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

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

Newcastle University, Newcastle-upon-Tyne, UK

[Received September 2013. Revised August 2014]

Summary. We present a non-homogeneous hidden Markov model for the spatiotemporal 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 on 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 non-homogeneous hidden Markov model, 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 12 sites located throughout the UK.

Keywords: Bayesian spatiotemporal analysis; Latent Gaussian variables; Non-homogeneous hidden Markov model; 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 by using general circulation models. These are complex, deterministic, mathematical models of the circulation of the atmosphere. Typically the resolution of general circulation model output is on a spatial scale of around 2–5° of longitude and latitude. However, questions of scientific interest, e.g. in hydrology or agriculture, often concern local patterns of precipitation over a much finer spatial scale. Addressing these questions by using general circulation model 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, which was introduced by Hay et al. (1991). These models assign each day to one of a small number of weather states by using observed atmospheric information. Typically these weather states are observable. Pre-
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 rainfall by 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 by using hidden Markov models (HMMs). In this paper we propose an 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 data set of daily precipitation measurements over 28 consecutive winters at 12 sites in the UK. The atmospheric data used are objective Lamb weather types (LWTs) which provide a classification of synoptic weather into 27 categories, based on surface pressure around the British Isles. To capture the spatiotemporal 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 models for precipitation occurrences. Non-zero precipitation amounts are given a multivariate log-normal distribution. The mean of the corresponding multivariate normal distribution depends linearly on the latent normal random vector underlying the multivariate probit model. This offers some advantages over the truncated, power-transformed multivariate normal distribution which is often used for multisite 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 data set that is analysed in subsequent sections. Section 3 surveys the literature on the spatiotemporal 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 data set, including a summary of the resulting posterior distribution and the results of model checks, which compare the posterior predictive distribution with data which were not used in model fitting.

The data that are analysed in the paper are available from the Met Office integrated data archive system MIDAS (Met Office, 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. For this, the precipitation data that are analysed in Section 6 are from a network of 12 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 data set comprises 2527 daily precipitation totals recorded at each of the 12 sites over the 28 winter periods from 1961–1962 to 1988–1989, which include seven leap years.
The sites were chosen to give good spatial coverage over the UK. Chapter 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.

Fig. 1(a) shows the locations of the sites whereas 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 in the data, with sites that are closer showing more similarity than those further apart. To illustrate, Fig. 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 $\log\left[\frac{n_{0,0}(i, j) n_{1,1}(i, j)}{n_{0,1}(i, j) n_{1,0}(i, j)}\right]$ where $n_{d,k}(i, j)$ is the observed number of days where the rainfall occurrence indicator is equal to $d$ at site $i$ and $k$ 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 LWTs have been used extensively for characterizing 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 used 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

![Fig. 1.](image-url)
Table 1. Summary of data from the UK network for winter periods from 1961–1962 to 1988–1989†

| Site            | Altitude (m) | Proportion wet days (%) | Mean daily precipitation (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                    |

†The mean and coefficient of variation for daily precipitation are based on wet days only.

Table 2. Labelling of the objective LWTs and frequencies from 1961–1962 to 1988–1989

| 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        |

specified height in the atmosphere. Under this scheme an expert analyst can use their judgement to determine the weather type on any day 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 by 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 categorization 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 un-classifiable (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, whereas days on which the vorticity is strongly positive or negative will be categorized 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 1961–1962 to 1988–1989 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 the on-line 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 or low amounts are typically associated respectively with cyclonic or 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 precipitation amounts are associated with LWTs 6 and 7, 13–16 and 22–25, the majority of which are westerly types. A different pattern is observed at Plymouth, which is a low altitude site on the south coast, where it seems that southerly and easterly LWTs tend to be associated with heavier precipitation.

3. Spatiotemporal rainfall modelling through empirical statistical models

In the terminology of Cox and Isham (1994) the model that is presented in this paper is an empirical statistical model for rainfall, describing only the observations of aggregated precipitation, and not the underlying physical phenomena. Spatiotemporal 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 spatiotemporal 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, e.g. 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 spatiotemporal 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 an HMM. Compared with 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 spatiotemporal 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-site and multisite 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 that was proposed by MacDonald and Zucchini (1997) is unique in allowing the logit of the hidden state transition probabilities to vary smoothly across seasons by 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 an NHMM. Hughes et al. (1999) presented a more sophisticated NHMM for precipitation occurrence in which within-weather-state spatial dependence was modelled explicitly by using an autologistic model.

Extensions to HMMs to include amounts of precipitation 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 an 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 amounts of precipitation on wet days. Thompson et al. (2007) presented a three-state (partially) HMM 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 the observed (given hidden) processes is built by using copulas. Finally, Ailliot et al. (2009) presented an HMM in which spatial dependence in the within-weather-state joint distributions for precipitation occurrence and amount was modelled explicitly by using a truncated, power-transformed multivariate normal 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 \( W_i = \mathbb{I}(Z_i > \alpha_0) g_1(Z_i, \alpha_i), \) for \( i = 1, \ldots, n, \) where \( \mathbb{I}(A) = 1 \) if \( A \) is true and \( \mathbb{I}(A) = 0 \) otherwise. \( W = (W_1, \ldots, W_n)^T \) denotes a vector of precipitation amounts at a network of \( n \) sites, \( Z = (Z_1, \ldots, Z_n)^T \) is a multivariate normal random vector, \( \alpha_0 \) is a threshold parameter and \( g_1(Z_i, \alpha_i) \) is a transformation function which may depend on parameters \( \alpha_i. \) Typically \( \alpha_0 = 0 \) and \( g_1(Z_i, \alpha_i) \) 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 \( \alpha_i = \alpha_i \) and \( g_1(Z_i, \alpha_i) = (Z_i)^{\alpha_i}. \) This induces a heavy-tailed distribution for non-zero rainfall and gives \( W \) a truncated, power-transformed multivariate normal distribution. Ailliot et al. (2009) introduced temporal dependence in their model by embedding the truncated power-transformed multivariate normal distribution in an HMM. Other researchers have taken different approaches, e.g. by incorporating \( Z \) in a vector auto-regression (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
Bayesian Modelling of Rainfall Data

structure were captured by allowing the logit of the probability of occurrence 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 that have been 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 spatiotemporal dependence. The latent normal variables enter through the expression \( W_t = (Z_0^t > 0) 2^t Z_1^t \) where \( Z_0 = (Z_0^1, \ldots, Z_0^n)^T \) and \( Z_1 = (Z_1^1, \ldots, Z_1^n)^T \) are two correlated multivariate normal vectors. This approach offers potential advantages over the truncated, power-transformed multivariate normal distribution with its single normal vector.

4. Model and prior

Consider a network of \( n \) sites. Let \( D_t = (D_t^1, \ldots, D_t^n)^T \) be a random vector for rainfall occurrences where \( D_t^i = 1 \) if there are 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 that inference is insensitive to this choice. Let \( W_t = (W_t^1, \ldots, W_t^n)^T \) 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 \( D_t \) and \( W_t \) be \( w \) and \( d \).

The number \( r \in \{1, \ldots, r_{\text{max}}\} \) 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, \( p(w, d|\theta_r, r) \). Each of the \( r_{\text{max}} \) (conditional) likelihoods has its own set of parameters, \( \theta_r \), to which we assign a prior, \( \pi(\theta_r|\theta) \). In Section 4.1 we describe the likelihood (i.e. the model) and then, in Section 4.2, the priors \( \pi(\theta_r|\theta) \), \( r \in \{1, \ldots, r_{\text{max}}\} \). For notational clarity, unless stated otherwise, dependence on \( r \) is assumed without explicit notational reference. For example, we generally refer to \( p(w, d|\theta, r) \) and \( \pi(\theta) \) rather than \( p(w, d|\theta_r, r) \) and \( \pi(\theta_r|\theta) \). 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 categorizes the rainfall patterns on each day. We interpret \( S_t \) as the weather state on day \( t \) and denote its state space by \( S_r = \{1, \ldots, r\} \). Denote by \( X_t \) the atmospheric data on day \( t \). Both \( S_t \) and \( X_t \) are common to all sites in the network. In their NHMMs for rainfall, other researchers (e.g. Hughes and Guttorp (1994), Hughes et al. (1999) and 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 \( X_t \) provides a summary of atmospheric conditions over the region of interest on day \( t \). In our case \( X_t = X_t \in \{1, \ldots, 27\} \) is a categorical covariate, namely the observed LWT on day \( t \), labelled according to Table 2.

We denote the parameters of our NHMM by \( \theta = (\theta_{\text{hid}}, \theta_{\text{obs}}) \), partitioned so that \( \theta_{\text{hid}} \) parameterizes the weather state process and \( \theta_{\text{obs}} \) parameterizes the observed process. Let \( y_{i,j} \) 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:
dependence structure so that dry spells that were observed in the data. We therefore allow a refinement to the temporal models with this temporal structure were often unable to predict the longer duration wet and earlier applications involving UK rainfall data (see Germain (2010), chapter 4) we found that related multivariate normal random vectors distribution of non-zero amounts are related over time and between sites, we introduce two cor-

amount goes to 0. We have adopted a distribution which satisfies this requirement, namely the distribution for the weather state $S_t$ to learn about the initial distribution $\nu$. We note that assumption (1) describes a conditional distribution for the weather state $S_t$ given $S_{t-1}$ and the LWT $X_t$. In statistical downscaling, a general circulation model 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 assumption (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 assumption (1) it can be seen that we have a model for the weather states such that, for every combination of lag $l$ weather state $S_{t-1} = j$ and current LWT $X_t = x$, a different stochastic vector $\lambda^x_j$ governs the probabilities of transition into the current weather state. This parameterization offers the possibility of a conjugate Dirichlet prior distribution for each $\lambda^x_j$. We denote the collection of transition probabilities by $\Lambda = (\lambda^1_1, \ldots, \lambda^x_j)$ and so $\theta_{\text{hid}} = (\Lambda, \nu)$. As some of the LWTs occur very infrequently (see Table 2), a prior for the $\lambda^x_j$ 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 0. We have adopted a distribution which satisfies this requirement, namely the log-normal 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 corre-

related multivariate normal random vectors $Z_{0,t} = (Z_{0_{1,t}}, \ldots, Z_{0_{n,t}})^T$ and $Z_{1,t} = (Z^1_{1,t}, \ldots, Z^n_{1,t})^T$ for $t = 1, \ldots, T$. We define the observable rainfall at site $i$ on day $t$ as

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

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) expressed rainfall as a function of two normal variables in the same way in their spatiotemporal 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 auto-correlation in the rainfall data. In the model that is described here, this would correspond to an assumption that the bivariate latent process \{$Z_{0,t}, Z_{1,t}$\} 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
Bayesian Modelling of Rainfall Data

\[ p(\mathbf{z}_{0:t}, \mathbf{z}_{1:t}| \mathbf{z}_{0:1:t-1}, \mathbf{z}_{1:1:t-1}, D_0, S_{0:T}, X_{1:T}, \theta_{\text{obs}}) = p(\mathbf{z}_{0:t}|d_{t-1}, S_t = k, \theta_{\text{obs}}) p(\mathbf{z}_{1:t}|\mathbf{z}_{0:t}, S_t = k, \theta_{\text{obs}}), \]

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

\[ \Pr(\mathbf{d}_0|S_{0:T}, X_{1:T}, \theta_{\text{obs}}) = \prod_{i=1}^{n} \Pr(D_{0i}^i) \quad D_{0i}^i \sim \text{Bern}(p_i), \]

where each \( p_i \in [0, 1] \) is fixed. Note that \( \mathbf{z}_{0:t} \) depends on the previous day’s rainfall occurrence indicator \( D_{t-1} \). We specify

\[ \mathbf{z}_{0:t}|D_{t-1} = d_{t-1}, S_t = k, \theta_{\text{obs}} \sim N_n(\mu_{t,k}, \Sigma_k), \]

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

\[ \mathbf{z}_{1:t}|\mathbf{z}_{0:t}, S_t = k, \theta_{\text{obs}} \sim N_n(\alpha_k + \Gamma_k \mathbf{z}_{0:t}, \Omega_k), \]

where \( \alpha_k \) is an \( n \)-vector, \( \Gamma_k \) is an \( n \times n \) matrix and \( \Omega_k \) is an \( n \times n \) symmetric positive definite matrix. Experience has shown that, unless the number of sites \( n \) is small, allowing a different \( \Gamma_k \)-matrix for each state \( k \) compromises the performance of the MCMC sampler. We therefore assume a constant matrix \( \Gamma_k = \Gamma \) for all \( k \in S_t \). In the special case where each element in \( \Gamma \) is 0 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 \( \text{var}(\mathbf{z}_{1:t}|\mathbf{d}_{t-1} = d_{t-1}, S_t = k, \theta_{\text{obs}}) = \Omega_k + \Gamma_k \Sigma_k \Gamma^T \) and so, because \( \Gamma \) is non-diagonal, we can make the covariance matrices \( \Omega_k \) diagonal and still allow within-state spatial dependence between the elements of \( \mathbf{z}_{1:t} \). Therefore, 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 \( \mathbf{z}_{0:t} \) and \( S_t \) and all the within-state spatial dependence is carried by \( \mathbf{z}_{0:t} \). Note, however, that \( Z_{1,t}^i \) is not conditionally independent of other sites given just \( \mathbf{z}_{0:t} \) and this captures the fact that, under some conditions, the amount of rain that 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 \( \mathbf{z}_{1:t} \), and then a conditional distribution for \( \mathbf{z}_{0:t} \), given \( \mathbf{z}_{1:t} \), in which the \( \mathbf{z}_{0:t} \), were conditionally independent. In this case all of the within-state spatial dependence would have been captured by \( \mathbf{z}_{1:t} \). However, we chose our formulation because rainfall modellers usually specify a distribution for rainfall occurrence and then a distribution for rainfall amounts given occurrence. Allowing \( \mathbf{z}_{0:t} \) 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 \( \{ \mathbf{z}_{0:t} \} \)) through

\[ p(\mathbf{z}_{1:t}|S_t, \mathbf{z}_{0:t}, \theta_{\text{obs}}) = \prod_{i=1}^{n} p(z_{1,t}^i|s_t, \mathbf{z}_{0:t}, \theta_{\text{obs}}), \]

where

\[ Z_{1,t}^i|S_t = k, \mathbf{z}_{0:t} = \mathbf{z}_{0:t} \sim N \left( \alpha_k^i + \sum_{j=1}^{n} \Gamma_i^j z_{0,t}^j, \omega_{k,i}^2 \right), \quad W_i^j = D_i^j \exp(Z_{1,t}^i). \]

This is the representation of the within-state model on which we shall 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 $W_t$, the model that was defined above would reduce to a Markov chain of multivariate probit models, conditional on the weather state. To ensure parameter identifiability in the observed data likelihood of multivariate probit models, constraints are necessary to prevent arbitrary rescaling of the linear predictor $\mu_{t,k}$ or the covariance matrix $\Sigma_k$. Although our model describes amounts of rainfall, as well as occurrences, this problem of non-identifiability is not remedied because changes to the scale of $Z_{0,t}$ could exactly compensate for changes in the scale of $\Gamma$. However, once the scale and location of $Z_{0,t}$ have been fixed, all parameters are identifiable; see the on-line supplementary materials for further explanation and a numerical demonstration.

With multivariate probit models, a common means of fixing the scale and location of $Z_{0,t}$ 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 owing to the complex constraints on the space of correlation matrices. These constraints also make sampling a correlation matrix during MCMC sampling 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 $\Sigma_k^{-1}$ (as described in Section 4.2) and then to fix the values of the conditional variances arising from this reparameterization; see, for example, Webb and Forster (2008). However, we found that this led to poor mixing during MCMC sampling. We therefore avoid placing constraints on the covariance matrices $\Sigma_k$ and instead fix the scale of the coefficients $\beta_{1,k} = (\beta_{1,k}^1, \ldots, \beta_{1,k}^m)^T$ so that $\beta_{1,k}^i \in \{-1, 1\}$ for each $i = 1, \ldots, n$ and each $k = 1, \ldots, n$. Note that, in weather state $k$, the marginal site $i$ probabilities of rain after no rain and rain after rain are $\Phi(\beta_{0,k}^i / \sqrt{\Sigma_k^{ii}})$ and $\Phi(\beta_{0,k}^i / \sqrt{\Sigma_k^{ii}} + \beta_{1,k}^i / \sqrt{\Sigma_k^{ii}})$ respectively. Therefore our constraint still allows the probability of rain after rain to be (any amount) more than $\beta_{1,k}^i = 1$, less than $\beta_{1,k}^i = -1$ or equal to $\Sigma_k^{ii}$ large the probability of rain after no rain.

We note that, whereas the transition probabilities $\lambda_{i,k}$ for changes between weather states described in Section 4.1.1 are, of course, common to all sites, the within-state model that is 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 factorization of the joint distribution of $\{(S_t, W_t, D_t, Z_{0,t}) : t = 1, \ldots, T\}$ conditionally on $\{X_t : t = 1, \ldots, T\}$, $D_0$ and $S_0$ is represented in the directed acyclic graph in Fig. 2(a) where the double circles show deterministic dependence. Fig. 2(b) shows the factorization of the joint distribution for $\{(S_t, W_t, D_t) : t = 1, \ldots, T\}$ that arises after marginalizing over $Z_{0,t}$. Note that $(W_t, D_t)$ are conditionally independent of the LWT $X_t$ given $S_t$ and $D_{t-1}$ and so the influence of the atmospheric data is only through the evolution of the weather states. From Fig. 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 spatiotemporal dependence structure.

Our use of a second multivariate normal random variable $Z_{1,t}$ means that we do not restrict the amount of rainfall to be a deterministic function of the latent variables $Z_{0,t}$ governing occurrence.
As discussed in Section 3, other researchers have captured spatial dependence between non-zero rainfall amounts by using generalized linear spatial process models. With $Z_{0,t}$ playing the role of the vector of spatial random effects, the model that is defined through equations (3) and (4) is similar to a generalized 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 $\pi(\theta_r|\theta_r) = \pi(\theta_{r,\text{hid}}|\theta_{r,\text{obs}})$. 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 $\pi(\theta_{r,\text{obs}}|\theta)$ for all $r \in \{1, \ldots, r_{\text{max}}\}$. 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 $Y = (Y^1, \ldots, Y^n)^T | \mu, \Sigma \sim \mathcal{N}_n(\mu, \Sigma)$. A transformation based on the square-root free Cholesky decomposition of the precision matrix is given by $\Sigma^{-1} = (I_n - \Phi)^T \Psi (I_n - \Phi)$ where $\Psi$ is a diagonal matrix with positive diagonal entries $\psi_i^2 \in \mathbb{R}^+$, $i = 1, \ldots, n$, and $\Phi$ 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 by using covariates. The parameters in $\Phi$ and $\Psi$ have an auto-regressive interpretation based on the marginal or conditional decomposition of the joint density of $Y = (Y^1, \ldots, Y^n)^T$. Specifically, for $i > 1$, $\phi_{i,1}, \ldots, \phi_{i,i-1}$ are the slope coefficients in the regression of $Y^i$ on its (mean-centred) predecessors $Y^1, \ldots, Y^{i-1}$ whereas $\psi_i^2$ is the conditional variance in the auto-regression. This generalized auto-regression requires an ordering of the elements in $Y$. Therefore, our strategy when applying this reparameterization to the state-specific covariance matrices $\Sigma_1, \ldots, \Sigma_r$ 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 $\tilde{\Sigma}_k = M \Sigma_k M^T$ into new parameters $\{\tilde{\phi}_k, \tilde{\psi}_k\}: k = 1, \ldots, r$ where $\tilde{\phi}_k = (\phi_{k,1}, \phi_{k,2}, \ldots, \phi_{k,n})^T$ and $\tilde{\psi}_k = \text{diag}(\tilde{\psi}_{k,1}, \ldots, \tilde{\psi}_{k,n})$. We can then
choose to make the slope coefficients \((\tilde{\phi}_1, \ldots, \tilde{\phi}_r)\) and the conditional variances \((\tilde{\Psi}_1, \ldots, \tilde{\Psi}_r)\) independent \(a\ priori\); for further details, see the on-line supplementary materials.

Prior uncertainty about the model parameters is expressed through a prior of the form 
\[
\pi(\theta) = \pi(\theta_{\text{obs}}) \pi(\theta_{\text{hid}})
\]
where
\[
\pi(\theta_{\text{obs}}) = \pi(\beta_{0,1}, \ldots, \beta_{0,r}) \pi(\beta_{1,1}, \ldots, \beta_{1,r}) \pi(\tilde{\phi}_1, \ldots, \tilde{\phi}_r) \pi(\tilde{\Psi}_1, \ldots, \tilde{\Psi}_r)
\]
and
\[
\pi(\theta_{\text{hid}}) = \pi(\Lambda) \pi(\nu).
\]
Note that the product \(\pi(\tilde{\phi}_1, \ldots, \tilde{\phi}_r)\pi(\tilde{\Psi}_1, \ldots, \tilde{\Psi}_r)\) induces a joint prior for \((\Sigma_1, \ldots, \Sigma_r)\).

Consider first the parameters \(\theta_{\text{obs}}\) of the observed process. We assume \(a\ priori\) independence between weather states for the parameter blocks \((\alpha_1, \ldots, \alpha_r)\) and \((\Omega_1, \ldots, \Omega_r)\) in the process \(\{W_i | D_i, S_i, Z_{0,i}\}\) and for the parameter blocks \((\beta_{0,1}, \ldots, \beta_{0,r}), (\beta_{1,1}, \ldots, \beta_{1,r})\) and \((\tilde{\Psi}_1, \ldots, \tilde{\Psi}_r)\) in the process \(\{Z_{0,i} | S_i, D_{i-1}\}\). Within each weather state, for each of these parameter blocks, we then adopt hierarchical priors which induce positive correlation between sites while maintaining semiconjugacy in the prior specification: \(\alpha_i^r \sim N(\alpha_{0i}, \sigma^2_{\alpha_i})\) and \(\sigma^2_{\alpha_i} \sim IG(h_{0\alpha}, h_{1\alpha})\) independently for \(i = 1, \ldots, n\). Then \(\mu_{u1i} \sim N(\mu_{0u1}, \sigma^2_{u1i})\) and \(\sigma^2_{u1i} \sim IG(h_{0\alpha}, h_{1\alpha})\), where \(‘IG’\) denotes an inverse gamma distribution. Similarly, for \(k = 1, \ldots, r\),
\[
\begin{align*}
\omega^2_{k,i} | \mu^2_{x_i} & \sim IG(v^2_{x_i} + 2, \mu^2_{x_i}(v^2_{x_i} + 1)), \\
\beta^2_{0,k} | \mu_{0i,k} & \sim N(\mu_{0i,k}, \sigma^2_{0i,k}), \\
\beta^2_{1,k} | \mu_{1i,k} & \sim \text{ScBern}(\mu_{1i,k}), \\
\psi^2_{k,i} & \sim \text{Bern}(p), \text{ scaled to have support on } \{-1, 1\}, \text{ i.e. } X = 2Y - 1 \text{ where } Y \sim \text{Bern}(p).
\end{align*}
\]
Here the notation \(X \sim \text{ScBern}(p)\) means that the random variable \(X\) has a Bernoulli distribution, \(\text{Bern}(p)\), scaled to have support on \(\{-1, 1\}\), i.e. \(X = 2Y - 1\) where \(Y \sim \text{Bern}(p)\).

The purpose of the fixed hyperparameters \(C_i > 0, i = 1, \ldots, n\), in the prior for the conditional variances \(\psi^2_{k,i}\) is to allow the marginal prior means \(E(\psi^2_{k,i})\) to differ across sites \(i = 1, \ldots, n\). This reflects the fact that the conditional variances are not exchangeable in our prior beliefs because, as \(i\) increases from \(1\) to \(n\), \(\psi^2_{k,i}\) represents the residual variance after \(Z_{0,i}\) has been regressed on an increasing number of predecessors \(Z_{1,i}, \ldots, Z_{n-1,i}\). We note that the prior that was chosen for the \(\beta^2_{0,k}\) 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 \((\tilde{\phi}_1, \ldots, \tilde{\phi}_r)\) and \((\tilde{\Psi}_1, \ldots, \tilde{\Psi}_r)\). 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 Alliott et al. (2009)) or by assuming a common covariance matrix \(\Sigma = \Sigma\) for all \(k \in S_r\). 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 while allowing the data to inform us of differences between them. It is achieved through the hierarchical specification
\[
\mu^2_{z_i} \sim \text{IG}(\sigma^2_{z_i}, 1), \quad \text{for } i = 1, \ldots, r, \text{ with } \mu^2_{z_i} \sim N(n(n-1)/2, \sigma^2_{z_i}), \quad \text{for } k = 1, \ldots, r, \text{ with } \mu^2_{z_i} \sim N(n(n-1)/2, \sigma^2_{z_i}), \quad \text{for } k = 1, \ldots, r, \text{ and } \text{cov}(\phi_k, \phi_l) = C_{\phi_k} \text{ for each } k \neq l.
\]

Marginalizing over \(\mu^2_{z_i}\) leads to a joint multivariate normal prior for \((\tilde{\phi}_1, \ldots, \tilde{\phi}_r)\) in which
\[
E(\phi_k) = \mu^2_{z_i}, \quad \text{var}(\phi_k) = V^2_{\phi} = V^2_{\phi} + C_{\phi_i} \quad \text{for each } k = 1, \ldots, r, \text{ and } \text{cov}(\phi_k, \phi_l) = C_{\phi_i} \text{ for each } k \neq l.
For simplicity, we elicit \( \tilde{V}_j \) and then take \( C_z = \rho_z \tilde{V}_z \), where \( \rho_z \in (0, 1) \) is fixed. This means that, for \( k \neq l \), \( \text{corr}(\tilde{\phi}_{k,i,v}, \tilde{\phi}_{l,v}) = \rho_{z_k} \text{corr}(\phi_{k,i,v}, \phi_{l,v}) \).

The \( i \)th row of \( \Gamma \) comprises coefficients \( \Gamma_{i,j}, j = 1, \ldots, n \), in the regression of \( \log(W^j_i) \) on the latent variables \( Z_{0,i}^j, j = 1, \ldots, n \). A priori, we believe that the effect of \( Z_{0,i}^j \) on \( \log(W^j_i) \) 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 \( \Gamma \). For the collections of on- and off-diagonal elements we then encourage borrowing of strength by choosing semiconjugate hierarchical priors with first-level specifications

\[
\Gamma_{i,i} \mid \mu_{\text{on}}, \sigma_{\text{on}}^2 \sim N(\mu_{\text{on}}, \sigma_{\text{on}}^2) \quad \text{independently for } i = 1, \ldots, n, \\
\Gamma_{i,j} \mid \mu_{\text{off}}, \sigma_{\text{off}}^2 \sim N(\mu_{\text{off}}, \sigma_{\text{off}}^2) \quad \text{independently for } i, j = 1, \ldots, n, i \neq j,
\]

and second-level specifications \( \mu_{\text{on}} \sim N(a_{0, \text{on}}, a_{0, \text{on}}^2) \), \( \sigma_{\text{on}}^2 \sim \text{IG}(h_{0, \text{on}}, h_{1, \text{on}}) \), \( \mu_{\text{off}} \sim N(a_{0, \text{off}}, a_{0, \text{off}}^2) \) and \( \sigma_{\text{off}}^2 \sim \text{IG}(h_{0, \text{off}}, h_{1, \text{off}}) \).

Consider now the parameters of the hidden process \( \theta_{\text{hid}} \). The initial distribution \( \nu \) is assigned a conjugate Dirichlet prior \( \nu \sim \mathcal{D}_r(Gg) \) where \( g = E(\nu) \in \mathcal{S}_r \), the \( r \)-dimensional simplex, and \( G \in \mathbb{R}^r \). The assumption of a priori exchangeability across weather states requires that \( g = (1/r, \ldots, 1/r) \). We then choose \( G = r \) to give a flat Dirichlet \( \mathcal{D}_r(1, \ldots, 1) \) 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 \( \lambda^x_j = (\lambda^x_{j,1}, \ldots, \lambda^x_{j,x}) \) where \( \lambda_{j,k}^x = \text{Pr}(S_t = k|S_{t-1} = j, X_t = x, \theta_{\text{hid}}) \). We can, again, facilitate (indirect) learning about some of the more rare \((j, x)\) combinations by adopting the hierarchical Dirichlet prior

\[
\lambda^x_j | \xi_j \overset{\text{ID}}{\sim} \mathcal{D}_r(\Xi_j \xi_j), \quad \xi_j \sim \mathcal{D}_r(E_j e_j), \quad (5)
\]

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

Analogously to the precision parameters in a normal hierarchical prior, the parameters \( \Xi_j \) and \( E_j \) reflect the amounts of specific and common information in \( \xi_j \). 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 \( \lambda^x_j \).

The parameters \( \{\mu_{0,k}, \mu_{1,x}, \mu_{2}, \sigma_{0,x}, \sigma_{0}^2, \mu_{\alpha}^x, \sigma_{\alpha}^2, \sigma_{\text{off}}^2, \sigma_{\text{on}}^2\} : k = 1, \ldots, r\) of \( \mu_{\alpha}^x, \mu_{\text{off}}, \sigma_{\text{off}}^2, \sigma_{\text{on}}^2 \) which were given distributions at the second level in the hierarchical prior specifications above are appended to \( \theta_{\text{hid}} \). For convenience we introduce the notation \( \mathcal{E} = (\xi_1, \ldots, \xi_r) \in \mathcal{S}_r \) and then append \( \mathcal{E} \) to the \( \theta_{\text{hid}} \). Our prior for the complete set of model parameters \( \theta = (\theta_{\text{obs}}, \theta_{\text{hid}}) \) may then be written as

\[
\pi(\theta) = \prod_y \left[ \prod_{k=1}^r \left\{ \pi(\beta_{0,k} | \mu_{0,k}) \pi(\beta_{1,k} | \mu_{1,k}) \pi(\beta_{2,k} | \mu_{2,k}) \pi(\tilde{\beta}_{k} | \mu_{2,k}) \pi(\alpha_k | \mu_{\alpha_k}, \sigma_{\alpha_k}^2) \right. \\
\times \pi(\mu_{\alpha_k} | \sigma_{\alpha_k}^2) \pi(\mu_{\text{off}} | \sigma_{\text{off}}^2) \pi(\mu_{\text{on}} | \sigma_{\text{on}}^2) \pi(\sigma_{\text{off}}^2 | \sigma_{\text{on}}^2) \pi(\sigma_{\text{off}}^2 | \sigma_{\text{on}}^2) \\
\left. \times \pi(\mu_{\text{off}} | \sigma_{\text{off}}^2) \pi(\sigma_{\text{off}}^2 | \sigma_{\text{on}}^2) \pi(\sigma_{\text{off}}^2 | \sigma_{\text{on}}^2) \right\} \pi(\Lambda | \mathcal{E}) \pi(\mathcal{E} | \nu) \right].
\]
Let the length of subseries realizations. In our application each subseries refers to the winter months in a particular year. We consider data in the form of a collection of time series which are treated as independent subseries (year) $y$.

We first consider inferences given a fixed value of $r$. For this 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 using data augmentation is non-standard and so we sample $(\Lambda_j, \xi_j)$ by using a Metropolis–Hastings step for $j = 1, \ldots, r$.

$\pi(\theta, s, s_0, d_0, z_0|w, d, x)$ in a series of Gibbs (or Metropolis-within-Gibbs) steps. The full conditional distributions for all parameters in $\theta_{obs}$ and for the initial distribution $\nu$ are standard distributions and can be sampled directly. The joint full conditional distribution is non-standard and so we sample $(\Lambda, \xi)$ by using a Metropolis–Hastings step for $j = 1, \ldots, r$. This is achieved using a forward–backward simulation algorithm (see, for example Frühwirth-Schnatter (2006), algorithm 11.5) in which $z_0$ and $d_0$ are treated in the same manner as observed data.

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

Step 1: sample $\theta$ from its conditional posterior distribution $\pi(\theta|w, d, d_0, s, s_0, z_0, x)$ in a series of Gibbs steps. The full conditional distributions for all parameters in $\theta_{obs}$ and for the initial distribution $\nu$ are standard distributions and can be sampled directly. The joint full conditional distribution is non-standard and so we sample $(\Lambda_j, \xi_j)$ by using a Metropolis–Hastings step for $j = 1, \ldots, r$.

Step 2: sample $(s, s_0)$ from its conditional posterior $\pi(s, s_0|w, d, d_0, z_0, \theta, x)$. This is achieved using a forward–backward simulation algorithm (see, for example Frühwirth-Schnatter (2006), algorithm 11.5) in which $z_0$ and $d_0$ are treated in the same manner as observed data.

The algorithm is applied separately to each subseries $y$.

Step 3: sample $z_0$ from its conditional posterior, $\pi(z_0|w, d, d_0, s, \theta_{obs})$. When $d_1, \ldots, d_T$, $y$ are all observed, the latent variables $Z_0, 1, \ldots, Z_0, T_y$ are independent in this joint distribution.

Step 4: sample $d_0$ from its conditional posterior $\pi(d_0|z_0, s, \theta_{obs})$. The initial occurrences $d_1, 0, \ldots, d_T, 0$ are conditionally independent in their joint full conditional distribution. For each subseries $y$, the components of $d_{y, 0}$ are sampled one at a time from their univariate Bernoulli conditionals.

Full details of this scheme can be found in the on-line 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 $\theta$ and the weather states $s$ conditional on $r$, and this requires a distinct labelling of the states. To overcome this problem we use an on-line relabelling algorithm which was 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 \( r \)

It is now necessary to introduce notational dependence on \( r \). For example, we denote the parameters of the hidden process in an \( r \)-state NHMM by \( \theta_{r,\text{hid}} = (\Lambda_r, \mathcal{E}_r, \nu_r) \).

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

\[
\pi_r(r|w,d,x) = \frac{p(w,d|x,r) \pi_r(r)}{\sum_{k=1}^{r_{\text{max}}} p(w,d|x,k) \pi_r(k)}
\]

in which the marginal likelihood, \( p(w,d|x,r) \), is the normalizing constant in the conditional posterior distribution of \( \theta_r \) given \( r \):

\[
p(w,d|x,r) = \int p(w,d|\theta_r, x, r) \pi(\theta_r|r) \, d\theta_r.
\]

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 \( \pi(\theta_r, w, d, x) \), whereas the latter approximate the marginal likelihood for each model \( r \) in turn and then compute the posterior for \( r \) through application of equation (6). Unfortunately we could not find a workable method of either kind; see the on-line 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 materialized. Gschlößl and Czado (2007) considered the use of proper scoring rules in the context of Bayesian model comparison in which observations from an *out-of-sample period*, i.e. 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 \((w^{\text{rep}}, d^{\text{rep}})\) that could have been observed under the model with \( r \) states is given by

\[
p(w^{\text{rep}}, d^{\text{rep}}|w,d,x,r) = \int p(w^{\text{rep}}, d^{\text{rep}}|\theta_r, x, r) \pi(\theta_r|w, d, x, r) \, d\theta_r.
\]

From this equation we can deduce, for example, the marginal posterior predictive distribution for site \( i \) and day \( t \): \( p(w_{i,\text{rep},t}, d_{t,\text{rep}}|w, d, x, r) \). Given the intractability of the posterior distribution for \( r \), we shall 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, which was presented in Gneiting and Raftery (2007), is the continuous ranked probability score (CRPS). Omitting notational reference to \( d \), this is defined as

\[
\text{CRPS}_i(F_{i,t}^j, w_{i,t}^j) = -\int_0^\infty \{ F_{i,t}^j(y|r) - I(y \geq w_{i,t}^j) \}^2 \, dy,
\]

for site \( i \) and time \( t \). Here \( F_{i,t}^j(\cdot|r) \) is the posterior predictive distribution for rainfall at site \( i \) on day \( t \) given an \( r \)-state model, i.e. the distribution function corresponding to the marginal density \( p(w_{i,\text{rep},t}^j, d_{t,\text{rep}}|w, d, x, r) \). This scoring rule assigns the highest rewards to predictive distribution functions which are very concentrated around the observation which ultimately materializes. The CRPS can also be written as
CRPS\(_r\)(F\(_t^i\), W\(_t^i\)) = \frac{1}{2} E_{F_t^i}(|W_{t}^{\text{rep},i} - W_{t}^{\text{rep},i'}|) - E_{F_t^i}(|W_{t}^{\text{rep},i} - W_{t}^i|), \tag{8}

in which \(W_{t}^{\text{rep},i}\) and \(W_{t}^{\text{rep},i'}\) 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 \(w_{t}^{\text{rep},i,1}, \ldots, w_{t}^{\text{rep},i,N}\) when the terms on the right-hand side of equation (8) can be approximated through, for example,

\[
E_{F_t^i}(|W_{t}^{\text{rep},i} - W_{t}^{\text{rep},i'}|) \simeq \frac{1}{N/2} \sum_{j=1}^{N/2} |w_{t}^{\text{rep},i,[2j-1]} - w_{t}^{\text{rep},i,[2j]}| \tag{9}
\]

and

\[
E_{F_t^i}(|W_{t}^{\text{rep},i} - W_{t}^i|) \simeq \frac{1}{N} \sum_{j=1}^{N} |w_{t}^{\text{rep},i,[j]} - w_{t}^i|. \tag{10}
\]

It is straightforward to generate the samples that are needed to evaluate approximations (9) and (10) given LWT data from an out-of-sample period and an approximately unauto-correlated sample \(\theta_{j}, j = 1, \ldots, N\), from the posterior distribution of the model parameters. This is achieved by generating a sample \(w_{t}^{\text{rep},i,j} = (w_{t}^{\text{rep},1,j}, \ldots, w_{t}^{\text{rep},n,j})\), \(t = 1, 2, \ldots\), from the model \(p(w_{t}^{\text{rep},i,j} | \theta_{j}, x, r)\) for each \(j = 1, \ldots, N\), in which the LWTs \(x\) 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 equation (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 practice 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 sitewise 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 on-line supplementary materials.

6. Application to UK winter rainfall data

In this section we apply the model and inferential procedures to the UK winter data set which was introduced in Section 2. The data set has \(Y = 28\) subseries, with one subseries 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 subseries so it seems reasonable to model them as independent realizations 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 on-line supplementary materials. Unlike the parameters in the observed process \(\theta_{r, \text{obs}}\), our priors for the parameters in \(\theta_{r, \text{hid}}\) differ with the number of states \(r\) in an effort to balance the amount of information that is 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 by 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 with 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 with observed data which were not used in model fitting.

For comparison, we also consider a reduced model in which we fix $\Gamma$ to be a matrix of 0s, the coefficients of the lag 1 rainfall occurrence indicators $\beta_{r,1,k}$ to be vectors of 0s and the covariance matrices $\Sigma_{r,k}$ to be identity matrices (i.e. every $\phi_{r,k} = 0$ and every $\psi_{r,k,i} = 1$). This produces a within-state model in which the rainfall occurrences $D_{y,t}^{r,k}$ and the non-zero amounts $W_{y,t}^{r,k}$ are independent in time and space with Bernoulli $\text{Bern}\{\Phi_{i,r,k}\}$ and log-normal $\log\text{N}\{\alpha_{i,r,k}, \omega_{r,k,i}\}$ distributions respectively. This reduced model is very similar to the pioneering model of Bellone et al. (2000), differing only in the use of log-normal, 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 Markov chain Monte Carlo scheme, convergence and mixing

For each fixed number of states $r = 1, \ldots, 5$, the MCMC algorithm was used to generate 2.5 million draws from the posterior, omitting the first 500000 as burn-in and thinning the remaining output to retain every 200th iterate, to give posterior samples of size $N = 10000$. Graphical diagnostic checks including trace and auto-correlation 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 $\beta_{r,0,k} = (\beta_{r,0,k}^1, \ldots, \beta_{r,0,k}^n)^T$ within the weather state $k \in \{1, \ldots, r\}$ that is associated with the largest probabilities of rain. At certain sites $i$ in these states $k$, trace plots revealed $\beta_{r,0,k}^i$ 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 that were 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 $\beta_{r,0,k}$, this causes the likelihood to favour arbitrarily large values of $\beta_{r,0,k}$. To prevent this from happening in the posterior, the prior for the $\beta_{r,0,k}$ needed to have small variance and reasonably short tails.

Initializing 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. Consequently, there was no evidence of any lack of convergence. On the basis of auto-correlation plots, thinning to every 200th iterate appeared to remove most of the auto-correlation 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 because of 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 that was required to generate 2.5 million posterior draws was around 130 h by using sequential C code on a 2.40-GHz Dell PowerEdge R410 server with two six-core Intel Xeon E5645 central processor units and 32 Gbytes of random-access memory. 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.
Table 3. Mean CRPS (averaged across sites and time points) for models with \( r = 1, \ldots, 5 \) states

| \( r \) | Mean CRPS |
|-------|-----------|
| 1     | -2.130    |
| 2     | -2.031    |
| 3     | -2.015    |
| 4     | -2.007†   |
| 5     | -2.013    |

†Best predictive performance.

6.2. Choice of \( r \)

To use the proper scoring rule method that was 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 we have precipitation and LWT data for the six winter periods that followed the 28 winter seasons that were used in model fitting, i.e. from the years 1989–1990 to 1994–1995. These subseries, 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 \( r \) grows 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 on-line 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 auto-correlation, this relatively small value of \( r \) is not surprising.

6.3. Parameter inference assuming four weather states

We now summarize the conditional posterior distribution given \( r = \hat{r} = 4 \). Figs 3(a) and 3(b) show the posterior means and 95% equitailed 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 \( D_{y,t-1} = d \). Marginalizing over \( Z_{0,y,t} \) in the joint distribution for \((Z_{0,y,t}, Z_{1,y,t} | D_{y,t-1}, S_{y,t})\) gives

\[
Z_{1,y,t} | D_{y,t-1} = d, y_{t-1}, S_{y,t} = k, \theta_{\text{obs}} \sim \mathcal{N}(\alpha_k + \Gamma \mu_{y,t,k}, \Omega_k + \Gamma \Sigma_k \Gamma^T),
\]

from which we can easily deduce the univariate log-normal distribution for wet day rainfall at each site conditionally on \( S_{y,t} = k \) and \( D_{y,t-1} = d \in \{0, 1\}^\alpha \). For the two most frequent rainfall occurrence indicators in the observed data, \( d = (0, \ldots, 0)^T \) and \( d = (1, \ldots, 1)^T \), and, for each state \( k = 1, \ldots, 4 \), the posteriors for the log-medians in these distributions are displayed in Figs 3(c) and 3(d) for sites \( i = 1, \ldots, 12 \). In Fig. 3, the effect of \( D_{y,t-1} \) can be seen clearly. This supports the regression of \( Z_{0,y,t} \) on \( D_{y,t-1} \).

Fig. 3 also shows that weather state 2 is a clear-cut wet state, characterized by high probabilities of rain at each of the sites and large rainfall amounts on wet days. Similarly, weather state 4 is clearly 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
Fig. 3. Conditional on \( r = 4 \), posterior means with 95% equitailed credible intervals at each site in weather states 1 (\( \star \)), 2 (\( \ast \)), 3 (\( \bullet \)) and 4 (\( \blacksquare \)): probabilities of rain following (a) a dry day and (b) a wet day, and log-medians in the log-normal distributions for rainfall amounts when (c) \( d_{y,t-1} = (0, \ldots, 0) \) and (d) \( d_{y,t-1} = (1, \ldots, 1) \).

Fig. 4 displays the marginal posterior distributions for the weather state transition probabilities \( \lambda_{x,4,j,k} = \Pr(S_{y,t}=k|S_{y,t-1}=j, X_t=x, \theta_{3,\text{obs}}, r=4), x=1, \ldots, 27 \), for two representative \( j \rightarrow k \) 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 \( \lambda_{4,j,k} \) (which is the same for all \( x \)). Fig. 4(b) shows the marginal posteriors for \( \lambda_{4,2,4}, x=1, \ldots, 27 \), and is typical of the posteriors for all probabilities of transition from the wet weather state (state 2). For these 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 Fig. 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. Fig. 4(a), for example, displays the marginal posterior distributions for \( \lambda_{4,4,3}, x=1, \ldots, 27 \), 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 \( \lambda_{4,4,3} \) and a couple of the \( \lambda_{4,4,3} \) 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 \( \Lambda_x \) be a \( 4 \times 4 \) stochastic transition matrix with \( j \)th row \( \lambda_{4,j} \). The solution \( \delta_x \) to the matrix equation \( \delta_x \Lambda_x = \delta_x \) for each LWT \( x \) can be interpreted as the stationary distribution of the
Fig. 4. Conditional on $r = 4$, posterior means with 95% equitailed credible intervals for $\lambda_{x,k}^{i}, x = 1, \ldots, 27$, \((a)\) and $\bar{\lambda}_{x,k}^{i,j}$ \((b)\) when \(j = 4\) and \(k = 3\) and \(j = 2\) and \(k = 4\): also shown are the marginal prior means with 95% equitailed credible intervals \((\lambda)\) for the corresponding transition probabilities for a (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 $\delta^{i}_{x} = \delta^{i}_{x,1}, \delta^{i}_{x,2}, \delta^{i}_{x,3}, \delta^{i}_{x,4} \in \mathcal{S}_{r}$ for each value of $x$. These are displayed in Fig. 5 and reveal a complex pattern among the LWTs with the dominant feature being the variation among the pure directional and among 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) whereas 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 on-line supplementary materials.

6.4. Model checking

Chapter 6 of Gelman et al. (1995) describes several 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 $\{w_{rep}, d_{rep}\}$ that could have been observed under the model. The predictive distributions that are used in this section are conditional on $r = \hat{r} = 4$.

Let $T(w, d)$ be a test quantity, i.e. a scalar summary representing an aspect of the data that we want to capture accurately, e.g. 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 $\theta_{r}[j], j = 1, \ldots, N$, to generate draws $\{w_{rep}[j], d_{rep}[j]\}$ from the posterior predictive distribution as discussed in Section 5.2. These draws are used to compute $T(w_{rep}[j], d_{rep}[j]), j = 1, \ldots, N$. The posterior predictive distributions of various test quantities can then be compared graphically with 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 that was introduced in Section 6.2, conditioning the posterior predictive distribution on LWT data from this period. Where model checking plots correspond to single
Bayesian Modelling of Rainfall Data 563

Fig. 5. Marginal posterior means and 95% equitailed credible intervals for the solution to the matrix equation \( \delta x = 1, \ldots, 27 \) (here \( \Lambda x \) is the \( 4 \times 4 \) stochastic matrix with \( j \)th row equal to \( \lambda x, j \) and \( \delta x = (\delta x,1; \delta x,2; \delta x,3; \delta x,4) \) \in \( \mathcal{X} \)). (a) \( \delta x,4; \) (b) \( \delta x,2; \) (c) \( \delta x,3; \) (d) \( \delta x,4 \)

sites, we show the results for two sites: site 5 (Kew) and site 9 (Plymouth), chosen because Kew generally represented a good fit, whereas Plymouth generally represented the poorest fit out of the 12 sites. In all plots, the posterior predictive distributions are summarized through their mean and 95% equitailed credible interval.

Marginal properties of the rainfall data, e.g. 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 on-line 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 auto-correlation between rainfall occurrences and non-zero rainfall amounts by using log-odds ratios and Spearman’s rank correlation coefficients respectively. For all pairs of sites, Figs 6(a) and 6(b) compare the observed values of these statistics with their posterior predictive distributions. Figs 6(a) and 6(b) also show, for comparison, summaries of the posterior predictive distributions for the simple conditional independence model that was discussed in Section 6.2. For clarity we show only 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 auto-correlations between rainfall occurrences. This is surprising because, when these model checks are performed by using the model fitting data, there is very good agreement between the observed statistics and their posterior predictive distributions; see Figs 6(c) and 6(d). Comparing Figs 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 period from 1989–1990 to 1994–1995 to be larger than any of those calculated from the period from 1961–1962 to 1988–1989 data used.
Fig. 6. (a)–(d) Observed versus posterior predictive means for (a), (c) log-odds ratios and (b), (d) Spearman’s rank correlations between each pair of sites based on (a), (b) out-of-sample data and (c), (d) model fitting data, and (e)–(h) observed (■) and posterior predictive mean (□) empirical survival distributions of (e), (g) wet spells and (f), (h) dry spells at (e), (f) Kew and (g), (h) Plymouth based on out-of-sample data: (■) posterior 95% credible intervals; (□) posterior predictive means for the simple conditional independence model.
Bayesian Modelling of Rainfall Data

in model fitting. This suggests that our combination of model, data and prior could not 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 auto-correlation 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 or dry spells is simply defined as respectively the proportion of runs of consecutive wet or dry days that persist for at least \( k \) days, where \( k = 1, 2, \ldots \). Figs 6(e) and 6(f) show the plots on a log-scale for wet and dry spells at Kew whereas Figs 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 a 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. Figs 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 on-line 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 auto-correlation between rainfall amounts within wet spells is assessed by comparing the observed Spearman rank correlation coefficients between rainfall amounts at various lags (within uninterrupted wet spells) with the corresponding posterior predictive distribution. Plots revealed good agreement and can be found in the supplementary materials, along with additional commentary.

7. Discussion

We have presented a model for the spatiotemporal 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,

(a) 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

(b) it avoids the singularity at zero in the conditional density of non-zero rainfall amounts.

Further discussion of this point is given in the on-line supplementary materials.

The model also has some 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 LWTs. This provides a link to models which might be used to predict or simulate LWTs, 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 auto-correlation. 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 that are summarized above, we feel that our model is more appropriate for UK rainfall data than an NHMM using the truncated, power-transformed multivariate normal 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 $\beta_{i,r,k}$; 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 doctoral 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 LWT data are freely available from the Climate Research Unit Web site (http://www.cru.uea.ac.uk/cru/data/lwt/).

References

Ailliot, P., Thompson, C. and Thomson, P. (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 Glasbey, C. A. (2003) A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation. Appl. Statist., 52, 487–498.

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

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

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

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

Boys, R. J. and Henderson, D. A. (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., Hurn, M. and Robert, C. P. (2000) Computational and inferential difficulties with mixture posterior distributions. J. Am. Statist. Ass., 95, 957–970.

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

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

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

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

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

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

Frühwirth-Schnatter, S. (2006) Finite Mixture and Markov Switching Models. New York: Springer.
Zhang, X., Boscardin, W. J. and Belin, T. R. (2006) Sampling correlation matrices in Bayesian models with correlated latent variables. *J. Computnl Graph. Statist.*, 15, 880–896.

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

*Supporting information*

Additional 'supporting information' may be found in the on-line version of this article:

*Supplementary material for "Bayesian modelling of rainfall data using non-homogeneous hidden Markov models and latent Gaussian variables".*