Bayesian modelling of rainfall data using non–homogeneous hidden Markov models and latent Gaussian variables

Sarah E. Heaps, Richard J. Boys and Malcolm Farrow

School of Mathematics & Statistics, Newcastle University, Newcastle Upon Tyne, UK.

Summary. We present a non–homogeneous hidden Markov model (NHMM) for the spatio–temporal analysis of rainfall data, within a subjective Bayesian framework. In this model, daily rainfall patterns are driven by a small number of unobserved states, interpreted as states of the weather, that evolve in time according to a first order non–homogeneous Markov chain, with transition probabilities dependent upon time–varying atmospheric data. The weather states alone do not account for all the space time structure in the data and so we introduce latent multivariate normal random variables in a flexible model for the probability of rain and the distribution of non–zero rainfall amounts. In the resulting hierarchical NHMM, rainfall occurrences and non–zero rainfall amounts are spatially dependent and conditionally Markov in time, given the weather state. We build a prior distribution that conveys genuine initial beliefs and apply the model and inferential procedures to data from a network of twelve sites located throughout the UK.

Keywords: Non–homogeneous hidden Markov model, latent Gaussian variables, Bayesian spatio–temporal analysis, rainfall, statistical downscaling.

1. Introduction

Concerns about the potential effects of climate change in recent years have led to an increasing interest in the relationship between rainfall and climate. For a set of initial conditions, realistic simulations of the earth's atmosphere can be generated using general circulation models (GCMs). These are complex, deterministic, mathematical models of the circulation of the atmosphere. Typically the resolution of GCM output is on a spatial scale of around 2–5° of longitude and latitude. However, questions of scientific interest, for example in hydrology or agriculture, often concern local patterns of precipitation over a much finer spatial scale. Addressing these questions using GCM simulations therefore presents the problem of how to turn these simulations into fine scale predictions of rainfall. *Statistical downscaling* provides a solution, generally by developing stochastic models which link the synoptic (large scale) atmospheric variables and small scale precipitation fields; see Wilby and Wigley (1997), Fowler et al. (2007) or Maraun et al. (2010) for a review.

One class of statistical downscaling models is the *weather state model*, introduced by Hay et al. (1991). These models assign each day to one of a small number of weather states using observed atmospheric information. Typically these weather states are observable. Precipitation is then modelled conditionally on the weather state which is generally assumed to evolve according to some temporal process. Hughes and Guttorp (1994) proposed modelling

Address for correspondence: Richard Boys, School of Mathematics & Statistics, Newcastle University, Newcastle Upon Tyne, U.K.
E-mail: richard.boys@newcastle.ac.uk

rainfall using a *non–homogeneous hidden Markov model* (NHMM) which differs from the classic weather state model in that the weather states are not observable. In their NHMM, the observed atmospheric data enter the model as explanatory variables whose role is to influence the probability of transition from one weather state to the next.

Aggregated (e.g. daily) precipitation has a mixed distribution, with a point mass at zero and a positively skewed density function on the positive real line. This makes it difficult to construct models for precipitation which can accommodate its space–time structure. The development of statistical downscaling models therefore represents a problem which is both topical and challenging. There have been few attempts taking a Bayesian approach. In particular we are not aware of any Bayesian approaches to modelling precipitation using hidden Markov models (HMMs). In this paper we propose a NHMM for daily precipitation which we formulate in a fully Bayesian framework. This allows the evaluation of all posterior uncertainty as well as the incorporation of useful prior beliefs.

We consider a dataset of daily precipitation measurements over 28 consecutive winters at twelve sites in the UK. The atmospheric data used are *objective Lamb weather types* which provide a classification of synoptic weather into 27 categories, based on surface pressure around the British Isles. In order to capture the spatio–temporal dependence in the precipitation data we found that a rather sophisticated within–weather–state model for precipitation was required. Our proposed within–weather–state model uses a Markov chain of multivariate probit (MVP) models for precipitation occurrences. Non–zero precipitation amounts are given a multivariate lognormal distribution. The mean of the corresponding multivariate normal distribution depends linearly on the latent normal random vector underlying the MVP model. This offers a number of advantages over the truncated power transformed multivariate normal distribution which is often used for multi–site rainfall; see, for example, Sansó and Guenni (2000). This will be discussed further in Section 3.

The remainder of this paper is structured as follows. In Section 2 we introduce the UK dataset analysed in subsequent sections. Section 3 surveys the literature on the spatio–temporal modelling of rainfall, particularly through HMMs and latent Gaussian variable models. In Section 4 we develop and describe our NHMM for rainfall and a prior distribution for the unknowns in the model which allows the incorporation of genuine initial beliefs. Section 5 outlines the Markov chain Monte Carlo (MCMC) scheme for generating posterior samples for a model with a fixed number of states r and discusses posterior inference for r. Finally Section 6 applies the model and inferential procedures to the UK dataset, including a summary of the resulting posterior distribution and the results of model checks, which compare the posterior predictive distribution to data which were not used in model fitting.

The data that are analysed in the paper are available on request from the Met Office MIDAS database (2012).

2. UK winter rainfall data

The main objective of this paper is the development of a Bayesian statistical downscaling model for UK rainfall data. To this end, the precipitation data analysed in Section 6 are from a network of twelve sites located throughout the UK. In common with other work on HMMs for precipitation, we consider data from one season only, namely calendar winter (December–February). The dataset comprises 2,527 daily precipitation totals recorded at each of the twelve sites over the 28 winter periods from 1961/2 to 1988/9, which includes seven leap years. The sites were chosen to give good spatial coverage over the UK. Chapter

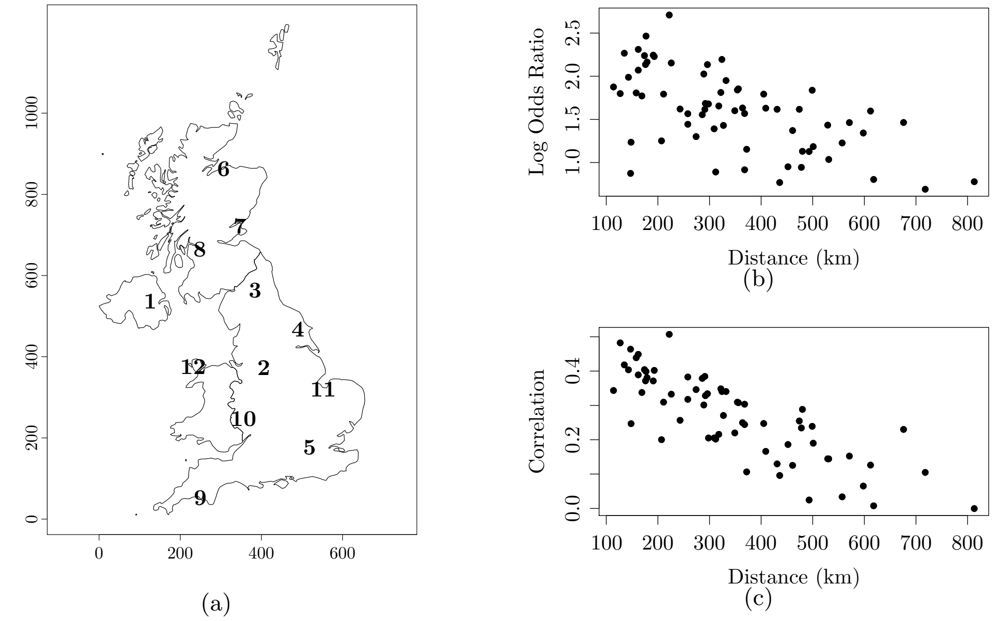

Fig. 1. (a) Locations of sites within the UK. Axes denote kilometres from the south western point of the British National Grid coordinate system (latitude $49°$ north, longitude $2°$ west). (b) Log odds ratios and (c) correlations against distance for this network.

7 of Germain (2010) provides complete details of an analogous application involving a more spatially dense network of sites, with missing values. Although all measurements refer to precipitation, the term rainfall will be used synonymously in the remainder of this paper.

Figure 1(a) shows the locations of the sites whilst Table 1 shows summaries of the proportion of wet days, precipitation on those wet days and the altitudes of the sites. The distances between sites range from 113.6 km to 813.1 km. In spite of these reasonably large distances, there is still clear spatial structure present in the data, with sites in closer proximity showing more similarity than those further apart. To illustrate, Figure 1 also displays plots of the log–odds ratios for occurrences and rank correlations for amounts against the distance between sites, both of which show a clear decreasing trend. Here the log–odds ratio for two sites i and j is $\log[\{n_{0,0}(i,j)n_{1,1}(i,j)\}/\{n_{0,1}(i,j)n_{1,0}(i,j)\}]$ where $n_{d^i,d^j}(i,j)$ is the observed number of days where the rainfall occurrence indicator is equal to d^i at site i and d^j at site j, where $d^k = 1$ if there is non–zero precipitation at site k and $d^k = 0$ otherwise.

In UK climatology, *objective Lamb weather types* (LWTs) have been used extensively for characterising atmospheric circulation patterns, making them a natural choice of atmospheric variable in downscaling models; see, for example, Fowler and Kilsby (2002), Conway and Jones (1998), and Bardossy and Plate (1992) who use *Grosswetterlagen*, the German equivalent. In this work, we investigate the use of LWTs in our NHMM for UK rainfall.

Lamb (1972) developed a subjective weather type classification scheme based on daily synoptic charts which depict the state of atmospheric flow over the British Isles at surface level and at a specified height in the atmosphere. Under this scheme an expert analyst can

Table 1. Summary of data from the UK network for winter periods from 1961/2 to 1988/9. The mean and coefficient of variation for daily precipitation are based only on wet days.

	Site	Altitude (m)	Proportion wet days (%)	Mean daily precip. (mm)	Coefficient of variation
1	Aldergrove	68	63.4	3.812	1.141
2	Buxton	307	65.7	6.201	1.149
3	Haydon Bridge	82	56.9	3.354	1.308
4	Highmow	175	57.5	3.684	1.288
5	Kew Gardens	6	43.9	3.375	1.086
6	Kinloss	5	53.9	2.914	1.330
7	Leuchars	10	50.4	3.683	1.300
8	Paisley	32	61.9	5.624	1.125
9	Plymouth	50	57.8	6.035	1.065
10	Preston Wynne	84	51.3	3.611	1.134
11	Terrington St. Clement	2	51.8	2.852	1.217
12	Valley	10	57.9	4.311	1.158

use their judgement to determine the weather type on any day in order to give an indication of the daily steering of circulation systems. Jenkinson and Collinson (1977) developed an automated (sometimes called "objective") method for identifying these LWTs using daily gridded mean–sea–level pressure charts. From these data it is possible to calculate estimates of the dominant direction and speed of the flow, as well as its vorticity. Particular values of these measures are then associated with specific LWTs so that the classification provides a categorisation of the direction and synoptic type of the surface flow over the British Isles.

The Jenkinson classification scheme contains eight main directional types: north (N), north–east (NE) *etc*; and three main non-directional types: anticyclonic (A), cyclonic (C) and unclassifiable (U). A further 16 hybrid types combine the eight main directional types with the anticyclonic or cyclonic non–directional type. This gives 27 objective LWTs, which are shown in Table 2. Days on which the vorticity is low and the flow is from the west, for example, will be classified as westerly types, whilst days on which the vorticity is strongly positive or negative will be categorised as cyclonic or anticyclonic, respectively. When the vorticity is only moderately positive or negative, the direction of air flow is also used to provide the classification into one of the hybrid types. Type U is provided for days on which the circulation is too complicated for it to be classified as any of the other types.

The objective Lamb classification scheme has been used to classify the weather type over the British Isles for every day from 1880 to the present. The frequencies of their occurrence in winters from 1961/2 to 1988/9 can be seen in Table 2. The most commonly occurring LWTs are pure anticyclonic (type 1), pure westerly (type 15) and pure cyclonic (type 18).

Preliminary graphical analysis using heat maps (see Supplementary Materials) shows clear patterns in the proportion of wet days. Lower proportions are associated with the anticyclonic types (1–9) and higher proportions with the cyclonic types (18–26). This pattern is also seen to a lesser extent in the variation in mean wet day precipitation amounts across LWTs where high (low) amounts are typically associated with cyclonic (anticyclonic) types. Some sites also show particular relationships with the directional classification of the LWTs. For example, at Buxton, a high elevation site in the Pennine Hills, higher wet day

Table 2. Labelling of the objective Lamb weather types and frequencies 1961/2-1988/9.

Label	LWT	Frequency	Label	LWT	Frequency	Label	LWT	Frequency
1	A	437	27	U	9	18	C	307
2	ANE	15	10	NE	41	19	CNE	5
3	AE	34	11	E	66	20	CE	13
4	ASE	38	12	SE	76	21	CSE	32
5	AS	58	13	S	186	22	CS	47
6	ASW	71	14	SW	228	23	CSW	73
7	AW	85	15	W	310	24	CW	58
8	ANW	51	16	NW	140	25	CNW	41
9	AN	23	17	N	68	26	CN	15

precipitation amounts are associated with LWTs 6–7, 13–16 and 22–25, the majority of which are westerly types. A different pattern is observed at Plymouth, a low-altitude site on the South coast, where it seems that southerly and easterly LWTs tend to be associated with heavier precipitation.

3. Spatio–temporal rainfall modelling through empirical statistical models

In the terminology of Cox and Isham (1994) the model presented in this paper is an *empirical statistical model* for rainfall, describing only the observations of aggregated precipitation, and not the underlying physical phenomena. Spatio–temporal models of this type must address the complication arising from the mixed nature of rainfall distributions. This is generally achieved through the introduction of weather states or by using latent Gaussian variables, often with truncation and transformation. In this section we survey the literature on spatio–temporal empirical statistical models for (broadly) daily precipitation.

The weather state model was first introduced by Hay et al. (1991). Each day is assigned to one of a finite number of weather states and then precipitation is modelled conditionally on the weather state. Classically the weather states are observable given atmospheric data. The weather states are then assumed to evolve according to some temporal model, for example, a homogeneous first order Markov chain (Katz and Parlange, 1993, 1996), a homogeneous semi–Markov chain (Bardossy and Plate, 1992; Fowler et al., 2000) or a non–homogeneous first order Markov chain with transition probabilities dependent on time–varying covariates (Vrac et al., 2007). It is often assumed that the weather state explains most of the space time structure in the data and this allows reasonably simple spatio–temporal structures to be adopted for the within–weather–state distributions.

Alternatively, the weather state is introduced as an unobserved ("hidden") variable which evolves in time as a homogeneous first order Markov chain, in a HMM. Compared to observed–weather–state models, this has the benefit of allowing the states themselves to define precipitation patterns, which should therefore provide a good description of the spatio–temporal structure in the data. However, this is at the cost of the potential loss of the meteorological interpretation of the states, and an increase in model complexity. Special cases of (two–state) HMMs for precipitation occurrence were presented by Foufoula-Georgiou (1987) and Smith (1987). HMMs were later introduced formally as a general means of modelling single and multi–site precipitation occurrence data by Zucchini and

Guttorp (1991). In their model, precipitation occurrences were assumed to be conditionally independent across the spatial network, given the weather state. In this and most other HMMs in the literature, to account for seasonality, parameters are assumed to be constant within, but different across, seasons or months. In this respect, the non–stationary two state HMM proposed by MacDonald and Zucchini (1997) is unique in allowing the logit of the hidden state transition probabilities to vary smoothly across seasons using partial sums of Fourier series. Hughes and Guttorp (1994) attempted to provide a link between large–scale atmospheric measures and small–scale precipitation fields by incorporating atmospheric explanatory variables in the transition probabilities of the HMM. Since the atmospheric variables were time–varying, this led to a NHMM. Hughes et al. (1999) presented a more sophisticated NHMM for precipitation occurrence in which within–weather–state spatial dependence was modelled explicitly using an autologistic model.

Extensions to HMMs to include precipitation amounts have been proposed by Charles et al. (1999), Bellone et al. (2000), Betro et al. (2008) and Ailliot et al. (2009). Charles et al. (1999) used a NHMM for precipitation occurrence to identify the most likely sequence of hidden states. Amounts were then introduced *a posteriori* by conditioning on this sequence and fitting a regression model with weather state specific parameters and precipitation occurrence at neighbouring sites as regressors. Other approaches have taken a more unified approach by modelling the precipitation occurrences and amounts jointly, thereby allowing both to influence the characteristics of the weather states. A simple model for the non–zero precipitation amounts is to assume them to be conditionally independent across time and space, given occurrence and the weather state. This assumption was adopted by Bellone et al. (2000) and Betro et al. (2008) who, respectively, chose gamma and a mixture of Weibull distributions for the precipitation amounts on wet days. Thompson et al. (2007) presented a three state (partially) hidden Markov model which included an observable dry state. This HMM differs from those discussed so far in that it is a local, as opposed to regional, weather state model in which each site is associated with its own state. A separate HMM is defined marginally for each site, then spatial dependence in both the hidden and observed (given hidden) processes is built using copulas. Finally, Ailliot et al. (2009) presented a HMM in which spatial dependence in the within–weather–state joint distributions for precipitation occurrence and amount was modelled explicitly using a *truncated, power transformed multivariate normal* (TPTMVN) distribution.

Truncating and transforming (partially) latent Gaussian variables is commonly used for inducing spatial covariance structure in mixed rainfall distributions. The basic idea is to define $W^i = \mathbb{I}(Z^i > \alpha_0) g_1(Z^i, \boldsymbol{\alpha}_i),\ i = 1, \ldots, n$, where $\mathbb{I}(A) = 1$ if A is true and 0 otherwise, $\boldsymbol{W} = (W^1, \ldots, W^n)^{\mathrm{T}}$ denotes a vector of precipitation amounts at a network of n sites, $\boldsymbol{Z} = (Z^1, \ldots, Z^n)^{\mathrm{T}}$ is a multivariate normal random vector, α_0 is a threshold parameter and $g_1(Z^i, \boldsymbol{\alpha}_i)$ is a transformation function which may depend on parameters $\boldsymbol{\alpha}_i$. Typically $\alpha_0 = 0$ and $g_1(Z^i, \boldsymbol{\alpha}_i)$ is a strictly increasing function so that Z^i is observable on wet days but on dry days we just observe that $Z^i < 0$. Often $\boldsymbol{\alpha}_i = \alpha_i$ and $g_1(Z^i, \boldsymbol{\alpha}_i) = (Z^i)^{\alpha_i}$. This induces a heavy–tailed distribution for non–zero rainfall and gives $\boldsymbol{W}$ a TPTMVN distribution. Ailliot et al. (2009) introduced temporal dependence in their model by embedding the TPTMVN distribution in a HMM. Other authors have taken different approaches, for example by incorporating $\boldsymbol{Z}$ in a vector autoregression (Bardossy and Plate, 1992), seasonal multivariate dynamic linear model (Sansó and Guenni, 2000) or Gaussian Markov random field (Allcroft and Glasbey, 2003). In each case a single multivariate normal random vector is used to induce a joint distribution for rainfall occurrences and non–zero amounts.

In other latent Gaussian variable models for rainfall, the unobserved normal variables have been used as random effects. Velarde et al. (2004) developed a model in which precipitation occurrences and amounts, given occurrences, were assumed to be conditionally independent over space, given some spatially varying random effects. Seasonality and short term temporal structure were captured by allowing the logit of the occurrence probability at each site and the logarithm of the parameter in the exponential distribution for non–zero amounts to depend linearly on lagged precipitation occurrences at that site. Spatial structure was then incorporated through spatial effects in the linear predictors.

There appears to have been a relatively slow uptake of the Bayesian approach in precipitation modelling. Notable exceptions include Sansó and Guenni (1999a,b, 2000) and Velarde et al. (2004). In this paper we contribute to this small literature and build on the NHMMs proposed by, for example, Hughes et al. (1999) and Bellone et al. (2000). The following section describes a novel hierarchical NHMM which uses latent multivariate normal variables, as well as weather states, to model spatio–temporal dependence. The latent normal variables enter through the expression $W^i = \mathbb{I}(Z_0^i > 0) g_2(Z_1^i)$ where $\boldsymbol{Z}_0 = (Z_0^1, \ldots, Z_0^n)^{\mathrm{T}}$ and $\boldsymbol{Z}_1 = (Z_1^1, \ldots, Z_1^n)^{\mathrm{T}}$ are two correlated multivariate normal vectors. This approach offers potential advantages over the TPTMVN distribution with its single normal vector.

4. Model and prior

Consider a network of n sites. Let $\boldsymbol{D}_t = (D_t^1, \ldots, D_t^n)^{\mathrm{T}}$ be a random vector for rainfall occurrences where $D_t^i = 1$ if there is at least c mm of rain on day t at site i and $D_t^i = 0$ otherwise, for some suitable cut–off c mm. According to Glickman (2000), in British climatology a *rain day* is defined with $c = 0.2$ mm. We use this standard value and note that we have found inference to be insensitive to this choice. Let $\boldsymbol{W}_t = (W_t^1, \ldots, W_t^n)^{\mathrm{T}}$ be an n–dimensional random vector for rainfall amounts on day t where we set $W_t^i = 0$ if $D_t^i = 0$, and let the collections of observed values of $\boldsymbol{D}_t$ and $\boldsymbol{W}_t$ be $\boldsymbol{w}$ and $\boldsymbol{d}$.

The number $r \in \{1, \ldots, r_{\max}\}$ of weather states in the NHMM is not known *a priori*. We formulate our NHMM with the likelihood specified conditionally on a fixed number r of states, $p(\boldsymbol{w}, \boldsymbol{d} | \boldsymbol{\theta}_r, r)$. Each of the $r_{\max}$ (conditional) likelihoods has its own set of parameters, $\boldsymbol{\theta}_r$, to which we assign a prior, $\pi(\boldsymbol{\theta}_r | r)$. In Section 4.1 we describe the likelihood (i.e. the model) and then in Section 4.2, the priors $\pi(\boldsymbol{\theta}_r | r)$, $r \in \{1, \ldots, r_{\max}\}$. For notational clarity, unless stated otherwise, dependence on r is assumed without explicit notational reference. For example, we generally refer to $p(\boldsymbol{w}, \boldsymbol{d} | \boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta})$ rather than $p(\boldsymbol{w}, \boldsymbol{d} | \boldsymbol{\theta}_r, r)$ and $\pi(\boldsymbol{\theta}_r | r)$. Inference for the value of r is discussed in Section 5.2.

4.1. Model specification

4.1.1. Distribution for the weather states given the atmospheric data

Let $\{S_t : t = 0, \ldots, T\}$ denote a *hidden* or unobservable discrete–valued stochastic process which categorises the rainfall patterns on each day. We interpret S_t as the *weather state* on day t and denote its state space by $\mathcal{S}_r = \{1, \ldots, r\}$. Denote by $\boldsymbol{X}_t$ the atmospheric data on day t. Both S_t and $\boldsymbol{X}_t$ are common to all sites in the network. In their NHMMs for rainfall, other authors (e.g. Hughes and Guttorp, 1994; Hughes et al., 1999; Bellone et al., 2000) have used continuous atmospheric data, typically comprising linear combinations of high dimensional atmospheric fields (e.g. sea–level pressure) so that $\boldsymbol{X}_t$ provides a summary of atmospheric conditions over the region of interest on day t. In our case $\boldsymbol{X}_t = X_t \in \{1, \ldots, 27\}$ is a categorical covariate, namely the observed LWT on day t, labelled according to Table 2.

We denote the parameters of our NHMM by $\boldsymbol{\theta} = (\boldsymbol{\theta}_{\text{hid}}, \boldsymbol{\theta}_{\text{obs}})$, partitioned so that $\boldsymbol{\theta}_{\text{hid}}$ parameterises the weather state process and $\boldsymbol{\theta}_{\text{obs}}$ parameterises the observed process. Let $\boldsymbol{y}_{i:j}$ denote the sequence $y_i, y_{i+1}, \ldots, y_j$. We develop the temporal structure of our NHMM hierarchically beginning with the following assumption for the weather states

$$\Pr(S_t = k | \boldsymbol{S}_{0:t-1}, \boldsymbol{X}_{1:T}, \boldsymbol{\theta}_{\text{hid}}) = \Pr(S_t = k | S_{t-1} = j, X_t = x, \Lambda) = \lambda^x_{j,k}, \tag{1}$$

for $j, k \in \mathcal{S}_r$, $x \in \{1, \ldots, 27\}$ and $t = 1, \ldots, T$ with $\Pr(S_0 = k | \boldsymbol{X}_{1:T}, \boldsymbol{\theta}_{\text{hid}}) = \Pr(S_0 = k | \boldsymbol{\nu}) = \nu_k$. We denote the initial distribution by $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_r)$ and write $\boldsymbol{\lambda}^x_j = (\lambda^x_{j,1}, \ldots, \lambda^x_{j,r})$ for every pair (j, x) where $j \in \mathcal{S}_r$, $x \in \{1, \ldots, 27\}$. In our application we model the rainfall data from each winter period as an independent realisation of the same NHMM and there is, therefore, some potential to learn about the initial distribution $\boldsymbol{\nu}$. We note that (1) describes a conditional distribution for the weather state S_t given S_{t-1} and the LWT X_t. In statistical downscaling, a GCM would typically be used to generate a projected time series of LWTs. Our model could then be used to predict rainfall conditionally on this time series.

If we neglected the conditioning on atmospheric data in (1), then we would simply have the Markov assumption for the weather states. The role of the atmospheric data is to adjust the transition probabilities which would prevail in a homogeneous model in light of the current atmospheric information X_t. Therefore, through their influence on patterns of rainfall, we might expect different atmospheric conditions to be associated with different weather states. From (1) it can be seen that we have a model for the weather states such that, for every combination of lag–one weather state $S_{t-1} = j$ and current LWT $X_t = x$, a different stochastic vector $\boldsymbol{\lambda}^x_j$ governs the probabilities of transition into the current weather state. This parameterisation offers the possibility of a conjugate Dirichlet prior distribution for each $\boldsymbol{\lambda}^x_j$. We denote the collection of transition probabilities by $\Lambda = (\boldsymbol{\lambda}^1_1, \boldsymbol{\lambda}^2_1, \ldots, \boldsymbol{\lambda}^{27}_r)$ and so $\boldsymbol{\theta}_{\text{hid}} = (\Lambda, \boldsymbol{\nu})$. As some of the LWTs occur very infrequently (see Table 2), a prior for the $\boldsymbol{\lambda}^x_j$ which encourages borrowing of strength between LWTs will be necessary.

4.1.2. Distribution for the observations given the weather states

The distribution of observed rainfall on day t at site i is mixed, with a positive probability of zero rainfall and a continuous distribution over positive values. Exploratory analysis suggested that the conditional density of rainfall amount, given that it is non–zero, should go to zero as the amount goes to zero. We have adopted a distribution which satisfies this requirement, namely the lognormal distribution. To build a flexible model for the way that the probability of zero and the distribution of non–zero amounts are related over time and between sites, we introduce two correlated multivariate normal random vectors $\boldsymbol{Z}_{0,t} = (Z^1_{0,t}, \ldots, Z^n_{0,t})^{\text{T}}$ and $\boldsymbol{Z}_{1,t} = (Z^1_{1,t}, \ldots, Z^n_{1,t})^{\text{T}}$ for $t = 1, \ldots, T$. We define the observable rainfall at site i on day t as

$$W^i_t = \mathbb{I}(Z^i_{0,t} > 0) \exp(Z^i_{1,t}) = D^i_t \exp(Z^i_{1,t}).$$

Thus the sign of $Z^i_{0,t}$ determines the occurrence or otherwise of rain. Its value, apart from the sign, is not observed but it helps to carry the correlation structure. When $W^i_t > 0$, we have $Z^i_{1,t} = \log W^i_t$; otherwise $Z^i_{1,t}$ plays no role. Rappold et al. (2008) express rainfall as a function of two normal variables in the same way in their spatio–temporal model for wet mercury deposition.

In the literature, other HMMs for rainfall have relied on the temporal dynamics of the hidden states to capture all the temporal autocorrelation in the rainfall data. In the model described here, this would correspond to an assumption that the bivariate latent process $\{(\boldsymbol{Z}_{0,t}, \boldsymbol{Z}_{1,t})\}$ is conditionally independent across time t given the weather state. However, in earlier applications involving UK rainfall data (see Germain, 2010, Chapter 4) we found that models with this temporal structure were often unable to predict the longer duration wet and dry spells that were observed in the data. We therefore allow a refinement to the temporal dependence structure so that

$$p(\boldsymbol{z}_{0,t}, \boldsymbol{z}_{1,t}|\boldsymbol{z}_{0,1:t-1},\boldsymbol{z}_{1,1:t-1}, D_0, \boldsymbol{S}_{0:T}, \boldsymbol{X}_{1:T}, \boldsymbol{\theta}_{\text{obs}}) \\ = p(\boldsymbol{z}_{0,t}|\boldsymbol{d}_{t-1}, S_t = k, \boldsymbol{\theta}_{\text{obs}})p(\boldsymbol{z}_{1,t}|\boldsymbol{z}_{0,t}, S_t = k, \boldsymbol{\theta}_{\text{obs}}),$$

for $t = 1, \ldots, T$, with a simple initial model

$$\Pr(\boldsymbol{D}_0|S_{0:T}, \boldsymbol{X}_{1:T}, \boldsymbol{\theta}_{\text{obs}}) = \prod_{i=1}^{n} \Pr(D_0^i) \quad \text{with} \quad D_0^i \sim \text{Bern}(p_i),$$

where each $p_i \in [0, 1]$ is fixed. Note that $\boldsymbol{Z}_{0,t}$ depends on the previous day's rainfall occurrence indicator $\boldsymbol{D}_{t-1}$. We specify

$$\boldsymbol{Z}_{0,t}|\boldsymbol{D}_{t-1} = \boldsymbol{d}_{t-1}, S_t = k, \boldsymbol{\theta}_{\text{obs}} \sim \text{N}_n(\boldsymbol{\mu}_{t,k}, \Sigma_k), \tag{2}$$

where Σ_k is a $n \times n$ symmetric positive definite matrix, $\boldsymbol{\mu}_{t,k} = (\mu_{t,k}^1, \ldots, \mu_{t,k}^n)^{\text{T}}$ and $\mu_{t,k}^i = \beta_{0,k}^i + \beta_{1,k}^i d_{t-1}^i$. Finally, given $\boldsymbol{Z}_{0,t}$ and S_t, the (partially) latent process of log rainfall amounts, $\{\boldsymbol{Z}_{1,t}\}$, are conditionally independent across time with

$$\boldsymbol{Z}_{1,t}|\boldsymbol{Z}_{0,t}, S_t = k, \boldsymbol{\theta}_{\text{obs}} \sim \text{N}_n(\boldsymbol{\alpha}_k + \Gamma_k \boldsymbol{Z}_{0,t} \,, \Omega_k),$$

where $\boldsymbol{\alpha}_k$ is a n–vector, Γ_k is a $n \times n$ matrix and Ω_k is a $n \times n$ symmetric positive definite matrix. Experience has shown that unless the number of sites n is small, allowing a different Γ_k matrix for each state k compromises the performance of the MCMC sampler. We therefore assume a constant matrix $\Gamma_k = \Gamma$ for all $k \in \mathcal{S}_r$. In the special case where each element in Γ is zero we obtain a model in which changes in the probability of rainfall occurrence have no effect on the distribution of non–zero rainfall amounts and *vice versa*.

It is straightforward to show that $\text{Var}(\boldsymbol{Z}_{1,t}|\boldsymbol{D}_{t-1} = \boldsymbol{d}_{t-1}, S_t = k, \boldsymbol{\theta}_{\text{obs}}) = \Omega_k + \Gamma\Sigma_k\Gamma^{\text{T}}$ and so because Γ is non–diagonal, we can make the covariance matrices Ω_k diagonal and still allow within–state spatial dependence between the elements of $\boldsymbol{Z}_{1,t}$. Therefore, in order to create a more parsimonious model we adopt this simplification and denote $\Omega_k = \text{diag}(\omega_{k,1}^2, \ldots, \omega_{k,n}^2)$. It follows that $(Z_{1,t}^1, \ldots, Z_{1,t}^n)$ are conditionally independent given $\boldsymbol{Z}_{0,t}$ and S_t and all the within–state spatial dependence is carried by $\boldsymbol{Z}_{0,t}$. Note however that $Z_{1,t}^i$ is not conditionally independent of other sites given just $Z_{0,t}^i$ and this captures the fact that, under some conditions, the amount of rain we might expect at site i, if it does rain, might be related to whether it also rains at some other sites.

As an alternative, we could have formulated our model by defining a marginal distribution for $\boldsymbol{Z}_{1,t}$ and then a conditional distribution for $\boldsymbol{Z}_{0,t}$ given $\boldsymbol{Z}_{1,t}$ in which the $Z_{0,t}^i$ were conditionally independent. In this case all of the within–state spatial dependence would have been captured by $\boldsymbol{Z}_{1,t}$. However, we chose the our formulation because rainfall modellers usually specify a distribution for rainfall occurrence and then a distribution for rainfall amounts given occurrence. Allowing $\boldsymbol{Z}_{0,t}$ to carry the spatial dependence enables us to represent the within–state model in similar terms as a model for rainfall occurrence (2) and then a model for rainfall amounts, given occurrences (and $\{\boldsymbol{Z}_{0,t}\}$) through

$$p(\boldsymbol{z}_{1,t}|s_t, \boldsymbol{z}_{0,t}, \boldsymbol{\theta}_{\text{obs}}) = \prod_{i=1}^{n} p(z_{1,t}^i|s_t, \boldsymbol{z}_{0,t}, \boldsymbol{\theta}_{\text{obs}}), \quad \text{where} \tag{3}$$

$$Z^i_{1,t}|S_t = k, \boldsymbol{Z}_{0,t} = \boldsymbol{z}_{0,t} \sim \mathrm{N}\left(\alpha^i_k + \sum_{j=1}^{n} \Gamma^{i,j} z^j_{0,t}\,,\, \omega^2_{k,i}\right) \quad \text{and} \quad W^i_t = D^i_t \exp(Z^i_{1,t}). \tag{4}$$

This is the representation of the within–state–model on which we will focus for the remainder of this paper.

It remains to introduce an identifiability constraint for the parameters in the occurrence model. If we were modelling rainfall occurrences only and therefore omitted $\boldsymbol{W}_t$, the model defined above would reduce to a Markov chain of multivariate probit (MVP) models, conditional on the weather state. To ensure parameter identifiability in the observed data likelihood of MVP models, constraints are necessary to prevent arbitrary rescaling of the linear predictor $\boldsymbol{\mu}_{t,k}$ or the covariance matrix Σ_k. Although our model describes rainfall amounts, as well as occurrences, this problem of non–identifiability is not remedied because changes to the scale of $\boldsymbol{Z}_{0,t}$ could exactly compensate for changes in the scale of Γ. However once the scale and location of $\boldsymbol{Z}_{0,t}$ is fixed, all parameters are identifiable; see the Supplementary Materials for further explanation and a numerical demonstration.

With MVP models, a common means of fixing the scale and location of $\boldsymbol{Z}_{0,t}$ is to constrain the covariance matrix to be a correlation matrix. However there are two main problems with this approach. Specifying a meaningful prior for a correlation matrix is difficult due to the complex constraints on the space of correlation matrices. These constraints also make sampling a correlation matrix during MCMC challenging, although efficient MCMC schemes have been developed which use the related ideas of parameter expansion (Liu, 2001; Liu and Daniels, 2006), marginal data augmentation (Berrett and Calder, 2012) and the introduction of dummy parameters (Zhang et al., 2006). An alternative solution would have been to decompose the covariance matrix according to the square–root free Cholesky decomposition of the precision matrix Σ_k^{-1} (as described in Section 4.2) and then to fix the values of the conditional variances arising from this reparameterisation; see, for example, Webb and Forster (2008). However, we found that this led to poor mixing during MCMC. We therefore avoid placing constraints on the covariance matrices, Σ_k, and instead fix the scale of the coefficients $\boldsymbol{\beta}_{1,k} = (\beta^1_{1,k}, \ldots, \beta^n_{1,k})^{\mathrm{T}}$ so that $\beta^i_{1,k} \in \{-1, 1\}$ for each $i = 1, \ldots, n$ and each $k = 1, \ldots, r$. Note that in weather state k, the marginal site–i probabilities of rain after no rain and rain after rain are $\Phi(\beta^i_{0,k}/\sqrt{\Sigma^{i,i}_k})$ and $\Phi(\beta^i_{0,k}/\sqrt{\Sigma^{i,i}_k} + \beta^i_{1,k}/\sqrt{\Sigma^{i,i}_k})$ respectively. Therefore our constraint still allows the probability of rain after rain to be (any amount) more than ($\beta^i_{1,k} = 1$), less than ($\beta^i_{1,k} = -1$) or equal to ($\Sigma^{i,i}_k$ large) the probability of rain after no rain.

We note that, while the transition probabilities $\lambda^x_{j,k}$ for changes between weather states described in 4.1.1 are, of course, common to all sites, the within-state model described in this section involves site-specific parameters so that the rainfall behaviour within a given weather state can vary between sites.

4.1.3. Joint distribution

The factorisation of the joint distribution of $\{(S_t, \boldsymbol{W}_t, \boldsymbol{D}_t, \boldsymbol{Z}_{0,t}) : t = 1, \ldots, T\}$ conditional on $\{X_t : t = 1, \ldots, T\}$, $\boldsymbol{D}_0$ and S_0 is represented in the directed acyclic graph (DAG) in Figure 2(a) where the double circles show deterministic dependence. Figure 2(b) shows the factorisation of the joint distribution for $\{(S_t, \boldsymbol{W}_t, \boldsymbol{D}_t) : t = 1, \ldots, T\}$ that arises after marginalising over $\boldsymbol{Z}_{0,t}$. Note that $(\boldsymbol{W}_t, \boldsymbol{D}_t)$ are conditionally independent of the LWT X_t given S_t and $\boldsymbol{D}_{t-1}$ and so the influence of the atmospheric data is only through the evolution of the weather states. From Figure 2(b) it can also be seen that both amounts and occurrences depend on the previous day's rainfall occurrence indicator. Additionally, at every time point t, neither occurrences nor amounts on wet days are conditionally independent across sites, given the weather state. Compared with other HMMs for rainfall from the literature, this represents a more sophisticated spatio–temporal dependence structure.

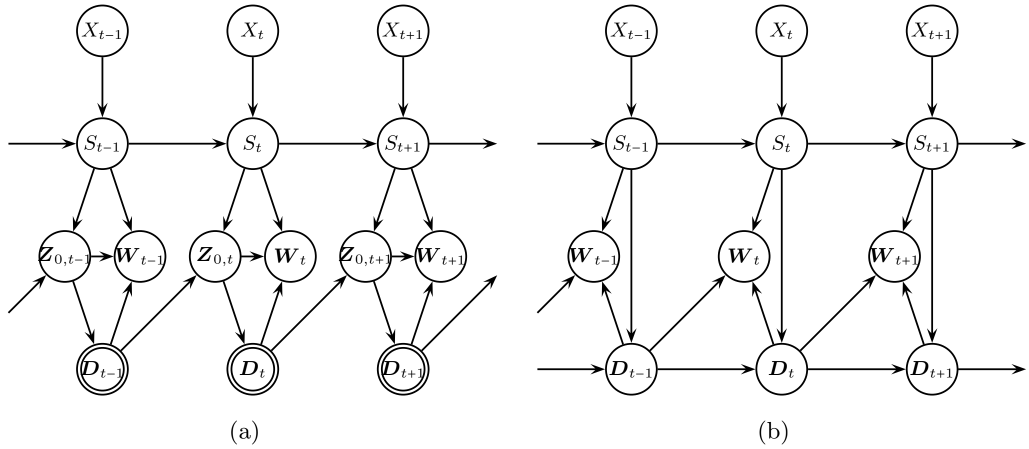

Fig. 2. DAGs showing the temporal dependence structure in the NHMM (a) before and (b) after omitting the $\boldsymbol{Z}_{0,t}$ nodes.

Our use of a second multivariate normal random variable $\boldsymbol{Z}_{1,t}$ means that we do not restrict the rainfall amount to be a deterministic function of the latent variables $\boldsymbol{Z}_{0,t}$ governing occurrence. As discussed in Section 3, other authors have captured spatial dependence between non–zero rainfall amounts by using generalised linear spatial process models. With $\boldsymbol{Z}_{0,t}$ playing the role of the vector of spatial random effects, the model defined through equations (3) and (4) is similar to a generalised linear spatial process model with normally distributed observables (log non–zero rainfall) and an identity link.

4.2. Prior distribution

We assume that the conditional prior for the model parameters in an r–state model takes the form $\pi(\boldsymbol{\theta}_r|r) = \pi(\boldsymbol{\theta}_{r,\text{hid}}|r)\pi(\boldsymbol{\theta}_{r,\text{obs}}|r)$. We further assume exchangeability with respect to the state labels as we do not wish to discriminate between any of the states *a priori*. For the parameters of the observed process, we choose the same hyperparameters in $\pi(\boldsymbol{\theta}_{r,\text{obs}}|r)$ for all $r \in \{1, \dots, r_{\max}\}$. This is in an effort to match the first and second moments in the prior predictive distribution of daily rainfall across models with different numbers of states.

The problem of incorporating genuine initial beliefs in a prior for the covariance matrix of spatial multivariate normal distributions becomes more straightforward if the covariance matrix is first transformed into a new set of parameters in a less constrained space. For illustration, consider a general random vector $\boldsymbol{Y} = (Y^1, \dots, Y^n)^{\mathrm{T}}|\boldsymbol{\mu}, \Sigma \sim \mathrm{N}_n(\boldsymbol{\mu}, \Sigma)$. A transformation based on the square–root free Cholesky decomposition of the precision matrix is given by $\Sigma^{-1} = (I_n - \Phi)^{\mathrm{T}}\Psi(I_n - \Phi)$ where Ψ is a diagonal matrix with positive diagonal entries $\psi_i^2 \in \mathbb{R}^+$, $i = 1, \dots, n$, and Φ is a strictly lower triangular matrix with (i, j)–th entry $\phi^{i,j} \in \mathbb{R}$ for $i > j$. This idea was proposed by Pourahmadi (1999) as a means of modelling a covariance matrix using covariates. The parameters in Φ and Ψ have an autoregressive interpretation based on the marginal/conditional decomposition of the joint density of $\boldsymbol{Y} = (Y^1, \dots, Y^n)^{\mathrm{T}}$. Specifically for $i > 1$, $\phi^{i,1}, \dots, \phi^{i,i-1}$ are the slope coefficients in the regression of Y^i on its (mean–centred) predecessors $Y^1, \dots, Y^{i-1}$ whilst ψ_i^2 is the conditional variance in the autoregression. This *generalised autoregression* requires an ordering of the elements in $\boldsymbol{Y}$. Therefore, our strategy when applying this reparameterisation to the state–specific covariance matrices $\Sigma_1, \dots, \Sigma_r$ is first to arrange the sites in a more natural order, described by a fixed permutation matrix M, and then to transform the permuted covariance matrix $\tilde{\Sigma}_k = M \Sigma_k M^{\mathrm{T}}$ into new parameters $\{(\tilde{\boldsymbol{\phi}}_k, \tilde{\Psi}_k) : k = 1, \dots, r\}$ where $\tilde{\boldsymbol{\phi}}_k = (\tilde{\phi}_k^{2,1}, \tilde{\phi}_k^{3,1}, \dots, \tilde{\phi}_k^{n,n-1})^{\mathrm{T}}$ and $\tilde{\Psi}_k = \mathrm{diag}(\tilde{\psi}_{k,1}^2, \dots, \tilde{\psi}_{k,n}^2)$. We can then choose to make the slope coefficients $(\tilde{\boldsymbol{\phi}}_1, \dots, \tilde{\boldsymbol{\phi}}_r)$ and the conditional variances $(\tilde{\Psi}_1, \dots, \tilde{\Psi}_r)$ independent *a priori*; for further details, see the Supplementary Materials.

Prior uncertainty about the model parameters is expressed through a prior of the form $\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}_{\mathrm{obs}}) \times \pi(\boldsymbol{\theta}_{\mathrm{hid}})$ where $\pi(\boldsymbol{\theta}_{\mathrm{obs}}) = \pi(\boldsymbol{\beta}_{0,1}, \dots, \boldsymbol{\beta}_{0,r}) \pi(\boldsymbol{\beta}_{1,1}, \dots, \boldsymbol{\beta}_{1,r}) \pi(\tilde{\boldsymbol{\phi}}_1, \dots, \tilde{\boldsymbol{\phi}}_r)$ $\times \pi(\tilde{\Psi}_1, \dots, \tilde{\Psi}_r) \pi(\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_r) \pi(\Gamma) \pi(\Omega_1, \dots, \Omega_r)$ and $\pi(\boldsymbol{\theta}_{\mathrm{hid}}) = \pi(\Lambda) \pi(\boldsymbol{\nu})$. Note that the product $\pi(\tilde{\boldsymbol{\phi}}_1, \dots, \tilde{\boldsymbol{\phi}}_r) \pi(\tilde{\Psi}_1, \dots, \tilde{\Psi}_r)$ induces a joint prior for $(\Sigma_1, \dots, \Sigma_r)$.

Consider first the parameters, $\boldsymbol{\theta}_{\mathrm{obs}}$, of the observed process. We assume *a priori* independence between weather states for the parameter blocks $(\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_r)$ and $(\Omega_1, \dots, \Omega_r)$ in the process $\{(\boldsymbol{W}_t | \boldsymbol{D}_t, S_t, \boldsymbol{Z}_{0,t})\}$ and for the parameter blocks $(\boldsymbol{\beta}_{0,1}, \dots, \boldsymbol{\beta}_{0,r})$, $(\boldsymbol{\beta}_{1,1}, \dots, \boldsymbol{\beta}_{1,r})$ and $(\tilde{\Psi}_1, \dots, \tilde{\Psi}_r)$ in the process $\{(\boldsymbol{Z}_{0,t} | S_t, \boldsymbol{D}_{t-1})\}$. Within every weather state, for each of these parameter blocks, we then adopt hierarchical priors which induce positive correlation between sites whilst maintaining semi–conjugacy in the prior specification: $\alpha_k^i | \mu_{\alpha_k}, \sigma_{\alpha_k}^2 \sim \mathrm{N}(\mu_{\alpha_k}, \sigma_{\alpha_k}^2)$ independently for $i = 1, \dots, n$. Then $\mu_{\alpha_k} \sim \mathrm{N}(a_{0,\alpha}, a_{1,\alpha}^2)$ and $\sigma_{\alpha_k}^2 \sim \mathrm{IG}(h_{0,\alpha}, h_{1,\alpha})$, where IG denotes an inverse gamma distribution. Similarly, for $k = 1, \dots, r$,

$$
\begin{aligned}
\omega_{k,i}^2 | \mu_{\omega_k^2} &\overset{iid}{\sim} \mathrm{IG}\{v_{\omega^2}^{-2} + 2, \mu_{\omega_k^2}(v_{\omega^2}^{-2} + 1)\}, & \mu_{\omega_k^2} &\sim \mathrm{Ga}(c_{0,\omega^2}, c_{1,\omega^2}), \\
\beta_{0,k}^i | \mu_{\beta_{0,k}} &\overset{iid}{\sim} \mathrm{N}(\mu_{\beta_{0,k}}, \sigma_{\beta_{0,k}}^2), & \mu_{\beta_{0,k}} &\sim \mathrm{N}(a_{0,\beta_0}, a_{1,\beta_0}^2), \\
\beta_{1,k}^i | \mu_{\beta_{1,k}} &\overset{iid}{\sim} \mathrm{ScBern}(\mu_{\beta_{1,k}}), & \mu_{\beta_{1,k}} &\sim \mathrm{Beta}(b_{0,\beta_1}, b_{1,\beta_1}), \\
\tilde{\psi}_{k,i}^2 | \mu_{\tilde{\psi}_k^2} &\overset{indep.}{\sim} \mathrm{IG}\{v_{\tilde{\psi}_i^2}^{-2} + 2, C_i \mu_{\tilde{\psi}_k^2}(v_{\tilde{\psi}_i^2}^{-2} + 1)\}, & \mu_{\tilde{\psi}_k^2} &\sim \mathrm{Ga}(c_{0,\tilde{\psi}^2}, c_{1,\tilde{\psi}^2}).
\end{aligned}
$$

Here the notation $X \sim \mathrm{ScBern}(p)$ means that the random variable X has a Bernoulli distribution, $\mathrm{Bern}(p)$, scaled to have support on $\{-1, 1\}$, i.e. $X = 2Y - 1$ where $Y \sim \mathrm{Bern}(p)$.

The purpose of the fixed hyperparameters $C_i > 0$, $i = 1, \dots, n$, in the prior for the conditional variances $\tilde{\psi}_{k,i}^2$ is to allow the marginal prior means $\mathrm{E}(\tilde{\psi}_{k,i}^2)$ to differ across sites $i = 1, \dots, n$. This reflects the fact that the conditional variances are not exchangeable in our prior beliefs because as i increases from 1 to n, $\tilde{\psi}_{k,i}^2$ represents the residual variance after $Z_{0,t}^i$ is regressed on an increasing number of predecessors $Z_{0,t}^1, \dots, Z_{0,t}^{i-1}$. We note that the prior chosen for the $\beta_{0,k}^i$ is pivotal in determining convergence of the MCMC sampler. This will be explained further in Section 6.1.

There are $rn(n+1)/2$ transformed covariance matrix parameters in $(\tilde{\boldsymbol{\phi}}_1, \dots, \tilde{\boldsymbol{\phi}}_r)$ and $(\tilde{\Psi}_1, \dots, \tilde{\Psi}_r)$. For $r > 1$ it is unlikely that all $rn(n+1)/2$ distinct parameters will be well identified in the likelihood, particularly if some of the states occur infrequently. We might reduce the number of parameters by assuming a parametric form for the covariance matrices (see Ailliot et al., 2009) or by assuming a common covariance matrix $\Sigma_k = \Sigma$ for all $k \in \mathcal{S}_r$. Instead we adopt the more flexible approach of using positive prior correlation between $(\Sigma_1, \ldots, \Sigma_r)$. This exploits "borrowing of strength" between the covariance matrices whilst allowing the data to inform us of differences between them. It is achieved through the hierarchical specification $\tilde{\boldsymbol{\phi}}_k | \boldsymbol{\mu}_{\tilde{\phi}} \overset{iid}{\sim} \mathrm{N}_{n(n-1)/2}(\boldsymbol{\mu}_{\tilde{\phi}}, V_{\tilde{\phi}})$, for $k = 1, \ldots, r$, with $\boldsymbol{\mu}_{\tilde{\phi}} \sim \mathrm{N}_{n(n-1)/2}(\mathbf{m}_{\tilde{\phi}}, C_{\tilde{\phi}})$.

Marginalising over $\boldsymbol{\mu}_{\tilde{\phi}}$ leads to a joint multivariate normal prior for $(\tilde{\boldsymbol{\phi}}_1, \ldots, \tilde{\boldsymbol{\phi}}_r)$ in which $\mathrm{E}(\tilde{\boldsymbol{\phi}}_k) = \mathbf{m}_{\tilde{\phi}}$, $\mathrm{Var}(\tilde{\boldsymbol{\phi}}_k) = \tilde{V}_{\tilde{\phi}} = V_{\tilde{\phi}} + C_{\tilde{\phi}}$ for each $k = 1, \ldots, r$ and $\mathrm{Cov}(\tilde{\boldsymbol{\phi}}_k, \tilde{\boldsymbol{\phi}}_\ell) = C_{\tilde{\phi}}$ for each $k \neq \ell$. For simplicity, we elicit $\tilde{V}_{\tilde{\phi}}$ and then take $C_{\tilde{\phi}} = \rho_{\tilde{\phi}} \tilde{V}_{\tilde{\phi}}$, where $\rho_{\tilde{\phi}} \in (0,1)$ is fixed. This means that for $k \neq \ell$, $\mathrm{Corr}(\tilde{\phi}_k^{s,t}, \tilde{\phi}_\ell^{u,v}) = \rho_{\tilde{\phi}} \mathrm{Corr}(\tilde{\phi}_k^{s,t}, \tilde{\phi}_k^{u,v}) = \rho_{\tilde{\phi}} \mathrm{Corr}(\tilde{\phi}_\ell^{s,t}, \tilde{\phi}_\ell^{u,v})$.

The i–th row of Γ comprises coefficients $\Gamma^{i,j}$, $j = 1, \ldots, n$, in the regression of $\log W_t^i$ on the latent variables $Z_{0,t}^j$, $j = 1, \ldots, n$. *A priori*, we believe that the effect of $Z_{0,t}^j$ on $\log W_t^i$ when $j = i$ will not be related to the effect when $j \neq i$ and so choose a prior with *a priori* independence between the on and off diagonal elements of Γ. For the collections of on and off diagonal elements we then encourage borrowing of strength by choosing semi–conjugate hierarchical priors with first–level specifications

$$\begin{aligned} \Gamma^{i,i} | \mu_{\Gamma_{\mathrm{on}}}, \sigma^2_{\Gamma_{\mathrm{on}}} &\sim \mathrm{N}(\mu_{\Gamma_{\mathrm{on}}}, \sigma^2_{\Gamma_{\mathrm{on}}}) \quad \text{independently for } i = 1, \ldots, n, \\ \Gamma^{i,j} | \mu_{\Gamma_{\mathrm{off}}}, \sigma^2_{\Gamma_{\mathrm{off}}} &\sim \mathrm{N}(\mu_{\Gamma_{\mathrm{off}}}, \sigma^2_{\Gamma_{\mathrm{off}}}) \quad \text{independently for } i, j = 1, \ldots, n,\ i \neq j, \end{aligned}$$

and second level specifications $\mu_{\Gamma_{\mathrm{on}}} \sim \mathrm{N}(a_{0,\Gamma_{\mathrm{on}}}, a^2_{1,\Gamma_{\mathrm{on}}})$, $\sigma^2_{\Gamma_{\mathrm{on}}} \sim \mathrm{IG}(h_{0,\Gamma_{\mathrm{on}}}, h_{1,\Gamma_{\mathrm{on}}})$, $\mu_{\Gamma_{\mathrm{off}}} \sim \mathrm{N}(a_{0,\Gamma_{\mathrm{off}}}, a^2_{1,\Gamma_{\mathrm{off}}})$ and $\sigma^2_{\Gamma_{\mathrm{off}}} \sim \mathrm{IG}(h_{0,\Gamma_{\mathrm{off}}}, h_{1,\Gamma_{\mathrm{off}}})$.

Consider now the parameters of the hidden process $\boldsymbol{\theta}_{\mathrm{hid}}$. The initial distribution $\boldsymbol{\nu}$ is assigned a conjugate Dirichlet prior $\boldsymbol{\nu} \sim \mathscr{D}_r(G\boldsymbol{g})$ where $\boldsymbol{g} = \mathrm{E}(\boldsymbol{\nu}) \in \mathscr{S}_r$, the r-dimensional simplex, and $G \in \mathbb{R}^+$. The assumption of *a priori* exchangeability across weather states requires that $\boldsymbol{g} = (1/r, \ldots, 1/r)$. We then choose $G = r$ to give a flat Dirichlet $\mathscr{D}_r(1, \ldots, 1)$ prior.

Table 2 shows that there are some LWTs which occur very infrequently. This means that the data are unlikely to be informative about some of the stochastic vectors $\boldsymbol{\lambda}_j^x = (\lambda_{j,1}^x, \ldots, \lambda_{j,r}^x)$ where $\lambda_{j,k}^x = \Pr(S_t = k | S_{t-1} = j, X_t = x, \boldsymbol{\theta}_{\mathrm{hid}})$. We can, again, facilitate (indirect) learning about some of the more rare (j, x) combinations by adopting a hierarchical Dirichlet prior

$$\boldsymbol{\lambda}_j^x | \boldsymbol{\xi}_j \overset{iid}{\sim} \mathscr{D}_r(\Xi_j \boldsymbol{\xi}_j), \quad \boldsymbol{\xi}_j \sim \mathscr{D}_r(E_j \boldsymbol{e}_j), \tag{5}$$

independently for each $j = 1, \ldots, r$, where $\Xi_j \in \mathbb{R}^+$, $E_j \in \mathbb{R}^+$ and $\boldsymbol{e}_j = \mathrm{E}(\boldsymbol{\xi}_j) \in \mathscr{S}_r$ are fixed hyperparameters and $\mathrm{E}(\boldsymbol{\lambda}_j^x | \boldsymbol{\xi}_j) = \boldsymbol{\xi}_j$ for each $x \in \{1, \ldots, 27\}$. In the prior induced for Λ, the blocks of stochastic vectors $(\boldsymbol{\lambda}_j^1, \ldots, \boldsymbol{\lambda}_j^{27})$ and $(\boldsymbol{\lambda}_k^1, \ldots, \boldsymbol{\lambda}_k^{27})$ are independent for each distinct pair of weather states $j \neq k$. However, within each block, the stochastic vectors $(\boldsymbol{\lambda}_j^1, \ldots, \boldsymbol{\lambda}_j^{27})$ are positively correlated, expressing the belief that if, for example, $\lambda_{j,k}^x$ was found to be larger (smaller) than its mean, this would lead to an upward (downward) revision of our beliefs about the mean of $\lambda_{j,k}^y$ for a different LWT, $y \neq x$. A benefit of (5) is that it is semi–conjugate to the multinomial form of the complete data likelihood.

Analogously to the precision parameters in a normal hierarchical prior, the parameters Ξ_j and E_j reflect the amounts of specific and common information in the prior. These can be considered in terms of the numbers of observations on transitions in the same LWT x or in another LWT x' which we would need to make a given change in our expectation of $\boldsymbol{\lambda}_j^x$.

The parameters $\{(\mu_{\beta_{0,k}}, \mu_{\beta_{1,k}}, \mu_{\tilde{\psi}_k^2}, \mu_{\alpha_k}, \sigma^2_{\alpha_k}, \mu_{\omega_k^2}) : k = 1, \ldots, r\}$, $\boldsymbol{\mu}_{\tilde{\phi}}$ and $(\mu_{\Gamma_{\mathrm{on}}}, \mu_{\Gamma_{\mathrm{off}}}, \sigma^2_{\Gamma_{\mathrm{on}}}, \sigma^2_{\Gamma_{\mathrm{off}}})$ which were given distributions at the second level in the hierarchical prior specifications above are appended to $\boldsymbol{\theta}_{\text{obs}}$. For convenience we introduce the notation $\mathcal{E} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_r) \in \mathscr{S}_r^r$ and then append $\mathcal{E}$ to the $\boldsymbol{\theta}_{\text{hid}}$. Our prior for the complete set of model parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_{\text{obs}}, \boldsymbol{\theta}_{\text{hid}})$ may then be written as

$$\pi(\boldsymbol{\theta}) = \Bigg[\prod_{k=1}^{r} \Big\{ \pi(\boldsymbol{\beta}_{0,k}|\mu_{\beta_{0,k}})\pi(\mu_{\beta_{0,k}})\pi(\boldsymbol{\beta}_{1,k}|\mu_{\beta_{1,k}})\pi(\mu_{\beta_{1,k}})\pi(\tilde{\Psi}_k|\mu_{\tilde{\psi}_k^2})\pi(\mu_{\tilde{\psi}_k^2})\pi(\boldsymbol{\alpha}_k|\mu_{\alpha_k}, \sigma^2_{\alpha_k}) \\ \times \pi(\mu_{\alpha_k})\pi(\sigma^2_{\alpha_k})\pi(\Omega_k|\mu_{\omega_k^2})\pi(\mu_{\omega_k^2})\pi(\tilde{\boldsymbol{\phi}}_k|\boldsymbol{\mu}_{\tilde{\phi}}) \Big\} \times \pi(\boldsymbol{\mu}_{\tilde{\phi}})\pi(\Gamma|\mu_{\Gamma_{\text{on}}}, \mu_{\Gamma_{\text{off}}}, \sigma^2_{\Gamma_{\text{on}}}, \sigma^2_{\Gamma_{\text{off}}}) \\ \times \pi(\mu_{\Gamma_{\text{on}}})\pi(\mu_{\Gamma_{\text{off}}})\pi(\sigma^2_{\Gamma_{\text{on}}})\pi(\sigma^2_{\Gamma_{\text{off}}}) \Bigg] \times \Big\{ \pi(\Lambda|\mathcal{E})\pi(\mathcal{E})\pi(\boldsymbol{\nu}) \Big\}.$$

5. Posterior inference via MCMC

5.1. *Posterior inference for fixed* r

We consider data in the form of a collection of time series which are treated as independent realisations. In our application each sub–series refers to the winter months in a particular year. Let the length of sub–series y be T_y with $y = 1, \ldots, Y$. The subscript y, t denotes day t within sub–series (year) y. For example $S_{y,t}$ denotes the weather state on day t in year y.

We first consider inferences given a fixed value of r. For this purpose we turn to MCMC techniques using data augmentation (Tanner and Wong, 1987) in which the latent variables are regarded as missing data and augmented to the state space of the sampler. In our case, the joint posterior of interest is then $\pi(\boldsymbol{\theta}, \boldsymbol{s}, \boldsymbol{s}_0, \boldsymbol{d}_0, \boldsymbol{z}_0|\boldsymbol{w}, \boldsymbol{d}, \boldsymbol{x})$, which we can write as

$$\pi(\boldsymbol{\theta}, \boldsymbol{s}, \boldsymbol{s}_0, \boldsymbol{d}_0, \boldsymbol{z}_0|\boldsymbol{w}, \boldsymbol{d}, \boldsymbol{x}) = p(\boldsymbol{w}, \boldsymbol{d}, \boldsymbol{d}_0, \boldsymbol{z}_0|\boldsymbol{s}, \boldsymbol{s}_0, \boldsymbol{\theta}_{\text{obs}})p(\boldsymbol{s}, \boldsymbol{s}_0|\boldsymbol{x}, \boldsymbol{\theta}_{\text{hid}})\pi(\boldsymbol{\theta}_{\text{hid}})\pi(\boldsymbol{\theta}_{\text{obs}}).$$

Samples from this distribution can be generated using a Gibbs scheme which iterates through the following four steps:

1. Sample $\boldsymbol{\theta}$ from its conditional posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{w}, \boldsymbol{d}, \boldsymbol{d}_0, \boldsymbol{s}, \boldsymbol{s}_0, \boldsymbol{z}_0, \boldsymbol{x})$ in a series of Gibbs (or Metropolis–within–Gibbs) steps. The full conditional distributions (FCDs) for all parameters in $\boldsymbol{\theta}_{\text{obs}}$ and for the initial distribution $\boldsymbol{\nu}$ are standard distributions and can be sampled directly. The joint FCD for $(\Lambda, \mathcal{E})$ is non–standard and so we sample $(\Lambda_j, \mathcal{E}_j)$ using a Metropolis–Hastings step for $j = 1, \ldots, r$.

2. Sample $(\boldsymbol{s}, \boldsymbol{s}_0)$ from its conditional posterior $\pi(\boldsymbol{s}, \boldsymbol{s}_0|\boldsymbol{w}, \boldsymbol{d}, \boldsymbol{d}_0, \boldsymbol{z}_0, \boldsymbol{\theta}, \boldsymbol{x})$. This is achieved using a forward–backward simulation algorithm (see, for example Frühwirth-Schnatter, 2006, Algorithm 11.5) in which $\boldsymbol{z}_0$ and $\boldsymbol{d}_0$ are treated in the same manner as observed data. The algorithm is applied separately to each sub–series y.

3. Sample $\boldsymbol{z}_0$ from its conditional posterior, $\pi(\boldsymbol{z}_0|\boldsymbol{w}, \boldsymbol{d}, \boldsymbol{d}_0, \boldsymbol{s}, \boldsymbol{\theta}_{\text{obs}})$. When $\boldsymbol{d}_{1,1}, \ldots, \boldsymbol{d}_{Y,T_Y}$ are all observed, the latent variables $\boldsymbol{Z}_{0,1,1}, \ldots, \boldsymbol{Z}_{0,Y,T_Y}$ are independent in this joint distribution. The conditional posterior for each $\boldsymbol{Z}_{0,y,t}$ is a truncated multivariate normal distribution and its components are sampled one–at–a–time from their univariate truncated normal conditionals; see, for example, Geweke (1991).

4. Sample $\boldsymbol{d}_0$ from its conditional posterior $\pi(\boldsymbol{d}_0|\boldsymbol{z}_0, \boldsymbol{s}, \boldsymbol{\theta}_{\text{obs}})$. The initial occurrences $\boldsymbol{d}_{1,0}, \ldots,$ $\boldsymbol{d}_{Y,0}$ are conditionally independent in their joint full conditional distribution. For each sub–series y, the components of $\boldsymbol{d}_{y,0}$ are sampled one–at–a–time from their univariate Bernoulli conditionals.

Full details of this scheme can be found in the Supplementary Materials. It can be regarded as an extension of the traditional two–stage Gibbs sampling strategy which is often employed in the analysis of more standard HMMs, in which the hidden states are the only latent variables (see, for example Frühwirth-Schnatter, 2006, Algorithm 11.3).

Inference in HMMs is complicated by the problem of label-switching which occurs because posterior probability is spread between the different possible permutations of the state labels; see, for example, Stephens (2000). This can be particularly problematic when priors are chosen which are exchangeable with respect to the state labels. In Section 6 we consider some properties of posterior distributions for the model parameters $\boldsymbol{\theta}$ and the weather states $\boldsymbol{s}$ conditional on r, and this requires a distinct labelling of the states. To overcome this problem we use an online relabelling algorithm which is described by Boys and Henderson (2002). After each MCMC iteration, the algorithm uses a scoring criterion to find the permutation of the labels which is most consistent with previous iterations.

5.2. Posterior inference for $\boldsymbol{r}$

It is now necessary to introduce notational dependence on r. For example, we denote the parameters of the hidden process in a r–state NHMM by $\boldsymbol{\theta}_{r,\text{hid}} = (\Lambda_r, \mathcal{E}_r, \boldsymbol{\nu}_r)$.

The posterior mass function for the number of states $r \in \{1, \ldots, r_{\max}\}$ is given by

$$\pi_r(r|\boldsymbol{w}, \boldsymbol{d}, \boldsymbol{x}) = \frac{p(\boldsymbol{w}, \boldsymbol{d}|\boldsymbol{x}, r)\pi_r(r)}{\sum_{k=1}^{r_{\max}} p(\boldsymbol{w}, \boldsymbol{d}|\boldsymbol{x}, k)\pi_r(k)} \tag{6}$$

in which the marginal likelihood, $p(\boldsymbol{w}, \boldsymbol{d}|\boldsymbol{x}, r)$, is the normalising constant in the conditional posterior distribution of $\boldsymbol{\theta}_r$ given r,

$$p(\boldsymbol{w}, \boldsymbol{d}|\boldsymbol{x}, r) = \int p(\boldsymbol{w}, \boldsymbol{d}|\boldsymbol{\theta}_r, \boldsymbol{x}, r)\pi(\boldsymbol{\theta}_r|r)\, d\boldsymbol{\theta}_r.$$

This integral cannot be evaluated in closed form. However, posterior model probabilities of the form (6) can be approximated by a variety of numerical methods. These methods can be divided into *across–* and *within–model–simulation* techniques. The former use Markov chains which target the joint posterior $\pi(\boldsymbol{\theta}_r, r|\boldsymbol{w}, \boldsymbol{d}, \boldsymbol{x})$, whilst the latter approximate the marginal likelihood for each model r in turn and then compute the posterior for r through application of (6). Unfortunately we were unable to find a workable method of either kind; see the Supplementary Materials for further details.

Proper scoring rules (Gneiting and Raftery, 2007) use a numerical score to quantify the quality of a probabilistic forecast on the basis of the predictive distribution from which the forecast was issued and the observation that ultimately materialised. Gschlößl and Czado (2007) consider the use of proper scoring rules in the context of Bayesian model comparison in which observations from an *out–of–sample period*, that is, a period which was not used in model–fitting, are compared with forecasts from the corresponding posterior predictive distribution. In our case, the posterior predictive distribution of data $(\boldsymbol{w}^{\text{rep}}, \boldsymbol{d}^{\text{rep}})$ that could have been observed under the model with r states is given by

$$p(\boldsymbol{w}^{\text{rep}}, \boldsymbol{d}^{\text{rep}}|\boldsymbol{w}, \boldsymbol{d}, \boldsymbol{x}, r) = \int p(\boldsymbol{w}^{\text{rep}}, \boldsymbol{d}^{\text{rep}}|\boldsymbol{\theta}_r, \boldsymbol{x}, r)\pi(\boldsymbol{\theta}_r|\boldsymbol{w}, \boldsymbol{d}, \boldsymbol{x}, r)\, d\boldsymbol{\theta}_r. \tag{7}$$

From this equation we can deduce, for example, the marginal posterior predictive distribution for site i and day t, $p(w_t^{\text{rep},i}, d_t^{\text{rep},i}|\boldsymbol{w}, \boldsymbol{d}, \boldsymbol{x}, r)$. Given the intractability of the posterior distribution for r, we will use these ideas as an alternative means of comparing models with different numbers r of states.

Various proper scoring rules are available but those defined in terms of predictive distribution functions, rather than predictive densities, have particular appeal in the context of rainfall modelling because they better suit the mixed nature of precipitation distributions. One such scoring rule, presented in Gneiting and Raftery (2007), is the *continuous–ranked probability score* (CRPS). Omitting notational reference to $\boldsymbol{d}$, this is defined as

$$\mathrm{CRPS}_r(F_t^i, w_t^i) = -\int_0^\infty \{F_t^i(y|r) - \mathbb{I}(y \geq w_t^i)\}^2 dy,$$

for site i and time t. Here $F_t^i(\cdot|r)$ is the posterior predictive distribution for rainfall at site i on day t given an r state model, that is, the distribution function corresponding to the marginal density $p(w_t^{\mathrm{rep},i}, d_t^{\mathrm{rep},i}|\boldsymbol{w}, \boldsymbol{d}, \boldsymbol{x}, r)$. This scoring rule assigns the highest rewards to predictive distribution functions which are very concentrated around the observation which ultimately materialises. The CRPS can also be written as

$$\mathrm{CRPS}_r(F_t^i, w_t^i) = \frac{1}{2}\,\mathrm{E}_{F_t^i}\left(|W_t^{\mathrm{rep},i} - W_t^{\mathrm{rep},i'}|\right) - \mathrm{E}_{F_t^i}\left(|W_t^{\mathrm{rep},i} - w_t^i|\right), \tag{8}$$

in which $W_t^{\mathrm{rep},i}$ and $W_t^{\mathrm{rep},i'}$ are independent replicates of the random variable W_t^i with distribution function $F_t^i(\cdot|r)$ and the expectations are with respect to this distribution. This representation is particularly useful when predictive distribution functions are numerically approximated through a random sample $w_t^{\mathrm{rep},i,[1]}, \ldots, w_t^{\mathrm{rep},i,[N]}$ when the terms on the right-hand-side of (8) can be approximated through, for example,

$$\mathrm{E}_{F_t^i}\left(|W_t^{\mathrm{rep},i} - W_t^{\mathrm{rep},i'}|\right) \simeq \frac{1}{N/2}\sum_{j=1}^{N/2} |w_t^{\mathrm{rep},i,[2j-1]} - w_t^{\mathrm{rep},i,[2j]}| \tag{9}$$

and

$$\mathrm{E}_{F_t^i}\left(|W_t^{\mathrm{rep},i} - w_t^i|\right) \simeq \frac{1}{N}\sum_{j=1}^{N} |w_t^{\mathrm{rep},i,[j]} - w_t^i|. \tag{10}$$

It is straightforward to generate the samples needed to evaluate (9) and (10) given LWT data from an out–of–sample period and an approximately un-autocorrelated sample $\boldsymbol{\theta}_r^{[j]}$, $j = 1, \ldots, N$, from the posterior distribution of the model parameters. This is achieved by generating a sample $\boldsymbol{w}_t^{\mathrm{rep},[j]} = (w_t^{\mathrm{rep},1,[j]}, \ldots, w_t^{\mathrm{rep},n,[j]})$, $t = 1, 2, \ldots$, from the model $p(\boldsymbol{w}^{\mathrm{rep}}, \boldsymbol{d}^{\mathrm{rep}}|\boldsymbol{\theta}_r^{[j]}, \boldsymbol{x}, r)$ for each $j = 1, \ldots, N$, in which the LWTs $\boldsymbol{x}$ are from the out–of–sample period. The draws for a particular time t and site i then correspond to a sample from the marginal posterior predictive distribution $F_t^i(\cdot|r)$ and can be used to evaluate (8). An average of the CRPS scores across all sites and all time points provides an overall measure of forecast quality for an r–state model. In practise we fit models with $r = 1, 2, \ldots$ states until increasing r leads to no further improvement (increase) in the score.

An alternative to these site–wise comparisons is to use a proper scoring rule which can assess whether forecasts are spatially consistent. We discuss such an extension of the CRPS which applies to vector forecasts in the Supplementary Materials.

6. Application to UK winter rainfall data

In this section we apply the model and inferential procedures to the UK winter dataset which was introduced in Section 2. The dataset has $Y = 28$ sub–series, with one sub–series of length $T_y = 90$ (or $T_y = 91$ in leap years) for each of the 28 calendar winters. There are very long (9 month) time periods separating consecutive sub–series so it seems reasonable to model them as independent realisations of the NHMM.

Our choice of hyperparameters in the prior distribution is based primarily on our subjective assessments about various aspects of the rainfall process; see Germain (2010) for an account of suitable elicitation strategies. These values, along with the permutation matrix M, are given in the Supplementary Materials. Unlike the parameters in the observed process $\boldsymbol{\theta}_{r,\text{obs}}$, our priors for the parameters in $\boldsymbol{\theta}_{r,\text{hid}}$ differ with the number of states r in an effort to balance the amount of information contained in the prior for each model.

We begin this section by describing the implementation of our MCMC scheme and then how we select a suitable value $\hat{r}$ for the number of weather states using the proper scoring rule method described previously. For reasons which will be explained in Section 6.2, we consider models with $r = 1, \ldots, 5$ states. This is followed by summaries of the posterior distribution for the parameters in the model with $\hat{r}$ states. We conclude with an assessment of the fit of the model, comparing the posterior predictive distribution to observed data which were not used in model fitting.

For comparative purposes, we also consider a reduced model in which we fix Γ to be a matrix of zeros, the coefficients of the lag–1 rainfall occurrence indicators $\boldsymbol{\beta}_{r,1,k}$ to be vectors of zeros and the covariance matrices $\Sigma_{r,k}$ to be identity matrices (i.e. every $\tilde{\phi}^{i,j}_{r,k} = 0$ and every $\tilde{\psi}^2_{r,k,i} = 1$). This produces a within-state model in which the rainfall occurrences $D^i_{y,t}$ and the non–zero amounts $W^i_{y,t}$ are independent in time and space with Bernoulli $\text{Bern}\{\Phi(\beta^i_{r,0,k})\}$ and lognormal $\log\text{N}(\alpha^i_{r,k}, \omega^2_{r,k,i})$ distributions respectively. This reduced model is very similar to the pioneering model of Bellone et al. (2000), differing only in the use of lognormal, rather than gamma, distributions for non–zero rainfall amounts. It is used as a benchmark in Section 6.4 when we consider the fit of the latent Gaussian variable NHMM.

6.1. Implementation of the MCMC scheme, convergence and mixing

For each fixed number of states $r = 1, \ldots, 5$, the MCMC algorithm was used to generate 2.5M draws from the posterior, omitting the first 500k as burn–in and thinning the remaining output to retain every 200–th iterate, to give posterior samples of size $N = 10\text{k}$. Graphical diagnostic checks including trace and autocorrelation plots were used to inspect the convergence and mixing properties of the chains.

When using a more diffuse prior than that detailed in Section 4.2, the MCMC chains for models with large numbers r of states failed to converge. In particular, problems arose with the parameter $\boldsymbol{\beta}_{r,0,k} = (\beta^1_{r,0,k}, \ldots, \beta^n_{r,0,k})^{\text{T}}$ within the weather state $k \in \{1, \ldots, r\}$ associated with the largest probabilities of rain. At certain sites i in these states k, trace plots revealed $\beta^i_{r,0,k}$ increasing without bound over the course of the MCMC run. It is largely these parameters which control the probability of rain at site i in state k. Investigation into the days typically assigned to these states revealed that the data suggested a probability of rain very close to 1 at these sites. Given the probit transformation mapping the rainfall probabilities to functions of the $\beta^i_{r,0,k}$, this causes the likelihood to favour arbitrarily large values of $\beta^i_{r,0,k}$. To prevent this from happening in the posterior, the prior for the $\beta^i_{r,0,k}$ needed to have small variance and reasonably short tails.

Initialising the chains at a variety of starting points and comparing trace plots, the various runs produced essentially the same results up to the labelling of the states. Conse-

Table 3. The mean CRPS (averaged across sites and time points) for models with $r = 1, \ldots, 5$ states.

r	1	2	3	4	5
Mean CRPS	-2.130	-2.031	-2.015	**-2.007**	-2.013

quently, there was no evidence of any lack of convergence. Based on autocorrelation plots, thinning to every 200–th iterate appeared to remove most of the autocorrelation in the chains for models with $r \leq 4$ states. Plots of the posterior densities and trace plots for some of the parameters in the model with $r = 5$ states displayed evidence of multimodality and this made mixing difficult to assess. Multimodality in the posterior distributions of parameters in mixture models and HMMs is not uncommon; see, for example, Richardson and Green (1997) and Celeux et al. (2000). It generally arises due to the existence of multiple competing descriptions of the data which are comparable in terms of their posterior support. Nevertheless, the sampler appeared to move readily between the different modes.

As an example, for the model with $r = 4$ states, the computing time required to generate 2.5M posterior draws was around 130 hours using sequential C code on a 2.40GHz Dell PowerEdge R410 server with two six–core Intel Xeon E5645 CPUs and 32GB RAM. In rough terms, the number of sampled unknowns increases between linearly and quadratically with the number n of sites for fixed r. The computing time scales in correspondence, in this case increasing by a factor of 2.4 and 5.1 when the number of sites doubles and triples, respectively.

6.2. Choice of r

In order to use the proper scoring rule method outlined in Section 5.2, we need to compare posterior predictive distributions with data from a period that was not used to construct them. For this purpose we have precipitation and LWT data for the six winter periods that followed the 28 winter seasons used in model–fitting, that is, from the years 1989/90 to 1994/95. These sub–series, of total length 541 days, contain no missing data. Table 3 shows the mean CRPS for models with various numbers of weather states r. Higher scores indicate better predictive performance and so it appears that the model's predictive performance improves as r gets larger until the number of weather states is $r = 4$. Increasing r further to $r = 5$ leads to no further gains. We note that when we considered the extension of the CRPS involving vector forecasts, the conclusion was the same; see the Supplementary Materials for more details. Therefore we chose the number of weather states to be $\hat{r} = 4$. Given the complexity of the within–state model, which itself captures spatial and temporal autocorrelation, this relatively small value of r is not surprising.

6.3. Parameter inference assuming four weather states

We now summarise the conditional posterior distribution given $r = \hat{r} = 4$. Figures 3(a) and 3(b) show the posterior means and 95% equi–tailed credible intervals for the conditional probability of rain at each site, given the weather state $S_{y,t} = k$ and the site's rainfall status the previous day $D_{y,t-1}^{i} = d$. Marginalising over $\boldsymbol{Z}_{0,y,t}$ in the joint distribution for $(\boldsymbol{Z}_{0,y,t}, \boldsymbol{Z}_{1,y,t} | \boldsymbol{D}_{y,t-1}, S_{y,t})$ gives

$$\boldsymbol{Z}_{1,y,t} | \boldsymbol{D}_{y,t-1} = \boldsymbol{d}_{y,t-1}, S_{y,t} = k, \boldsymbol{\theta}_{\text{obs}} \sim \text{N}_n(\boldsymbol{\alpha}_k + \Gamma \boldsymbol{\mu}_{y,t,k} \,, \, \Omega_k + \Gamma \Sigma_k \Gamma^{\text{T}}),$$

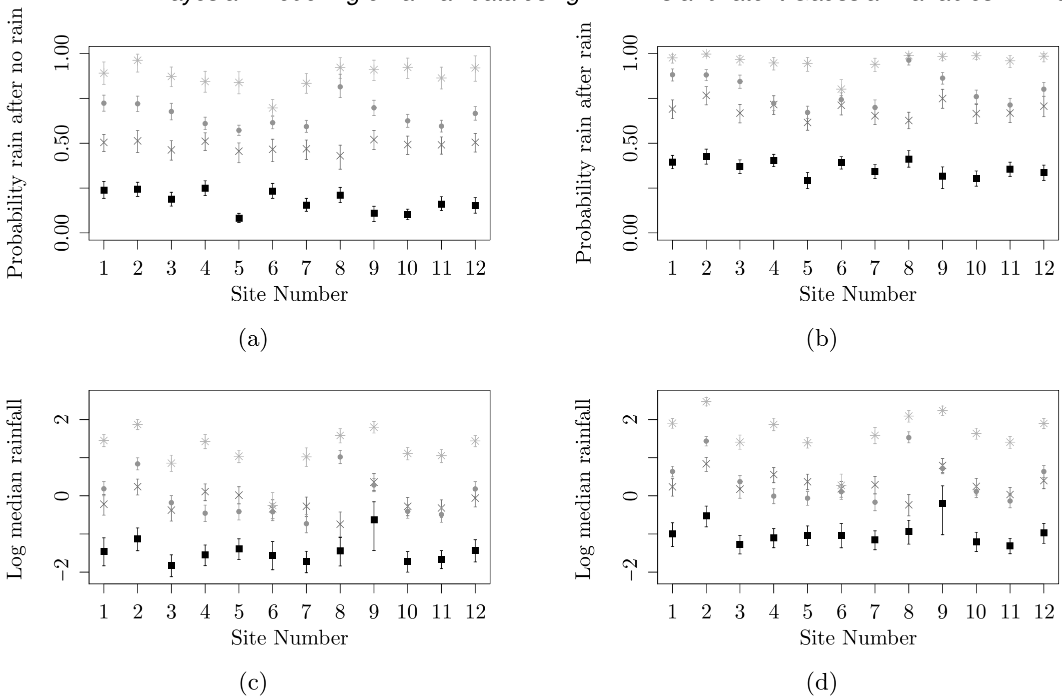

Fig. 3. Conditional on $r = 4$, posterior means with 95% equi–tailed credible intervals at each site in weather states 1 (—×—), 2 (—✳—), 3 (—•—) and 4 (—■—). Top row: probabilities of rain following (a) a dry day and (b) a wet day. Bottom row: log medians in the lognormal distributions for rainfall amounts when (c) $\boldsymbol{d}_{y,t-1} = (0, \ldots, 0)^{\mathrm{T}}$ and (d) $\boldsymbol{d}_{y,t-1} = (1, \ldots, 1)^{\mathrm{T}}$.

from which we can easily deduce the univariate lognormal distribution for wet day rainfall at each site conditional on $S_{y,t} = k$ and $\boldsymbol{D}_{y,t-1} = \boldsymbol{d} \in \{0,1\}^n$. For the two most frequent rainfall occurrence indicators in the observed data, $\boldsymbol{d} = (0, \ldots, 0)^{\mathrm{T}}$ and $\boldsymbol{d} = (1, \ldots, 1)^{\mathrm{T}}$, and for each state $k = 1, \ldots, 4$, the posteriors for the log medians in these distributions are displayed in Figures 3(c) and 3(d) for sites $i = 1, \ldots, 12$. In Figure 3, the effect of $\boldsymbol{D}_{y,t-1}$ can be seen clearly. This supports the regression of $\boldsymbol{Z}_{0,y,t}$ on $\boldsymbol{D}_{y,t-1}$.

Figure 3 also shows that weather state 2 is a clear–cut wet state, characterised by high probabilities of rain at each of the sites and large rainfall amounts on wet days. Similarly, weather state 4 is clear–cut dry. States 1 and 3 are intermediate between states 2 and 4, although state 1 represents drier conditions than state 3 at most sites. A plot of the posterior for the coefficients of variation in the daily rainfall distributions at each site also reveals that the wet weather state displays the least variation of the four, representing only the wettest days; see the Supplementary Materials for more details.

Figure 4 displays the marginal posterior distributions for the weather state transition probabilities $\lambda^x_{4,j,k} = \Pr(S_{y,t} = k | S_{y,t-1} = j, X_t = x, \boldsymbol{\theta}_{3,\mathrm{obs}}, r = 4)$, $x = 1, \ldots, 27$, for two representative $j \to k$ transitions. The LWTs are labelled according to Table 2. Both plots also show the marginal posterior distribution for the corresponding $\xi_{4,j,k}$ and the marginal prior distribution for the transition probability $\lambda^x_{4,j,k}$ (which is the same for all x). Figure 4(b) shows the marginal posteriors for $\lambda^x_{4,2,4}$, $x = 1, \ldots, 27$, and is typical of the posteriors for all probabilities of transition from the wet weather state (state 2). For these

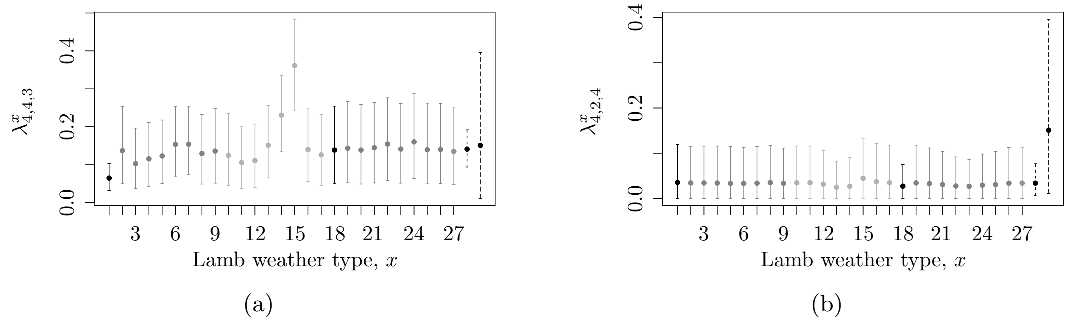

Fig. 4. Conditional on $r = 4$, posterior means with 95% equi–tailed credible intervals for $\lambda^x_{4,j,k}$, $x = 1,\ldots,27$, (———) and $\xi_{4,j,k}$ (·········) when (a) $j = 4$, $k = 3$ and (b) $j = 2$, $k = 4$. Also shown are the marginal prior means with 95% equi–tailed credible intervals (------) for the corresponding transition probabilities $\lambda^x_{4,j,k}$, $x = 1,\ldots,27$.

transition probabilities there is considerable overlap in the marginal posteriors across LWTs, indicating that the atmospheric data are not particularly helpful in explaining transitions *from* the wet weather state. This may be due to the transient nature of this state, possibly representing a frontal depression which typically passes in a day; see Figure S5 of the Supplementary Materials. In contrast, the transition probabilities from the other three weather states (1, 3 and 4) and in particular the clear–cut dry state (state 4) are much more markedly influenced by the LWT. Figure 4(a), for example, displays the marginal posterior distributions for $\lambda^x_{4,4,3}$, $x = 1,\ldots,27$ which is the probability of moving from the dry state to the wetter of the two intermediate states, given that the current LWT is x. The central 95% of the posteriors for $\lambda^1_{4,4,3}$ and a couple of the $\lambda^x_{4,4,3}$ corresponding to the pure directional types (x =10–17) do not overlap. The information gained from using LWTs is reinforced by considering the (hypothetical) stationary distributions for each LWT.

Let Λ^x_4 be a 4×4 stochastic transition matrix with j–th row $\boldsymbol{\lambda}^x_{4,j}$. The solution $\boldsymbol{\delta}^x_4$ to the matrix equation $\boldsymbol{\delta}^x_4 \Lambda^x_4 = \boldsymbol{\delta}^x_4$ for each LWT x can be interpreted as the stationary distribution of the (homogeneous) HMM that would prevail if the LWT was always x. Therefore a good summary of the effect of the LWTs is given by the posterior distributions for $\boldsymbol{\delta}^x_4 = (\delta^x_{4,1}, \delta^x_{4,2}, \delta^x_{4,3}, \delta^x_{4,4}) \in \mathscr{S}_r$ for each value of x. These are displayed in Figure 5 and reveal a complex pattern amongst the LWTs with the dominant feature being the variation amongst the pure directional and amongst the pure vortical types.

The pure south–westerly ($x = 14$) and pure westerly ($x = 15$) LWTs seem to favour state 3, the wetter of the two intermediate states, but offer little support to the driest two states (states 1 and 4). Southerly types, in particular x =12,13,14, seem to favour the wettest weather state (state 2) whilst the pure cyclonic ($x = 18$) and pure anticyclonic ($x = 1$) types overwhelmingly support states 1 and 4 respectively. It is especially noticeable that the anticyclonic type favours the driest state (state 4). Given that this LWT is typically associated with dry conditions, this result confirms expectations.

An analysis of the posterior distribution of the weather states, given $r = 4$, is given in the Supplementary Materials.

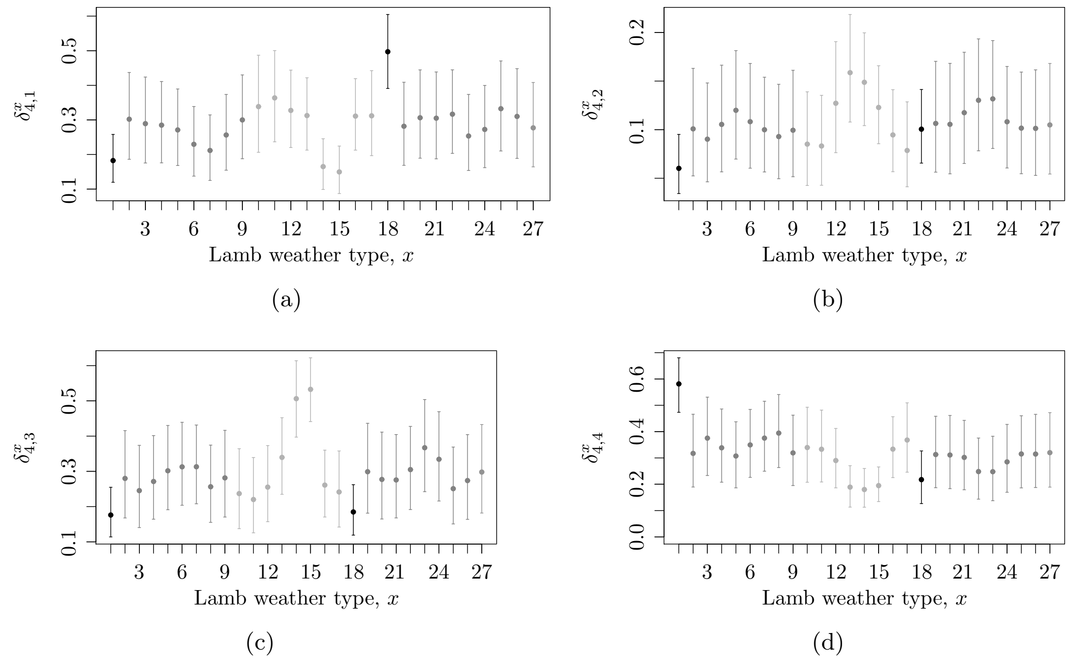

Fig. 5. Marginal posterior means and 95% equi–tailed credible intervals for the solution to the matrix equation $\boldsymbol{\delta}_4^x \boldsymbol{\Lambda}_4^x = \boldsymbol{\delta}_4^x$, $x = 1, \ldots, 27$. Here $\boldsymbol{\Lambda}_4^x$ is the 4×4 stochastic matrix with j–th row equal to $\boldsymbol{\lambda}_{4,j}^x$, $\boldsymbol{\delta}_4^x = (\delta_{4,1}^x, \delta_{4,2}^x, \delta_{4,3}^x, \delta_{4,4}^x) \in \mathscr{S}_4$ and the plots show (a) $\delta_{4,1}^x$, (b) $\delta_{4,2}^x$, (c) $\delta_{4,3}^x$ and (d) $\delta_{4,4}^x$.

6.4. Model checking

Chapter 6 of Gelman et al. (1995) describes a number of Bayesian model checking procedures. We use some of the graphical checks in this section to assess the ability of the NHMM to capture some important properties of the joint rainfall distribution. These checks are based on the posterior predictive distribution (7) of a hypothetical replicate of data $(\boldsymbol{w}^{\text{rep}}, \boldsymbol{d}^{\text{rep}})$ that *could* have been observed under the model. The predictive distributions used in this section are conditional on $r = \hat{r} = 4$.

Let $T(\boldsymbol{w}, \boldsymbol{d})$ be a *test quantity*, that is, a scalar summary representing an aspect of the data that we want to capture accurately, for example the proportion of wet days at one of the sites. We can simulate from the posterior predictive distribution of the test quantity by using the MCMC output $\boldsymbol{\theta}_r^{[j]}$, $j = 1, \ldots, N$, to generate draws $(\boldsymbol{w}^{\text{rep},[j]}, \boldsymbol{d}^{\text{rep},[j]})$ from the posterior predictive distribution as discussed in Section 5.2. These draws are used to compute $T(\boldsymbol{w}^{\text{rep},[j]}, \boldsymbol{d}^{\text{rep},[j]})$, $j = 1, \ldots, N$. The posterior predictive distributions of various test quantities can then be compared graphically to their observed values. If the model fits well, then the observed test quantities should look plausible under the corresponding posterior predictive distributions. To avoid using the same data for both model fitting and model checking, we base these comparisons on data from the out–of–sample period introduced in Section 6.2, conditioning the posterior predictive distribution on LWT data from this period. Where model–checking plots correspond to single sites, we show the results for two sites: site 5 (Kew) and site 9 (Plymouth), chosen because Kew generally represented good fit, whilst Plymouth generally represented the poorest fit out of the twelve sites. In all plots, the posterior predictive distributions are summarised through their mean and 95% equi–tailed credible interval.

Marginal properties of the rainfall data, for example the relative frequencies of rainfall occurrence and the sample quantiles in the distribution of non–zero amounts at each site, showed good agreement with their posterior predictive distributions. Further comments and plots can be found in the Supplementary Materials. The remainder of this section focuses on the spatial and temporal characteristics of the data.

As discussed in Section 2, we can measure spatial autocorrelation amongst rainfall occurrences and non–zero rainfall amounts using log–odds ratios and Spearman's rank correlation coefficients respectively. For all pairs of sites, Figures 6(a) and 6(b) compare the observed values of these statistics to their posterior predictive distributions. These figures also show, for comparison, summaries of the posterior predictive distributions for the simple conditional independence model discussed in Section 6.2. For clarity we only show the posterior predictive means although plots which also show 95% credible intervals can be found in the Supplementary Materials. Although the fit of the latent Gaussian variable model offers a considerable improvement over the simpler model, it still seems to underestimate the larger spatial autocorrelations between rainfall occurrences. This is surprising because, when these model checks are performed using the model–fitting data, there is very good agreement between the observed statistics and their posterior predictive distributions; see Figures 6(c) and 6(d). Comparing Figures 6(a) and 6(c) it appears that there has been a change in the joint patterns of rainfall occurrence causing some of the log odds ratios for the 1989/90 to 1994/95 period to be larger than any of those calculated from the 1961/62 to 1988/89 data used in model fitting. This suggests that our combination of model, data and prior was unable to explain this medium–term change in the precipitation behaviour, even after accounting for the observed LWTs during this period.

To assess the model's ability to capture the temporal autocorrelation in the occurrence process, we compare the observed empirical survivor functions of wet and dry spells at each of the sites and their posterior predictive distributions. The empirical survivor function of wet (dry) spells is simply defined as the proportion of runs of consecutive wet (dry) days that persist for at least k days, $k = 1, 2, \ldots$. Figures 6(e) and 6(f) show the plots on a log–scale for wet and dry spells at Kew whilst Figures 6(g) and 6(h) show the corresponding plots for Plymouth. The wet spells plot for Kew is representative of those for the majority of other sites, showing a close correspondence between the observed distribution and its posterior predictive mean. The plot for dry spells actually shows poorer fit than that displayed at most of the other sites, although for most durations the observed statistics still lie within the central 95% of their posterior predictive distributions. The corresponding plots for Plymouth depict the worst fit of any of the sites and indicate that at a small number of sites, the model fails to predict longer duration wet and dry spells as frequently as they are observed. Figures 6(g) and 6(h) also show summaries of the posterior predictive distribution for Plymouth obtained under the simple conditional independence model. Again, for clarity we show only the posterior predictive means, with plots showing 95% credible intervals given in the Supplementary Materials. Compared with the simpler model, we see that the latent Gaussian variable model offers a noticeable improvement in fit.

The ability of the model to capture the temporal autocorrelation between rainfall amounts within wet spells is assessed by comparing the observed Spearman's rank correlation coefficients between rainfall amounts at various lags (within uninterrupted wet spells) to the corresponding posterior predictive distribution. Plots revealed good agreement and can be found in the Supplementary Materials, along with additional commentary.

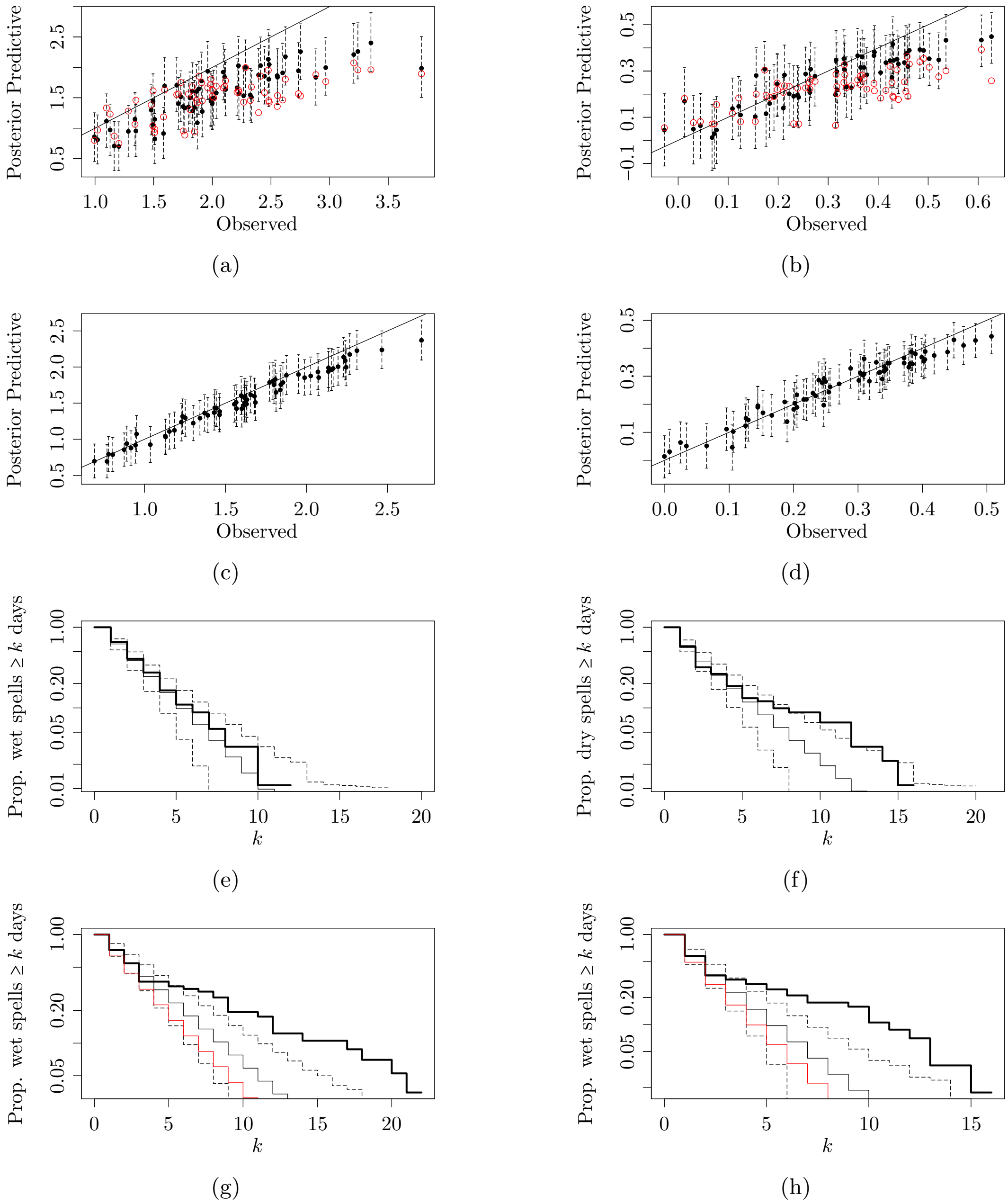

Fig. 6. Rows 1 and 2: observed versus posterior predictive means for log odds ratios (column 1) and Spearman's rank correlations (column 2) between each pair of sites based on out–of–sample data (row 1) and model–fitting data (row 2). Rows 3 and 4: observed (———) and posterior predictive mean (———) empirical survival distributions of wet spells (column 1) and dry spells (column 2) at Kew (row 3) and Plymouth (row 4) based on out–of–sample data. In each case ------ indicate posterior predictive 95% credible intervals. In rows 1 and 4, ∘ and ——— indicate the posterior predictive means for the simple conditional independence model.

7. Discussion

We have presented a model for the spatio-temporal analysis of daily rainfall data. A key feature of the model is that a bivariate latent variable (Z_0, Z_1) is used to govern occurrence and amount rather than, for example, using a single variable and truncation. We believe that this new model offers advantages. In particular (i) by using two multivariate normal random vectors, it allows dependence between sites in non–zero rainfall amounts by a mechanism which is additional to that which governs dependence in rainfall occurrences and (ii) it avoids the singularity at zero in the conditional density of non–zero rainfall amounts. Further discussion of this point is given in the Supplementary Materials.

The model also has a number of other distinguishing features. The process governing transitions in the latent weather states is a non–homogeneous Markov process, with the transition probabilities depending on observed atmospheric data, in the form of objective Lamb weather types. This provides a link to models which might be used to predict or simulate Lamb weather types, for example under possible future climatic conditions. Given the hidden weather states, the rainfall observations on consecutive days are not independent as there is a direct dependence on the previous day's occurrence. This gives an additional mechanism for representing the temporal autocorrelation. We have used relatively complicated within-state models to represent the weather process. The benefit of this is that only a small number of weather states were required. Real prior information about the rainfall process is available and we have deliberately made provision for its use, and indeed used it, in specifying both the structure of the model and the prior distributions. For the reasons summarised above, we feel that our model is more appropriate for UK rainfall data than a NHMM using the TPTMVN distribution within weather–states. We have also demonstrated the superiority of our model to a reduced version which assumes conditional independence in space and time. In this latter case, it is likely that an impractically large number of weather states would be required to model the spatial structure in the data.

This is a complicated model, for a complicated phenomenon, and there is scope for further research to improve the model and methods. In particular it may be beneficial to include at least one hidden weather state in which, with certainty, it rains at every site. This would avoid the posterior distribution "trying" to replicate the effect with very large values of $\beta^i_{r,0,k}$; see Section 6.1. Similarly, it may be advantageous to include a state in which, with certainty, all sites are dry.

Acknowledgements

This paper arose from work in the PhD project of Sarah E. Heaps (née Germain) which was funded by the Engineering and Physical Sciences Research Council. Suggestions and remarks by the Joint Editor, an Associate Editor and two reviewers have improved the paper and are gratefully acknowledged. We thank the UK Meteorological Office for access to their MIDAS database. The Lamb weather type data are freely available from the Climate Research Unit website (`http://www.cru.uea.ac.uk/cru/data/lwt/`).

References

Ailliot, P., C. Thompson, and P. Thomson (2009). Space–time modelling of precipitation by using a hidden Markov model and censored Gaussian distributions. *Appl. Statist. 58*, 405–426.

Allcroft, D. J. and C. A. Glasbey (2003). A latent Gaussian Markov random–field model for spatiotemporal rainfall disaggregation. *Appl. Statist. 52*, 487–498.

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

Bellone, E., J. P. Hughes, and P. Guttorp (2000). A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts. *Climate Research 15*, 1–12.

Berrett, C. and C. A. Calder (2012). Data augmentation strategies for the Bayesian spatial probit regression model. *Comp. Statist. Data Anal. 56*(3), 478–490.

Betro, B., A. Bodini, and Q. A. Cossu (2008). Using a hidden Markov model to analyse extreme rainfall events in Central–East Sardinia. *Environmetrics 19*, 702–713.

Boys, R. J. and D. A. Henderson (2002). On determining the order of Markov dependence of an observed process governed by a hidden Markov model. *Sci. Program. 10*, 241–251.

Celeux, G., M. Hurn, and C. P. Robert (2000). Computational and inferential difficulties with mixture posterior distributions. *J. Amer. Statist. Ass. 95*, 957–970.

Charles, S. P., B. C. Bates, and J. P. Hughes (1999). A spatiotemporal model for downscaling precipitation occurrence and amounts. *J. Geophys. Res. 104*, 31657–31669.

Conway, D. and P. D. Jones (1998). The use of weather types and air flow indices for GCM downscaling. *J. Hydrol. 213*, 348–361.

Cox, D. R. and V. Isham (1994). Stochastic models of precipitation. In V. Barnett and K. F. Turkman (Eds.), *Statistics for the Environment 2: Water Related Issues*, Chichester, pp. 3–18. John Wiley & Sons Ltd.

Foufoula-Georgiou, E. (1987). A Markov renewal model for rainfall occurrences. *Water Resour. Res. 23*, 875–884.

Fowler, H. G. and C. G. Kilsby (2002). A weather–type approach to analysing water resource drought in the yorkshire region from 1881 to 1998. *J. Hydrol. 262*, 177–192.

Fowler, H. G., C. G. Kilsby, and P. E. O'Connell (2000). A stochastic rainfall model for the assessment of regional water resource systems under changed climatic conditions. *Hydrol. Earth Syst. Sci. 4*, 263–282.

Fowler, H. J., S. Blenkinsop, and C. Tebaldi (2007). Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology 27*(12), 1547–1578.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models.* New York: Springer.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1995). *Bayesian Data Analysis.* London: Chapman & Hall.

Germain, S. E. (2010). *Bayesian spatio-temporal modelling of rainfall through non-homogeneous hidden Markov models.* Ph. D. thesis, School of Mathematics & Statistics, Newcastle University, UK.

Geweke, J. (1991). Efficient simulation from the multivariate normal and student–t distributions subject to linear constraints and the evaluation of constraint probabilities. *Computer Sciences and Statistics 23*, 571–578.

Glickman, T. S. (2000). *Glossary of Meteorology* (Second ed.). Boston: American Meteorological Society.

Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Ass. 102*(477), 359–378.

Gschlößl, S. and C. Czado (2007). Spatial modelling of claim frequency and claim size in non–life insurance. *Scandinavian Actuarial Journal 3*, 202–225.

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

Hughes, J. P. and P. Guttorp (1994). A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resour. Res. 30*, 1535–1546.

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

Jenkinson, A. F. and B. P. Collinson (1977). An initial climatology of gales over the north sea. Synoptic climatology branch memorandum no. 62, Meteorological Office, London.

Katz, R. W. and M. B. Parlange (1993). Effects of a index of atmospheric circulation on stochastic properties of precipitation. *Water Resour. Res. 29*, 2335–2344.

Katz, R. W. and M. B. Parlange (1996). Mixtures of stochastic processes: application to statistical downscaling. *Climate Research 7*, 185–193.

Lamb, H. H. (1972). British Isles weather types and a register of daily sequence of circulation patterns, 1861–1971. Geophysical memoir 116, HMSO, London.

Liu, C. (2001). Comment on "The art of data augmentation". *J. Comput. Graph. Stat. 10*, 75–81.

Liu, X. and M. J. Daniels (2006). A new algorithm for simulating a correlation matrix based on parameter expansion and reparameterization. *J. Comput. Graph. Stat. 15*, 897–914.

MacDonald, I. L. and W. Zucchini (1997). *Hidden Markov and Other Models for Discrete–Valued Time Series*. London: Chapman & Hall.

Maraun, D., F. Wetterhall, A. M. Ireson, R. E. Chandler, E. J. Kendon, M. Widmann, S. Brienen, H. W. Rust, T. Sauter, M. Themeßl, V. K. C. Venema, K. P. Chun, C. M. Goodess, R. G. Jones, C. Onof, M. Vrac, and I. Thiele-Eich (2010). Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics 48*(3).

Met Office MIDAS database (2012). Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853-current). NCAS British Atmospheric Data Centre, cited October 2014. Available from `http://catalogue.ceda.ac.uk/uuid/220a65615218d5c9cc9e4785a3234bd0`.

Pourahmadi, M. (1999). Joint mean–covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika 86*, 677–690.

Rappold, A. G., A. E. Gelfand, and D. M. Holland (2008). Modelling mercury deposition through latent space–time processes. *Appl. Statist. 57*, 187–205.

Richardson, S. and P. J. Green (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *J. R. Statist. Soc. B 59*, 731–792.

Sansó, B. and L. Guenni (1999a). A stochastic model for tropical rainfall at a single location. *J. Hydrol. 214*, 64–73.

Sansó, B. and L. Guenni (1999b). Venezuelan rainfall data analysed by using a Bayesian space–time model. *Appl. Statist. 48*, 345–362.

Sansó, B. and L. Guenni (2000). A nonstationary multisite model for rainfall. *J. Amer. Statist. Ass. 95*, 1089–1100.

Smith, J. A. (1987). Statistical modeling of daily rainfall occurrences. *Water Resour. Res. 23*, 885–893.

Stephens, M. (2000). Dealing with label switching in mixture models. *J. R. Statist. Soc. B 62*, 795–809.

Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation (with discussion). *J. Amer. Statist. Ass. 82*, 528–550.

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

Velarde, L. G. C., H. S. Migon, and B. de Bragança Pereira (2004). Space–time modeling of rainfall data. *Environmetrics 15*, 561–576.

Vrac, M., M. Stein, and K. Hayhoe (2007). Statistical downscaling of precipitation through a nonhomogeneous stochastic weather typing approach. *Climate Research 34*, 169–184.

Webb, E. L. and J. J. Forster (2008). Bayesian model determination for multivariate ordinal and binary data. *Comp. Statist. Data Anal. 52*, 2632–2649.

Wilby, R. L. and T. M. L. Wigley (1997). Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography 21*(4), 530–548.

Zhang, X., W. J. Boscardin, and T. R. Belin (2006). Sampling correlation matrices in Bayesian models with correlated latent variables. *J. Comput. Graph. Stat. 15*, 880–896.

Zucchini, W. and P. Guttorp (1991). A hidden Markov model for space–time precipitation. *Water Resour. Res. 27*, 1917–1923.