WATER RESOURCES RESEARCH, VOL. 45, XXXXXX, doi:10.1029/2008WR007526, 2009

Simulation of multisite precipitation using an extended chain-dependent process

Xiaogu Zheng,[1,2] James Renwick,[1] and Anthony Clark[1]

Received 15 October 2008; revised 13 May 2009; accepted 12 August 2009; published XX Month 2009.

[1] The chain-dependent process is a popular stochastic model for precipitation sequence data. In this paper, the effect of daily regional precipitation occurrence is incorporated into the stochastic model. This model is applied to analyze the daily precipitation at a small number of sites in the upper Waitaki catchment, New Zealand. In this case study, the probability distributions of daily precipitation occurrence and intensity, spatial dependences, and the relation between precipitation and atmospheric forcings are simulated quite well. Specifically, some behaviors which are not well modeled by existing models, such as the extremal behavior of daily precipitation intensity, the lag 1 cross correlation of daily precipitation occurrence, spatial intermittency, and spatial correlation of seasonal precipitation totals, are significantly improved. Moreover, a new and simpler approach is proposed which successfully eliminates overdispersion, i.e., underestimation of the variance of seasonal precipitation totals.

Citation: Zheng, X., J. Renwick, and A. Clark (2009), Simulation of multisite precipitation using an extended chain-dependent process, *Water Resour. Res.*, *45*, XXXXXX, doi:10.1029/2008WR007526.

1. Introduction

[2] Stochastic models for observed precipitation data sequences are useful in applications such as drainage system design and hydrological design. They make up the most important step in construction of weather generators, which have wide applications in agriculture and ecosystem simulations [*Richardson*, 1981] and have application in climate change studies [*Wilks*, 1992; *Furrer and Katz*, 2007; *Brissette et al.*, 2007]. Although much progress has been achieved in the development of precipitation simulation tools, current challenges include the accurate representation of extremal behavior, the generation of multisite sequences with realistic spatial dependence, the need to represent realistic levels of interannual variability in the generated sequences, and the representation of complex dynamical structures within a relatively cheap computational framework [e.g., *Wheater et al.*, 2005].

[3] *Katz* [1977] proposed a stochastic model for single-site precipitation data called a chain-dependent process. Precipitation occurrence is modeled as a first-order Markov chain, and precipitation intensity is simulated using a power-transformed Gaussian distribution. During the 30 years since Katz introduced it, this stochastic precipitation model has been improved considerably. First, external forcing, internal cycles, and trends were incorporated by introducing threshold models [*Katz and Parlange*, 1993] and generalized linear models [e.g., *Furrer and Katz*, 2007]. Overdispersion was eliminated by introducing mixture models [e.g., *Katz and Zheng*, 1999; *Zheng and Katz*, 2008a] and other approaches [e.g., *Katz and Parlange*, 1998]. Several approaches for modeling the spatial dependence of precipitation were also proposed [*Wilks*, 1998; *Zheng and Katz*, 2008b].

[4] Despite all of this progress, the traditional chain-dependent process still has considerable shortcomings. First, it appears that the extremal behavior of precipitation is poorly modeled. It is widely believed that the assumption of a power-transformed Gaussian distribution is largely responsible. Use of other distributions, such as the mixture of the exponential [*Wilks*, 1998; *Brissette et al.*, 2007] and the gamma distribution [*Furrer and Katz*, 2007], has brought about some improvements, but extremal behavior is still underestimated. Second, spatial dependence is not well modeled. Specifically, spatial intermittence [*Wilks*, 1998] is still significant, and the lag 1 cross correlations of daily precipitation occurrence are often significantly underestimated [*Wilks*, 1998]. In this paper, we will further show that traditional chain-dependent models tend to underestimate the spatial dependence of seasonal precipitation totals.

[5] A possible reason for the existing chain-dependent process model not simulating these properties well is that the model is oversimplified. In fact, the existing multisite chain-dependent process models [e.g., *Zheng and Katz*, 2008b] assume that the marginal precipitation distribution at a single site is determined by the data at that site only and is independent of precipitation occurrences at other sites. However, multisite precipitation in a region is often forced by the same atmospheric circulation feature. So the distributions of occurrence and intensity at any single site are likely related to the precipitation occurrences at other sites.

[6] In this study, the traditional chain-dependent process is extended to include an index which represents the effect of regional precipitation occurrence for modeling

[1]National Institute of Water and Atmospheric Research, Wellington, New Zealand.

[2]College of Global Change and Earth System Science, Beijing Normal University, Beijing, China.

0043-1397/09/2008WR007526$09.00

both precipitation occurrence and intensity. Specifically, the precipitation intensity is still assumed to be power-transformed Gaussian, but its error variance is dependent on the precipitation occurrences at neighboring sites. We will show, through a case study, that the extended chain-dependent process can significantly improve the simulation of extremal behavior, spatial dependence, and interannual variability. Atmospheric forcing can be easily incorporated into the new model. Furthermore, overdispersion can be eliminated by introducing a random seasonal forcing.

[7] The paper is arranged as follows. The extended chain-dependent process is introduced in section 2. Section 3 describes the case study of a long-term daily precipitation data series using the proposed model. Finally, the discussion on the extended model and our conclusions are given in section 4.

2. Methodology

[8] Let $\mathbf{J}_t = (J_t(1), \ldots J_t(M))$ denote daily multisite precipitation occurrences (i.e., $J_t(m) = 1$ indicates a "wet day" and $J_t(m) = 0$ indicates a "dry day"), where $t = (1, \ldots, T)$ is a day within a season (for example, December–February) in a year and $m(=1, \ldots, M)$ is a geographic location. Let x_t denote a forcing variable on day t.

[9] To model daily precipitation at a single site, *Katz* [1977] introduced the chain-dependent process, and *Zheng and Katz* [2008a] introduced the generalized chain-dependent process. The main innovation of the new stochastic model proposed in this study is to introduce the following index into the generalized chain-dependent process,

$$K_t(m) \equiv \frac{1}{M-1} \sum_{m' \neq m} J_t(m') c(m', m), \tag{1}$$

where $c(m', m)$ is the correlation of precipitation occurrence between site pair m' and m, which can be estimated by observations. K_t(m) is referred to as the effect of regional precipitation occurrences around site m. A larger K_t(m) indicates more wet sites around the site m.

[10] For the new model, the conditional probability of daily precipitation occurrences at a single site given the multisite precipitation occurrence on the previous day is assumed to be the logistic regression form

$$\begin{aligned}
&\Pr(J_t(m) = 1|\mathbf{J}_{t-1}) = \\
&1 - 1/[1 + \exp(\alpha_0(m) + \alpha_1(m)K_{t-1}(m) + \alpha_2(m)x_t)] J_{t-1}(m) = 0 \\
&\Pr(J_t(m) = 1|\mathbf{J}_{t-1}) = \\
&1 - 1/[1 + \exp(\beta_0(m) + \beta_1(m)K_{t-1}(m) + \beta_2(m)x_t)] J_{t-1}(m) = 1,
\end{aligned} \tag{2}$$

where Pr indicates the probability function. Since a larger $K_{t\text{-}1}(m)$ indicates more wet sites around site m on the previous day, the site m is more likely to be wet on day t because of day-to-day persistence of atmospheric circulation. Therefore, the parameters $\alpha_1(m)$ and $\beta_1(m)$ are expected to be positive.

[11] Let $R_t(m)$ denote daily precipitation amounts on day t and at site m. It is further assumed that on a wet day (i.e., $J_t(m) = 1$), the transformed variable $R_t^{q(m)}(m)$ has a Gaussian distribution. The values $q = 1/2$, 1/3, and 1/4 are commonly employed to account for the high degree of positive skewness in the distribution of daily precipitation amounts. In this study, $q(m)$ is initially assigned to be 1/4 and later may be adjusted to fit the extremes of daily precipitation intensity at individual sites. Moreover, the mean of $R_t^{q(m)}(m)$ is assumed to be

$$\begin{aligned}
&E_t(m) \equiv \\
&\mu_0(m) + \mu_1(m)K_t(m) + \mu_2(m)K_{t-1}(m) + \mu_3(m)x_t + \mu(m)\gamma_t,
\end{aligned} \tag{3}$$

where γ_t is a seasonal random Gaussian variable with zero mean and unit variance, which remains a constant over a season and is statistically independent with respect to season. The standard deviation of $R_t^{q(m)}(m)$ is assumed to be

$$S_t(m) \equiv \sigma_0(m) + \sigma_1(m)K_t(m). \tag{4}$$

[12] To investigate the relation between $K_t(m)$ and $R_t^{q(m)}(m)$, a scatterplot of their values for a site (i.e., Franz Josef; see Figure 1) is shown in Figure 2. Figure 2 shows that $K_t(m)$ and $R_t^{q(m)}(m)$ are positively correlated. Hence, $\mu_1(m)$ is expected to be positive. Figure 2 also shows that as $K_t(m)$ increases, the error in $R_t^{q(m)}(m)$ expands. Therefore, the standard deviation of $R_t^{q(m)}(m)$ is assumed to be in the linear form of (4), and $\sigma_1(m)$ is expected to be positive. Parameters $\alpha_2(m)$, $\beta_2(m)$, and $\mu_3(m)$ represent the effect of a single atmospheric forcing x_t. Finally, $\mu(m)\,\gamma_t$ is a seasonal random variable which forces the variance of simulated seasonal precipitation total close to that observed.

[13] Equations (1)–(4) define a daily precipitation model which we referred to as the extended chain-dependent process forced by x_t and seasonal random forcing γ_t because under the constraints $\alpha_1 = \alpha_2 = 0$, $\beta_1 = \beta_2 = 0$, $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$, and $\sigma_1 = 0$, it is a standard multisite chain-dependent process [*Zheng and Katz*, 2008b]. A major difference between a multisite chain-dependent process and the extended chain-dependent process is that for the former model, the marginal probability distribution functions of precipitation occurrence and precipitation intensity are independent of the precipitation occurrences at other sites. This is not the case for the latter model, as $K_t(m)$ is related to the precipitation occurrence around site m.

[14] Practical estimation approaches for α_0, α_1, α_2, β_0, β_1, and β_2 are documented in Appendix A, and practical estimation approaches for μ_0, μ_1, μ_2, μ_3, and μ are documented in Appendix B.

[15] In generating a multisite precipitation time series, we further generate standard Gaussian vectors $\{W_t(1), \ldots, W_t(\mathrm{M})\}$ and $\{Z_t(1), \ldots, Z_t(M)\}$ for precipitation occurrence and intensity, respectively, and a standard Gaussian random variable γ_t, which is unchanged within every season. To correctly simulate the spatial dependence of precipitation occurrence and of precipitation intensity, the $\{W_t(1), \ldots, W_t(M)\}$ and $\{Z_t(1), \ldots, Z_t(M)\}$ must be spatially correlated [e.g., *Wilks*, 1998]; our methodologies for estimation of the spatial correlation coefficients are documented in Appendix D. Moreover, $\{W_t(1), \ldots, W_t(M)\}$, $\{Z_t(1), \ldots, Z_t(M)\}$, and γ_t are statistically independent of each other and of day t.

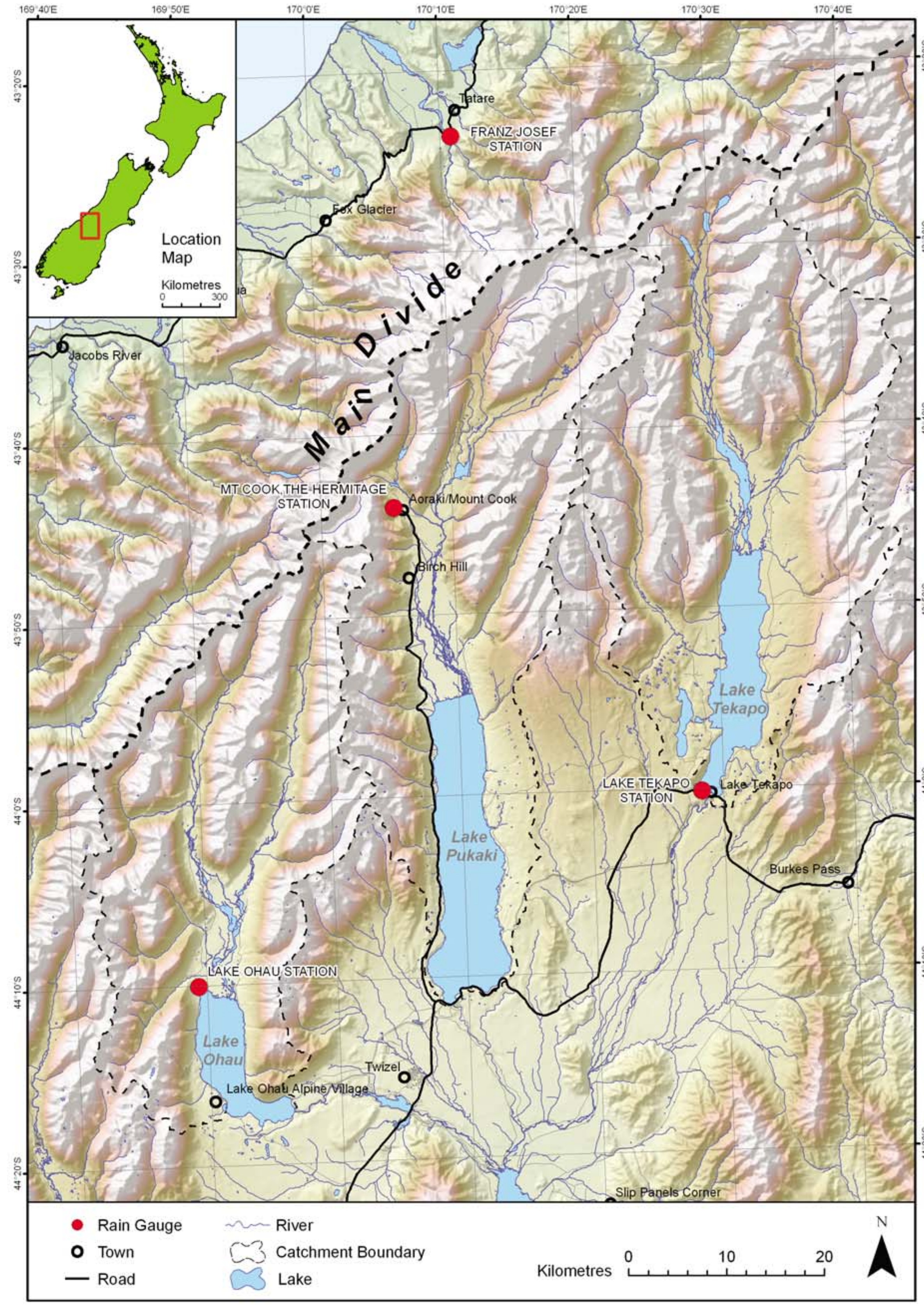

Figure 1. The geographical features of the Waitaki catchment, Southland, New Zealand.

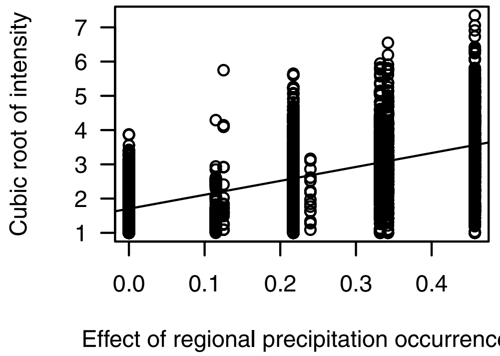

Figure 2. Plot of the effect of regional precipitation occurrences for all wet days (i.e., K_t, equation (1)) versus the cube root of daily precipitation intensity at Franz Josef (see Figure 1).

Table 2. Estimated Parameters for Model 4 for Single Site[a]

	Site m			
	1	2	3	4
α_0 (m)	−1.222	−2.365	−2.265	−0.746
α_1 (m)	1.268	2.225	1.199	0.000
α_2 (m)	0.089	−0.097	0.000	0.093
β_0 (m)	−0.198	−1.739	−1.531	0.477
β_1 (m)	0.557	1.227	0.840	0.243
β_2 (m)	0.113	0.000	0.000	0.104
μ_0 (m)	1.275	1.451	1.198	1.694
μ_1 (m)	0.752	0.261	0.268	1.361
μ_2 (m)	0.000	0.112	0.052	0.000
μ_3 (m)	0.030	0.031	0.000	0.067
μ (m)	0.200	0.100	0.100	0.200
σ_0 (m)	0.427	0.398	0.271	0.587
σ_1 (m)	0.189	0.025	0.103	0.534

[a]The site numbers are shown in Figure 3.

[16] Knowing the estimated parameters and the generated random fields, multisite precipitation time series can be generated by Monte Carlo simulation (Appendix C).

3. A Simulation Study

[17] The upper Waitaki catchment is situated in and east of the Southern Alps, South Island, New Zealand. There are three lakes: Lake Tekapo, Lake Pukaki, and Lake Ohau (see Figure 1). These lakes supply water for hydroelectric power generation; they provide about one fourth of the electricity generation capacity in New Zealand. For better management of these water resources, the hydrological catchment model TOPNET [*Bandaragoda et al.*, 2004] is used to simulate the inflow into the lakes and then the outflow from the lakes. Since daily precipitation is the most important forcing for TOPNET, we aim to simulate an ensemble of regional daily precipitation to force TOPNET for the upper Waitaki catchment.

[18] The simulated daily precipitation must correctly represent the spatial variability at basin scale and the temporal variability at all time scales, specifically, the decadal time scale. In order to estimate the rainfall variability over the next 2–3 decades, a climate variable is needed that is both predictable and significantly associated with precipitation on a decadal time scale. Fortunately, the Interdecadal Pacific Oscillation (IPO) may be such a climate variable. The IPO has significant impacts on precipitation and river flows in the upper Waitaki catchment, particularly for the austral summer season (December–January–February (DJF)). The negative IPO phase is generally associated with lower rainfall and inflows, and the positive IPO phase is generally associated with higher rainfall and inflows [*Zheng and Thompson*, 2007]. For this reason, the forcing variable x_t used in this study is the low-frequency IPO index, provided by the Hadley Centre of the United Kingdom Meteorological Office [*Folland et al.*, 1999]. It is derived from the third empirical orthogonal function pattern of 13 year low-pass-filtered global SST [see *Zheng and Thompson*, 2007, Figure 2].

Table 1. Model Hierarchy

Name	Constraints on Parameters
Model 1: multisite chain-dependent process	$\alpha_1 = \alpha_2 = 0$, $\beta_1 = \beta_2 = 0$, $\mu_1 = \mu_2 = \mu_3 = \mu = 0$, $\sigma_1 = 0$
Model 2: extended chain-dependent process	$\alpha_2 = \beta_2 = \mu_3 = \mu = 0$
Model 3: extended chain-dependent process forced by IPO	$\mu = 0$
Model 4: extended chain-dependent process forced by IPO and random seasonal forcing	none

[19] There are only four rainfall stations in or near the upper Waitaki catchment with records covering the period 1953–2000: Lake Tekapo, Lake Ohau, Mount Cook, and Franz Josef (see Figure 1 for locations). Their record lengths cover the period 1953–2000, which roughly spans one complete cycle of the IPO, i.e., one positive and negative phase. The daily precipitation has been power transformed. Values for q of 1/4, 1/4, 1/4, and 1/3 were adopted for Lake Tekapo, Lake Ohau, Mount Cook, and Franz Josef, respectively. All these values were initially chosen as 1/4. However, for Franz Josef, it was found that the tail of daily precipitation intensity is overestimated for q = 1/4. This may be due to Franz Josef being the only station west of the main divide, so larger rainfalls appear more frequently. 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.

[20] A hierarchy of four models was fitted to the austral summer season daily precipitation for the four long-term rainfall stations: (1) the multisite chain-dependent process, (2) the extended chain-dependent process, (3) the extended chain-dependent processes forced by the IPO, and (4) the extended chain-dependent processes forced by IPO and seasonal random forcing. Their names and the constraints on the parameters are listed in Table 1. We will investigate model 4 in the simulation, while models 1–3 are treated as alternatives for comparison. All models are fitted to the daily precipitation at the four sites during austral summer for the period 1953–2000. On the basis of the fitted parameters (shown in Tables 2 and 3) and the observed seasonal IPO index, 100 independent simulations of the DJF

Table 3. Estimated Spatial Correlation for the Gaussian Fields $\{W_t(m), m = 1, \ldots, 4\}$ and $\{Z_t(m), m = 1, \ldots, 4\}$[a]

	1	2	3	4
1	1	0.712	0.822	0.897
2	0.526	1	0.711	0.623
3	0.363	0.492	1	0.714
4	0.742	0.397	0.274	1

[a]The site numbers are shown in Figure 3. Correlations for $\{W_t(m), m = 1, \ldots, 4\}$ are in the top right, and correlations for $\{Z_t(m), m = 1, \ldots, 4\}$ are in the bottom left.

daily precipitation over the 47 year period are generated using the four models.

3.1. Daily Precipitation Intensity

[21] The Q-Q plots of observed daily precipitation intensity versus that simulated for each site and each model are shown in Figure 3. Generally speaking, model 1 underestimates the distribution, specifically, for the extremes. All other models (2–4) simulate the distribution of daily precipitation intensity quite well.

[22] The Q-Q plots of observed regional daily precipitation totals versus those simulated using models 1–4 are shown in Figure 4. Figure 4 shows that models 2–4 simulate the distribution quite well, while model 1 tends to underestimate the distribution, specifically, for extremal behavior.

3.2. Spatial Dependence of Daily Precipitation

[23] The correlations of the two Gaussian random fields are estimated using model 1 (see Appendix D) and applied in the simulation study using models 2–4. For all models, the spatial dependence of precipitation occurrence is overestimated, and the spatial dependence of precipitation intensity is underestimated, except the spatial dependence of precipitation occurrence for model 1. However, after the initially estimated correlations are adjusted (see Appendix D), the biases of the precipitation occurrence and intensity are strongly reduced. The final estimated correlations of the two Gaussian random fields are shown in Table 3.

[24] The lag 1 cross-correlation coefficients of precipitation occurrence observed and simulated are shown in Table 4. The coefficients simulated using models 2–4 are very close to the observed. However, the coefficients simulated by model 1 are negatively biased. The improvement is mainly to the east of the main divide. This is consistent with the fact that $\alpha_1(m)$ and $\beta_1(m)$ are much more significant to the east of the main divide than to the west (see Table 2).

[25] Accurate simulation of the dependence between precipitation intensity and occurrence at other sites is important in several applications, for example, drainage system design and simulation of regional agricultural yields. To estimate whether the spatial intermittence problem [*Wilks*, 1998] is handled appropriately, *Wilks* [1998] defined

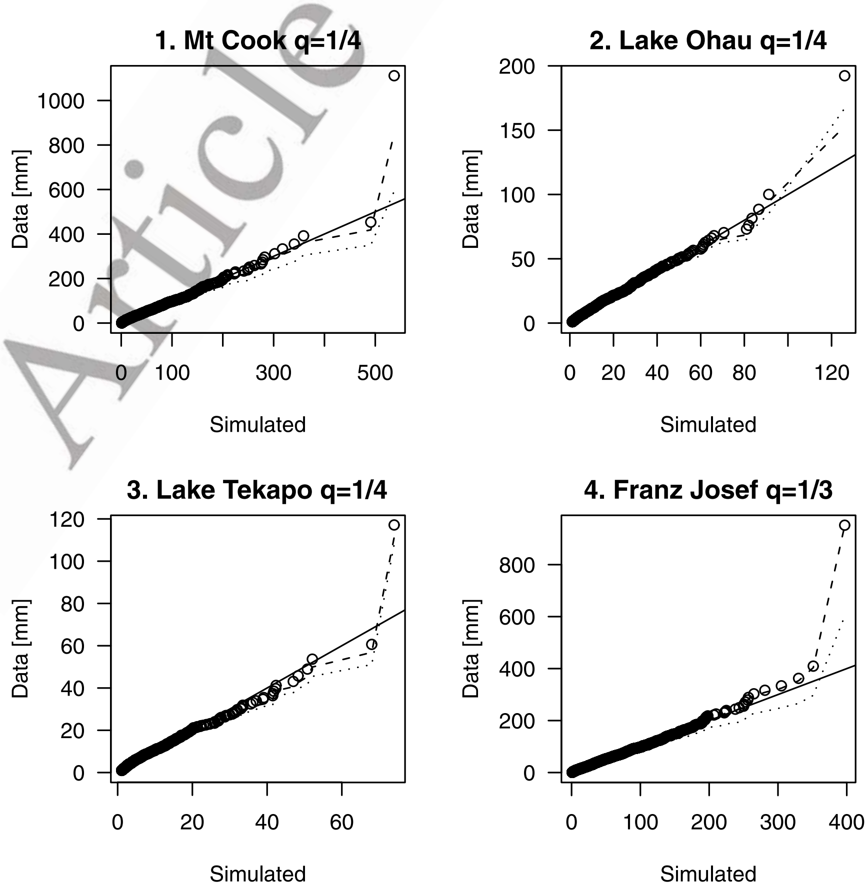

Figure 3. Q-Q plots of observed versus simulated daily precipitation intensity. Dotted line is model 1, dashed line is model 2, and open symbols are model 4. The Q-Q plot for model 3 (not show here) is very close to that for model 4.

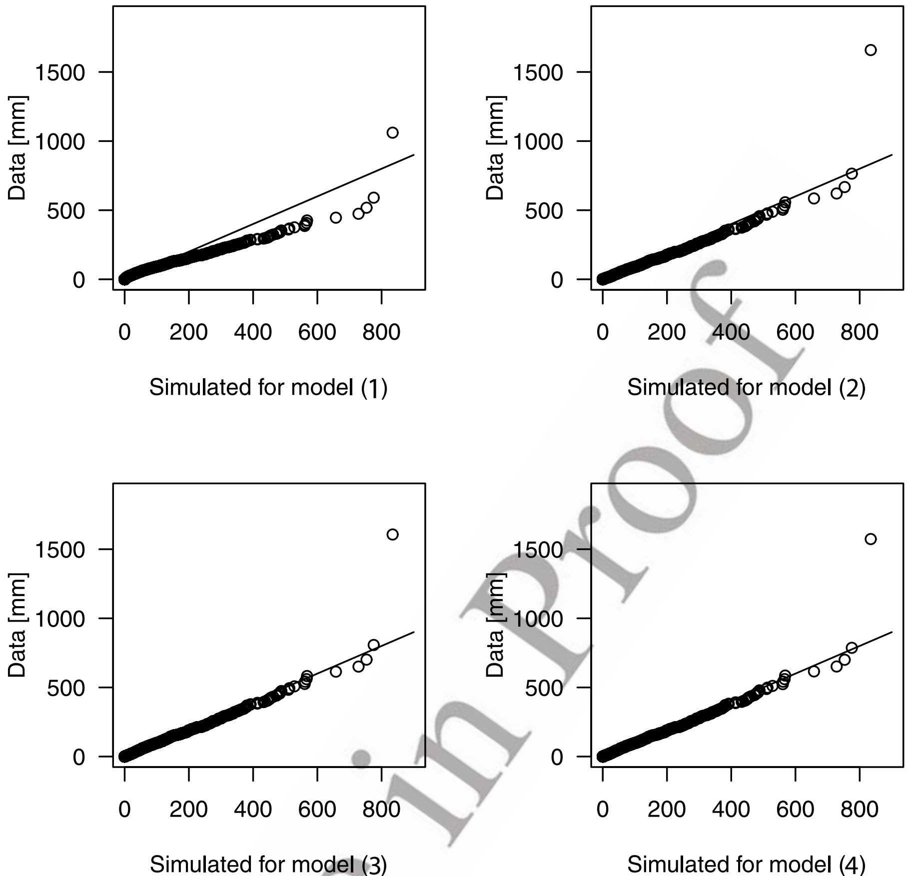

Figure 4. Q-Q plots of observed regional daily precipitation total versus that simulated.

an index of the spatial intermittence called the continuity ratio between two sites m and m':

$$C(m, m') \equiv E(R_t(m)|J_t(m) = 1, J_t(m') = 0)/ E(R_t(m)|J_t(m) = 1, J_t(m') = 1). \quad (5)$$

It is a measure of the dependence of the mean of precipitation intensity at site m on the precipitation occurrence at site m'.

[26] Figure 5 shows the plots of the continuity ratios observed versus simulated for all 12 site pairs. It shows that the continuity ratios simulated by model 1 are all close to 1. This indicates that, regardless of whether the other sites are wet or dry, the mean of the precipitation intensity at any single site is not changed much. However, this is not the case for the observations. Figure 5 also shows that the continuity ratios simulated using models 2–4 are quite comparable to those observed.

3.3. Interannual Variability

[27] Correlations between seasonal precipitation totals and the IPO index are shown in Table 5. The correlations are reasonably significant, especially for Mount Cook and Franz Josef (for the total of 47 samples, a correlation of 0.28 is at the 5% significant level, and a correlation of 0.35 is at the 1% significant level). While the correlation was simulated quite well by models 3 and 4, it was completely missed by models 1 and 2.

[28] Spatial correlations of seasonal precipitation totals for all site pairs are shown in Table 6. The correlation is very strong in the observations. The correlation simulated using model 1 is weak (negatively biased). The correlations simulated by models 2 and 3 are improved but still fall short of the observed. However, the correlation simulated by model 4 is further improved and is close to that observed.

[29] Standard deviations of seasonal precipitation totals are shown in Table 7. Table 7 shows that the standard deviations are significantly underestimated by model 1. This phenomenon is referred to as overdispersion [*Katz and Zheng*, 1999]. Overdispersion is reduced to some extent by model 2 and is further eliminated by model 3, but not completely. Finally, the overdispersion is almost fully eliminated by model 4.

[30] Q-Q plots of the regional seasonal precipitation totals simulated by models 1–4 are shown in Figure 6. Figure 6 shows that model 1 tends to underestimate the wet extremes

Table 4. Lag 1 Cross Correlation of Daily Precipitation Occurrence[a]

Site Pair	Model 1	Model 2	Model 3	Model 4	Observed
1–2	0.06	0.31	0.31	0.31	0.32
1–3	0.08	0.22	0.22	0.22	0.21
1–4	0.21	0.22	0.22	0.22	0.24
2–3	0.09	0.18	0.18	0.18	0.16
2–4	0.11	0.13	0.13	0.13	0.12
3–4	0.12	0.14	0.14	0.14	0.13
Average	0.11	0.20	0.20	0.20	0.20

[a]The site numbers are shown in Figure 3.

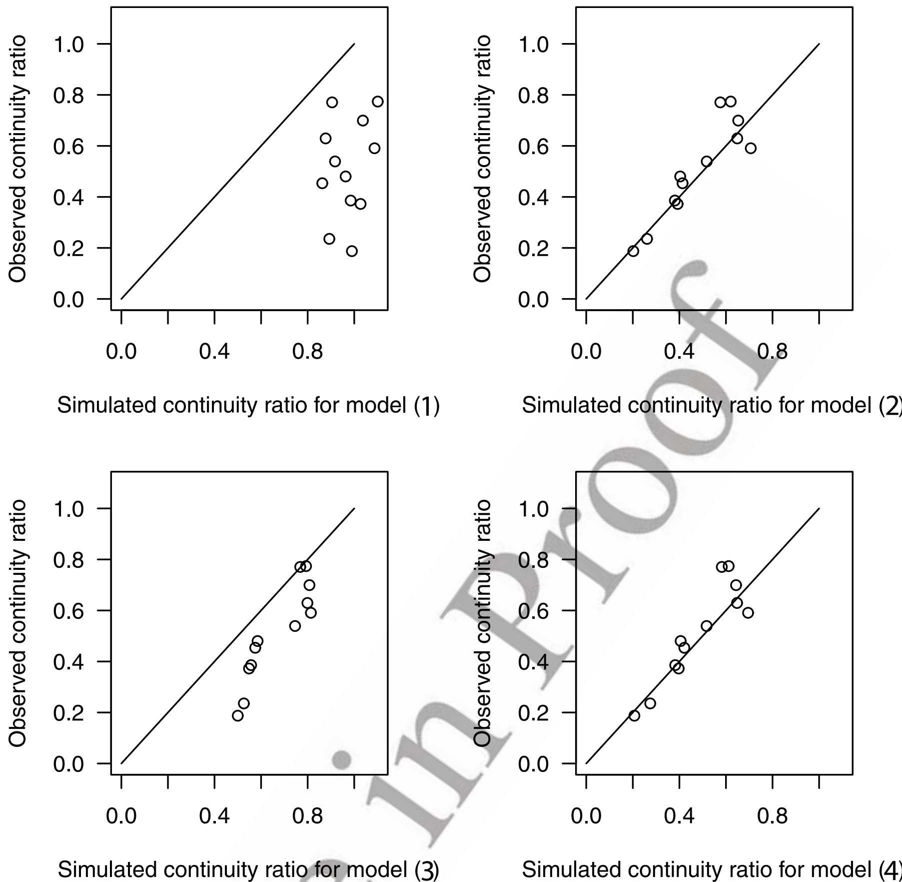

Figure 5. Scatterplot of the continuity ratios of the observed versus those simulated.

of total precipitation by about 1000 mm and to overestimate the dry extremes by about 500 mm. The situation is progressively improved from model 2 to model 3 and is modeled quite well by model 4. *Zheng and Katz* [2008a] showed that the probability distribution of the seasonal precipitation totals can be correctly simulated by the mixture chain-dependent process. Here we provide an alternative model to eliminate the overdispersion.

[31] The distributions of dry runs and wet spells of precipitation were also examined. Generally speaking, the distributions simulated using all models 1–4 coincide well with the observed.

4. Discussion and Conclusions

[32] We have demonstrated several advantages of the extended chain-dependent process over the multisite chain-dependent process. To investigate the roles played by individual parameters, these parameters are dropped in turn from the extended chain-dependent process forced by IPO and seasonal random forcing, and the analysis in section 3 is repeated. As a result, the following conclusions emerge.

[33] The parameter $\sigma_1(m)$ plays the most important role in improving the extremal behavior of precipitation, suggesting some spatial coherence in extreme behavior. The parameters $\mu_1(m)$ and $\mu_2(m)$ also play some role. The intermittence problem can be solved only by introducing $\mu_1(m)$. The parameters $\alpha_1(m)$ and $\beta_1(m)$ play the dominant role in correctly modeling the lag 1 cross correlation of daily precipitation occurrence. The parameters $\alpha_1(m)$, $\beta_1(m)$, $\mu_1(m)$, and $\mu(m)$ are all important for improving the spatial dependence of seasonal precipitation totals. The reason $\alpha_1(m)$ and $\beta_1(m)$ played a role may be because of the dependence of seasonal totals on the daily lag 1 cross

Table 5. Correlations Between Seasonal Precipitation Totals and the IPO Index[a]

Site	Model 1	Model 2	Model 3	Model 4	Observed
Mt. Cook	0.02	0.02	0.33	0.24	0.29
Ohau	0.00	0.02	0.22	0.18	0.16
Tekapo	0.00	0.01	0.09	0.07	0.15
Franz Josef	0.00	0.02	0.42	0.36	0.40
Average	0.01	0.02	0.27	0.22	0.25

[a]The site numbers are shown in Figure 3.

Table 6. Similar to Table 2, but for Cross Correlation of Seasonal Precipitation Totals

Site Pair	Model 1	Model 2	Model 3	Model 4	Observed
1–2	0.36	0.72	0.73	0.80	0.89
1–3	0.29	0.61	0.60	0.71	0.73
1–4	0.58	0.85	0.86	0.88	0.86
2–3	0.35	0.65	0.65	0.74	0.88
2–4	0.28	0.66	0.67	0.74	0.80
3–4	0.24	0.58	0.57	0.66	0.66
Average	0.35	0.68	0.68	0.76	0.80

Table 7. Similar to Table 3, but for Standard Deviation of Seasonal Precipitation Totals

Site	Model 1	Model 2	Model 3	Model 4	Observed
Mt. Cook	281	339	361	502	495
Ohau	80	89	92	107	115
Tekapo	44	47	49	57	61
Franz Josef	316	382	424	483	528
Average	181	215	233	287	300

precipitation field [e.g., *Zheng*, 1996], and $\alpha_1(m)$ and $\beta_1(m)$ help to improve the daily lag 1 spatial dependence. Finally, as expected, $\mu(m)$ plays the key role in eliminating overdispersion.

[34] The cases when extremes are not well modeled by stochastic precipitation models were often attributed to the tails of statistical distributions not being heavy enough or atmospheric forcing being neglected. In this case, the general extreme value distribution is recommended for modeling the extremal behavior of precipitation [e.g., *Koutsoyiannis*, 2004; *Furrer and Katz*, 2008]. In this study, we showed that excluding the effect of the precipitation occurrence at the regional scale may be a major reason for extremes being underestimated. As shown here, when such an index is appropriately incorporated, the extremes of precipitation can be modeled quite well, even using the power-transformed Gaussian distribution and without introducing any atmospheric forcing. Adjustment of the power transform parameter q would further improve the simulated extremal behavior. Moreover, by appropriately introducing spatial dependence of daily precipitation, extremes of the regional daily precipitation total can be correctly estimated (Figure 6).

[35] All the improvement in extremal behavior and spatial dependence can be achieved by using precipitation data only, that is, by model 2, without any atmospheric forcing. Therefore, model 2 is useful because forcing is not always available or necessary, for example, in application to drainage system design.

[36] In this study, we have demonstrated that a single atmospheric forcing can be effectively modeled by assuming $\alpha_2(m) \neq 0$, $\beta_2(m) \neq 0$, and $\mu_3(m) \neq 0$. However, as with other rainfall generators based on generalized linear models [e.g., *Furrer and Katz*, 2007], this approach can be easily generalized to incorporate multiple atmospheric forcing variables, seasonal cycles, and trends.

[37] In this study, the parameters are estimated in an ad hoc manner, and neither the robustness nor the precision of the estimates has been fully investigated. However, the results seem acceptable because all the basic statistics are correctly simulated with these parameters in this case study. In the future, we plan to further improve the parameter estimation and to investigate the impact of the ad hoc

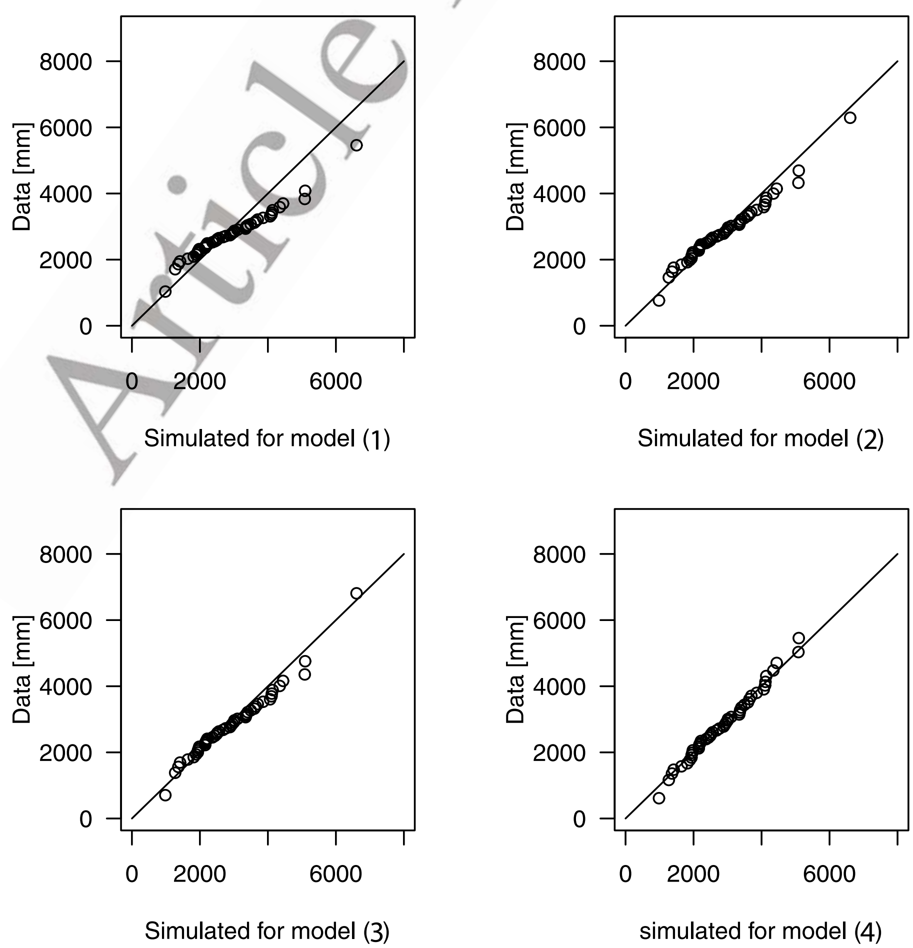

Figure 6. Q-Q plots of observed regional seasonal precipitation totals versus that simulated.

estimation. We also plan to fit this model to precipitation data at more sites, using more forcing data, to further test the efficacy of the model. Specifically, we plan to use atmospheric forcing generated from global circulation model output to downscale climate change scenarios for estimation of regional rainfall in impact studies.

[38] In conclusion, the introduction of a regional precipitation index into a multisite chain-dependent process improved the simulation of extremes in rainfall intensity, spatial correlations of occurrence, and seasonal totals. In addition, introduction of atmospheric forcing, in this case the IPO and random seasonal effects, led to a reduction in overdispersion. The models investigated offer several advantages over the traditional chain-dependent process. In this case study, the new stochastic precipitation model significantly improves the quality of precipitation simulation.

Appendix A: Estimation of $\alpha_0(m)$, $\alpha_1(m)$, $\alpha_2(m)$, $\beta_0(m)$, $\beta_1(m)$, and $\beta_2(m)$

[39] Equation (2) is in a typical logistic regression form [*McCullagh and Nelder*, 1989]. So $\alpha_0(m)$, $\alpha_1(m)$, and $\alpha_2(m)$ can be estimated using all the precipitation occurrence observations where the previous day was dry. Similarly, $\beta_0(m)$, $\beta_1(m)$, and $\beta_2(m)$ can be estimated using all the precipitation occurrence observations where the previous day was wet. In this study, they are estimated using the function glm in the open source statistical package R.

Appendix B: Estimation of $\mu_0(m)$, $\mu_1(m)$, $\mu_2(m)$, $\mu_3(m)$, $\sigma_0(m)$, $\sigma_1(m)$, and $\mu(m)$

[40] We have assumed that the power-transformed precipitation intensity $R_t^{q(m)}(m)$ has a Gaussian distribution with the mean represented by expression (3) and the standard deviation represented by expression (4). Since there is a random effect term γ_t in expression (3), $R_t^{q(m)}(m)$ can be modeled by a general linear mixed model [e.g., *Jones*, 1992, chapter 2.1]. In principle, the parameters of $R_t^{q(m)}(m)$ can be estimated by the maximum likelihood estimation [see *Jones*, 1992, chapters 2.2–2.6]. However, the reason for introducing the random effect term γ_t here is to correctly estimate the seasonal mean precipitation. Since the simulated seasonal mean precipitation is not power transformed and is related to the simulated precipitation occurrence, fitting the general linear mixed model by the maximum likelihood estimation may not achieve our goal.

[41] In this study, we use an alternative empirical approach to estimate the parameters in expressions (3) and (4). Since the random effect term γ_t is with mean zero, $\mu_0(m)$, $\mu_1(m)$, $\mu_2(m)$, $\mu_3(m)$, $\sigma_0(m)$, and $\sigma_1(m)$ are estimated under the assumption $\mu(m) = 0$. In this case, $R_t^{q(m)}(m)$ has a Gaussian distribution, and the -2 log likelihood function of $R_t^{q(m)}(m)$ on wet days is

$$L(m) \equiv \sum_t J_t(m)\left\{ \ln\left[(\sigma_0(m) + \sigma_1(m)K_t(m))^2\right] - \frac{(R_t^q(m) - \mu_0(m) - \mu_1(m)K_t(m) - \mu_2(m)K_{t-1}(m) - \mu_3(m)x_t)^2}{[\sigma_0(m) + \sigma_1(m)K_t(m)]^2}\right\}. \tag{B1}$$

In principle, the parameters can be estimated by minimizing function (B1). However, since there are six parameters in (B1), direct optimization may be difficult. For this reason, we use the following approximate estimation. First, $\mu_0(m)$, $\mu_1(m)$, $\mu_2(m)$, and $\mu_3(m)$ are estimated by the stepwise regression assuming $R_t^{q(m)}(m)$ has constant error variance. Then $\sigma_0(m)$ and $\sigma_1(m)$ are estimated by minimizing (B1), but with $\mu_0(m)$, $\mu_1(m)$, $\mu_2(m)$, and $\mu_3(m)$ being fixed as estimated previously. In this study, the function nlminb in the open source statistical package R is applied for the optimization.

[42] After the parameters $\mu_0(m)$, $\mu_1(m)$, $\mu_2(m)$, $\mu_3(m)$, $\sigma_0(m)$, and $\sigma_1(m)$ have been estimated, $\mu(m)$ is determined by moment estimation. To obtain more details, we introduce the term $\mu(m)\gamma_t$ into model 3 to force the variance of the simulated seasonal precipitation total close to that observed. For each site m, $\mu(m)$ increases at step 0.05 from zero until the two variances become sufficiently close.

Appendix C: Generating Multisite Precipitation

[43] Knowing the generated spatially correlated random Gaussian fields $\{W_t(1), \ldots, W_t(M)\}$ and $\{Z_t(1), \ldots, Z_t(M)\}$ (see Appendix D) and initial occurrence states $J_0(m)$, $m = 1, \ldots, M$, we can generate, by Monte Carlo simulation, a multisite rainfall time series iteratively with day t.

C1. Occurrence

[44] For every $m = 1, \ldots, M$, construct the precipitation occurrences transition probability $\Pr(J_t(m) = 1|\mathbf{J}_{t-1})$ using equation (2) (where $K_{t-1}(m)$ has been constructed at previous time step day $t-1$). Then the precipitation occurrence is constructed by using

$$J_t(m) = \begin{cases} 1, \Phi(W_t(m)) \le \Pr(J_t(m) = 1|\mathbf{J}_{t-1}) \\ 0, \Phi(W_t(m)) > \Pr(J_t(m) = 1|\mathbf{J}_{t-1}), \end{cases} \tag{C1}$$

where Φ is the standard Gaussian probability distribution function, so $\Phi(W_t(m))$ is a uniform random variable on the interval [0, 1]. In this study, $\Phi(W_t(m))$ is calculated by using the function pnorm in the open source statistical package R.

C2. Intensity

[45] For every $m = 1, \ldots, M$, construct the effect of regional precipitation occurrence on day t $K_t(m)$ using equation (1). Then the precipitation intensity can be constructed by using

$$R_t^{q(m)}(m) = J_t(m)[(\sigma_0(m) + \sigma_1(m)K_t(m))Z_t(m) + \mu_0(m) + \mu_1(m)K_t(m) + \mu_2(m)K_{t-1}(m) + \mu_3(m)x_t + \mu(m)\gamma_t]. \tag{C2}$$

Appendix D: Estimating Correlations of Gaussian Fields

[46] In this study, the cross correlations of the Gaussian fields $\{W_t(1), \ldots, W_t(M)\}$ and $\{Z_t(1), \ldots, Z_t(M)\}$ are

initially estimated under the constraints $\alpha_1 = \alpha_2 = 0$, $\beta_1 = \beta_2 = 0$, $\mu_1 = \mu_2 = \mu_3 = \mu = 0$, and $\sigma_1 = 0$ (i.e., model 1), assuming the site-specific parameters $q(m)$, $\alpha_0(m)$, $\beta_0(m)$, $\mu_0(m)$, and $\sigma_0(m)$ are estimated using methodology documented in Appendixes A and B.

[47] The correlation between $Z_t(m)$ and $Z_t(n)$ is denoted by $\psi(m, n)$ and can be estimated as

$$\hat{\psi}(m,n) = \frac{\sum_{t:r_t(m)r_t(n)>0} \left(r_t^{q(m)}(m) - \hat{\mu}_0(m)\right)\left(r_t^{q(n)}(n) - \hat{\mu}_0(n)\right)}{\hat{\sigma}_0(m)\hat{\sigma}_0(n)}, \tag{D1}$$

where $r_t(m)$ is the observed precipitation intensity on day t at site m [e.g., *Zheng and Katz*, 2008b].

[48] The correlation between $W_t(m)$ and $W_t(n)$ is denoted by $\omega(m, n)$ and can be estimated as follows. Note that $\{J_{y,t}(m), J_{y,t}(n), t = 1, \ldots, T\}$ is a bivariate Markov chain [*Zheng and Katz*, 2008b]. By equation (C1), the transition probability from $\{J_{t-1}(m) = k, J_{t-1}(n) = k'\}$ to $\{J_t(m) = j, J_t(n) = j'\}$ (denoted by $P_{kk',jj'}(m, n)$) is

$$P_{kk',11}(m,n) = \Pr\{\Phi(W_t(m)) \le P_{k,1}(m); \Phi(W_t(n)) \le P_{k',1}(n)\}, \tag{D2}$$

$$P_{kk',10}(m,n) = P_{k,1}(m) - P_{kk',11}(m,n), \tag{D3}$$

$$P_{kk',01}(m,n) = P_{k',1}(n) - P_{kk',11}(m,n), \tag{D4}$$

$$P_{kk',00}(m,n) = 1 - P_{k,1}(m) - P_{k',1}(n) + P_{kk',11}(m,n). \tag{D5}$$

where $P_{k,j}(m)$ denotes the transition probability from $\{J_{t-1}(m) = k\}$ to $\{J_t(m) = j\}$.

[49] By the ergodic theory of Markov chains [e.g., *Feller*, 1971], the bivariate invariant probability measure Pr $(J_{t-1}(m) = j, J_{t-1}(n) = j')$ of the transition probability matrix $\mathbf{P}$ is the last row of $\mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{A}^{\mathrm{T}})^{-1}$, where the partitioned matrix $\mathbf{A} = [\mathbf{I} - \mathbf{P}, 1]$; $\mathbf{I}$ is the identity matrix, and all elements of column vector 1 are 1. Since $\mathbf{P}$ is uniquely determined by $\omega(m, n)$, the invariant probability measure Pr $(J_{t-1}(m) = 1, J_{t-1}(n) = 1)$ is uniquely determined. The function for calculating multivariate Gaussian probability distribution (i.e., Pr in equation (D2)) is available, for example, the function dmvnorm in the library mvtnorm of the open source statistical package R. So, given $\omega(m, n)$, Pr $(J_t(m) = 1, J_t(n) = 1)$ can be computed, and the modeled daily cross correlation of precipitation occurrence between site pair m and n is

$$C(J_t(m), J_t(m)) = \frac{\Pr(J_t(m)=1, J_t(n)=1) - \Pr(J_t(m)=1)\Pr(J_t(n)=1)}{\sqrt{\Pr(J_t(m)=1)\Pr(J_t(m)=0)\Pr(J_t(n)=1)\Pr(J_t(n)=0)}}. \tag{D6}$$

Finally, $\omega(m, n)$ is chosen such that the modeled cross correlation (expression (D6)) is equal to the cross correlation of the observed occurrence.

[50] When the initially estimated correlations are applied to model 2, correlations of precipitation occurrence (intensity) are likely to be overestimated (underestimated). To correct this bias, for every site pair m and n, the initially estimated correlation between $W_t(m)$ and $W_t(n)$ is multiplied by the ratio of the correlation of the observed occurrence to the correlation of the simulated occurrence (see Appendix C) using model 2 with initially estimated correlations of $\{W_t(1), \ldots, W_t(M)\}$. A similar approach can be applied to correct the bias of the initially estimated correlations of the Gaussian field $\{Z_t(1), \ldots, Z_t(M)\}$.

[51] **Acknowledgments.** This work was supported by the New Zealand Foundation for Research, Science and Technology (contract C01X0302), the project sponsored by SRF for ROCS, SEM China, and projects 40875062 and 40975062 supported by NSFC, China. We thank Demetris Koutsoyiannis, three anonymous reviewers, Peter Thomson for useful comments, James Sturman for producing Figure 1, and Hadley Centre UKMO for the IPO time series.

References

Bandaragoda, C., D. G. Tarboton, and R. Woods (2004), Application of TOPNET in the distributed model intercomparison project, *J. Hydrol.*, *298*, 178–201, doi:10.1016/j.jhydrol.2004.03.038.

Brissette, F. P., M. Khalili, and R. Loconte (2007), Efficient stochastic generation of multi-site synthetic precipitation data, *J. Hydrol.*, *345*, 121–133, doi:10.1016/j.jhydrol.2007.06.035.

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, vol. 1, 3rd ed., John Wiley, New York.

Folland, C. K., D. E. Parker, A. W. Colman, and R. Washington (1999), Large scale modes of ocean surface temperature since the late nineteenth century, in *Beyond El Niño: Decadal and Interdecadal Climate Variability*, edited by A. Navarra, pp. 73–102, Springer, Berlin.

Furrer, E. M., and R. W. Katz (2007), Generalized linear modeling approach to stochastic weather generators, *Clim. Res.*, *34*, 129–144, doi:10.3354/cr034129.

Furrer, E. M., and R. W. Katz (2008), Improving the simulation of extreme precipitation events by stochastic weather generators, *Water Resour. Res.*, *44*, W12439, doi:10.1029/2008WR007316.

Jones, R. H. (1992), *Longitudinal Data with Serial Correlation: A State-Space Approach*, Chapman and Hall, London.

Katz, R. W. (1977), An application of chain-dependent processes to meteorology, *J. Appl. Probab.*, *14*, 598–603, doi:10.2307/3213463.

Katz, R. W., and M. B. Parlange (1993), Effects of an index of atmospheric circulation on stochastic properties of precipitation, *Water Resour. Res.*, *29*, 2335–2344, doi:10.1029/93WR00569.

Katz, R. W., and M. B. Parlange (1998), Overdispersion phenomenon in stochastic modeling of precipitation, *J. Clim.*, *11*, 591–601, doi:10.1175/1520-0442(1998)011<0591:OPISMO>2.0.CO;2.

Katz, R. W., and X. Zheng (1999), Mixture model for overdispersion of precipitation, *J. Clim.*, *12*, 2528–2537, doi:10.1175/1520-0442(1999)012<2528:MMFOOP>2.0.CO;2.

Koutsoyiannis, D. (2004), Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation, *Hydrol. Sci. J.*, *49*, 575–590, doi:10.1623/hysj.49.4.575.54430.

McCullagh, P., and J. A. Nelder (1989), *Generalized Linear Models*, 2nd ed., Chapman and Hall, London.

Richardson, D. W. (1981), Stochastic simulation of daily precipitation, temperature, and solar radiation, *Water Resour. Res.*, *17*, 182–190, doi:10.1029/WR017i001p00182.

Wheater, H. S., R. E. Chandler, C. J. Onof, V. S. Isham, E. Bellone, C. Yang, D. Lekkas, G. Lourmas, and M. L. Segond (2005), Spatial-temporal rainfall modeling for flood risk estimation, *Stochastic Environ. Res. Risk Assess.*, *19*, 403–416, doi:10.1007/s00477-005-0011-8.

Wilks, D. S. (1992), Adapting stochastic weather generation algorithms for climate change studies, *Clim. Change*, *22*, 67–84, doi:10.1007/BF00143344.

Wilks, D. S. (1998), Multisite generalization of a daily stochastic precipitation generation model, *J. Hydrol.*, *210*, 178–191, doi:10.1016/S0022-1694(98)00186-3.

Zheng, X. (1996), Unbiased estimation of autocorrelations of daily meteorological variables, *J. Clim.*, *9*, 2197–2203, doi:10.1175/1520-0442(1996)009<2197:UEOAOD>2.0.CO;2.

Zheng, X., 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, doi:10.1016/j.jhydrol.2007.10.061.

Zheng, X., and R. W. Katz (2008b), Simulation of spatial dependence in daily rainfall using multisite generators, *Water Resour. Res.*, *44*, W09403, doi:10.1029/2007WR006399.

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, doi:10.1016/j.jhydrol.2006.12.020.

A. Clark and J. Renwick, National Institute of Water and Atmospheric Research, Private Bag 14901, Wellington, NA 6003, New Zealand

X. Zheng, College of Global Change and Earth System Science, Beijing Normal University, Beijing, 100875, China. (x.zheng@niwa.co.nz)