Simulation of Spatial Dependence in Daily Precipitation Using a Mixture of Generalized Chain-Dependent Processes at Multisites

XIAOGU ZHENG

College of Global Change and Earth System Science, Beijing Normal University, Beijing, China

CRAIG S. THOMPSON

National Institute of Water and Atmospheric Research, Wellington, New Zealand

ABSTRACT

Recently, a single-site stochastic precipitation model called "the mixture of generalized chain-dependent processes conditioned on a climate variable" was developed. The model can effectively eliminate overdispersion—that is, underestimation in variance of seasonal precipitation total. In this paper, the single-site model is further developed into a multisite stochastic precipitation model by driving a collection of individual single-site models, but with spatial dependence following a method proposed by D. S. Wilks. Specifically, a computationally effective algorithm for estimating the spatial dependence of precipitation occurrence is developed to replace the construction of the empirical curves in the Wilks method. An effective and straightforward approach for correcting the bias of the spatial correlation of precipitation intensity is also proposed. This model is tested on a small network of sites from a significant hydroelectric power generation region of South Island, New Zealand.

1. Introduction

In 2003, the National Institute of Water and Atmospheric Research (NIWA) of New Zealand launched a project called "Climate-Related Risks for Energy Supply and Demand" (Zheng and Thompson 2007). In this project, daily precipitation at four sites in the upper Waitaki catchment, New Zealand, was simulated under several climate scenarios. The outflows from the catchment were simulated using a hydrological catchment model forced by the simulated precipitation. Since the outflows dominate the electricity generation capacity of the hydropower stations near the catchment, the climate-related risks for the hydroenergy supply could be accessed.

An outcome of this project was the development of a single-site stochastic precipitation model called "the mixture of generalized chain-dependent processes conditioned on a climate variable" (Zheng and Katz 2008a; Zheng and Thompson 2007). Unlike other existing stochastic precipitation models, this model can effectively

eliminate overdispersion (i.e., underestimation of the variance in seasonal precipitation totals; Katz and Zheng 1999). However, the capacity of hydroelectric power is dominated by the seasonal precipitation total at the catchment scale. Correctly simulating the distribution of the seasonal precipitation total is most important in this project. In this paper, we will demonstrate that, without introducing spatial dependence of daily precipitation, the overdispersion at the catchment scale still exists, despite the fact that overdispersion for every single site is eliminated. In this case, the risk for energy supply related to wet or dry seasons could still be underestimated.

The literature on multisite precipitation models is rather limited (e.g., Wilks 1998; Qian et al. 2002; Thompson et al. 2007; Brissette et al. 2007), with the majority of the work based on the method proposed by Wilks (1998). Wilks' approach demands significant effort to construct the empirical curves corresponding to all possible station pairs. In the present paper, we propose an alternative algorithm for the estimation of the spatial dependence of precipitation occurrence and intensity. This algorithm is faster and easier to implement.

The paper is arranged as follows. First the proposed estimation method for the spatial dependence of daily

Corresponding author address: Xiaogu Zheng, College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China.
E-mail: x.zheng@bnu.edu.cn

precipitation is introduced. Then the results of a simulation study of long-term daily precipitation in the upper Waitaki catchment, New Zealand, using the proposed multisite precipitation model are described. Finally, discussions and conclusions are presented.

2. Methodology

In this section, we briefly introduce a multisite daily precipitation model called the mixture of generalized chain-dependent processes conditional on a climate variable at multisites.

a. Generalized chain-dependent processes conditioned on a climate variable

A "generalized chain-dependent processes conditioned on a climate variable" is a stochastic model for single-site daily precipitation. The model can be summarized as follows [see Zheng and Katz (2008a) for the more detail]:

Let X_y be a climate variable in year y; for example, a seasonal Interdecadal Pacific Oscillation (IPO) index whose state remains constant in a year y over the time period of T days (i.e., a season). Let $J_{yt}, t = 1, \ldots, T$ denote the sequence of indicators of daily precipitation occurrence in year y (i.e., $J_{yt} = 1$ indicates a "wet day" and $J_{yt} = 0$ a "dry day"). It is assumed that this process is a first-order Markov chain whose transition probability is related to X_y by

$$p_{y,j} \equiv \Pr(J_{yt} = 1 | J_{y(t-1)} = j, X_y)$$
$$= \begin{cases} 1 - 1/[1 + \exp(\alpha_0 + \alpha_1 X_y)] & \text{if} \quad j = 0 \\ 1 - 1/[1 + \exp(\beta_0 + \beta_1 X_y)] & j = 1 \end{cases}, \quad (1)$$

where "Pr" is being used as a generic "probability function" (i.e., both for discrete probabilities and for density functions) (Chandler and Wheater 2002).

Let $R = (R_{yt}, t = 1, \ldots, T)$ denote the time series of daily precipitation amounts in year y. The "intensity" $R_{yt} > 0$ (i.e., days for which $J_{yt} = 1$) is taken to be conditionally independent and identically distributed. It is further assumed that daily precipitation intensity has a power-transform distribution. That is,

$$\Pr(R_{yt}^q = x | J_{yt} = 1) = \phi(x; \mu, \sigma), \quad (2)$$

where $\phi(x; \mu, \sigma)$ denotes the Gaussian density function with mean μ and standard deviation σ. The values of $q = \frac{1}{2}, \frac{1}{3}$, or $\frac{1}{4}$ are commonly employed to account for the high degree of positive skewness in the distribution of daily precipitation amounts (Katz and Zheng 1999).

TABLE 1. Summary of parameters of the mixture model.

Parameter	For	Reference
$w, {}_i m_i s,$	$w_y \equiv \Pr(I_y = 1 | X_y)$	Eq. (4)
${}_i\alpha_0, {}_i\alpha_1,$	${}_i p_{y,j} \equiv \Pr(J_{yt} = 1 | J_{y(t-1)} = j, I_y = i, X_y)$	Eq. (5)
${}_i\beta_0, {}_i\beta_1$		
${}_i\mu, {}_i\sigma, q$	Rainfall intensity conditioned on $I_y = i$	

b. Mixture of generalized chain-dependent processes conditioned on a climate variable

Let $I_y(I_y = i, i = 0, 1)$ denote a two-state index—usually hidden, unobserved, or only partially observed—that remains constant over the time period of T days. The annual time series of index states is assumed to be identically distributed with respect to year y, with a common distribution

$$w = \Pr(I_y = 1). \quad (3)$$

For a given year, the time series of daily precipitation amounts $(R_{yt}, t = 1, \ldots, T)$ is called a mixture of generalized chain-dependent processes conditional on X_y; if for given $I_y = i$, it is a generalized chain-dependent process conditioned on X_y with parameters $w, {}_i\alpha_0, {}_i\alpha_1, {}_i\beta_0, {}_i\beta_1, {}_i\mu, {}_i\sigma$, and q (the left subscript is for the state $I_y = i$). Specifically, similar to Eqs. (1) and (2),

$$_i p_{y,j} \equiv \Pr(J_{yt} = 1 | J_{y(t-1)} = j, I_y = i, X_y)$$
$$= \begin{cases} 1 - 1/[1 + \exp({}_i\alpha_0 + {}_i\alpha_1 X_y)] & \text{if} \quad j = 0 \\ 1 - 1/[1 + \exp({}_i\beta_0 + {}_i\beta_1 X_y)] & j = 1 \end{cases}, \quad \text{and}$$
$$(4)$$

$$\Pr(R_{yt}^q = x | J_{yt} = 1, I_y = i) = \phi(x; {}_i\mu, {}_i\sigma). \quad (5)$$

The climate variable X_y is assumed to be a mixture of two Gaussian random variables with mean ${}_i m$ that depends on $I_y = i$ and a common standard deviation s. Specifically,

$$\Pr(X_y | I_y = i) = \phi(X_y; {}_i m, s) \quad (6)$$

(Zheng and Katz 2008a). Under this assumption, parameters listed in Table 1 can be estimated using the expectation maximization (EM) algorithm (McLachlan and Krishnan 1997; Zheng and Katz 2008a). Moreover, for simulating daily precipitation, we need to know the probability of I_y conditioned on X_y. This can be derived by Bayes' Theorem and Eqs. (3), (5), and (6),

$$w_y \equiv \Pr(I_y = 1 | X_y) = \frac{\Pr(X_y | I_y = 1)\Pr(I_y = 1)}{\Pr(X_y | I_y = 0)\Pr(I_y = 0) + \Pr(X_y | I_y = 1)\Pr(I_y = 1)}$$
$$= [\phi(X_y; {}_1 m, s)w] / [\phi(X_y; {}_0 m, s)(1 - w) + \phi(X_y; {}_1 m, s)w]. \quad (7)$$

Given a parameter set listed in Table 1 and an initial occurrence state J_{y0}, we can generate by a Monte Carlo simulation a daily precipitation time series from a mixture of chain-dependent processes conditioned on X_y. The procedure is as follows (Zheng and Katz 2008a):

(i) Generate statistically independent uniform random variables [over the range (0, 1)] U_y, and standard Gaussian random variables ($W_{yt}, t = 1, \ldots, T$) and ($Z_{yt}; t = 1, \ldots, T$).

(ii) Simulate the state index

$$I_y = \begin{cases} 1 & \text{if} \quad U_y \le w_y \\ 0 & \quad\quad U_y > w_y \end{cases}. \quad (8)$$

(iii) Simulate the precipitation occurrences

$$J_{y,t} = \begin{cases} 1 & \text{if} \quad \Phi(W_{yt}) \le {}_{I_y}p_{y,J_{y,t-1}} \\ 0 & \quad\quad \Phi(W_{yt}) > {}_{I_y}p_{y,J_{y,t-1}} \end{cases}, \quad (9)$$

where Φ is the standard Gaussian probability distribution function, so $\Phi(W_{yt})$ is uniformly distributed.

(iv) Transform the standard Gaussian variable Z_{yt} into the general Gaussian variable with mean ${}_{I_y}\mu$ and standard deviation ${}_{I_y}\sigma$ to get the power-transformed daily precipitation amount

$$R_{yt}^q = J_{yt}({}_{I_y}\sigma \times Z_{yt} + {}_{I_y}\mu). \quad (10)$$

If R_{yt}^q is negative, truncate it to zero.

c. Mixture of generalized chain-dependent processes at multisites

The mixture of generalized chain-dependent processes conditioned on X_y at multisites $[R_{y1}(m), \ldots, R_{yT}(m)]$ can be simply represented by extending notions of mixture of generalized chain-dependent processes to a higher dimension to have additional parameters for every site (listed in Table 1) and to add spatial correlations to the two Gaussian fields. Here $[R_{y1}(m), \ldots, R_{yT}(m)]$ can be simulated from a uniform random variable U_y that is common for all sites, and temporally independent but spatially correlated standard Gaussian fields $[W_{yt}(m), t = 1, \ldots, T]$ and $[Z_{yt}(m); t = 1, \ldots, T]$, following the steps (ii)–(iv) documented in section 2b.

To estimate the spatial correlation between $W_{yt}(m)$ and $W_{yt}(n)$ [denoted by $\omega(m, n)$], we represent the modeled correlation between $J_{yt}(m)$ and $J_{yt}(n)$ [denoted by $C_y^J(m, n)$] as a function of $\omega(m, n)$, then we adjust $\omega(m, n)$ to make $C_y^J(m, n)$ close to the observed correlation, following the idea of Wilks (1998). However, Wilks' approach demands significant effort to simulate the relation between $C_y^J(m, n)$ and $\omega(m, n)$. Here, we propose an alternative algorithm with much less computational cost.

By conditional probability formula, the modeled correlation can be expressed as

$$\sum_{i,i'=0}^{1} \frac{[{}_{ii'}\pi_y(m,n) - {}_i\pi_y(m){}_{i'}\pi_y(n)]\Pr[I_y(m) = i, I_y(n) = i'|X_y]}{\sqrt{{}_i\pi_y(m)[1 - {}_i\pi_y(m)]{}_{i'}\pi_y(n)[1 - {}_{i'}\pi_y(n)]}}, \quad (11)$$

where by Eq. (8) and the fact that X_y is common for all stations, we have

$$\Pr[I_y(m) = 1, I_y(n) = 1|X_y]$$
$$= \Pr[U_y \le w_y(m), U_y \le w_y(n)]$$
$$= \min[w_y(m), w_y(n)], \quad (12)$$

$$\Pr[I_y(m) = 1, I_y(n) = 0|X_y]$$
$$= \Pr[U_y \le w_y(m), U_y > w_y(n)]$$
$$= \max[w_y(m) - w_y(n), 0], \quad (13)$$

$$\Pr[I_y(m) = 0, I_y(n) = 1|X_y]$$
$$= \Pr[U_y > w_y(m), U_y \le w_y(n)]$$
$$= \max[w_y(n) - w_y(m), 0], \quad (14)$$

$$\Pr[I_y(m) = 0, I_y(n) = 0|X_y]$$
$$= \Pr[U_y > w_y(m), U_y > w_y(n)]$$
$$= \min[1 - w_y(m), 1 - w_y(n)], \quad \text{and} \quad (15)$$

$${}_i\pi_y(m) \equiv {}_ip_{y,0}(m)/[1 - {}_ip_{y,1}(m) + {}_ip_{y,0}(m)] \quad (16)$$

is the invariant measure of the Markov chain $(J_{yt}(m), t = 1, \ldots, T)$ conditioned on $I_y(m) = i$ (Feller 1971), and ${}_{i,i'}\pi_y(m, n)$ is the invariance measure of the bivariate Markov chain $\{[J_{yt}(m), J_{yt}(n)], t = 1, \ldots, T\}$ conditioned on $I_y(m) = i$ and $I_y(j) = j$. The quantity ${}_{i,i'}\pi_y(m, n)$ can be uniquely determined by the parameter $\omega(m, n)$ [a computationally effective algorithm for deriving ${}_{ii'}\pi_y(m, n)$ given $\omega(m, n)$ is documented in the appendix]. Therefore, $\omega(m, n)$ can be estimated by minimizing the mean square error (over all years)

between modeled correlations [Eq. (11)] and observed correlations.

The spatial correlation $\zeta(m, n)$ is initially estimated as

$$\hat{\zeta}(m, n) = \frac{1}{Y} \sum_{y=1}^{Y} C_y^r(m, n), \qquad (17)$$

where $C_y^r(m, n)$ is the spatial correlation of observed power-transformed precipitation intensity between the site pair m and n in year y. However, because of the power transformation, it is likely that the spatial correlations of simulated precipitation intensity using $\hat{\zeta}(m, n)$ are biased (Zheng and Katz 2008b). To correct the bias, the $\hat{\zeta}(m, n)$ estimated by Eq. (14) is adjusted by applying a factor that is the ratio of the observed spatial correlation of precipitation intensity to the simulated spatial correlation of precipitation intensity using $\hat{\zeta}(m, n)$ estimated by Eq. (17).

3. Data and results

The upper Waitaki catchment (Fig. 1) is the one of the most important hydroelectric power generation catchments in New Zealand. It is crucial for New Zealand to understand the variability of the outflow from the three lakes (Lake Pukaki, Lake Tekapo, and Lake Ohau; see Fig. 1) in the upper Waitaki catchment and to estimate this variability for the next two to three decades. To this end, a climate variable is needed that is both predictable and significantly associated with rainfall on a decadal time scale. Fortunately, the IPO may be such a climate variable. The IPO has significant impacts on rainfall and river flows in certain regions of New Zealand. In the west and south of South Island, the negative IPO phase is generally associated with lower rainfall and inflows, and vice versa for the positive IPO phase (Salinger et al. 2001; Zheng and Thompson 2007). An aspect of this NIWA project is to assess how precipitation in the upper Waitaki catchment, on interannual time scales, is linked to the IPO, and in particular the correlation between the IPO and the seasonal rainfall total in the upper Waitaki catchment.

There are only four precipitation stations near the catchment—Mt. Cook, Lake Ohau, Lake Tekapo, and Franz Josef (see Fig. 1 for the locations)—with long record lengths. A common period to all four sites was 1953–2000. A wet day, in the context of this study, occurs when at least 1 mm of precipitation was recorded by the rain gauge; otherwise, the day is treated as dry. Only the summer season (December–February, with the seasonal length $T = 90$) is analyzed in this study, because generally there is sufficient precipitation for other seasons in the upper Waitaki catchment. Power-transform values for q of ⅓, ⅖, ½, and 4/7 are adopted for Lake Tekapo, Lake Ohau,

Mt. Cook, and Franz Josef, respectively. These values of q were empirically chosen to make the simulated seasonal mean totals similar to the observed (Zheng and Thompson 2007; Zheng and Katz 2008a). This is because the seasonal variability of the precipitation is related to both daily and decadal variability, so correctly simulating the seasonal mean precipitation at each site is most essential. Zheng and Thompson (2007) showed that the selected q values can simulate reasonable daily, seasonal, annual, and decadal variability of the precipitation in the upper Waitaki catchment.

The IPO index is provided by the Hadley Centre from the Met Office. It is derived from the 3rd empirical orthogonal function (EOF) pattern of 13-year low-pass-filtered global SST (see Fig. 2 in Zheng and Thompson 2007). There have been three major phases during the twentieth century: positive phases from 1920 to 1945 and 1978–99, and a negative phase between 1946 and 1977. The IPO appears to modulate the frequency and intensity of the El Niño–Southern Oscillation (ENSO) phenomenon: during the last positive phase of the IPO, there were more El Niño episodes and fewer La Niña ones than during the preceding negative IPO phase (e.g., Salinger et al. 2001). The climate variable X_y is the observed mean IPO index over the austral summer season [December–February (DJF)].

As described in section 2, parameters for each site are estimated by fitting the single-site mixture of generalized chain-dependent processes conditioned on the IPO index to each site's data. All the estimated parameters are listed in Table 2. The methodology of section 2c is used to estimate the spatial dependence of precipitation occurrence and precipitation intensity. These estimates are listed in Table 3. For each site pair m and n, $\hat{\omega}(m, n)$ can be estimated within 10 s by a standard PC with a 3.2-GHz P4 processor and 2 Gb of RAM. The computational cost for estimation of $\hat{\zeta}(m, n)$ is negligible.

Based on the fitted parameters and the observed seasonal IPO index, 100 independent simulations of the DJF daily precipitation over a 47-year period are generated using Eqs. (8)–(10).

The observed and simulated correlations between daily precipitation occurrences at all site pairs are shown in Table 4. It shows that the correlations simulated are virtually identical to the correlations observed. This indicates that for this dataset, the proposed estimation procedure works well.

The observed and simulated correlations between daily precipitation intensity at all site pairs are shown in Table 5. It is clear that, without the factor adjustment for $\hat{\zeta}(m, n)$ in Eq. (17), the simulated correlations are all underestimated. However, after $\hat{\zeta}(m,n)$ is bias corrected, the remaining bias in the simulated correlations is negligible.
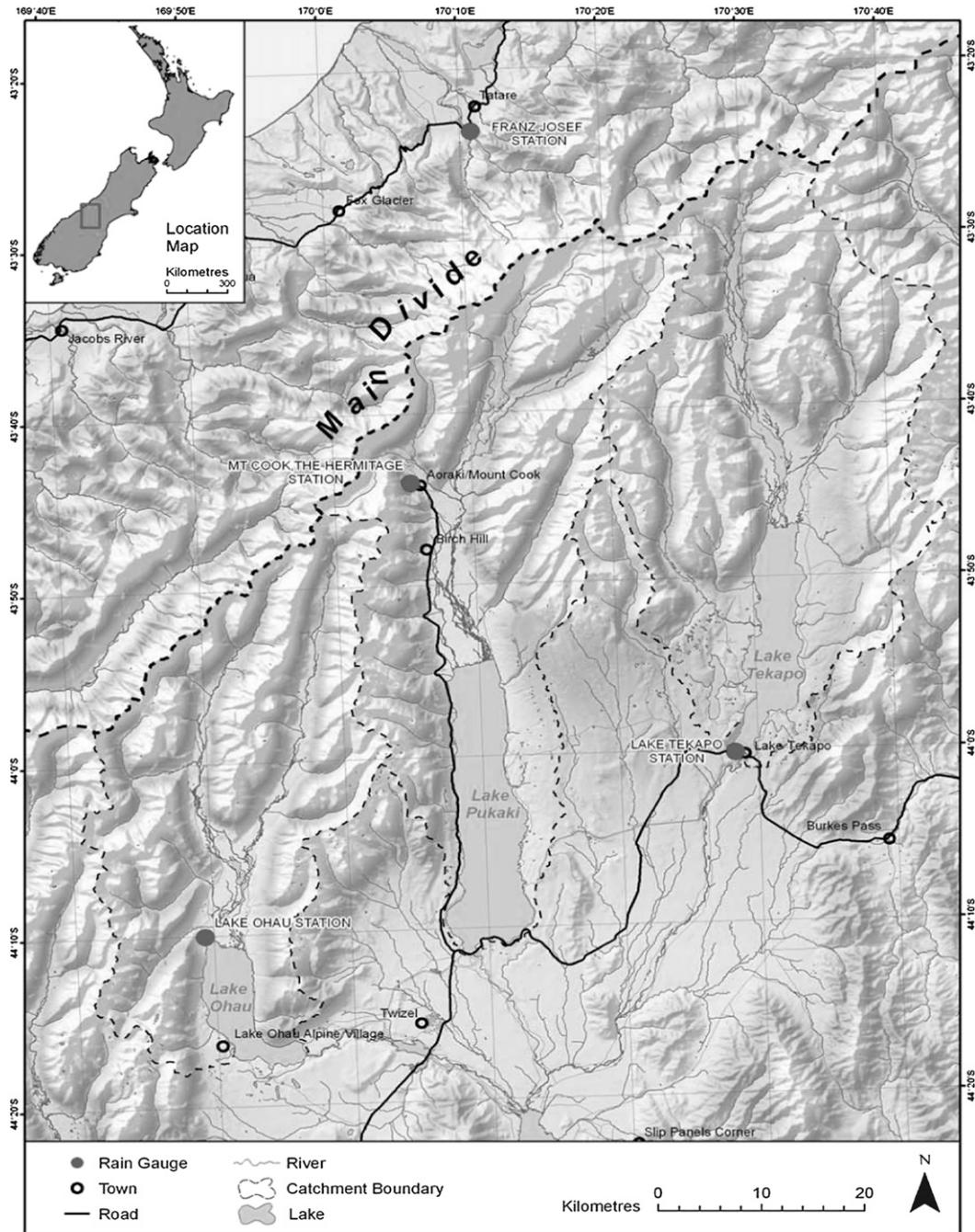
FIG. 1. The geographical features of the upper Waitaki catchment region of central South Island. Locations of rainfall sites used in this study are also shown.

A major advantage of the single-site mixture of generalized chain-dependent processes conditioned on a climate variable over single-site chain-dependent processes is in the better modeling of the interannual and decadal variability (Zheng and Katz 2008a). Zheng and Katz (2008a) showed that the variance of seasonal mean precipitation and the correlation between the seasonal mean precipitation and the climate variable can be well captured by such a mixture model. In this study, we further investigate simulated seasonal regional precipitation totals because they are closely related to the capacity of hydroelectric power generation in each year and are therefore crucial in estimating the climate-related risk of energy supply.

TABLE 2. Estimated parameters of the mixture model, according to the hidden index state I ($=0, 1$). See Table 1 for the summary of the parameters.

	Mt. Cook		Lake Ohau		Lake Tekapo		Franz Josph	
	$i = 0$	$i = 1$	$i = 0$	$i = 1$	$i = 0$	$i = 1$	$i = 0$	$i = 1$
$\Pr(I_y = i)$	0.41	0.59	0.3	0.7	0.75	0.25	0.73	0.27
$_i m$		0.28	−0.13	0.27	−0.06	0.17	−0.33	0.35	−0.37
$_i s$	1.14	1.14	1.14	1.14	1.16	1.16	1.14	1.14
$_i \alpha_0$	−0.86	−1.01	−1.5	−1.46	−1.68	−1.93	−0.75	−0.64
$_i \alpha_1$	0.1	0.07	0.04	0.02	0.05	0.02	0.06	0.23
$_i \beta_0$	0.4	0.26	−0.58	−0.69	−0.54	−1.00	0.69	0.64
$_i \beta_1$	0.18	0.03	0.05	−0.13	−0.01	0.02	0.06	0.22
$_i \mu$	5.04	4.03	2.85	2.39	1.81	1.6	6.51	5.06
$_i \sigma$	3.64	2.67	1.11	0.88	0.64	0.43	4.88	3.59

The qqplot of the seasonal regional precipitation totals simulated by the mixture of the generalized chain-dependent processes at multisites against those observed is shown in the left panel of Fig. 2. It shows that the distributions of the simulated and observed seasonal precipitation total are similar since they lie on or very close to the one-to-one line in the figure.

4. Discussion and conclusions

In this paper, the mixture of generalized chain-dependent processes conditioned on a climate variable at multisites has been proposed for simulating daily precipitations in order to capture from daily-to-decadal precipitation data, variability at the catchment scale. The proposed model is similar in concept to the Wilks (1998) precipitation model. But in the estimation of spatial dependence of the daily precipitation occurrence and intensity, the construction of the empirical curves used in the Wilks method have been replaced by a computationally effective optimization algorithm.

In this study, the distribution of transformed precipitation intensity is assumed to be power-transformed Gaussian. An advantage of the power-transformed Gaussian assumption is that when the mixture of statistical processes is

TABLE 4. Observed and simulated correlations of daily precipitation occurrence from sites in/near the upper Waitaki catchment, South Island, New Zealand.

Site pairs m and n	Observed	Simulated
Mt. Cook–Lake Ohau	0.43	0.44
Mt. Cook–Lake Tekapo	0.48	0.47
Lake Ohau–Lake Tekapo	0.45	0.46
Mt. Cook–Franz Josef	0.65	0.64
Lake Ohau–Franz Josef	0.34	0.35
Lake Tekapo–Franz Josef	0.38	0.37

introduced, the overdispersion (i.e., underestimation of the variance of seasonal precipitation total) can be effectively eliminated. However, this is not necessarily the case for other distributions (Zheng and Katz 2008a), such as the gamma or mixed exponential. In the present paper, we have shown that the spatial dependence of precipitation intensity can be effectively modeled under the power-transformed Gaussian assumption. Moreover, the estimation of the spatial correlation of precipitation intensity is very simple and the bias arising from the power transformation can be easily corrected.

To investigate why the distribution of seasonal regional precipitation total is well simulated by the proposed model, we further consider the spatially independent mixture of generalized chain-dependent processes [i.e., $\omega(m, n) = \xi(m, n) = 0$]. The qqplots of the alternative model is shown in the right panel of Fig. 2. The model tends to overestimate the smaller precipitation totals but underestimate the larger precipitation totals. This suggests that if the spatial correlation is not introduced, it would underestimate the risks to hydroelectric supply associated with wet and dry seasons.

A reason for the spatially independent mixture of generalized chain-dependent processes underestimating the risks associated with wet and dry seasons may be as follows. Since the four sites are close, we expect (Tables 4 and 5) that the spatial dependence of the precipitation between them is positive and significant. If one site has a low precipitation total in any given summer season,

TABLE 3. Pairwise estimation of spatial correlations of the two Gaussian fields—that is, $\omega(m, n)$ and $\zeta(m, n)$ from the site in/near to the upper Waitaki catchment, South Island, New Zealand.

Site pairs m and n	$\hat{\omega}(m, n)$	Initial $\hat{\zeta}(m, n)$	Adjusted $\hat{\zeta}(m, n)$
Mt. Cook–Lake Ohau	0.699	0.507	0.561
Mt. Cook–Lake Tekapo	0.801	0.299	0.320
Lake Ohau–Lake Tekapo	0.724	0.425	0.505
Mt. Cook–Franz Josef	0.889	0.740	0.792
Lake Ohau–Franz Josef	0.626	0.393	0.435
Lake Tekapo–Franz Josef	0.702	0.229	0.265

TABLE 5. As for Table 4, but for daily precipitation intensity.

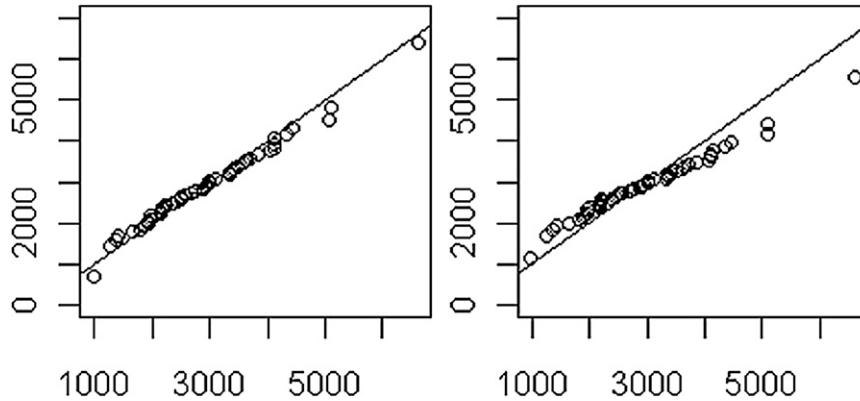| Site pairs m and n	Observed	Initially simulated	Simulated after bias correction
Mt. Cook–Lake Ohau	0.53	0.48	0.53
Mt. Cook–Lake Tekapo	0.28	0.26	0.29
Lake Ohau–Lake Tekapo	0.43	0.38	0.43
Mt. Cook–Franz Josef	0.73	0.68	0.74
Lake Ohau–Franz Josef	0.39	0.35	0.39
Lake Tekapo–Franz Josef	0.25	0.22	0.26
Average	0.434	0.39	0.443

FIG. 2. The qqplot of the simulated DJF regional precipitation total (horizontal axis) against that observed (vertical axis). (left) The multisite mixture of generalized chain-dependent processes conditioned on IPO proposed in this paper. (right) The spatially independent mixture of generalized chain-dependent processes. The solid line shows the equivalence between two distributions of observed and simulated precipitations (mm).

then there is a high likelihood, through its spatial dependency, for the other sites to also have low precipitation totals. However, as the model has no spatial dependencies, the probability of the simulated precipitation at all four sites to have above- or below-average precipitation in any single season is actually smaller. Therefore, the likelihood of large or small regional totals in any single season is also smaller.

In conclusion, the stochastic model proposed in the study has simulated the spatial correlations of daily precipitation occurrence and intensity quite well, and has also correctly simulated the distribution of the total. This is particularly vital for regional water budget, and for any risk analysis of hydroelectric generation potential during wet and dry years.

APPENDIX

Why $_{ii'}\pi_y(m, n)$ Is Uniquely Determined by $\omega(m, n)$

The probability of $[J_{yt}(m) = j, J_{yt}(n) = j']$ conditioned on $[J_{y(t-1)}(m) = k, J_{y(t-1)}(n) = k'; I_y(m) = i, I_y(n) = i'; X_y]$ is denoted by $_{ii'}P_{kk'jj'}(m, n)$ and the corresponding

transition probability matrix is denoted by $\mathbf{P}$. Given $[I_y(m) = i, I_y(n) = i'; X_y]$, $[J_{yt}(m), J_{yt}(n)], t = 1, \ldots, T]$ is a bivariate Markov chain. By Eq. (8), its transition probabilities are

$$
_{ii'}P_{kk'11}(m, n) \\
= \Pr\{\Phi[W_{yt}(m)] \leq {}_iP_{k1}(m); \Phi[W_{yt}(n)] \leq {}_{i'}P_{k'1}(n)\}, \tag{A1}
$$

$$
_{ii'}P_{kk'10}(m, n) = {}_iP_{k1}(m) - {}_{i,i'}P_{kk'11}(m, n), \tag{A2}
$$

$$
_{ii'}P_{kk'01}(m, n) = {}_{i'}P_{k'1}(n) - {}_{ii'}P_{kk'11}(m, n), \quad \text{and} \tag{A3}
$$

$$
_{ii'}P_{kk'00}(m, n) = 1 - {}_iP_{k1}(m) - {}_{i'}P_{k'1}(n) + {}_{ii'}P_{kk'11}(m, n), \tag{A4}
$$

where $_iP_{kj}(m)$ denotes the probability of the event $[J_{yt}(m) = j]$ conditioned on $[J_{y(t-1)}(m) = k; I_y(m) = i; X_y]$. The function for calculating multivariate Gaussian probability distribution [i.e., Φ in Eq. (A1)] is available, for example, the function ''dmvnorm'' in the library ''mvtnorm'' of the statistical package ''R'' (http://cran.r-project.org). So $_{ii'}P_{kk'll}(m, n)$ can be quickly computed.

By the ergodic theory of Markov chains (e.g., Feller 1971), the invariant probability measure of the transition probability matrix $\mathbf{P}$ is the last row of $\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}$; where the partitioned matrix $\mathbf{A} = [\mathbf{I} - \mathbf{P}, \mathbf{1}]$, $\mathbf{I}$ is the identity matrix and all elements of column vector $\mathbf{1}$ are 1. Since $\mathbf{P}$ is uniquely determined by $\omega(m, n)$, so it is an invariant measure. Finally, since $_{ii'}\pi(m, n)$ is the value of the binary invariant at the binary state $J_{yt}(m) = 1, J_{yt}(n) = 1$, it is also uniquely determined by $\omega(m, n)$.

REFERENCES

Brissette, F. P., M. Khalili, and R. Leconte, 2007: Efficient stochastic generation of multi-site synthetic precipitation data. *J. Hydrol.,* **345,** 121–133.

Chandler, R. E., and H. S. Wheater, 2002: Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland. *Water Resour. Res.,* **38,** 1992, doi:10.1029/2001WR000906.

Feller, W., 1971: *An Introduction to Probability Theory and its Applications*. Vol. 1. 3rd ed. Wiley, 509 pp.

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

McLachlan, G. J., and T. Krishnan, 1997: *The EM Algorithm and Extensions*. Wiley, 274 pp.

Qian, B., J. Corte-Real, and H. Xu, 2002: Multisite stochastic weather models for impact studies. *Int. J. Climatol.,* **22,** 1377–1397.

Salinger, M. J., J. A. Renwick, and A. B. Mullan, 2001: Interdecadal Pacific Oscillation and South Pacific climate. *Int. J. Climatol.,* **21,** 1705–1721.

Thompson, C. S., P. J. Thomson, and X. Zheng, 2007: Fitting a multisite daily rainfall model to New Zealand data. *J. Hydrol.,* **340,** 25–39.

Wilks, D. S., 1998: Multisite generalization of a daily stochastic precipitation generation model. *J. Hydrol.,* **210,** 178–191.

Zheng, X., and C. S. Thompson, 2007: Simulation of precipitation in the upper Waitaki catchment, New Zealand, and its relation to the interdecadal Pacific oscillation: Interannual and intraseasonal variability. *J. Hydrol.,* **339,** 105–117.

——, and R. W. Katz, 2008a: Mixture model of generalized chain-dependent processes and its application to simulation of interannual variability of daily rainfall. *J. Hydrol.,* **349,** 191–199.

——, and ——, 2008b: Simulation of spatial dependence in daily rainfall using multisite generators. *Water Resour. Res.,* **44,** W09403, doi:10.1029/2007WR006399.