

Nonparametric statistical temporal downscaling of daily precipitation to hourly precipitation and implications for climate change scenarios



Taesam Lee^{a,1}, Changsam Jeong^{b,*}

^aDepartment of Civil Engineering, ERI, Gyeongsang National University, 501 Jinju-daero, Jinju, Gyeongnam 660-701, Republic of Korea

^bDepartment of Civil & Environmental Engineering, Induk University, Seoul, Republic of Korea

ARTICLE INFO

Article history:

Received 23 April 2013

Received in revised form 9 September 2013

Accepted 15 December 2013

Available online 21 December 2013

This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Uwe Haberlandt, Associate Editor

SUMMARY

Hydro-meteorological time series on finer temporal scales, such as hourly, are essential for assessing the hydrological effects of land use or climate change on medium and small watersheds. However, these time series are, in general, available at no finer than daily time intervals. An alternative method of obtaining finer time series is temporal downscaling of daily time series to hourly time series. In the current study, a temporal downscaling model that combines a nonparametric stochastic simulation approach with a genetic algorithm is proposed. The proposed model was applied to Jinju station in South Korea for a historical time period to validate the model performance. The results revealed that the proposed model preserves the key statistics (i.e., the mean, standard deviation, skewness, lag-1 correlation, and maximum) of the historical hourly precipitation data. In addition, the occurrence and transition probabilities are well preserved in the downscaled hourly precipitation data. Furthermore, the RCP 4.5 and RCP 8.5 climate scenarios for the Jinju station were also analyzed, revealing that the mean and the wet-hour probability (P1) significantly increased and the standard deviation and maximum slightly increased in these scenarios. The magnitude of the increase was greater in RCP 8.5 than RCP 4.5. Extreme events of different durations were also tested. The downscaled hourly precipitation adequately reproduced the statistical behavior of the extremes of the historical hourly precipitation data for all durations considered. However, the inter-daily relation between the 1st hour of the present day and the last hour of the previous day was not preserved. Overall, the results demonstrated that the proposed temporal downscaling model is a good alternative method for downscaling simulated daily precipitation data from weather generators or RCM outputs.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

General climate models (GCMs) and their spatially downscaled versions, regional climate models (RCMs), provide superb areal coverage (Orlowsky et al., 2008; Kendon et al., 2010). Hydrologists employ the outputs of these models to characterize the interaction between the land surface and the atmosphere and to assess the hydrological effects of climate change (Prudhomme and Nick Reynard, 2002). While time series data are essential for assessing the hydrological effects of climate change on medium- or small-sized watersheds, time series of hydro-meteorological variables of interest such as precipitation (or rainfall) are not always available at the desired time interval (Alfieri et al., 2012). Therefore, hydro-meteorological data with time scales on the order of 1 h or less are urgently required (Krajewski et al., 1991).

Weather generators, one of the popular statistical downscaling methods, can provide daily outputs. In the literature, weather generators for statistical downscaling are used to simulate the time series of daily hydro-meteorological variables (Soltani and Hoogenboom, 2003; Mason, 2004; Boe et al., 2007; Lee et al., 2012). However, physical rainfall-runoff simulations with daily rainfall underestimate the magnitude of runoff by averaging out the characteristics of short and intense rainfall (Eagleson, 1978). To obtain precipitation time series with finer temporal resolution, the creation of hourly time series from daily time series using disaggregation rules such as preserving the diurnal cycle and the additive condition is recommended (Jones and Harpham, 2009).

Several approaches for statistical temporal downscaling of precipitation time series have been suggested in the literature, including methods based on a point-process model (Rodríguez-Iturbe et al., 1988; Glasbey et al., 1995; Koutsoyiannis and Onof, 2001; Marani and Zanetti, 2007). Koutsoyiannis and Onof (2001) used the Bartlett-Lewis model to simulate rainfall time series by

* Corresponding author. Tel.: +82 2 950 7567; fax: +82 2 950 7579.

E-mail addresses: tae3lee@gnu.ac.kr (T. Lee), csjeong@induk.ac.kr (C. Jeong).

¹ Tel.: +82 55 772 1797; fax: +82 55 772 1799.

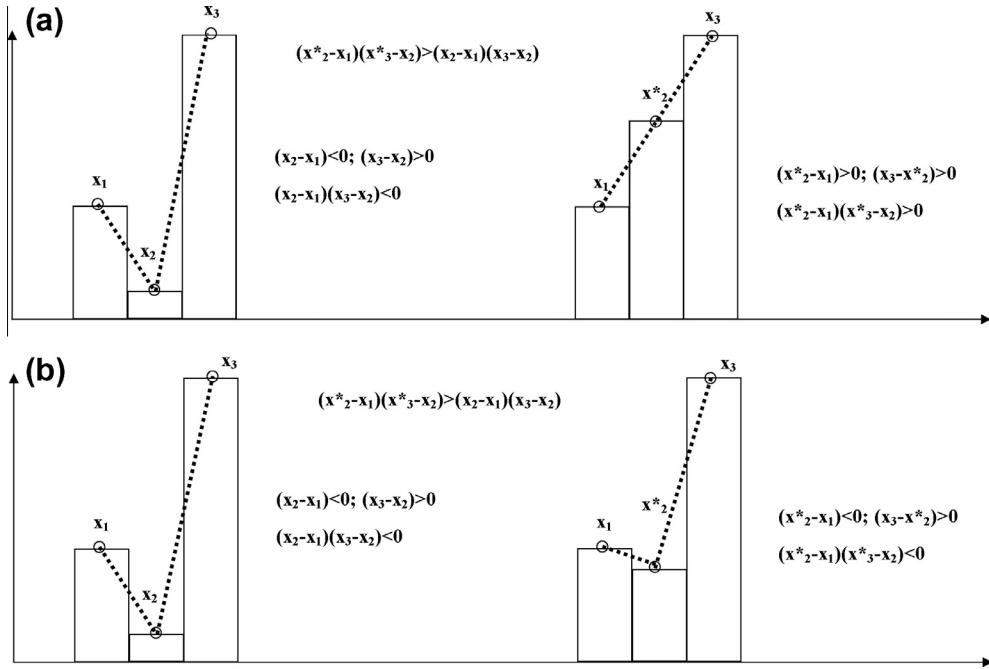


Fig. 1. Selection of crossover element with respect to the rule of gradual variation in a precipitation event. Note that for both (a and b) cases, it is supposed to substitute x_2 into x_2^* since x_2^* more likely follows the rule of precipitation gradual variation.

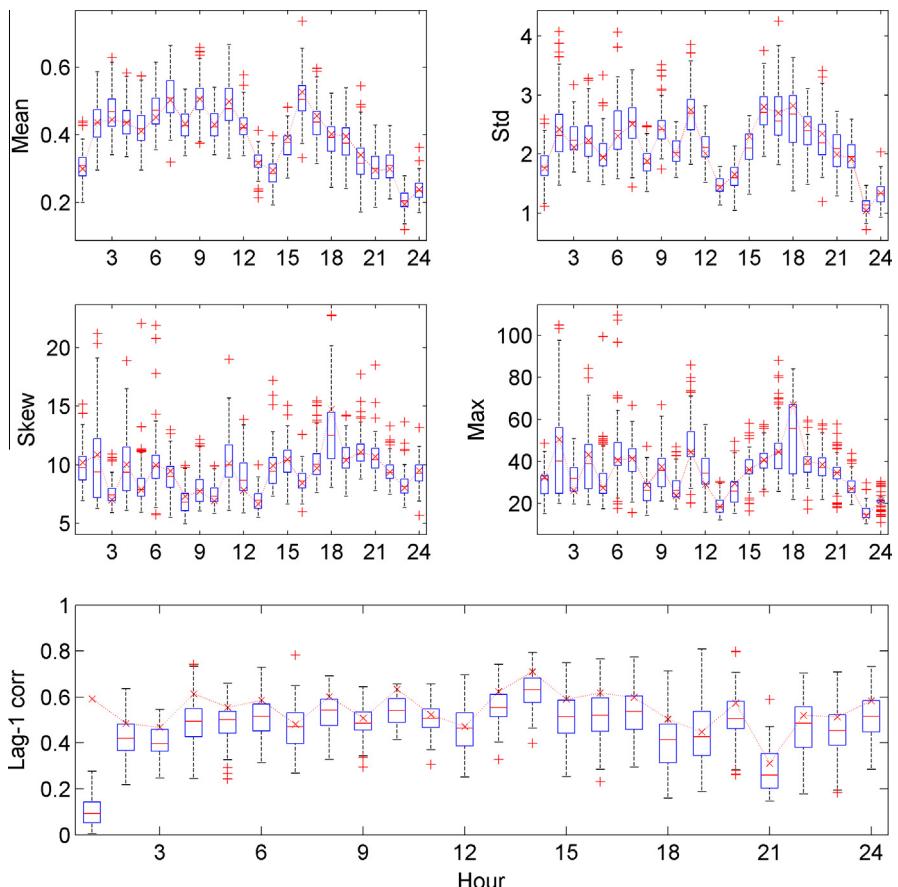


Fig. 2. Key statistics of the historical data (\times) and downscaled data (boxplots) with the proposed model without considering the inter-daily connection for month 8 of Jinju station. The employed values of the GA mixture parameters, P_c and P_m are 0.3 and 0.1, respectively.

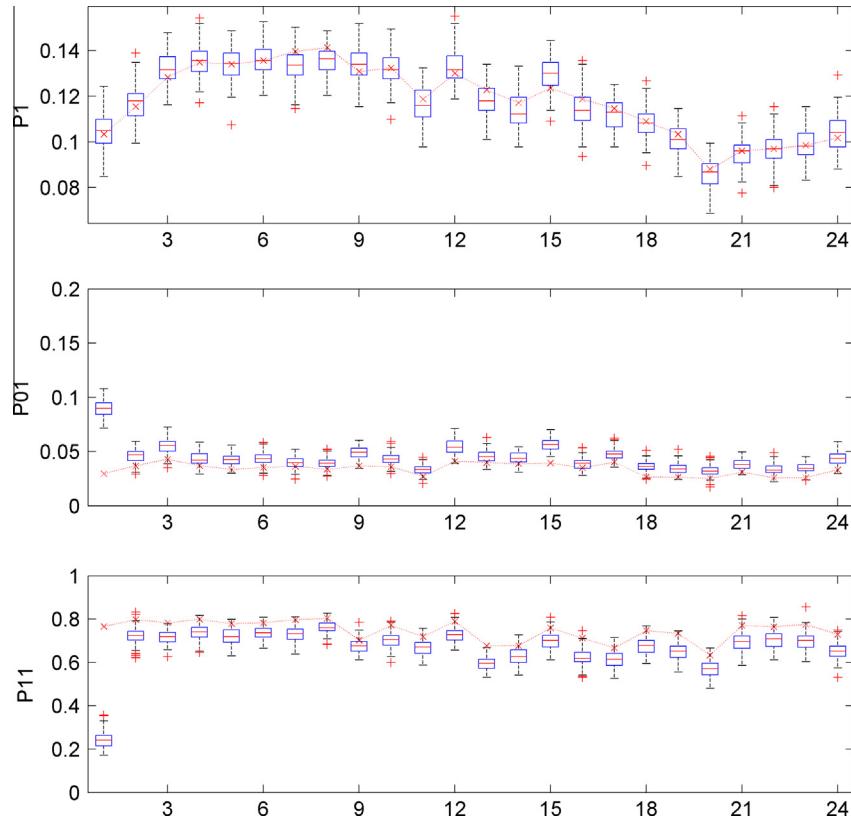


Fig. 3. Occurrence (P_1) & transition probabilities (P_{01} and P_{11}) of historical (− \times −) and downscaled data (boxplots) with the same model as in Fig. 2 for month 8 of Jinju station.

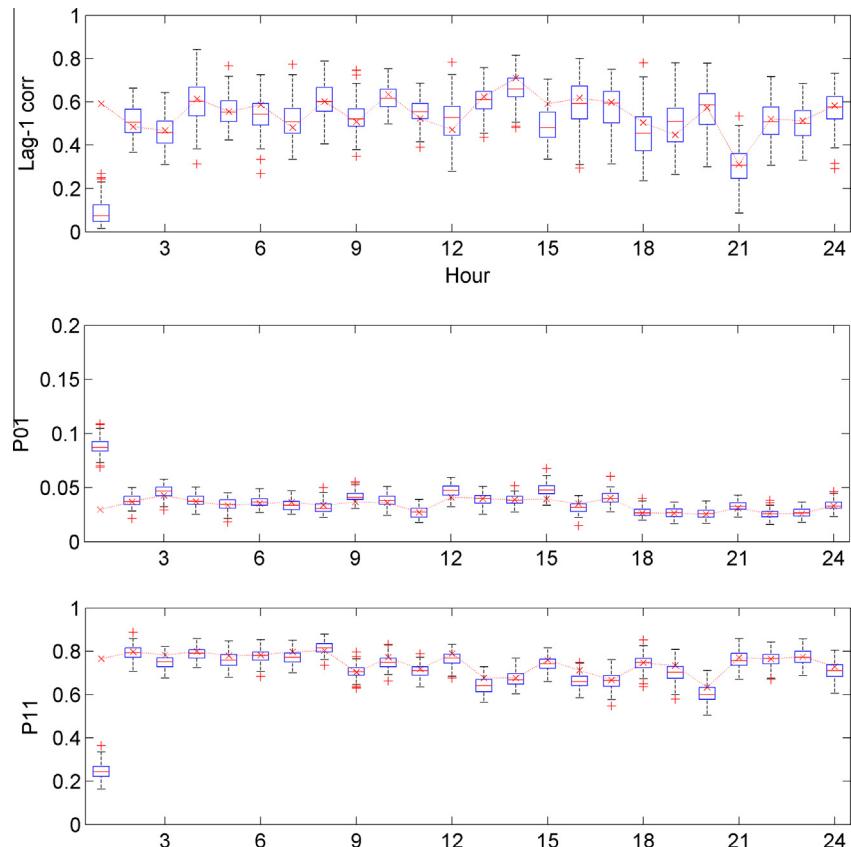


Fig. 4. Lag-1 correlation and transition probabilities of historical (− \times −) and downscaled data (boxplots) with the same model as in Fig. 2 for month 8 of Jinju station except that the tuned P_c (=0.1) and P_m (=0.01) are used.

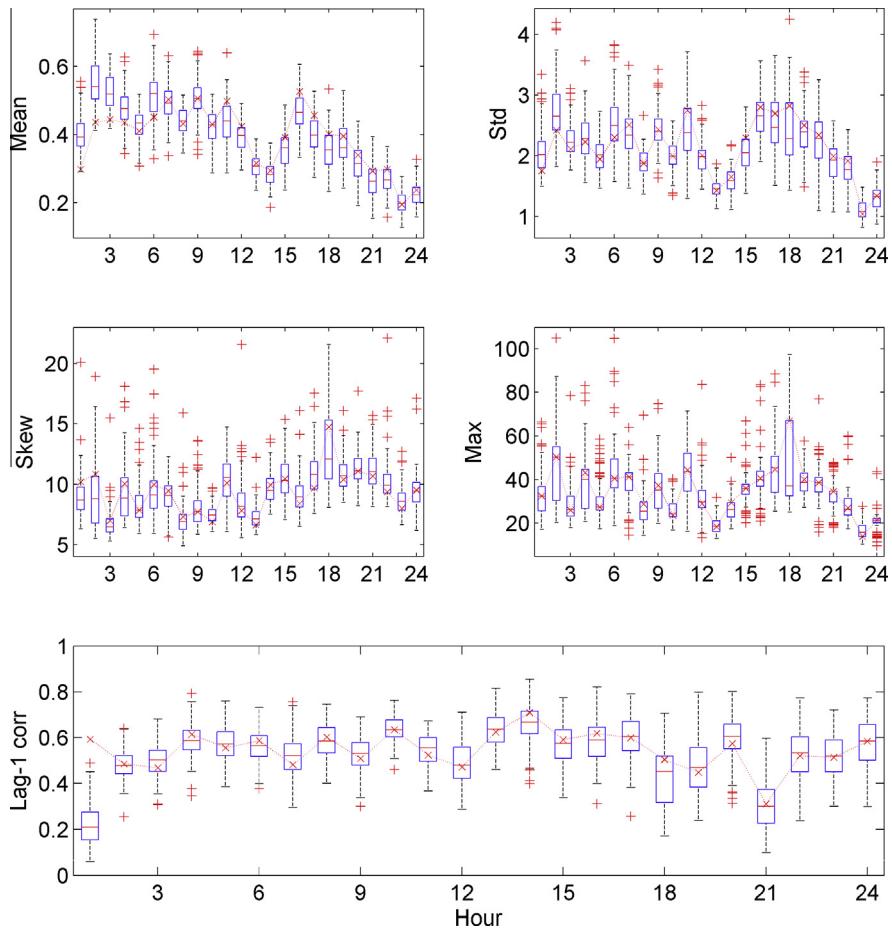


Fig. 5. Key statistics of the historical data ($-x-$) and downscaled data (boxplot) with the proposed model considering inter-daily connection for month 8 of Jinju station.

adjusting simulated data to meet the additive condition. Marani and Zanetti (2007) presented a temporal downscaling method based on a point-process model that employs theoretically based estimates of rainfall variability on an hourly scale derived from daily statistics to produce approximately unbiased estimates of rainfall variance on an hourly time scale.

Machine learning techniques (e.g., neural networks) have also been applied to temporally downscale precipitation data (Coulibaly et al., 2005; Kumara et al., 2012). Kumara et al. (2012) downscaled multiple meteorological variables from monthly to sub-daily while retaining consistent correlations between variables.

In addition, some naïve methods have been introduced (Kondo and Xu, 1997; Wang and Liu, 2006; Yang et al., 2009; Chen et al., 2012). Kondo and Xu (1997) suggested that the hourly rainfall amount follows a sinusoidal distribution in which the peak density values and the last time of rainfall are proportional to the daily rainfall amounts. Wang and Liu (2006) employed a random distribution method, but this method overestimates the amount of rainfall when the rainfall intensity is high. Chen et al. (2011) suggested a proportional method in which the daily rainfall amount is divided into four periods and distributed proportionally. These methods are direct and easy to use. However, they cannot reproduce key hourly statistics (e.g., the hourly mean, standard deviation, and skewness) and the intermittency of rainfall events.

One of the most important characteristics of hourly precipitation is the diurnal cycle of the local and global weather and climate. For example, summertime rainfall may evaporate more quickly in the afternoon hours but can easily permeate into the soil and be stored overnight. Specifically, Dai et al. (1999) argued that over the Rocky Mountains, precipitation is twice as likely to occur

in approximately the 15–18th hours of the day (i.e., 3 PM–6 PM) than at any other times. The diurnal cycle can strongly influence the hydrological cycle and agriculture (Yin et al., 2009; He and Zhang, 2010) as well as the surface air temperature range (Dai, 1999). Reproduction of the diurnal cycle and the key statistics of hourly precipitation data are crucial for exploring future climate scenarios and assessing the impacts of climate change for water resource management and flood control.

However, these existing models do not consider the diurnal cycle and the specific details of key hourly statistics such as the mean, standard deviation, skewness, and lag-1 correlation for downscaled hourly precipitation data. Lee et al. (2010) proposed a nonparametric disaggregation model and applied it to the Colorado River streamflow to disaggregate the yearly streamflow to monthly streamflow. They adapted a meta-heuristic optimization technique, a genetic algorithm (GA) (Goldberg, 1989), and k-nearest neighbor resampling (KNNR) for nonparametric disaggregation. KNNR is a nonparametric resampling technique that has been employed for stochastic simulations, particularly for hydro-meteorological variables (Lee et al., 2007; Salas and Lee, 2010). GAs are a class of adaptive stochastic optimization algorithms and a programming technique that mimics biological evolution as a problem-solving strategy; these algorithms are commonly used in search and optimization. The GA process contains three operators, reproduction, crossover and mutation. These three GA operators are employed in nonparametric resampling to mix the historical patterns and subsequently obtain new patterns.

Therefore, the objective of the current study is to propose a nonparametric temporal downscaling model to reproduce statistical characteristics on an hourly time scale and the hourly statistical

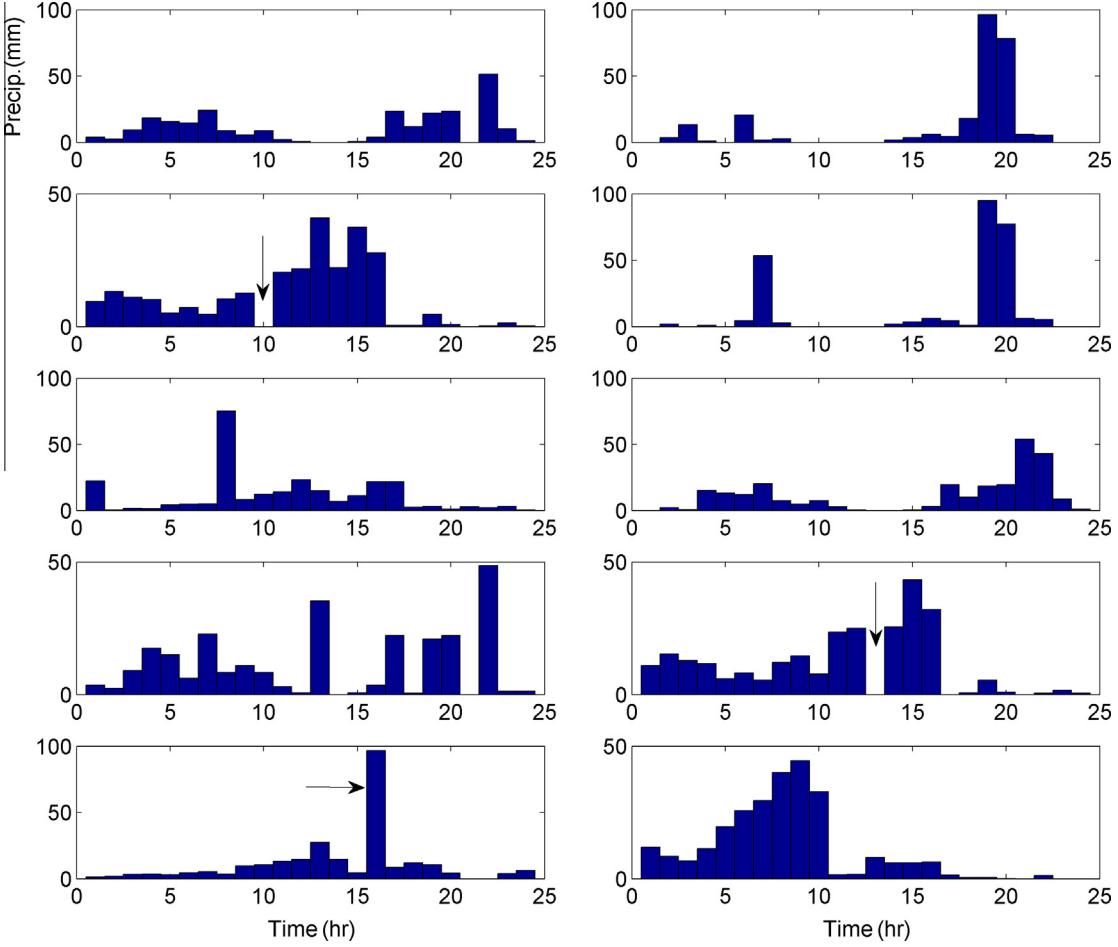


Fig. 6. Examples of downscaled hourly precipitation without considering the gradual variation in Eq. (5) for month 8 of Jinju station.

pattern for a day (i.e., the diurnal cycle) based on the temporal downscaling method suggested by Lee et al. (2010).

The temporal disaggregation of yearly to monthly streamflow is very analogous to the temporal downscaling of daily to hourly precipitation because, in both cases, the lower temporal scale exhibits cyclic oscillation. The primary difference is the strong intermittency in the hourly precipitation time series. In addition, hourly precipitation events tend to increase and decrease gradually.

This paper is organized as follows. A mathematical description of the proposed model is presented in Section 2. A detailed explanation of the weather data employed is presented in Section 3. The results of the application of the proposed model are illustrated in Section 4. Finally, the summary and conclusions are presented in Section 5.

2. Methodology

Employing KNNR and the GA mixture, the proposed approach to downscale daily time series to hourly time series is presented in the following subsections.

2.1. Data-based temporal downscaling using a GA

Consider the daily, y_i , and hourly observations $\mathbf{x}_i = [x_{i,1}, x_{i,2}, \dots, x_{i,24}] = [x_{i,h}]_{h \in \{1,24\}}$ and $i = 1, \dots, n$, where n is the record length and h indicates the h th hour. In addition, Y_t is the target daily precipitation value for time t for $t = 1, \dots, T$, where T is the length of the target

daily precipitation. The objective is to downscale the daily time series Y_t to the hourly time series $\mathbf{X}_t = [X_{t,1}, X_{t,2}, \dots, X_{t,24}]$.

Provided that the number of nearest neighbors, k , is known, the temporal downscaling procedure is as follows.

- (1) Estimate the distances between the target daily precipitation Y_t and the observed daily precipitation $y_i = \sum_{h=1}^{24} x_{i,h}$. Here, the distances are measured for $i = 1, \dots, n$ as

$$D_i = (Y_t - y_i)^2 \quad (1)$$

- (2) Arrange the estimated distances from step (1) in ascending order, select the first k distances (i.e., the smallest k values), and reserve the time indices of the smallest k distances.
- (3) Randomly select one of the stored k time indices with the weighting probability given by

$$w_m = \frac{1/m}{\sum_{j=1}^k 1/j}, \quad m = 1, \dots, k \quad (2)$$

- (4) Assign the hourly values of the selected time index from step (3) as $\mathbf{x}_p = [x_{p,h}]_{h \in \{1,24\}}$. Here, it is assumed that the selected time index is p .

- (5) Execute the following steps for GA mixing:
 - (5-1) Reproduction: select one additional time index using steps (1) through (4) and denote this index as p^* . Obtain the corresponding hourly precipitation values, $\mathbf{x}_{p^*} = [x_{p^*,h}]_{h \in \{1,24\}}$. The subsequent two GA operators employ the two selected vectors, \mathbf{x}_p and \mathbf{x}_{p^*} .

- (5-2) Crossover: replace each element $x_{p,h}$ with $x_{p^*,h}$ at probability P_c , i.e.,

$$X_{t,h}^* = \begin{cases} x_{p^*,h} & \text{if } \varepsilon < P_c \\ x_{p,h} & \text{otherwise} \end{cases},$$

where ε is a uniformly distributed random number between 0 and 1.

- (5-3) Mutation: replace each element (i.e., each hour, $h = 1, \dots, 24$) with one selected from all the observations of this element for $i = 1, \dots, n$ (hourly precipitation data) with probability P_m , i.e.,

$$X_{t,h}^* = \begin{cases} x_{i,h} & \text{if } \varepsilon < P_m \\ x_{p,h} & \text{otherwise} \end{cases},$$

where $x_{i,h}$ is selected from $[x_{i,h}]_{i \in \{1, \dots, n\}}$ with equal probability for $i = 1, \dots, n$.

- (6) Adjust the GA mixed hourly values as follows to preserve the additive condition:

$$X_{t,h} = \frac{X_{t,h}^*}{\sum_{j=1}^{24} X_{t,j}^*} Y_t, \quad (3)$$

where $X_{t,h}$ is the resampled hourly time series including the GA mixture process, $h = 1, \dots, 24$.

- (7) Repeat steps (1–5) until the required data are generated.

The roles of the crossover probability P_c and the mutation probability P_m were studied by Lee et al. (2010). In the current study, these two probabilities were employed as tuning parameters to improve

the reproduction of the historical statistics for the generated hourly time series. The selection of the number of nearest neighbor (k) has been studied (Lall and Sharma, 1996; Lee and Ouarda, 2011). The most common and simplest selection method was applied in the current study by setting $k = \sqrt{n}$. This heuristic approach has often been used in stochastic simulations with KNNR (Lall and Sharma, 1996; Rajagopalan and Lall, 1999; Prairie et al., 2006; Lee et al., 2010; Lee and Ouarda, 2012).

2.2. Consideration of inter-day connections

The temporal downscaling of daily precipitation is performed without consideration of the previous day. However, hourly precipitation between the last hour of the previous day, $X_{t-1,24}$, and the first hour of the following day, $X_{t,1}$, is almost certainly connected. To account for this inter-day relation, the following procedure was tested in the current study.

If $X_{t-1,24} > 0$ and $Y_t > 0$, then select the possible k time indices from the observations only when $x_{i-1,24} > 0$ and $y_i > 0$ for steps (1 and 2) of the approach presented in the previous section. The following distance measurement is employed as a substitute for the following equation:

$$D_i = \mathbf{Z}^T \mathbf{C}^{-1} \mathbf{Z}, \quad (4)$$

where $\mathbf{Z}_i = [Y_t - y_i, X_{t-1,24} - x_{i-1,24}]$ and \mathbf{C}^{-1} is the inverse of the covariance of $[y_i, x_{i-1,24}]$ for $i = 2, \dots, n$. This remediation is proposed to create a similar environment to historical precipitation events in which there is rain during the last hour of the previous day followed by rain in the present day.

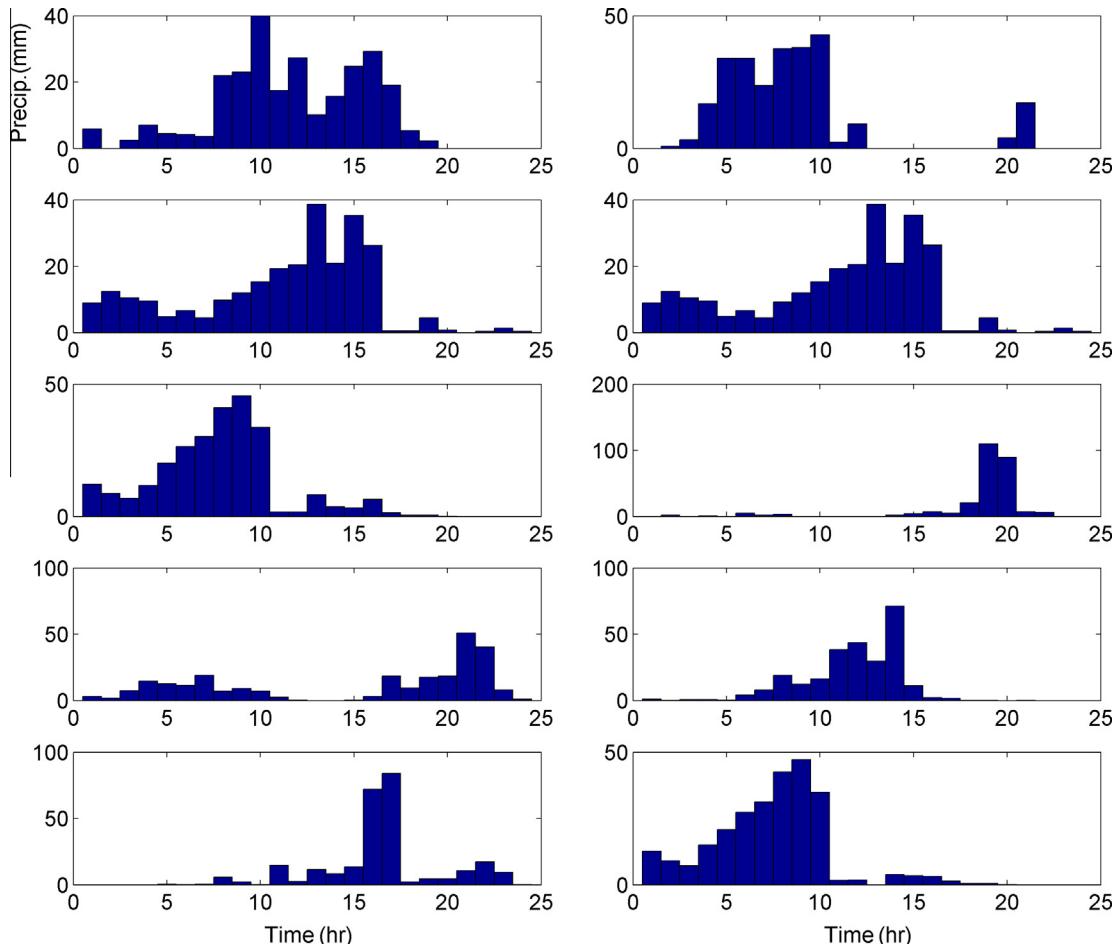


Fig. 7. Examples of downscaled hourly precipitation with considering the gradual variation in Eq. (5) for month 8 of Jinju station.

2.3. Reproduction of the gradual variation in precipitation events

Precipitation tends to increase or decrease gradually during an event. A precipitation event is defined as a period during which precipitation occurs without interruption. When performing the crossover in the downscaling procedure (see steps (5–2) in the previous section), the gradual variation cannot be preserved because of the random replacement of components in an event. To preserve the gradual variation of the precipitation event when selecting $X_{i,h}$, the crossover (steps (5–2)) is performed by substituting $x_{p,h}$ with $x_{p^*,h}$ only when the following condition is met:

$$(X_{i,h-1} - x_{p^*,h})(x_{p^*,h} - x_{p,h+1}) > (X_{i,h-1} - x_{p,h})(x_{p,h} - x_{p,h+1}), \quad (5)$$

where $X_{i,h-1}$ is the generated hourly precipitation of the h th hourly precipitation components and $x_{p,h}$ and $x_{p^*,h}$ are the candidates for $X_{i,h}$. The candidate for $X_{i,h+1}$ is $x_{p,h+1}$ because the p th time index is selected. $X_{i,h+1}$ must be known to evaluate the gradual variation so that the most probable candidate $x_{p,h+1}$ for $X_{i,h+1}$ is employed. The order of the precipitation is $X_{i,h-1}, x_{p^*,h}, x_{p,h+1}$ when the condition of Eq. (5) is met, and it is $X_{i,h-1}, x_{p,h}, x_{p,h+1}$ otherwise. If the precipitation amount of the current event is gradually varied, the magnitude order is $X_{i,h-1} > x_{p,h} > x_{p,h+1}$ or $X_{i,h-1} < x_{p,h} < x_{p,h+1}$. Therefore,

gradual variation of precipitation events can be considered in the crossover procedure if the condition in Eq. (5) is met.

As an example (see Fig. 1), consider three hourly precipitation amounts, x_1, x_2 , and x_3 . Here, we must decide whether to substitute x_2 into x_2^* . In Fig. 1(a), $(x_1 - x_2^*) > 0$ and $(x_1 - x_2) < 0$, whereas $(x_2^* - x_3) > 0$ and $(x_2 - x_3) > 0$. Subsequently, $(x_1 - x_2^*)(x_2^* - x_3) > 0$ and $(x_1 - x_2)(x_2 - x_3) < 0$. Finally, x_2^* is selected because it satisfies Eq. (5) because $(x_1 - x_2^*)(x_2^* - x_3) > (x_1 - x_2)(x_2 - x_3)$. In Fig. 1(b), $(x_1 - x_2^*) < 0$ and $(x_1 - x_2) < 0$, whereas $(x_2^* - x_3) > 0$ and $(x_2 - x_3) > 0$.

Subsequently, $(x_1 - x_2^*)(x_2^* - x_3) < 0$ and $(x_1 - x_2)(x_2 - x_3) < 0$. However, it is obvious that the value of $(x_1 - x_2)(x_2 - x_3)$ is less than the value of $(x_1 - x_2^*)(x_2^* - x_3)$. Finally, x_2^* is selected because it satisfies Eq. (5) because $(x_1 - x_2^*)(x_2^* - x_3) > (x_1 - x_2)(x_2 - x_3)$. The crossover operator with the required condition given in Eq. (5) allows the method to preserve the gradual variation of hourly precipitation events. Note that this crossover operator is randomly performed with probability P_c . Therefore, the simulated hourly precipitation is not always gradually varied, which is consistent with natural patterns. In addition, this random performance is helpful when the hourly precipitation is at a maximum such that $x_{i,h-1} < x_{p,h} > x_{p,h+1}$. No further consideration of this peak condition has been made.

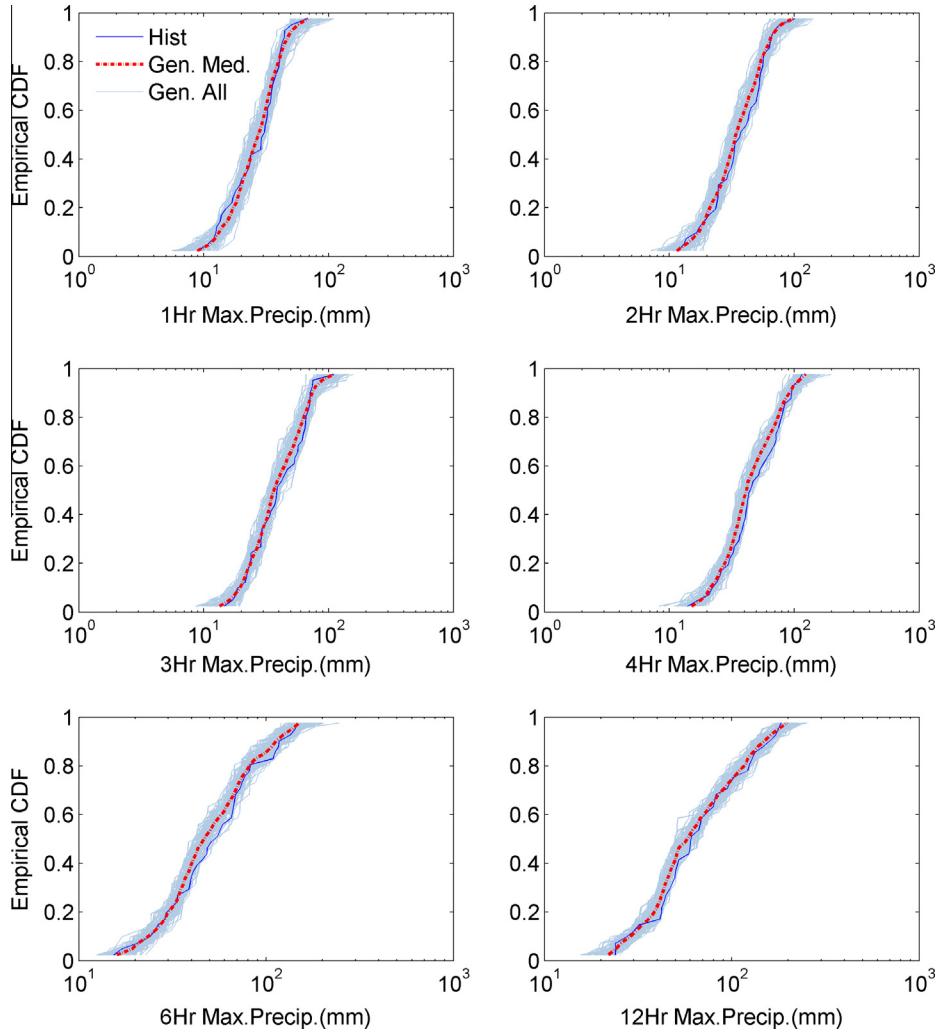


Fig. 8. Exceedance probability of annual maximum precipitation with different duration (i.e. 1,2,4,6,12 h) for historical data (thick solid line) and 100 downscaled series of the proposed model (thick dashed lines) and the median (thin solid line) of the simulated series for month 8 of Jinju station.

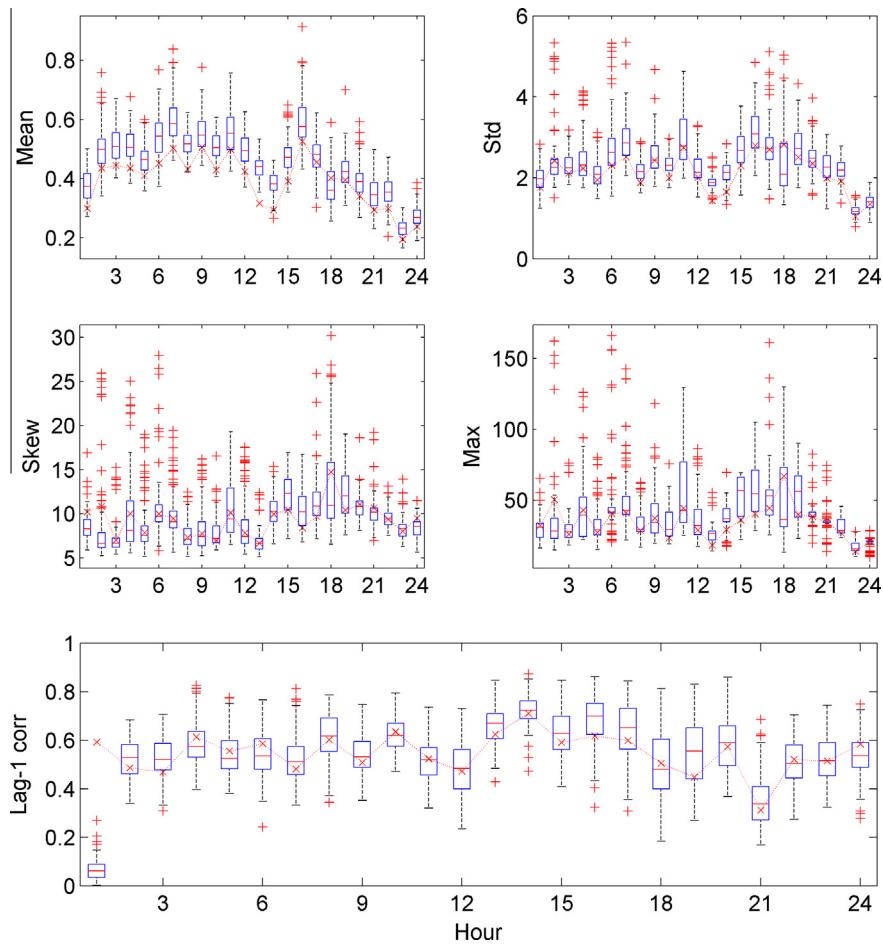


Fig. 9. Key statistics of the historical data (−×−) and downscaled data (boxplots) of RCP 4.5 for year 2011–2099 with the proposed model without considering the inter-daily connection for month 8 of Jinju station. The employed values of the GA mixture parameters, P_c and P_m are 0.1 and 0.01, respectively.

3. Data description and application methodology

In the current study, daily and hourly precipitation data from the Jinju station, which is located in the southern middle region of South Korea (35.16° latitude and 128.04° longitude), covering the time range 1971–2010 were employed to present the performance of the suggested model for temporal downscaling. The Korean peninsula is climatologically influenced by the Siberian air mass during winter and the Maritime Pacific High during summer. Approximately two thirds of the total precipitation is concentrated in the summer. Therefore, the flood control system is an important aspect of this study. In particular, Namgang Dam was constructed in 1969 on the Nam River, a tributary of the Nakdong River, to prevent flood damage and provide irrigation and living water. The dam is located just above the city of Jinju. The dam was reinforced because of continuous flood damage and increasing needs for urbanization (Kim et al., 2009). The weather of the Jinju area is of particular concern because when the discharge from the Namgang Dam is high, the region downstream of the Nakdong River may be easily flooded. Therefore, Jinju station was selected for the current study, and its annual maximum rainfall during different durations was also analyzed for the climate change scenarios. Furthermore, in order to compare the model performance for different locations in South Korea, two stations have been added such as Seoul (37.57° latitude and 126.97° longitude) and Daegu (35.89° latitude and 128.62° longitude) stations.

The climate change scenarios of daily precipitation were provided by the Korea Meteorological Administration (KMA) to the

Intergovernmental Panel on Climate Change (IPCC) Representative Concentration Pathways 8.5 and 4.5 (RCP 8.5 and RCP 4.5) from the year 2011 to the year 2099. KMA produced the regional climate projections for Korea with 12.5-km resolution using the dynamical downscaling method (the atmospheric regional climate model HadGEM3-RA, i.e., Hadley Centre Global Environment Model version 3) from a global climate change projection obtained by using a coupled atmosphere–ocean general circulation model (GCM), HadGEM2-AO (version 2 of the atmosphere–ocean coupled model of the Hadley Centre Global Environment Model), with approximately 135-km resolution for RCP 8.5 and RCP 4.5. The HadGEM3-RA of the RCP projection was bias-corrected using linear scaling method (Lenderink et al., 2007) for the employed KMA weather stations (i.e. Jinju, Seoul and Daegu) on a daily time scale.

As mentioned previously, daily precipitation maxima occur during summer, particularly in August. Therefore, the results for the month of August are displayed. The statistical characteristics of the results of the other months are not presented due to space limitations.

The statistics of the downscaled data are presented using a boxplot in which the box displays the interquartile range (IQR). When the extrema are higher than 1.5IQR, the whiskers extend up to 1.5IQR, and the excess values are displayed with crosses and are considered outliers. Otherwise, the whiskers extend to the extrema and feature horizontal lines at their ends. The horizontal line inside the box depicts the median of the data. In addition, the value of the statistic corresponding to the historical data is represented by a cross connected with a dotted line.

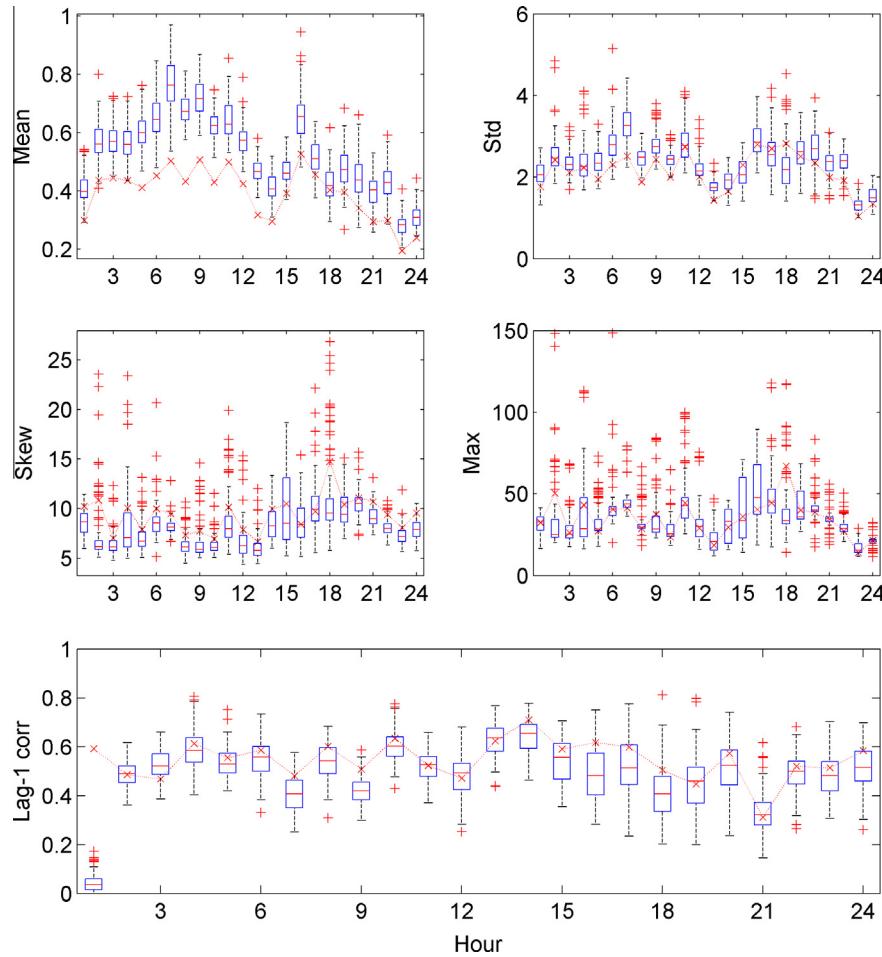


Fig. 10. Key statistics of the historical data ($-x-$) and downscaled data (boxplots) of RCP 8.5 for year 2011–2099 with the same model as in Fig. 9 for month 8 of Jinju station.

4. Results

4.1. Test for the observed time period

The proposed method was tested to determine whether the downscaled hourly precipitation properly reproduced the statistical characteristics of the observed hourly precipitation. A total of 100 series of hourly precipitation were obtained by downscaling the observed daily precipitation for the years 1971–2010 using the proposed nonparametric temporal downscaling model. A few parameters were set, such as the crossover probability P_c , the mutation probability P_m , and the distance measurement. As a default, $P_c = 0.3$ and $P_m = 0.1$, as given by Lee et al. (2010), were used, and the distance measurement of Eq. (1) was also employed (i.e., there was no consideration of inter-day connections). Note that inter-day connections and gradual variations of precipitation events (see Sections 2.2 and 2.3, respectively) are not considered in the default model to investigate the effect of their inclusion, as explained in Sections 2.2 and 2.3. The effects of considering inter-day connections and gradual variations are discussed below.

The key statistics of the downscaled hourly data are compared with the statistics of the observed data, as shown in Fig. 2. The key historical statistics, such as the mean, standard deviation, skewness, and maximum, are reproduced well in the downscaled data. The lag-1 correlation is slightly underestimated in the downscaled data, and the lag-1 correlation of the 1st hour for the downscaled data is significantly underestimated. Note that the minimum is not presented because it is always zero for hourly precipitation.

The occurrence probabilities, such as P_1 (the probability of rain), P_{01} (the probability of rain in an hour adjacent to a clear hour), and P_{11} (the probability of rain on a day adjacent to a rainy day), were also estimated for the historical and downscaled data and compared in Fig. 3. P_1 and P_0 ($1-P_1$) are the occurrence probabilities (or limiting probabilities), whereas P_{01} and P_{11} are the transition probabilities. The results demonstrate that the probability of rain P_1 is well reproduced for all hours, whereas the transition probabilities of P_{01} and P_{11} are somewhat biased. The P_{01} values for the downscaled data are slightly overestimated, whereas the P_{11} are slightly underestimated. These biases of the transition probabilities might be due to the crossover and mutation operations of the GA mixture process.

The effects of the probabilities of crossover (P_c) and mutation (P_m) were extensively tested by evaluating different combinations of the two probabilities. As the probabilities increased, the downscaled data contained more diverse new patterns, but the biases for the lag-1 correlation and transition probabilities (P_{01} and P_{11}) increased. To circumvent these biases while still producing a sufficiently large number of new patterns, P_c and P_m were tuned to be 0.1 and 0.01, respectively. The results in Fig. 4 indicate that the lag-1 correlation and transition probabilities are much better preserved by tuning the two GA parameters (P_c and P_m). These two values (0.1 and 0.01) were employed in the subsequent experiments.

However, the significant biases of the lag-1 correlation and transition probabilities for the 1st hour could not be reduced because the downscaling is performed independently of the

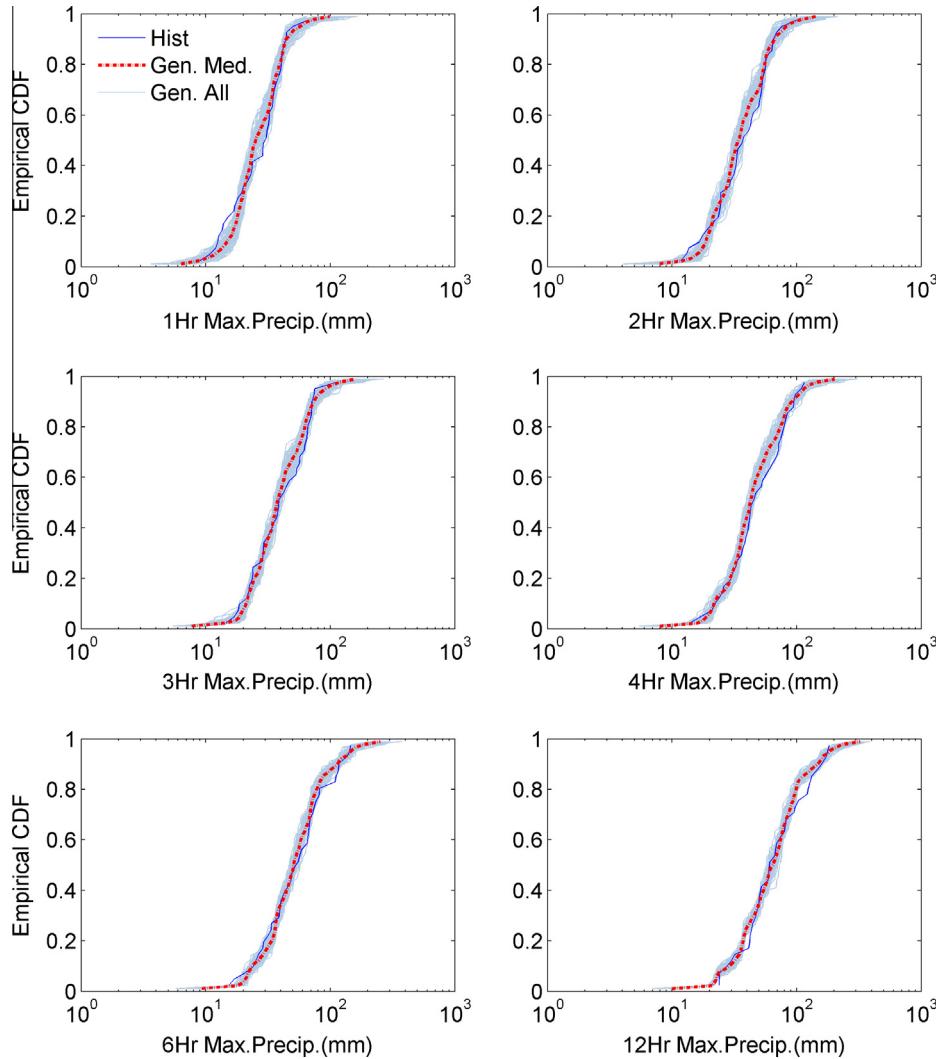


Fig. 11. Exceedance probability of annual maximum precipitation with different duration (i.e. 1, 2, 4, 6, 12 h) for historical data (thick solid line) and 100 downscaled series (thick dashed lines) of RCP 4.5 for year 2011–2099 with the same model as in Fig. 9 and their median (thin solid line) for month 8 of Jinju station.

preceding day. To remedy this bias, inter-day connections were considered by including the term of the last hour's precipitation of the previous day in the distance measurement by using Eq. (4) rather than Eq. (1) (see Section 2.2). This special remedy was only considered when the last hourly precipitation of the previous day was greater than zero (i.e., $X_{t-1,24} > 0$). As mentioned in Section 2.2, the k-neighbors were only selected from those hours that satisfied this condition of $x_{i-1,24} > 0$.

As shown in Fig. 5, the median lag-1 correlation of the 1st hour slightly increased from 0.1 (see Fig. 2) to 0.2. However, there was still a significant underestimation. A slight overestimation of the mean and standard deviation for the first two hours is also evident. The bias of the transition probabilities (P01 and P11) (data not shown) was reduced by accounting for inter-day connections using Eq. (4). A significant improvement of P11 from 0.2 (see Fig. 3) to 0.5 was observed. However, the probabilities were still underestimated. Furthermore, there were some biases in the occurrence probabilities (P1) for all hours. It can be concluded that the remedy to account for inter-day connections was not successful and further improvement is still required. Therefore, the model that does not considering inter-day connections was employed for the subsequent experiment.

Fig. 2 shows that the mean and standard deviation in the approximately 14th and 24th hours are less than the same statistics at the other hours, which suggests a typical pattern of a diurnal cycle (see Fig. 2). The 14th hour is the time when the temperature reaches its daily maximum, particularly during the summer. The diurnal cycle of the mean and standard deviation statistics is more obvious in the afternoon and evening than in the morning. A different cycle is observed for the occurrence probability (P1; see Fig. 3). There is a higher probability of rainy hours during the hours of 03:00–10:00 than during the evening hours of 18:00–24:00. The diurnal cycle is well reproduced by the proposed temporal downscaling model. The reproduction of this diurnal cycle is critical for assessing hydrological and environmental impacts, particularly from climate change.

In Fig. 6, examples of downscaled hourly precipitation for extreme event daily data are presented. As mentioned in Section 2.3, intermittency during a continuous precipitation event (see the left panel in the second row and right panel in the fourth row) and sudden peak values (see the bottom-left panel) in the middle of light precipitation are observed, due to the random crossover operation. To avoid this phenomenon, Eq. (5) was applied in the crossover selection. Fig. 7 illustrates examples of the downscaled data

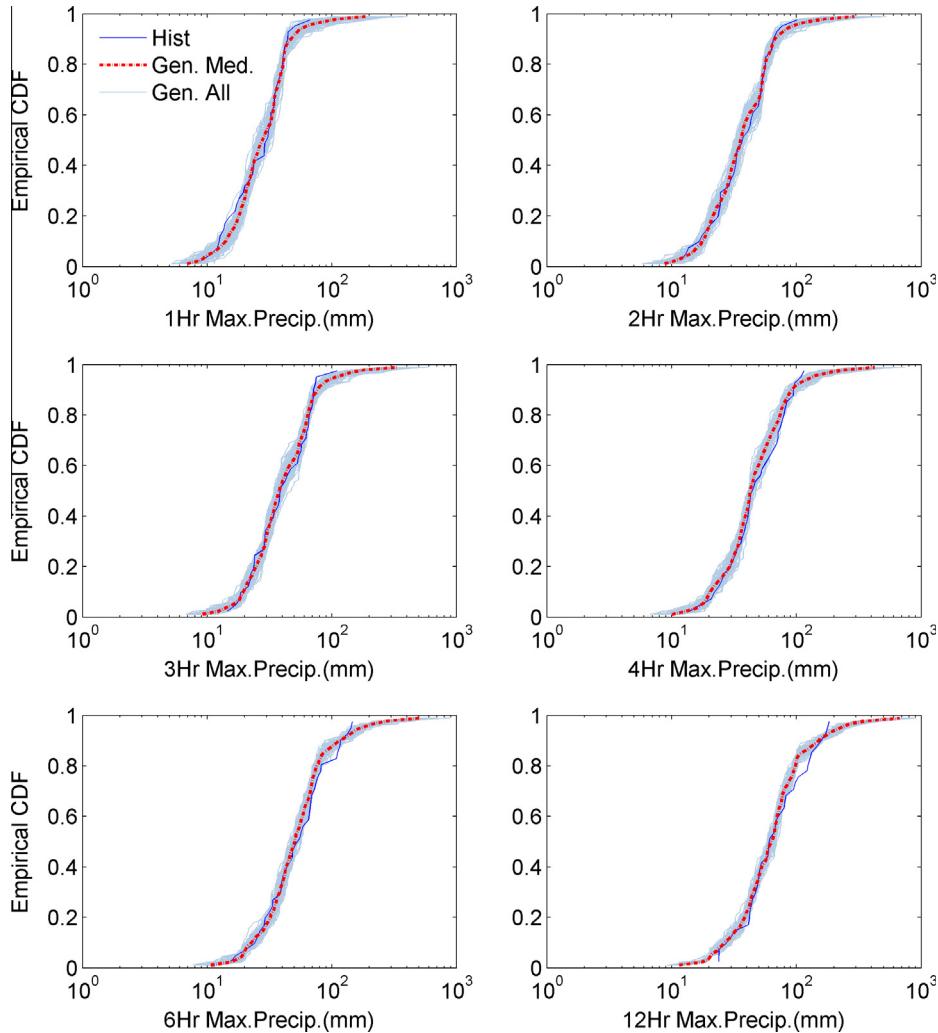


Fig. 12. Exceedance probability of annual maximum precipitation with different duration (i.e. 1, 2, 4, 6, 12 h) for historical data (thick solid line) and 100 downscaled series (thick dashed lines) of RCP 8.5 for year 2011–2099 with the same model as in Fig. 9 and their median (thin solid line) for month 8 of Jinju station.

yielded by employing Eq. (5) in the crossover selection. The downscaled data present no interruption of zero values during continuous hourly precipitation. No significant differences were observed for the key statistics and occurrence probabilities obtained using Eq. (5) for the crossover selection (data not shown).

In Fig. 8, empirical cumulative distribution functions (CDF) for different rainfall durations (1, 2, 3, 4, 6, and 12 h) during the extreme events are presented for the historical data (dark solid line) and 100 downscaled series (light solid lines); their median is also shown (thick dotted line). The results indicate that the downscaled data reproduce the empirical CDFs of the historical data for all of the different durations.

4.2. Temporal downscaling of climate scenarios

The climate scenarios of the RCM daily precipitation data for RCM4.5 and RCM8.5 for Jinju station until the year 2099 were downscaled using the proposed temporal downscaling model. The parameters of $P_c = 0.1$ (crossover probability) and $P_m = 0.01$ (for mutation probability) were employed. In addition, gradual variation of precipitation amounts was applied using Eq. (5), whereas inter-daily connections were not considered.

Fig. 9 present the key statistics of the downscaled data of the RCM 4.5 climate scenario. The mean of the downscaled data (see

Fig. 9) is significantly higher (see the boxplot) than that of the observed data (the dotted line with a cross), whereas the standard deviation and maximum of the downscaled data are slightly higher. No difference in the lag-1 correlation can be observed between the historical data and the downscaled data from RCM4.5. Note that the underestimation of the 1st hour lag-1 correlation is inherited from the weakness of the proposed downscaling model, as shown in Fig. 2. The occurrence (P_1) probability significantly increased for all hours, and the transition probabilities (P_{01} and P_{11}) also significantly increased (data not shown).

Similar statistical characteristics can be observed in the downscaled RCM8.5 data, as shown in Fig. 10. A slightly greater increase in the mean of the RCM8.5 downscaled precipitation data compared with the mean of the RCM4.5 downscaled data was observed. The median of the mean of the RCM8.5 downscaled data (see the top-left panel of Fig. 10) increases to 0.8 mm at the 7th hour, whereas the median of the mean of the RCM4.5 data is 0.6 (see the top-left panel of Fig. 9). The occurrence probability (P_1) exhibits a greater increase for the RCM8.5 downscaled data than for the RCM4.5 downscaled data (data not shown). No significant difference in the transition probabilities can be observed for the RCM4.5 and RCM8.5 downscaled data.

As shown in Figs. 11 and 12, the empirical CDFs of the RCM4.5 and RCM8.5 downscaled data exhibit patterns similar to the

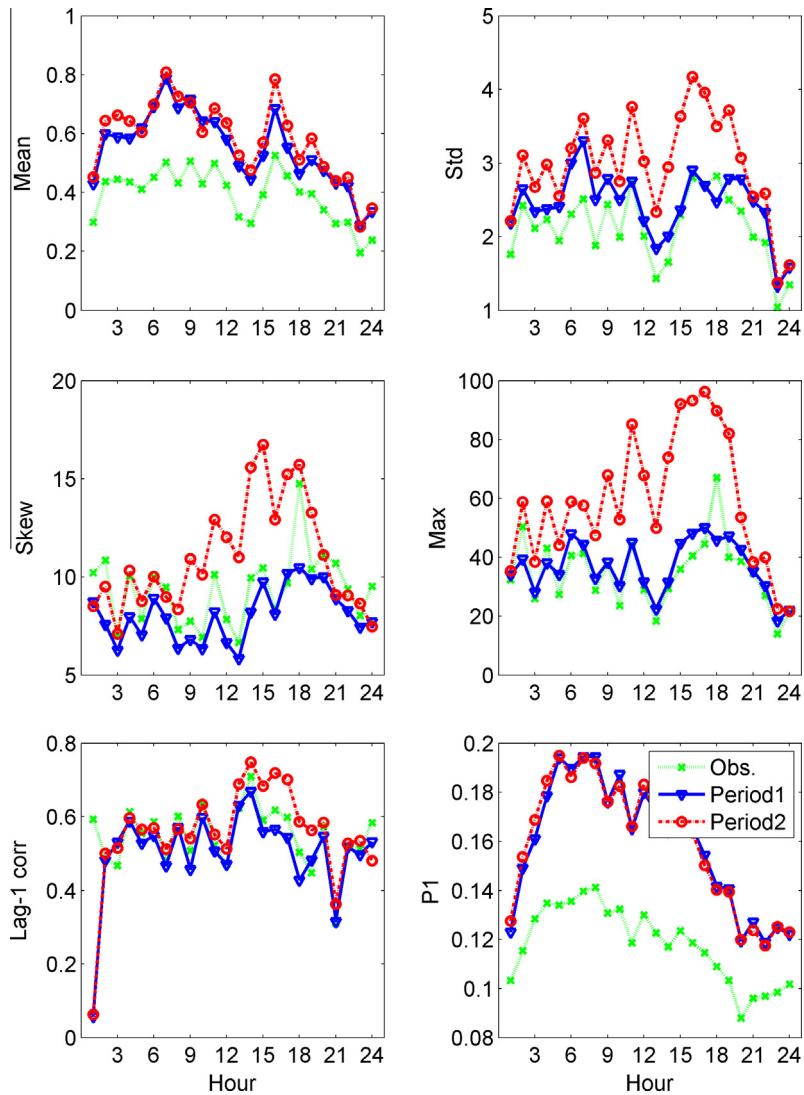


Fig. 13. The key statistics of the observed data (dotted line with \times marker) and downscaled RCP 8.5 data for the Period 1 (2011–2055; solid line with inverse triangle marker) and Period 2 (2056–2099; dash-dotted line with circle) of Jinju station. Note that the estimated 100 statistics from 100 downscaled series are averaged.

historical data, with the exception of the upper and lower tails. The upper tails of the RCM downscaled data are much flatter than the observation data over all the time durations. The RCM data exhibit much greater values than the observed data with the same probability. Consequently, more extreme events occur more frequently in RCP 8.5 than in the current climate. The RCM8.5 downscaled data (see Fig. 12) indicate that the more extreme events will occur more frequently than for the RCM4.5 data (see Fig. 11).

The RCP 8.5 data was split into two periods as Period 1 (2011–2055) and Period 2 (2056–2099) to investigate future trends of the key statistics and the results were presented in Fig. 13. In addition, two stations such as Seoul and Daegu were included to present the spatial difference of the future trends in South Korea and the results are illustrated in Figs. 14 and Fig. 15, respectively.

In Fig. 13, the six key statistics of the observed data (dotted with \times marker) and the RCP 8.5 data of Period 1 (solid line with inverse triangle marker) and Period 2 (dash-dotted line with circle) are presented for Jinju station. The mean and occurrence probability (P_1) are higher than the observed statistics with no difference between Period 1 and Period 2. The standard deviation, skewness, and maximum for Period 2 are higher than those for Period 1 with

not much difference from the statistics of the observed data. The lag-1 correlations for all three datasets are not much from each other.

The behaviors of the tested key statistics are different in Seoul station from the same in Jinju station shown in Fig. 14. The mean, standard deviation, P_1 for the RCP 8.5 data of Period 1 are higher than the same of the observed and Period 2 data. Those statistics for Period 1 is slightly higher than for Period 2. The skewness and maximum are often lower in the RCP 8.5 data (Period 1 and Period 2) than in the observed data. The lag-1 correlation and standard deviation show no difference among all the datasets.

In Fig. 15, the key statistics of the Seoul station are presented. The mean and P_1 are higher for the RCP 8.5 data of Period 1 than the same of the observed and Period 2 data. The standard deviation, skewness, maximum for the RCP 8.5 data of Period 2 present highest compared to the observed and Period 1 data. The lag-1 correlations of the RCP data (Period 1 and Period 2) are not significantly different from the observed data.

Overall, the key statistics of the RCP 8.5 data show different trend patterns compared to the statistics of the observed data especially in the mean and P_1 while not much difference can be

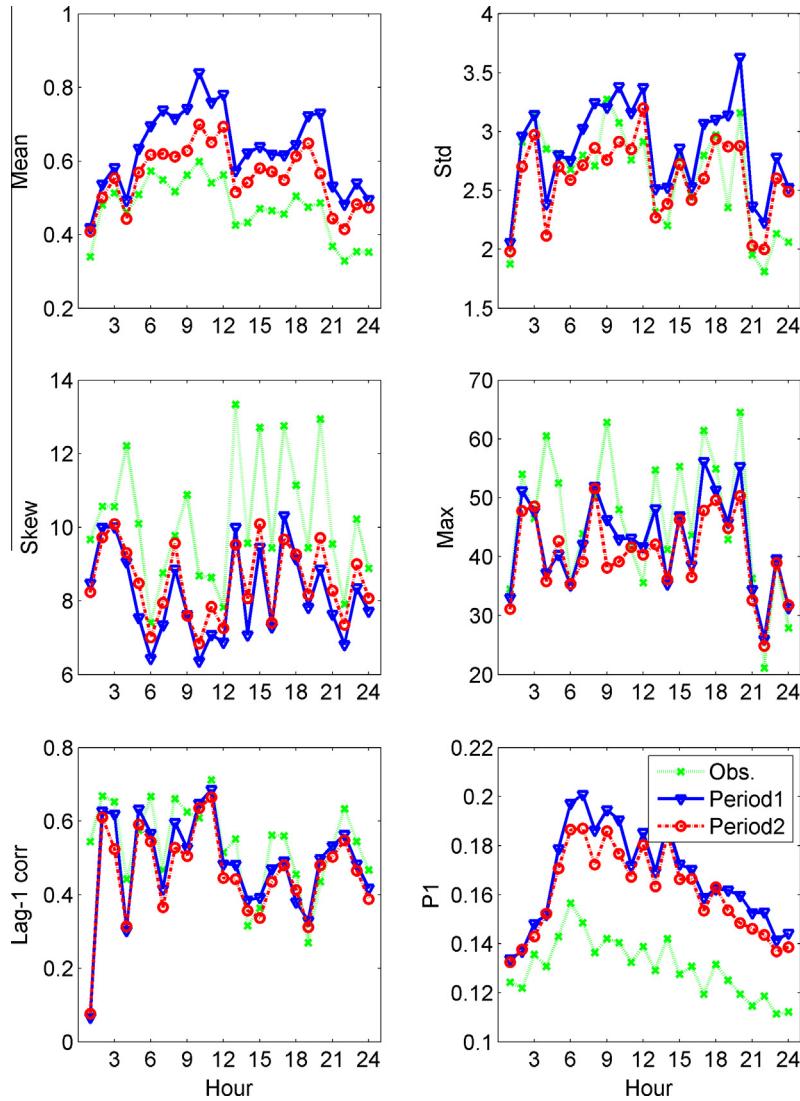


Fig. 14. The same as Fig. 13 except the employed data is from the Seoul station (37.57 latitude and 126.97 longitude).

observed in lag-1 correlation. No significant trend patterns that are spatially consistent can be detected from the results. Further analysis of the spatial trends in hourly precipitation data will be tested with applying the proposed model to the entire area of South Korea and will be reported as a following study.

5. Summary and conclusions

Finer time-resolution precipitation data are required for climate research, particularly for assessing the hydro-meteorological impacts of climate change. Although it is important to reproduce the key statistics and their diurnal cycle in temporally downscaled hourly precipitation data, temporal downscaling has not received much attention. The focus of the current study was to present a new nonparametric model for temporal downscaling from daily to hourly time series that reproduces the key statistics of the historical data and their daily pattern. To illustrate the performance of the proposed model, a precipitation dataset from Jinju station in South Korea was employed.

The results demonstrated that the proposed nonparametric temporal downscaling model faithfully reproduces the key statistics of the historical data and that the probability distribution of

extreme events in the downscaled data is quite similar to that for the historical data. The proposed downscaling method was also applied to the RCP 4.5 and RCP 8.5 climate scenarios. The results indicated that the mean and occurrence probability of the hourly precipitation data increased significantly in the climate scenarios. The standard deviation and maximum also increased, but the skewness and lag-1 correlation were unchanged. The magnitude of the change was greater for RCP 8.5 than for RCP 4.5. The results also demonstrated that a significant increase in extreme events occurs in the RCP 8.5 climate scenario. These results can be used for rainfall-runoff physical simulations, flood control, and dam management, particularly to mitigate the impacts of climate change.

The RCP data was split into two periods as well as including two other stations in order to investigate future trends in the key statistics and to illustrate the spatial difference of the future trends. The results show that not much spatially consistent trend can be detected while some future trends in the mean and P1 can be observed.

Finally, the results demonstrate that the proposed model can be a good alternative for downscaling daily precipitation to hourly precipitation while reproducing the key statistics of the hourly time scale. Although the proposed model exhibited good reproduction of key statistics, the inter-day relation (i.e., the correlation

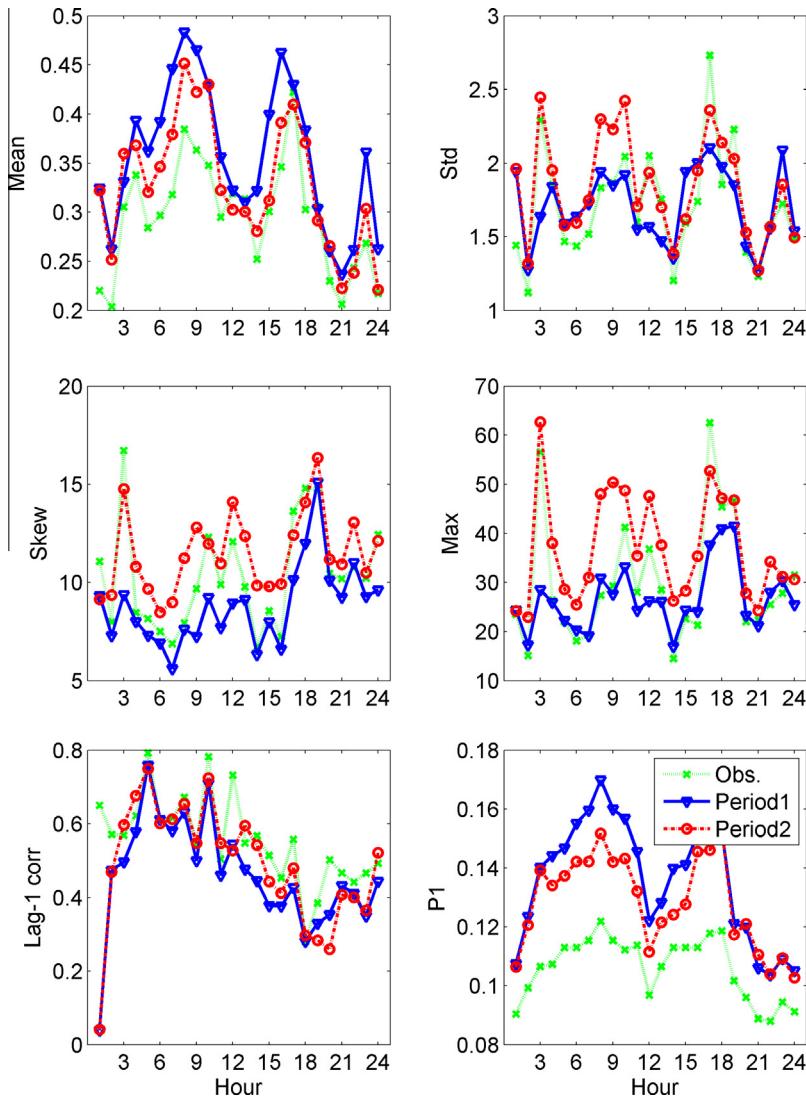


Fig. 15. The same as Fig. 13 except the employed data is from the Daegu station (35.89 latitude and 128.62 longitude).

between the 1st hour of the current day and the last hour of the previous day) is still significantly underestimated. Multiple modifications were tried, but the results were not successful. Future studies should focus on reproducing this inter-day relation.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MEST) (2012-0434). The first author also acknowledges that this research was supported by a grant 'Construction of a flood forecasting and warning system for the medium and small stream' (NEMA-NH-2011-45) from the Natural Hazard Mitigation Research Group, National Emergency Management Agency of Korea.

References

- Alfieri, L., Thielen, J., Pappenberger, F., 2012. Ensemble hydro-meteorological simulation for flash flood early detection in southern Switzerland. *J. Hydrol.* 424–425, 143–153.
- Boe, 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 (12), 1643–1655.
- Chen, G.F., Qin, D.Y., Ye, R., Guo, Y.X., Wang, H., 2011. A new method of rainfall temporal downscaling: a case study on sannmenxia station in the Yellow River Basin. *Hydrol. Earth Syst. Sci. Discuss.* 8 (2), 2323–2344.
- Chen, H., Sun, J., Wang, H., 2012. A statistical downscaling model for forecasting summer rainfall in China from DEMETER hindcast datasets. *Weather Forecasting* 27 (3), 608–628.
- Coulibaly, P., Dibike, Y.B., Anctil, F., 2005. Downscaling precipitation and temperature with temporal neural networks. *J. Hydrometeorol.* 6 (4), 483–496.
- Dai, A., 1999. Recent changes in the diurnal cycle of precipitation over the United States. *Geophys. Res. Lett.* 26 (3), 341–344.
- Dai, A., Trenberth, K.E., Karl, T.R., 1999. Effects of clouds, soil moisture, precipitation, and water vapor on diurnal temperature range. *J. Clim.* 12 (8 Part 2), 2451–2473.
- Eagleson, P.S., 1978. Climate, soil and vegetation: I introduction to water balance dynamics. *Wat. Resour. Res.* 14 (5), 705–712.
- Glasbey, C.A., Cooper, G., McGechan, M.B., 1995. Disaggregation of daily rainfall by conditional simulation from a point-process model. *J. Hydrol.* 165 (1–4), 1–9.
- Goldberg, D.E., 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Pub. Co., Reading, MA, 432 pp.
- He, H., Zhang, F., 2010. Diurnal variations of warm-season precipitation over Northern China. *Mon. Weather Rev.* 138 (4), 1017–1025.
- Jones, P., Harpham, C., 2009. UK Climate Projections Science Report: Projections of Future Daily Climate for the UK from the Weather Generator. University of Newcastle, Newcastle.
- Kendon, E.J., Jones, R.G., Kjellström, E., Murphy, J.M., 2010. Using and designing GCM–RCM ensemble regional climate projections. *J. Clim.* 23 (24), 6485–6503.
- Kim, W.-G., Park, J.-H., Lee, E.-R., 2009. Analysis of Namgang Dam Flood Control During Intensive Rianfall in 2006, 28th Annual USSD Conference, USSD, Portland, Oregon.
- Kondo, J., Xu, J., 1997. Seasonal variations in the heat and water balances for non-vegetated surfaces. *J. Appl. Meteorol.* 36, 1671–1696.

- Koutsoyiannis, D., Onof, C., 2001. Rainfall disaggregation using adjusting procedures on a Poisson cluster model. *J. Hydrol.* 246 (1–4), 109–122.
- Krajewski, W.F., Lakshmi, V., Georgakakos, K.P., Jain, S.C., 1991. A Monte Carlo study of rainfall sampling effect on a distributed catchment model. *Water Resour. Res.* 27 (1), 119–128.
- Kumara, J., Brooks, B.-G.J., Thornton, P.E., Dietze, M.C., 2012. Sub-daily Statistical Downscaling of Meteorological Variables Using Neural Networks. *Procedia Computer Science*, vol. 9.
- Lall, U., Sharma, A., 1996. A nearest neighbor bootstrap for resampling hydrologic time series. *Water Resour. Res.* 32 (3), 679–693.
- Lee, T., Ouarda, T.B.M.J., 2011. Identification of model order and number of neighbors for k-nearest neighbor resampling. *J. Hydrol.* 404 (3–4), 136–145.
- Lee, T., Ouarda, T.B.M.J., 2012. Stochastic simulation of nonstationary oscillation hydro-climatic processes using empirical mode decomposition. *Water Resour. Res.* 48 (W02514), 1–15.
- Lee, T., Ouarda, T.B.M.J., Jeong, C., 2012. Nonparametric multivariate weather generator and an extreme value theory for bandwidth selection. *J. Hydrol.* 452 (25), 161–171.
- Lee, T., Salas, J.D., Frevert, D.K., Fulp, T., 2007. Stochastic Simulation Study of the Colorado River System Utilizing Alternative Models, World Env. and Water Res. Congress 2007, Tempe FL.
- Lee, T., Salas, J.D., Prairie, J., 2010. An enhanced nonparametric streamflow disaggregation model with genetic algorithm. *Water Resour. Res.* 46, W08545.
- Lenderink, G., Buishand, A., Van Deursen, W., 2007. Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach. *Hydrol. Earth Syst. Sci.* 11 (3), 1145–1159.
- Marani, M., Zanetti, S., 2007. Downscaling rainfall temporal variability. *Water Resour. Res.* 43 (9).
- Mason, S.J., 2004. Simulating climate over western North America using stochastic weather generators. *Climatic Change* 62 (1–3), 155–187.
- Orlowsky, B., Gerstengarbe, F.W., Werner, P.C., 2008. A resampling scheme for regional climate simulations and its performance compared to a dynamical RCM. *Theoret. Appl. Climatol.* 92 (3–4), 209–223.
- Prairie, J.R., Rajagopalan, B., Fulp, T.J., Zagora, E.A., 2006. Modified K-NN model for stochastic streamflow simulation. *J. Hydrol. Eng.* 11 (4), 371–378.
- Prudhomme, C., Nick Reynard, S.C., 2002. Downscaling of global climate models for flood frequency analysis: where are we now? *Hydrol. Process.* 16 (6), 1137–1150.
- Rajagopalan, B., Lall, U., 1999. A k-nearest-neighbor simulator for daily precipitation and other weather variables. *Water Resour. Res.* 35 (10), 3089–3101.
- Rodriguez-Iturbe, I., Cox, D.R., Isham, V., 1988. A point process model for rainfall: further developments. *Proc. – R. Soc. London Ser. A* 417 (1853), 283–298.
- Salas, J.D., Lee, T., 2010. Nonparametric simulation of single-site seasonal streamflows. *J. Hydrol. Eng.* 15 (4), 284–296.
- Soltani, A., Hoogenboom, G., 2003. A statistical comparison of the stochastic weather generators WGEN and SIMMETEO. *Clim. Res.* 24 (3), 215–230.
- Wang, G., Liu, J., 2006. The Digital Watershed Model for the Yellow River Basin. China Science Press, Beijing.
- Yang, B.J., Ren, L., Chen, F., Shi, J., 2009. Uncertainty research of hydrological prediction in ungauged basis. *Water Res. Power* 27 (4), 7–10.
- Yin, S., Chen, D., Xie, Y., 2009. Diurnal variations of precipitation during the warm season over China. *Int. J. Climatol.* 29 (8), 1154–1170.