JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 104, NO. D24, PAGES 31,657–31,669, DECEMBER 27, 1999

A spatiotemporal model for downscaling precipitation occurrence and amounts

Stephen P. Charles and Bryson C. Bates
Commonwealth Scientific and Industrial Research Organisation (CSIRO) Land and Water, Wembley, Australia

James P. Hughes
Department of Biostatistics, University of Washington, Seattle

Abstract. A stochastic model that relates synoptic atmospheric data to daily precipitation at a network of gages is presented. The model extends the nonhomogeneous hidden Markov model (NHMM) of Hughes et al. by incorporating precipitation amounts. The NHMM assumes that multisite, daily precipitation occurrence patterns are driven by a finite number of unobserved weather states that evolve temporally according to a first-order Markov chain. The state transition probabilities are a function of observed or modeled synoptic scale atmospheric variables such as mean sea level pressure. For each weather state we evaluate the joint distribution of daily precipitation amounts at n sites through the specification of n conditional distributions. The conditional distributions consist of regressions of transformed amounts at a given site on precipitation occurrence at neighboring sites within a set radius. Results for a network of 30 daily precipitation gages and historical atmospheric circulation data in southwestern Australia indicate that the extended NHMM accurately simulates the wet-day probabilities, survival curves for dry- and wet-spell lengths, daily precipitation amount distributions at each site, and intersite correlations for daily precipitation amounts over the 15 year period from 1978 to 1992.

1. Introduction

The need for improved quantitative precipitation forecasts, and realistic assessments of the regional impacts of natural climate variability and possible climate change due to the enhanced greenhouse effect, has generated increased interest in regional climate simulation. Although current general circulation models (GCMs) and limited area models (LAMs) perform reasonably well in simulating synoptic scale atmospheric fields, they tend to overestimate the frequency and underestimate the intensity of daily precipitation and thus fail to reproduce the statistics of historical records at local scales [e.g., *Mearns et al.*, 1995; *Walsh and McGregor*, 1995, 1997; *Bates et al.*, 1998].

The above limitations have led to the development of statistical downscaling techniques to derive sub-grid-scale weather from the coarse spatial resolution atmospheric data available from GCMs and LAMs [*Hewitson and Crane*, 1996]. Early techniques classified large-scale atmospheric circulation patterns and then modeled the daily precipitation process through multivariate probability distributions conditional on the derived patterns [e.g., *Bardossy and Plate*, 1991, 1992; *Bogardi et al.*, 1993; *Matyasovszky et al.*, 1993a, b; *Bartholy et al.*, 1995]. These schemes produce weather patterns that are defined independently of precipitation, and they have had only limited success in reproducing wet and dry spell length statistics [*Hay et al.*, 1991; *Wilson et al.*, 1991, 1992; *Zorita et al.*, 1995]. Recent approaches include regressions on continuous atmospheric circulation indices, geographic location and topographical variables [e.g., *Enke and Spekat*, 1997; *Huth*, 1997; *Kidson and Thompson*, 1998; *Kilsby et al.*, 1998; *Wilby et al.*, 1998], and neural networks [e.g., *Crane and Hewitson*, 1998]. However, the success of these methods in simulating the historical process is not always high. For example, *Huth* [1997] and *Enke and Spekat* [1997] were able to explain only 20% of the at-site variance of historical daily precipitation.

The downscaling method described here uses the nonhomogeneous hidden Markov model (NHMM) of *Hughes et al.* [1999] to simulate precipitation occurrence and multiple linear regression to simulate precipitation amounts in southwestern Australia. The NHMM determines the most distinct patterns in a daily multisite precipitation occurrence record rather than patterns in atmospheric circulation. These patterns are then defined as conditionally dependent on a set of atmospheric predictor variables. In this way, the NHMM captures much of the spatial and temporal variability of the precipitation occurrence process. The joint distribution of daily precipitation amounts at n sites is evaluated through the specification of n conditional distributions. The conditional distributions consist of regressions of transformed amounts at a given site on precipitation occurrence at neighboring sites within a set radius. An automatic variable selection procedure is used to identify the key neighboring sites.

The paper is divided into seven sections. Section 2 describes the study area and the historical precipitation and atmospheric data used, while section 3 details the NHMM and the multiple linear regression and variable selection procedures used. Sections 4 and 5 present the approach and the results of the analysis, respectively. Section 6 contains a general discussion of the issues raised in the paper, with final conclusions presented in section 7.

Paper number 1999JD900119.
0148-0227/99/1999JD900119$09.00

2. Study Area and Data

The study area, the southwestern Australia (SWA) region, extends approximately from 30° to 35° south and from 115° to 120° east (Figure 1). SWA is bounded by the Indian Ocean to the west and the Southern Ocean to the south. The region experiences a "mediterranean" climate with abundant winter rains, which are nearly double that of any similarly exposed location in any other continent, and intense summer drought. Eighty percent of annual precipitation falls in the period from May to October, and most of the winter rains come from low-pressure frontal systems [*Gentilli*, 1972; *Wright*, 1974]. Thus we divided the year into "winter" (May–October) and "summer" (November–April) seasons. A full set of winter results are presented here.

Atmospheric data on a Lambert conformal grid were obtained from the Commonwealth Bureau of Meteorology, Australia, for the period from 1978 to 1992. The data at 1100 UT were chosen on the basis that they are close to the midpoint of the daily precipitation recording period for SWA (2400–0100 UT). The available variables included mean sea level pressure (mslp, hPa) and, at the 850 and 500 hPa levels, geopotential height (GPH, m), air temperature (T, °K), dew point temperature (T_d, °K), and meridional and zonal wind components (knots). *Bates et al.* [1998, Table 1] derived 24 variables from this data set. These included the raw variables listed above, the north-south and east-west gradients of the raw variables, and lagged raw variables. Here we add the dew point temperature depression at 850 hPa:

$$DT_d^{850} = T^{850} - T_d^{850} \quad (1)$$

where T^{850} and T_d^{850} denote the air and dew point temperature at 850 hPa, respectively. DT_d^{850} is a measure of how close the atmosphere at 850 hPa is to saturation with water vapor. The data for the 25 atmospheric variables were interpolated to a rectangular 3.75° longitudinal by 2.25° latitudinal grid (Figure 1).

Daily precipitation data for 30 sites were obtained from the same data source for the same period. The locations of the sites are shown in Figure 1. These sites were chosen because they had no missing records over this period. The 0, 25, 50, 75, and 100% quantiles for the 30 site mean annual precipitation are 309 (at site 20), 371, 514, 824, and 1280 mm (at site 12) and for elevation are 2 (at site 7), 102, 252, 293, and 353 m (at site 17). These ranges are representative of conditions in SWA. There is a 300 m escarpment arising from the coastal plain along the line of sites 5, 4, 12, 15, 11, and 14 with rain shadow effects eastward.

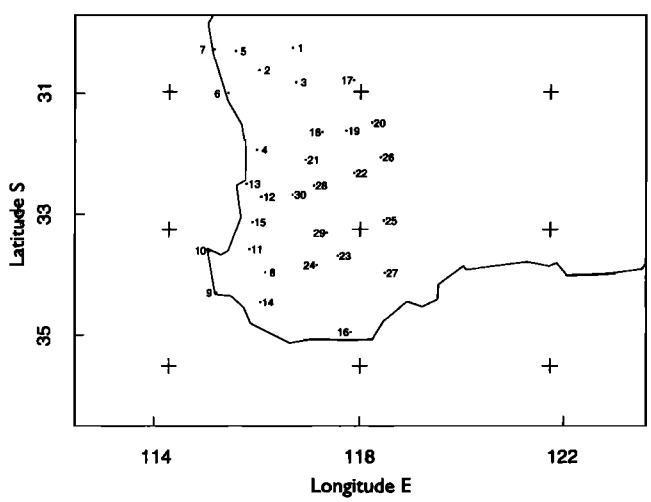

Figure 1. Map of study area showing site locations. Pluses denote grid points for atmospheric circulation data.

3. Model Description

3.1. Precipitation Occurrence

A detailed description and split-sample validation of the NHMM is given by *Hughes et al.* [1999] and thus only a brief description will be given here. The NHMM relates M synoptic scale, atmospheric circulation variables through a finite number of "hidden" (unobserved) weather states (N) to multisite, daily precipitation occurrence data. The atmospheric circulation data may include raw variables such as mslp or derived variables such as north-south mslp gradient. Unlike other downscaling techniques based on classification schemes, the weather states are not defined a priori.

The NHMM defines the stochastic conditional relationships that link the multisite, daily precipitation occurrence patterns to the discrete set of weather states. A first-order Markov process defines the daily transitions from weather state to weather state, with its transition probabilities conditional on a set of atmospheric circulation predictors. Throughout this paper, vectors and matrices will be written in boldface and all vectors are row vectors. Also, lowercase letters will be used to indicate the actual value of a random variable. Let $\mathbf{R}_t = \{R_t^{(1)}, \cdots, R_t^{(n)}\}$ denote the precipitation occurrence pattern at n sites on day t with observed values $\mathbf{r} = \{r^{(1)}, \cdots, r^{(n)}\}$ where $r^{(i)} = 1$ if precipitation greater than or equal to 0.3 mm occurs at site i and 0 otherwise, S_t denote the unobserved weather state at day t, and $\mathbf{X}_t$ denote a summary of historical or modeled atmospheric circulation data at day t, for $t = 1, \cdots, T$.

The NHMM makes two assumptions:

$$P(\mathbf{R}_t|S_1^T, \mathbf{R}_1^{t-1}, \mathbf{X}_1^T) = P(\mathbf{R}_t|S_t) \quad (2)$$

Table 1. Comparison of Conditional Independence and Spatial Models

NHMM*	Number of Hidden Parameters†	Number of Output Parameters‡	BIC
3 ~ .	6	90	61484
4 ~ .	12	120	59110
4 ~ 1, 4	36	120	57168
5 ~ .	20	150	57067
5 ~ 1, 4	60	150	54954
6 ~ .	30	180	56356
6 ~ 1, 4	90	180	54249
6 ~ 1, 3, 4	120	180	54365
6 ~ 1, 4, 8	120	180	54195
6 ~ 1, 4, 8§	120	192	50450
6 ~ 1, 4, 25	120	180	54217
6 ~ 1, 4, 25§	120	192	50178
7 ~ .	42	210	55731
7 ~ 1, 4	126	210	53836
7 ~ 1, 4, 8	168	210	53869

*1 is mslp, 3 is T_d^{850}, 4 is north-south mslp gradient, 8 is east-west GPH gradient at 850 hPa, 25 is DT_d^{850}.

†$(M + 1)[N(N - 1)]$.

‡nN for conditional independence model; $nN + 2N$ for spatial model.

§Spatial model.

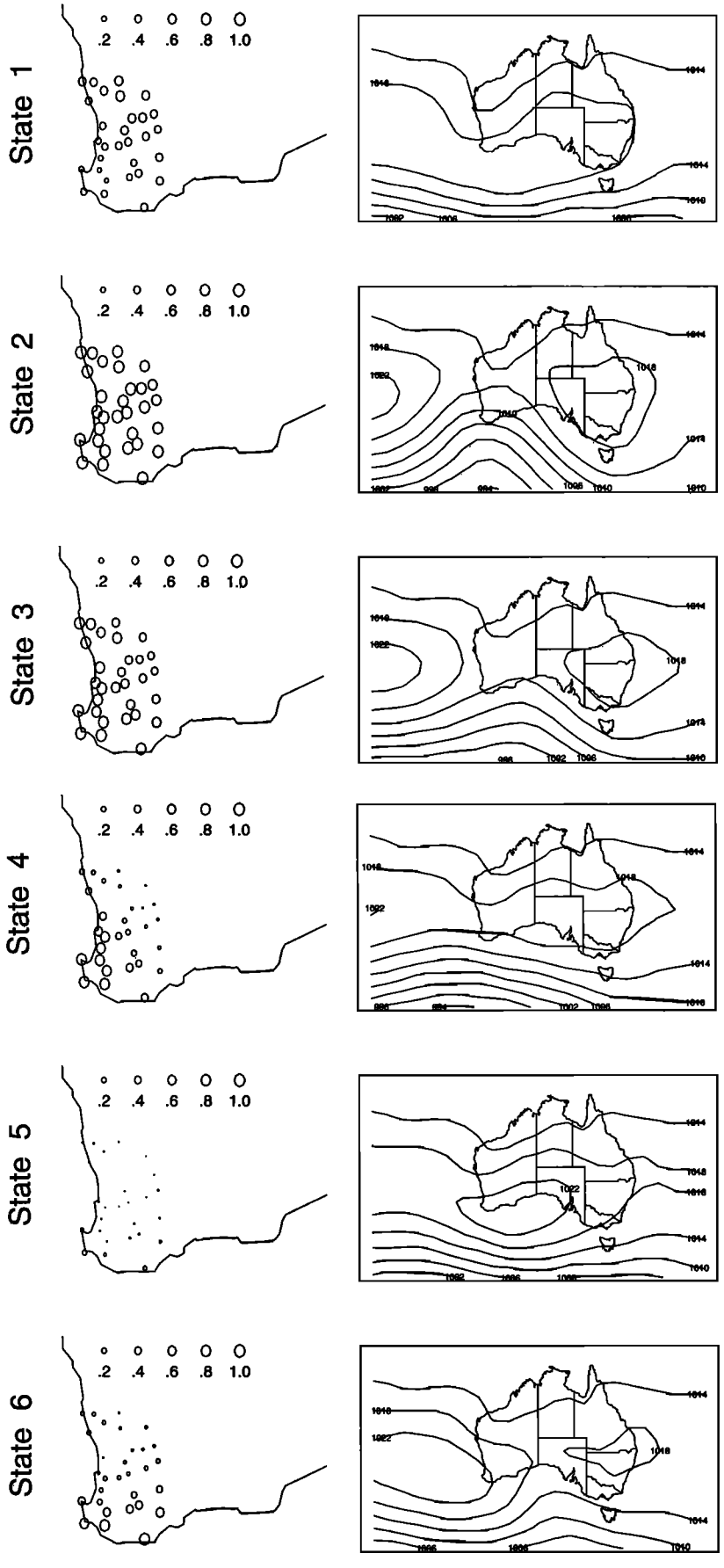

Figure 2. Precipitation occurrence patterns and mslp averaged over all days classified under each weather state for the 6 ~ 1, 4, 25 spatial model.

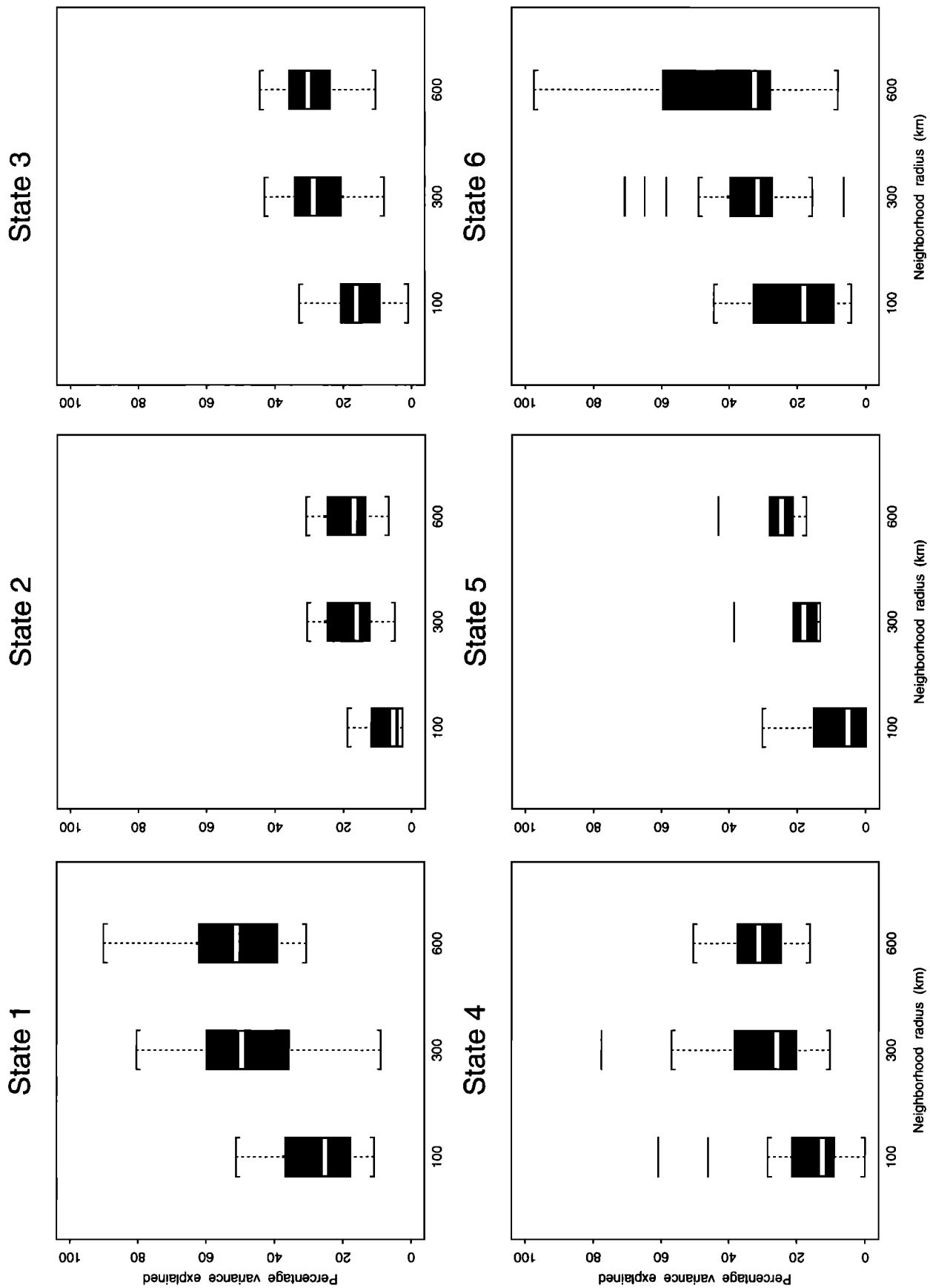

Figure 3. Percentage variance explained by at-site, daily precipitation amount models for $\delta = 100, 300$, and 600. Edges of boxes mark upper and lower quartiles, and horizontal white bar depicts the median.

$$P(S_t|S_1^{t-1}, \mathbf{X}_1^T) = P(S_t|S_{t-1}, \mathbf{X}_t) \quad (3)$$

where the notation $\mathbf{X}_1^T$ means all values of $\mathbf{X}_t$ from day 1 to day T (similarly for S_1^T). That is, (2) states that the n site precipitation occurrence pattern on day t ($\mathbf{R}_t$) is conditional only on the weather state of day t (S_t); and (3) states that this weather state is conditional on both the weather state of the preceding day (S_{t-1}) and the values of the atmospheric data ($\mathbf{X}_t$) on day t. A particular NHMM is defined by parameterizations of the precipitation occurrence probability distribution $P(\mathbf{R}_t|S_t)$ and the weather state transition matrix $P(S_t|S_{t-1}, \mathbf{X}_t)$.

The autologistic model [*Besag*, 1974; *Cressie*, 1991] for multivariate binary data is used to parameterize $P(\mathbf{R}_t|S_t)$:

$$P(\mathbf{R}_t = \mathbf{r}|S_t = s) \propto \exp\left(\sum_{i=1}^{n} \alpha_{si} r^{(i)} + \sum_{j<i} \beta_{sij} r^{(i)} r^{(j)}\right) \quad (4)$$

where both α_{si} and β_{sij} must be finite. To reduce the number of model parameters, β_{sij} can be modeled as a function of the distance and direction between sites i and j:

$$\beta_{sij} = b_{0s} + b_{1s} \log\left[d_{ij}\sqrt{\cos(\phi_s + h_{ij})^2 + \sin(\phi_s + h_{ij})^2/e_s}\right] \quad (5)$$

where d_{ij} and h_{ij} are the distance and direction between sites i and j, respectively. For computational reasons the parameters ϕ_s and e_s were fixed at values determined by a prior nonlinear least squares regression. See *Hughes et al.* [1999] for more details. The b parameters were determined by the maximum likelihood estimation procedure described below. Equations (4) and (5) define a "spatial model" for multisite rainfall occurrence (conditional on the weather state). When $\beta_{sij} = 0$ for all i, j, and s,

$$P(\mathbf{R}_t = \mathbf{r}|S_t = s) = \prod_{i=1}^{n} p_{si}^{r^{(i)}}(1 - p_{si})^{1-r^{(i)}}, \quad (6)$$

where $p_{si} = \exp(\alpha_{si})/[1 + \exp(\alpha_{si})]$ is the probability of precipitation at site i in weather state s. Although the $R_t^{(i)}$ in (6) are assumed to be spatially independent, conditional on the weather state, they are unconditionally correlated due to the influence of the common weather state. Thus (6) is referred to as a "conditional independence model."

Table 2. Summary of Weather States

State	Description of Synoptic Situation	Number of Days, %	Rainfall, %*
1	low-pressure trough associated with a midlevel moisture source or spring thunderstorms	6.5	7.8
2	ridging low-pressure system, frontal westerly winds	19.8	65.9
3	coastal convergence, light winds	14.1	16.4
4	postfrontal, southwesterly winds	20.2	6.9
5	high-pressure system centered east of SWA, east to northeast winds	27.8	0.5
6	ridging high-pressure system, moist southerly winds	11.6	2.5

*Percentage of mean winter rainfall (average of the 30 sites shown in Figure 1).

Suppose $\mathbf{X}_t$ is multivariate normal. Then the state transition probability matrix can be parameterized as

$$P(S_t = j|S_{t-1} = i, \mathbf{X}_t) = \frac{P(S_t = j|S_{t-1} = i)P(\mathbf{X}_t|S_{t-1} = i, S_t = j)}{P(\mathbf{X}_t|S_{t-1} = i)}$$

$$\propto \gamma_{ij} \exp\left[-\tfrac{1}{2}(\mathbf{X}_t - \mu_{ij})\mathbf{V}^{-1}(\mathbf{X}_t - \mu_{ij})'\right] \quad (7)$$

where μ_{ij} is the mean of $\mathbf{X}_t$, conditional on S_{t-1} and S_t, and $\mathbf{V}$ is the corresponding covariance matrix. The γ parameters contain information about the baseline state transition probability matrix corresponding to the transition matrix of a standard hidden Markov model (HMM), with the exponential term quantifying the effect of the atmospheric circulation data on the baseline transition probabilities.

Maximum likelihood estimates of the NHMM parameters are obtained using a modified EM algorithm [*Baum et al.*, 1970; *Dempster et al.*, 1977] and the method of Monte Carlo maximum likelihood [*Geyer and Thompson*, 1992]. See *Hughes et al.* [1999] for details. The selection of a NHMM involves sequential fitting of several NHMMs with an increasing number of weather states and atmospheric variables. The fit is evaluated in terms of the physical realism and distinctness of the identified weather states as well as the Bayes information criterion:

$$\text{BIC} = -2L + p \log(T) \quad (8)$$

where L is the log-likelihood and p is the number of model parameters. The objective is to select an NHMM that minimizes the BIC, thus identifying a relatively parsimonious model that fits the data well. Throughout this paper, we use the compact notation $N \sim .$ to represent a standard N-state HMM and $N \sim i.j$ to represent an N-state NHMM with $M = 2$ observed covariates that correspond to the ith and jth atmospheric variables. The most likely weather state sequence is obtained from the selected NHMM using the Viterbi algorithm [*Forney*, 1978]. This permits the assignment of each day to its respective state [*Hughes et al.*, 1999].

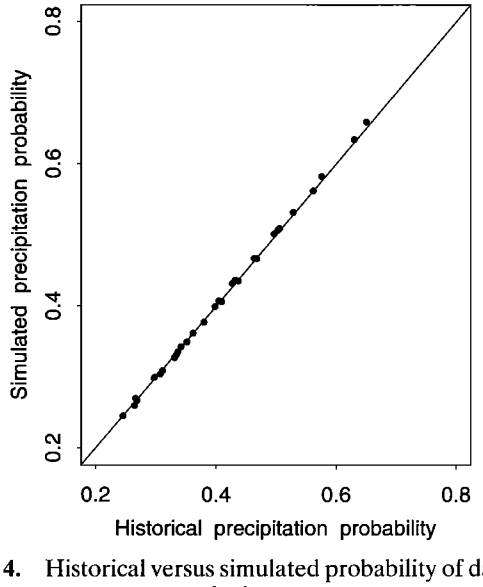

Figure 4. Historical versus simulated probability of daily precipitation occurrence at each site.

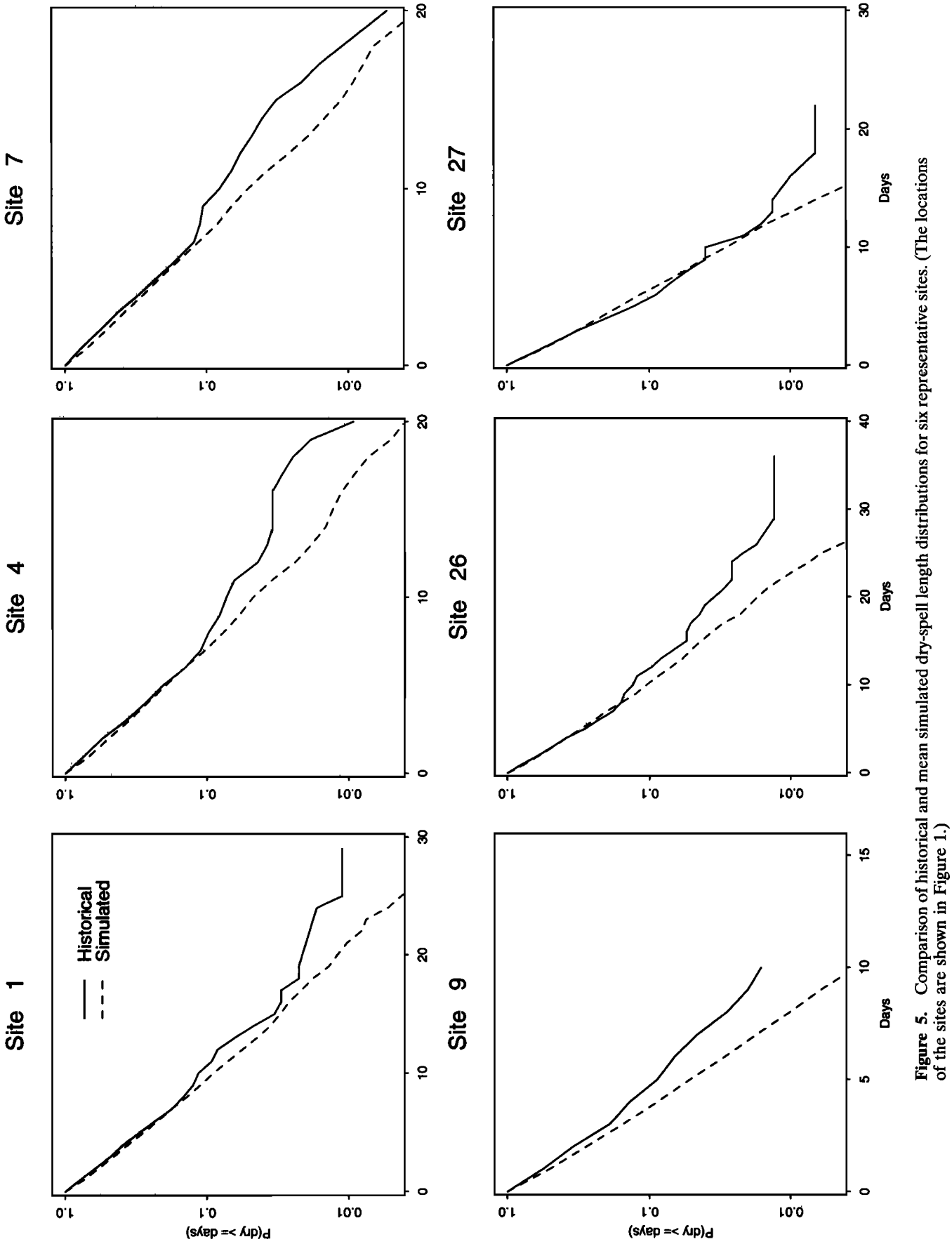

Figure 5. Comparison of historical and mean simulated dry-spell length distributions for six representative sites. (The locations of the sites are shown in Figure 1.)

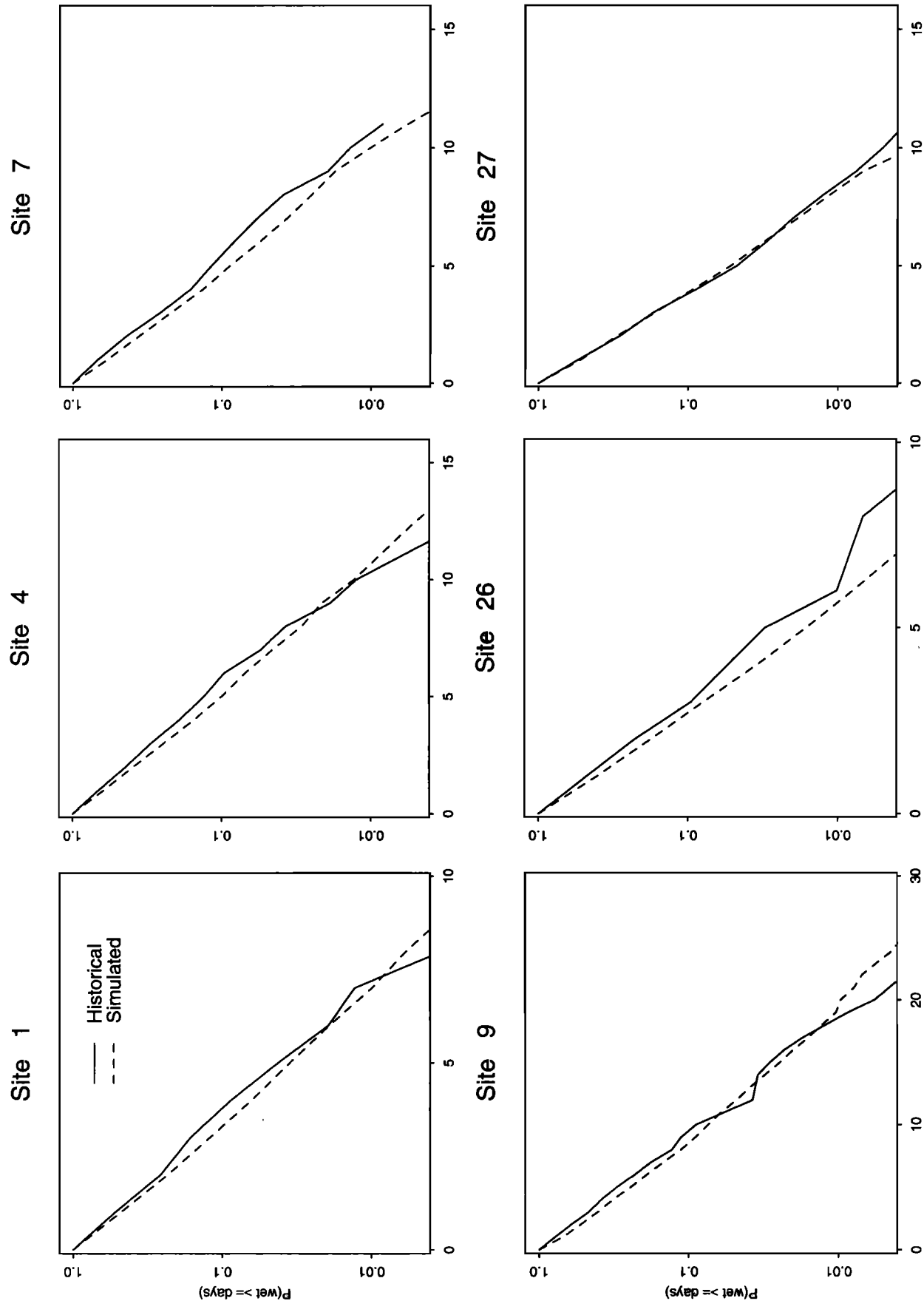

Figure 6. Comparison of historical and mean simulated wet-spell length distributions for six representative sites. (The locations of the sites are shown in Figure 1.)

3.2. Precipitation Amounts

The joint distribution of daily precipitation amounts at n sites is evaluated through the specification of n conditional distributions for each weather state ($s = 1, \cdots, N$). The conditional distributions consist of regressions of inverse normal transformed amounts at a given site on precipitation occurrence at neighboring sites within a given radius (δ, km). An automatic variable selection procedure is used to identify the key neighboring sites. Thus the precipitation amounts model can be expressed as

$$z_s^{(t)} = \theta_{0s}^{(t)} + \sum_{k \in n_i(\delta)} \theta_{ks}^{(t)} r^{(k)} + \varepsilon_s^{(t)}, \qquad i = 1, \cdots, n \tag{9}$$

where the θ_s are regression parameters, $n_i(\delta)$ denotes the set of indices of the key neighboring sites for site i, $\varepsilon_s^{(t)}$ is an error term modeled stochastically by assuming $\varepsilon_s^{(t)} \sim N(0, \sigma_s^2(i))$, and

$$z_s^{(t)} = \Phi^{-1}(\hat{F}(y_s^{(t)})) \tag{10}$$

in which Φ denotes the normal cumulative distribution function and $\hat{F}(y_s^{(t)})$ is the empirical distribution function of $y_s^{(t)}$, the precipitation amounts on days with $r^{(t)} = 1$.

Radii of 0, 100, 300, and 600 km were investigated in this study. Setting $\delta = 0$ produces a "conditional independence amounts model." The greatest distance between any two sites is 576 km. Thus setting $\delta = 600$, initially puts all $n - 1$ surrounding sites in the neighborhood of site i.

4. Approach

Hughes et al. [1999] report the results of a successful split-sample test of the NHMM using the historical atmospheric and precipitation data set described in section 2, and this test will not be repeated here. The atmospheric variables used did not exhibit any departures from normality. We applied a suite of NHMMs to the entire (15 years) data set and used the BIC from (8) and the distinctness and realism of the weather states as guides to appropriate values for M and N. Preference was given to NHMMs using a greater number of atmospheric variables (and hence more atmospheric information) rather than a greater number of weather states.

Least squares estimates of parameters in the at-site precipitation amount models from (9) were obtained for each site and weather state combination. For $0 < \delta \leq 600$, an automated backward elimination procedure was used to determine the adjacent sites that provided useful information about at-site precipitation amounts. The procedure used tends to be cautious in rejecting and generous in accepting terms [*Venables and Ripley*, 1994].

Ten 15 year sequences of daily precipitation occurrence patterns were generated from the selected NHMM, conditionally on the 15 year sequence of historical atmospheric data, using the procedure described by *Hughes and Guttorp* [1994]. At-site daily precipitation amounts were then simulated conditionally on the weather state and daily precipitation occurrences at neighboring sites. Summary statistics for the simulated daily precipitation series were then compared with their historical counterparts.

5. Results

5.1. Parameter Estimation

Table 1 reports a subset of the results of fitting the conditional independence model (6) and the spatial model (equations (4) and (5)) to the data set described above. Consider the results for the conditional independence model. The BIC criterion (8) suggests that the use of the atmospheric circulation data results in a marked improvement in the model fit. Although the BIC decreases with increasing N, the rate of decrease declines sharply at $N = 5$ regardless of the number of covariates used. Also, the use of three covariates rather than two leads to a marginal decrease in the BIC.

Comparison of the precipitation occurrence patterns and corresponding composite mslp fields for the weather states of the six state NHMMs indicated that the states are reasonably distinct (see, for example, Figure 2). Moreover, the six states have a high degree of physical realism in terms of the synoptic situation, their relative frequency, and precipitation amounts (as summarized in Table 2). In contrast, the seven state NHMMs have two weather states that are almost indistinguishable. These two states have precipitation occurrence patterns and composite mslp fields that are very similar to those for state 4 in Figure 2. Thus a six state NHMM is appropriate. Further experimentation with six state, conditional independence models indicated that the $6 \sim 1, 4, 8$ and $6 \sim 1, 4, 25$ NHMMs provided the lowest values of the BIC. (A key to the atmospheric variables nominated herein appears in Table 1.) Finally, comparison of the corresponding spatial models led to the selection of the $6 \sim 1, 4, 25$ spatial model.

Figure 3 shows box plots of the variances explained by the at-site precipitation amounts models (9) on a weather state and neighborhood radius basis. Each box and its associated whiskers represents the range of variances explained for the 30 sites. Overall, there is a marked increase in explained variance when δ is increased from 100 to 300 km and a relatively slight increase when δ is increased from 300 to 600 km. The explained variance is low for state 2 and relatively high for state 1. The comparatively poor result for state 2 suggests that information in addition to daily precipitation occurrence at neighboring sites is required to capture the total variance in at-site precipitation amounts for days in that state.

5.2. Simulated Precipitation Occurrence

Figure 4 compares the mean simulated daily precipitation occurrence probabilities with historical values for the 30 sites.

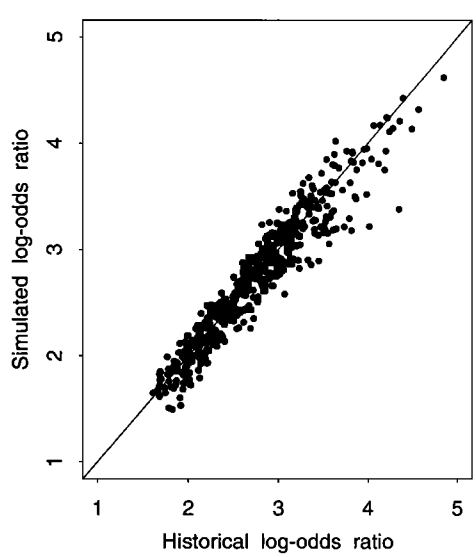

Figure 7. Historical versus mean simulated log-odds ratio.

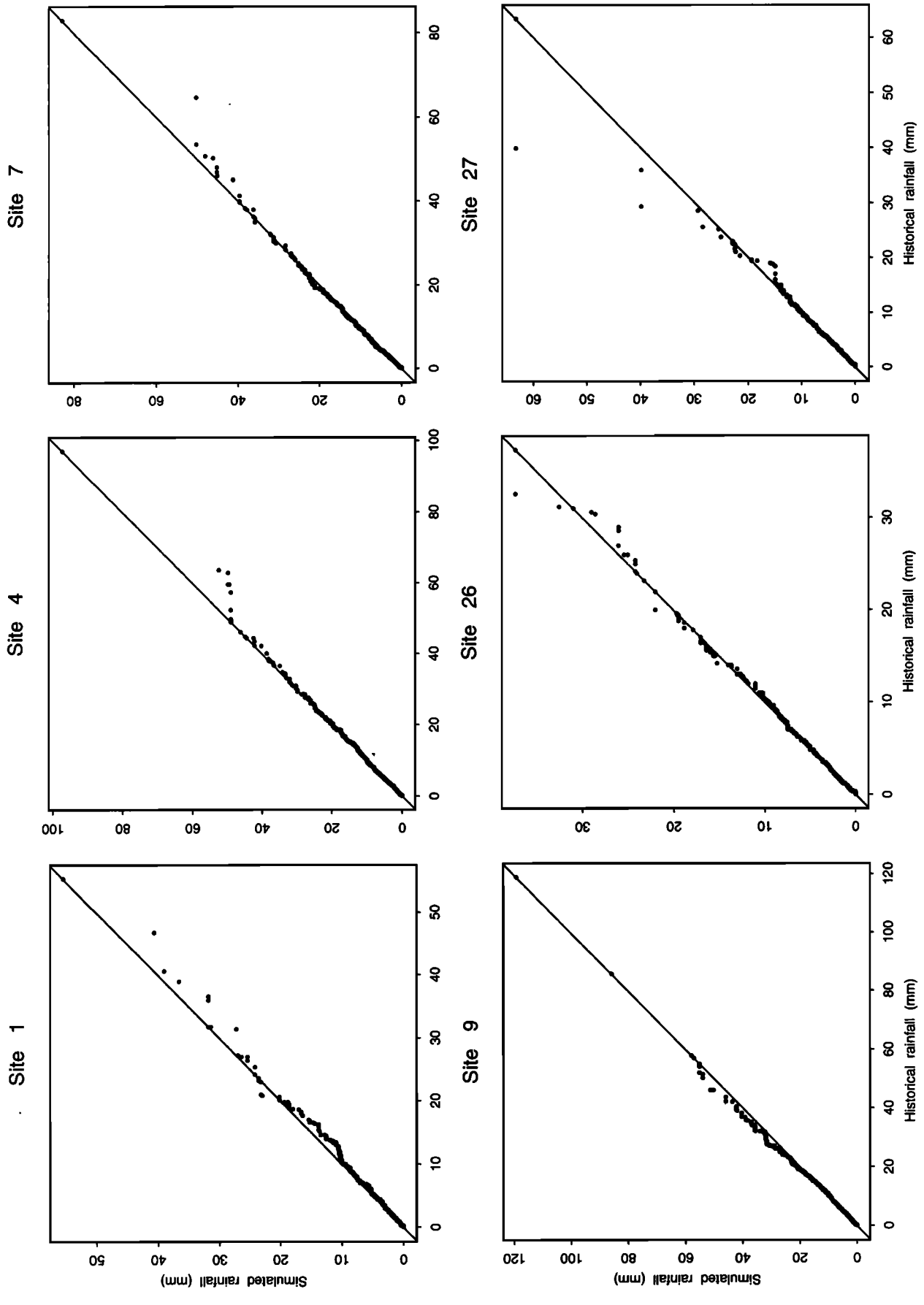

Figure 8. Quantile-quantile plots of historical versus simulated precipitation amounts for six representative sites. (The locations of the sites are shown in Figure 1.)

The simulated precipitation probabilities are close to the observed across all sites.

Figure 5 compares the historical distributions and mean simulated distributions of dry spell lengths at six representative sites spread across SWA. (A dry spell is defined as a sequence of consecutive days during which daily precipitation remains below 0.3 mm.) The mean simulated distributions provide good approximations to the historical distributions, particularly for the short-duration spells that encompass about 90% of dry events. The discrepancies for the remaining 10% of events may be attributable to uncertainty in their probability estimates due to the small sample sizes involved. Also, the tendency to underestimate the lengths of extreme dry spells is exaggerated by the use of a log probability scale.

Figure 6 compares the historical distributions and the mean simulated distributions for wet-spell lengths at the six representative sites. (A wet spell is defined as a sequence of consecutive days during which the daily precipitation equals or exceeds 0.3 mm.) The mean simulated distributions provide good approximations to the historical distributions, particularly for the short-duration spells that encompass ~99% of wet events.

Figure 7 compares the mean simulated log-odds ratio for precipitation occurrence with the observed. (The log-odds ratio is a measure of association for binary data, which is analogous to spatial correlation of continuous variables [*Kotz and Johnson*, 1985].) The selected NHMM clearly captures the spatial correlation between sites, which is induced by the weather states and other effects such as local orography and aspect.

5.3. Simulated Precipitation Amounts

Figure 8 compares the distributions of the historical and simulated precipitation amounts (for $\delta = 300$) at the six representative sites, for one of the 10 generated sequences chosen at random. Some discrepancies in the upper tails of the distributions are evident due to the deficiencies in simulating amounts for days in state 2. However, the model does well in reproducing all but the largest quantiles of at-site precipitation.

Figure 9 compares historical versus simulated Spearman rank intersite correlations. There is a bias of −0.03 in the simulated correlations across the range of historical correlations. Comparison with the corresponding plot for the $\delta = 600$ case revealed no noticeable reduction in the bias. For the $\delta = 0$ (independence amounts model) case, however, the bias was −0.06 for historical correlations between 0.4 and 0.75, increasing to −0.08 for historical correlations greater than 0.75. The nine other simulated sequences produced similar results.

The major source of the bias was identified by examining the correlation plots on a state-by-state basis. The plot for the above generated sequence with $\delta = 300$ is illustrated in Figure 10. Clearly the reproduction of intersite correlations for state 2 is poor, and this probably reflects the low percentage variance explained by the at-site precipitation amounts model for this state (Figure 3). (The simulated intersite correlations for state 2 were no greater than 0.2 for the $\delta = 0$ case.) There is also a large degree of scatter in the intersite correlations plot for state 5 (the "dry" weather state). This is to be expected given that precipitation will be highly localized for this state. However, the scatter is not serious given that the precipitation amounts associated with state 5 are small relative to other states (Table 2). The reproduction of intersite correlations is good for states 1, 3, 4, and 6. The correlation plots for $\delta = 600$ showed no noticeable improvement.

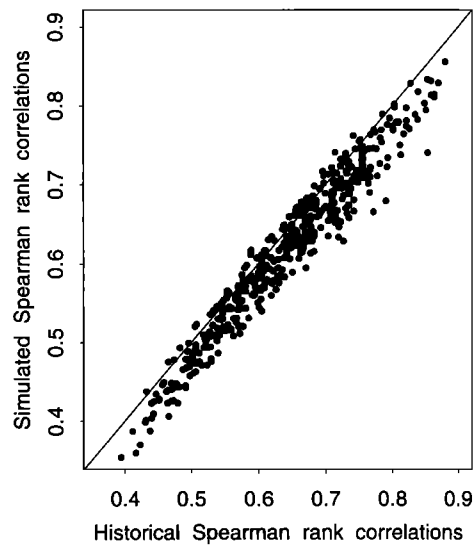

Figure 9. Historical versus simulated Spearman rank intersite correlations ($\delta = 300$).

6. Discussion

The NHMM examined here does well in reproducing the frequency characteristics of wet and dry spell lengths. In the past this has proven to be the most difficult characteristic of precipitation to reproduce using downscaling models. Although good agreement was observed between the historical and the simulated at-site quantile functions for precipitation amounts (Figure 8), deficiencies in (9) are apparent, particularly with regard to intersite correlations for state 2 (Figure 10). In state 2, all sites are likely to be wet so there is little information in the precipitation occurrence pattern for this state. A similar situation applies for state 5 as all sites are likely to be dry. While the practical importance of these deficiencies remains an open question, a more complex approach to the simulation of precipitation amounts may be warranted for state 2. Future work will examine (1) regression models that consider precipitation amounts at neighboring sites as well as occurrences and interaction terms involving products of two or more occurrences and amount variables, and (2) a new NHMM framework that considers precipitation amounts and occurrences jointly. The latter requires the specification of a mixed discrete-continuous model for $P(\mathbf{R}_t|S_t)$ [*Bellone et al.*, 1998]. Such a framework will also account for any potential errors caused by the misclassification of days to weather states.

The weather states derived from the selected NHMM are analogous to the weather type classifications obtained from traditional synoptic climatology [e.g., *Yarnal*, 1993]. However, the weather states are defined on the basis of the most distinct multisite precipitation occurrence patterns conditional on the atmospheric variables that maximize their predictability. The ability to classify days into states that are distinct in terms of precipitation as well as synoptic situation means that the realism of the states is directly interpretable in terms of regional hydroclimatology. Consider state 2 (high precipitation probabilities at all sites). This precipitation occurrence pattern can be expected when a strong frontal system traverses SWA. The

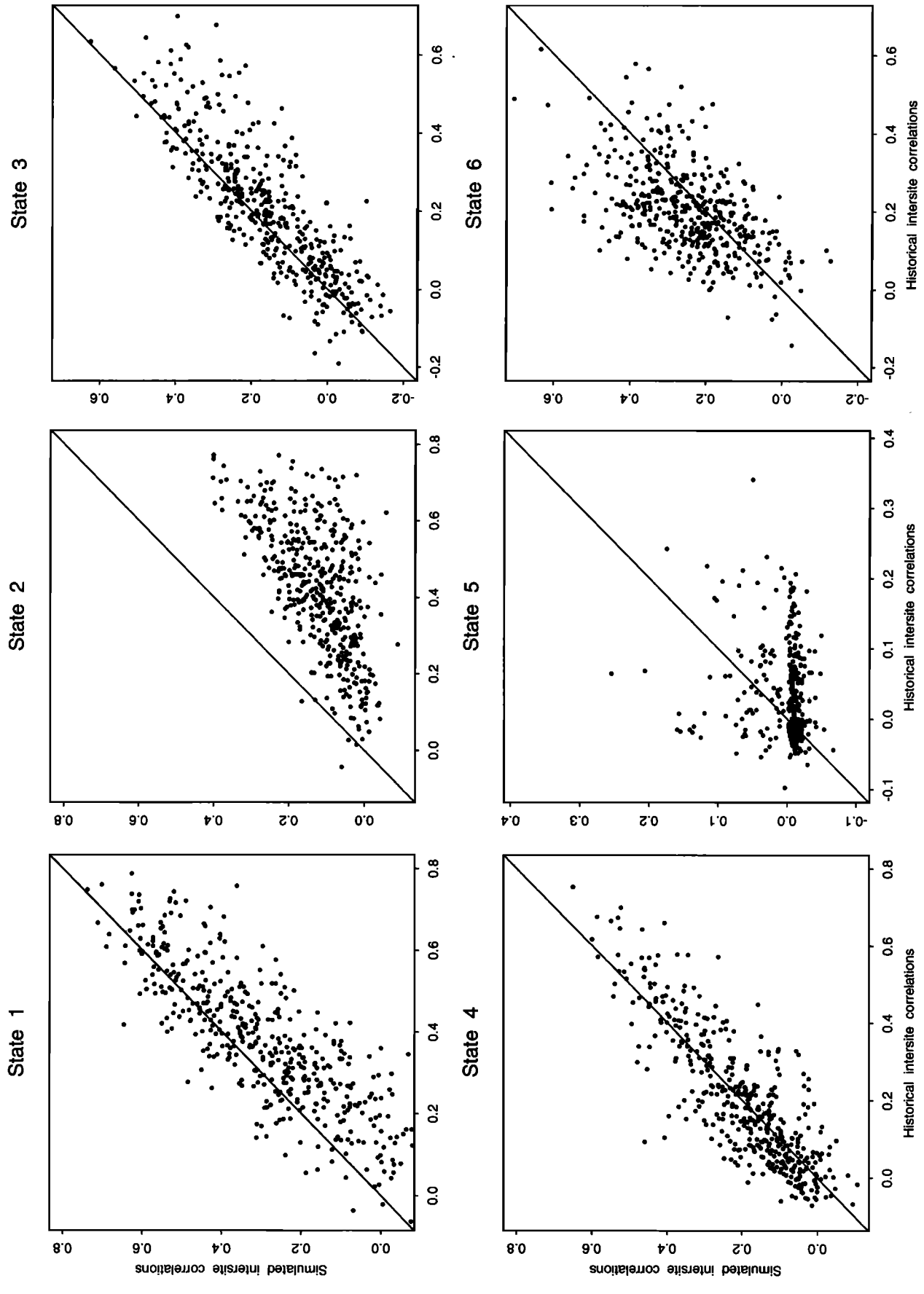

Figure 10. Historical versus simulated Spearman rank intersite correlations ($\delta = 300$) according to weather state.

corresponding synoptic situation in the fitted NHMM is indicative of such a system, showing a ridging low-pressure system dominating the region (Figure 2). In comparison, state 3 (widespread precipitation occurrence across SWA but high precipitation probabilities confined to the coast) is indicative of a weak frontal system associated with a low-pressure system located south of the region. Thus frontal rains are less likely to spread into the hinterland. Similar interpretations can be made for the four remaining states.

It is noteworthy that the 6 ~ 1, 4, 25 NHMM is preferable to the 6 ~ 1, 3, 4 NHMM according to the BIC (Table 1). Although variable 3 (T_d^{850}) is a proxy for the absolute moisture content of the lower atmosphere, it can be argued that the probability of precipitation occurrence would be better related to a measure of saturation, and so probability of cloud formation, such as variable 25 (DT_d^{850}). This result has important implications for climate change research. *Wilby and Wigley* [1997] found that the use of airflow variables (such as mslp) alone in downscaling schemes may not lead to realistic precipitation changes under projected future climates. They recommended the incorporation of measures of atmospheric moisture.

7. Conclusions

Our downscaling study was motivated by the need for improved quantitative precipitation forecasts, and realistic assessments of the regional impacts of natural climate variability and possible climate change due to the enhanced greenhouse effect, at local and regional scales. We restricted our attention to a dense network of 30 precipitation gages in the southwest region of Western Australia. Our main findings may be summarized as follows:

1. The nonhomogeneous hidden Markov model (NHMM) of *Hughes et al.* [1999] reproduces the key statistics of daily precipitation occurrence in the network.

2. Regression models conditioned on the weather states identified by the NHMM provide a simple and parsimonious method of simulating daily precipitation amounts in the network. The performance of the conditional independence amounts model confirms that the weather states capture much of the spatial and temporal variability of precipitation and that the regression models incorporate local spatial dependencies in precipitation amounts. There is scope for refinement of precipitation amount simulation, and this will be the subject of future research efforts.

3. Given that the extended NHMM is successful in downscaling historical atmospheric circulation data to local and regional scales, it may be useful for identifying weather types for traditional synoptic climatology, downscaling dynamical climate model simulations, and diagnosing problems with scale relationships in climate models. However, a fundamental caveat of all statistical downscaling models is that they are not necessarily valid beyond the range of the atmospheric and precipitation data used to fit the model.

4. Our findings apply to one geographical region and season where precipitation is generated by frontal systems. The generality of these findings needs to be tested in other regions where a mix of generation mechanisms is responsible for precipitation.

Acknowledgments. The atmospheric and precipitation data were provided by the National Climate Centre, Commonwealth Bureau of Meteorology, Australia. This work contributes to the CSIRO Climate Change Research Program and is part funded through the Australian Government's National Greenhouse Research Program. SPC and BCB are grateful for their visiting scholarships held at the National Research Center for Statistics and the Environment, University of Washington, during which the work on the precipitation amounts model was carried out. We also appreciate the constructive comments made by the anonymous reviewers.

References

Bardossy, A., and E. J. Plate, Modeling daily rainfall using a semi-Markov representation of circulation pattern occurrence, *J. Hydrol.*, *122*, 33–47, 1991.

Bardossy, A., and E. J. Plate, Space-time model for daily rainfall using atmospheric circulation patterns, *Water Resour. Res.*, *28*(5), 1247–1259, 1992.

Bartholy, J., I. Bogardi, and I. Matyasovszky, Effect of climate change on regional precipitation in Lake Balaton watershed, *Theor. Appl. Climatol.*, *51*, 237–250, 1995.

Bates, B. C., S. P. Charles, and J. P. Hughes, Stochastic downscaling of numerical climate model simulations, *Environ. Modell. Software*, *13*(3–4), 325–331, 1998.

Baum, L. E., T. Petrie, G. Soules, and N. Weiss, A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, *Ann. Math. Stat.*, *41*, 164–171, 1970.

Bellone, E., P. Guttorp, and J. P. Hughes, A stochastic model for precipitation amounts at multiple stations, paper presented at Sixth International Conference on Precipitation: Predictability of Rainfall at the Various Scales, Natl. Sci. Found. Mauna Lani Bay, Hawaii, June 29–July 1, 1998.

Besag, J. E., Spatial interaction and the statistical analysis of lattice systems, *J. R. Stat. Soc., Ser. B*, *36*, 192–236, 1974.

Bogardi, I., I. Matyasovszky, A. Bardossy, and L. Duckstein, Application of a space-time stochastic model for daily precipitation using atmospheric circulation patterns, *J. Geophys. Res.*, *98*, 16,653–16,667, 1993.

Crane, R. G., and B. C. Hewitson, Doubled CO_2 precipitation changes for the Susquehanna basin: Down-scaling from the GENESIS general circulation model, *Int. J. Climatol.*, *18*, 65–76, 1998.

Cressie, N. A. C., *Statistics for Spatial Data*, 900 pp., John Wiley, New York, 1991.

Dempster, A. P., N. M. Laird, and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm (with discussion), *J. R. Stat. Soc., Ser. B*, *39*, 1–38, 1977.

Enke, W., and A. Spekat, Downscaling climate model outputs into local and regional weather elements by classification and regression, *Clim. Res.*, *8*, 195–207, 1997.

Forney, G. D., Jr., The Viterbi algorithm, *Proc. IEEE*, *61*, 268–278, 1978.

Gentilli, J., *Australian Climate Patterns*, 285 pp., Nelson, Melbourne, 1972.

Geyer, C. J., and E. A. Thompson, Constrained Monte Carlo maximum likelihood for dependent data, *J. R. Stat. Soc., Ser. B*, *54*, 657–699, 1992.

Hay, L. E., G. J. McCabe, D. M. Wolock, and M. A. Ayers, Simulation of precipitation by weather type analysis, *Water Resour. Res.*, *27*(4), 493–501, 1991.

Hewitson, B. C., and R. G. Crane, Climate downscaling: Techniques and application, *Clim. Res.*, *7*, 85–95, 1996.

Hughes, J. P., and P. Guttorp, Incorporating spatial dependence and atmospheric data in a model of precipitation, *J. Appl Meteorol.*, *33*(12), 1503–1515, 1994.

Hughes, J. P., P. Guttorp, and S. P. Charles, A non-homogeneous hidden Markov model for precipitation occurrence, *Appl. Stat.*, *48*(1), 15–30, 1999.

Huth, R., Potential of continental-scale circulation for the determination of local daily surface variables, *Theor. Appl. Climatol.*, *56*, 165–186, 1997.

Kidson, J. W., and C. S. Thompson, A comparison of statistical and model-based downscaling techniques for estimating local climate variations, *J. Clim.*, *11*(4), 735–753, 1998.

Kilsby, C. G., P. S. P. Cowpertwait, P. E. O'Connell, and P. D. Jones, Predicting rainfall statistics in England and Wales using atmospheric circulation variables, *Int. J. Climatol.*, *18*, 523–539, 1998.

Kotz, S., and N. L. Johnson, *Encyclopedia of Statistical Sciences*, vol. 6, 758 pp., John Wiley, New York, 1985.

Matyasovszky, I., I. Bogardi, A. Bardossy, and L. Duckstein, Estimation of local precipitation statistics reflecting climate change, *Water Resour. Res.*, *29*(12), 3955–3968, 1993a.

Matyasovszky, I., I. Bogardi, A. Bardossy, and L. Duckstein, Space-time precipitation reflecting climate change, *Hydrol. Sci. J.*, *38*(6), 539–558, 1993b.

Mearns, L. O., F. Giorgi, L. McDaniel, and C. Shields, Analysis of daily variability of precipitation in a nested regional climate model: Comparison with observations and doubled CO_2 results, *Global Planet. Change*, *10*, 55–78, 1995.

Venables, W. N., and B. D. Ripley, *Modern Applied Statistics With S-Plus*, 462 pp., Springer-Verlag, New York, 1994.

Walsh, K. J. E., and J. L. McGregor, January and July climate simulations over the Australian region using a limited-area model, *J. Clim.*, *8*(10), 2387–2403, 1995.

Walsh, K. J. E., and J. L. McGregor, An assessment of simulations of climate variability over Australia with a limited area model, *Int. J. Climatol.*, *17*, 201–223, 1997.

Wilby, R. L., and T. M. L. Wigley, Downscaling general circulation model output: A review of methods and limitations, *Prog. Phys. Geogr.*, *21*(4), 530–548, 1997.

Wilby, R. L., H. Hassan, and K. Hanaki, Statistical downscaling of hydrometeorological variables using general circulation model output, *J. Hydrol.*, *205*, 1–19, 1998.

Wilson, L. L., D. P. Lettenmaier, and E. F. Wood, Simulation of daily precipitation in the Pacific Northwest using a weather classification scheme, *Surv. Geophys.*, *12*, 127–142, 1991.

Wilson, L. L., D. P. Lettenmaier, and E. Skyllingstad, A hierarchical stochastic model of large-scale atmospheric circulation patterns and multiple station daily precipitation, *J. Geophys. Res.*, *97*, 2791–2809, 1992.

Wright, P. B., Seasonal rainfall in Southwestern Australia and the general circulation, *Mon. Weather Rev.*, *102*, 219–232, 1974.

Yarnal, B., *Synoptic Climatology in Environmental Analysis: A Primer.*, 195 pp., Belhaven, London, 1993.

Zorita, E., J. P. Hughes, D. P. Lettenmaier, and H. Von Storch, Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation, *J. Clim.*, *8*(5), 1023–1042, 1995.

B. C. Bates, and S. P. Charles, CSIRO Land and Water, Private Bag P.O., Wembley, Western Australia 6014, Australia. (stephen.charles@per.clw.csiro.au; bryson.bates@per.clw.csiro.au)

J. P. Hughes, Department of Biostatistics, University of Washington, Seattle, WA 98195. (hughes@diamond.cfas.washington.edu)

(Received October 15, 1998; revised February 18, 1999; accepted February 23, 1999.)