ORIGINAL PAPER

Stochastic simulation of precipitation data for preserving key statistics in their original domain and application to climate change analysis

Taesam Lee

Abstract We propose a new method to estimate autoregressive model parameters of the precipitation amount process using the relationship between original and transformed moments derived through a moment generating function. We compare the proposed method with the traditional parameter estimation method, which uses transformed data, by modeling precipitation data from Denver International Airport (DIA), CO. We test the applicability of the proposed method (M2) to climate change analysis using the RCP 8.5 scenario. The modeling results for the observed data and future climate scenario indicate that M2 reproduces key historical and targeted future climate statistics fairly well, while M1 presents significant bias in the original domain and cannot be applied to climate change analysis.

1 Introduction

Stochastic modeling and simulation of daily rainfall has been a prominent subject in hydrology and water resources for several decades. The simulation of daily rainfall can be used for agricultural operations, for the design of irrigation systems, and as an input for rainfall-runoff studies. These simulation models have also been employed in climate change studies (Lee et al. 2012; Mezghani and Hingray 2009). An important criterion for stochastic modeling is reproducibility of the statistical characteristics of observed data. Target characteristics for daily rainfall would be the occurrence and amount processes. The scaling behavior and over-dispersion (tendency to

T. Lee (✉)
Department of Civil Engineering, ERI, Gyeongsang National University, 501 Jinju-daero, Jinju, Kyungnam, South Korea
e-mail: tae3lee@gnu.ac.kr

underestimate the observed variance of larger time scale data) of generated data have also been considered (Burlando and Rosso 1996; Katz and Zheng 1999).

A number of models have been developed to explain these processes. Traditionally, stochastic modeling of precipitation separates daily rainfall into two processes to account for intermittency.

The occurrence process was first modeled by Gabriel and Neumann (1962) using a Markov Chain (MC) for Tel Aviv daily rainfall data. The MC model assumes that the probability of rainfall for a certain day depends on that for the previous day. The MC model has been further developed in several studies (e.g., Dennett et al. 1983; Guttorp and Minin 1995; Yoon et al. 2013; Berne et al. 2004).

Meanwhile, Burlando and Rosso (1991) applied the discrete autoregressive moving average (DARMA) model, which was originally formulated by Eagleson (1978), to the occurrence process of precipitation data. An alternative model for the occurrence process is renewal, defined as a sequence of alternating wet and dry intervals (Hingray and Ben Haha 2005; Roldan and Woolhiser 1982). Several other models have been developed to explain the occurrence process, such as the non-homogeneous hidden Markov model (Hughes et al. 1999). Lall et al. (1996) used a nonparametric approach for wet- and dry-spell length that is similar to the renewal model but uses a discrete fernel estimator.

Several probability distributions have been used to model daily rainfall amount. Rainfall amount is so highly skewed that an appropriate distribution must be chosen, or the data should be transformed to fit Gaussian-based time series models. Todorovic andf Woolhiser (1975) chose the exponential distribution to describe the amount process; this distribution has been applied by other authors (Richardson 1981; Wilby 1994). A two-parameter gamma distribution, with scale and shape parameters, has also been studied (e.g., Eagleson

1978; Katz 1996; Koutsoyiannis and Onof 2001). An alternative has been the use of a mixture of two single-parameter exponential distributions (Lebel et al. 1998; Wilks 1999; Woolhiser and Roldan 1982).

Transformations have been applied to skewed data for normalization, and the autoregressive moving average (ARMA) has been used in the study of successive rainfall amounts (Katz and Parlange 1993, 1995). Katz (1999) experimented with the power transformation and derived a direct relationship between the original and transformed data for moments and autocorrelation functions. This study determined that the power transformation stabilizes the variance and amplifies the autocorrelation of the amount of consecutive rainy days. Hannachi (2012) applied the autoregressive-1 (AR-1) model to daily precipitation from the Northern Ireland Armagh Observatory. Aronica and Bonaccorso (2013) applied a first-order Markov Chain and ARMA model to qualitatively assess the impact of climate change on the hydrological regime of the Alcantara River Basin in Italy.

Time-dependent models for precipitation data are not favored due to model complexity. In the current study, we normalized the data on the amount of precipitation with a power transformation to test a time-dependent AR model. We applied a simple Markov Chain to precipitation occurrence (X process). We used a novel method to estimate AR model parameters using the relationship between original and transformed moments through the moment-generating function. Finally, we compared our proposed method with the traditional parameter estimation method from the transformed domain.

This paper is organized as follows: we present the mathematical description in Section 2, followed by data description and application methodology in Section 3. The results are shown in Section 4. Finally, the summary and conclusions are presented in Section 5.

2 Mathematical description

The model we applied to describe the daily rainfall process had the following form:

$$Y = X \cdot Z \tag{1}$$

where Y represents a positive intermittent variable of the daily rainfall process, X is the occurrence variable, and Z is the amount variable. X and Z are assumed to be independent and periodic on a monthly basis.

2.1 Modeling the amount of precipitation

We applied a simple first-order AR-1 model because daily precipitation data do not present long-term memory and

involve intermittency (Hannachi 2012). We tested two-parameter estimation methods, the traditional direct method (M1) and the indirect method (M2), each month to determine how well key statistics were preserved on daily, monthly, and yearly scales.

2.1.1 AR-1 model with direct parameter estimation (M1)

Katz and Parlange (1995) applied an ARMA model to account for the time dependency of precipitation events; because rainfall amount is highly positively skewed, these data are normalized. Power and log transformations are generally used. Katz (1999) applied a power transformation to daily rainfall amount. This transformation enables choosing of the most appropriate power value to match the magnitude of skewness.

The following equations describe the AR-1 model with the power transformation for the daily rainfall variable at time t (Z_t):

$$N_t = Z_t^{1/C} \tag{2}$$

$$N_t = \mu_N + \phi_{1,N}(N_{t-1} - \mu_N) + \varepsilon_t \tag{3}$$

where μ_N, σ_ε, and $\phi_{1,N}$ are the model parameters indicating the mean, standard deviation of the random component, and the first-order autocorrelation coefficient, respectively, for the transformed domain variable N. ε_t represents the white noise or error term with a normal distribution $N(0, \sigma_\varepsilon^2)$, where σ_ε^2 is the unexplained variance from the autoregressive model in Eq. (3). Additionally, c represents the power exponent, for which only an integer number is generally used (i.e., $c = 1, 2, 3, \ldots$).

Note that the transformed variable N_t is assumed to be normal with the condition that precipitation occurs (i.e., $X_t = 1$). Therefore, the parameters are also conditional. For example, the mean μ_N is $E(N_t | X_t = 1)$ rather than $E(N_t)$. The condition was not included in the notation for simplification in the current work.

The model parameters have been estimated through the method of moments, which uses the following relationship between model moments and calculated moments from observed data (Salas 1993):

$$\widehat{\mu}_N = \frac{1}{n} \sum_{t=1}^{n} N_t \tag{4}$$

$$\widehat{\sigma}_N^2 = \frac{1}{n-1} \sum_{t=1}^{n} \left(N_t - \widehat{\mu}_N \right)^2 \tag{5}$$

$$\widehat{\phi}_{1,N} = \widehat{\rho}_{1,N} = \frac{\frac{1}{n-1} \sum_{t=2}^{n} (N_t - \mu_N)(N_{t-1} - \mu_N)}{\widehat{\sigma}N^2} \tag{6}$$

$$\hat{\sigma}_{\varepsilon}^2 = \left(1 - \hat{\phi}_{1,N}^2\right)\hat{\sigma}_N^2 \quad (7)$$

Note that model parameters were estimated each month; i.e., a different parameterization was applied each month.

Parameters estimated from Eqs. (4)–(7) are used to generate N_t in the simulation. The generated N_t data should be back-transformed to obtain $Z_t = N_t^c$. Finally, Z_t is multiplied by X_t to obtain Y_t.

2.1.2 AR-1 model with the indirect parameter estimation method (M2)

Bias in key statistics has been reported when using M1 for the AR-1 model. In this section, we describe an indirect method to estimate model parameters from the relationship between the Z and N processes. This relationship is explicitly described by Katz (1999) as follows:

$$E(Z_t) = E\left(N_t^c\right) \quad (8)$$

$$E\left(Z_t^2\right) = E\left(N_t^{2c}\right) \quad (9)$$

$$E(Z_t Z_{t+l}) = E\left(N_t^c N_{t+l}^c\right) \quad (10)$$

where $l = 0, 1, 2, \ldots$ is the time lag. Note that the proposed method is only valid for cases in which the exponent c in Eq. (2) is an integer.

The moment relationship between variables Z and N can be derived from the moment generating function of a bivariate normal distribution with two dummy random variables U_1 and U_2:

$$M_{U_1 U_2}(\mu_1, u_2) =$$
$$\exp\left[\mu_1 u_1 + \mu_2 u_2 + \frac{1}{2}\left(\sigma_1^2 u_1^2 + \sigma_2^2 u_2^2 + 2\rho\sigma_1\sigma_2 u_1 u_2\right)\right] \quad (11)$$

The product moments are denoted as follows:

$$\gamma(j,k) = E\left(U_1^j, U_2^k\right) \quad (12)$$

The repeated partial differentiation of Eq. (11) with $u_1 = u_2 = 0$ yields the following two-dimensional recursion:

$$\gamma(j,0) = \mu_1\gamma(j-1,0) + (j-1)\sigma_1^2\gamma(j-2,0) \quad (13)$$

$$\gamma(0,k) = \mu_1\gamma(0,k-1) + (k-1)\sigma_2^2\gamma(0,k-2) \quad (14)$$

$$\gamma(1,k) = \mu_2\gamma(1,k-1) + \sigma_1\sigma_2\rho\gamma(0,k-1) + (k-1)\sigma_2^2\gamma(1,k-2) \quad (15)$$

$$\gamma(j,k) = \mu_1\gamma(j-1,k) + k\sigma_1\sigma_2\rho\gamma(j-1,k-1) + (j-1)\sigma_1^2\gamma(j-2,k) \quad (16)$$

The initial condition for this recursion is $\gamma(0,0)=1, \gamma(1,0)=\mu_1, \gamma(0,1)=\mu_2, \gamma(1,1)=\mu_1\mu_2+\sigma_1\sigma_2\rho$. Equations (13) and (14) will provide the univariate moments, while Eqs. (15) and (16) will provide the lagged correlation. By substituting the derivatives of Eqs. (13)–(16) into Eqs. (8)–(10), the high moments of the power transformed variable, N, are described as follows:

$$E(Z_t) = E\left(N_t^c\right) = \gamma(c,0) = \mu_N\gamma(c-1,0) + (c-1)\sigma_N^2\gamma(c-2,0) \quad (17)$$

$$E\left(Z_t^2\right) = E\left(N_t^{2c}\right) = \gamma(2c,0)$$
$$= \mu_N\gamma(2c-1,0) + (2c-1)\sigma_N^2\gamma(2c-2,0) \quad (18)$$

$$E(Z_t, Z_{t+1}) = E\left(N_t^c, N_{t+1}^c\right) = \gamma(c,c) = \mu_N\gamma(c-1,c) + c\sigma_N^2\rho_{1,N}\gamma(c-1,c-1) + (c-1)\gamma(c-2,c) \quad (19)$$
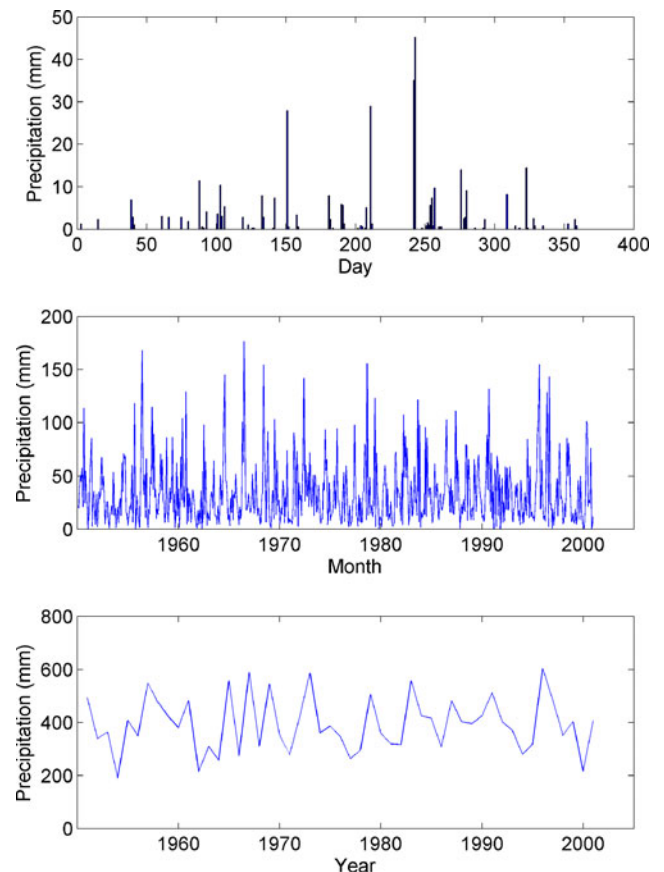

Fig. 1 Time series of daily (*top panel*) and monthly (*middle panel*) precipitation for 1999 and yearly (*bottom panel*) precipitation for all data (1950–2003)

Table 1 Estimated parameters from the AR-1 model for method 1 (M1) and method 2 (M2)

Month	M1			M2		
	μ_N	σ_ε	$\phi_{1,N}$	μ_N	σ_ε	$\phi_{1,N}$
1	1.029	0.152	0.005	1.027	0.164	0
2	1.054	0.153	0.033	1.059	0.151	0
3	1.095	0.171	0.179	1.074	0.18	0.448
4	1.127	0.191	0.006	1.127	0.197	0
5	1.12	0.204	0.204	1.131	0.201	0.273
6	1.101	0.182	0.31	1.096	0.196	0.303
7	1.114	0.205	0.074	1.125	0.207	0.023
8	1.087	0.201	0.113	1.098	0.187	0.427
9	1.104	0.186	0.12	1.117	0.18	0.041
10	1.126	0.191	0.152	1.151	0.155	0.362
11	1.099	0.168	0.157	1.119	0.137	0.339
12	1.055	0.147	0.169	1.027	0.184	0.129

Note that from the relationship between Z and N, N variable statistics, such as the mean, standard deviation, and autocorrelation, can be estimated from Z variable statistics. To complete the relationship in Eqs. (17)–(19), recursion is required; that is, $\gamma(c-2,0)$ and $\gamma(c-1,0)$ are required for $\gamma(c,0)$, $\gamma(c-3, 0)$ and $\gamma(c-2,0)$ are required for $\gamma(c-1,0)$, and so on. The moments of the Z variable can be estimated from the arithmetic mean of sampled data with the given formulation.

Instead of using moments, common statistics such as variance and autocorrelation coefficients can be used for the relationship, as follows:
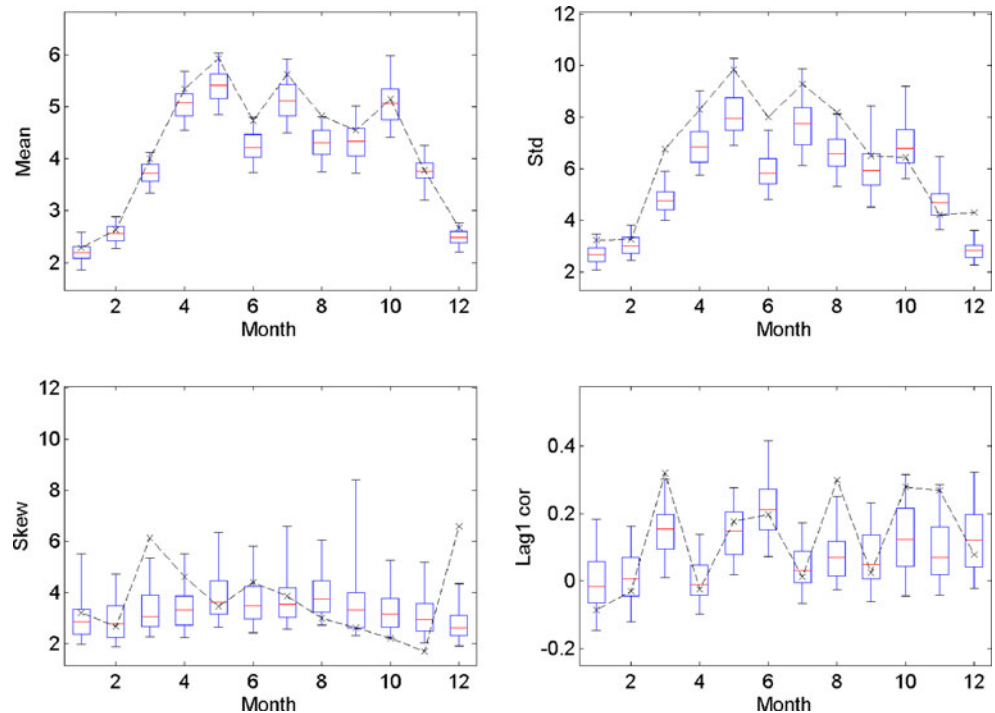
$$\sigma_Z^2 = EZ^2 - (EZ)^2 = E(N^{2c}) - (E(N^c))^2 = \gamma(2c,0) - (\gamma(c,0))^2 \tag{20}$$

$$\rho_{1,z} = \frac{\text{Cov}(N_t^c, N_{t+1}^c)}{\text{Var}(N_t^c)} = \frac{E(N_t^c, N_{t+1}^c) - (E(N_t^c))^2}{E(N_t^{2c}) - (E(N_t^c))^2} \tag{21}$$

$$= \frac{\gamma(c,c) - (\gamma(c,0))^2}{\gamma(2c,0) - (\gamma(c,0))^2}$$

where $\widehat{\mu}_N$, $\widehat{\sigma}_N$, and $\widehat{\rho}_N$ are estimated from the moment relationship.

Note that to perform the stochastic simulation of daily rainfall, $\widehat{\mu}_N$, $\widehat{\phi}_{1,N}$, and $\widehat{\sigma}_\varepsilon^2$ are required. $\widehat{\mu}_N$ is estimated from numerically solving Eqs. (17) and (20) as well as $\widehat{\sigma}_N^2$. $\widehat{\phi}_{1,N}$ is calculated from Eq. (21). $\widehat{\phi}_\varepsilon^2$ is estimated from $\widehat{\sigma}_N^2$ and $\widehat{\phi}_{1,N}$ as in Eq. (7). One example of $c=2$ is presented in Appendix 1 for further clarification.

2.2 Modeling precipitation occurrence

Several models have been developed to simulate the occurrence process (X) of daily rainfall. Gabriel and Neumann (1962) applied a binary occurrence process for wet and dry days by Discrete ARMA (DARMA(p,q)). The simple DARMA(1,0) is described as follows:

$$X_t = V_t X_{t-1} + (1 - V_t) W_t \tag{22}$$

Fig. 2 Key statistics of precipitation amount for historical (DIA station, *dotted line with cross mark*) and M1 simulated (*boxplot*) data. The *box* represents the interquartile range (IQR), *whiskers* are maximum and minimum, and the *horizontal line* is the median for 100 simulations

Fig. 3 Key statistics of precipitation amount for historical (DIA station, *dotted line with cross mark*) and M2 simulated (*boxplot*) data. The *box* represents the interquartile range (IQR), *whiskers* are maximum and minimum, and the *horizontal line* is the median for 100 simulations
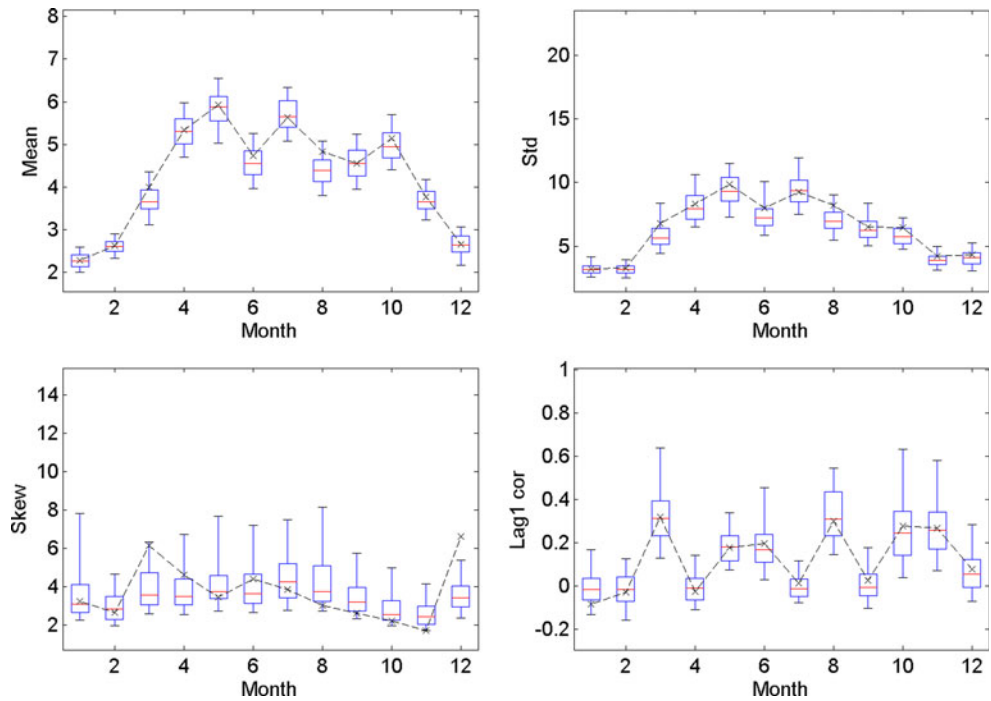
where all variables represent binary processes, i.e., W_t, V_t, and $X_t \in \{0,1\}$,

$$\Pr[W_t = 1] = \xi \quad \text{and} \quad \Pr[V_t = 1] = \lambda \qquad (23)$$

Therefore, the process X_t can be written as follows:

$$X_t = \begin{cases} X_{t-1} & \text{with probability } \lambda \\ W_t & \text{with probability } 1-\lambda \end{cases} \qquad (24)$$

The DARMA(1,0) model is equivalent to the Markov Chain, which has been applied broadly to model the rainfall process since Gabriel and Neumann (1962). The Markov Chain model is expressed with the following transition probability matrix:

$$P = \begin{vmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{vmatrix} \qquad (25)$$

where $p_{ab} = \Pr[X_t = b | X_{t-1} = a]$ and $a, b \in \{0,1\}$, with the limiting distributions $p_0 = \Pr[X_t = 0]$ and $p_1 = \Pr[X_t = 1]$.

Fig. 4 Transition and limiting probabilities of precipitation occurrence for historical (DIA station, *dotted line with cross mark*) and M1 and M2 simulated (*boxplot*) data. The occurrence process (*X*) model was the same for M1 and M2, making the results the same. The *box* represents the interquartile range (IQR), *whiskers* are maximum and minimum, and the *horizontal line* is the median for 100 simulations
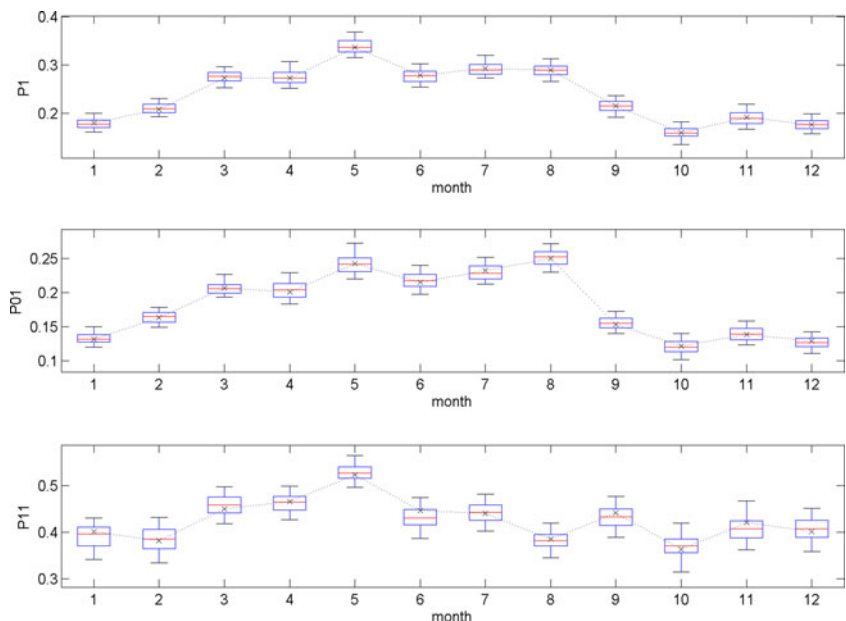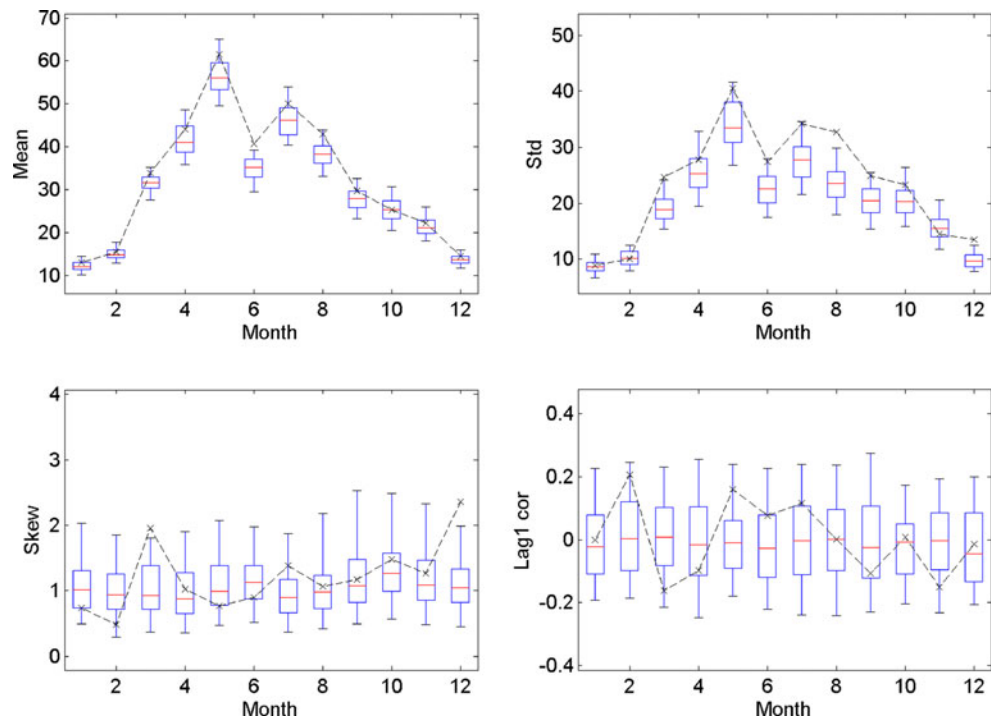
Fig. 5 Key statistics of monthly precipitation for historical (*dotted line with cross mark*) and M1 simulated (*boxplot*) data. The *box* represents the interquartile range (IQR), *whiskers* are maximum and minimum, and the *horizontal line* is the median for 100 simulations

These two models (DARMA(1,0) and Markov Chain) are equivalent to the following parametric relationship:

$$\begin{vmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{vmatrix} = \begin{vmatrix} \lambda + (1-\lambda)(1-\xi) & (1-\lambda)\xi \\ (1-\lambda)(1-\xi) & \lambda + (1-\lambda)\xi \end{vmatrix} \quad (26)$$

The components of the transition probability matrix are estimated using maximum likelihood (Marani 2003) from precipitation occurrence data as follows:

$$\widehat{p}_{ab} = \frac{n(a,b)}{n(a)} \quad (27)$$

Fig. 6 Key statistics of monthly precipitation for historical (*dotted line with cross mark*) and M2 simulated (*boxplot*) data. The *box* represents the interquartile range (IQR), *whiskers* are maximum and minimum, and the *horizontal line* is the median for 100 simulations
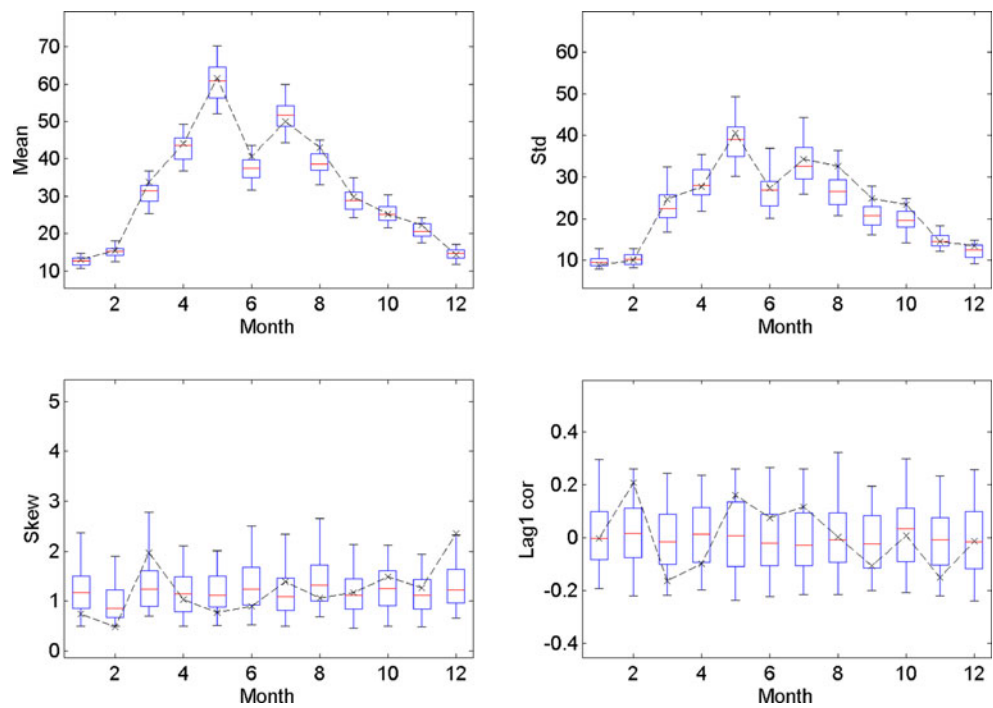
Fig. 7 Cumulative distribution function (CDF) of annual maximum daily precipitation for historical (*dashed red line*) and M1 (*solid black line*) and M2 (*dotted gray line*) simulated data. Note that longer series were used for the simulation (i.e., 500 instead of 51 years; equal to the historical record length) to evaluate the overall CDF behavior. Simulation CDFs are averaged over 100 simulated CDFs
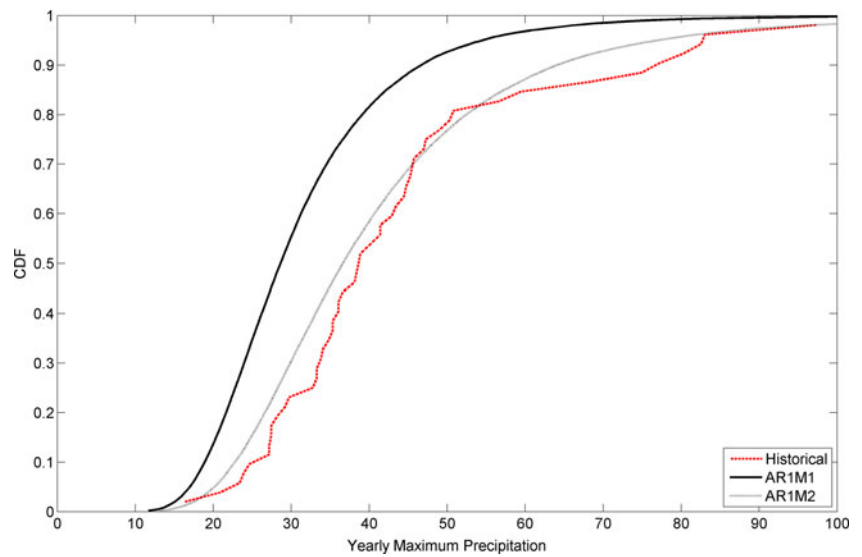
where $n(a, b)$ is the number of times that rainfall in state a transitions to state b, and $n(a)=n(a,0)+n(a,1)$.

3 Data description and application methodology

3.1 Observational data

We used daily precipitation data measured between 1950 and 2003 from the Denver International Airport (DIA) station (39.83° N, 104.66° W); we did not consider more recent data because the station had been moved. A year-long daily precipitation time series is presented in the top panel of Fig. 1; monthly and annual precipitation for all the data are shown in the middle and bottom panels, respectively. Daily precipitation indicates higher occurrence and amount during the summer (days 150–250). Therefore, we used different parameters each month to account for this annual cycle while applying the

same power transformation exponent (c) for Eq. (2). Monthly precipitation data show a similar annual cycle, whereas yearly precipitation shows stationary variation.

3.2 Employed climate scenario

A climate scenario was used to illustrate how our parameter estimation method applies to climate change analysis. We selected RCP 8.5 from the Coupled Model Intercomparison Project phase 5 (CMIP5) from NOAA's Geophysical Fluid Dynamics Laboratory (GFDL) to validate the performance of our method. The Bias-Correction Constructed Analogue (BCCA) method (Maurer et al. 2010) was applied for the
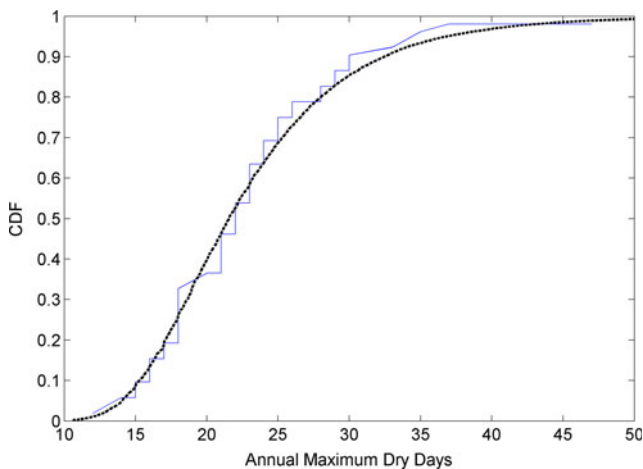

Fig. 8 Annual maximum dry days from historical (*solid line*) and simulated (*dashed line*) data

Table 2 Percent differences in the mean and standard deviation of precipitation data between the reference period (1950–1999) and three future periods (period 1, 2010–2039; period 2, 2040–2069; and period 3, 2070–2099)

	Mean			Standard deviation		
	Period 1	Period 2	Period 3	Period 1	Period 2	Period 3
Mon1	34.1	29.6	53.6	55.1	32.6	51.3
Mon2	−1.0	22.0	32.6	−16.5	12.2	51.4
Mon3	18.8	31.4	20.1	9.3	1.7	4.1
Mon4	5.3	28.9	19.4	17.4	31.6	27.8
Mon5	21.7	8.9	29.0	26.8	−1.8	25.7
Mon6	−13.0	−7.8	−14.4	−15.8	−15.3	−15.3
Mon7	−12.4	−0.7	6.1	−13.6	−0.4	−6.1
Mon8	1.2	23.3	13.4	−3.5	2.1	−10.0
Mon9	15.4	4.2	−3.9	23.6	21.1	−5.8
Mon10	2.3	24.5	−5.9	22.5	36.1	16.8
Mon11	11.4	6.1	26.6	10.0	2.1	22.5
Mon12	28.9	51.0	104.2	31.4	51.3	77.4

Fig. 9 Key statistics of monthly precipitation for historical (DIA station, *dotted line with cross*), climate change scenario (RCP 8.5, *solid line with circle*), and M2 simulated (*boxplot*) data for 2010–2039 (period 1). The *box* represents the interquartile range (IQR), *whiskers* are maximum and minimum, and the *horizontal line* is the median for 100 simulations
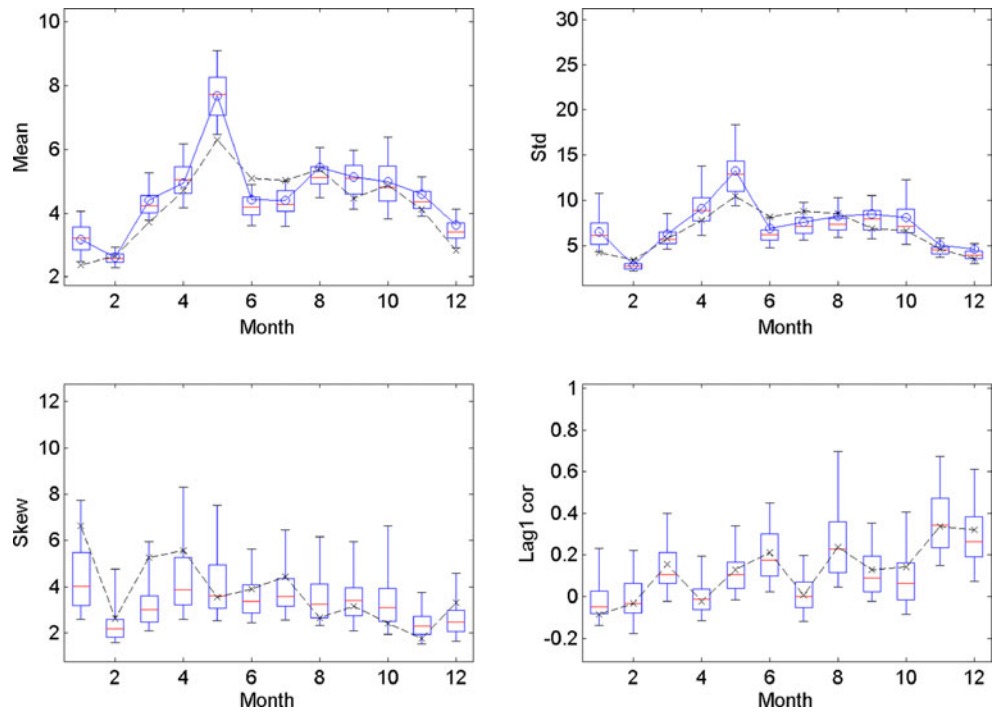

downscaled precipitation data of RCP 8.5 (Brekke et al. 2013) (http://gdo-dcp.ucllnl.org/downscaled_cmip_projections/).

The variation in mean and standard deviation statistics was estimated from the climate scenario as follows:

$$A\theta(\%) = \frac{\phi_{\text{future}} - \theta_{\text{ref}}}{\theta_{\text{ref}}} \times 100 \qquad (28)$$

where θ is the statistic for the future (θ_{future}) and reference (θ_{ref}) period. The reference period is the portion of the observed climate with which climate change information is combined to create a climate scenario; we selected 1950–1999 as our reference period. The future period, 2010–2099, was separated into the following three parts: period 1, from 2010–2039; period 2, from 2040–2069; and period 3, from 2070–2099.

Fig. 10 Key statistics of monthly precipitation for historical (DIA station, *dotted line with cross*), climate change scenario (RCP 8.5, *solid line with circle*), and M2 simulated (*boxplot*) data for 2040–2069 (period 2). The *box* represents the interquartile range (IQR), *whiskers* are maximum and minimum, and the *horizontal line* is the median for 100 simulations
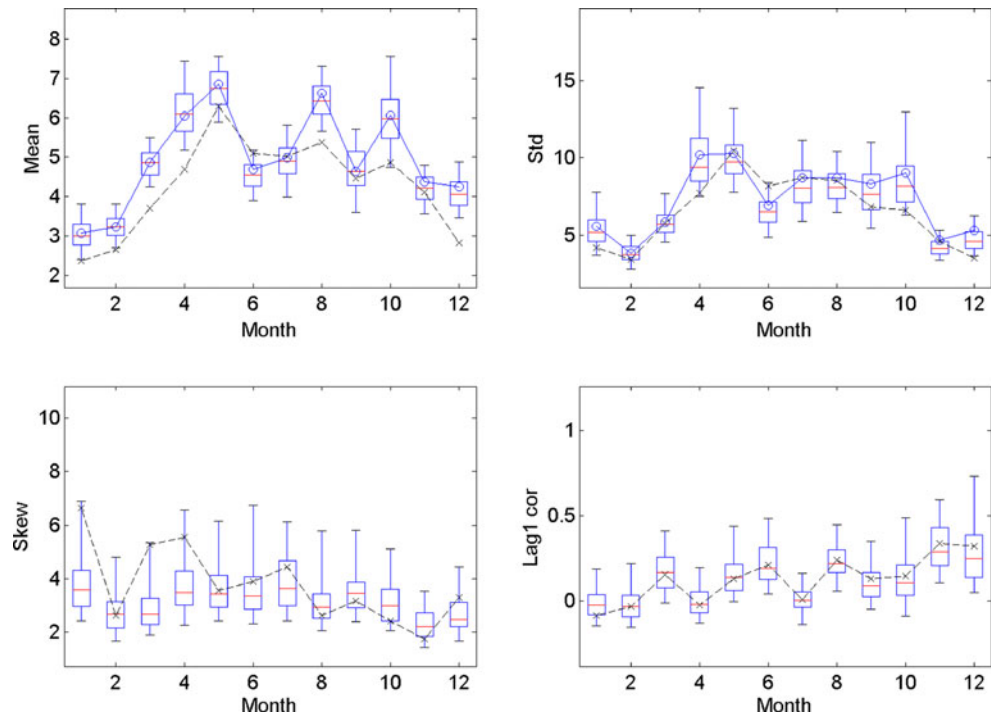
3.3 Application methodology

To validate the performance of our model, we simulated 100 sets of daily precipitation data with the same record length as the historical data. We compared the mean, standard deviation, skewness, lag-1 correlation, and marginal and transition probabilities of the occurrence process for historical and simulated daily precipitation data. We also compared the monthly and yearly aggregated levels of historical and simulated data for validation.
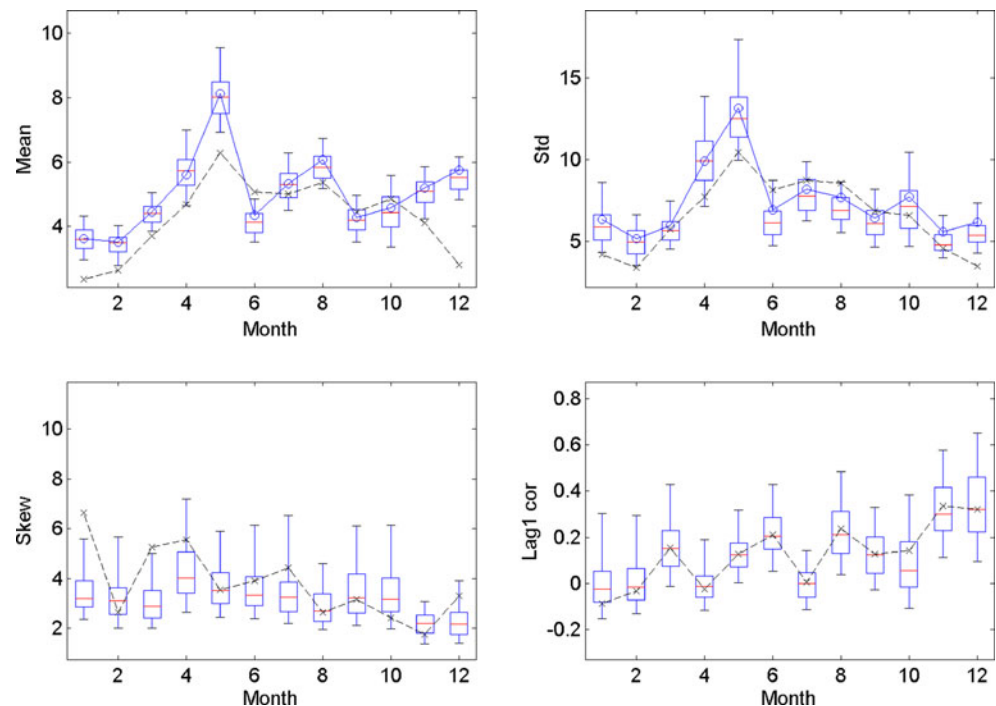
The statistics from the simulated data are presented with a boxplot, in which the box displays the interquartile range (IQR), and the whiskers extend to both extrema, with horizontal lines at the maximum and minimum of the estimated statistics of the 100 generated sequences. The horizontal line inside the box indicates the median of the data. In addition, the value of the statistic corresponding to the historical data is represented with a cross marker.

4 Results

4.1 Validation with observed data

To apply our method, we needed to choose an exponent for Eq. (2). We tested several exponents; "6" most reliably reproduced the key daily statistics of the observed data, including skewness. The power transformation was applied to eliminate the skewed distributional behavior of daily precipitation data; our results were produced with an exponent of $c=6$.

We used two different parameter estimation methods (i.e., a direct method (M1) and an indirect method (M2)). The estimated parameters are presented in Table 1 for M1 and M2. Key statistics for the amount variable (Z) of historical and simulated data are presented in Figs. 2 and 3 for M1 and M2, respectively. Figure 2 shows that the mean, standard deviation, and lag-1 correlation for the simulated data are underestimated for almost all months with M1. In contrast, Fig. 3 shows that with M2, the mean and standard deviation are well reproduced. Skewness is not well preserved with either method. We tested several other exponents to improve skewness modeling unsuccessfully; there may not be a parameterization for skewness that resolves this bias.

Figure 4 shows the transition and limiting probabilities for the occurrence variable; the occurrence process is well explained by the model. Note that using different parameter estimation methods for the amount process does not affect the occurrence process because the occurrence (X) and amount (Z) processes are modeled independently (see Eq. (1)).

As shown in Fig. 5, underestimation of daily statistics from M1 propagates to the monthly time scale. The mean and standard deviation of simulated monthly data from M1 are significantly underestimated in most months. Nonetheless, the mean and standard deviation from M2 are well explained even for aggregate time scale (monthly) precipitation data, as presented in Fig. 6. Note that neither method explains the lag-1 correlation of historical monthly data.

Figure 7 shows the cumulative distribution functions (CDFs) estimated for annual maximum daily precipitation for historical and simulated data. We used the Weibull plotting position formula to estimate corresponding CDFs for annual

Fig. 11 Key statistics of monthly precipitation for historical (DIA station, *dotted line with cross*), climate change scenario (RCP 8.5, *solid line with circle*), and M2 simulated (*boxplot*) data for 2070–2099 (period 3). The *box* represents the interquartile range (IQR), *whiskers* are maximum and minimum, and the *horizontal line* is the median for 100 simulations

maximum daily precipitation. M2 reproduces the historical CDF shape better than M1, specifically the tail, whereas the CDF produced by M1 is shifted to the left. We believe that the significant underestimation of daily mean and standard deviation statistics may have caused this shift. Figure 8 shows the CDF of maximum run days for historical and simulated data, indicating that the simple occurrence model properly explains the extreme statistics of the upper level time scale (i.e., annual).

4.2 Downscaled climate precipitation data simulation

As mentioned, the change of the mean and standard deviation in the future downscaled climate scenario (RCP8.5) was estimated with the reference period of 1950–1999. Table 2 shows the differences in the mean and standard deviation between the reference and future periods. There is an increase in the mean and standard deviation of the future compared to the reference period, except in summer months (months 6 and 7) and month 2. The variations of the mean and standard deviation are similar for all periods. Mean differences are significantly increased during the winter months (months 11, 12, 1, and 2) in later future periods.

Figures 9, 10, and 11 show the historical, future, and simulated (M2) statistics of the mean and standard deviation for periods 1, 2, and 3, respectively. The targeted future statistics are estimated by applying observational statistics to the reference statistics as follows:

$$\theta_{\text{future}} = \Delta\theta \Big/ 100 \times \theta_{\text{obs}} - \theta_{\text{obs}} \tag{29}$$

Note that the traditional parameter estimation method (M1) cannot reproduce the targeted future climate statistics because the parameters in M1 are estimated from the transformed data. When Eq. (29) is applied to the transformed parameters, the back-transformed data cannot be used because they produce inadequate values, such as those too large to be precipitation values (data not shown).

Fig. 12 Mean (*a-1*, *b-1*, and *c-1*) and standard deviation (*a-2*, *b-2*, and *c-2*) of monthly precipitation for historical (DIA station, *dotted line with cross*) and M2 simulated (*boxplot*) data for (*a*) period 1, 2010–2039; (*b*) period 2, 2040–2069; and (*c*) period 3, 2070–2099. The *box* represents the interquartile range (IQR), *whiskers* are maximum and minimum, and the *horizontal line* is the median for 100 simulations
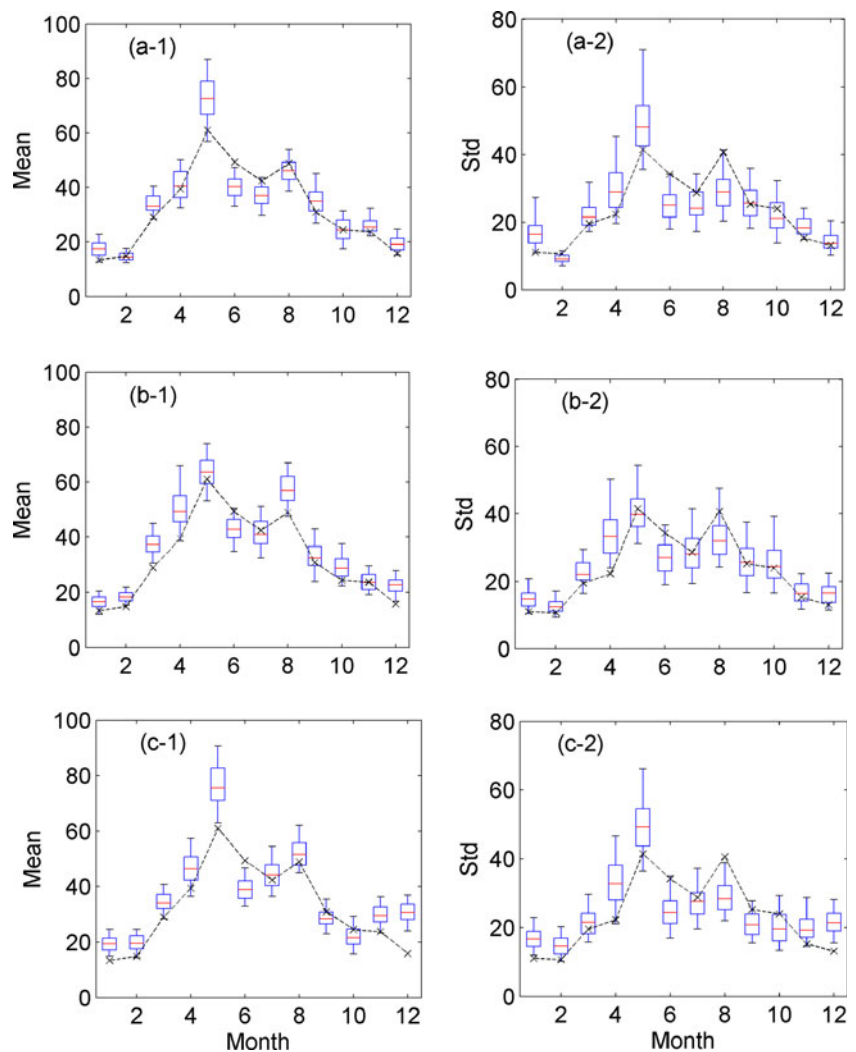
These figures indicate that our parameter estimation method adequately models key statistics (especially the mean and standard deviation) of the climate change scenario. The lag-1 correlation, which should not differ from that of the reference period, is reproduced well in all future periods. Skewness is still not modeled correctly. A significant increase in the mean is seen in winter months, especially for period 3 (see Fig. 11).

Figure 12 shows the mean and standard deviation of monthly precipitation data for future periods. There is a significant increase in the mean and standard deviation, compared to historical statistics, for month 5 of periods 1 and 3; there is no difference for period 2. The increase seen during winter months for daily precipitation (see Table 2) is not noticeable.

5 Summary and conclusions

We compared rainfall simulation models to DIA data. To adequately capture the intermittency of daily rainfall, the daily rainfall (Y process) is modeled separately from the occurrence (X) and amount (Z) processes. A simple Markov Chain is applied to the X process, and we focus mainly on the amount process, Z. We used the AR-1 model for each month of the study, applying a power transformation to normalize data. We compared the traditional parameter estimation method (M1) from transformed data with the indirect parameter estimation method (M2) from the relationship between the original and transformed moments.

We modeled historical precipitation data from DIA, CO, to test the performance of M1 and M2. Data simulated with M1 underestimate the mean, standard deviation, and lag-1 correlation, whereas the M2 simulated data model those statistics fairly well.

Further investigation for the applicability of M1 and M2 to climate change analysis was performed using downscaled RCP 8.5 data. The proposed M2 parameter estimation method models future climate statistics fairly well. M1 cannot be applied in the climate change study because it requires estimating parameters in the transformed domain, but the values produced are highly inadequate.

The M2 method is comparable to M1 in preserving key historical statistics and surpasses M1 in the climate change analysis. Additionally, M2 can also be applied to higher order ARMA models as well as to seasonal ARMA models. Although M2 was applied to daily precipitation data, it can easily be extended to other applications that use ARMA models with the power transformation.

Appendix 1: Example of M2 parameter estimation procedure

As an example, the parameter estimation procedure of M2 for the exponent $c=2$ is presented below. The mean, variance, and lag-1 correlation for the original variable Z can be expressed as:

$$\mu_Z = E(Z_t) = E(N_t^2) = \gamma(2,0) = \mu_N\gamma(1,0) \qquad (A1)$$
$$+ (2-1)\sigma_N^2\gamma(0,0) = \mu_N^2 + \sigma_N^2$$

$$\sigma_Z^2 = EZ^2 - (EZ)^2 = E(N^4) - \left(E(N^2)\right)^2 = \gamma(4,0) - (\gamma(2,0))^2$$
$$= \mu_N\gamma(3,0) + 3\sigma_N^2\gamma(2,0) - \left(\mu_N^2 + \sigma_N^2\right)^2 \qquad (A2)$$

$$\rho_{1,z} = \frac{\gamma(2,2) - (\gamma(2,0))^2}{\gamma(4,0) - (\gamma(2,0))^2} \qquad (A3)$$

$$\gamma(2,2) = \mu_N\gamma(1,2) + 2\sigma_N^2\rho_{1,N}\gamma(1,1) + \gamma(0,2) \qquad (A4)$$

Here, μ_z, σ_Z^2, and ρ_{1z} are known values from the original amount data employing Eqs. (4)–(6), but for the original data before transformation. $\widehat{\mu}_N$ and $\widehat{\sigma}_N^2$ are estimated with two known values of μ_Z and σ_Z^2 by numerically solving Eqs. (A1) and (A2). Note that Eq. (A2) is a function of μ_Z and σ_Z^2.

Furthermore, $\rho_{1,N}$ is numerically solved with Eqs. (A3) and (A4) after estimating $\widehat{\mu}_N$ and $\widehat{\sigma}_N^2$. Additionally, $\widehat{\sigma}_\varepsilon^2$ is estimated from Eq. (7).

References

Aronica GT, Bonaccorso B (2013) Climate change effects on hydropower potential in the Alcantara River Basin in Sicily (Italy). Earth Interact 17:1–22. doi:10.1175/2012EI000508.1

Berne A, Delrieu G, Creutin JD, Obled C (2004) Temporal and spatial resolution of rainfall measurements required for urban hydrology. J Hydrol 299:166–179

Brekke L, Thrasher BL, Maurer EP, Pruitt T (2013) Downscaled CMIP3 and CMIP5 climate and hydrology projections. Bureau of Reclamation, Washington

Burlando P, Rosso R (1991) Extreme storm rainfall and climatic change. Atmos Res 27:169–189

Burlando P, Rosso R (1996) Scaling and multiscaling models of depth-duration-frequency curves for storm precipitation. J Hydrol 187:45–64. doi:10.1016/S0022-1694(96)03086-7

Dennett MD, Rodgers JA, Keatinge JDH (1983) Simulation of a rainfall record for the site of a new agricultural development: an example from Northern Syria. Agric Meteorol 29:247–258. doi:10.1016/0002-1571(83)90086-9

Eagleson PS (1978) Climate, soil and vegetation: I. Introduction to water balance dynamics. Water Resour Res 14:705–712

Gabriel KR, Neumann J (1962) A Markov chain model for daily rainfall occurrence at Tel Aviv. Q J R Meteorol Soc 88:90–95. doi:10.1002/qj.49708837511

Guttorp P, Minin VN (1995) Stochastic Modeling of Scientific Data. Taylor & Francis, Boca Ratan, p 384

Hannachi A (2012) Intermittency, autoregression and censoring: a first-order AR model for daily precipitation. Meteorol Appl. doi:10.1002/met.1353

Hingray B, Ben Haha M (2005) Statistical performances of various deterministic and stochastic models for rainfall series disaggregation. Atmos Res 77:152–175

Hughes JP, Guttorp P, Charles SP (1999) A non-homogeneous hidden Markov model for precipitation occurrence. J R Stat Soc Ser C Appl Stat 48:15–30

Katz RW (1996) Use of conditional stochastic models to generate climate change scenarios. Clim Chang 32:237–255

Katz RW (1999) Moments of power transformed time series. Environmetrics 10:301–307

Katz RW, Parlange MB (1993) Effects of an index of atmospheric circulation on stochastic properties of precipitation. Water Resour Res 29:2335–2344

Katz RW, Parlange MB (1995) Generalizations of chain-dependent processes: application to hourly precipitation. Water Resour Res 31:1331–1341

Katz RW, Zheng XG (1999) Mixture model for overdispersion of precipitation. J Clim 12:2528–2537

Koutsoyiannis D, Onof C (2001) Rainfall disaggregation using adjusting procedures on a Poisson cluster model. J Hydrol 246:109–122

Lall U, Rajagopalan B, Tarboton DG (1996) A nonparametric wet/dry spell model for resampling daily precipitation. Water Resour Res 32:2803–2823

Lebel T, Braud I, Creutin JD (1998) A space-time rainfall disaggregation model adapted to Sahelian mesoscale convective complexes. Water Resour Res 34:1711–1726

Lee T, Ouarda TBMJ, Jeong C (2012) Nonparametric multivariate weather generator and an extreme value theory for bandwidth selection. J Hydrol 452–453:161–171. doi:10.1016/j.jhydrol.2012.05.047

Marani M (2003) On the correlation structure of continuous and discrete point rainfall. Water Resour Res 39:SWC21–SWC28

Maurer EP, Hidalgo HG, Das T, Dettinger MD, Cayan DR (2010) The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in California. Hydrol Earth Syst Sci 14:1125–1138. doi:10.5194/hess-14-1125-2010

Mezghani A, Hingray B (2009) A combined downscaling-disaggregation weather generator for stochastic generation of multisite hourly weather variables over complex terrain: Development and multiscale validation for the Upper Rhone River Basin. J Hydrol 377:245–260. doi:10.1016/j.jhydrol.2009.08.033

Richardson CW (1981) Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resour Res 17:182–190

Roldan J, Woolhiser DA (1982) Stochastic daily precipitation models: 1. A comparison of occurrence processes. Water Resour Res 18:1451–1459

Salas JD (1993) In: Maidment DR (ed) Handbook of Hydrology. McGraw-Hill, New York 19.1-19.72.

Todorovic P, Woolhiser DA (1975) Stochastic model of n-day precipitation. J Appl Meteorol 14:17–24

Wilby RL (1994) Stochastic weather type simulation for regional climate change impact assessment. Water Resour Res 30:3395–3402

Wilks DS (1999) Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. Agric For Meteorol 93:153–169

Woolhiser DA, Roldan J (1982) Stochastic daily precipitation models: 2. A comparison of distributions of amounts. Water Resour Res 18:1461–1468

Yoon S, Jeong C, Lee T (2013) Application of harmony search to design storm estimation from probability distribution models. J Appl Math 2013:11