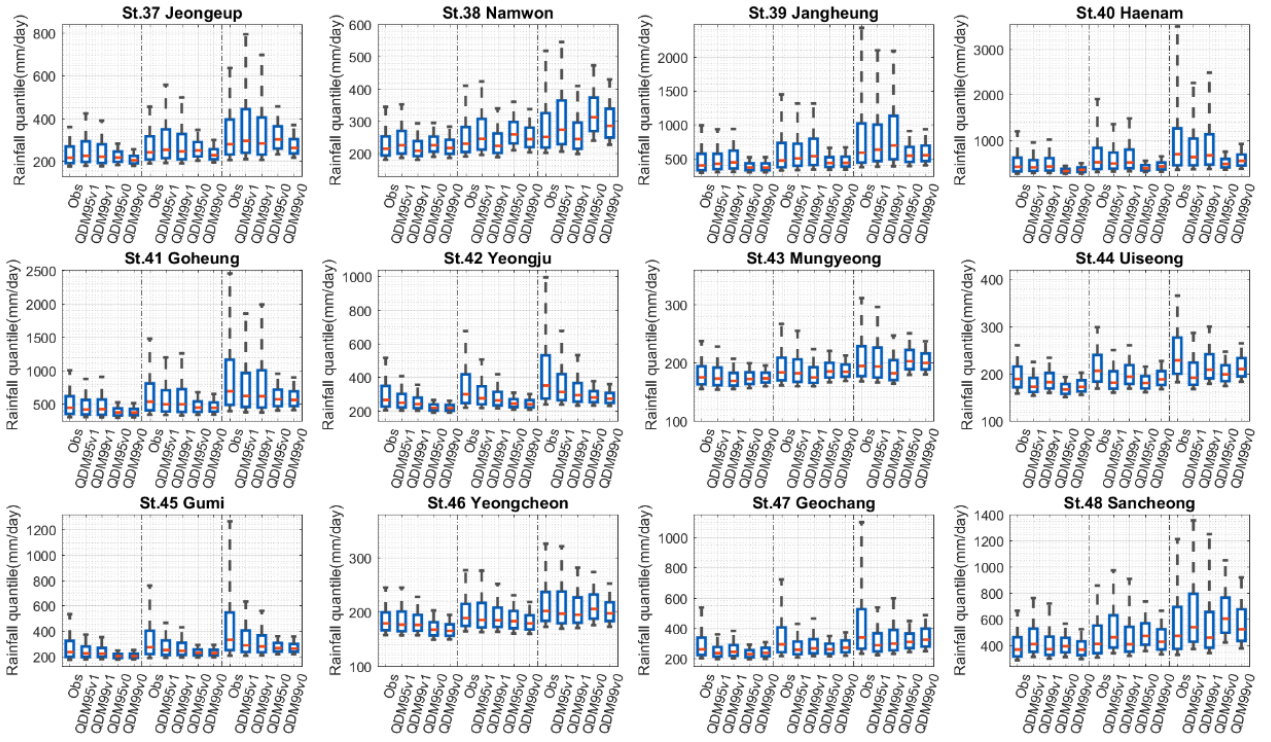


Figure A2. Comparison of the uncertainties for design rainfalls with different return levels (i.e. 30-year, 50-year and 100-year return period) for both the observed (Obs) and the bias corrected ERA-20c by QDM approaches over 48 stations. Here, QDM95v1 and QDM99v1 represent the values estimated for the reference period (i.e. 1974-2010) while QDM95v0 and QDM99v0 are derived from 1900 to 2010.

Appendix C. List of Abbreviations

ID	Definitions
AMR	Annual maximum rainfall
CDF	Cumulative distribution functions
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA-20c	ECWMF's 20th century reanalysis assimilated by surface observations only
ERA-20cm	ECMWF's 20th century atmospheric model ensemble
GEV	Generalized extreme value distribution
GPD	Generalized Pareto distribution
IDW	Inverse distance weighting method
KMA	Kora Meteorological Administration
NOAA	National Oceanic and Atmospheric Administration
Obs95a/ Obs99a	Uncertainty of the observation-based design rainfall with the prior distribution for shape parameter, informed by the bias-corrected values based on QDM95/QDM99
Obs95b/ Obs99b	Uncertainty of the observation-based design rainfall with the prior distributions for location and shape parameters, informed by the bias-corrected values based on QDM95/QDM99
Obs95c/ Obs99c	Uncertainty of the observation-based design rainfall with the prior distributions for all parameters, informed by the bias-corrected values based on QDM95/QDM99
NSE	Nash-Sutcliffe efficiency
RMSE	Root mean square error
QM	Quantile mapping
QDM	Quantile delta mapping
QDM-GP	QDM method of combining a composite Gamma-Pareto distribution
QDM95/ QDM99	QDM with the upper tail of 95th/99th percentile
QDM95v1/ QDM99v1	Design rainfall by using the bias-corrected AMRs based on QDM95/QDM99 for the reference period (1974-2010)
QDM95v0/ QDM99v0	Design rainfall by using the bias-corrected AMRs based on QDM95/QDM99 for the whole data period (1900-2010)
SQM	Quantile mapping approach based on the stationary assumption
SQM95/ SQM99	SQM with the upper tail of 95th/99th percentile
SQM95v1/ SQM99v1	Design rainfall by using the bias-corrected AMRs based on SQM95/SQM99 for the reference period (1974-2010)
SQM95v0/ SQM99v0	Design rainfall by using the bias-corrected AMRs based on SQM95/SQM99 for the whole data period (1900-2010)
TH	Cut-off threshold for a composite distribution
20CR	The 20th century reanalysis by the NOAA

Appendix D. List of Symbols

ID	Definitions
α	shape parameter of a gamma distribution
β	scale parameter of a gamma distribution
ξ	Shape parameter of a GPD
θ	Scale parameter of a GPD
u	High upper threshold for a GPD

References

Acero FJ, García JA, Gallego MC. 2011. Peaks-over-threshold study of trends in extreme rainfall over the Iberian Peninsula. *Journal of Climate* **24** (4): 1089–1105 DOI: 10.1175/2010JCLI3627.1

Amidror I. 2002. Scattered data interpolation methods for electronic imaging systems: a survey. *Journal of Electronic Imaging* **11** (2): 157 DOI: 10.1117/1.1455013

Befort DJ, Wild S, Kruschke T, Ulbrich U, Leckebusch GC. 2016. Different long-term trends of extra-tropical cyclones and windstorms in ERA-20C and NOAA-20CR reanalyses. *Atmospheric Science Letters* **17** (11): 586–595 DOI: 10.1002/asl.694

Brands S, Gutiérrez JM, Herrera S, Cofiño AS. 2012. On the use of reanalysis data for downscaling. *Journal of Climate* **25** (7): 2517–2526 DOI: 10.1175/JCLI-D-11-00251.1

Bürger G, Sobie SR, Cannon AJ, Werner AT, Murdock TQ. 2013. Downscaling extremes: An intercomparison of multiple methods for future climate. *Journal of Climate* **26** (10): 3429–3449 DOI: 10.1175/JCLI-D-12-00249.1

Cannon AJ, Sobie SR, Murdock TQ. 2015. Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *Journal of Climate* **28** (17): 6938–6959

Chan SC, Kendon EJ, Roberts NM, Fowler HJ, Blenkinsop S. 2015. Downturn in scaling of UK extreme rainfall with temperature for future hottest days. *Nature Geoscience* **9** (1): 24–28 DOI: 10.1038/ngeo2596

Chang H, Kwon W-T. 2007. Spatial variations of summer precipitation trends in South Korea, 1973–2005. *Environmental Research Letters* **2** (4): 45012

Choi G, Collins D, Ren G, Trewin B, Baldi M, Fukuda Y, Afzaal M, Pianmana T, Gomboluudev P, Huong PTT. 2009. Changes in means and extreme events of temperature and precipitation in the Asia-Pacific Network region, 1955–2007. *International Journal of Climatology* **29** (13): 1906–1925

Chung YS, Yoon MB. 2000. Interpretation of recent temperature and precipitationtrends observed in Korea. *Theoretical and Applied Climatology* **67** (3–4): 171–180

Coles S, Pericchi LR, Sisson S. 2003. A fully probabilistic approach to extreme rainfall modeling. *Journal of Hydrology* **273** (1–4): 35–50 DOI: 10.1016/S0022-1694(02)00353-0

Coles SG. 2001. *An introduction to Statistical Modeling of Extreme Values*. Springer: London. DOI: 10.1007/978-1-4471-3675-0

Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, Gleason BE, Vose RS, Rutledge G, Bessemoulin P. 2011. The twentieth century reanalysis project. *Quarterly Journal of*

the royal meteorological society **137** (654): 1–28

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P. 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society* **137** (656): 553–597

Donat MG, Alexander L V, Herold N, Dittus AJ. 2016. Temperature and precipitation extremes in century-long gridded observations, reanalyses, and atmospheric model simulations. *Journal of Geophysical Research: Atmospheres* **121** (19)

Environment Agency. 2017. Flood risk assessments: climate change allowances. *Department for Enviroment, Food and Rural Affairs* Available at: https://www.gov.uk/guidance/flood-risk-assessments-climate-change-allowances [Accessed 25 June 2018]

Eum H Il, Cannon AJ. 2017. Intercomparison of projected changes in climate extremes for South Korea: application of trend preserving statistical downscaling methods to the CMIP5 ensemble. *International Journal of Climatology* **37** (8): 3381–3397 DOI: 10.1002/joc.4924

Fang G, Yang J, Chen YN, Zammit C. 2015. Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in China. *Hydrology and Earth System Sciences* **19** (6): 2547–2559

Gao L, Bernhardt M, Schulz K, Chen XW, Chen Y, Liu MB. 2016. A First Evaluation of ERA-20CM over China. *Monthly Weather Review* **144** (1): 45–57 DOI: 10.1175/Mwr-D-15-0195.1

Gutjahr O, Heinemann G. 2013. Comparing precipitation bias correction methods for high-resolution regional climate simulations using COSMO-CLM. *Theoretical and Applied Climatology* **114** (3): 511–529 DOI: 10.1007/s00704-013-0834-z

Hersbach H, Peubey C, Simmons A, Berrisford P, Poli P, Dee D. 2015. ERA-20CM: a twentieth-century atmospheric model ensemble. *Quarterly Journal of the royal meteorological society* **141** (691): 2350–2375

Ho C, Lee J, Ahn M, Lee H. 2003. A sudden change in summer rainfall characteristics in Korea during the late 1970s. *International Journal of Climatology* **23** (1): 117–128

Huard D, Mailhot A, Duchesne S. 2010. Bayesian estimation of intensity-duration-frequency curves and of the return period associated to a given rainfall event. *Stochastic Environmental Research and Risk Assessment* **24** (3): 337–347 DOI: 10.1007/s00477-009-0323-1

Jung HS, Lim GH, Oh JH. 2001. Interpretation of the transient variations in the time series of precipitation amounts in Seoul, Korea. Part I: Diurnal variation. *Journal of Climate* **14** (13): 2989–3004 DOI: 10.1175/1520-0442(2001)014<2989:IOTTVI>2.0.CO;2

Jung IW, Bae DH, Kim G. 2011. Recent trends of mean and extreme precipitation in Korea. *International Journal of Climatology* **31** (3): 359–370 DOI: 10.1002/joc.2068

Kim D, Kwon H, Han D. 2018. Exploring the Long-Term Reanalysis of Precipitation and the Contribution of Bias Correction to the Reduction of Uncertainty over South Korea : A Composite Gamma-Pareto Distribution Approach to the Bias Correction. *Hydrology and Earth System Sciences Discussions* **36** (February): 1–53

Kim D-I, Han D. 2018. Comparative study on long term climate data sources over South Korea. *Journal of Water and Climate Change* **(in press)** DOI: 10.2166/wcc.2018.032

Kim KB, Bray M, Han D. 2015a. An improved bias correction scheme based on comparative precipitation characteristics. **2266** (October 2014): 2258–2266 DOI: 10.1002/hyp.10366

Kim KB, Kwon HH, Han D. 2015b. Bias correction methods for regional climate model simulations considering the distributional parametric uncertainty underlying the observations. *Journal of Hydrology* **530**: 568–579 DOI: 10.1016/j.jhydrol.2015.10.015

Krueger O, Schenk F, Feser F, Weisse R. 2013. Inconsistencies between long-term trends in storminess derived from the 20CR reanalysis and observations. *Journal of Climate* **26** (3): 868–874 DOI: 10.1175/JCLI-D-12-00309.1

Kwon H-H, Brown C, Lall U. 2008. Climate informed flood frequency analysis and prediction in Montana using hierarchical Bayesian modeling. *Geophysical Research Letters* **35** (5): L05404 DOI: 10.1029/2007GL032220

Kwon HH, Sivakumar B, Moon Y Il, Kim BS. 2011. Assessment of change in design flood frequency under climate change using a multivariate downscaling model and a precipitation-runoff model. *Stochastic Environmental Research and Risk Assessment* **25** (4): 567–581 DOI: 10.1007/s00477-010-0422-z

Lawrence D, Hisdal H. 2011. *Hydrological projections for floods in Norway under a future climate*. Norwegian Water Resources and Energy Directorate.

Li H, Sheffield J, Wood EF. 2010. Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *Journal of Geophysical Research: Atmospheres* **115** (D10)

Li J, Evans J, Johnson F, Sharma A. 2017. A comparison of methods for estimating climate change impact on design rainfall using a high-resolution RCM. *Journal of Hydrology* **547**: 413–427 DOI: 10.1016/j.jhydrol.2017.02.019

Madsen H, Lawrence D, Lang M, Martinkova M, Kjeldsen TR. 2014. Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. *Journal of Hydrology* **519** (PD): 3634–3650 DOI: 10.1016/j.jhydrol.2014.11.003

Manton MJ, Haylock MR, Hennessy KJ, Nicholls N, Chambers LE, Collins DA, Daw G, Finet A, Gunawan D, Inape K, et al. 2001. Trends in Extreme Daily Rainfall and Temperature in Southeast Asia and the South Pacific : 1961 – 1998. *International Journal of Climatology* **21**: 269–284 DOI: 10.1002/joc.610

Maraun D. 2016. Bias Correcting Climate Change Simulations - a Critical Review. *Current Climate Change Reports* **2** (4): 211–220 DOI: 10.1007/s40641-016-0050-x

Maraun D, Widmann M. 2018. *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge University Press.

Mason S, Waylen P, Mimmack G, Rajaratnam B, Harrison M. 1999. Changes in Extreme Rainfall Events in South Africa Simon. *Clmatic Change* **41**: 249–257

Miao C, Su L, Sun Q, Duan Q. 2016. A nonstationary bias-correction technique to remove bias in GCM simulations. *Journal of Geophysical Research* **121** (10): 5718–5735 DOI: 10.1002/2015JD024159

Nahar J, Johnson F, Sharma A. 2017. Assessing the extent of non-stationary biases in GCMs. *Journal of Hydrology* **549**: 148–162 DOI: 10.1016/j.jhydrol.2017.03.045

Nyunt CT, Koike T, Yamamoto A. 2016. Statistical bias correction for climate change impact on the basin scale precipitation in Sri Lanka , Philippines , Japan and Tunisia. (January) DOI: 10.5194/hess-2016-14

Overeem A, Buishand A, Holleman I. 2008. Rainfall depth-duration-frequency curves and their uncertainties. *Journal of Hydrology* **348** (1–2): 124–134 DOI: 10.1016/j.jhydrol.2007.09.044

Park JS, Kang HS, Lee YS, Kim MK. 2011. Changes in the extreme daily rainfall in South Korea. *International Journal of Climatology* **31** (15): 2290–2299 DOI: 10.1002/joc.2236

Piani C, Haerter JO, Coppola E. 2010. Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoretical and Applied Climatology* **99** (1): 187–192 DOI: 10.1007/s00704-009-0134-9

Poli P, Hersbach H, Dee DP, Berrisford P, Simmons AJ, Vitart F, Laloyaux P, Tan DGH, Peubey C, Thépaut J-N. 2016. ERA-20C: An Atmospheric Reanalysis of the Twentieth Century. *Journal of Climate* **29** (11): 4083–4097

Poli P, Hersbach H, Tan D, Dee D, Thépaut J-N, Simmons A, Peubey C, Laloyaux P, Komori T, Berrisford P, et al. 2013. The data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C)

Rabiei E, Haberlandt U. 2015. Applying bias correction for merging rain gauge and radar data. *Journal of Hydrology* **522**: 544–557

Reis DS, Stedinger JR. 2005. Bayesian MCMC flood frequency analysis with historical information.

Journal of Hydrology **313** (1–2): 97–116 DOI: 10.1016/j.jhydrol.2005.02.028

Schmidli J, Frei C, Vidale PL. 2006. Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods. *International Journal of Climatology* **26** (5): 679–689

So BJ, Kwon HH, Kim D, Lee SO. 2015. Modeling of daily rainfall sequence and extremes based on a semiparametric Pareto tail approach at multiple locations. *Journal of Hydrology* **529**: 1442–1450 DOI: 10.1016/j.jhydrol.2015.08.037

Teutschbein C, Seibert J. 2012. Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology* **456**: 12–29

Themeßl MJ, Gobiet A, Leuprecht A. 2011. Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *International Journal of Climatology* **31** (10): 1530–1544 DOI: 10.1002/joc.2168

Tung Y koung, Wong C leung. 2014. Assessment of design rainfall uncertainty for hydrologic engineering applications in Hong Kong. *Stochastic Environmental Research and Risk Assessment* **28** (3): 583–592 DOI: 10.1007/s00477-013-0774-2

Volosciuk C, Maraun D, Vrac M, Widmann M. 2017. A combined statistical bias correction and stochastic downscaling method for precipitation. *Hydrology and Earth System Sciences* **21** (3): 1693–1719 DOI: 10.5194/hess-21-1693-2017

Vrac M, Naveau P. 2007. Stochastic downscaling of precipitation : From dry events to heavy rainfalls. **43** (July): 1–13 DOI: 10.1029/2006WR005308

Van de Vyver H. 2015. Bayesian estimation of rainfall intensity-duration-frequency relationships. *Journal of Hydrology* **529**: 1451–1463 DOI: 10.1016/j.jhydrol.2015.08.036

Westra S, Fowler HJ, Evans JP, Alexander L V, Berg P, Johnson F, Kendon EJ, Lenderink G, Roberts NM. 2014. Future changes to the intensity and frequency of short-duration extreme rainfall. *Reviews of Geophysics* **52** (3): 522–555 DOI: 10.1002/2014RG000464

Wilson PS, Toumi R. 2005. A fundamental probability distribution for heavy rainfall. *Geophysical Research Letters* **32** (14): 1–4 DOI: 10.1029/2005GL022465

Yilmaz AG, Hossain I, Perera BJC. 2014. Effect of climate change and variability on extreme rainfall intensity-frequency-duration relationships: A case study of Melbourne. *Hydrology and Earth System Sciences* **18** (10): 4065–4076 DOI: 10.5194/hess-18-4065-2014

Zhang Q, Körnich H, Holmgren K. 2013. How well do reanalyses represent the southern African precipitation? *Climate Dynamics* **40** (3–4): 951–962 DOI: 10.1007/s00382-012-1423-z

Tables

Table 1. The local rainfall stations used in this study

Station No.	Name	Latitude (°N)	Longitude (°E)	Elevation (m. asl)	Data period
St. 1	Sokcho	38.2508	128.5644	19.5	1974-2010
St. 2	Daegwallyeong	37.6769	128.7181	774.0	1974-2010
St. 3	Chuncheon	37.9025	127.7356	79.1	1974-2010
St. 4	Gangneung	37.7514	128.8908	27.4	**1912**-2010
St. 5	Seoul	37.5714	126.9656	11.1	**1910**-2010
St. 6	Incheon	37.4775	126.6247	69.6	**1910**-2010
St. 7	Wonju	37.3375	127.9464	150.0	1974-2010
St. 8	Suwon	37.2700	126.9875	38.3	1974-2010
St. 9	Chungju	36.9700	127.9525	116.5	1974-2010
St. 10	Seosan	36.7736	126.4958	30.3	1974-2010
St. 11	Cheongju	36.6361	127.4428	58.6	1974-2010
St. 12	Daejeon	36.3689	127.3742	70.3	1974-2010
St. 13	Chupungyeong	36.2197	127.9944	246.1	1974-2010
St. 14	Andong	36.5728	128.7072	141.5	1974-2010
St. 15	Pohang	36.0325	129.3794	3.7	1974-2010
St. 16	Gunsan	36.0019	126.7631	24.6	1974-2010
St. 17	Daegu	35.8850	128.6189	65.5	**1910**-2010
St. 18	Jeonju	35.8214	127.1547	54.8	**1919**-2010
St. 19	Ulsan	35.5600	129.3200	36.0	1974-2010
St. 20	Gwangju	35.1728	126.8914	73.8	1974-2010
St. 21	Busan	35.1044	129.0319	71.0	**1910**-2010
St. 22	Mokpo	34.8167	126.3811	39.4	**1910**-2010
St. 23	Yeosu	34.7392	127.7406	66.0	1974-2010
St. 24	Jinju	35.1636	128.0400	31.6	1974-2010
St. 25	Yangpyeong	37.4886	127.4944	49.4	1974-2010
St. 26	Icheon	37.2639	127.4842	79.4	1974-2010
St. 27	Inje	38.0600	128.1669	201.6	1974-2010
St. 28	Hongcheon	37.6833	127.8803	142.3	1974-2010
St. 29	Jecheon	37.1592	128.1942	265.0	1974-2010
St. 30	Boeun	36.4875	127.7339	176.4	1974-2010
St. 31	Cheonan	36.7794	127.1211	24.0	1974-2010
St. 32	Boryeong	36.3269	126.5572	16.9	1974-2010
St. 33	Buyeo	36.2722	126.9206	12.7	1974-2010
St. 34	Geumsan	36.1056	127.4817	171.7	1974-2010
St. 35	Buan	35.7294	126.7164	13.4	1974-2010
St. 36	Imsil	35.6122	127.2853	249.3	1974-2010
St. 37	Jeongeup	35.5631	126.8658	46.0	1974-2010
St. 38	Namwon	35.4053	127.3328	91.7	1974-2010
St. 39	Jangheung	34.6886	126.9194	46.4	1974-2010
St. 40	Haenam	34.5533	126.5689	14.4	1974-2010
St. 41	Goheung	34.6181	127.2756	54.5	1974-2010
St. 42	Yeongju	36.8717	128.5167	212.2	1974-2010
St. 43	Mungyeong	36.6272	128.1486	172.0	1974-2010
St. 44	Uiseong	36.3558	128.6883	83.2	1974-2010
St. 45	Gumi	36.1306	128.3206	50.3	1974-2010
St. 46	Yeongcheon	35.9772	128.9514	95.0	1974-2010
St. 47	Geochang	35.6711	127.9108	222.4	1974-2010
St. 48	Sancheong	35.4128	127.8789	0.8	1974-2010

Figures

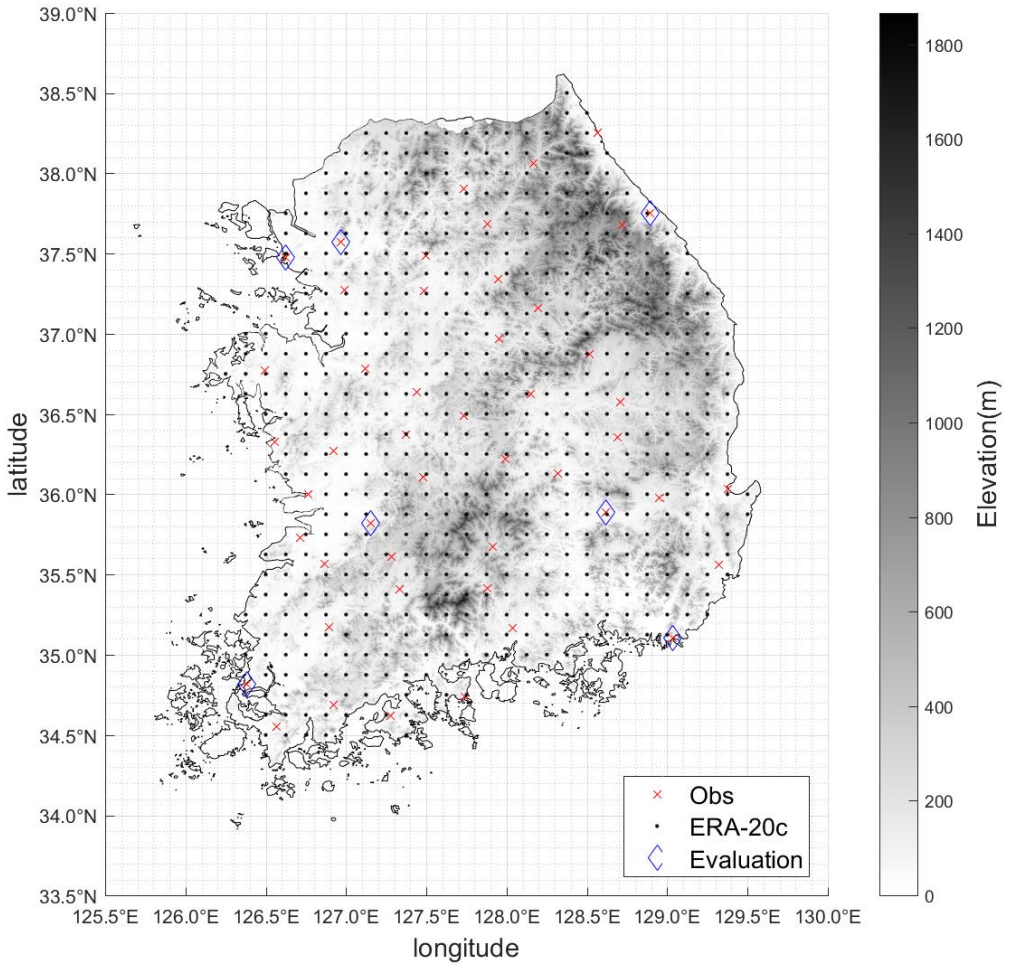


Figure 1. A map showing the study area, local gauging stations, grid points of ERA-20c and evaluation points. The grey shading on the map indicates elevations

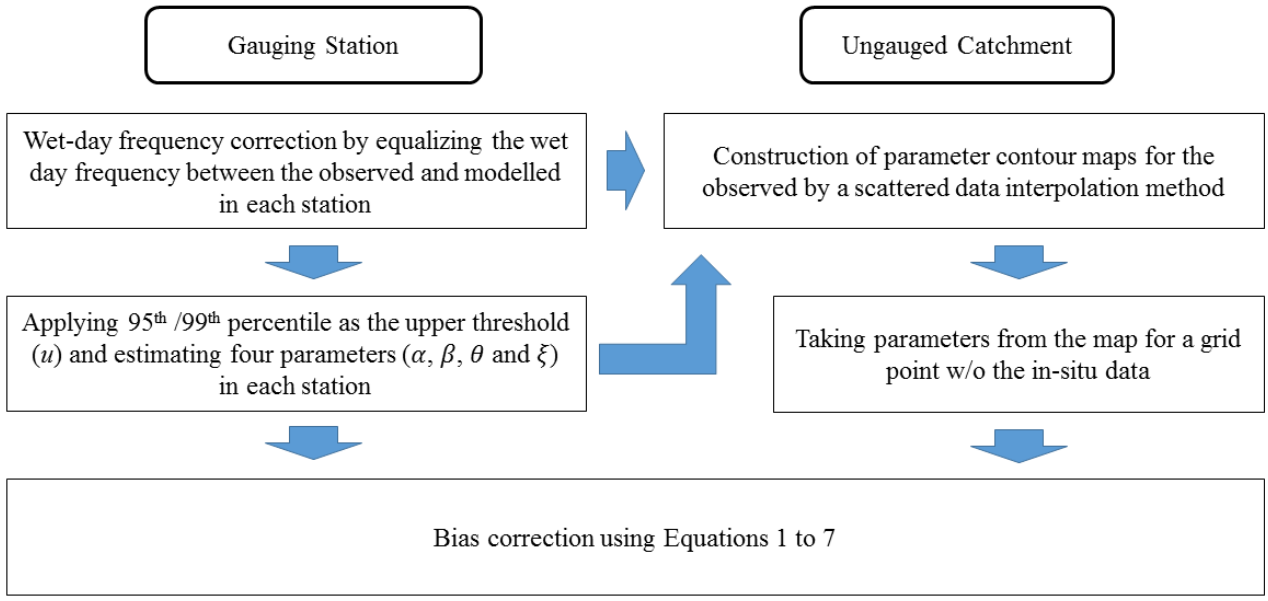
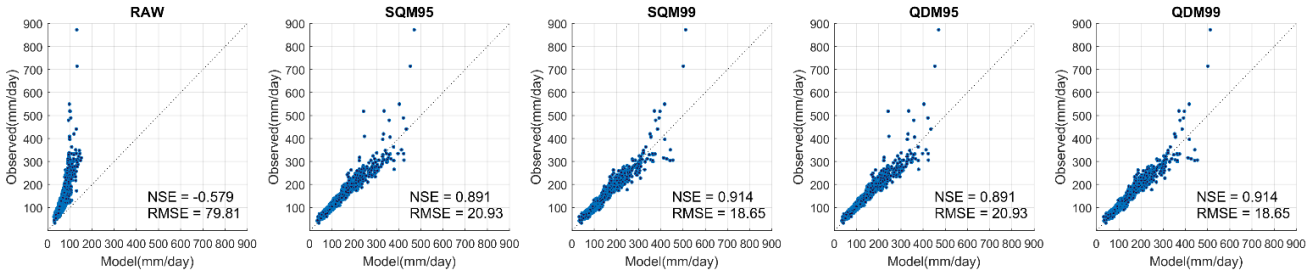


Figure 2. A flowchart of the quantile mapping approach with a composite distribution in gauging stations and ungauged catchment

(a)



(b)

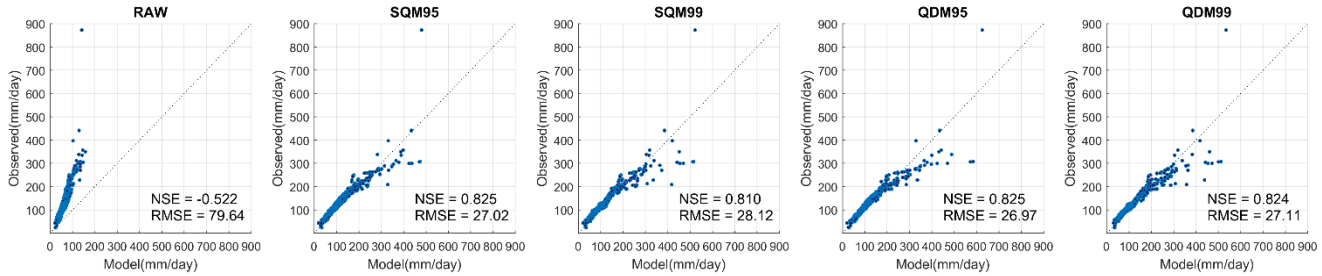


Fig 3. Scatter plot between the Annual maximum rainfalls of the observed and the bias-corrected ERA-20c over (a) all 48 stations for the reference period (1974-2010) and (b) 7 stations from 1910 to 2010

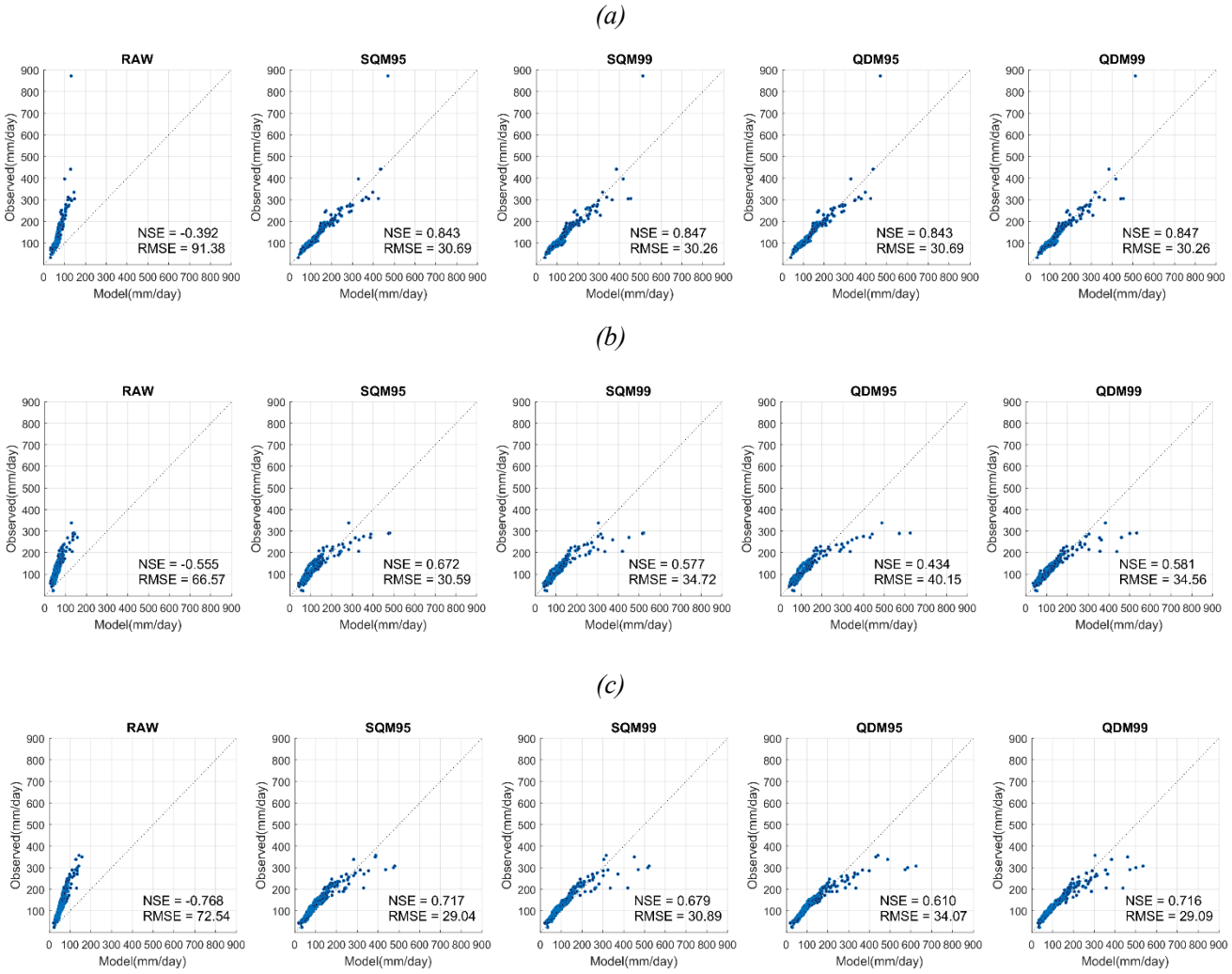


Fig 4. Scatter plot between the Annual maximum rainfalls of the observed and the bias-corrected ERA-20c during (a) 1974-2010, (b) 1937-1973 and (c) 1910-1973 in 7 stations

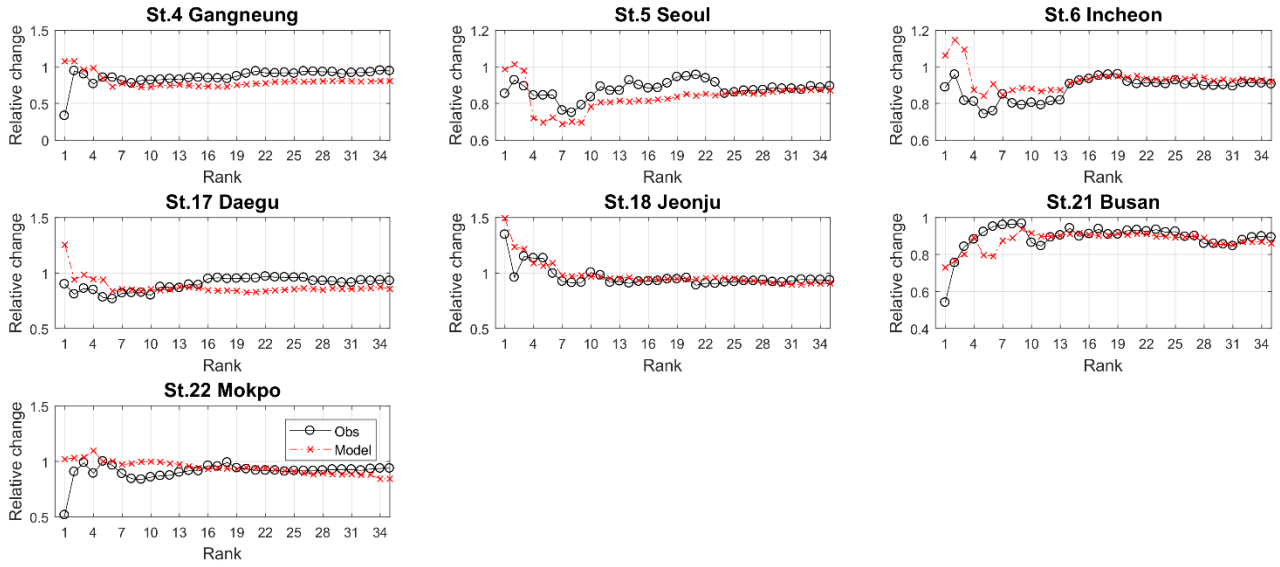


Fig 5. Relative changes in the descending-ordered extreme rainfalls between the reference period (1974-2010) and past period (1937-1973) for the observation (Obs.) in 7 stations and the raw ERA-20c (Model) in the corresponding 7 grid points

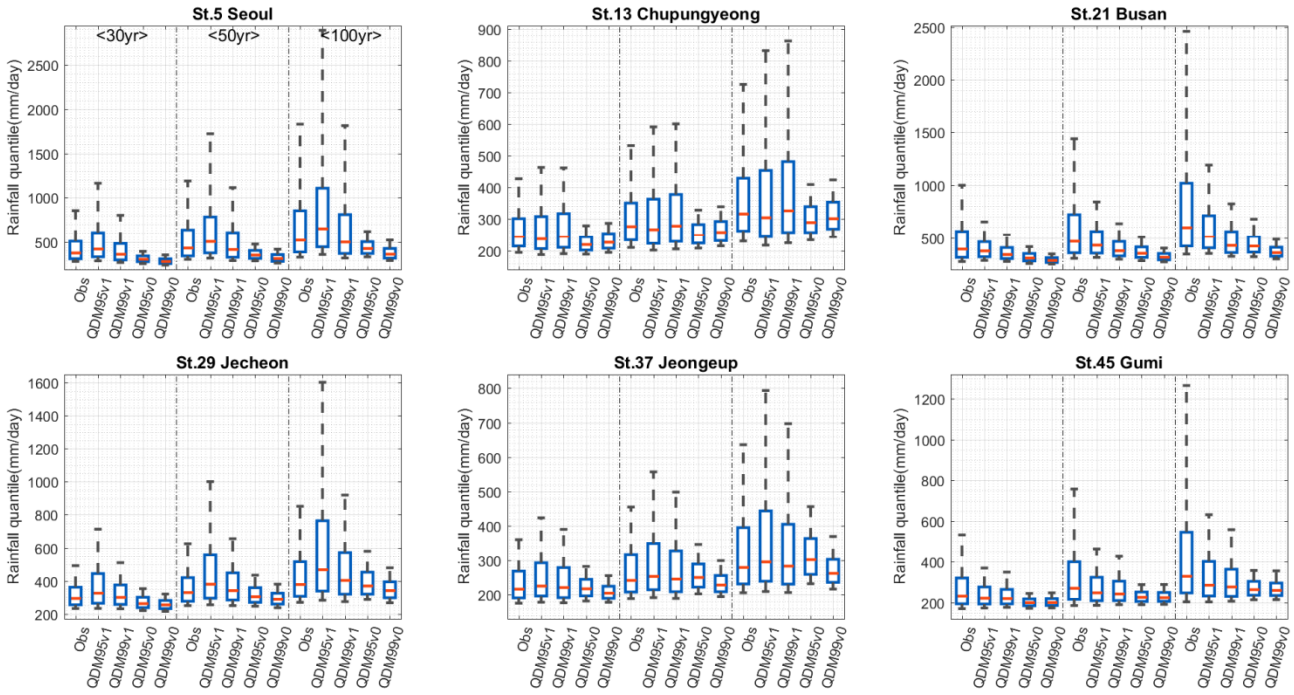


Fig 6. Boxplot for the uncertainties of design rainfalls with 30-year, 50-year and 100-year return period for the observation (Obs) and the bias corrected ERA-20c by QDM approaches in 6 stations (St5. Seoul, St.13 Chupungyeong, St.21 Busan, St.29 Jecheon, St.37 Jeongeup and St.45 Gumi). QDM95v1 and QDM99v1 represent the values estimated for the reference period (i.e. 1974-2010) while QDM95v0 and QDM99v0 are derived from 1900 to 2010. Note that the ends of the whiskers in boxplots mean 9% and 91% of the simulated by MCMC approach

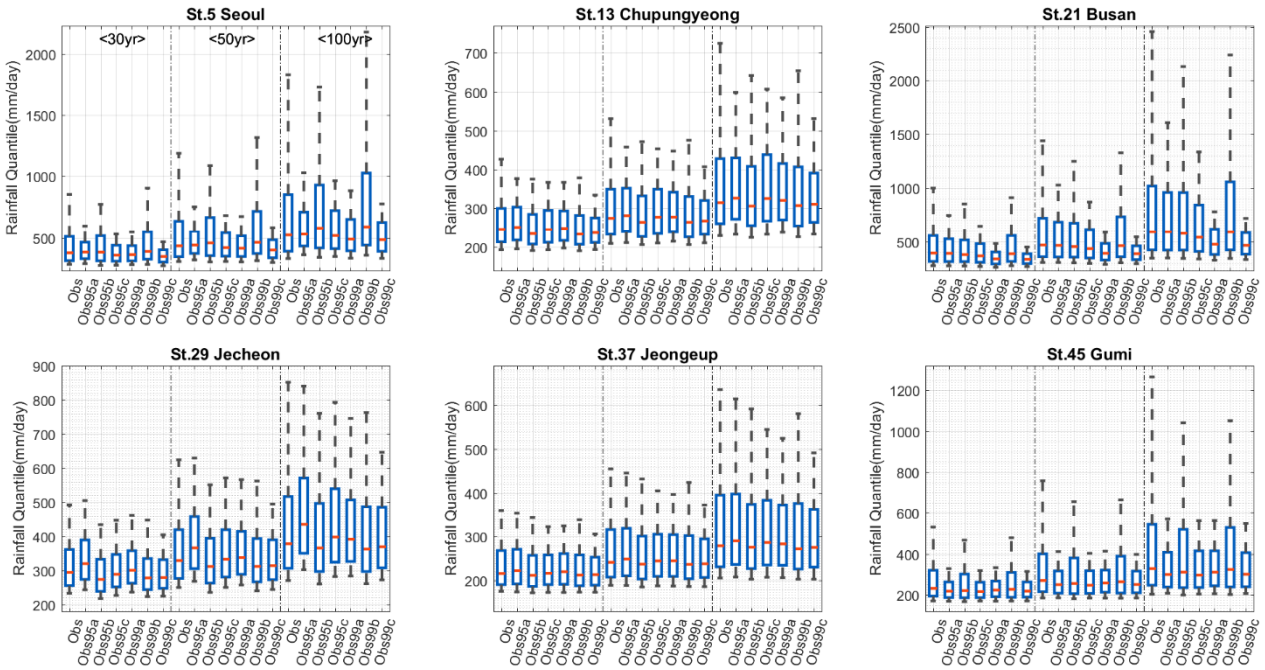


Fig 7. Boxplot for the uncertainties of design rainfalls with 30-year, 50-year and 100-year return period for the observation using the prior information from the bias corrected ERA-20c by QDM approaches in 6 stations (St5. Seoul, St.13 Chupungyeong, St.21 Busan, St.29 Jecheon, St.37 Jeongeup and St.45 Gumi). Here, Obs indicates the values based on the non-informative prior distribution, Obs95a and Obs99a were estimated by the shape parameter information from QDM95v0 and QDM99v0, respectively, Obs95b and Obs99b were based on the corresponding scale and location parameter information, and Obs95c and Obs99c were derived from the prior informations of the all parameters.

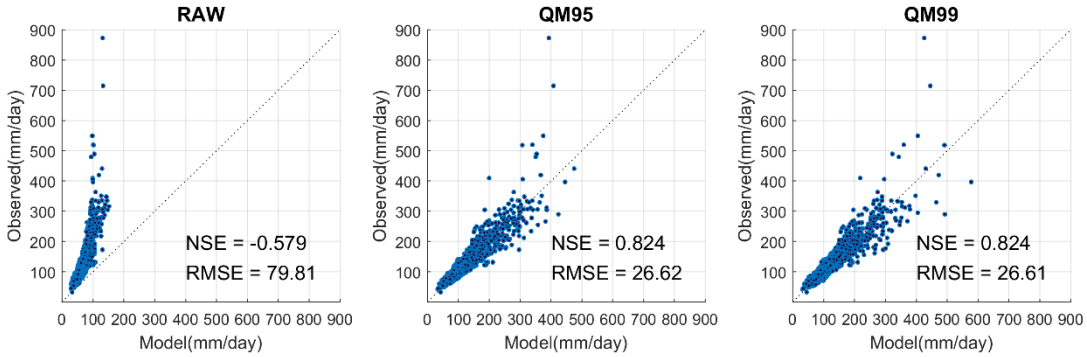


Fig 8. Scatter plot between the Annual maximum rainfalls of the observed and the bias-corrected ERA-20c by QM approaches (QM95 and QM99) over all 48 stations for the reference period (1974-2010). The result presented here are obtained by leave-one-out cross validation.

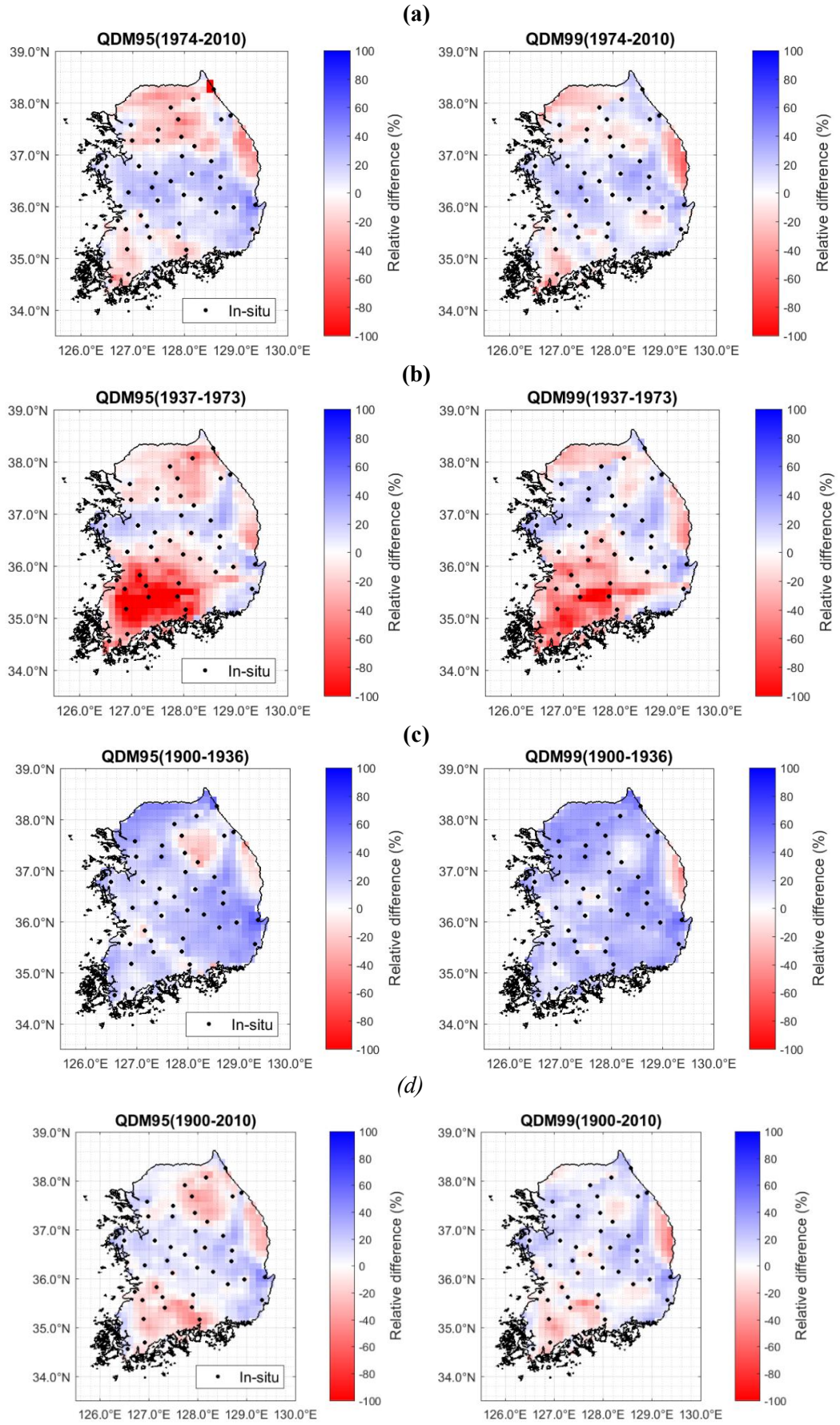


Fig 9. Relative change (%) in design rainfalls of the modelled in four different periods, (a) 1974-2010, (b) 1937-1973, (c) 1900-1936 and (d) 1900-2010, compared with those of the observed for the reference period (1974-2010).