Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Geoscientific Model Development Discussions
Open Access

Discrete k-nearest neighbor resampling for simulating multisite precipitation occurrence and adaption to climate change

: Discrete KNNR for Multisite Occurrence (DKMO version1.0) - model development

Keywords: daily precipitation, discrete, k-nearest neighbor, Markov chain, multisite, occurrence

Taesam Lee[1] and Vijay P. Singh[2]

[1] Department of Civil Engineering, ERI, Gyeongsang National University,
501 Jinju-daero, Jinju, Gyeongnam, South Korea, 660-701

[2] Department of Biological and Agricultural Engineering & Zachry Department of Civil Engineering, Texas A&M University, 321 Scoates Hall, College Station, Texas, United States, 77843

Corresponding Author :

Taesam Lee, Ph.D.
Gyeongsang National University, Dept. of Civil Engineering
Tel)+82-55-772-1797, Fax)+82-55-772-1799
Email) tae3lee@gnu.ac.kr

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Geoscientific Model Development Discussions
Open Access

Abstract

Stochastic weather simulation models are commonly employed in water resources management and agricultural applications. The data simulated by these models, such as precipitation, temperature, and wind, are used as input for hydrological and agricultural models. Stochastic simulation of multisite precipitation occurrence is a challenge because of its intermittent characteristics as well as spatial and temporal cross-correlation. Employing a nonparametric technique, k-nearest neighbor resampling (KNNR), and coupling it with Genetic Algorithm (GA), this study proposes a novel simulation method for multisite precipitation occurrence. The proposed discrete version of KNNR (DKNNR) model is compared with an existing parametric model, called multisite occurrence model with standard normal variate (MONR). The datasets simulated from both the DKNNR model and the MONR model are tested using a number of statistics, such as occurrence and transition probabilities as well as temporal and spatial cross-correlations. Results show that the proposed DKNNR model can be a good alternative for simulating multisite precipitation occurrence. We also tested the model capability to adapt climate change. It is shown that the model is capable but further improvement is required to have specific variations of the occurrence probability due to climate change. Combining with the generated occurrence, the multisite precipitation amount can then be simulated by any multisite amount model.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

1. Introduction

Stochastic simulation of weather variables has been employed for water resources management, hydrological design, agricultural applications, filling in missing historical data, extending observed records, simulating data, and simulating different weather conditions. Stochastic simulation models play a key role in producing weather sequences, while preserving the statistical characteristics of observed data. A number of stochastic weather simulation models have been developed using parametric and nonparametric approaches (Lee, 2017; Lee et al., 2012; Wilby et al., 2003; Wilks, 1999; Wilks and Wilby, 1999).

Parametric approaches summarize the statistical characteristics of observed weather data with a parameter set (Jeong et al., 2012; Lee, 2016; Zheng and Katz, 2008). The parameters fitted with observed weather data are employed in simulation. In nonparametric approaches, historical analogs with current conditions are searched following the weather simulation data (Buishand and Brandsma, 2001; Lee et al., 2012). Furthermore, combinations of parametric and nonparametric models have also been proposed (Apipattanavis et al., 2007; Frost et al., 2011).

Among weather variables, the precipitation variable possesses intermittency and zero values between precipitation events, and to properly reproduce them is difficult and remains a challenge (Beersma and Buishand, 2003; Hughes et al., 1999; Katz and Zheng, 1999). Due to this difficulty, precipitation is simulated separately from other variables. The main method for reproducing intermittency has been the multiplication of precipitation occurrence and an amount as $Z=X\cdot Y$, where X is the occurrence (binary as either 0 or 1) and Y is the amount (Jeong et al., 2013; Lee and Park, 2017; Todorovic and Woolhiser, 1975). The spatial and temporal dependence in the occurrence and amount of precipitation introduces further complexity multisite simulation.

Wilks (1998) presented a multisite simulation model for the occurrence process (i.e. *X*) using the standard normal variable that is spatially dependent, representing the relation between the occurrence variable and the standard normal variable with simulation data. Even though the multisite occurrence data simulated by this model preserves various statistical characteristics of the observed data well, some drawbacks still exist, such as underestimation of lagged cross-correlation. Furthermore, the relation between standard normal variable and occurrence variable relies on long stochastic simulation.

Lall and Sharma (1996) proposed a nonparametric simulation model, called k-nearest neighbor resampling (KNNR). The model has been updated to simulate multivariate hydro-meteorological variables (Brandsma and Buishand, 1998; Mehrotra et al., 2006; St-Hilaire et al., 2012). One of the major drawbacks of this multivariate KNNR model is that the simulated data cannot produce patterns different from those of the observed data. Lee et al. (2010a) overcame this shortcoming by mixing the simulated dataset with Genetic Algorithm (GA) that led to the reproduction of similar populations. A number of variants of KNNR-GA have since been applied (Lee et al., 2012; Lee and Park, 2017).

Therefore, in the current study we propose a novel simulation method for multisite occurrence of the precipitation variable with a nonparametric approach. The proposed nonparametric model is compared with the existing multisite model (Wilks, 1998). The paper is organized as follows. The next section presents a mathematical background of existing multisite occurrence modeling. The modeling procedure is discussed in section 3. The study area and data are reported in section 4. The model is applied in section 5. Results of the proposed model are discussed in section 6, and summary and conclusions are presented in section 7.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

2. Background

2.1. Single site occurrence modeling

Let X_t^s represent the occurrence of daily precipitation for a location *s* (*s*=1,…, *S*) on day *t* (*t*=1,…, *n*; *n* is the number observed days) and let X_t^s be either zero for dry day or one for wet day. The first order Markov chain model for X_t^s is defined with the assumption that the occurrence probability of a wet day is fully defined by the previous day as

$$\Pr\{X_t^s = 1 \mid X_{t-1}^s = 0\} = p_{01}^s \tag{1}$$

$$\Pr\{X_t^s = 1 \mid X_{t-1}^s = 1\} = p_{11}^s \tag{2}$$

Also $p_{00}^s = 1 - p_{01}^s$ and $p_{10}^s = 1 - p_{11}^s$, since the summation of zero and one should be unity with the same previous condition. This consists of a transition probability matrix (TPM) as

$$TPM^s = \begin{bmatrix} p_{00}^s & p_{01}^s \\ p_{10}^s & p_{11}^s \end{bmatrix} = \begin{bmatrix} 1 - p_{01}^s & p_{01}^s \\ 1 - p_{11}^s & p_{11}^s \end{bmatrix} \tag{3}$$

The marginal distributions of TPM (i.e. p_0 and p_1) can be expressed with TPM and its condition of $p_0 + p_1 = 1$ as:

$$p_0^s = \frac{p_{01}^s}{1 + p_{01}^s - p_{11}^s} \tag{4}$$

$$p_1^s = \frac{1 - p_{11}^s}{1 + p_{01}^s - p_{11}^s} \tag{5}$$

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Note that p_1 represents the probability of precipitation occurrence for a day, while p_0 does non-occurrence. The lag-1 autocorrelation of precipitation occurrence is the combination of transition probabilities as:

$$\rho_1(s,s) = p_{11}^s - p_{01}^s \tag{6}$$

The simulation can be done by comparing TPM with a uniform random number (u_t^s) as

$$X_t^s = \begin{cases} 1 & \text{if } u_t^s \le p_{i1}^s \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where p_{i1}^s is the selected probability from TPM regarding the previous condition i (i.e. either 0 or 1). Wilks (1998) suggested a different method using a standard normal random number w_t^s ~N[0,1] as

$$X_t^s = \begin{cases} 1 & \text{if } w_t^s \le \Phi^{-1}(p_{i1}^s) \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where Φ^{-1} indicates the inverse of the standard normal cumulative function Φ.

2.2. Multisite occurrence modeling

Wilks (1998) suggested a multisite occurrence model using a standard normal random number (here, denoted as MONR) that is spatially dependent but serially independent. The correlation of the standard normal variate for a site pair of q and s can be expressed as:

$$\tau(q,s) = corr[w_t^q, w_t^s] \tag{9}$$

Also, the correlation of the original occurrence variate is

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Geoscientific Model Development Discussions

$$\rho(q,s) = corr[X_t^q, X_t^s] \tag{10}$$

Once the correlation of the standard normal variate is known, the simulation of multisite precipitation occurrence is straightforward. Multivariate standard normal distribution is used with the parameter set of [**0**, **T**] where **0** is the zero vector (*S*x1) and **T** is the correlation matrix with the elements of $\tau(q,s)$ for $q \in \{1,\ldots,S\}$ and $s \in \{1,\ldots,S\}$.

Since direct estimation of $\tau(q,s)$ is not applicable, a simulation technique is used to estimate $\tau(q,s)$ from $\rho(q,s)$. A long sequence of the occurrence process is simulated with different values of $\tau(q,s)$ and its corresponding correlation of the original domain $\rho(q,s)$ is estimated with the simulated long sequence by the inverse standard normal cumulative function (i.e. Φ^{-1}). A curve between $\tau(q,s)$ and $\rho(q,s)$ is derived from this long simulation with the MONR model and is employed for the parameter estimation for real application.

3. DKNNR

3.1. DKNNR modeling procedure

In the current study, a novel multisite simulation model for discrete occurrence of precipitation variable with k-nearest neighbor resampling (KNNR) technique (Lall and Sharma, 1996; Lee and Ouarda, 2011; Lee et al., 2017) for discrete case (denoted as Discrete KNNR; DKNNR) is proposed by combining a mixture mechanism with Genetic Algorithm (GA).

Provided the number of nearest neighbors, *k*, is known, the discrete k-nearest neighbor resampling with genetic algorithm is done as follows:

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

(1) Estimate the distance between the current (i.e. time index: c) multisite occurrence X_c^s and the observed multisite occurrence x_i^s. Here, the distance is measured for i=1,…,n-1 as

$$D_i = \sum_{s=1}^{S} \left| X_c^s - x_i^s \right| \tag{11}$$

(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

$$w_m = \frac{1/m}{\sum_{j=1}^{k} 1/j}, \qquad m = 1,\ldots,k \tag{12}$$

(4) Assume the selected time index from step (3) as p. Note that there are a number of values that have the same distance as the selected D_p, since D_p is a natural number between 0 and S. A random selection procedure is required to take into account the cases with the same quantity. One particular time index is randomly selected with the equal probabilities among the time indices of the same distances.

(5) Assign the binary vector of the proceeding index of the selected time as $\mathbf{x}_{p+1} = [x_{p+1}^s]_{s\in\{1,S\}}$. Here, p is the finally selected time index from step (4).

(6) Execute the following steps for GA mixing if GA mixing is selected. Otherwise, skip this step.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

(6-1) Reproduction: Select one additional time index using steps (1) through (4) and denote this index as p^*. Obtain the corresponding precipitation occurrence values, $\mathbf{x}_{p^*+1} = [x^s_{p^*+1}]_{s\in\{1,\ldots,S\}}$. The subsequent two GA operators employ the two selected vectors, $\mathbf{x}_{p+1}$ and $\mathbf{x}_{p^*+1}$.

(6-2) Crossover: Replace each element x^s_{p+1} with $x^s_{p^*+1}$ at probability P_{cr}, i.e.,

$$X^s_{c+1} = \begin{cases} x^s_{p^*+1} & \text{if } \varepsilon < P_{cr} \\ x^s_{p+1} & \text{otherwise} \end{cases} \tag{13}$$

where ε is a uniform random number between 0 and 1.

(6-3) Mutation: Replace each element (i.e., each station, s=1,…, S) with one selected from all the observations of this element for i=1,…,n with probability P_m, i.e.,

$$X^s_{c+1} = \begin{cases} x^s_{\xi+1} & \text{if } \varepsilon < P_m \\ x^s_{p+1} & \text{otherwise} \end{cases} \tag{14}$$

where $x^s_{\xi+1}$ is selected from $[x^s_i]_{i\in\{1,\ldots,n\}}$ with equal probability for i=1,…,n and ε is a uniform random number between 0 and 1.

(7) Repeat steps (1)-(6) until the required data are generated.

The selection of the number of nearest neighbors (k) has been investigated by Lall and Sharma (1996) and Lee and Ouarda (2011). A simple selection method was applied in the current study as $k = \sqrt{n}$. For hydrometeorological stochastic simulations, this heuristic approach of k selection has been employed (Lall and Sharma, 1996; Lee and Ouarda, 2012; Lee et al., 2010b; Prairie et al., 2006; Rajagopalan and Lall, 1999). The roles of crossover probability P_{cr} and

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

mutation probability P_m were studied by Lee et al. (2010b). Lee et al. (2010b) showed that P_{cr} =0.1 and P_m=0.01 can be a reasonable parameter set which does not critically affect the performance. Therefore, this parameter set was applied in the current study. In Appendix A, an example of the DKNNR simulation procedure is explained in detail.

3.2. Adaptation to climate change

The capability of model to take climate change into account is critical. For example, the marginal distributions and transition probabilities in Eqs. (5) and (3) can change in future climate scenarios. It is known that nonparametric simulation models have a difficulty to adapt to climate change, since the models employ in general the current observation sequences. However, the proposed model in the current study possesses the capability to adapt to the variations of probabilities by tuning the crossover and mutation probabilities in P_{cr} (13) and P_m (14) , adding the condition when applied.

For example, the probability of P_{11} can be increased with the cross-over probability P_{cr} by adding the condition to increase the probability of P_{11} as:

$$X_{c+1}^{s} = \begin{cases} x_{p^*+1}^{s} & \text{if} \quad \varepsilon < P_{cr} \ \& \ x_{p^*+1}^{s} = 1 \ \& \ X_{c}^{s} = 1 \\ x_{p+1}^{s} & \text{otherwise} \end{cases} \tag{15}$$

It is obviously possible to increase the probability of P_1 by removing the condition of $X_c^s = 1$.

In addition, further adjustment can be made with the mutation process in Eq. (14) as

$$X_{c+1}^{s} = \begin{cases} x_{\xi+1}^{s} & \text{if} \quad \varepsilon < P_{m} \text{ and } x_{\xi+1}^{s} = 1 \\ x_{p+1}^{s} & \text{otherwise} \end{cases} \tag{16}$$

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

This adjustment of adding the condition $x^{s}_{\xi+1}=1$ can increase the marginal distribution as much as $P_m \times P_1$. This has been tested in the case study.

4. Study area and data description

For testing the occurrence model, 12 weather stations were selected from Yeongnam province which is located in the southeastern part of South Korea, as shown in Figure 1. Information on longitude and latitude (fourth and fifth columns) as well as order index and the identification number (first and second columns) of these stations operated by Korea Meteorological Administration with the area name (third column) is shown at Table 1.

Figure 1 illustrates the locations of the selected weather stations. All the stations are inside Yeongnam province which consists of two different regions as north and south Gyeongsang as well as the self-governing cities of Busan, Deagu, and Ulsan. Most of the Yeongnam region is drained to Nakdong River. It is important to analyze the impact of weather conditions for planning agricultural operations and water resources management especially during the summer season, because around 50-60 percent of the annual precipitation occurs during the summer season from June to September. The length of daily precipitation data record ranges from 1976 to 2008 and the summer season record was employed since a large number of rainy days occurs during summer and it is important to preserve these characteristics. Also, the whole year dataset was tested and other seasons were further applied but the correlation coefficient was relatively high and its correlation matrix estimated was not a positive semi-definite matrix for the MONR model.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

5. Application

To analyze the performance of the proposed DKNNR model, the occurrence of precipitation was simulated. The DKNNR simulation was compared with that of the MONR model. For each model, 100 series of daily occurrence with the same record length were simulated. The key statistics of observed data and each generated series, such as transition probabilities (P_{11}, P_{01}, and P_1) and cross-correlation (see Eq.(10)), were determined. The MONR model underestimated the lag-1 cross-correlation, as indicated by Wilks (1998). In the current study, this statistic was analyzed, since a synoptic scale weather system could result in lagged cross-correlation (Wilks, 1998). It was formulated as

$$\rho_1(q,s) = corr[X_{t-1}^q, X_t^s] \tag{17}$$

Statistics from 100 generated series were evaluated by the root mean square error (RMSE) expressed as below:

$$RMSE = \left(\frac{1}{N} \sum_{m=1}^{N} (\Gamma_m^G - \Gamma^h)^2 \right)^{1/2} \tag{18}$$

where N is the number of series (here 100), Γ_m^G is the statistic estimated from the m^{th} generated series, while Γ^h is the statistic for the observed data. Note that lower RMSE indicates better performance representing the summarized error of a given statistic of generated series from the statistic of the observed data.

The 100 simulated statistic values were illustrated with boxplots to show their variability as shown in Figure 2 - Figure 4. The box of boxplot represents the interquartile range (IQR) ranging 25 percentile to 75 percentile. The whiskers extend to up and down 1.5×IQR. Data beyond the

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

whiskers (1.5×IQR) are indicated by a plus sign (+). The horizontal line inside the box represents the median of the data. The statistics of the observed data are denoted by a cross (x). The closer a cross is to the horizontal line inside the box, the better the simulated data from a model reproduces the statistical characteristics of the observed data.

6. Results

6.1. Occurrence and transition probabilities

The data simulated from the proposed DKNNR model and the existing MONR model were analyzed. The estimated transition probabilities (P_{11} and P_{01} in Eq. (3)) as well as the occurrence probability (P_1 in Eq. (5)) are shown in Table 2 and Figure 2 - Figure 4 for the observed data and the data generated from the DKNNR and MONR models. In Table 2, the observed statistic shows that P_{11} is always higher than P_{01} and P_1 is between P_{11} and P_{01}. Site 6 shows the lowest P_{11} and P_1 and site 12 shows the highest P_{11}.

As shown in Figure 2, the probability P_{11} of the observed data shows that sites 6, 7, 8, and 9 located in the northern part of the region exhibited lower consistency (i.e. consecutive rainy days) than did the other sites, while sites 5 and 12 had higher probability of P_{11} than did other sites. Both models preserved well the observed P_{11} statistic. It seems that the MONR model had a slight better performance since this statistic is parameterized in the model as shown in the section 2.2. Note that the MONR model employed the transition probabilities in simulating rainfall occurrence, while DKNNR model did not. The occurrence probability P_1 can be described with the combination of transition probabilities as in Eq. (5). Even though the transition probabilities were not employed in simulating rainfall occurrence, the DKNNR model preserved this statistic fairly well.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

As shown in Figure 3, the P_{01} probability showed a slightly different behavior such that sites 1, 2, and 3 located in the middle part of the Yeongnam province showed a higher probability than did other sites. A slight underestimation was seen for sites 2 and 11 but it was not critical, since its observed value with a cross mark was close to the upper IQR representing 75 percentile.

The behavior of P_1 was found to be same as that of the P_{11} probability. It can be seen in Figure 4 that no significant underestimation is seen for the DKNNR model (top panel). The P_1 statistic was fairly preserved by both DKNNR and MONR models. Note that the MONR model parameterized the P_1 statistic through the transition probabilities as in Eq. (5), while DKNNR model did not. Although the DKNNR model did not use any parameter for simulation, the P_1 statistic was preserved fairly well.

6.2. Cross-correlation

Cross-correlation is a measure of relationship between sites. Preservation of cross-correlation is important for the simulation of precipitation occurrence and is required in the regional analysis for water resources management or agricultural applications. Furthermore, lagged cross-correlation is also essential as much as is cross-correlation (i.e. contemporaneous correlation). For example, the amount of streamflow for a watershed from a certain precipitation event is highly related with lagged cross-correlation. It is accepted that precipitation event is not significantly correlated with more than one day. Therefore, only lag-1 cross-correlation was analyzed in the current study.

The cross-correlation of observed data is shown in Table 3. High cross-correlation among grouped sites, such as sites 6, 7, and 8 (northern part) and sites 3, 4, and 5 as well as 12 (southeast coastal area, 0.68-0.87), was found. As expected, sites 5 and 12 had the highest cross-correlation

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Geoscientific Model Development Discussions
Open Access

(0.87) due to the proximity. The northern sites and coastal sites showed low cross-correlation. This observed cross-correlation was well preserved in the data generated from both DKNNR and MONR models, as shown in Figure 5 as well as Table 4 and Table 5. However, consistently slight but significant underestimation of cross-correlation was seen for the data generated by the MONR model (see the bottom panel of Figure 5). Note that the errobars are extended to upper and lower lines of the circles to 1.95×standard deviation. The difference of RMSE in Table 6 showed this characteristic, as most of the values were positive, to be indicating that the proposed DKNNR model performed better for cross-correlation.

The lag-1 cross-correlation of observed data, as shown in Table 7, ranged from 0.22-0.35. The lag-1 cross-correlation for the same site (i.e. $\rho_1(q,s)$, q=s) was autocorrelation and was highly related with P_{01} and P_{11} as in Eq. (6). All the lag-1 cross-correlations exhibited similar magnitudes even for autocorrelation. This implies that the lag-1 cross-correlation among the selected sites was as strong as the autocorrelation and as much as the transition probabilities P_{01} and P_{11}, thereof. Relatively low lag-1 cross-correlation was observed between northern sites (6, 7, and 8) and coastal sites (3, 4, and 5), as shown in Table 7.

The observed cross-correlation was well preserved in the data generated by the DKNNR model, as shown in the top panel of Figure 6, while the MONR model showed significant underestimation, as seen in the bottom panel of Figure 6. The difference of RMSE shown in Table 8 reflects this behavior. In the bottom panel of Figure 6, some of the lag-1 cross-correlations were well preserved, that was aligned with the base line. From Table 8, the MONR model reproduced the autocorrelations well with the shaded values. It is because the lag-1 autocorrelation was indirectly parameterized with the transition probabilities of P_{11} and P_{01} as in Eq. (6). Other than this autocorrelation, the lag-1 cross-correlation was not reproduced with the MONR model. This shortcoming was mentioned by Wilks (1998). Meanwhile, the proposed DKNNR model preserved this statistic without any parameterization.

Also, the whole year data instead of the summer season data was tested for model fitting. Note that all the results presented above were with the summer season data (June-September) as mentioned in section 4 on the data description. The lag-1 cross-correlation is shown in Figure 7 which indicates that the same characteristic was observed as for the summer season, such that the proposed DKNNR model preserved better the lagged cross-correlation than did the existing MONR model. Other statistics, such as correlation matrix and transition probabilities, exhibited the same results (not shown). Also, other seasons were tried but the estimated correlation matrix was not a positive semi-definite matrix and its inverse cannot be made for multivariate normal distribution in the MONR model. It was because the selected stations were close to each other (around 50-100 km) and produced high cross-correlation, especially in the occurrence during dry seasons. Special remedy should be applied, such as decreasing cross-correlation by force, but further remedy was not applied in the current study since it was not within the current scope and focus.

6.3. Adaptation to climate change

Model adaptability to climate change in hydro-meteorological simulation models is a critical factor, since one of the major applications of the models is to assess the impact of climate change. Therefore, we tested the capability of the proposed model in the current study by adjusting the probabilities of cross-over and mutation as in Eqs.(15) and (16). A number of variations can be made with different conditions.

In Figure 8, the changes of transition and marginal probabilities are shown along with increasing the crossover probability P_{cr} from 0.01 to 0.2 with the condition that that the candidate value is one and the previous value is also one as in Eq. (15) for the selected 5 stations among the 12 stations (from station 1 to station 5, see Table 1 for the detail). The stations were limited in this analysis due to computational time. At each case 100 series were simulated. The average value of the simulated statistics is presented in the figure. It is obvious that the transition probability P_{11} increased as intentioned along with the increase of P_{cr} . As expected from Eq. (5), P_1 presents that the change of P_1 is highly related to P_{11}. However, the probability P_{01} fluctuated along with the increase of P_{cr}. Elaborate work to adjust all the probabilities is however required.

The changes in transition and marginal probabilities are presented in Figure 9 with increasing mutation probability P_m from 0.01 to 0.2 under the condition that the candidate value is one so that the marginal probability P_1 increased. P_{01} also increased along with increasing P_1. The change of P11 was not related with other probabilities. The combination of the adjustment of P_{cr} and P_m with a certain condition to the previous state will allow the specific adaptation for simulating future climatic scenarios.

7. Conclusions

In the current study, a nonparametric simulation model, based on discrete KNNR and DKNNR, is proposed. The proposed DKNNR model is compared with the existing MONR model. Occurrence and transition probabilities and cross-correlation as well as lag-1 cross-correlation are estimated for both models. Better preservation of cross-correlation and lag-1 cross-correlation with the DKNNR model than the MONR model is observed. For some cases (i.e., the whole year data and other seasons than the summer season), the estimated cross-correlation matrix is not a positive semi-definite matrix so the multivariate normal simulation is not applicable for the MONR model because the tested sites are close to each other with high cross-correlation.

Results of this study indicate that the proposed DKNNR model reproduces the occurrence and transition probabilities fairly well and preserves the cross-correlations better than the existing MONR model. Thus, the proposed DKNNR model can be a good alternative for simulating multisite precipitation occurrence.

We tested further enhancement of the proposed model for adapting climate change through modifying the mutation and crossover probability P_m and P_{cr} with the current and previous states. The results show that the current model has the capability to adapt to the climate change scenarios but elaborate work is required however. Further study on improving the model adaptability to climate change will be followed in near future.

Also, the simulated multisite occurrence can be coupled with a multisite amount model to produce precipitation events, including zero values. Further development can be made for multisite amount models with a nonparametric technique, such as KNNR and bootstrapping.

Code and Data Availability

DKNNR code is written in Matlab and is available at the supplement.

The precipitation data employed in the current study is downloadable through http://www.weather.go.kr/weather/main.jsp

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant (NRF-2018R1A2B6001799) funded by the Korean Government (MEST).

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Appendix A: Example of DKNNR

In this appendix, one example of DKNNR simulation is presented with observed dataset in Table A 1 (i.e. $\mathbf{x}_i = [x_i^s]_{s\in\{1,S\}}$ for i=1,…,n; here S=12 and n=16). The upper part of the table presents the observed precipitation (unit: mm). Its occurrence data is presented in the bottom part of this table. The current precipitation occurrence $\mathbf{X}_c = [X_c^s]_{s\in\{1,\dots,12\}}$ is shown in the second row of Table A 2. The number of nearest neighbors $k = \sqrt{n} = \sqrt{16} = 4$ and the parameters for GA (i.e. P_c and P_m) are 0.1 and 0.01, respectively. The simulation can be made as follows:

(1) Estimate the distance D_i between $\mathbf{x}_i$ and $\mathbf{X}_c$ for i=1,…,n-1 as in Eq.(11). For example, for i=1,

$$D_1 = \sum_{s=1}^{S} \left| X_c^s - x_1^s \right| = |0-1| + |1-1| + \ldots + |0-1| = 6$$

All the estimated distances are shown in the last column of Table A 2.

(2) The daily index values are sorted according to the smallest distances shown in the first two columns of Table A 3. The sorted day indices and their corresponding distances are shown in the third and fourth columns of Table A 3. Among k number of sorted indices, one is selected with the weight probability (see Eq.(12)), which is shown in the last column of Table A 3.

(3) Simulate a uniform random number (u) between 0 and 1. Say u=0.321. This value must be compared with the cumulative weighted probabilities in the last column of Table A 3 as [0 0.48 0.72 0.88 1.0]. The corresponding day index is assigned as to where the simulated uniform number falls in the cumulative weighted probabilities, here [0 0.48].

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Geoscientific Model Development
Discussions
Open Access

Therefore, the selected day, p, is 14. The occurrences of the following day p+1=15 for 12 stations are selected as in the second row of Table A 4.

(4) For GA mixture, another set must be chosen as in step (3). Say u=0.561, which falls in [0.48 0.72]. The second one should be selected. However, there are a number of days with the same distances. Specifically, six days have the same distances with D_i=4. In this case, one among all six days is selected with equal probability. Assume that p=4 is selected and the following occurrences are selected as shown in the third row of Table A 4.

(5) With two sets, crossover and mutation process is performed as follows:

(5-1) Crossover: For each station, a uniform random number (ε) is generated and compared with P_c=0.1 here. Say ε =0.345, then skip since ε =0.345> P_c=0.1. For s=6, assume the generated random number, ε (=0.051)< P_c(=0.1) and then switch the 6th station value of Set 1 into the value of Set 2 (see Table A 4). The occurrence state of X_{c+1}^{s} turns into 1 from 0 as shown in the fourth row of Table A 4 as well as station 8.

(5-2) Mutation: For each station, a uniform random number (ε) is generated and compared with P_m=0.01. For s=12, assume ε =0.009< P_m=0.01 and switch the 12th station value of Set 1 with the one selected among all the observed 12th station values with equal probability (here the last column, s=12, of the bottom part of Table A 1, [1 1 0 0 … 1]). The occurrence state of X_{c+1}^{12} turns into 0 from 1 as shown in the fourth column of Table A 4.

(6) Repeat steps (1)-(5) until the target simulation length is reached.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Geoscientific Model Development Discussions
Open Access

References

Apipattanavis, S., Podesta, G., Rajagopalan, B., and Katz, R. W.: A semiparametric multivariate and multisite weather generator, Water Resources Research, 43, Artn W11401, 2007.

Beersma, J. J. and Buishand, A. T.: Multi-site simulation of daily precipitation and temperature conditional on the atmospheric circulation, Climate Research, 25, 121-133, 2003.

Brandsma, T. and Buishand, T. A.: Simulation of extreme precipitation in the Rhine basin by nearest-neighbour resampling, Hydrology and Earth System Sciences, 2, 195-209, 1998.

Buishand, T. A. and Brandsma, T.: Multisite simulation of daily precipitation and temperature in the Rhine basin by nearest-neighbor resampling, Water Resources Research, 37, 2761-2776, 2001.

Frost, A. J., Charles, S. P., Timbal, B., Chiew, F. H. S., Mehrotra, R., Nguyen, K. C., Chandler, R. E., McGregor, J. L., Fu, G., Kirono, D. G. C., Fernandez, E., and Kent, D. M.: A comparison of multi-site daily rainfall downscaling techniques under Australian conditions, Journal of Hydrology, 408, 1-18, 2011.

Hughes, J. P., Guttorp, P., and Charles, S. P.: A non-homogeneous hidden Markov model for precipitation occurrence, Journal of the Royal Statistical Society. Series C: Applied Statistics, 48, 15-30, 1999.

Jeong, D. I., St-Hilaire, A., Ouarda, T. B. M. J., and Gachon, P.: A multi-site statistical downscaling model for daily precipitation using global scale GCM precipitation outputs, International Journal of Climatology, 33, 2431-2447, 2013.

Jeong, D. I., St-Hilaire, A., Ouarda, T. B. M. J., and Gachon, P.: Multisite statistical downscaling model for daily precipitation combined by multivariate multiple linear regression and stochastic weather generator, Climatic Change, 114, 567-591, 2012.

Katz, R. W. and Zheng, X.: Mixture model for overdispersion of precipitation, Journal of Climate, 12, 2528-2537, 1999.

Lall, U. and Sharma, A.: A nearest neighbor bootstrap for resampling hydrologic time series, Water Resources Research, 32, 679-693, 1996.

Lee, T.: Multisite stochastic simulation of daily precipitation from copula modeling with a gamma marginal distribution, Theoretical and Applied Climatology, doi: 10.1007/s00704-017-2147-0, 2017. 1-10, 2017.

Lee, T.: Stochastic simulation of precipitation data for preserving key statistics in their original domain and application to climate change analysis, Theoretical and Applied Climatology, 124, 91-102, 2016.

Lee, T. and Ouarda, T. B. M. J.: Identification of model order and number of neighbors for k-nearest neighbor resampling, Journal of Hydrology, 404, 136-145, 2011.

Lee, T. and Ouarda, T. B. M. J.: Stochastic simulation of nonstationary oscillation hydro-climatic processes using empirical mode decomposition, Water Resources Research, 48, 1-15, 2012.

Lee, T., Ouarda, T. B. M. J., and Jeong, C.: Nonparametric multivariate weather generator and an extreme value theory for bandwidth selection, Journal of Hydrology, 452-453, 161-171, 2012.

Lee, T., Ouarda, T. B. M. J., and Yoon, S.: KNN-based local linear regression for the analysis and simulation of low flow extremes under climatic influence, Climate Dynamics, doi: 10.1007/s00382-017-3525-0, 2017. 1-19, 2017.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Lee, T. and Park, T.: Nonparametric temporal downscaling with event-based population generating algorithm for RCM daily precipitation to hourly: Model development and performance evaluation, Journal of Hydrology, 547, 498-516, 2017.

Lee, T., Salas, J. D., and Prairie, J.: An enhanced nonparametric streamflow disaggregation model with genetic algorithm, Water Resources Research, 46, 2010a.

Lee, T., Salas, J. D., and Prairie, J.: An Enhanced Nonparametric Streamflow Disaggregation Model with Genetic Algorithm, Water Resources Research, 46, W08545, 2010b.

Mehrotra, R., Srikanthan, R., and Sharma, A.: A comparison of three stochastic multi-site precipitation occurrence generators, Journal of Hydrology, 331, 280-292, 2006.

Prairie, J. R., Rajagopalan, B., Fulp, T. J., and Zagona, E. A.: Modified K-NN model for stochastic streamflow simulation, Journal of Hydrologic Engineering, 11, 371-378, 2006.

Rajagopalan, B. and Lall, U.: A k-nearest-neighbor simulator for daily precipitation and other weather variables, Water Resources Research, 35, 3089-3101, 1999.

St-Hilaire, A., Ouarda, T. B. M. J., Bargaoui, Z., Daigle, A., and Bilodeau, L.: Daily river water temperature forecast model with a k-nearest neighbour approach, Hydrological Processes, 26, 1302-1310, 2012.

Todorovic, P. and Woolhiser, D. A.: Stochastic model of n-day precipitation Journal of Applied Meteorology, 14, 17-24, 1975.

Wilby, R. L., Tomlinson, O. J., and Dawson, C. W.: Multi-site simulation of precipitation by conditional resampling, Climate Research, 23, 183-194, 2003.

Wilks, D. S.: Multisite downscaling of daily precipitation with a stochastic weather generator, Climate Research, 11, 125-136, 1999.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Wilks, D. S.: Multisite generalization of a daily stochastic precipitation generation model, Journal of Hydrology, 210, 178-191, 1998.

Wilks, D. S. and Wilby, R. L.: The weather generation game: a review of stochastic weather models, Progress in Physical Geography, 23, 329-357, 1999.

Zheng, X. and Katz, R. W.: Simulation of spatial dependence in daily rainfall using multisite generators, Water Resources Research, 44, 2008.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Table 1. Information on 12 selected stations from Yeongnam province, South Korea.

Order	Station Number[†]	Name	Longitude	Latitude
1	138	Pohang	129.3797	36.0327
2	143	Daegu	128.6189	35.8850
3	152	Ulsan	129.3200	35.5600
4	159	Busan	129.0319	35.1044
5	162	Tongyeong	128.4356	34.8453
6	277	Youngdeok	129.4092	36.5331
7	278	Uisung	128.6883	36.3558
8	279	Gumi	128.3206	36.1306
9	281	Youngcheon	128.9514	35.9772
10	285	Hapcheon	128.1697	35.5650
11	288	Milyang	128.7439	35.4914
12	294	Geojae	128.6044	34.8881

[†]The station number indicates the identification number operated by Korea Meteorological Administration (KMA).

Table 2. Occurrence and transition probabilities of observed data and data simulated by DKNNR and MONR for 12 stations from Yeongnam province, South Korea, during the summer season. Note that 100 sets with the same record length as the observed data were simulated and the statistics of 100 sets were averaged.

	Obs			DKNNR			MONR		
	P11	P01	P1	P11	P01	P1	P11	P01	P1
S1	0.57	0.26	0.38	0.55	0.26	0.37	0.58	0.27	0.39
S2	0.56	0.27	0.38	0.57	0.26	0.37	0.58	0.27	0.39
S3	0.57	0.26	0.38	0.56	0.25	0.37	0.58	0.27	0.39
S4	0.58	0.25	0.37	0.56	0.24	0.36	0.60	0.25	0.39
S5	0.58	0.25	0.37	0.57	0.24	0.36	0.60	0.25	0.38
S6	0.52	0.24	0.33	0.50	0.24	0.32	0.53	0.25	0.35
S7	0.54	0.25	0.35	0.53	0.24	0.34	0.56	0.26	0.37
S8	0.55	0.25	0.36	0.52	0.24	0.34	0.57	0.26	0.38
S9	0.54	0.24	0.35	0.54	0.23	0.34	0.55	0.26	0.36
S10	0.58	0.24	0.37	0.56	0.23	0.35	0.57	0.26	0.38
S11	0.56	0.24	0.36	0.55	0.23	0.34	0.57	0.25	0.37
S12	0.59	0.24	0.37	0.58	0.24	0.36	0.61	0.25	0.39

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Geoscientific Model Development Discussions

Table 3. Cross-correlation of observed data for 12 stations from Yeongnam province, South Korea.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
S1	1.00	0.71	0.71	0.65	0.58	0.71	0.65	0.64	0.76	0.65	0.67	0.60
S2	0.71	1.00	0.67	0.64	0.61	0.65	0.69	0.72	0.80	0.72	0.73	0.62
S3	0.71	0.67	1.00	0.75	0.68	0.62	0.57	0.58	0.68	0.67	0.75	0.70
S4	0.65	0.64	0.75	1.00	0.79	0.57	0.56	0.56	0.66	0.67	0.74	0.82
S5	0.58	0.61	0.68	0.79	1.00	0.52	0.54	0.56	0.61	0.65	0.70	0.87
S6	0.71	0.65	0.62	0.57	0.52	1.00	0.70	0.66	0.68	0.59	0.60	0.55
S7	0.65	0.69	0.57	0.56	0.54	0.70	1.00	0.79	0.71	0.64	0.63	0.56
S8	0.64	0.72	0.58	0.56	0.56	0.66	0.79	1.00	0.71	0.68	0.65	0.57
S9	0.76	0.80	0.68	0.66	0.61	0.68	0.71	0.71	1.00	0.69	0.72	0.62
S10	0.65	0.72	0.67	0.67	0.65	0.59	0.64	0.68	0.69	1.00	0.77	0.66
S11	0.67	0.73	0.75	0.74	0.70	0.60	0.63	0.65	0.72	0.77	1.00	0.71
S12	0.60	0.62	0.70	0.82	0.87	0.55	0.56	0.57	0.62	0.66	0.71	1.00

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Table 4. Averaged cross-correlation of the 100 simulated series from the DKNNR model for 12 stations from Yeongnam province, South Korea.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
S1	1.00	0.69	0.69	0.63	0.57	0.69	0.64	0.63	0.74	0.63	0.66	0.59
S2	0.69	1.00	0.65	0.63	0.61	0.63	0.68	0.70	0.77	0.71	0.72	0.61
S3	0.69	0.65	1.00	0.73	0.66	0.60	0.56	0.57	0.67	0.66	0.73	0.68
S4	0.63	0.63	0.73	1.00	0.77	0.56	0.55	0.56	0.64	0.65	0.72	0.80
S5	0.57	0.61	0.66	0.77	1.00	0.51	0.53	0.55	0.60	0.64	0.69	0.84
S6	0.69	0.63	0.60	0.56	0.51	1.00	0.68	0.65	0.66	0.58	0.59	0.54
S7	0.64	0.68	0.56	0.55	0.53	0.68	1.00	0.76	0.70	0.63	0.61	0.55
S8	0.63	0.70	0.57	0.56	0.55	0.65	0.76	1.00	0.70	0.67	0.64	0.56
S9	0.74	0.77	0.67	0.64	0.60	0.66	0.70	0.70	1.00	0.68	0.71	0.61
S10	0.63	0.71	0.66	0.65	0.64	0.58	0.63	0.67	0.68	1.00	0.75	0.65
S11	0.66	0.72	0.73	0.72	0.69	0.59	0.61	0.64	0.71	0.75	1.00	0.70
S12	0.59	0.61	0.68	0.80	0.84	0.54	0.55	0.56	0.61	0.65	0.70	1.00

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Geoscientific Model Development Discussions
Open Access

Table 5. Averaged cross-correlation of 100 simulated series from the MONR model for 12 stations from Yeongnam province.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
S1	1.00	0.69	0.69	0.59	0.57	0.67	0.63	0.62	0.74	0.62	0.63	0.57
S2	0.69	1.00	0.63	0.62	0.60	0.64	0.66	0.69	0.76	0.70	0.70	0.60
S3	0.69	0.63	1.00	0.71	0.65	0.59	0.55	0.56	0.64	0.64	0.72	0.68
S4	0.59	0.62	0.71	1.00	0.78	0.54	0.54	0.54	0.63	0.62	0.70	0.78
S5	0.57	0.60	0.65	0.78	1.00	0.51	0.52	0.55	0.59	0.62	0.66	0.84
S6	0.67	0.64	0.59	0.54	0.51	1.00	0.67	0.63	0.67	0.56	0.59	0.52
S7	0.63	0.66	0.55	0.54	0.52	0.67	1.00	0.76	0.67	0.61	0.59	0.53
S8	0.62	0.69	0.56	0.54	0.55	0.63	0.76	1.00	0.69	0.65	0.62	0.54
S9	0.74	0.76	0.64	0.63	0.59	0.67	0.67	0.69	1.00	0.65	0.70	0.59
S10	0.62	0.70	0.64	0.62	0.62	0.56	0.61	0.65	0.65	1.00	0.73	0.62
S11	0.63	0.70	0.72	0.70	0.66	0.59	0.59	0.62	0.70	0.73	1.00	0.68
S12	0.57	0.60	0.68	0.78	0.84	0.52	0.53	0.54	0.59	0.62	0.68	1.00

Table 6. The difference of RMSE of cross-correlation between MONR and DKNNR. Note that the positive value indicates that the DKNNR model better performs in preserving the cross-correlation, while a negative value (underlined) shows that the MONR model better performs.

MONR-DKNNR	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
S1	0.000	0.000	0.007	0.040	0.005	0.016	0.004	0.008	0.006	0.016	0.026	0.018
S2	0.000	0.000	0.016	0.016	0.005	0.005†	0.016	0.011	0.018	0.009	0.016	0.010
S3	0.007	0.016	0.000	0.016	0.011	0.009	0.004	0.005	0.025	0.020	0.014	0.001
S4	0.040	0.016	0.016	0.000	0.002	0.018	0.013	0.015	0.016	0.027	0.023	0.020
S5	0.005	0.005	0.011	0.002	0.000	0.007	0.012	0.007	0.006	0.016	0.021	0.007
S6	0.016	0.005	0.009	0.018	0.007	0.000	0.009	0.014	0.006	0.019	0.001	0.016
S7	0.004	0.016	0.004	0.013	0.012	0.009	0.000	0.008	0.023	0.014	0.018	0.010
S8	0.008	0.011	0.005	0.015	0.007	0.014	0.008	0.000	0.010	0.017	0.024	0.015
S9	0.006	0.018	0.025	0.016	0.006	0.006	0.023	0.010	0.000	0.023	0.007	0.017
S10	0.016	0.009	0.020	0.027	0.016	0.019	0.014	0.017	0.023	0.000	0.018	0.026
S11	0.026	0.016	0.014	0.023	0.021	0.001	0.018	0.024	0.007	0.018	0.000	0.020
S12	0.018	0.010	0.001	0.020	0.007	0.016	0.010	0.015	0.017	0.026	0.020	0.000

†Underline represents a negative value implying that the MONR model better performs.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Table 7. Lag-1 cross-correlation of observed data for 12 stations from Yeongnam province, South Korea.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
S1	0.30[‡]	0.27	0.31	0.27	0.25	0.30	0.26	0.25	0.28	0.27	0.29	0.26
S2	0.28	0.29	0.29	0.27	0.25	0.29	0.28	0.27	0.30	0.29	0.32	0.26
S3	0.29	0.26	0.31	0.30	0.27	0.27	0.25	0.24	0.27	0.27	0.30	0.27
S4	0.29	0.28	0.32	0.34	0.31	0.29	0.27	0.26	0.28	0.29	0.31	0.32
S5	0.30	0.29	0.32	0.34	0.34	0.29	0.27	0.27	0.30	0.30	0.34	0.35
S6	0.25	0.22	0.26	0.24	0.23	0.28	0.24	0.22	0.25	0.23	0.25	0.23
S7	0.26	0.26	0.27	0.26	0.25	0.29	0.30	0.27	0.27	0.27	0.28	0.26
S8	0.29	0.29	0.29	0.27	0.26	0.30	0.31	0.30	0.30	0.30	0.31	0.27
S9	0.29	0.29	0.30	0.28	0.26	0.29	0.27	0.27	0.30	0.29	0.32	0.27
S10	0.29	0.31	0.32	0.31	0.29	0.29	0.30	0.30	0.32	0.33	0.34	0.29
S11	0.27	0.29	0.31	0.30	0.27	0.28	0.27	0.26	0.29	0.29	0.32	0.28
S12	0.30	0.30	0.33	0.35	0.33	0.30	0.28	0.27	0.30	0.31	0.35	0.35

[‡]Shaded values represents lag-1 autocorrelation (i.e. the one lagged correlation for the same site).

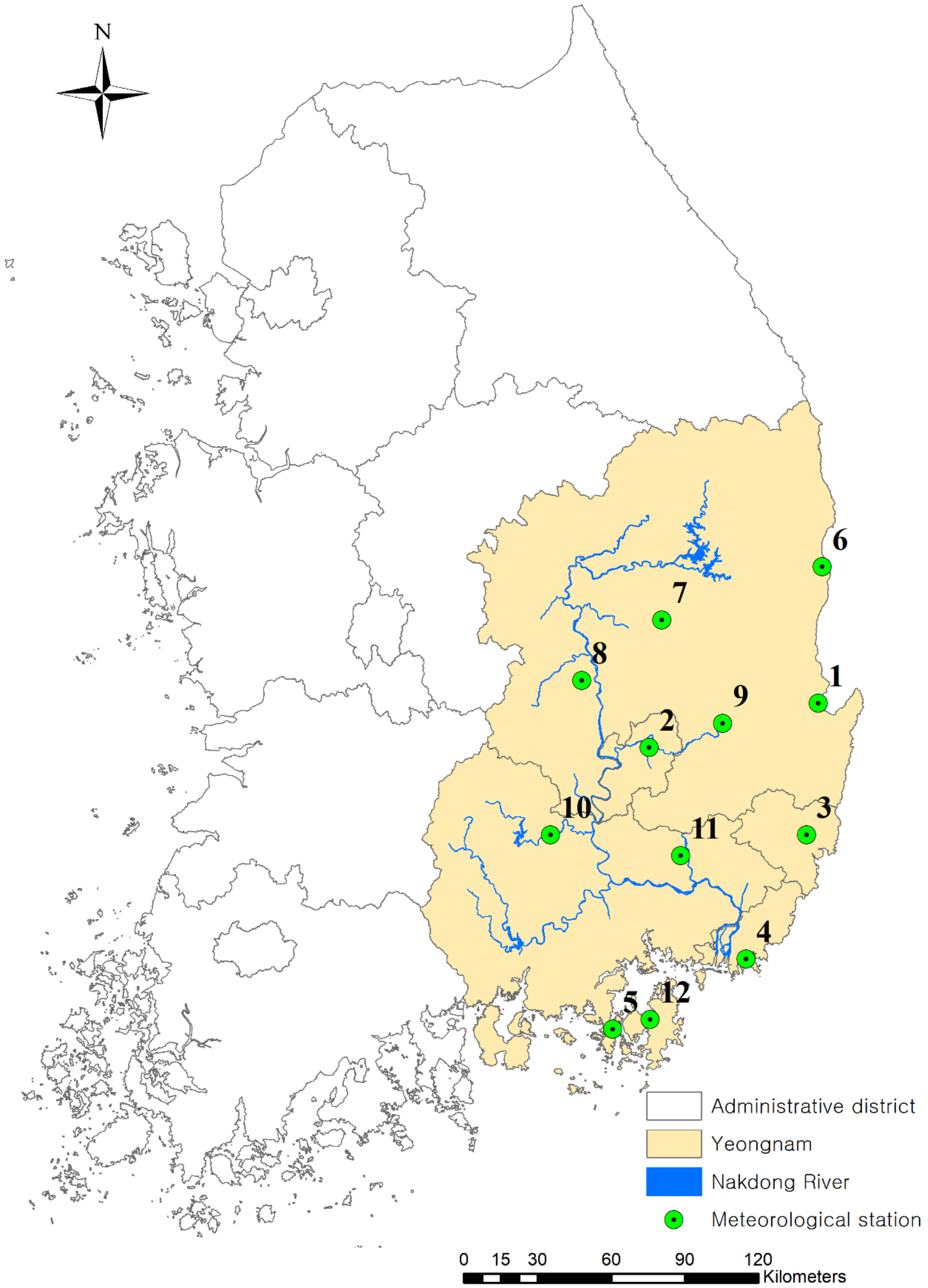
Table 8. The difference of RMSE of lag-1 cross-correlation between MONR and DKNNR. Note that a positive value indicates that the DKNNR model better performs in preserving lag-1 cross-correlation, while a negative value (underlined) shows that the MONR model better performs.

MONR-DKNNR	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
S1	0.003	0.050	0.081	0.062	0.044	0.098	0.060	0.046	0.048	0.050	0.076	0.046
S2	0.056	0.004†	0.078	0.053	0.036	0.092	0.065	0.053	0.064	0.043	0.078	0.037
S3	0.065	0.053	0.002	0.048	0.041	0.096	0.070	0.054	0.062	0.045	0.060	0.019
S4	0.093	0.084	0.087	0.001‡	0.040	0.123	0.089	0.083	0.081	0.065	0.078	0.034
S5	0.109	0.096	0.111	0.074	0.002	0.129	0.106	0.088	0.110	0.076	0.120	0.045
S6	0.031	0.016	0.062	0.043	0.044	0.001	0.020	0.017	0.031	0.029	0.046	0.029
S7	0.053	0.048	0.081	0.063	0.057	0.085	0.003	0.025	0.060	0.048	0.078	0.056
S8	0.089	0.077	0.096	0.080	0.063	0.111	0.070	0.001	0.084	0.070	0.101	0.063
S9	0.049	0.047	0.091	0.064	0.052	0.088	0.055	0.050	0.004	0.064	0.084	0.055
S10	0.085	0.094	0.107	0.090	0.065	0.123	0.107	0.093	0.106	0.000	0.095	0.061
S11	0.065	0.064	0.076	0.054	0.036	0.096	0.081	0.062	0.064	0.032	0.001	0.034
S12	0.118	0.102	0.105	0.080	0.043	0.138	0.108	0.096	0.115	0.093	0.120	0.000

†Underline represents a negative value implying that the MONR model better performs.

‡Shaded values represent lag-1 autocorrelation (i.e. the lagged-1 correlation for the same site).

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018



532

533 Figure 1. Locations of 12 selected weather stations at the Yeongnam province. See Table 1 for
534 further information about the stations.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Geoscientific Model Development Discussions
Open Access

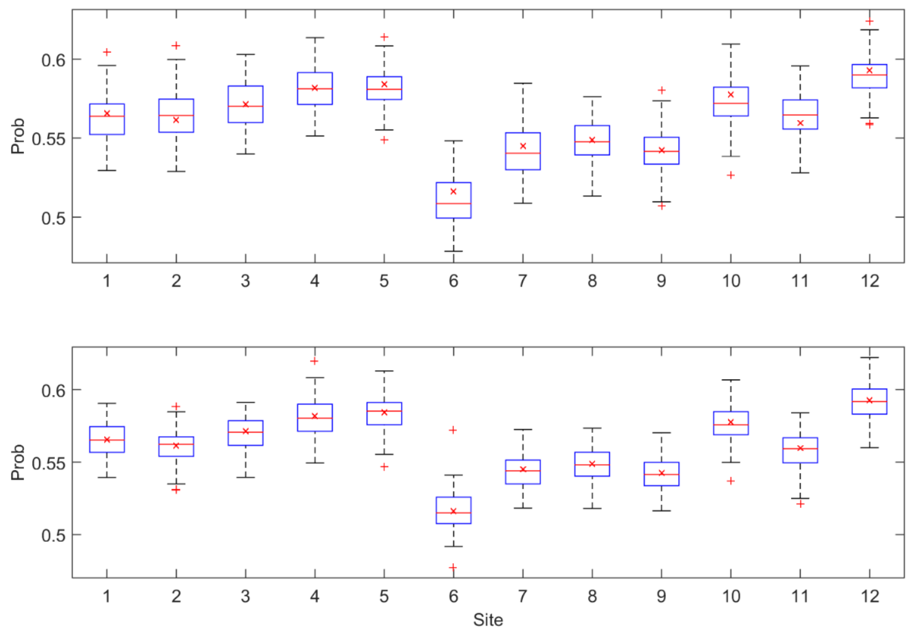


Figure 2. Boxplots of the P11 probability for the simulated data from the DKNNR model (top panel) and the MONR model (bottom panel) as well as the observed (x marker) for the 12 selected weather stations from the Yeongnam province.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Geoscientific Model Development Discussions
Open Access

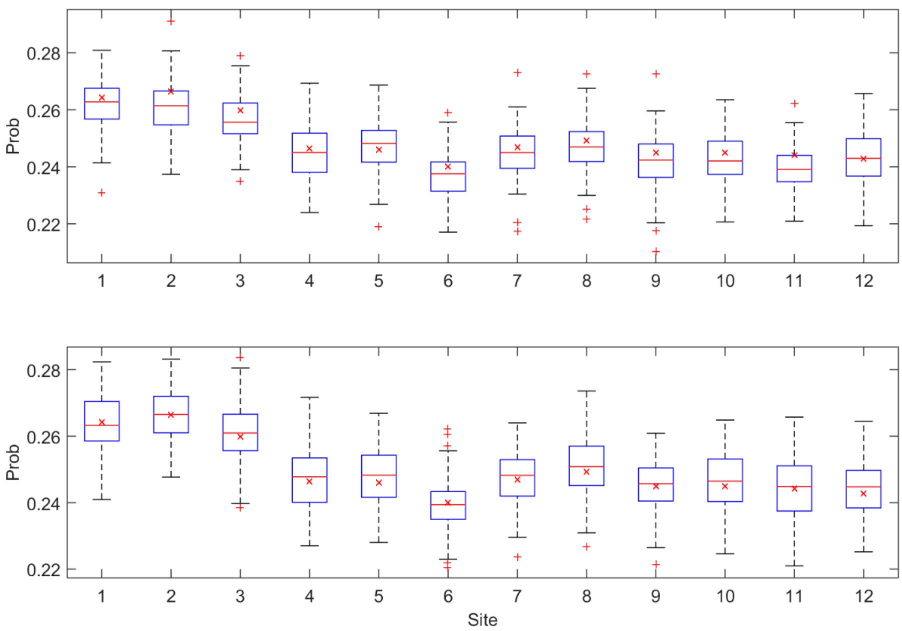


Figure 3. Boxplots of the P01 probability for the data simulated from the DKNNR model (top panel) and the MONR model (bottom panel) as well as the observed (x marker) for the 12 selected weather stations from the Yeongnam province.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Geoscientific Model Development
Discussions
Open Access

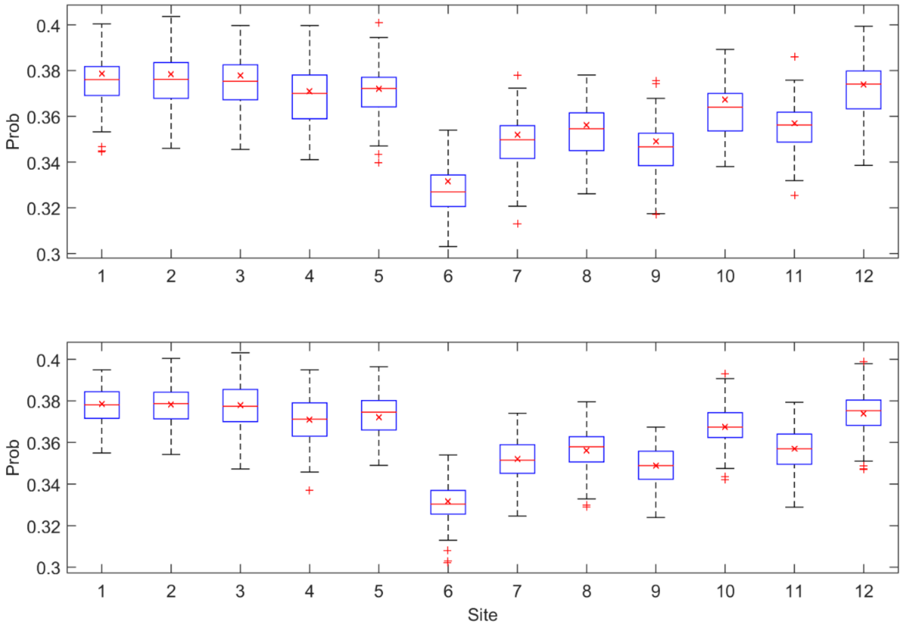


Figure 4. Boxplots of the P1 probability for the data simulated from the DKNNR model (top panel) and the MONR model (bottom panel) as well as the observed (x marker) for the 12 selected weather stations from the Yeongnam province.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Geoscientific Model Development Discussions
Open Access

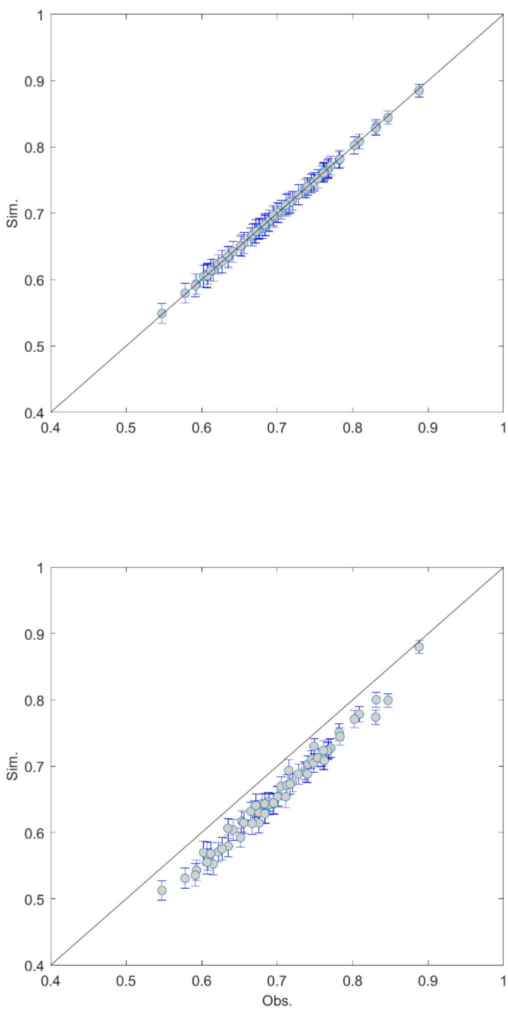


Figure 5. Scatterplot of cross-correlations between 12 weather stations for the observed data (X coordinate) and the generated data (Y coordinate) generated from the DKNNR model (top panel) and the MONR model (bottom panel). The cross-correlations from 100 generated series are averaged for the filled circle and the errorbars upper and lower extended lines indicate the range of 1.95×standard deviation.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Geoscientific
Model Development
Discussions
Open Access

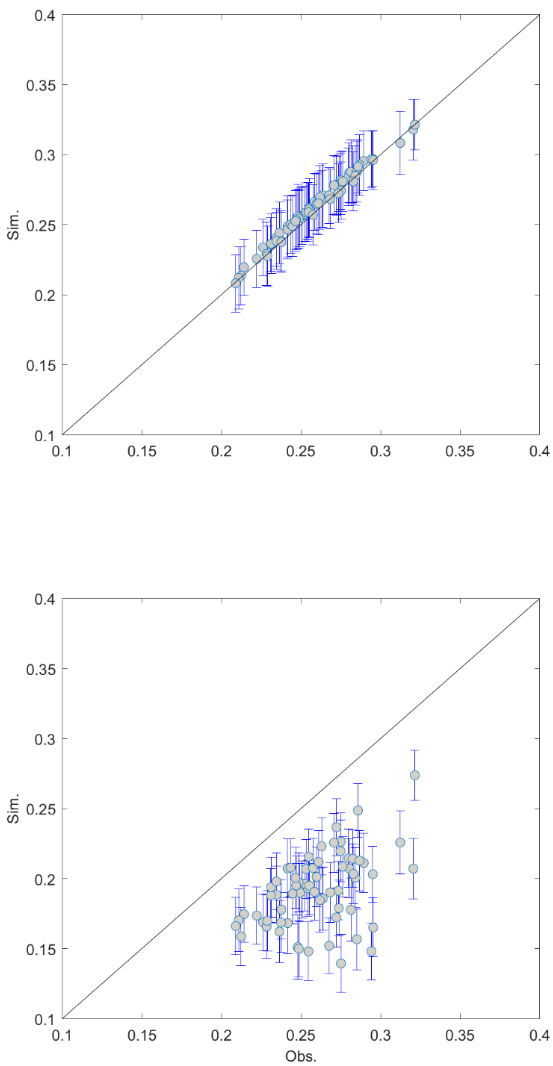


Figure 6. Scatterplot of lag-1 cross-correlations between 12 weather stations for the observed data (X coordinate) and the generated data (Y coordinate) generated from the DKNNR model (top panel) and the MONR model (bottom panel). The cross-correlations from 100 generated series are averaged for the filled circle and the errorbars upper and lower extended lines indicate the range of 1.95×standard deviation.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Geoscientific Model Development Discussions
Open Access

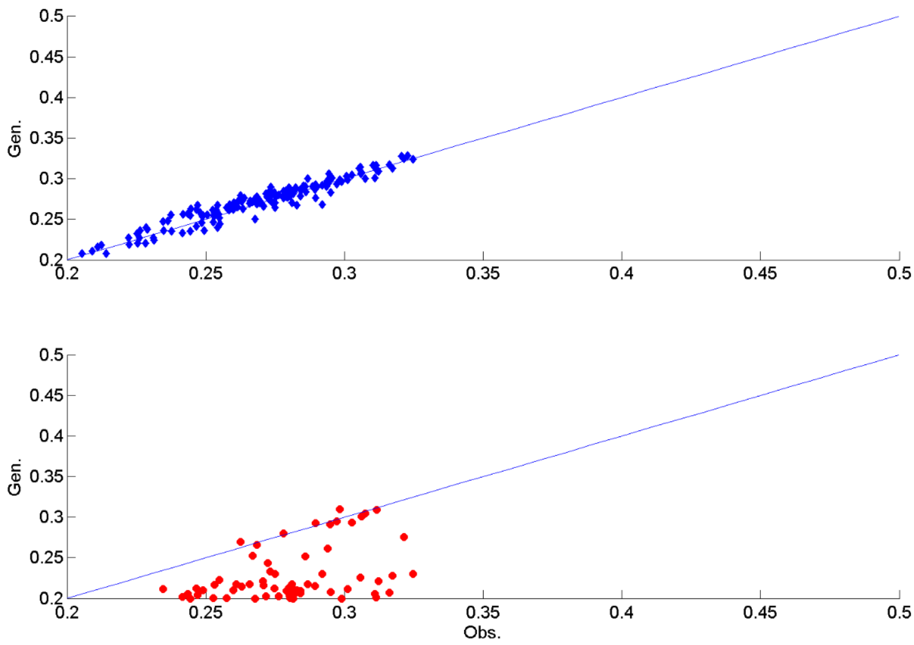


Figure 7. Scatterplot of lag-1 cross-correlations between 12 weather stations for the observed data (X coordinate) and the generated data (Y coordinate) generated from the DKNNR model (top panel) and the MONR model (bottom panel) with the whole year data not with the summer season. The cross-correlations from 100 generated series are averaged.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Geoscientific Model Development
Discussions

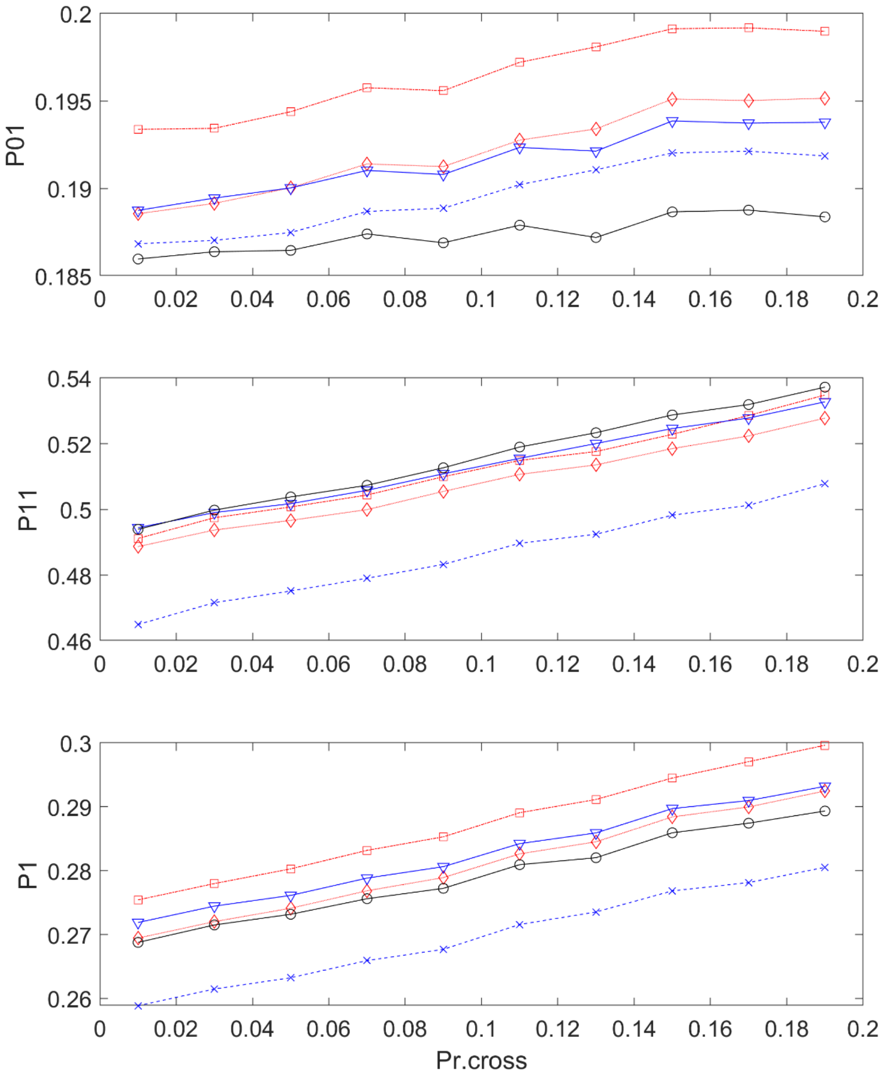


Figure 8. Transition probabilities and marginal distribution for the selected five stations along with changing the cross-over probability P_{cr} with the condition that the candidate value is one and the previous value is also one. See Eq.(15) for the detail.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Geoscientific Model Development Discussions

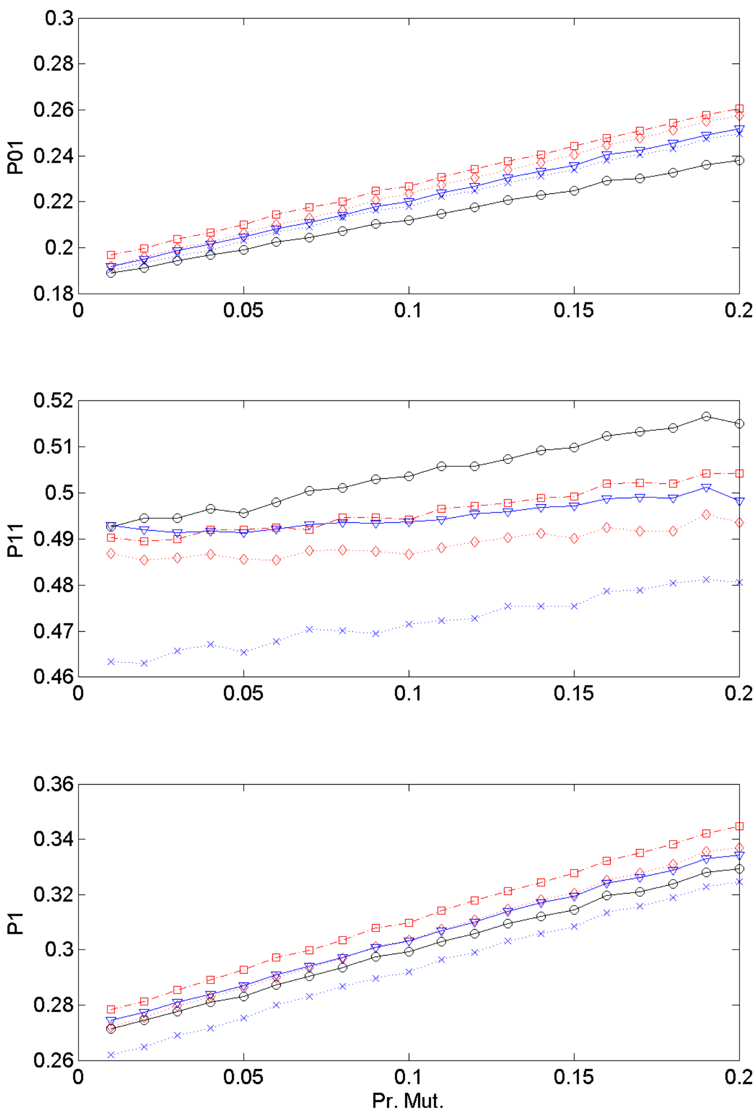


Figure 9. Transition probabilities and marginal distribution along with changing the cross-over probability with the condition that the mutation is processed only if the candidate value is one. See Eq.(16) for the detail.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Geoscientific Model Development
Discussions
Open Access

Table A 1. Example dataset of daily rainfall with 12 weather stations and 16 days for measured rainfall (mm) in the upper part of this table and its corresponding occurrences in the bottom part of this table.

Day	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
1	2.0	2.9	1.2	0.0	0.0	1.8	4.0	8.9	2.0	4.6	1.3	0.6
2	52.6	39.8	47.2	17.4	11.8	31.0	30.0	33.7	52.0	57.8	37.0	17.5
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.2	1.0	1.4	1.9	12.3	0.0	0.0	0.0	0.7	3.1	3.5	8.1
6	14.8	0.2	0.8	0.2	5.0	0.0	0.0	18.0	0.0	0.0	0.6	3.1
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	1.0	0.0	0.4	0.0	3.8	0.0	0.1	0.0	0.0	0.0	0.0
11	7.1	6.4	12.8	12.8	13.6	2.3	2.0	5.4	6.0	7.3	16.4	20.3
12	0.0	0.0	0.0	0.0	5.5	0.0	0.0	0.0	0.0	0.0	0.0	4.3
13	10.0	1.6	11.6	14.3	1.5	5.4	0.0	0.0	2.5	0.0	2.7	16.1
14	2.3	0.0	0.7	0.0	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.0
15	31.5	4.3	30.6	12.7	14.4	25.8	3.5	0.8	5.0	2.7	6.5	20.3
16	37.0	7.8	30.1	11.2	9.6	36.8	2.5	4.7	13.5	1.7	10.1	14.1
Day	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
1	1	1	1	0	0	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1	1	1
3	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0
5	1	1	1	1	1	0	0	0	1	1	1	1
6	1	1	1	1	1	0	0	1	0	0	1	1
7	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0
10	0	1	0	1	0	1	0	1	0	0	0	0
11	1	1	1	1	1	1	1	1	1	1	1	1
12	0	0	0	0	1	0	0	0	0	0	0	1
13	1	1	1	1	1	1	0	0	1	0	1	1
14	1	0	1	0	0	1	0	0	0	0	0	0
15	1	1	1	1	1	1	1	1	1	1	1	1
16	1	1	1	1	1	1	1	1	1	1	1	1

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Geoscientific Model Development
Discussions
Open Access

Table A 2. Example dataset for estimating distances. The second row presents the current daily precipitation occurrences for 12 stations and the rows below show the absolute difference between the current occurrences (**Xc**) and the observed data in Table A 1. The last column presents the distances in Eq. (11).

day	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	Dist
Xc	*0*	*1*	*1*	*0*	*0*	*1*	*1*	*0*	*0*	*0*	*0*	*0*	
1	1	0	0	0	0	0	0	1	1	1	1	1	**6**
2	1	0	0	1	1	0	0	1	1	1	1	1	**8**
3	0	1	1	0	0	1	1	0	0	0	0	0	**4**
4	0	1	1	0	0	1	1	0	0	0	0	0	**4**
5	1	0	0	1	1	1	1	0	1	1	1	1	**9**
6	1	0	0	1	1	1	1	1	0	0	1	1	**8**
7	0	1	1	0	0	1	1	0	0	0	0	0	**4**
8	0	1	1	0	0	1	1	0	0	0	0	0	**4**
9	0	1	1	0	0	1	1	0	0	0	0	0	**4**
10	0	0	1	1	0	0	1	1	0	0	0	0	**4**
11	1	0	0	1	1	0	0	1	1	1	1	1	**8**
12	0	1	1	0	1	1	1	0	0	0	0	1	**6**
13	1	0	0	1	1	0	1	0	1	0	1	1	**7**
14	1	1	0	0	0	0	1	0	0	0	0	0	**3**
15	1	0	0	1	1	0	0	1	1	1	1	1	**8**
16	1	0	0	1	1	0	0	1	1	1	1	1	**8**

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Geoscientific Model Development Discussions

Table A 3. Example for selecting one sequence for $\mathbf{X}_{c+1}$. The second row presents the distances in Table A 2. The third and fourth columns show the sorted days and distances for the smallest distances to the largest in the second column. The fourth row presents the probabilities estimated with Eq. (12). Note that there are six days whose distances are the same with each other. In this case all the days are included and among six days, one is selected with equal probabilities.

Day	Dist.	Sorted Day	Sorted Dist	Prob
1	6	14	3	0.48
2	8	3	**4**	0.24
3	4	4	**4**	0.16
4	4	7	**4**	0.12
5	9	8	**4**	
6	8	9	**4**	
7	4	10	**4**	
8	4	1	6	
9	4	12	6	
10	4	13	7	
11	8	2	8	
12	6	6	8	
13	7	11	8	
14	3	15	8	
15	8	16	8	
16	8	5	9	

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-181
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 12 October 2018

Table A 4. Example for GA mixture for $\mathbf{X_{c+1}}$. The second and third rows present two selected sets, while the third row shows the final set for $\mathbf{X_{c+1}}$ with the crossover at S6 and S8 and the mutation for S12.

	Assigned day, *p*	Selected day, *p*+1	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
Set1	14	**15**	1	0	0	1	1	0	0	1	1	1	1	1
Set2	4	**5**	1	0	0	1	1	1	1	0	1	1	1	1
Final			1	0	0	1	1	<u>1</u>	0	<u>0</u>	1	1	1	**0**