Nonparametric quantile mapping using the response surface method – bias correction of daily precipitation

Taeho Bong, Young-Hwan Son, Seung-Hwan Yoo and Sye-Woon Hwang

ABSTRACT

Currently, regional climate models are widely used to provide projections of how climate may change locally. However, they sometimes have a spatial resolution that is too coarse to provide an appropriate resolution for the local scale. In this paper, a new nonparametric quantile mapping method based on the response surface method was proposed to perform an efficient and robust bias correction. The proposed method was applied to correct the bias of the simulated precipitation for the period of 1976–2005, and the performance and uncertainty were subsequently assessed. As a result, the proposed method was effectively able to reduce the biases of the entire distribution range, and to predict new extreme precipitation. The future precipitation based on representative concentration pathways of RCP 4.5 and 8.5 were bias corrected using the proposed method, and the impacts of the climate scenarios were compared. It was found that the average annual precipitations increased compared to the past for both scenarios, and they tended to increase over time in the three studied areas. The uncertainty of future precipitation was slightly higher than in the past observation period.

Key words | bias correction, moving least squares, precipitation, quantile mapping, response surface method

Taeho Bong
Department of Civil and Construction Engineering,
Oregon State University,
Corvallis,
USA

Young-Hwan Son (corresponding author)
Department of Rural Systems Engineering, and
Research Institute for Agriculture & Life Sciences,
Seoul National University,
Seoul,
Republic of Korea
E-mail: *syh86@snu.ac.kr*

Seung-Hwan Yoo
Department of Rural and Bio-Systems Engineering,
College of Agriculture and Life Sciences,
Chonnam National University,
Gwangju,
Republic of Korea

Sye-Woon Hwang
Department of Agricultural Engineering, Institute of
Agriculture and Life Science,
Gyeongsang National University,
Jinju,
Republic of Korea

INTRODUCTION

Global climate models (GCMs) have a spatial resolution (typical grid cells are 100–300 km) that is too coarse to provide an appropriate resolution of the local scale, and hydrology processes typically occur on finer scales than those provided by the output of the GCMs. Thus, there is a general mismatch between the spatial resolution of the output from GCMs and the local scale of interest in an area. In particular, precipitation is one of the most difficult atmospheric variables to downscale, primarily because of its high spatial and temporal variability and its nonlinear nature (Kallache *et al.* 2011). To overcome these problems, downscaling techniques are used to link the coarse resolution GCM outputs to the catchment-scale climatic variables. Downscaling methods are classified into two main approaches: dynamical and statistical downscaling (Mearns *et al.* 1999). Dynamical downscaling uses regional climate models (RCMs) that transform the outputs from GCMs into outputs with finer spatial and temporal resolutions. However, dynamical downscaling requires high computational costs owing to the complex physics-based structure of the RCMs, and RCMs are also known to exhibit systematic biases (Benestad 2010; Ahmed *et al.* 2013). Hence, bias correction is required to obtain reliable results for local scale climates before being used for climate impact assessment (Christensen *et al.* 2008; Maraun *et al.* 2010).

Many studies have been performed related to the bias correction technique for precipitation (Cannon 2011; Raje & Mujumdar 2011; Goyal *et al.* 2012; Hu *et al.* 2012; Nasseri *et al.* 2013; Tareghian & Rasmussen 2013). Maraun *et al.* (2010) classified those methods into three categories:

doi: 10.2166/wcc.2017.127

(1) perfect prognosis, (2) model output statistics (MOS), and (3) weather generators. Regression methods (known as perfect prognosis) are easy to implement and require relatively lower costs of pre-processing (Chen *et al.* 2010). Thus, most of the statistical downscaling approaches employ regression methods to estimate the empirical relationships between large-scale and local-scale climate circulation. However, these methods are ill-suited for predicting extreme values of the climate variables.

An alternative approach is to use techniques, such as quantile regression, to estimate the particular quantile that corresponds to the extreme values. Chen *et al.* (2013) compared the performance of six bias correction methods and summarized their advantages and disadvantages. They noted that distribution-based methods are consistently better than mean-based methods, and that all six bias correction methods exhibit an inability to specifically correct the temporal structure of daily precipitation occurrence. Themeßl *et al.* (2011) also compared several statistical downscaling approaches, and quantile mapping was determined to be the most efficient in removing the biases of precipitation.

In many studies and practices, quantile mapping is commonly used to correct the systematic biases in precipitation outputs from climate models and it is considered an effective method for removing GCM or RCM biases with relatively low computational cost (Maurer & Pierce 2014). The quantile mapping bias correction algorithms have been also reviewed in the context of hydrological impacts studies on streamflow, and have been found to outperform simpler bias correction methods that correct only the mean or mean and variance of precipitation series in the statistical bias correction (Teutschbein & Seibert 2012; Chen *et al.* 2013). Additionally, frequency analysis using bias-corrected future climate data would enhance the management of water resources applications as well as the effective utilization of water resources (Biniyam & Kemal 2017).

The limitation of quantile mapping based bias-correction is that temporal errors of major circulation systems cannot be corrected because bias is defined as the time-independent component of the error (Chen *et al.* 2011; Haerter *et al.* 2011). Quantile mapping does not alter the temporal sequence of the simulated output. For this reason, weather generators are frequently applied to produce time series for impact studies. However, hydrological processes depend on the entire distribution function of precipitation intensity and temperature, and it is well known that pre-processing is necessary to remove biases present in the simulated climate outputs (Haerter *et al.* 2011). From this perspective, quantile mapping is widely accepted for the correction of climate data that are to serve as input data for impact models considering temporally integrated statistics, such as hydrological (e.g. floods and drought) or crop models. Rajczak *et al.* (2016) compared the capabilities of the quantile mapping and weather generator method for RCM precipitation data at weather stations across Switzerland. They found that the application of quantile mapping provides an obvious added value for the representation of transition probabilities and multiday precipitation characteristics in climate model data at the local scale, although temporal characteristics are not explicitly corrected by the quantile mapping. However, it should be noted that the performance of the bias-corrected historical simulation results in reproducing observational temporal characteristics is needed to be checked beforehand. Another limitation is that it is not possible to reflect changes in the statistical characteristics of the biases between the observed and simulated data. The basic assumption for quantile mapping is that the biases are stationary in the future. However, if the time series used to train the statistical model is long enough, many different situations, including the altered climate, may be included (Zorita & von Storch 1999). Therefore, quantile mapping techniques are still widely used, and various techniques have been studied to solve the quantile mapping task. Gudmundsson *et al.* (2012) considered three types of quantile mapping methods, and quantile mapping by nonparametric transformations was recommended for most applications of statistical bias correction. Lafon *et al.* (2013) also compared the performance of four statistical downscaling techniques (linear, nonlinear, γ-based quantile mapping, and empirical quantile mapping), and the most comprehensive correction was achieved using the empirical quantile mapping methods. They noted that the accuracy of this method is controlled by the number of quantile divisions; however, using too many quantiles might result in overfitting the model to the data. When the observed climate data are sufficiently large, the cumulative distribution shows a smooth curve, and the value between

the percentiles can be estimated effectively by linear interpolation with small quantile intervals. However, if climate data is insufficient, the empirical quantile mapping can be sensitive to the number of data, and data for the extreme precipitation are relatively scarce. Therefore, high cumulative probability changes are sensitive to some observed precipitations. Another major drawback of quantile mapping is that the greatest observed value can never be exceeded in the regional climate scenario, which could be problematic for extremely high precipitation. Thus, it is important to overcome this drawback because the amount of daily precipitation and extreme precipitation has been increasing (Westra *et al.* 2013; Asadieh & Krakauer 2015). To overcome this drawback of quantile mapping, Boé *et al.* (2007) used a simple extrapolation, which is a constant correction. However, the correction constant is quite variable for high precipitation and does not take into account the trend of extreme precipitation.

In this study, a new empirical quantile mapping technique using the response surface combined with the moving least squares method (QM-RSMLS) was proposed to overcome the drawbacks of the conventional empirical quantile mapping. The conventional response surface method (RSM) is modified by changing the response surface function (RSF) using the transcendental function to consider the non-linearity between percentiles and to prevent overfitting, and the percentiles for the unobserved precipitation were estimated by considering the trend of certain probability intervals. Therefore, it is possible to reproduce the empirical cumulative distribution function (CDF) of precipitation having a relatively robust and smooth curve, because the proposed method is not sensitive to specific precipitation as it reflects the trend of distribution, and the new extreme precipitation can be extrapolated. The proposed method was applied to correct the bias of the simulated historical precipitation for the period 1976–2005. The performance of the QM-RSMLS was assessed by the mean absolute error (MAE) and by cross validation for the precipitation on the past observation period. For future precipitation, two representative concentration pathways (RCPs)-based precipitation data (RCP 4.5 and 8.5), established by the Korean Meteorological Administration (KMA), were bias corrected for the period 2011–2900. The impacts of climate change on precipitation for the RCP 4.5 and 8.5 climate scenarios were compared, and the uncertainties of precipitation were assessed.

METHODOLOGY

Quantile mapping method

The quantile mapping method, which is a generalized approach for the MOS proposed by Panofsky & Brier (1968), was first used to remove the systematic bias in GCM simulations (Wood *et al.* 2002). In this method, the CDF of GCM outputs for the past were mapped onto the CDF of the past observations. Therefore, all of the statistical moments of the simulated outputs can be matched with those of the observations. The empirical CDFs of the observed and simulated data are used to remove the biases, and the transformation form is defined as:

$$P_O = F_O^{-1}(F_S(P_S)) \tag{1}$$

where P_O is the observed value, P_S is the simulated value, and F_O^{-1} and F_S are the inverse CDF of P_O and the CDF of P_S, respectively. Figure 1 shows a schematic of the quantile mapping approach.

Various methods for performing quantile mapping were presented by Gudmundsson *et al.* (2012). In their study, the quantile mapping methods were classified into three groups. The first method is the distribution derived transformation

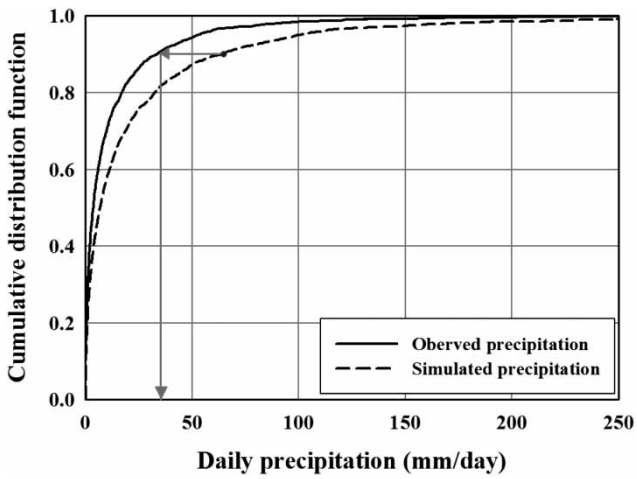

Figure 1 | Schematics of the quantile mapping approach.

based on the theoretical distribution, with the generalized extreme value (GEV) or the general Pareto distribution (GPD) commonly used for modeling extremes because either approach provides a sophisticated basis for analyzing the upper tails of the distribution. The gamma distribution, in contrast, provides a good fit to the entire data sample, but not necessarily to the upper tail (Benestad 2010). The second method is the parametric transformation, which uses the regression equation directly. The quantile-quantile relation was modeled using various regression equations (such as linear, nonlinear, power-law, or exponential). The last method is the nonparametric transformation. The general method of a nonparametric transformation is performed using a correction table of the two CDFs (simulated and observed data), and a linear interpolation is applied between the two percentiles.

Proposed QM-RSMLS

To formulate the empirical CDFs, the RSM which was first developed by Box & Draper (1987) was used. The RSM is a method used to formulate the relationship between the input and output of a physical experiment using a simple mathematical expression. Currently, the RSM has become one of the better known meta-modeling techniques. When considering only one variable, the simplest model that can be used in RSM is based on the first-degree model:

$$Y = a_0 + a_1 x + \varepsilon \tag{2}$$

where Y and x represent the output and variable, respectively. a_0 is the constant term, a_1 represents the coefficients of the linear parameter, and ε is the residual. The second-degree model is expressed as follows:

$$Y = a_0 + a_1 x + a_2 x^2 + \varepsilon \tag{3}$$

where a_2 represents the coefficients of the secondary polynomial. Equations (2) and (3) can be written in a simpler form as:

$$Y = aX + \varepsilon \tag{4}$$

where aX is called the response surface and the coefficients (a) are generally estimated using the least squares method to minimize the errors of the entire set of data points. Applying the least squares method, the analytical solution can be expressed as:

$$a = (X^T X)^{-1} X^T Y \tag{5}$$

In the conventional RSM, polynomials are used to generate a RSF. To consider the non-linearity between the percentiles and obtain a more accurate prediction, second or higher degree polynomial models should be used. However, the CDF (ascending cumulative plot) is an increasing function, and when using a second or higher degree polynomial to minimize only the prediction error for sampling points (observed values), overfitting can occur in scarce data regions. To overcome this problem, the conventional RSF has been replaced by a natural logarithmic function because the relationship between the precipitation and standard Gaussian variate is similar to a logarithmic function, and the logarithmic function is also an increasing function:

$$Y = a_0 + a_1 \ln(x) + \varepsilon \tag{6}$$

In Equation (6), the input is the precipitation and the output is the standard Gaussian variate corresponding to that precipitation. If the standard Gaussian variate is the input and the precipitation is the output, then the exponential function that is the inverse of the natural logarithmic function should be used as the RSF. However, the coefficients of the exponential function cannot be directly estimated using the least squares method. Therefore, the RSF is generated using Equation (6), and the precipitation is calculated using the inverse function of Equation (6) as follows:

$$x = e^{(Y - a_0)/a_1} + \varepsilon \tag{7}$$

As mentioned above, the coefficients of the RSF can be easily calculated using the least squares method. However, there are approximation errors inherent in the RSM. In particular, when the relationship between the input and output is highly nonlinear, the approximation errors increase. Therefore, many studies have been conducted to reduce the approximation errors, and the moving least squares (MLS) is one of the improved approximation methods. The MLS method was introduced by Shepard (1968) in the lowest

order case and was generalized to a higher degree by Lancaster & Salkauskas (1986). The MLS method uses a weight function to estimate the coefficients of the RSF, with the weight function depending on the distance between the approximation point and a point in an influence domain. In this study, the quartic-spline function was used as the weight function:

$$w(x-x_i) = w\left(r = \frac{\|x-x_i\|}{d_m}\right) = \begin{Bmatrix} 1-6r^2+8r^3-3r^4 & for\ r \geq 1 \\ 0 & for\ r < 1 \end{Bmatrix} \quad (8)$$

where x is the approximation point, x_i is the observed point, and d_m is the distance of influence. The size of the response surface is determined by d_m. If d_m is the nth smallest distance value, then the $n-1$ sample points closest to the approximation point are selected and used to generate the local RSF. The matrix of the weight function is a diagonal matrix and is expressed as:

$$[W] = diag[w(x-x_1), w(x-x_2), \cdots, w(x-x_n)] \quad (9)$$

The coefficients of the local RSF can be estimated by minimizing the weighted residuals as follows:

$$\beta = (X^T W X)^{-1} X^T W Y \quad (10)$$

In the process of quantile mapping, some percentiles of precipitation may exceed the range of the probability of the observed precipitation. In this case, precipitation was corrected to the maximum value of the observations. Therefore, the bias-corrected values cannot exceed the maximum observed precipitation. Boé *et al.* (2007) used a simple extrapolation, which is a constant correction. The ratio of simulated and bias-corrected precipitation for the last quantile was used as a correction constant. However, this correction constant depends on the nonlinearity of the relationship between the simulated and bias-corrected precipitation, and its variability is greater for the range of high precipitation (quantile), as shown in Figure 2.

Therefore, this method will be poorly extrapolated for extreme precipitation. In this study, the tails of the simulated and observed cumulative distributions were predicted considering the non-linear trend of distribution. Extrapolation may lead to an incorrect estimation that is further away from the observed range. Therefore, the maximum value was limited, so the increasing rate of bias-corrected precipitation could not exceed that of the simulated precipitation from the last quantile. Figure 3 shows an example of extrapolation to a new extreme precipitation.

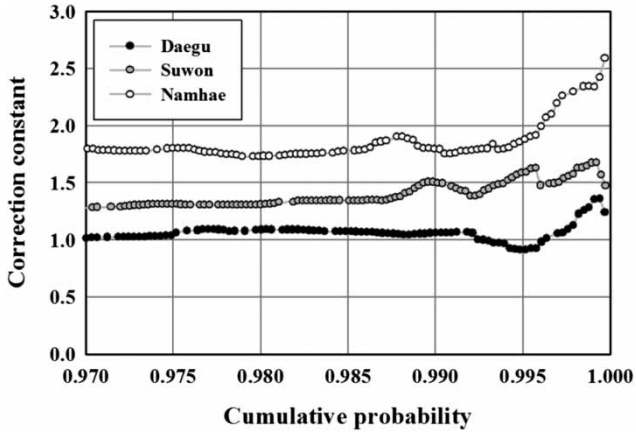

Figure 2 | Correction constant according to the cumulative probability.

In general, the data for extreme precipitation is very scarce, and the difference between their values is relatively large. Therefore, linear extrapolation using only two points may provide a poor estimation because its trend is largely influenced by the difference between the two largest values, as indicated by a comparison of Figure 3(a) and 3(b).

In addition, the cumulative probability of precipitation over a certain size was sometimes greater than one. On the other hand, the proposed method uses several precipitation values to consider the trend of the cumulative probability distribution, and a smooth CDF curve can be generated. Therefore, more robust extrapolation is possible, and the cumulative probability of precipitation never exceeds one.

Implementation of the QM-RSMLS

In the procedure of the proposed method, a frequency analysis is first performed to eliminate errors owing to duplicate values, and then the empirical CDF is generated (Figure 4).

Figure 5 shows a schematic depiction of the QM-RSMLS implementation.

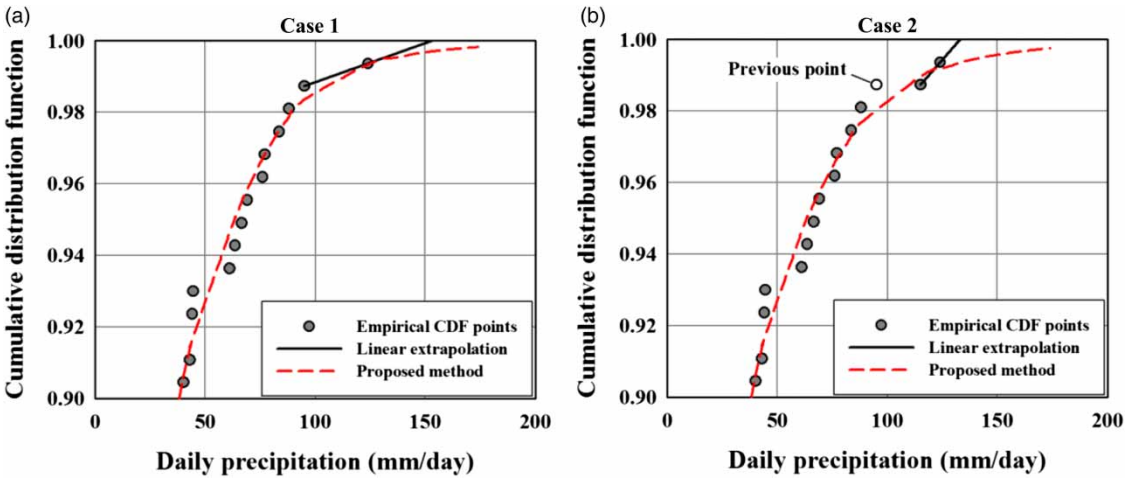

Figure 3 | Extrapolation to extreme precipitation.

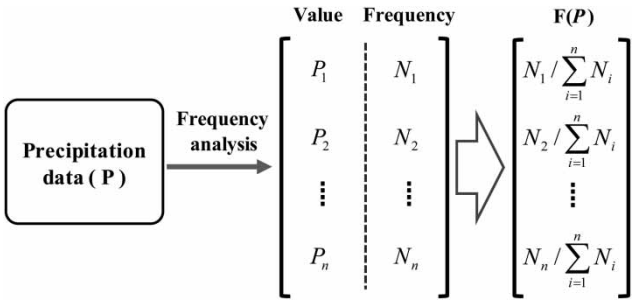

Figure 4 | Generation of the empirical CDF of precipitation.

There are two transformations for which RSF is required. First, the $RSF_{Sim.\text{-}SGV}$ is generated for the simulated precipitation, where $RSF_{A\text{-}B}$ represents the response surface function and the subscript A-B indicates the transformation from A to B. The subscripts Sim., Obs., and SGV represent the simulated precipitation, observed precipitation, and standard Gaussian variate, respectively. $F(\cdot)$ and Φ^{-1} are the empirical CDF and the inverse of the cumulative standardized Gaussian distribution, respectively. In $RSF_{Sim.\text{-}SGV}$, the input is the simulated precipitation and the output is the standard Gaussian variate corresponding to that precipitation. The reason for using a standard Gaussian variate instead of the percentile is the low non-linearity between precipitation and the standard Gaussian variate. Additionally, it is possible to reasonably fit the tails of the cumulative, because the standard Gaussian variate was converted into a probability with a value that was always less than one.

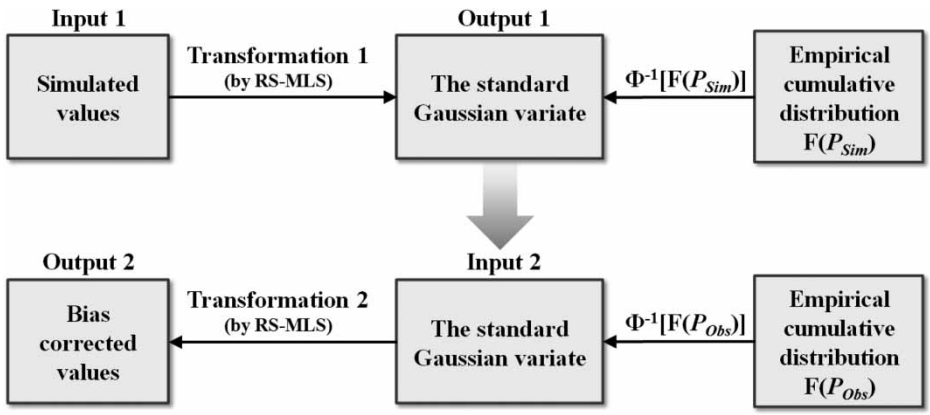

Figure 5 | Schematic depiction of the QM-RSMLS.

Second, the $RSF_{SGV\text{-}Obs.}$ is generated for the observed values of precipitation. However, as previously mentioned, the coefficients of the exponential function cannot be directly estimated using the MLS. Therefore, the $RSF_{Obs.\text{-}SGV}$ is generated for the observed values of precipitation, and then, $RSF_{SGV\text{-}Obs.}$ is obtained using Equation (7).

The d_m was set to be the 11th smallest distance value, so 10 points were automatically selected closest to the approximation point to create the RSF. The coefficients of each RSF and the weighted function according to the approximation precipitation are estimated by the MLS. Finally, the bias-corrected precipitations that exceed the maximum observed precipitation were fixed as the maximum observed precipitation.

STUDY AREA AND DATA

Study area

In this study, the 67 weather stations of South Korea were divided into three groups based on the average annual precipitation (period 1976–2005), and three representative areas were selected from each group considering the station locations. The selected study areas are Daegu, Suwon, and Namhae, with average annual precipitations of 1058.4, 1268.3, and 1817.1 mm, respectively. Figure 6 shows the location of the study area, and the points indicate the locations of the weather stations.

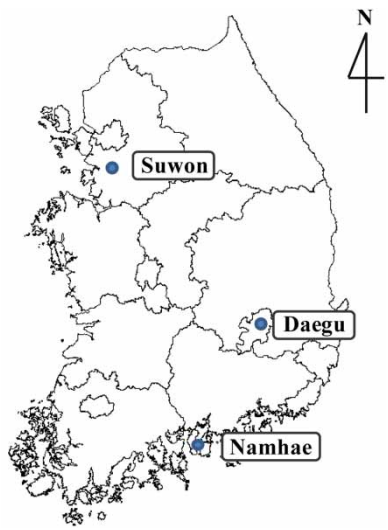

Figure 6 | Locations of the weather stations.

Daily precipitation data

The daily precipitation data from three weather stations were used in our study. The observation period was the past 30 years (1976–2005). For RCM data the HadGEM3-RA model, which is a regional atmospheric model based on the atmospheric component of the latest Earth System Model developed by the Met Office Hadley Centre (HadGEM2-AO), was adopted. The KMA has utilized the HadGEM3-RA for dynamical downscaling, and established RCP-based precipitation data with 12.5 km resolution in Korea. The schematic for generating a climate change scenario is shown in Figure 7.

The historical period was the same as the observation period (1976–2005), and the future simulated precipitation data of the two RCP scenarios (RCP 4.5 and 8.5) during the period from 2011 to 2100 were used for future assessment after bias correction of daily precipitation.

RESULTS AND DISCUSSION

QM-RSMLS assessment

Most RCMs tend to overestimate the occurrence of wet days (Murphy 1999; Maraun *et al.* 2010). This overestimation is called the drizzle effect, and a simple method to correct the drizzle effect is to set all the simulated precipitation values below a certain threshold to zero. Fowler *et al.* (2007) and Maraun (2013) used a wet day threshold of 1 mm/day to correct the drizzle effect. In this study, the simulated and observed precipitation data were sorted in descending order, and the simulated precipitation was set to zero, with ranks of less than or equal to the number of dry days in observed precipitation (Hay & Clark 2003). The threshold precipitations were calculated as 1.2, 1.4, and 1.2 mm/day in Daegu, Suwon, and Namhae, respectively. As a result, the number of dry days for both the simulated and observed precipitation data were the same, and the empirical CDFs were only estimated from the wet

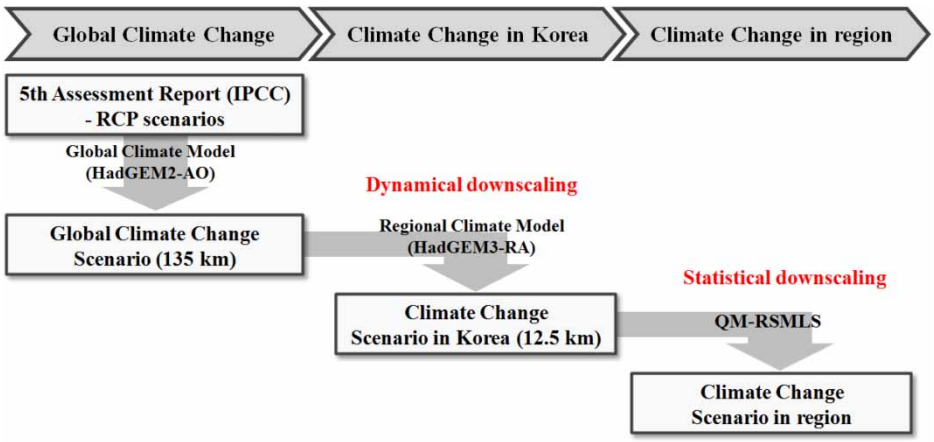

Figure 7 | Generation process of a climate change scenario.

days (the lower limit was 0.1 mm/day). Figure 8 shows the empirical CDFs of the precipitation for the three areas.

The empirical CDFs of Daegu and Suwon were almost identical, whereas Namhae exhibited a different distribution compared with the other two points and had a higher probability of daily precipitation than the other two areas. To remove the bias, the QM-RSMLS was conducted for the historical RCM precipitation data. The average annual precipitation, statistical properties of the observed precipitation, raw RCM, and corrected RCM are summarized in Table 1.

After bias correction over the past 30 years of precipitation, the average daily precipitations of RCM of Daegu

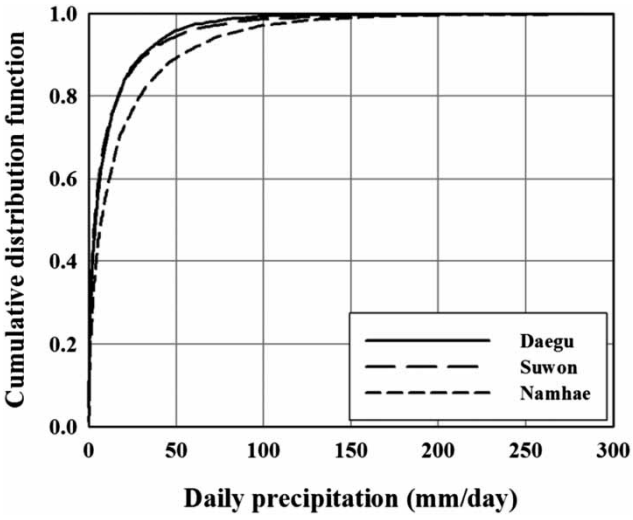

Figure 8 | Empirical CDFs of the observed precipitation (1976–2005).

Table 1 | Observed and bias-corrected precipitation over the past 30 years (1976–2005)

Weather station	Daily average[a] (mm/day)	Annual average (mm/year)	Standard deviation	COV (%)
Daegu				
Observed	11.3	1058.4	264.8	25.0
Raw RCM	13.0	1227.2	243.3	19.8
Bias-corrected	11.2	1058.2	258.7	24.4
Suwon				
Observed	12.0	1268.3	252.8	19.9
Raw RCM	12.1	1285.4	227.6	17.7
Bias-corrected	12.0	1267.8	266.5	21.0
Namhae				
Observed	19.6	1817.1	480.7	26.5
Raw RCM	12.4	1154.0	236.0	20.5
Bias-corrected	19.6	1818.4	462.4	25.4

[a]For the wet days only.

and Suwon were decreased by 15.3 and 0.8%, respectively, and Namhae was increased by 54.4%. The bias-corrected precipitations were approximately equal to the observed average daily precipitation. In addition, the standard deviation was also corrected, and the variability of the RCM precipitation was similar to the observed precipitation. Figure 9 shows the Q-Q plots of the raw and corrected RCM data against the observed precipitation.

The raw RCM data showed a large difference compared to the observations, and the greatest error was exhibited for high precipitation. After bias correction, the three corrected

RCM data sets were in good agreement with the observed precipitation. For comparison, a variety of quantile mapping methods were conducted, and their performances were compared. To assess the performance of the quantile mapping methods, Gudmundsson *et al.* (2012) described a variety of scores that were previously used, and they suggested using a set of scores based on the MAE for the overall performance, while keeping track of the many relevant properties of the distribution. In this study, the same scores were used to evaluate the performance of the QM-RSMLS. MAE_i represents the MAE of the specific probability interval, and the subscript indicates the upper bounds of the 0.1 wide probability intervals. Figure 10 shows the total MAE and MAE_i of the quantile mapping methods.

The QM-RSMLS clearly showed the best performance among the quantile mapping methods, with total MAEs for Daegu, Suwon, and Namhae of 0.04, 0.04, and 0.06 mm, respectively. In particular, the QM-RSMLS can reduce the biases from the RCM precipitation in the entire range of the distribution. For high precipitation, $MAE_{1.0}$ exhibited a relatively large error compared to other probability intervals because the empirical distribution of high precipitation is not a smooth curve. However, each of the three areas had a very small error of less than 0.4 mm, and the proposed method exhibited good performance compared with other quantile mapping methods. Therefore, the proposed method is better suited to predict extreme precipitation. When using the GEV distribution, the total MAE values were small in the three areas, but a very large MAE was found for high precipitation of more than 90%.

The uncertainty of annual precipitation over the past 30 years (1976–2005) was assessed using a box plot (Figure 11).

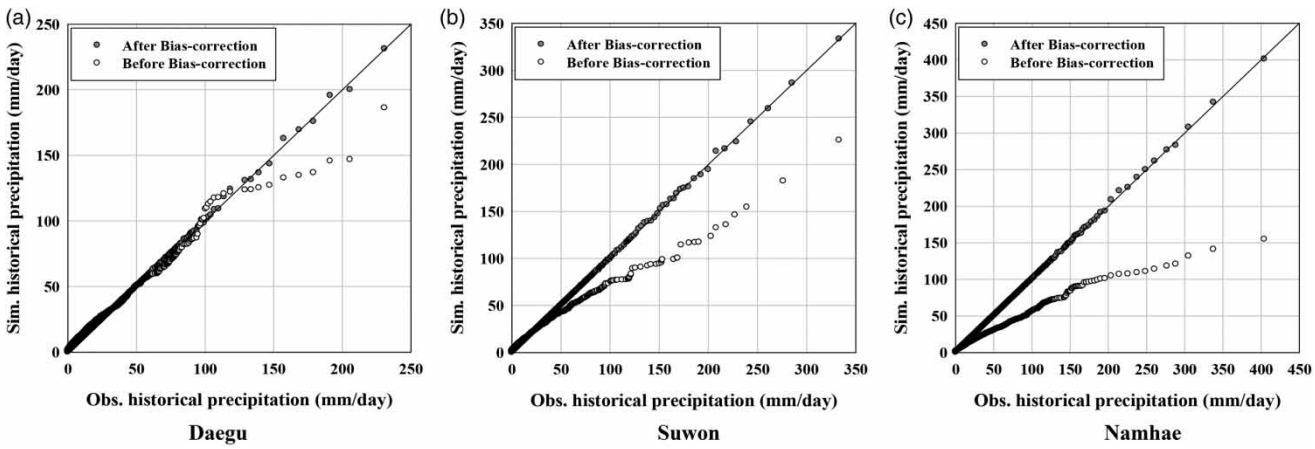

Figure 9 | Q-Q Plots of the simulated and the bias-corrected precipitation.

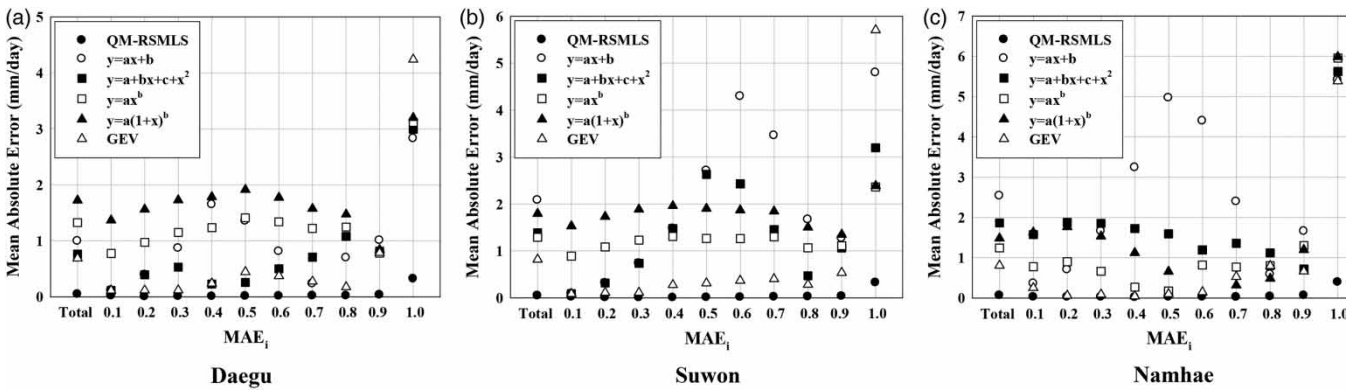

Figure 10 | Total MAE and the MAE for the specific probability interval.

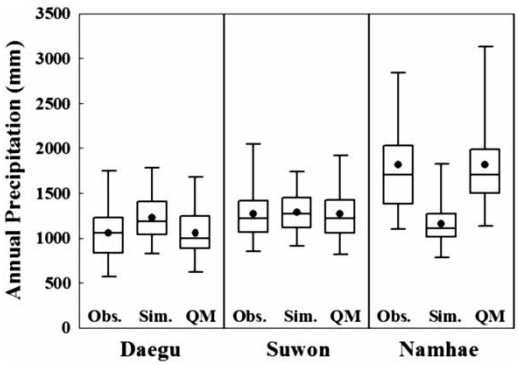

Figure 11 | Variability of annual precipitation over the past 30 years (Obs.: Observed precipitation, Sim.: Simulated precipitation, QM: Bias-corrected precipitation).

The top and bottom of the box indicate the top 75% (3rd quartile) and the bottom 25% (1st quartile) of the data, respectively. The horizontal line inside the box is the median, and the dot is the average annual precipitation. The vertical lines extending from the box reach out to the minimum and maximum values.

The precipitation showed a large uncertainty owing to high temporal variability, and the difference between the maximum and minimum values was highly significant. In particular, the uncertainty of Namhae is much greater than that of the other areas because Namhae has a relatively high daily precipitation and a wide range of precipitation is observed compared to other areas. Comparing the uncertainty according to the precipitation data shows that the simulated precipitation uncertainty is different from the observed precipitation uncertainty. However, it became similar to the observed precipitation uncertainty after bias correction by the proposed method.

Cross validation

To assess the performance of the downscaling models for the independent period, K-fold cross validation, which is one way to improve over the holdout method, was conducted. The data set was divided into six subsets: one of the six subsets was used as the testing set (5 years), and the other five subsets were put together to form a training set (25 years). The cross validation process was repeated six times for each area, and the error in the cross validation was estimated via MAE for the entire range of the distribution. The total error was obtained as the average MAE for each case as follows:

$$Total\_error = \frac{1}{K}\sum_{i=1}^{K} MAE_i \tag{11}$$

Figure 12 shows the total error in the cross validation for each area.

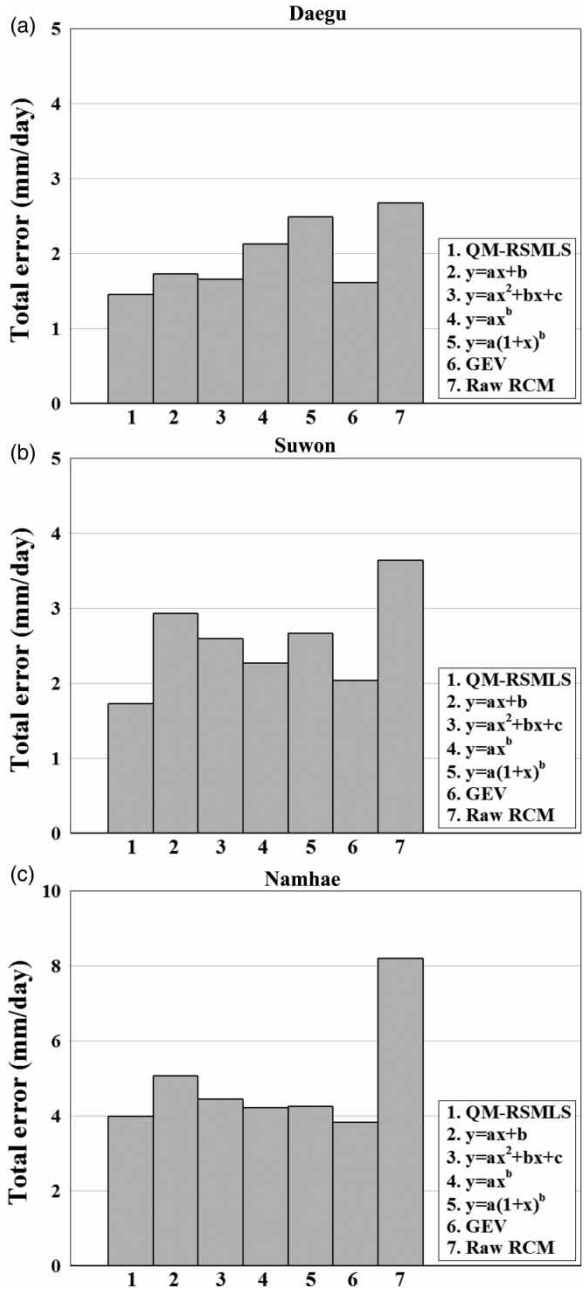

Figure 12 | Total error in the cross validation.

In Daegu and Suwon, the proposed method shows the smallest error, and the errors were reduced to 45.7 and 52.6% compared to the raw RCM data, respectively. In Namhae, the GEV shows the smallest error, and the proposed method has the second smallest error (the error was reduced by 51.4%). Comparing the average error of cross validation, Namhae exhibited a relatively large error. The reason for this is the difference in the probabilistic distribution of the precipitation data for the training and validation periods. The quantile mapping implicitly assumes the stationarity of climate data in their transfer functions. Therefore, the validation error can be varied depending on the stationarity of climate data. Figure 13 represents the average difference between the distributions of the precipitation data for the training and validation periods.

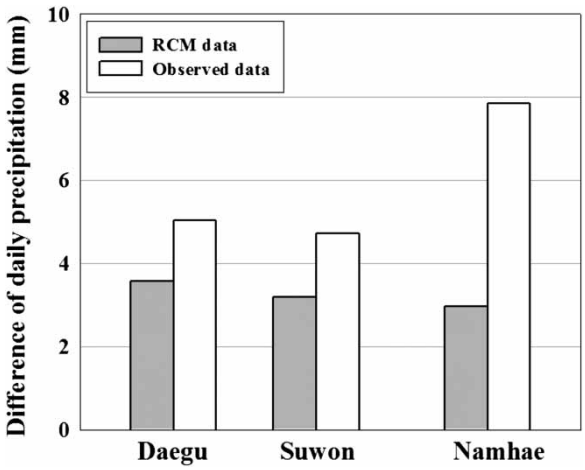

Figure 13 | Precipitation difference between the precipitation probabilistic distributions of training period and validation period.

The change in the probabilistic distribution of RCM data did not significantly differ depending on the area, but that of the observed precipitation data was the largest in Namhae. Accordingly, it can be deduced that Namhae has a large error in cross validation.

Bias correction of the future precipitation

Future RCM precipitation data from 2011 to 2100 were bias corrected using quantile mapping for the relationship between the historical RCM data and the observed data over the past 30 years. The drizzle effect was preferentially corrected using the threshold of each area, and then QM-RSMLS was performed to remove the bias. After performing the drizzle effect correction, the statistical properties of the raw and bias-corrected RCM are summarized in Table 2.

From the results, the average annual precipitations were expected to increase in the future, and the average annual precipitation for the RCP 8.5 scenario exhibited a larger value than that of the RCP 4.5 scenario. For the RCP 4.5 scenario, the average annual precipitation in the future period (2011–2100) increased by 4.6, 4.2, and 19.5% in Daegu, Suwon, and Namhae, respectively, compared to the past

Table 2 | The statistical properties of bias-corrected future precipitations (2011–2100)

	Daegu		Suwon		Namhae	
Scenario	**Raw RCM**	**Corrected RCM**	**Raw RCM**	**Corrected RCM**	**Raw RCM**	**Corrected RCM**
RCP 4.5						
Average[a] (mm/day)	13.7	12.0	12.6	12.7	14.1	23.6
Standard deviation	19.0	21.0	19.6	27.7	20.6	55.2
COV (%)	138.4	174.8	156.1	218.1	146.8	233.6
Annual average (mm/year)	1276.7	1119.9	1329.3	1342.0	1362.9	2291.2
RCP 8.5						
Average[a] (mm/day)	14.0	12.3	12.9	13.1	14.2	24.2
Standard deviation	20.7	22.5	22.2	31.4	21.7	62.1
COV (%)	147.9	182.8	172.6	240.1	152.4	256.8
Annual average (mm/year)	1294.2	1138.2	1378.8	1405.4	1403.9	2382.8

[a]For the wet days only.

period (1976–2005). For the RCP 8.5 scenario, the average annual precipitation increased by 6.4, 8.8, and 23.3% in Daegu, Suwon, and Namhae, respectively (Figure 14).

When comparing the three areas, Namhae showed the largest bias and average annual precipitation. The coefficient of variation, which is the ratio of the standard deviation to the mean, of the daily precipitation was the largest in Suwon. The new extreme precipitation was estimated using QM-RSMLS, constant correction, and linear extrapolation (Figure 15).

The extreme precipitation was significantly increased compared to the past, and Namhae exhibited the highest probable maximum precipitation. When using a constant correction, the probable maximum precipitation showed the highest values for all areas, and the linear interpolation exhibited the lowest extreme precipitation, although the future simulated precipitation increased significantly. On the other hand, the proposed method was able to estimate the extreme precipitation over a more reasonable range, considering the trend in the probabilistic distribution of extreme precipitation.

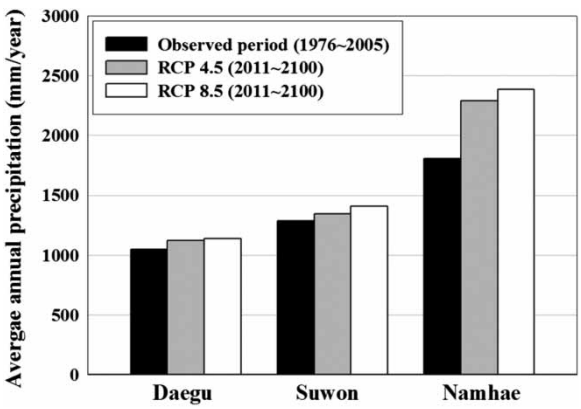

Figure 14 | Comparison of the average annual precipitation during the historical and two climate scenarios.

To analyze the trend of precipitation over time, a moving average was used to smooth out the short-term fluctuations and highlight the longer-term trends. The size of the subset was set to 20 years, and the subset was moved forward one year. Figure 16 shows the moving average of annual precipitation and its trend line.

The average annual precipitation was found to increase with time for all three areas. In Daegu and Namhae, the increasing trend of average annual precipitation for the RCP 8.5 scenario was higher than that for the RCP 4.5 scenario. However, the increasing trend of average annual precipitation for the RCP 4.5 scenario was higher than that for the RCP 8.5 scenario in Suwon, although the average annual precipitation for the future period for the RCP 8.5 scenario was higher. The uncertainties of bias-corrected precipitation for future climate scenarios are shown in Figure 17.

To assess the uncertainty of future precipitation, the annual precipitation variability was analyzed for the entire future period (2011–2100) and for each 30-year period. The uncertainty of future precipitation was slightly higher than in the past observation period. Comparing the results of the scenarios, the average annual precipitation of RCP 8.5 was larger, and the uncertainty was also slightly larger.

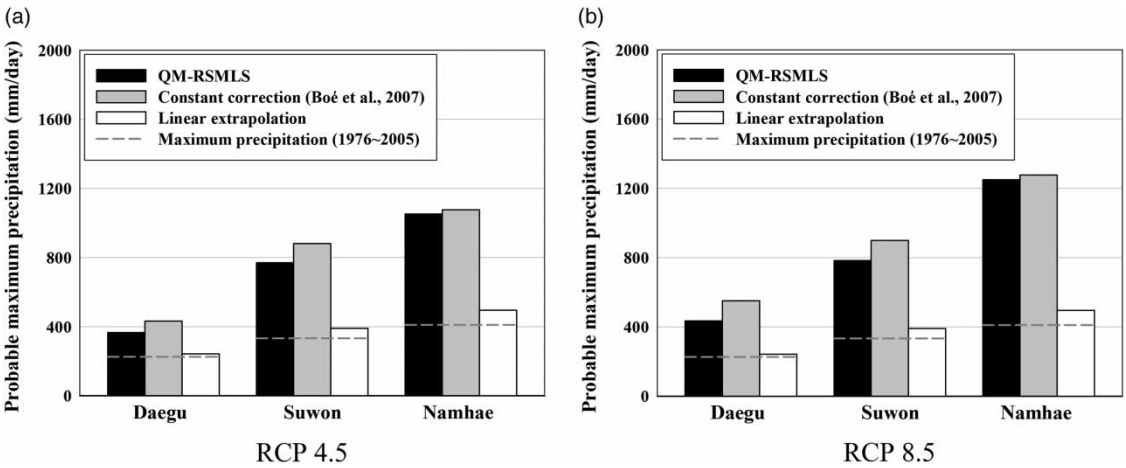

Figure 15 | Estimation of maximum precipitation in the future period.

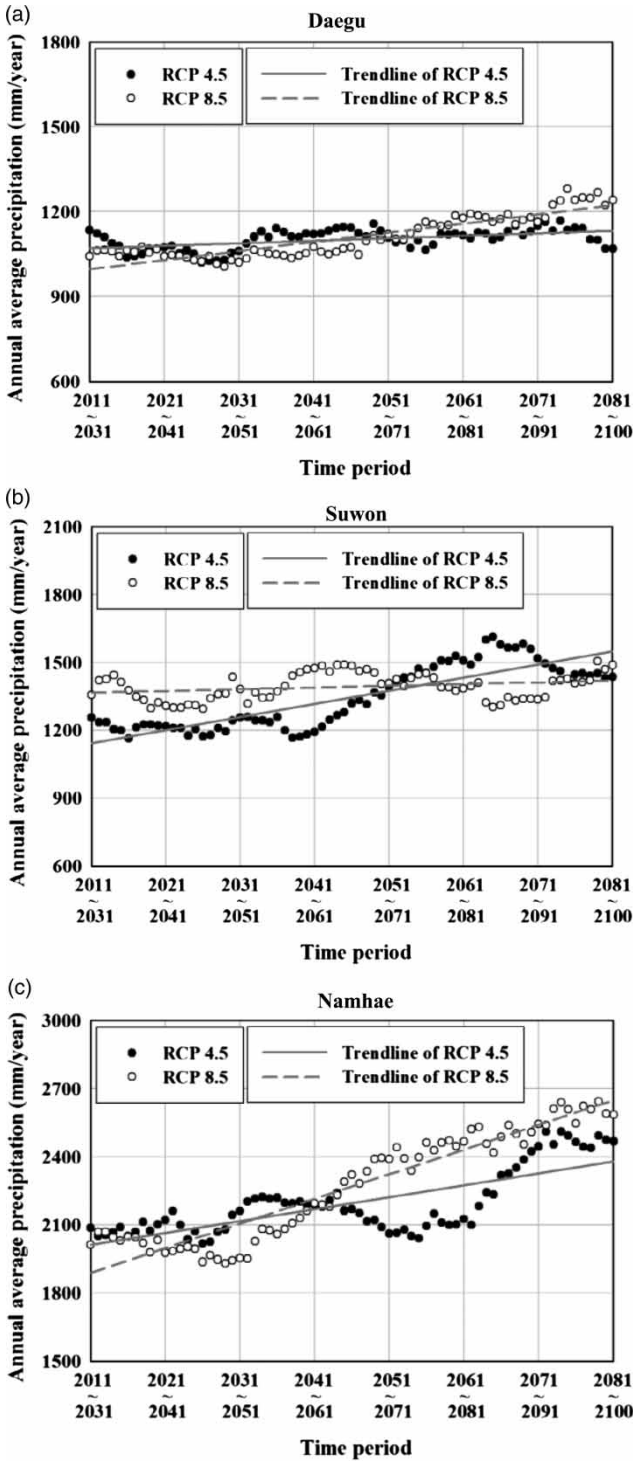

Figure 16 | Trends of the average annual precipitation in the future scenarios.

However, there was no significant difference. Comparing the uncertainty of the areas, Namhae shows the largest uncertainty for the past observation period.

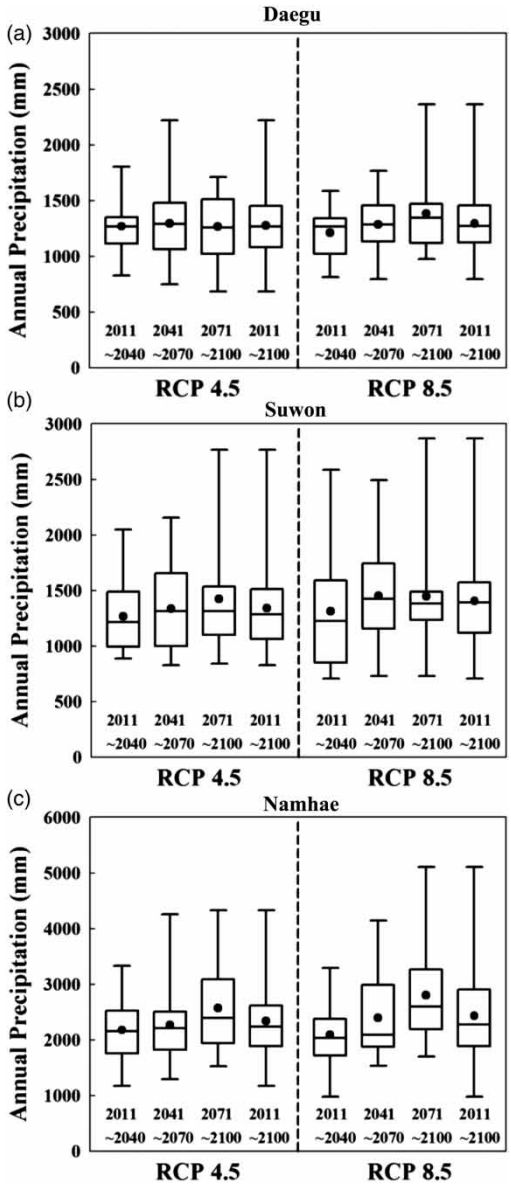

Figure 17 | Variability of future average annual precipitation.

CONCLUSIONS

In this study, the QM-RSMLS method, which is a nonparametric quantile mapping method, was proposed. Compared to the conventional nonparametric quantile mapping methods, the QM-RSMLS is a nonlinear transformation, and quantile division or a correction table is not required. The empirical CDFs are generated by the RSF, which is automatically changed according to the approximation

precipitation; this approach guarantees a certain number of setting sample points. The RSF was modified using a transcendental function instead of polynomials. Therefore, the non-linearity between the percentiles can be considered, and overfitting of the model does not occur. In addition, the proposed method can estimate the new extreme precipitation via extrapolation while considering the trend of the extreme precipitation distribution. These advantages of the proposed method may also be effectively applied to generate the CDF of precipitation in areas with insufficient climate data, such as in developing countries or new weather stations.

The proposed approach was applied to the RCP-based RCM precipitation of three weather stations with different average annual precipitations in Korea. The resulting Q-Q plots revealed a good estimation of the entire range of the daily precipitation distribution. Compared to other quantile mapping methods, the proposed method exhibited the best performance. The total MAEs of the three areas were small, approximately 0.05 mm/day, and the MAE in each probability interval also exhibited small values. In particular, for high precipitation, the proposed method exhibited good performance compared to other quantile mapping methods. Therefore, the QM-RSMLS can effectively reduce the biases of the entire range and is also suited for predicting extreme precipitation events. K-fold cross validation was also conducted to assess the performance of the downscaling models for the independent period. The proposed method exhibited a relatively small error, and this error depends on the stationarity of data. For future RCM precipitation of the two scenarios, the average annual precipitations were increased compared to the past, with the average annual precipitation for the RCP 8.5 exhibiting a larger value than that of the RCP 4.5. A moving average was used to analyze the trends of change in average annual precipitation, and it was found to increase with time in all three areas. Notably, Namhae showed the greatest increasing trend of average annual precipitation.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF), grant funded by the Korean Government (MSIP) (NRF-2017R1C1B5017866 and No. 2016010074).

REFERENCES

Ahmed, K. F., Wang, G., Silander, J., Wilson, A. M., Allen, J. M., Horton, R. & Anyah, R. 2013 Statistical downscaling and bias correction of climate model outputs for climate change impact assessment in the U.S. northeast. *Glob. Planet. Change* **100**, 320–332.

Asadieh, B. & Krakauer, N. Y. 2015 Global trends in extreme precipitation: climate models versus observations. *Hydrol. Earth. Syst. Sci.* **19**, 877–891.

Benestad, R. E. 2010 Downscaling precipitation extremes. *Theor. Appl. Climatol.* **100**, 1–21.

Biniyam, Y. & Kemal, A. 2017 The impacts of climate change on rainfall and flood frequency: the case of Hare Watershed, Southern Rift Valley of Ethiopia. *J. Earth Sci. Clim. Change.* **8**, 383.

Boé, J., Terray, L., Habets, F. & Martin, E. 2007 Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. *Int. J. Climatol.* **27**, 1643–1655.

Box, G. P. & Draper, N. R. 1987 *Empirical Model-Building and Response Surface*. Wiley, New York.

Cannon, A. J. 2011 Quantile regression neural networks: implementation in R and application to precipitation downscaling. *Comp. Geosci.* **37** (9), 1277–1284.

Chen, S. T., Yu, P. S. & Tang, Y. H. 2010 Statistical downscaling of daily precipitation using support vector machines and multivariate analysis. *J. Hydrol.* **385**, 13–22.

Chen, C., Haerter, J. O., Hagemann, S. & Piani, C. 2011 On the contribution of statistical bias correction to the uncertainty in the projected hydrological cycle. *Geophys. Res. Lett.* **38**, 1–6.

Chen, J., Brissette, F. P., Chaumont, D. & Braun, M. 2013 Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. *Water Resour. Res.* **49**, 4187–4205.

Christensen, J. H., Boberg, F., Christensen, O. B. & Lucas-Picher, P. 2008 On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophys. Res. Lett.* **35**, L20709.

Fowler, H. J., Ekström, M., Blenkinsop, S. & Smith, A. P. 2007 Estimating change in extreme European precipitation using a multimodel ensemble. *J. Geophys. Res.* **112**, D18104.

Goyal, M. K., Ojha, C. S. P. & Burn, D. H. 2012 Nonparametric statistical downscaling of temperature, precipitation and evaporation for semi-arid region in India. *J. Hydrol. Eng.* **17**, 615–627.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E. & Skaugen, T. E. 2012 Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods. *Hydrol. Earth Syst. Sci.* **16**, 3383–3390.

Haerter, J. O., Hagemann, S., Moseley, C. & Piani, C. 2011 Climate model bias correction and the role of timescales. *Hydrol. Earth Syst. Sci.* **15**, 1065–1079.

Hay, L. E. & Clark, M. P. 2003 Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western United States. *J. Hydrol.* **282**, 56–75.

Hu, Y., Maskey, S. & Uhlenbrook, S. 2012 Downscaling daily precipitation over the Yellow River source region in China: a comparison of three statistical downscaling methods. *Theor. Appl. Climatol.* **112**, 447–460.

Kallache, M., Vrac, M., Naveau, P. & Michelangeli, P. A. 2011 Nonstationary probabilistic downscaling of extreme precipitation. *J. Geophys. Res.* **116**, D05113.

Lafon, T., Dadson, S., Buysa, G. & Prudhomme, C. 2013 Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods. *Int. J. Climatol.* **33**, 1367–1381.

Lancaster, P. & Salkauskas, K. 1986 *Curve and Surface Fitting. An Introduction*. Academic Press, San Diego.

Maraun, D. 2013 Bias correction, quantile mapping, and downscaling: revisiting the inflation issue. *J. Clim.* **26**, 2137–2143.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themessl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M. & Thiele-Eich, I. 2010 Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys.* **48** (3), RG3003.

Maurer, E. P. & Pierce, D. W. 2014 Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean. *Hydrol. Earth Syst. Sci.* **18**, 915–925.

Mearns, L. O., Bogardi, I., Giorgi, F., Matyasovszky, I. & Palecki, M. 1999 Comparison of climate change scenarios generated from regional climate model experiments and statistical downscaling. *J. Geophys. Res.* **104**, 6603–6621.

Murphy, J. M. 1999 An evaluation of statistical and dynamical techniques for downscaling local climate. *J. Clim.* **12**, 2256–2284.

Nasseri, M., Tavakol-Davani, H. & Zahraie, B. 2013 Performance assessment of different data mining methods in statistical downscaling of daily precipitation. *J. Hydrol.* **492**, 1–14.

Panofsky, H. A. & Brier, G. W. 1968 *Some Applications of Statistics to Meteorology*. The Pennsylvania State University, University Park, USA.

Rajczak, J., Kotlarski, S. & Schär, C. 2016 Does quantile mapping of simulated precipitation correct for biases in transition probabilities and spell lengths? *J. Climate.* **29**, 1605–1615.

Raje, D. & Mujumdar, P. P. 2011 A comparison of three methods for downscaling daily precipitation in the Punjab region. *Hydrol. Process.* **25** (23), 3575–3589.

Shepard, D. D. 1968 A two dimensional interpolation function for irregularly spaced data. In: *Proc 23rd ACM Nat Conf*. ACM, New York, pp. 517–524.

Tareghian, R. & Rasmussen, P. F. 2013 Statistical downscaling of precipitation using quantile regression. *J. Hydrol.* **487**, 122–135.

Teutschbein, C. & Seibert, J. 2012 Bias correction of regional climate model simulations for hydrological climate-change impact studies: review and evaluation of different methods. *J. Hydrol.* **456–457**, 12–29.

Themeßl, M. J., Gobiet, A. & Leuprecht, A. 2011 Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *Int. J. Climatol.* **31**, 1530–1544.

Westra, S., Alexander, L. V. & Zwiers, F. W. 2013 Global increasing trends in annual maximum daily precipitation. *J. Clim.* **26**, 3904–3918.

Wood, A. W., Maurer, E. P., Kumar, A. & Lettenmaier, D. P. 2002 Long-range experimental hydrologic forecasting for the eastern United States. *J. Geophys. Res.* **107** (D20), 4429.

Zorita, E. & Von Storch, H. 1999 The analog method as a simple statistical downscaling technique: comparison with more complicated methods. *J. Clim.* **12**, 2474–2489.

First received 14 February 2017; accepted in revised form 14 October 2017. Available online 28 November 2017