A Nonhomogeneous Hidden Markov Model for Precipitation

James P. Hughes[1]
Peter Guttorp[2]
Stephen P. Charles[3]

December 11, 1996

[1]Dept. of Biostatistics 357232, University of Washington, Seattle, WA 98195 USA

[2]Dept. of Statistics 354322, University of Washington, Seattle, WA 98195 USA

[3]CSIRO Division of Water Resources, Private Bag, P.O. Wembley, Western Australia 6014

Address correspondence to James P. Hughes (email: hughes@biostat.washington.edu)

Abstract

Improvements in computing power, data gathering and our understanding of atmospheric dynamics have lead to the availability of spatially and temporally extensive sets of data on the atmospheric processes that affect precipitation. However, these two processes (atmospheric circulation and precipitation) operate on very different spatial scales. Recently, considerable effort has been devoted to developing "downscaling" models which condition local precipitation on broad-scale atmospheric circulation. In this article, we develop a stochastic model for relating precipitation occurrences at multiple rain gauge stations to atmospheric circulation patterns.

The proposed model is an example of a nonhomogeneous hidden Markov model, and generalizes existing downscaling models in the literature. The model assumes that atmospheric circulation can be classified into a small number of (unobserved) discrete patterns (called "weather states"). The weather states are assumed to follow a Markov chain in which the transition probabilities depend on observable characteristics of the atmosphere (e.g. mean sea-level pressure). Precipitation is assumed to be conditionally temporally, but not spatially, independent given the weather state. An autologistic model for multivariate binary data is used to model rainfall occurrences and capture local spatial dependencies. However, the usual approach to estimation in hidden Markov models — exact likelihood using the EM algorithm — is computationally intractable if there are large numbers of rain gauge stations. Therefore, two alternative estimation procedures are developed which combine (an approximation to) the usual E-step with a modified M-step based on either maximum pseudolikelihood or Monte Carlo maximum likelihood. Both techniques yield models which fit the data well, although the pseudolikelihood is seen to be ill-behaved in certain situations.

This approach is used to model a 15 year sequence of winter data from 30 rain stations in southwestern Australia. The first 10 years of data are used for model development and the remaining 5 years are used for model evaluation. The fitted model is able to accurately reproduce the observed rainfall statistics in the reserved

data, even in the face of a small non-stationary shift in atmospheric circulation (and, consequently, rainfall) between the two periods. The fitted model also provides some useful insights into the processes driving rainfall in this region. We discuss the role that models such as this might play in assessing the impact of climate change.

Keywords: hidden Markov model, climate change, precipitation model, pseudolikelihood, Monte Carlo maximum likelihood, EM algorithm

1 Introduction

Stochastic models for precipitation have long been used to aid in understanding the probabilistic structure of rainfall and for simulation studies. In particular, precipitation simulations are often used as input into hydrologic models of flooding, runoff, water supply, agricultural models of crop growth, and other applications. In the past these models considered only the rainfall process, without reference to the atmospheric processes that drive precipitation. In part, this reflected the absence of good, long-term records of atmospheric circulation. Thus, Gabriel and Neuman (1962) used a Markov chain with homogeneous transition matrix to model daily wet/dry occurrences at a single rain gauge station in Israel. Stern and Coe (1984) extended this model by making the (logits of) transition probabilities a Fourier series to represent seasonal variations. Others developed more mechanistic models. For example, LeCam (1961) described rainfall using a cluster point process whereby cyclonic storms were assumed to contain "bands" (areas of high rainfall intensity) and the bands contained rain cells where precipitation activity occurs. Waymire and Gupta (1981), Kavvas and Delleur (1975, 1981) and others expanded on the point process approach.

These models have several limitations, however. In developing hydrologic models researchers use information on temperature, solar radiation and other climatic factors in addition to precipitation. Ideally, the precipitation model should produce simulations which are consistent with these other inputs into the hydrologic model. In addition, precipitation models which exclude atmospheric information can only be used to simulate rainfall under climatic conditions which are stochastically similar to those used to fit the model. Yet the atmospheric processes that drive precipitation may be nonstationary, even over relatively short time periods (i.e. decades). Thus, the ability of these models to produce precipitation simulations for periods other than those used to fit the model (or even for subintervals of this period) is limited. In particular, a model which fails to incorporate atmospheric information would not be useful in studies of climate variability or climate change.

Over the past few decades advances in data gathering and our understanding of atmospheric circulation have lead to the availability of high quality sets of atmospheric data of variable length (typically, 15-40 years). In addition, the development of physically-based, three-dimensional, dynamic models of global circulations — general circulation models (GCMs) — has lead to the creation of realistic simulations of atmospheric circulation of essentially unlimited duration (some background on GCMs can be found in IPCC (1995)). To take advantage of these types of data, and to address the problems noted above, a new class of stochastic precipitation models known as "weather state models", has been developed. Recent efforts include papers by Hay et. al (1991), Bardossy and Plate (1992), Kidson (1994) and others. Weather state models condition precipitation on available atmospheric information. These models can be thought of as "conditionally stationary" in the sense that any nonstationarity in large-scale atmospheric circulation is (hopefully) captured by the conditioning variables.

Weather state models can be used to generate realistic precipitation simulations by using historical sequences of atmospheric data. Such an approach guarantees that the precipitation simulations will be consistent with the observable atmospheric information. In addition, weather state models can be used with atmospheric simulations from general circulation models to study the effects of climate variability on precipitation. In this respect, weather state models provide important data that cannot, at present, be obtained from GCM simulations. The spatial resolution of GCM's is constrained by both computational considerations as well as our understanding of atmospheric dynamics to scales of approximately 2° to 5° of longitude and latitude. Precipitation, however, varies on much more local scales. For this reason, GCMs have been unable to generate realistic simulations of rainfall (Giorgi and Mearns, 1991). Weather state models provide one solution to this so-called downscaling problem. Using the GCM atmospheric simulations as input, a weather state model can be used to generate realistic simulations of local precipitation.

A final, much more speculative, application of weather state models is to investigate the effect of hypothesized climate changes on precipitation. One such effect of particular

interest is the theory (popularly termed the "greenhouse effect") that observed increases (along with predicted future increases) in atmospheric CO_2 will lead to a global rise in temperature. These predictions are based on experiments with GCMs in which the model is run with an increased (typically doubled) atmospheric concentration of CO_2. Under the strong assumption that the historical relationship between precipitation and large-scale circulation would still apply, a weather state model could be used to access the impact of the altered climate on precipitation.

Hughes and Guttorp (1994a, 1994b) describe a class of models, which they term non-homogeneous hidden Markov models (NHMM), that can be used to model the relationship between atmospheric circulation and precipitation and to generate conditional simulations of precipitation. In a basic hidden Markov model (HMM), one assumes the existence of two processes — an observed process and a hidden process. The observed process (such as rain occurrence at a fixed set of stations) is assumed to be conditionally temporally independent given the hidden process; the hidden process is assumed to evolve according to a first order Markov chain (see Juang and Rabiner (1991) for a review of hidden Markov models). A nonhomogeneous hidden Markov model (NHMM) extends this idea by allowing the transition matrix of the hidden states to depend on a set of observed covariates. In the present application the covariates are derived from the atmospheric data. This approach provides a general framework for the development of weather state models, since Hughes and Guttorp (1994a) show that most existing weather state models can be written as special cases of the NHMM.

In this article we illustrate the use of NHMMs by developing a model for precipitation at 30 rain gauge stations in southwestern Australia. We extend our previous methodological work by developing improved methods of parameter estimation, variance computation and handling missing data. Techniques for model selection and evaluation are discussed and compared. Some interesting insights into the behavior of the pseudolikelihood (Besag, 1975) are also provided. Finally, we show that the proposed model is able to capture the effect

of a (small) non-stationary shift in atmospheric circulation on precipitation, a necessary condition if models such as this are to be used to assess the impact of hypothesized climate changes.

2 Model

To relate observed synoptic (large-scale) atmospheric measures to observed local or regional precipitation patterns we postulate the existence of an unobserved discrete valued process — the "weather state" — which acts as a link between the two disparate scales. Formally, let $\mathbf{R}_t$ be a multivariate vector giving rainfall amounts or occurrences at a network of sites at time t, S_t be the weather state at time t, and $\mathbf{X}_t$ the vector of atmospheric measures at time t for $1 \leq t \leq T$. The $\mathbf{X}_t$ will usually consist of one or more derived measures from the available atmospheric data (e.g. north-south gradient in sea-level pressure). The notation $\mathbf{X}_1^T$ will be used to indicate the sequence of atmospheric data from time 1 to T and similarly for $\mathbf{R}_1^T$ and S_1^T. Lower case will be used to indicate realized values of random variables (i.e. $P(\mathbf{R}_t = \mathbf{r})$). All vectors are row vectors. All vectors and matrices will be written in bold type.

In its most general form, the NHMM is defined by the following assumptions:

$$(M1) \qquad P(\mathbf{R}_t \mid S_1^T, \mathbf{R}_1^{t-1}, \mathbf{X}_1^T) = P(\mathbf{R}_t \mid S_t)$$

$$(M2) \qquad P(S_t \mid S_1^{t-1}, \mathbf{X}_1^T) = P(S_t \mid S_{t-1}, \mathbf{X}_t)$$

and $P(S_1 \mid \mathbf{X}_1^T) = P(S_1 \mid \mathbf{X}_1)$. Specific NHMM's are defined by parameterizing $P(\mathbf{R}_t \mid S_t)$ and $P(S_t \mid S_{t-1}, \mathbf{X}_t)$ as discussed below.

The first assumption (M1) states that the rainfall process, $\mathbf{R}_t$, is conditionally independent given the current weather state. In other words, all the temporal persistence in precipitation is captured by the persistence in the weather state described in (M2). Assumption (M2) states that, given the history of the weather state up to time $t-1$ and the

entire sequence of the atmospheric data (past and future), the weather state at time t depends only on the previous weather state and the current atmospheric data. In the absence of the atmospheric data this is simply the Markov assumption applied to the hidden process. The atmospheric data, when included, are used to modify the transition probabilities of the Markov process — hence the term "nonhomogeneous". Conceptually, it may seem that there is little need for the Markov assumption on S_t given the current atmospheric measurements. In practive, however, the $\mathbf{X}_t$ are typically collected at a point in time while $\mathbf{R}_t$ represents rainfall accumulation over a 24 hour period. The Markov assumption helps to make up for this temporal mismatch. Most weather state models in the literature define the weather states as deterministic functions of the atmospheric variables. These models can be written as special cases of the NHMM by forcing $P(S_t \mid S_{t-1}, \mathbf{X}_t)$ to be degenerate.

There are many possible parameterizations for $P(\mathbf{R}_t \mid S_t)$. We discuss two models for rainfall occurrence since occurrences are often of primary interest; in section 8 we discuss approaches to modelling amounts.

For an n-station network, let $\mathbf{R}_t = \{R_t^1, \ldots, R_t^n\}$ with observed value of $\mathbf{r}_t = \{r_t^1 \ldots, r_t^n\}$. Let $r_t^i = 1$ if rain occurs on day t at station i and 0 otherwise. Then the "independence model" for $P(\mathbf{R}_t \mid S_t)$ is defined as

$$P(\mathbf{R}_t = \mathbf{r} \mid S_t = s) = \prod_{i=1}^{n} p_{si}^{r^i}(1 - p_{si})^{1-r^i} \tag{1}$$

The parameters, p_{si}, may be interpreted as the probability of rain at station i in weather state s. The rainfall occurrences, R_t^i, are assumed to be spatially independent conditional on the weather state (unconditionally, however, the R_t^i will be correlated due to the influence of the common weather state). Hughes and Guttorp (1994a) present an example of a spatially dispersed network of rain gauge stations for which the independence model works well. If there are m weather states then there are nm parameters to estimate in this model.

The second model which we will consider is the autologistic model for multivariate binary data. This model generalizes the independence model by including second order (spatial)

interactions between the stations. The autologistic model is defined as

$$P(\mathbf{R}_t = \mathbf{r} \mid S_t = s) \propto \exp\left(\sum_{i=1}^{n} \alpha_{si} r^i + \sum_{j<i} \beta_{sij} r^i r^j\right) \tag{2}$$

where both α_{si} and β_{sij} must be finite and $\beta_{sii} = 0$. β_{sij} is the "conditional log odds ratio" of rain at station i to rain at station j (in state s) based on the probability distribution $P(r^i, r^j \mid \mathbf{r^{-i,-j}}, S_t = s)$. When β_{sij} is positive, stations i and j are positively associated (within weather state s). A negative value for β_{sij} implies a negative association between stations i and j (within weather state s). When $\beta_{sij} = 0$ for all i, j, and s, equation (2) reduces to the independence model with $\alpha_{si} = \log(p_{si}/(1 - p_{si}))$. To reduce the number of parameters in this model it will often be reasonable to model β_{sij} as a function of the distance and direction between stations i and j.

Two parameterizations for $P(S_t \mid S_{t-1}, \mathbf{X}_t)$ have been investigated. The parameterization which we prefer is motivated by Bayes formula and uses the normal kernel for the joint distribution of the atmospheric data:

$$\begin{aligned} P(S_t = j \mid S_{t-1} = i, \mathbf{X}_t) &\propto P(S_t = j \mid S_{t-1} = i) P(\mathbf{X}_t \mid S_{t-1} = i, S_t = j) && (3) \\ &= \gamma_{ij} \exp(-\frac{1}{2}(\mathbf{X}_t - \mu_{ij})\mathbf{\Sigma^{-1}}(\mathbf{X}_t - \mu_{ij})') && (4) \end{aligned}$$

where μ_{ij} is the mean of $\mathbf{X}_t$ and $\mathbf{\Sigma}$ is the corresponding covariance matrix. This model shows clearly how the NHMM is a general version of the simpler HMM. The γ_{ij} may be thought of as the baseline transition matrix of the weather state process and corresponds to the transition matrix of an HMM. The exponential term quantifies the effect of the atmospheric data on the baseline transition matrix. To ensure identifiability of the parameters, the constraints $\sum_j \gamma_{ij} = 1$ and $\sum_j \mu_{ij} = \mu_i = \mathbf{0}$ are imposed. In this formulation $\mathbf{\Sigma}$ is merely a scaling factor to aid in parameter interpretability and is not estimated as part of the model.

An alternative parameterization of $P(S_t \mid S_{t-1}, \mathbf{X}_t)$ is as a logistic model:

$$P(S_t \mid S_{t-1}, \mathbf{X}_t) \propto \exp(a_{s_{t-1},s_t} + \mathbf{X}_t \mathbf{b}'_{s_{t-1},s_t}) \tag{5}$$

Both (3) and (5) are just different parameterizations of the same underlying model. The choice of which parameterization to use will depend, in part, on the desired interpretation of the parameters in the application at hand.

3 Likelihood

Letting θ denote the model parameters, the likelihood can be written as

$$\begin{aligned} L(\theta) &= P(\mathbf{R}_1^T \mid \mathbf{X}_1^T, \theta) \\ &= \sum_{S_1,\ldots,S_T} P(\mathbf{R}_1^T, S_1^T \mid \mathbf{X}_1^T, \theta) \\ &= \sum_{S_1,\ldots,S_T} P(S_1 \mid \mathbf{X}_1) \prod_2^T P(S_t \mid S_{t-1}, \mathbf{X}_t) P(\mathbf{R}_t \mid S_t) \end{aligned} \tag{6}$$

which appears to be computationally intractable, even for a short sequence of data. However, the forward-backward procedure, a recursive algorithm developed to solve the standard hidden Markov model (e.g. Juang and Rabiner, 1991) can be extended to the NHMM and makes the calculation possible. The basic idea is to successively pass each of the multiple summations in the likelihood as far to the right as possible. For example, the summation over S_T may be passed through all terms in the product except the T'th term. Then, by defining the matrices $\mathbf{A}(\mathbf{x})$, and $\mathbf{B}(\mathbf{r})$ as in table 1, and $\delta(\mathbf{x})$ as the solution to $\mathbf{A}(\mathbf{x})\delta'(\mathbf{x}) = \delta'(\mathbf{x})$, the likelihood can be expressed in the matrix form

$$L(\theta) = \delta(\mathbf{x}_1)\mathbf{B}(\mathbf{r}_1)\mathbf{A}(\mathbf{x}_2)\mathbf{B}(\mathbf{r}_2)\ldots\mathbf{A}(\mathbf{x}_T)\mathbf{B}(\mathbf{r}_T)\mathbf{1}'. \tag{7}$$

If one has several independent sequences of data (for instance, multiple years of data) then the likelihoods for each sequence are multiplied together to form the overall likelihood. To simplify the computations define the recursive relationships

$$\begin{aligned} \mathbf{f}_1 &= \delta(\mathbf{x}_1)\mathbf{B}(\mathbf{r}_1) \\ \mathbf{f}_t &= \mathbf{f}_{t-1}\mathbf{A}(\mathbf{x}_t)\mathbf{B}(\mathbf{r}_t) \end{aligned} \tag{8}$$

Table 1: Definitions used in writing the likelihood.

$$
\begin{aligned}
\pi_{\mathbf{r}_t}(i) &= P(\mathbf{R}_t = \mathbf{r}_t \mid S_t = i) \\
B_{ij}(\mathbf{r}_t) &= \pi_{\mathbf{r}_t}(i) \quad i = j \\
&= 0 \qquad i \neq j \\
\gamma_{ij} &= P(S_t = j \mid S_{t-1} = i) \\
h_{ij}(\mathbf{x}_t) &= P(\mathbf{X}_t = \mathbf{x}_t \mid S_t = j, S_{t-1} = i) \\
A_{ij}(\mathbf{x}_t) &= P(S_t = j \mid S_{t-1} = i, \mathbf{X}_t = \mathbf{x}_t) \\
&= \frac{h_{ij}(\mathbf{x}_t) * \gamma_{ij}}{\sum_j (h_{ij}(\mathbf{x}_t) * \gamma_{ij})}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{b}_T &= \mathbf{1} \\
\mathbf{b}_t &= \mathbf{A}(\mathbf{x}_{t+1})\mathbf{B}(\mathbf{r}_{t+1})\mathbf{b}'_{t+1}. \qquad (9)
\end{aligned}
$$

Note that $\mathbf{f}_t = P(\mathbf{R}_1^t = \mathbf{r}_1^t, S_t \mid \mathbf{X}_1^T)$ (the "forward" probabilities) and $\mathbf{b}_t = P(\mathbf{R}_{t+1}^T = \mathbf{r}_{t+1}^T \mid S_t, \mathbf{X}_1^T)$ (the "backward" probabilities). Since $\mathbf{R}_1^t$ and $\mathbf{R}_{t+1}^T$ are conditionally independent given S_t, the likelihood may be written compactly as

$$
\begin{aligned}
L(\theta) = P(\mathbf{R}_1^T \mid \mathbf{X}_1^T) &= \mathbf{f}_t \mathbf{b}'_t \ \ \forall t \\
&= \mathbf{f}_T \mathbf{1}'.
\end{aligned}
$$

4 Parameter Estimation

Baum et al. (1970) developed an iterative algorithm to obtain maximum likelihood estimates for hidden Markov models by considering the hidden states, S_1^T, to be "missing" data. Let

$\theta = (\theta_R, \theta_S)$, the parameters of the observed and hidden processes, respectively. Then, using the notation from table 1, write

$$
\begin{aligned}
\Psi(\theta' \mid \theta) &= \sum_{\mathbf{S}} P(\mathbf{S} \mid \mathbf{R}_1^T, \theta, \mathbf{X}_1^T) \ln P(\mathbf{R}_1^T, \mathbf{S} \mid \theta', \mathbf{X}_1^T) && (10)\\
&= \sum_{\mathbf{S}} P(\mathbf{S} \mid \mathbf{R}_1^T, \theta, \mathbf{X}_1^T) \left[\sum_{t=1}^{T} \ln \pi'_{r_t}(s_t) + \ln \delta'_{s_1}(\mathbf{X}_1) + \sum_{t=2}^{T} \ln A'_{s_{t-1}, s_t}(\mathbf{X}_t) \right] \\
&= \sum_{\mathbf{S}} P(\mathbf{S} \mid \mathbf{R}_1^T, \theta, \mathbf{X}_1^T) \sum_{t=1}^{T} \ln \pi'_{r_t}(s_t) \\
&\quad + \sum_{\mathbf{S}} P(\mathbf{S} \mid \mathbf{R}_1^T, \theta, \mathbf{X}_1^T) \left[\ln \delta'_{s_1}(\mathbf{X}_1) + \sum_{t=2}^{T} \ln A'_{s_{t-1}, s_t}(\mathbf{X}_t) \right] \\
&= \sum_{t=1}^{T} \sum_{i_0=1}^{m} P(S_t = i_0 \mid \mathbf{R}_1^T, \theta, \mathbf{X}_1^T) \ln \pi'_{r_t}(i_0) \\
&\quad + \sum_{i_0=1}^{m} P(S_1 = i_0 \mid \mathbf{R}_1^T, \theta, \mathbf{X}_1^T) \ln \delta'_{i_0}(\mathbf{X}_1) \\
&\quad + \sum_{t=2}^{T} \sum_{i_1=1}^{m} \sum_{i_0=1}^{m} P(S_{t-1} = i_1, S_t = i_0 \mid \mathbf{R}_1^T, \theta, \mathbf{X}_1^T) \ln A'_{i_1 i_0}(\mathbf{X}_t) && (11)\\
&= \Psi(\theta'_R \mid \theta) + \Psi(\theta'_S \mid \theta) && (12)
\end{aligned}
$$

Baum et al. (1970) recommended iteratively maximizing $\Psi(\theta' \mid \theta)$ as a function of θ' to estimate the parameters. Since $\Psi(\theta' \mid \theta)$ is the expected complete data log likelihood, given the observed data and the current parameters (θ), this procedure constitutes an EM algorithm (Dempster et al., 1977) and will converge to the maximum likelihood estimates. Equation (12) shows that $\Psi(\theta'_R \mid \theta)$ and $\Psi(\theta'_S \mid \theta)$ can be maximized separately. Closed form solutions for these maximization problems are available for some parameterizations of $\Psi(\theta'_R \mid \theta)$ but maximization of $\Psi(\theta'_S \mid \theta)$ usually requires numerical optimization. Further details of the EM algorithm (including closed form solutions to the M-step where possible and first derivatives of $\Psi(\theta'_R \mid \theta)$ and $\Psi(\theta'_S \mid \theta)$ otherwise) for the parameterizations used in this paper are given in the appendix.

Implementation of the EM algorithm requires repeated computation of $P(\mathbf{R}_t \mid S_t)$. When the autologistic model for $P(\mathbf{R}_t \mid S_t)$ (equation 2) is used, the normalization constant of the

distribution (which is a sum over 2^n terms) becomes increasingly burdensome to compute as the number of stations increases. Hughes and Guttorp (1994b) developed an ad-hoc method for parameter estimation. Briefly, they first fit a model with $\beta_{sij} = 0$ (independence model). Using this model they restored (estimated) the state sequence, estimated the parameters of $P(\mathbf{R}_t \mid S_t = s)$ for each s using the restored states and maximum pseudolikelihood estimation (see section 4.1), and then iterated between these two steps. This is similar in spirit to the EM algorithm with two exceptions: the parameters of $P(S_t \mid S_{t-1})$ were not reestimated at each step; and the pseudolikelihood was used in the state restoration process. The iteration was stopped when a "good" match was obtained between the sample rainfall statistics and the model predicted statistics.

This approach has several drawbacks. First, a fully EM approach would reestimate the hidden model parameters at each step. Second, there is no obvious optimality or convergence criterion for the procedure. In fact, as noted by Hughes and Guttorp (1994b), the procedure appears to continue past the point where a "good" match is obtained between the observed and predicted rainfall statistics. In part, this deficiency results from the use of pseudolikelihood to estimate $P(\mathbf{R}_t \mid S_t)$ during the state restoration process, since the resulting "pseudo-probabilities" do not form a true probability distribution. Third, it is not possible to use likelihood-based procedures for model comparison. These problems have lead us to seek a more systematic approach to estimation in applications with many spatially-correlated stations. We now describe two techniques which will be used to modify the EM algorithm to obtain a more computationally feasible procedure.

4.1 Pseudo-likelihood

An alternative to maximum likelihood estimation that has been used successfully to fit the autologistic model is *maximum pseudolikelihood estimation* (MPLE – Besag, 1975). For the

autologistic model (2), the pseudolikelihood has the following form:

$$\begin{aligned} PL(\mathbf{R}_t = \mathbf{r} \mid S_t = s) &= \prod_{i=1}^{n} P(R_t^i \mid \mathbf{R}_t^{-i}, S_t = s) \\ &= \prod_{i=1}^{n} \frac{\exp(\alpha_{si} r^i + \sum_{j \neq i} \beta_{sij} r^i r^j)}{1 + \exp(\alpha_{si} + \sum_{j \neq i} \beta_{sij} r^j)}. \end{aligned} \tag{13}$$

where $\mathbf{R}_t^{-i}$ is the set of rain occurrences for all stations other than station i. Thus, the pseudolikelihood is simply the product of the n conditional probabilities of each station given all the others. Each of the conditional probabilities in (13) can be easily computed since there is no complicated normalizing constant. In the case of $\beta_{sij} = 0$ the pseudolikelihood is equivalent to the likelihood. When $\beta_{sij} \neq 0$ use of the pseudolikelihood is less efficient than the true likelihood. Besag (1977) showed, via simulation studies, that the efficiency loss was small when the spatial autocorrelation was small but could be substantial when the spatial autocorrelation was large. Grenander (1989) showed that maximum pseudo-likelihood estimation is consistent when the data consist of independent identically distributed samples from the autologistic model. Estimates of the parameters of the pseudolikelihood can be obtained using existing software for logistic regression (Cressie, 1993).

4.2 Monte Carlo maximum likelihood

Geyer and Thompson (1992) describe another technique that can be used when computation of the exact likelihood is intractable — *Monte Carlo maximum likelihood* (MCML).

Note that the autologistic model (2) has the form of an exponential family. Letting $< w, \eta >$ denote the inner product of w and η, we can write

$$P(\mathbf{R}_t = \mathbf{r} \mid S_t = s; \eta) = \frac{1}{c(\eta)} \exp < w(\mathbf{r}), \eta >$$

where the sufficient statistic is $w(\mathbf{r}) = (r^1, r^2, \ldots, r^1 r^2, \ldots)$ and the natural parameter is $\eta = (\alpha_{s1}, \ldots, \beta_{s12}, \ldots)$ each of dimension $n(n+1)/2$. The constant $c(\eta)$ is

$$c(\eta) = \sum_{w} \exp < w(\mathbf{r}), \eta > .$$

It is well-known that the moment generating function for the exponential family has the form

$$\begin{aligned} M_W(\tau;\eta) &\equiv \mathbf{E}_\eta(\exp < W, \tau >) \\ &= \frac{1}{c(\eta)} \sum_w \exp < w, \tau > \exp < w, \eta > \\ &= \frac{c(\eta+\tau)}{c(\eta)} \end{aligned} \tag{14}$$

provided $\eta+\tau$ is in the parameter space. Suppose that there is at least one η in the parameter space, say η_0, for which we can compute the normalizing constant $c(\eta_0)$ (for the autologistic model, this can be achieved by setting $\beta_{sij} = 0$). Using (14) gives

$$c(\eta) = c(\eta_0) M_W(\eta - \eta_0; \eta_0) \tag{15}$$

The Monte Carlo maximum likelihood approach replaces $M_W(\eta - \eta_0; \eta_0)$ in (15) by a Monte Carlo estimate based on a series of samples from $P(\mathbf{R}_t \mid S_t; \eta_0)$ (the Gibbs sampler can be used to generate these samples – see below). Denote these samples by $\mathbf{r}_1, \ldots, \mathbf{r}_N$. Then, the value of the the normalizing constant at η may be approximated as

$$c(\eta) \approx \frac{c(\eta_0)}{N} \sum_{i=1}^{N} \exp < w(\mathbf{r}_i), \eta - \eta_0 > . \tag{16}$$

When $P(\mathbf{R}_t \mid S_t; \eta)$ is based on (16), the resulting probability will be denoted by $\hat{P}(\mathbf{R}_t \mid S_t)$, and similarly for other probabilities involving $c(\eta)$.

The first and second moments of $\mathbf{R}_t$ may also be computed using Monte Carlo methods and the same sample, $\mathbf{r}_1, \ldots, \mathbf{r}_N$ (the moments are used in the numerical maximization of the autologistic model likelihood). For instance,

$$\begin{aligned} \mathbf{E}_\eta(\mathbf{R}_t^k) &= \frac{1}{c(\eta)} \sum_w r^k \exp < w(\mathbf{r}), \eta > \\ &= \frac{1}{c(\eta)} \sum_w r^k \exp < w(\mathbf{r}), \eta_0 > \exp < w(\mathbf{r}), \eta - \eta_0 > \\ &= \frac{\mathbf{E}_{\eta_0} r^k \exp < w(\mathbf{r}), \eta - \eta_0 >}{\mathbf{E}_{\eta_0} \exp < w(\mathbf{r}), \eta - \eta_0 >} \end{aligned}$$

$$\approx \frac{\sum_{i=1}^{N} r_i^k \exp < w(\mathbf{r}_i), \eta - \eta_0 >}{\sum_{i=1}^{N} \exp < w(\mathbf{r}_i), \eta - \eta_0 >} \tag{17}$$

Similarly,

$$\mathbf{E}_\eta(\mathbf{R}_t^k \mathbf{R}_t^h) \approx \frac{\sum_{i=1}^{N} r_i^k r_i^h \exp < w(\mathbf{r}_i), \eta - \eta_0 >}{\sum_{i=1}^{N} \exp < w(\mathbf{r}_i), \eta - \eta_0 >} \tag{18}$$

4.3 A modified EM procedure for the autologistic model

As can be seen in equation (11), the M-step of the EM algorithm amounts to finding the root of a weighted sum of the complete data scores for each weather state (see the appendix for further details). When the autologistic model is used for $P(\mathbf{R}_t \mid S_t)$ both the E-step and the M-step are computationally intractable: in the E-step $P(\mathbf{R}_t \mid S_t)$ is used to compute the weights (namely, $P(S_t = i_0 \mid \mathbf{R}_1^T, \theta, \mathbf{X}_1^T)$ and $P(S_{t-1} = i_1, S_t = i_0 \mid \mathbf{R}_1^T, \theta, \mathbf{X}_1^T)$ in (11)), so the normalizing constant of the distribution is needed; in the M-step the normalizing constant as well as the first and second moments of $\mathbf{R}_t$ are required to compute the scores associated with $\Psi(\theta'_R \mid \theta)$. Additionally, the M-step computations may be performed several times for each EM iteration since numerical optimization methods are required to maximize $\Psi(\theta'_R \mid \theta)$.

We have developed two modifications of the EM algorithm to circumvent the computational intractibility of $P(\mathbf{R}_t \mid S_t)$. In both approaches equation (16) is used to estimate the normalizing constant of the autologistic distribution. This allows us to estimate $P(\mathbf{R}_t \mid S_t)$ and, hence, the weights in the E-step. The approaches differ in their treatment of the M-step. In the first approach (which we term EM/MPLE) the pseudolikelihood (13) scores are substituted for the scores of $P(\mathbf{R}_t \mid S_t)$ in the M-step. In the second approach (which we term EM/MCML) the scores of the true complete data likelihood are approximated by MCML estimates (based on equations 17 and 18). In both cases, Newton-Raphson iteration is used maximize $\Psi(\theta'_R \mid \theta)$ in the M-step. Table 2 summarizes these two algorithms and compares them to the standard MLE.

The computational efficiency of these approaches depends on the choice of η_0. An η_0 which is far from η will require a much larger N to achieve a stable estimate of $M_W(\eta - \eta_0; \eta_0)$

Table 2: Summary of EM algorithms for maximizing $\Psi(\theta'_R \mid \theta)$.

	MLE	EM/MPLE	EM/MCML
E-step	compute $w_t(s) = P(S_t = s \mid \mathbf{R}_1^T, \mathbf{X}_1^T, \theta)$ using (2)	estimate $\hat{w}_t(s) = \hat{P}(S_t = s \mid \mathbf{R}_1^T, \mathbf{X}_1^T, \theta)$ using (2) and (16)	estimate $\hat{w}_t(s) = \hat{P}(S_t = s \mid \mathbf{R}_1^T, \mathbf{X}_1^T, \theta)$ using (2) and (16)
M-step	maximize $\sum_{ts} w_t(s) \ln P(\mathbf{R}_t \mid s, \theta')$ as a function of θ'	maximize $\sum_{ts} \hat{w}_t(s) \ln PL(\mathbf{R}_t \mid s, \theta')$ as a function of θ'	maximize $\sum_{ts} \hat{w}_t(s) \ln \hat{P}(\mathbf{R}_t \mid s, \theta')$ as a function of θ'

than would be required for η_0 near to η. Therefore, we use the following operational procedure for parameter estimation. First, an NHMM is fit using the conditional independence model for $P(\mathbf{R}_t \mid S_t)$ (equation 1). The parameters from this fit serve as η_0. The norm is easily computed (since $\beta_{sij} = 0$, the norm is $\prod_i(1 + \exp(\alpha_{si}))$) and it is straightforward to generate random variates $\mathbf{r}_1, \ldots, \mathbf{r}_N$. Starting from these initial values one cycle of the EM algorithm is run using numerical maximization (Newton-Raphson) of either the pseudolikelihood (EM/MPLE) or the Monte Carlo likelihood (EM/MCML) in the M-step. At the next E-step (beginning of cycle 2) equation (16) is used to estimate the normalizing constant $c(\eta_1)$ and hence the weights $\hat{P}(S_t = s \mid \mathbf{R_t}, \theta)$. As before, the pseudolikelihood or the Monte Carlo likelihood is maximized in the M-step. At the beginning of cycle 3, η_1 and $c(\eta_1)$ replace η_0 and $c(\eta_0)$ in equation (16) to estimate $c(\eta_2)$. Subsequent iterations proceed in a similar manner.

Simulation of $\mathbf{r}_1, \ldots, \mathbf{r}_N$ from $P(\mathbf{R}_t \mid S_t; \eta_0)$ is simple when η_0 represents the conditional independence model (i.e. all $\beta_{sij} = 0$) but simulation based on an arbitrary η_0 is less straightforward. However, the Gibbs sampler (Geman and Geman, 1984) provides an efficient way of sampling from the general autologistic distribution. To implement the Gibbs sampler for this problem we successively sample from the conditional distributions $P(R^i \mid \mathbf{R}^{-i}, S = s, \eta_0)$ (the form of this distribution was given in equation 13). One random value of $\mathbf{R}$ is generated for each pass through the subscripts $i = 1 \ldots n$. A value of $\mathbf{R}$ from the previous

iteration can be used to seed the sampler (that is, a value from $P(\mathbf{R} \mid S, \eta_0)$ can be used to start the sampler for $P(\mathbf{R} \mid S, \eta_1)$), although a burn-in period may be necessary before samples from the new distribution are obtained. The Gibbs sampler can also be used to generate simulations from a fitted NHMM, although in that case it is important to use only every k'th sample (e.g. $k = 10$) to avoid introducing artifical temporal correlation in the rainfall simulation.

Additional tuning can improve the computational efficiency of these algorithms in any given problem. For instance, in any EM procedure in which the M-step depends on numerical optimization, it is not necessary to fully maximize the expected complete data likelihood to ensure self-consistency (Rai and Matthews, 1993). In the modified EM procedures described above we have found that it is often advantageous to limit the number of Newton-Raphson iterates in the M-step. This prevents the new parameters from moving too far from the current values and reduces the number of samples needed to update the normalizing constant and moments via MCML. If the distance between η_0 and η_1 is still too large then the interval can be broken into several smaller intervals and (16) can be applied repeatedly.

5 Missing data

Missing values will commonly be found in the rainfall record and less often in the atmospheric data record. Missing rainfall data may be incorporated into the estimation procedures by considering the missing rainfall observations to be unobserved just as the weather states are unobserved. Formally, write $\{\mathbf{R}_1^T\} = \{\mathbf{R}^{obs}, \mathbf{R}^{miss}\}$ where $\{\mathbf{R}^{obs}\}$ is the observed data and $\{\mathbf{R}^{miss}\}$ is the missing data. Then equation (10) can be rewritten as

$$\Psi(\theta' \mid \theta) = \sum_{\mathbf{S}, \mathbf{R}^{miss}} P(\mathbf{S}, \mathbf{R}^{miss} \mid \mathbf{R}^{obs}, \theta, \mathbf{X}_1^T) \ln P(\mathbf{R}_1^T, \mathbf{S} \mid \theta', \mathbf{X}_1^T) \tag{19}$$

and the development of the EM algorithm proceeds along the same lines as before. In the end the changes are minor and involve replacing $P(\mathbf{R}_t \mid S_t)$ by $\sum_{\mathbf{R}_t^{miss}} P(\mathbf{R}_t \mid S_t)$ in the

computation of the forward-backward probabilities (equations 8 and 9).

When the pseudolikelihood is used, the approach to missing data is slightly different. The pseudolikelihood (equation (13)) conditions on only those R_t^j that are observed and the product is over only the observed R_t^i. That is, rewrite (13) as

$$PL(\mathbf{R}_t \mid S_t = s) = \prod_{i \notin miss} P(R_t^i \mid \mathbf{R}_t^{-i,-miss}, S_t = s) \tag{20}$$

where $\{miss\}$ represents the indicies of the missing sites. To derive the distribution of $P(R_t^i \mid R_t^{-i,-miss}, S_t = s)$ note that

$$\frac{P(R^i = 1 \mid R^{-i,-miss})}{P(R^i = 0 \mid R^{-i,-miss})} = \frac{P(R^i = 1, R^{-i,-miss})}{P(R^i = 0, R^{-i,-miss})} = \frac{\sum_{R^{miss}} P(R^i = 1, R^{-i})}{\sum_{R^{miss}} P(R^i = 0, R^{-i})} \tag{21}$$

This expression involves only the numerator of the autologistic model (2) and is therefore readily computable. The conditional probabilities required in (20) follow directly from (21).

In the case of missing atmospheric data, values may be imputed from a separate model or the $\mathbf{X}_t$ may be set equal to their mean, μ_{ij}, which effectively reduces $P(S_t \mid S_{t-1}, \mathbf{X}_t)$ to the baseline transition matrix, γ (see equation (4)).

6 Model selection and evaluation

There are several issues involved in selecting the "best" NHMM for any given set of data. These include order selection (selecting the number of hidden states), selecting candidate atmospheric variables for inclusion in the model, testing various constrained submodels and evaluating the goodness-of-fit of the model. These issues are discussed below.

Note that lower order NHMMs are nested in higher order models and that a model with one atmospheric variable is nested within a model with additional atmospheric variables. A standard statistical technique for choosing between two nested models is the likelihood ratio criterion. Such a procedure might be used to test, for example, the hypotheses of no spatial correlation in the autologistic model ($\beta_{sij} = 0$ in equation 2). However, there are objections

to using the likelihood ratio criterion to decide among a series of models, especially when no one model is favored a priori. A cogent discussion of these issues is given by Kass and Raftery (1995) who describe an alternative approach to model selection based on Bayes factors. Specifically, they suggest selecting the model which maximizes the Bayes factor $2\ell + \log|V|$ where ℓ is the log-likelihood and V is the covariance matrix of the parameters. To estimate $|V|$, the approach described by Hughes (1996) to approximate the information matrix for hidden Markov models can be adapted to the NHMM. Kass and Raftery (1995) also describe a simpler approximation which can be used in model selection:

$$2\ell - \nu \log T \tag{22}$$

where T is the sample size and ν is the number of independent parameters in the model. This is referred to as the *Bayes Information Criterion* (BIC).

Our experience has been that both the BIC and Bayes factors usually yield a reasonable model in the sense that the final model is interpretable and provides a good fit to the data. There are theoretical reasons for prefering Bayes factors but the BIC is much easier to compute. Further research on the utility of these measures for model selection in the context of NHMM's is required.

The atmospheric variables to be considered for inclusion in a model (the $\mathbf{X}$'s) will typically be determined by climate patterns and rainfall generating mechanisms specific to the application at hand. A variety of derived measures (e.g. mean level, gradients in various directions, laplaceans, etc.) on several fields may be considered potentially relevant. Since it is usually not computationally practicable to try fitting all possible models, we have found it is useful to screen the candidate measures in some simple way, such as correlating the atmospheric measures with rainfall at each station. Only measures that show a relatively high correlation are considered for inclusion in the model. Bayes factors or the BIC criterion can then be used to choose a set of measures to be included in the final model.

Ultimately, a good model is one that reproduces the observed precipitation statistics

on reserved data. The key statistics of interest are usually the first and second (spatial and temporal) moments, and the distribution of storm lengths and storm interarrival times (defined as the number of consecutive days of rain and no rain, respectively). These duration distributions are of particular interest to hydrologists because they strongly influence flood magnitude and frequency. They have also proven to be the most difficult characteristics of rainfall to reproduce using weather state models. The predicted values of these statistics can be computed from the fitted NHMM by simulation.

7 Example

A fifteen year record (1978–1992) of daily winter (May–October) rainfall occurrences (2760 days, total) at 30 stations in southwestern Australia was made available by the Australian Bureau of Meteorology. The locations of the stations are shown in figure 1. While total daily rainfall is available, we restrict our modelling efforts to the binary measure, rainfall below/above 0.3 mm, since this measure is often of greatest interest to hydrologists. Approaches to developing an amounts model are described in the discussion. Each rainfall value represents the total rainfall over a 24 hour period ending at 0900 (local standard time). Atmospheric data were obtained from the Australian Bureau of Meteorology on a Lambert conformal grid and interpolated to a rectangular grid of similar scale—2.25^o latitude by 3.75^o longitude (also shown in figure 1). Available atmospheric measures included sea-level pressure, geopotential height at 850 hPa (hectoPascals) and 500 hPa, air temperature, dew point temperature and u (north–south) and v (east– west) wind speed components. The atmospheric measurements were taken at 1900 (local standard time) on the preceeding day. Consultation with atmospheric scientists produced a list of 24 summary measures of the atmospheric data that might influence rainfall in this area. These included measures such as mean sea-level pressure and geopotential height over the region of interest, north–south and east–west gradients, etc. Some preliminary analyses were conducted to get a rough idea of

the ability of each of these summary measures to predict rainfall. These analyses included simple procedures such as correlating each summary atmospheric measure with rainfall at each station, as well as more complex multivariate procedures such as using tree-based classification (Breiman et al., 1984) to determine which summary atmospheric measures best predicted rainfall occurrence patterns at a subset (stations 7, 9, 16, and 17) of the stations. These preliminary analyses were used to provide a tentative ranking of the 24 measures for inclusion in the NHMM.

Using the results of the preliminary analyses for guidance, a series of NHMM's were fit to the first 10 years of data using the conditional spatial independence model (equation 1) for $P(\mathbf{R_t}|S_t)$ and the Bayes model (equation 3) for $P(S_t \mid S_{t-1}, \mathbf{X}_t)$. The remaining 5 years of data were reserved for model validation. Bayes factors (BF) (section 6) were used to guide model selection. Among the models that assume conditional spatial independence for $P(\mathbf{R_t}|S_t)$, the "best" model (having the lowest BF) included 6 weather states and 3 atmospheric measures (mean sea-level pressure (MSLP), north-south gradient in sea-level pressure and the east-west gradient in 850 hPa geopotential height (GPH)) and had a log-likelihood of -17110 (table 3). The BIC gave generally similar results, although that criterion would have lead us to chose a 7 weather state model with only 2 atmospheric measures (MSLP and north-south gradient in SLP). In general, we have observed some trade-off between the number of weather states identified and the number of atmospheric variables included in the model—models with fewer weather states achieve a minimum BF with more atmospheric variables while models with more weather states achieve a minimum BF with fewer atmospheric variables. Further research is necessary to assess the efficacy of these two measures for model selection in this context. However, one would expect that a model with more atmospheric information would produce precipitation simulations which are more responsive to shifts in atmospheric conditions.

Figures 2 and 3 illustrate the fit of the 6 state, 3 atmospheric variable model to observed rainfall statistics (the model-based statistics are computed by generating multiple simulations

Table 3: Comparison of the log-likelihood, BIC and Bayes Factor for several nonhomogeneous hidden Markov models using the conditional spatial independence model for $P(\mathbf{R}_t \mid S_t)$. Covariates are 1 = mean sea-level pressure; 2 = Mean geopotential height at 500mb; 4 = N-S gradient in sea level pressure; 8 = E-W gradient in geopotential height at 850mb.

no. states	covariates	log-likelihood	df	BIC	Bayes factor
4	—	19225	132	39442	39589
5	—	18414	170	38106	38229
6	—	18040	210	37659	37798
6	1	17592	240	36988	36769
6	1,4	17214	270	36458	35853
6	1,4,8	17110	300	36475	35528*
6	1,2,4,8	17031	330	36542	35852
7	—	17698	252	37290	37406
7	1	17161	294	36532	36287
7	1,4	16876	336	35751	35751
7	1,4,8	16747	378	36336	35756

from the model, conditional on the observed atmospheric data, and then averaging over the simulations so that variability in the predicted quantities is negligible). From these figures it is clear that the conditional spatial independence model does well in reproducing the observed probability of rainfall at each station and the distribution of "storm durations" (number of consecutive days with rain). However, this model does less well at reproducing the observed patterns of spatial correlation between stations, particularly for stations that are highly correlated. This makes sense: most of the spatial correlation between stations is induced by the common weather state and this source of correlation is captured by the model. However, additional correlation between nearby stations is created by local orographic and

other "sub weather state" scale effects and this source of correlation is not captured in the independence model for $P(\mathbf{R}_t \mid S_t)$.

To include these local effects, an NHMM was fit using the autologistic model (eq. 2) for $P(\mathbf{R}_t \mid S_t)$. The conditional log-odds ratios, β_{sij}, were modelled as a function of the distance and direction between the stations to reduce the number of parameters. To determine an appropriate functional form for the β_{sij}, each day was first classified into its most likely weather state using the 6 state, 3 atmospheric variables, conditional spatial independence model described above (a procedure known as the Viterbi algorithm is used to classify each day into a weather state; see, for example, Juang and Rabiner, 1991). Then, for each state, empirical estimates of the pairwise (unconditional) log-odds ratios were generated and plotted against the distance and direction between the stations. These plots (see figure 4 for an example) suggested that the within-state spatial correlation declined as the distance between stations increased and varied elliptically with direction. Using a nonlinear least-squares regression analysis, the following functional form was found to give a good fit to the empirical log-odds ratios and was, therefore, adopted as a model for the conditional log-odds ratios:

$$\beta_{sij} = b_{0s} + b_{1s} \log(d_{ij} \sqrt{\cos(\phi_s + h_{ij})^2 + \sin(\phi_s + h_{ij})^2 / e_s}) \tag{23}$$

where d_{ij} and h_{ij} are, respectively, the distance and direction between stations i and j. For each state, s, there are 4 parameters in this model. Although, theoretically, all four parameters could be estimated by the methods outlined in section 4, estimation of the nonlinear parameters, ϕ_s and e_s, slows down the computations substantially. Therefore, these parameters were fixed at the values obtained from the nonlinear regression analyses of the empirical log-odds ratios. The b's were then estimated using both the EM/MPLE and EM/MCML procedures described in section 4.

Both approaches significantly improved the fit of the model to the empirical log-odds ratios, as seen in figure 2. The EM/MCML algorithm converged to a model with estimated

log-likelihood equal to -15688 while the estimated log-likelihood of the final model obtained using the EM/MPLE algorithm was -15817. Both represent a significant improvement over the conditional spatial independence model. The computational performance of EM/MCML was somewhat slower than EM/MPLE. In part, this results from the need to estimate the moments of $\mathbf{R}_t$ by Monte Carlo, but we also observed that the EM/MCML algorithm was more likely than the EM/MPLE algorithm to overshoot the maximum of the objective function on the first few M-steps, resulting in a greater computational effort to compute the new normalizing constant.

The ability of the NHMM to reproduce key precipitation statistics conditional on the observed atmospheric data suggests that this model could be useful for generating conditional rainfall simulations for the period 1978–1987. However, if the model is to be used to generate precipitation simulations for other periods or alternative atmospheric datasets (e.g. to investigate the effects of climate change) then it is important to test the model on reserved data. Figure 5 compares various observed rainfall statistics to those predicted by the model for the 5 years of reserved data. Results for the spatial model fit using the EM/MCML algorithm are shown. Results obtained from the EM/MPLE algorithm (not shown) are similar. Figure 5 shows increased variability when the model is applied to reserved data (as expected) but no systematic biases. This latter point is important since a small but measureable shift in the mean atmospheric data fields occurred during the 5 year period of reserved data (table 4). If this shift is deliberately removed from the atmospheric data, but not the rainfall data (e.g. by recentering the atmospheric measures in the 5 year period around the same means as were observed in the 10 year period) then a small but noticable (about 3 percentage points) downward bias is observed in the predicted rainfall probabilities (table 4). In other words, the lack of bias seen in figure 5 indicates that the model was able to adjust the rainfall probabilities to account for the (slight) nonstationary shift in the atmospheric data. This is clearly a necessary condition if the model is to be able to make useful predictions about rainfall under altered climates.

Table 4: Mean change in the three atmospheric measures included in the final model between the 10 year period used for model fitting and the 5 year period used for model validation. Also shown is the bias in the predicted precipitation probability (observed - predicted) during the 5 year evaluation period if the model does/does not adjust for the mean change. SLP = sea level pressure (hPa); GPH850 = geopotential height (m) at 850 hPa.

			Precipitation bias (%)	
Mean SLP	N-S SLP gradient	E-W GPH850 gradient	adjusted	not adjusted
-0.81	0.47	0.45	0.014	3.1

Although the EM/MPLE and EM/MCML algorithms perform comparably in the example presented above, we have noted some unusual behavior in the EM/MPLE algorithm in other situations. We will discuss one such example that occurred when the full 15 years of data were used for model fitting. Initially, the same basic model was obtained (6 states with the same 3 atmospheric variables) using the conditional spatial independence model for $P(\mathbf{R}_t \mid S_t)$. As in figure 2, certain deficiencies were noted in the fit of the model to the observed log-odds ratios. Therefore, as described previously, we fit an autologistic model for $P(\mathbf{R}_t \mid S_t)$, using the functional form for the β_{sij} shown in (23). After a single EM iteration of the EM/MPLE algorithm, the estimated log-likelihood was observed to *decrease* substantially. Further iterations did not improve the fit. When the overall (marginal) observed probability of rainfall at each station was compared to that predicted from the EM/MPLE model the fit was very poor. It was not until we looked at the fit within each weather state that a possible explanation for this anomalous behavior presented itself. Each weather state is characterized by a particular rainfall pattern. For instance, state 1 is associated with greater probability of rainfall in the northern and eastern stations and lower probability of rainfall along the coast. State 2, in contrast, exhibits a very homogeneous pattern of high probability of rain over the entire study area. States 3, 4 and 6 are similar to state 1 in the

sense that they present spatially variable patterns of rainfall over the region while state 5 is associated with a very homogeneous pattern of low probability of rainfall over the entire region. When the model for $P(\mathbf{R}_t \mid S_t)$ obtained after the first EM/MPLE iteration was used to simulate rainfall from each of the 6 weather states separately, we found that the simulations for state 2 had a *low* probability of rain at all stations while the simulations for state 5 had a *high* probability of rain at all stations (figure 6). Both results are exactly opposite the observed data for those states. In contrast, the predicted rain probabilities from the other states closely matched the observed probabilities of rainfall in those states.

The failure of the EM/MPLE algorithm in this situation may result from a lack of information about the log-odds ratios in states 2 and 5. That is, the homogeneous patterns of rainfall seen in these states mean that there will be relatively few discordant rainfall pairs and, hence, little information about the spatial parameters b_{0s} and b_{1s}. In addition, although asymptotically consistent, the pseudolikelihood tends to overestimate the spatial correlation in finite samples (Geyer and Thompson, 1992, who also report poor performance by the MPLE in some situations). In combination, these two factors apparently produce an extremely poor fit to the data in these two states. If the spatial parameters for states 2 and 5 are fixed at 0 (i.e. in equation (23) set $b_{02} = b_{05} = b_{12} = b_{15} = 0$), which corresponds to conditional spatial independence, and the remaining parameters are estimated using EM/MPLE, the algorithm converges smoothly. No examples of inconsistent behavior have been observed using the EM/MCML algorithm.

Although the weather states are abstract constructs of the model, they can be visualized by first classifying each day into its most likely state and then averaging the values of sea-level pressure, geopotential height or other atmospheric measures over all days in a given state at each node of the atmospheric data grid. The resulting "composite" field can then be contoured to give a visual representation of the average field in that state. This procedure is illustrated in figure 7 for the final EM/MCML model. The other models give substantially similar plots. The upper row of plots characterizes state 2, which has high probability of rain

at all stations. In this state, the sea-level pressure pattern leads to a strong westerly flow of moist marine air over the entire study area. In contrast, state 5 (bottom row of figures) is associated with low probability of rain at all stations. This state is characterized by a high pressure system in the great Australian bight which blows dry air from inland over the region and forces moist marine air to the south. The remaining four states are characterized by rainfall in particular regions of the study area. For instance, state 4 (shown in the middle row of plots in figure 7) exhibits high probability of rain at the southwest stations but low probability of rain in the north and western stations. In this case the sea-level pressure pattern seen in state 2 is shifted to the south. The patterns seen in the composite MSLP plots are also present in the 850 hPa GPH composite plots, supporting our view that the model has identified the dominant synoptic scale features of precipitation in this region.

The weather states defined by this model are also relatively more homogeneous with respect to the atmospheric variables compared to the full dataset. The pooled, within-state variances for the three atmospheric measures used in the model are, respectively, 56%, 67% and 39% smaller than the total variances for those measures.

In addition to defining weather states, the fitted model can provide additional information to researchers, including the percentage of days which fall into each weather state, the average duration of each weather state and the pattern of transitions between weather states.

8 Discussion

Nonhomogeneous hidden Markov models can provide hydrologists and atmospheric scientists with a useful tool for generating realistic simulations of precipitation and understanding the relationships between atmospheric circulation patterns and rainfall. This approach to precipitation modelling will be most successful in areas and/or seasons where precipitation is driven by synoptic-scale systems. It is unlikely that these models will be successful in areas or seasons in which rainfall is driven primarily by convective activity (e.g. thunderstorms) since

these processes evolve on relatively small scales and may not be predictable from synoptic circulation patterns.

NHMMs generalize the concept of a weather state model as described by Hay et al. (1991), Bardossy and Plate (1992), Kidson (1994) and others. In these models, however, the investigators explicitly defined the weather states. The resulting states, while reflecting meteorological intuition, may not have been optimal for modelling rainfall. An important advantage of the NHMM approach is that one need not define the weather states a priori. Instead, one need only identify those atmospheric measures which are thought to influence precipitation. Then the weather states are defined automatically by the model. Plots such as figure 7 can provide insight into the interpretation of the weather states and the relationship between atmospheric circulation patterns and precipitation. A comparison of the total variance in the atmospheric measures to the within-weather state variance can be used to assess the homogeneity of the weather states.

Another important distinction of the NHMM approach is the use of the Markov assumption in the definition of the weather states. In previous work (Hughes and Guttorp, 1994a), we noted that the inclusion of the Markov assumption improves the fit of the model to observed rainfall statistics, particularly the observed duration distribution. Although it is conceptually appealing to assume that the current weather state (and, therefore, the current rainfall pattern) should depend only on current atmospheric conditions, the temporal discordance between the atmospheric data and the precipitation data described in section 2 makes such an assumption untenable.

NHMMs represent a completely stochastic approach to the downscaling problem. Thus far, more mechanistic approaches, such as GCM-based simulations of precipitation, have proved to be deficient at the spatial and temporal scales of relevance to regional and local hydrology (Grotch and MacCraken, 1991). Although it is, at present, computationally impossible to implement an entire GCM at local scales, some progress has been made in developing "nested" GCMs which implement phenomenologic models for rainfall on a finer

grid over a restricted area and use the coarse scale GCM data as boundary conditions. These limited area models are able to achieve grid spacings on the order of 1^o by 1^o. Even at this scale, however, deficiencies in the precipitation simulations have been noted (Mearns et al., 1995). Additional studies to compare the NHMM approach with the nested GCM approach in terms of ability to accurately reproduce current climate precipitation patterns are ongoing (Charles et al., 1996). However, even if GCMs are, at some point, able to accurately characterize local precipitation patterns, downscaling models would still be valuable for modelling phenomena that are not explicitly included in the GCMs (e.g. air pollution patterns).

We believe that future research in this area should focus on both conceptual and methodological issues. The outstanding conceptual issue in research on downscaling is making predictions under altered climate regimes. Predictions of the effects of hypothesized changes in climate (e.g. global warming) are based on GCM simulations and are, therefore, restricted to large scale effects. As described in IPCC (1995, sec. 6.6), there are considerable discrepancies between predictions of different GCMs in terms of changes in precipitation that would occur on a sub-continental scale under a doubled CO_2 climate. In addition, there are substantial biases in precipitation between GCM control runs and observations. At present, therefore, assessment of the local hydrologic effects of climate change necessitates the use of models to downscale the (altered climate) GCM circulation patterns. However, this means that the downscaling models must be used under very different conditions than they were fit under. Although the validity of a downscaling model under a radically different climate regime is impossible to determine a priori, some insight into the behavior/validity of the model under altered climates is possible from studies on reserved datasets, as has been presented here. In addition, data from "natural experiments" such as the eruption of Mt. Pinatubo in 1991, which caused measureable changes in global climate, can provide another approach to model validation. Of course, the validity of downscaling models for impact assessment also depends on the validity of the GCM model which provides the atmospheric information that drives precipitation. Assessment of GCM models is an active area of research — the interested

reader is referred to the IPCC report (1995, chapter 5) for a summary of the current state of knowledge.

Several methodological issues remain. The model developed here deals with rainfall occurrences only. For many applications this is sufficient. However, for some applications it is also necessary to model amounts. One approach is to first fit an NHMM to the occurrence data and then fit a model to the amounts, conditional on occurrence (and, possibly, weather state), a posteriori. However, this means that the amounts do not influence the definitions of the weather states. To fully integrate an amounts model into the NHMM would require specification of a multivariate mixed discrete-continuous model for $P(\mathbf{R}_t \mid S_t)$. In the context of an explicit weather state model, Bardossy and Plate (1992) have used a transformed multivariate normal distribution to model amounts at multiple stations. To extend this idea further, models based on multivariate observations (e.g. precipitation and temperature) could be developed and would be useful for input into hydrologic models.

Selecting the "correct" order (number of states) of a HMM or a NHMM is a problem similar to selecting the proper number of components in a mixture model. Titterington (1990) reviews some approaches to this problem but concludes that further research is needed. We have found that Bayes factors (and the related BIC) yield useful models but further research, including simulation studies, are needed to completely evaluate the efficacy of these procedures for model selection.

Finally, to further extend the utility of the weather state approach, methods could be developed to simulate rainfall occurrence at locations that have not been explicitly included in the model. In the context of the autologistic model this could be accomplished by spatially interpolating the α_{si} (note that a spatially smooth model for β_{sij} has already been included in the present analysis). For the example presented in section 7 we observed that the α_{si} from the best-fitting autologistic model were small and showed little variation within weather state in the interior of the network (e.g. -2.0 to -6.0, depending on the weather state; note that $\exp(\alpha_{si})/1 + \exp(\alpha_{si})$ is the probability of rain at station i given no rain at all other

stations). Thus, to generate rainfall probabilities at a new location, i', in the interior of the network one could set $\alpha_{si'}$ equal to the mean value of α_{si} from other stations in the interior and compute $\beta_{si'j}$ from (23).

Acknowledgements: Partial support was provided by National Science Foundation grant DMS-9524770 and by a University of Washington Royalty Research Fund award to the first author. Portions of this work were performed while the first author was Visiting Scientist at CSIRO, Perth, Australia. The Australian Bureau of Meteorology kindly provided the atmospheric and rainfall data for this study. The authors are grateful to Bryson Bates, Mick Fleming, and Tom Lyons for helpful discussions. Software (written in Fortran with calls to NAG library routines) to implement the methods discussed in this paper is available from the first author.

Appendix

In terms of the forward-backward probabilities (equations 8 and 9) and the definitions in table 1, $\Psi(\theta'_S \mid \theta)$ and $\Psi(\theta'_R \mid \theta)$ may be written as

$$\Psi(\theta'_S \mid \theta) = \sum_{s=1}^{m} \frac{\mathbf{f}_1(s)\mathbf{b}_1(s) \ln \delta'_s(\mathbf{X}_1)}{\mathbf{f}_T \mathbf{1}} + \sum_{t=2}^{T} \sum_{s_1=1}^{m} \sum_{s_0=1}^{m} \frac{\mathbf{f}_{t-1}(s_1) A_{s_1 s_0}(\mathbf{X}_t) \pi_{r_t}(s_0) \mathbf{b}_t(s_0) \ln A'_{s_1 s_0}(\mathbf{X}_t)}{\mathbf{f}_T \mathbf{1}'} \tag{A1}$$

and

$$\Psi(\theta'_R \mid \theta) = \sum_{t=1}^{T} \sum_{s=1}^{m} \frac{\mathbf{f}_t(s)\mathbf{b}_t(s)}{\mathbf{f}_T \mathbf{1}'} \ln \pi'_{r_t}(s). \tag{A2}$$

The EM algorithm consists of alternately computing the forward-backward probabilities (the E step), then maximizing $\Psi(\theta'_S \mid \theta)$ and $\Psi(\theta'_R \mid \theta)$ (the M step) until convergence is achieved. We discuss each of these separately.

$\underline{\Psi(\theta'_S \mid \theta)}$

In general, $\Psi(\theta'_S \mid \theta)$ must be maximized numerically using a routine such as the NAG library E04UCF. Specification of the algebraic form of the derivatives of $\Psi(\theta'_S \mid \theta)$ increases efficiency and are presented here. Let $\Psi_S \equiv \Psi(\theta'_S \mid \theta)$ and $\theta = (\gamma, \mu)$. Then, using the Bayes model (equation 3) for $P(S_t \mid S_{t-1}, \mathbf{X}_t)$

$$\frac{\partial \Psi_S}{\partial \theta'} = \sum_{s=1}^{m} \frac{C_1(s)}{\delta'_s(\mathbf{X}_1)} \frac{\partial \delta'_s(\mathbf{X}_1)}{\partial \theta'} + \sum_{t=2}^{T} \sum_{s_1=1}^{m} \sum_{s_0=1}^{m} \frac{C_{t-1,t}(s_1, s_0)}{A'_{s_1,s_0}(\mathbf{X}_t)} \frac{\partial A'_{s_1,s_0}(\mathbf{X}_t)}{\partial \theta'} \tag{A3}$$

where

$$\begin{aligned}
C_t(s) &= \frac{\mathbf{f}_t(s)\mathbf{b}_t(s)}{\mathbf{f}_T \mathbf{1}} \\
C_{t-1,t}(s_1, s_0) &= \frac{\mathbf{f}_{t-1}(s_1) A_{s_1 s_0}(\mathbf{X}_t) \pi_{r_t}(s_0) \mathbf{b}_t(s_0)}{\mathbf{f}_T \mathbf{1}'} \\
Q_{ij} &= -\frac{1}{2}(\mathbf{X} - \mu_{ij})\Sigma^{-1}(\mathbf{X} - \mu_{ij})^t \\
\frac{\partial A'_{ij}(\mathbf{X})}{\partial \gamma_{gh}} &= \exp(Q_{ij}) \sum_{l \neq j} \gamma_{il} \exp(Q_{il}) / (\sum_{l} \gamma_{il} \exp(Q_{il}))^2 \quad i = g; j = h
\end{aligned}$$

$$
\begin{aligned}
& = -\gamma_{ij}\exp(Q_{ij})\exp(Q_{ih})/(\sum_l \gamma_{il}\exp(Q_{il}))^2 \quad i=g; j\neq h \\
& = 0 \quad i\neq g
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial A'_{ij}(\mathbf{X})}{\partial \mu_{ghk}} & = \gamma_{ij}[\Sigma^{-1}(\mathbf{X}-\mu_{ij})^t]_k \exp(Q_{ij})\sum_{l\neq j}\gamma_{il}\exp(Q_{il})/(\sum_l \gamma_{il}\exp(Q_{il}))^2 \quad i=g; j=h \\
& = -\gamma_{ij}\gamma_{ih}\exp(Q_{ij})\exp(Q_{ih})[\Sigma^{-1}(\mathbf{X}-\mu_{ih})^t]_k/(\sum_l \gamma_{il}\exp(Q_{il}))^2 \quad i=g; j\neq h \\
& = 0 \quad i\neq g
\end{aligned}
$$

$\Psi(\theta'_R \mid \theta)$

When the independence model (equation 1) is used for $P(\mathbf{R}_t \mid S_t)$, $\Psi(\theta'_R \mid \theta)$ is maximized by

$$
p'_{si} = \frac{\sum_t \mathbf{f}_t(s)\mathbf{b}_t(s) r_t^i}{\sum_t \mathbf{f}_t(s)\mathbf{b}_t(s)} \tag{A4}
$$

for each station i and weather state s.

There is no closed form solution for maximizing $\Psi(\theta'_R \mid \theta)$ when the autologistic model (equation 2) is used so a numerical maximization procedure must be implemented. The derivatives of $\Psi(\theta'_R \mid \theta)$ with respect to α_{si} and β_{sij} are

True likelihood:

$$
\frac{\partial \Psi_R}{\partial \alpha_{si}} = \sum_t C_t(s)\left[r_t^i - \frac{\sum_{\mathbf{r}} r^i \exp(\sum_h \alpha_{sh} r^h + \sum_{hk}\beta_{shk} r^h r^k)}{\sum_{\mathbf{r}} \exp(\sum_h \alpha_{sh} r^h + \sum_{hk}\beta_{shk} r^h r^k)}\right] \tag{A5}
$$

$$
\frac{\partial \Psi_R}{\partial \beta_{sij}} = \frac{1}{2}\sum_t C_t(s)\left[r_t^i r_t^j - \frac{\sum_{\mathbf{r}} r_t^i r_t^j \exp(\sum_h \alpha_{sh} r^h + \sum_{hk}\beta_{shk} r^h r^k)}{\sum_{\mathbf{r}} \exp(\sum_h \alpha_{sh} r^h + \sum_{hk}\beta_{shk} r^h r^k)}\right] \tag{A6}
$$

Pseudolikelihood:

$$
\frac{\partial \Psi_R}{\partial \alpha_{si}} = \sum_t C_t(s)\left(r_t^i - \frac{\exp(\alpha_{si} + \sum_h \beta_{sih} r_t^h)}{1+\exp(\alpha_{si} + \sum_h \beta_{sih} r_t^h)}\right) \tag{A7}
$$

$$
\frac{\partial \Psi_R}{\partial \beta_{sij}} = \sum_t C_t(s)\left(r_t^i r_t^j - \frac{r_t^i r_t^j \exp(\alpha_{si} + \sum_h \beta_{sih} r_t^h)}{1+\exp(\alpha_{si} + \sum_h \beta_{sih} r_t^h)}\right) \tag{A8}
$$

References

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

[2] Baum, L.E., T. Petrie, G. Soules, N. Weiss (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statist., 41*, 164-171.

[3] Besag, J. (1975) Statistical analysis of non-lattice data. *Statistician, 24*, 179-195.

[4] Besag, J. (1977) Efficiency of pseudolikelihood estimation for simple Gaussian fields. *Biometrika, 64*, 616-618.

[5] Breiman L., Friedman J.H., Olshen R.A., and Stone, C.J. (1984) *Classification and Regression Trees*, Wadsworth International Group, Belmont CA.

[6] Charles, S.P., J.P. Hughes, B.C. Bates and T.J. Lyons (1996) Assessing downscaling models for atmospheric circulation — local precipitation linkage. *Proceedings of the International Conference on Water Resources and Environmental Research: Towards the 21st Century*, Tokyo, Japan.

[7] Cleveland, W.S., and Devlin, S.J., (1988) Locally-weighted regression: An approach to regression analysis by local fitting. *J. Am. Statist. Assoc., 83*, 596-610.

[8] Cressie, N.A.C. (1993) *Statistics for Spatial Data.* John Wiley and Sons, Inc., New York, New York, 900pp.

[9] Dempster, A.P., N.M. Laird, D.B. Rubin (1977) Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc. B, 39*, 1-38.

[10] Gabriel, K. R. and J. Neumann (1962) A Markov chain model for daily rainfall occurrences at Tel-Aviv. *Q. J. R. Meteorol. Soc., 88*, 85-90.

[11] Geyer, C.J., and E.A. Thompson (1992) Constrained Monte Carlo maximum likelihood for dependent data. *J. R. Statist. Soc. B, 54*, 657-699.

[12] Geman, S and Geman, D. (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *I.E.E.E. Trans. Pattern Anal. Machine Intell., 6*, 712-741.

[13] Giorgi, F. and L. O. Mearns (1991) Approaches to the simulation of regional climate change: A review. *Reviews of Geophysics, 29*, 191- 216.

[14] Grenander, U. (1989) Advances in pattern theory. *Annals of Statistics, 17*, 1-30.

[15] Grotch, S.L. and MacCracken, MC (1991) The use of general circultion models to predict climate change. *J. Climate, 4*, 286-303.

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

[17] Hughes, J. P. and P. Guttorp (1994a) A Class of Stochastic Models for Relating Synoptic Atmospheric Patterns to Regional Hydrologic Phenomena. *Water Resour. Res., 30*, 1535-1546.

[18] Hughes, J. P. and P. Guttorp (1994b) Incorporating Spatial Dependence and Atmospheric Data in a Model of Precipitation. *J. Applied Meteor., 33*, 1503-1515.

[19] Hughes, J. P. (1996) Computing the observed information in the hidden Markov model using the EM algorithm. *Probability and Statistics Letters*, in press.

[20] Intergovermental Panel on Climate Change (1995) *Climate Change 1995, The Science of Climate Change*, (ed. J.T. Houghton, L.G. Meira Filho, B.A. Callander, N. Harris, A. Kattenberg, and K. Maskell) Cambridge University Press, Cambridge.

[21] Juang, B.H. and L.R. Rabiner (1991) Hidden Markov models for speech recognition. *Technometrics, 33*, 251-272.

[22] Kass and Raftery (1995) Bayes factors. J. Am. Statist. Assoc., 90, 773-795.

[23] Kavvas, M. L. and J. W. Delleur (1975) The stochastic and chronological structure of rainfall sequences: Application to Indiana, *Water Resour. Res. Center Rep. 57,* Purdue Univ., West Lafayette, Ind.

[24] Kavvas, M. L. and J. W. Delleur (1981) A stochastic cluster model for daily rainfall sequences. *Water Resour. Res., 17*, 1151-1150.

[25] Kidson, J.W. (1994) The relation of New Zealand daily and monthly weather patterns to synoptic weather types. *Int. J. Climatol., 14*, 723-737.

[26] LeCam, L. (1961) A stochastic theory of precipitation. Fourth Berkeley Symposium on Mathematics, Statistics, and Probability, Univ. of Calif., Berkeley, Calif.

[27] Mearns, L.O., Giorgi, F., McDaniel, L. and Shields, C. (1995) Analysis of daily variability of precipitation in a nested regional climate model: comparison with observations and doubled CO2 results. *Glob. Planet. Change, 10*, 55-78.

[28] Rai and Matthews (1993) Improving the EM algorithm. *Biometrics, 49*, 587-591.

[29] Stern, R. D. and R. Coe (1984) A model fitting analysis of daily rainfall data. *J. R. Statist. Soc. A, 147*, 1-34.

[30] Titterington, D.M. (1990) Some recent research in the analysis of mixture distributions. *Statistics, 21*, 619-641.

[31] Waymire, E. and V. K. Gupta (1981) The mathematical structure of rainfall representations 2. A review of the theory of point processes. *Water Resour Res., 17*, 1273-1285.

List of Figures

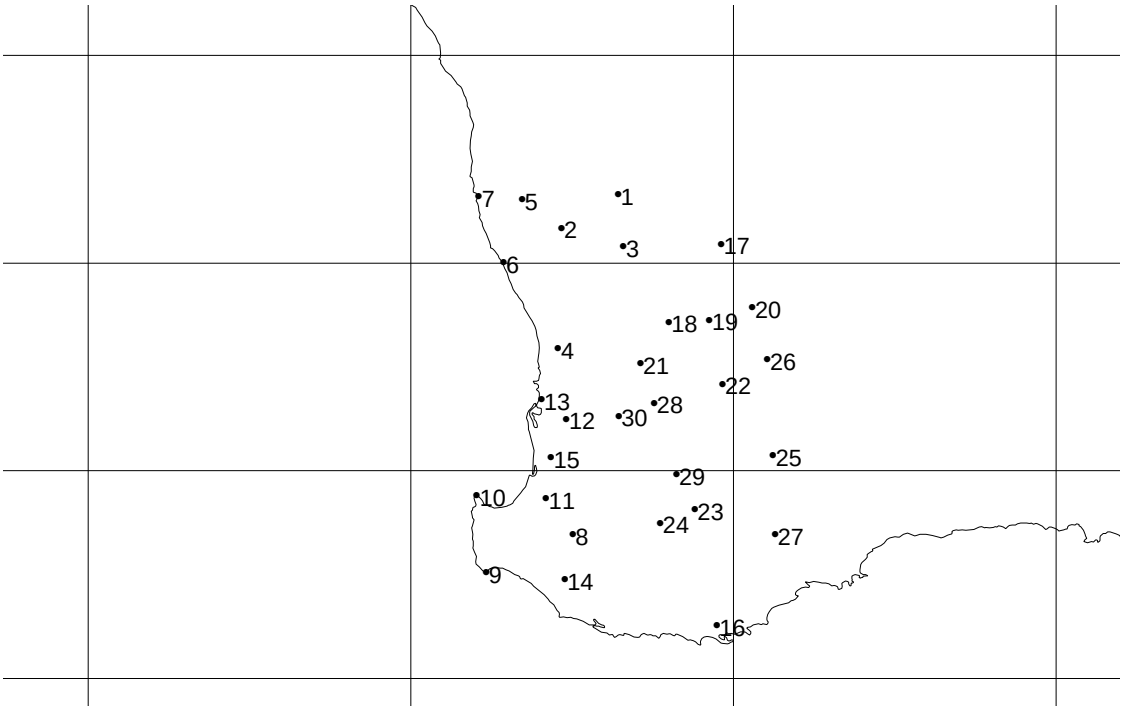

Figure 1: Map of study area showing the locations of the atmospheric data grid and rain gauge stations in southwestern Australia. Atmospheric data are interpolated to the verticies of the grid as described in the text.

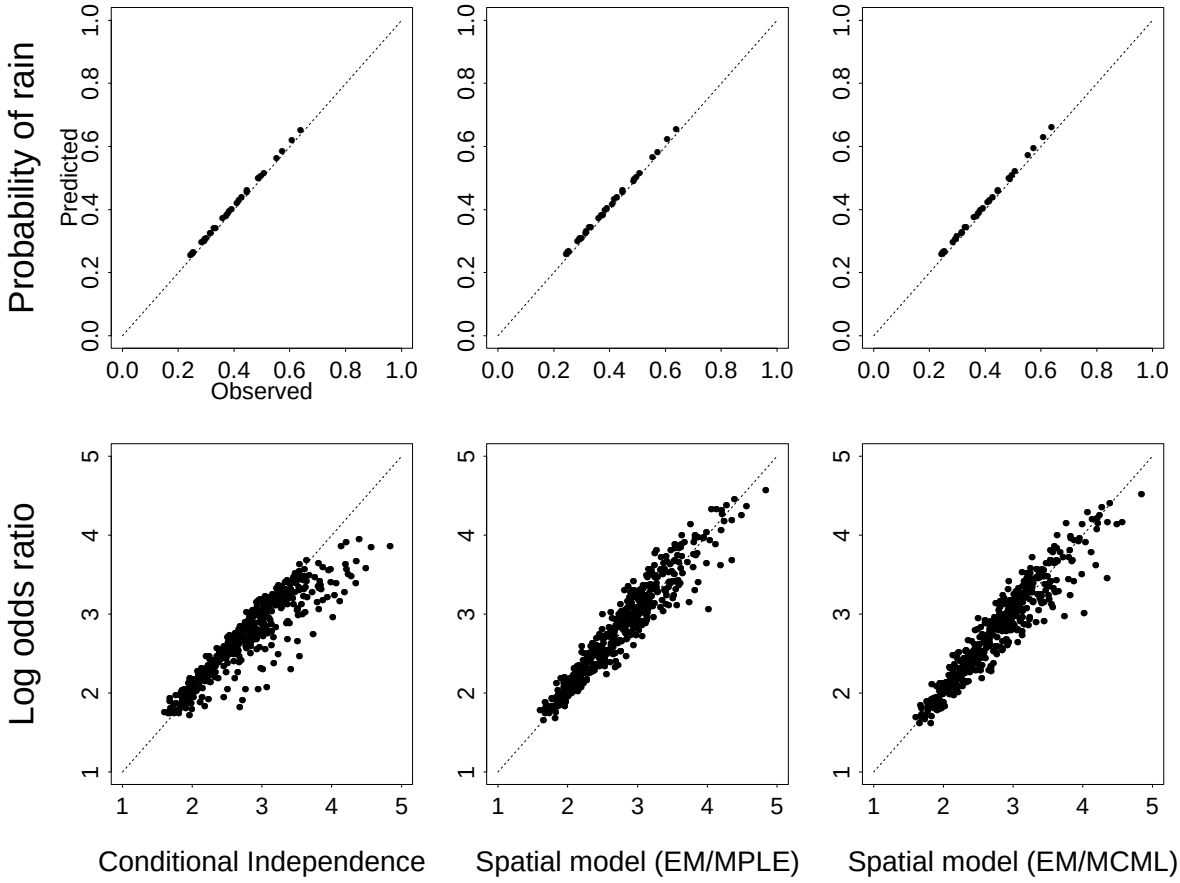

Figure 2: Comparison of observed and model-predicted rainfall statistics based on the 10 years of data used for model fitting. Model-predicted statistics are generated by simulation from the fitted model using the observed atmospheric data. Observed statistics are on the x-axis; model-predicted statistics are on the y-axis.

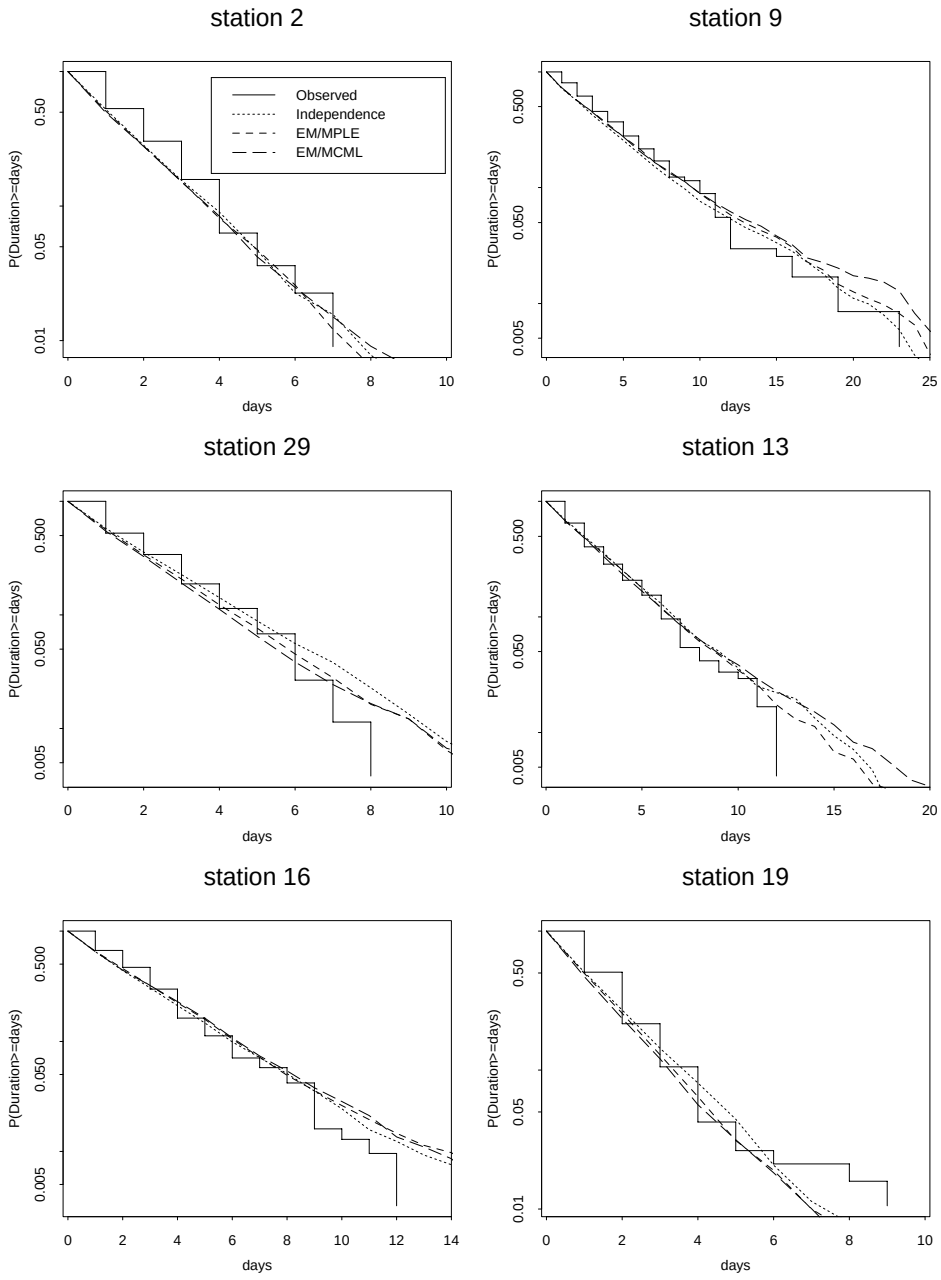

Figure 3: Comparison of observed and model-predicted rainfall statistics: duration distribution. Results are presented for 6 representative stations (see figure 1). Observed and model-predicted statistics are based on the 10 years of data used for model fitting.

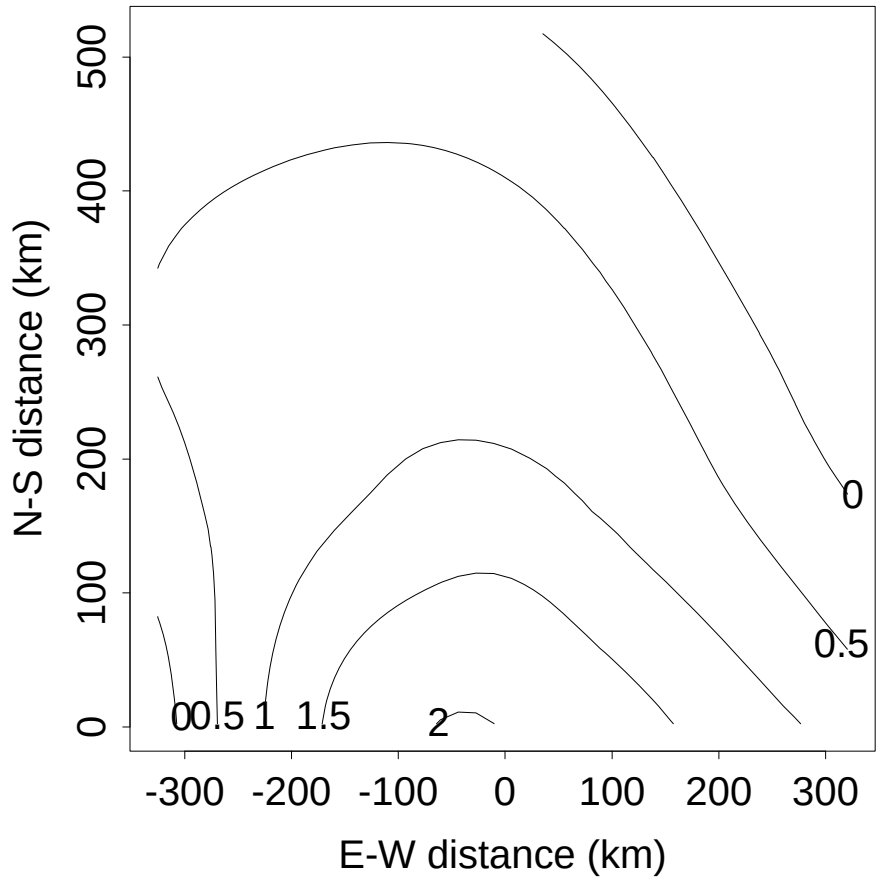

Figure 4: Empirical log-odds ratios as a function of the distance and direction between pairs of rain gauge stations for all days classified as weather state 2 using the 6 state, 3 atmospheric variable, conditional spatial independence model described in the text. The data were smoothed using the loess procedure (Cleveland and Devlin, 1988) prior to contouring. This plot is based on the 10 years of data used for model fitting.

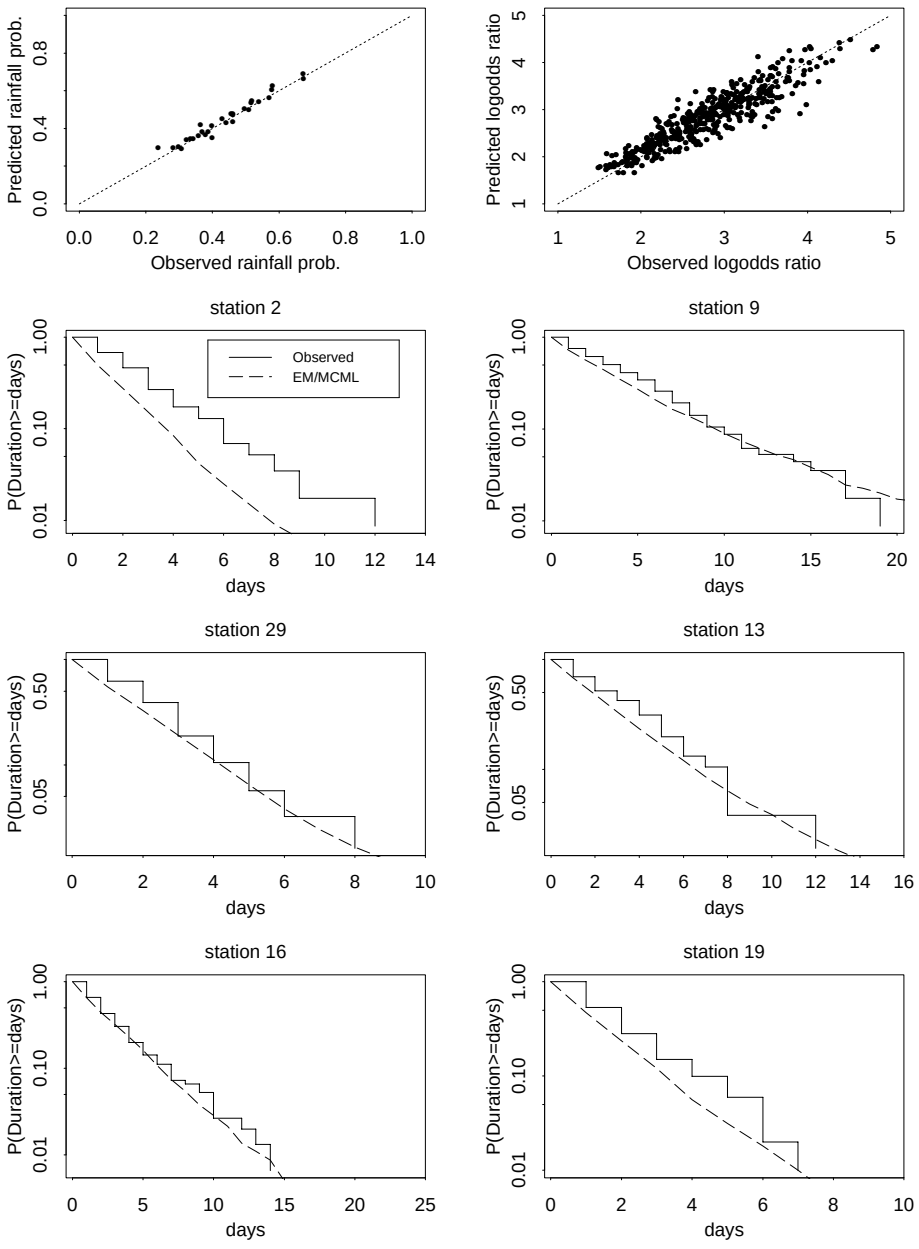

Figure 5: Comparison of observed and model-predicted rainfall statistics on the 5 years of reserved data. Model-predicted statistics are generated by simulation from the fitted model using the observed atmospheric data. Duration distributions are shown at a representative subset of stations. Station 2 represents the poorest fit seen.

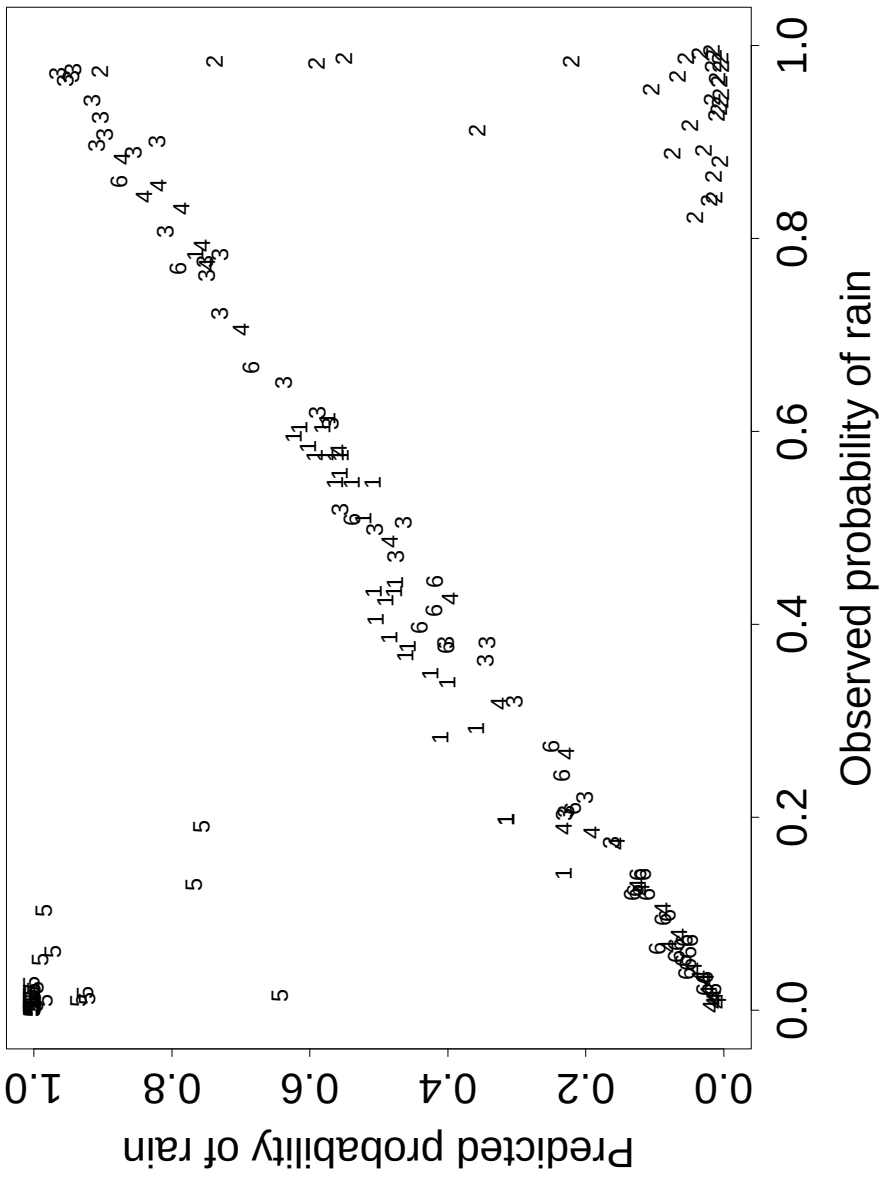

Figure 6: Observed probability of rain versus predicted probability of rain at each station in each weather state after one iteration of the EM/MPLE algorithm on the full 15 year dataset, starting at the best fitting conditional spatial independence model. Numbers correspond to the estimated weather states just prior to this iteration.

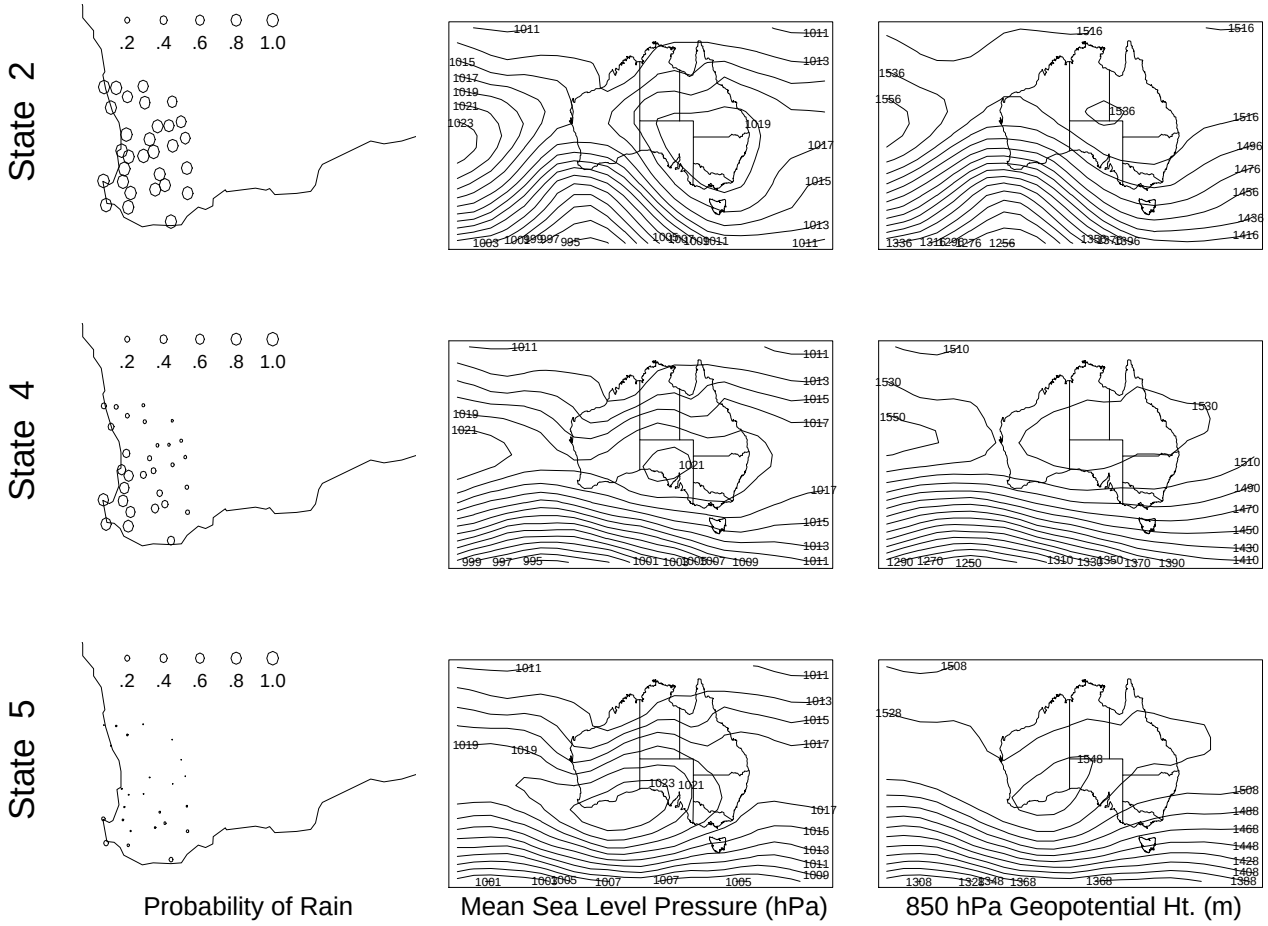

Figure 7: Probability of rain, composite sea-level pressure (hPa) and 850 hPa geopotential height (m) fields for three states from the six state spatial model estimated using EM/MCML. Each day is first classified into its most likely state using the Viterbi algorithm. All days in a particular state are then averaged at each station (for rainfall) or grid node (for the atmospheric variables) to obtain the composite fields.