Bayesian methods for time series with mixed and mixture distributions with an application to daily rainfall data

Muhammad Safwan Bin Ibrahim

Thesis submitted for the degree of
Doctor of Philosophy

School of Mathematics, Statistics & Physics
Newcastle University
Newcastle upon Tyne
United Kingdom

November 2018

I dedicate this thesis to my beloved father, Ibrahim Bin Othman and my late mother, Rokiyah Bt Ismail, who kept me going with their unconditional support and love. I am also grateful to my lovely family, supervisor, lecturers and friends for their support and encouragement.

Acknowledgements

First and foremost, I am very grateful to Allah The Almighty for giving me the strength to finish this research. I would like to thank my supervisor, Dr Malcolm Farrow for his patience, guidance, comments, encouragement and kindness which led to the completion of this thesis. I am really appreciating his support and very glad to have a good supervisor like him.

I would like to extend my most sincere gratitude to my family especially my beloved father, Ibrahim Bin Othman, for his support, encouragement, and advice. Not forgotten to all my siblings for their support and love.

Besides, I would like to thank and acknowledge the financial support I have received from the Ministry of Higher Education (Malaysia) and Islamic Science University of Malaysia (USIM).

Last but no least, I would like to express my special appreciation to all my teachers, lecturers, and friends for their encouragement, help and support towards the completion of this thesis. Thanks for all your support and understanding.

Abstract

A mixture model can be used to represent two or more sub-populations. A special case is when one of the sub-populations has a degenerate (or discrete) distribution, and another has a continuous distribution. This leads to a mixed distribution. For example, daily rainfall data contain zero and positive values. This can be represented using two processes: an amount process and an occurrence process.

This thesis is concerned with Bayesian time series models for non-independent mixture-distributed data, especially in the case of mixed distributions. Particular attention is given to the relationship between the occurrence and amount processes.

The main application in the thesis is to daily rainfall data from weather stations in Italy and the United Kingdom. Firstly, the models for univariate rainfall series are developed. These are then extended to multivariate models by developing spatiotemporal models for rainfall at several sites, giving attention to how the spatiotemporal dependencies affect both the occurrence and amount processes. For the case of the British data, the models involve dependence on the Lamb weather types. Seasonal effects are included for all models.

Posterior distributions of model parameters are computed using Markov chain Monte Carlo (MCMC) methods.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

This research is concerned with the development of Bayesian approaches to the modelling and analysis of univariate and multivariate time series data. The specific idea of this research is to look at the case where the observational distribution is a mixture distribution and, in particular, the special case of a mixture distribution where one of its sub-populations has a degenerate, or discrete, distribution and another component follows a continuous distribution. In some time series data, the observations of this special case do not only take positive values but may also include many zero values. This leads to a mixed distribution. A well-known example is daily rainfall data which contain zero (discrete) values if rain is absent and positive values (continuous) if rain occurs.

Daily rainfall data can be modelled using a two-stage approach which has two processes: the amount and occurrence processes. The amount process is a process that models the amount of rainfall which occurs during a rainy day. On the other hand, the occurrence process is a process that governs the probability of rainfall occurrences. In this thesis, particular attention is given to the relationship between the occurrence and amount processes so that no important information about rainfall amount and occurrence is lost. This has already been emphasized by Tooze *et al.* (2002) who stressed that the relationship between the amount and occurrence processes is important to improve the accuracy and adequacy of the model. Another important feature in modelling daily rainfall data is the way we deal with seasonal effects, especially when the rainfall data are expected to have a cyclical pattern. As a consequence, we will utilise a truncated Fourier series to capture the seasonal variability over the year.

The main application in this research is to daily rainfall data from weather stations in Italy and the United Kingdom. Firstly, the univariate models for rainfall data will be

developed. These are then extended to multivariate models by developing spatiotemporal models for rainfall at several sites, with a particular attention to how the spatiotemporal dependencies affect both the occurrence and amount processes. For the British daily rainfall application, the atmospheric circulation patterns will be directly incorporated into the model to provide a relationship between rainfall and climate. This is in contrast with the previous works by Heaps *et al.* (2015) and Germain (2010) who linked the atmospheric circulation patterns to latent weather states to influence the daily rainfall pattern. Hence, the model for the British daily rainfall is an alternative to the models developed by Heaps *et al.* (2015) and Germain (2010).

1.2 Research objectives

An important feature in this research is that we aim to develop and modify the work of Heaps *et al.* (2015) and Germain (2010) on spatiotemporal models for daily rainfall at a network of weather stations. The following developments will be investigated:

1. Removing the hidden weather states from the model and using the Lamb Weather Type (LWT) as the actual weather states.

2. The modelling approach will cover rainfall data over the whole year including the seasonal effects, extending the work of Heaps *et al.* (2015) and Germain (2010) who only used the winter data in their model development. When looking at daily rainfall data for the whole year, the transition probabilities in a Markov chain for the LWT will depend on the particular time of year.

3. A careful examination of the conditional distribution of the rainfall amount and how the mixed distribution of rainfall should be parameterised, how the parameters should change under different conditions and how the dependence between sites and between times should be modelled in terms of these parameterisations.

The methods are illustrated using daily rainfall data from the UK and Italy.

The main objectives of this research are:

1. To investigate Bayesian time series modelling in mixed or mixture distribution applications.

2. To develop novel approaches in modelling dependence, on both covariates and between realisations, and how this affects both the between-component distribution (*eg* the occurence process in the rainfall case) and the conditional within-component distribution (*eg* for the amount process in the rainfall case).

3. To investigate the computation of posterior distributions in univariate and multivariate models for mixture and mixed distributions.

4. To apply the developed methods to a number of practical problems and assess the strengths and weaknesses of each approach.

1.3 Outline of the thesis

The structure of this thesis is organised as follows. Chapter 2 describes the general idea of Bayesian inference and time series. We introduce some suitable models that can be used for modelling time series data. Particular attention is given to Markov chains and state space models. The state space models can be regarded as an alternative approach to autoregressive integrated moving average (ARIMA) models in the time series context. After that, we review Bayesian inference including the use of Bayes' theorem and some Markov chain Monte Carlo (MCMC) techniques which include the Metropolis-Hastings algorithm and Gibbs sampling. The data augmentation approach is also introduced in this chapter that can be used in the MCMC scheme to make the posterior sampling for some models more tractable.

Chapter 3 describes the introduction of Bayesian inference in mixture models with particular emphasis given to finite mixture models. The MCMC scheme for finite mixture models is then introduced with some discussion on the label switching problem. In addition, we also introduce the mixed distribution which is a special case of mixture models. Using the Bayesian framework, we then develop a mixture model for some ultrasound data with some discussion on the prior elicitation and posterior distribution.

In Chapter 4, we present the idea of modelling daily rainfall data within the Bayesian framework. We start with the general idea of how to develop a model for the amount and occurrence processes. We then discuss the seasonal effects and their utility in capturing the seasonal variations of daily rainfall over the year. To assess the adequacy of the model, we introduce a diagnostic checking procedure for mixed distribution time series. For the Italian daily rainfall dataset, three different probability density functions (pdfs) are considered for modelling the amount process, and these three pdfs are subsequently compared by using the posterior predictive distribution. A first-order Markov chain is also used to model the occurrence process. For a British daily rainfall dataset, we propose an alternative model to the models developed by Heaps *et al.* (2015) and Germain (2010). Instead of using hidden weather states, we incorporate atmospheric circulation patterns directly to the amount and occurrence processes. The atmospheric circulation patterns for the British daily rainfall application are represented by the Lamb weather types (LWTs). For preliminary investigation, we will solely focus on a single site before extending the

work to multiple sites in Chapter 5.

Chapter 5 introduces a spatiotemporal model for the daily rainfall data at multiple sites within the United Kingdom. This model is an extended version of the univariate model in Chapter 4. The multivariate model is more challenging to develop than the univariate model since it involves a large number of parameters and a high computational cost is required to fit the model. The exploratory examination of the dataset and the summary of investigations into the spatial and temporal characteristics of the data will also be discussed in this chapter. The adequacy of the model will then be assessed using a similar diagnostic checking method as introduced in Chapter 4.

Finally, Chapter 6 summarises our conclusions about the whole thesis and these include suggestions for future work.

Chapter 2

Introduction to Bayesian Inference and Time Series

2.1 Introduction

In this chapter, we will introduce some models that are suitable to use for modelling time series data. We will introduce Markov chains in Section 2.2.2 and state space models in Section 2.2.3. The Markov chain has two different types of states: discrete (Section 2.2.2.2) and continuous (Section 2.2.2.3). The state space model presents an alternative approach to autoregressive integrated moving average (ARIMA) models in building time series models. If the model is linear and the observations follow a Gaussian distribution, then we can use dynamic linear models (DLMs) to analyse the data as in Section 2.2.3.2. Alternatively, we can use dynamic generalised linear models (DGLMs) to model both non-normal and nonlinear time series (see Section 2.2.3.3). Hidden Markov models (HMMs) are another type of state space model which contain two different layers of a system: an observed process and an unobserved process. The behaviour of the observed process can be reproduced by modelling the unobserved process as a Markov chain (see Section 2.2.3.4).

We will also describe briefly Bayesian inference including Bayes' theorem (Section 2.3.2) and prior distributions (Section 2.3.3). After that, we will discuss the use of sampling techniques, specifically Markov chain Monte Carlo (MCMC) to obtain posterior samples for parameters in Section 2.3.4. We will also introduce the data augmentation technique in Section 2.3.5 which can be used in the MCMC scheme when the direct computation of the posterior samples for some models is difficult to handle. In Section 2.3.6, we review the application of the Bayesian approach in time series models.

2.2 Time Series Models

2.2.1 Introduction

A time series is a collection of observations that are generated sequentially in time. It can also be considered as a realisation of a stochastic process. The order of a time series is crucial because the observations are ordered with respect to time. The applications of time series can be found widely in various fields such as monthly rainfall, the number of accidents in a week, daily stock market prices, and annual series of the number of cancer patients in a hospital. The observations of a time series are given as

$$y_1, y_2, ..., y_T$$

where $t \in \{1, 2, \cdots, T\}$. The observation y_t is regarded as a realisation of a random variable Y_t and can be either continuous or discrete. For the time index t, it can also have continuous and discrete values. However, in this thesis we only consider a discrete time index t, where the observations are drawn at fixed intervals.

There are some reasons why we look at time series analysis. Firstly, we want to analyse the data and describe what has happened in the past in terms of the components of interest such as trends, seasonal effects, patterns and fluctuations. From here, the information will be gathered and then we will make some inferences from the data. Time series can also be used to assess the effects of other variables and interventions based on the behaviour of data, for example the effect of weather on gas consumption. Of course, time series can also be used to make some predictions about the future based on the observed data.

The modelling and analysis of time series are mainly focused on two main approaches which are autoregressive integrated moving average (ARIMA) models (Box & Jenkins, 1976) and state space models, e.g. dynamic linear models (DLMs) (West & Harrison, 1997). We will use some of these approaches in our analysis. For instance, we will use the first order autoregressive model AR(1) for the random effect in our daily rainfall model in Chapter 4. In the next section, we will review some basic theory about state space models. For an introductory survey of approaches to time series analysis, see Chatfield (2003).

2.2.2 Markov chains

2.2.2.1 Introduction

In this section, we review some basic ideas about Markov chains. A Markov chain is a sequence of random variables, $X^{(0)}, X^{(1)}, X^{(2)}, \cdots$ where the future states are conditionally

independent of the past states given the present state. This process is also known as a "memoryless" process because it has no information about where it has been in the past. Beyond the present state, the process must satisfy the Markov property for all $t \in \mathbb{N}$ and the conditional probabilities can be written as

$$\begin{aligned} &\Pr(X_{t+1} = x_{t+1} | X_t = x_t, X_{t-1} = x_{t-1}, \cdots, X_0 = x_0) \\ &= \Pr(X_{t+1} = x_{t+1} | X_t = x_t). \end{aligned} \tag{2.1}$$

Figure 2.1 shows the directed acyclic graph (DAG) for the dependence structure of a Markov chain where $X_t \in S$ denotes the state of the process at time t and S is a state space.

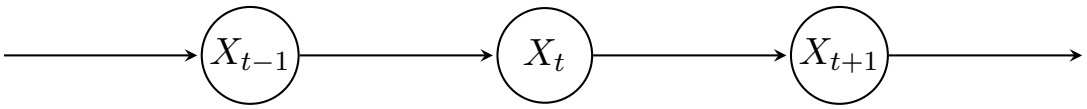


Figure 2.1: A directed acyclic graph (DAG) showing the dependence structure of a Markov chain

2.2.2.2 Discrete state

When dealing with discrete state spaces, the notation for the states of a Markov chain in (2.1) can be substituted by a single letter such as $i, j, k \in S$:

$$\Pr(X_{t+1} = j | X_t = i) = p_{ij}.$$

The conditional probabilities for a Markov chain are known as transition probabilities from state i to state j. If we have a finite state system, that is a system with J states where $J < \infty$, then the transition matrix can be written as

$$P(t) = \begin{pmatrix} p_{1,1}(t) & \cdots & p_{1,J}(t) \\ \vdots & \ddots & \vdots \\ p_{J,1}(t) & \cdots & p_{J,J}(t) \end{pmatrix}.$$

The elements of the matrix must satisfy these two properties:

1. $0 \leq p_{ij}(t) \leq 1$
2. $\sum_{j=1}^{J} p_{ij}(t) = 1$ where $i, j \in J$.

The matrix $P(t)$ is called a Markov matrix or stochastic matrix at time t if it fulfills these properties. If these probabilities depend on time t, the process is called a nonhomogeneous

Markov chain where, for example,

$$p_{ij}(1) = P[X_1 = j|X_0 = i] \neq P[X_2 = j|X_1 = i] = p_{ij}(2).$$

On the other hand, the process is described as a homogeneous Markov chain if these probabilities do not depend on time t where

$$P[X_1 = j|X_0 = i] = P[X_2 = j|X_1 = i] = \cdots = P[X_{t+1} = j|X_t = i] = p_{ij}.$$

with constant transition matrix

$$P = \begin{pmatrix} p_{1,1} & \cdots & p_{1,J} \\ \vdots & \ddots & \vdots \\ p_{J,1} & \cdots & p_{J,J} \end{pmatrix}.$$

Subject to certain conditions, a homogeneous Markov chain can be described further by finding the unconditional probabilities of the Markov chain. Let $\boldsymbol{\pi}^{(t)} = \left(\pi_1^{(t)}, \pi_2^{(t)}, \cdots, \pi_J^{(t)}\right)$ represents the row vector of probabilities for states at time t where

$$\pi_j^{(t)} = \Pr(X_t = j); \qquad j \in J.$$

Using the theorem of total probability, the probability distribution at time $t+1$ is given as

$$\pi_1^{(t+1)} = p_{11}\pi_1^{(t)} + p_{21}\pi_2^{(t)} + \cdots + p_{J1}\pi_J^{(t)},$$

and similarly for $\pi_2^{(t+1)}, \cdots, \pi_J^{(t+1)}$. These probability distributions can be converted into matrix form as

$$\left(\pi_1^{(t+1)}, \pi_2^{(t+1)}, \cdots, \pi_J^{(t+1)}\right) = \left(\pi_1^{(t)}, \pi_2^{(t)}, \cdots, \pi_J^{(t)}\right) \times \begin{pmatrix} p_{1,1} & \cdots & p_{1,J} \\ \vdots & \ddots & \vdots \\ p_{J,1} & \cdots & p_{J,J} \end{pmatrix}$$

which is equivalent to

$$\begin{aligned} \boldsymbol{\pi}^{(t+1)} &= \boldsymbol{\pi}^{(t)} P \\ &= \boldsymbol{\pi}^{(0)} P^{t+1} \end{aligned}$$

where $\boldsymbol{\pi}^{(0)}$ is the initial distribution and P represent the transition probability matrix. This is known as the Chapman-Kolmogorov equation. Hence, a homogeneous Markov chain with transition matrix P is said to have stationary distribution $\boldsymbol{\pi}$ if there exists $\boldsymbol{\pi}$

such that

$$\boldsymbol{\pi} = \boldsymbol{\pi} P$$

with $\boldsymbol{\pi} \times \mathbf{1}^{'} = 1$. It is called a stationary distribution because the distribution $\boldsymbol{\pi}^{(t)}$ stays the same as the time moves at any number k-steps:

$$\boldsymbol{\pi}^{(t)} = \boldsymbol{\pi}^{(t+k)} = \boldsymbol{\pi} \qquad \forall k > 0.$$

Under certain conditions (see Stewart (2009)),

$$\boldsymbol{\pi}^{(0)} P^t \longrightarrow \boldsymbol{\pi} \quad \text{as} \quad t \longrightarrow \infty. \tag{2.2}$$

However, in some cases, for example, $P = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, (2.2) does not hold. The unconditional probability of a homogeneous Markov chain can be found by solving

$$\boldsymbol{\pi}(I - P) = 0$$

where I is the $J \times J$ identity matrix. A comprehensive description of Markov chains can be found in Zucchini (2009) and Stewart (2009).

In a hidden Markov model (see Section 2.2.3.4), the application of a Markov chain is important to model the values of the hidden process. The stationary distribution of a Markov chain is also used in Bayesian inference to draw samples from the posterior distribution. Markov chains are widely used in applications such as analysing rainfall occurrence (Gabriel & Neumann, 1962), queues of customers (Kendall, 1953) and exchange rates of currencies (Masson, 2001). In the rainfall case, the homogeneous Markov chain is a favoured method to find the probability of rainfall occurrence. However, the non-homogeneous Markov chain is also employed if we believe that the probability of rainfall at time t may depend, for example, on seasonal factors. In Chapter 3, we will use both homogeneous and nonhomogeneous Markov chains to analyse the rainfall occurrence and the Lamb weather types (LWTs).

2.2.2.3 Continuous state

Now, let the state space S be continuous ($S \subset \mathbb{R}$) where the time is still discrete. For any $A \subset S$ and $x \in S$, a homogeneous chain for continuous state space is defined in a similar way to the discrete case as follows:

$$P(x, A) = \Pr(X_{t+1} \in A | X_t = x).$$

In the discrete case, this is equivalent to

$$P(x, \hat{x}) = \Pr(X_{t+1} = \hat{x} | X_t = x)$$

for $x, \hat{x} \in S$. For the continuous case, we can not use this form because $P(x, \{\hat{x}\}) = 0$. Instead, the notation $P(x, \hat{x})$ can be defined as:

$$P(x, \hat{x}) = \Pr(X_{t+1} \leq \hat{x} | X_t = x) = \Pr(X_1 \leq \hat{x} | X_0 = x)$$

for $x, \hat{x} \in S$. This form represents the transition matrix for the continuous case and the conditional density is given by

$$p(x, \hat{x}) = \frac{\partial}{\partial \hat{x}} P(x, \hat{x})$$

which defines the density form of the transition kernel of the chain. Assume that the conditional transition probability moves to k-steps as follows:

$$P^k(x, \hat{x}) = \Pr(X_{t+k} \leq \hat{x} | X_t = x),$$

where the transition kernel is given by

$$p^k(x, \hat{x}) = \frac{\partial}{\partial \hat{x}} P^k(x, \hat{x})$$

for $x, \hat{x} \in S$. This can be further extended by using the theorem of total probability where the conditional distribution at any step $t + k$ has the form

$$P^{t+k}(x, \hat{x}) = \int_S P^k(\ddot{x}, \hat{x}) p^t(x, \ddot{x}) d\ddot{x}$$

which is a continuous version of the Chapman-Kolmogorov equations. At $k = 1$, we have

$$P^{t+1}(x, \hat{x}) = \int_S P(\ddot{x}, \hat{x}) p^t(x, \ddot{x}) d\ddot{x}.$$

Then, the marginal distribution at any time $t + 1$ can be obtained as follows:

$$\pi^{(t+1)}(\hat{x}) = \int_S p(x, \hat{x}) \pi^{(t)}(x) dx.$$

A distribution $\pi(\hat{x})$ on continuous state spaces is said to be a stationary distribution if it satisfies

$$\pi(\hat{x}) = \int_S p(x, \hat{x}) \pi(x) dx.$$

A full comprehensive discussion about the continuous state space Markov chains can be found in Gamerman & Lopes (2006).

2.2.3 State space models

2.2.3.1 Introduction

In this section, we review some basic theory of state space models in time series analysis. State space models are an alternative to ARIMA models for constructing time series models and forecasting. It is a flexible model where computations for monitoring and forecasting can be done recursively. Originally, it was developed in the engineering field for updating information about the system continuously from the current position (Kalman & Bucy, 1961). This is one of the reasons why the state space model literature can be found more in engineering applications. However, it is still applicable to use in many other fields such as economics and computer sciences. For example, state space models can be used to forecast the unemployment rate as an "output" of the underlying state of the economy. This technique is also very popular among Bayesian statisticians and some of the notations used in state space models are expressed in Bayesian terms. For instance, the expectation of future values can be considered as our belief about the future and can be updated after we observe the data.

The basic idea of a state space model is to have a probabilistic dependence between the observation, Y_t and an unobserved state variable X_t:

$$Y_t = h(X_t) + \eta_t$$

where $h(.)$ is some function. Often $h(.)$ is a linear function that contains a linear combination of several components such as trend, seasonal or regressive components. The unobserved state variable X_t is also known as the signal and the random variable η_t is called the noise. In most cases, η_t is assumed to be independent from η_s when $s \neq t$. Therefore Y_t is conditionally independent of Y_s given X_t where $s \neq t$. The observations Y_t can be either univariate or multivariate. Typically, the sequence $\{X_t\}$ is a Markov chain. In this case, in terms of the joint probability density, the state space model can be written as

$$\begin{aligned}\pi(x_{1:T}, y_{1:T}) =& \pi(x_1)\pi(x_2|x_1)\pi(x_3|x_2)\cdots \times \pi(y_1|x_1)\pi(y_2|x_2)\pi(y_3|x_3)\cdots \\ =& \pi(x_1)\pi(y_1|x_1)\prod_{t=2}^{T}[\pi(x_t|x_{t-1})\pi(y_t|x_t)]\end{aligned}$$

where the conditional distributions often stay the same over time. By introducing the

initial distribution $\pi(x_0)$, the state space model can be simplified to

$$\pi(x_{0:T}, y_{1:T}) = \pi(x_0) \prod_{t=1}^{T} \left[\pi(x_t|x_{t-1})\pi(y_t|x_t)\right]. \tag{2.3}$$

The state space model can help to solve a broad range of time series problems such as inference for unknown parameters, smoothing and prediction. For details, see Kitagawa (1998), Durbin & Koopman (2000), and Christensen *et al.* (2012). It is considered to be one of the best approaches in time series at handling missing values. A particular example of a state space model in time series is a hidden Markov model. We will discuss hidden Markov models in Section 2.2.3.4. In addition, the parameterisation of ARIMA models can be converted into state space form.

2.2.3.2 Dynamic linear models

Dynamic linear models (DLM) form an important class of state space models to represent linear Gaussian processes. The observations are assumed to follow normal distributions where the mean μ is a linear combination of the elements of a system vector. The system vector either consists of an intercept term only or includes some other terms representing trend or seasonal components. The DLM can also be defined as a system of equations that has two stages: an observation equation and system equation. The equations are given as follows:

$$\text{Observational equation}: \quad Y_t = F_t' X_t + \upsilon_t; \qquad \upsilon_t \sim N_r(0, V_t) \tag{2.4a}$$

$$\text{System equation}: \quad X_t = G_t X_{t-1} + \omega_t; \qquad \omega_t \sim N_n(0, W_t) \tag{2.4b}$$

with X_t is defined as the state vector where F_t' and G_t are known matrices (of order $r \times n$ and $n \times n$ respectively) and υ_t and ω_t are vectors of normal errors with mean zero and known variance matrices V_t and W_t. In this case, $\mu_t = F_t' X_t$ represents the mean response of the model given X_{t-1} and the error sequences υ_t and ω_t are independent. So, υ_t and ω_s are independent, υ_t and υ_s are independent, and ω_t and ω_s are independent for all t and s with $t \neq s$. The observation Y_t could be a vector or a scalar.

In general, all of the information can be characterised by a quadruple, (F_t, G_t, V_t, W_t) for each time t. However, in some circumstances, each of the elements is constant for all time t as (F, G, V, W). The DLM structure can be simplified:

$$(Y_t|X_t) \sim N(F_t' X_t, V_t); \quad (X_t|X_{t-1}) \sim N(G_t X_{t-1}, W_t). \tag{2.5}$$

Inference about the current state and forecasting of a DLM can be obtained recursively

by using the Kalman filter. Detailed explanations and examples of DLM can be found in West & Harrison (1997) and Petris *et al.* (2009).

2.2.3.3 Dynamic generalised linear models

Dynamic generalised linear models (DGLM) are an extension of the DLM for application in non-normal and nonlinear time series. Originally, they were proposed by West *et al.* (1985) based on the framework of generalised linear models (Nelder & Wedderburn, 1972; McCullagh & Nelder, 1989) to provide an alternative dynamic approach when the data distribution is non-normal. The non-normal distribution could be from another member of exponential family distributions such as gamma, binomial or Poisson distributions that can be characterised by

$$f(y_t|\gamma_t, \phi) \propto \exp\left\{\frac{[y_t\gamma_t - b(\gamma_t)]}{\phi}\right\}$$

where γ_t is the natural parameter and ϕ is known as the scale parameter. Let $E(Y_t|\gamma_t, \phi) = \mu_t$, then the DGLM structure is defined by the following components:

$$\text{Observational model}: \quad f(Y_t|\gamma_t, \phi) \quad \text{and} \quad g(\mu_t) = \eta_t = F_t' X_t \tag{2.6a}$$

$$\text{Evolution equation}: \quad X_t = G_t X_{t-1} + \omega_t \quad \text{with} \quad \omega_t \sim [0, W_t]. \tag{2.6b}$$

In this structure, the mean μ_t of the observation model is determined by relating it through a link function $g(.)$ to the linear predictor η_t. Typically, the choice of the link function depends on the error distribution. For example, if the observations Y_t follow a Poisson distribution, then we can use a log-link function to provide the relationship between the linear predictor and the mean as:

$$\log(\mu) = \eta_t = F_t' X_t.$$

The underlying state equation for a DGLM is precisely the same as the standard DLM in (2.4b).

The analysis of a DGLM is more challenging than that of a DLM since we have lost the normality and linear framework, and there is no closed form solution for updating inferences. However, there are some possible approximations that can be used to solve this problem. For instance, West *et al.* (1985) suggested using a conjugate prior and posterior distribution for the exponential family parameters which can help to simplify the computational problem. The authors used a linear Bayes model for the system vector with only moments specified. We might now give the system vector a Gaussian model and lose the conjugacy but use MCMC to fit the model and possible methods such as a

particle filter for filtering. The full comprehensive development of DGLMs can found in West *et al.* (1985) and West & Harrison (1997).

2.2.3.4 Hidden Markov models

A hidden Markov model (HMM) presents a framework to develop a complex model for time series data as one of the ingenious statistical modelling approaches. It can be used in many areas such as communications, engineering, bioinformatics, finance, medicine, meteorology and speech recognition. A HMM is a bivariate discrete-time stochastic process, $\{X_t, \boldsymbol{Y}_t\}_{t\geq 1}$ that is split into two components: firstly, the observed process where the sequence of observations can be discrete or continuous and secondly, the unobserved (hidden) process which is usually assumed to be discrete with a finite state space that can be denoted by integers $\{1, \cdots, J\}$. The unobserved variable of the hidden process, X_t is called the state and can also be regarded as missing data or a latent variable. The inference on the parameters of the model can only be achieved through $\boldsymbol{Y}_t$ where the conditional distribution of $\boldsymbol{Y}_t$ is dependent on the state X_t. This means that every observation at time t was produced by some process whose state X_t is hidden from the observer. However, the output of $\boldsymbol{Y}_t$ given X_t is visible to the observer with each state having a probability distribution and the sequence of states X_t comprising a Markov chain. Then, the states $\{X_t\}_{t\geq 1}$ are said to fulfill the Markov property if the future state, X_{t+1} is conditionally independent of $X_{t-1}, X_{t-2}, \cdots$ given the current state, X_t. The assumptions of a "standard" hidden Markov model are given as follows:

1. The assumption for the hidden process can be written as

$$\Pr(X_t = j|X_{1:t-1}, \boldsymbol{\theta}) = \Pr(X_t = j|X_{t-1} = i, P) = p_{ij} \qquad i, j \in \{1, \cdots, J\}$$

   where $\boldsymbol{\theta}$ is the collection of parameters and the states $\{X_t\}_{t\geq 1}$ are a first order homogeneous J-state Markov chain with transition matrix P and the summation of each row $\boldsymbol{p}_i = (p_{i1}, \cdots, p_{ij})$ is equal to 1.

2. The assumption for the observation process can be summarised as

$$(\boldsymbol{Y}_t|\boldsymbol{Y}_{1:t-1}, X_{1:T}, \boldsymbol{\theta}) \equiv (\boldsymbol{Y}_t|X_t = j, \boldsymbol{\theta}) \sim \mathcal{F}(\boldsymbol{\theta}_j)$$

   where $\mathcal{F}(.)$ is a parametric distribution family that has density $p(\boldsymbol{y}_t|X_t = j, \boldsymbol{\theta}) \equiv p(\boldsymbol{y}_t|\boldsymbol{\theta}_j)$. From this assumption, the outputs $\{\boldsymbol{Y}_t\}_{t\geq 1}$ are considered to be conditionally independent given the hidden states $\{X_t\}_{t\geq 1}$.

The relationship between variables in a HMM can be understood by illustrating the dependence structure using a graphical model as in Figure 2.2. In this graph, each node

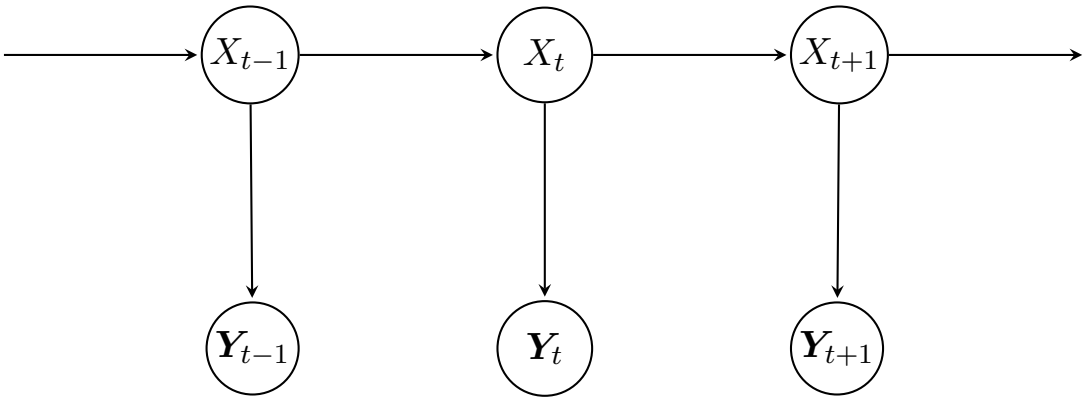


Figure 2.2: A DAG showing the dependence structure of a "standard" hidden Markov model

represents a random variable. The arrows correspond to the structure of the joint probability distribution. A hidden Markov model is said to be homogeneous when the transition probabilities of the Markov chain for the underlying process are constant over time and the conditional distributions of $Y_t|X_t$ remain constant over time.

There are a few extensions for HMM which have been reviewed by Germain (2010) and Cappé *et al.* (2005). For example, the order of the Markov chain for the hidden process can be more than one. Let the hidden state sequence $\{X_t\}_{t\geq 1}$ be a d-th order Markov process. Then, we can say that the conditional distribution of the future state, X_{t+1} depends on the previous d values, $X_{t-d}, X_{t-d+1}, \cdots, X_{t-1}$. Conceptually, models with $d > 1$ are as straightforward as in a standard hidden Markov model but we will not consider this case in our study. The hidden Markov model can be further extended by allowing the model to have non-homogeneous transition probabilities for the hidden process or non-homogeneous observation distributions for the observation process. Then, we can say that the distribution of X_{t+1} given X_t or $\mathbf{Y}_t$ given X_t is changed for every time-step t.

2.2.3.5 Markov switching models

A notable generalisation for hidden Markov models is Markov switching models. For this generalisation, the observed process allows the conditional distribution of $\mathbf{Y}_{t+1}$ given the history of past variables to depend not only on X_{t+1} but also on the previous q observations, $\mathbf{Y}_{t-q}, \mathbf{Y}_{t-q+1}, \cdots, \mathbf{Y}_{t-1}$. The dependence structure of this model is illustrated as in Figure 2.3. The statistical analysis for a Markov switching model is more complicated than for a HMM since $\mathbf{Y}_t$ does not simply depend on X_t only but also on $\mathbf{Y}_{t-q}, \cdots, \mathbf{Y}_{t-1}$.

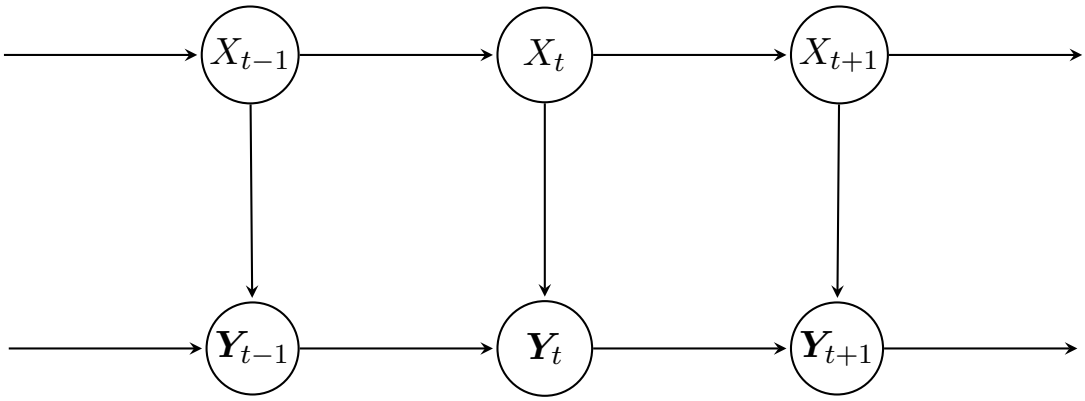


Figure 2.3: A DAG showing the dependence structure of Markov switching model with $q = 1$

2.3 Bayesian Inference

2.3.1 Introduction

In the Bayesian framework, probability distributions are used to describe uncertainty about the values of unknown quantities, whether these are parameters, missing data, latent variables or future observations. This is in contrast with the frequentist approach where, for example, the value of a parameter is regarded as fixed but unknown and no distinction is made between possible values which are more or less likely.

In Bayesian inference, the distribution given to the value of an unknown quantity before data are observed is called a prior distribution. When data are observed, the likelihood function from these data is combined with the prior distribution to give a posterior distribution which describes the new state of uncertainty about the value of the unknown quantity. The information from the posterior distribution can be summarised to make statistical inferences.

2.3.2 Bayes' theorem

Bayes' theorem or Bayes' rule plays a central role in Bayesian inference where it combines prior belief with observed data to obtain the conditional probability of a given hypothesis. Let $A_1, A_2, ..., A_n$ be a set of mutually exclusive events that form the sample space S. Given an additional event B from the same sample space with $\Pr(B) > 0$, Bayes' theorem can be written as:

$$\Pr(A_j|B) = \frac{\Pr(A_j \cap B)}{\Pr(A_1 \cap B) + \Pr(A_2 \cap B) + ... + \Pr(A_n \cap B)} = \frac{\Pr(A_j \cap B)}{\sum_{i=1}^{n} \Pr(A_i \cap B)}$$

Using the fact that $\Pr(A_j \cap B) = \Pr(A_j)\Pr(B|A_j)$, Bayes' theorem can then be rewritten as:

$$\Pr(A_j|B) = \frac{\Pr(A_j)\Pr(B|A_j)}{\sum_{i=1}^{n}\Pr(A_i)\Pr(B|A_i)}$$

Suppose we have a set of observations $\mathbf{y}=(y_1, y_2, ..., y_n)$ which depend on a set of k unknown quantities, $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_k)$ and we are interested in making inference about $\boldsymbol{\theta}$. The likelihood of parameter $\boldsymbol{\theta}$ can be expressed as

$$L(\boldsymbol{\theta}|\boldsymbol{y}) = f(\boldsymbol{y}|\boldsymbol{\theta}).$$

If $y_1, \cdots, y_n$ are conditionally independent given $\boldsymbol{\theta}$, then we can write

$$L(\boldsymbol{\theta}|\boldsymbol{y}) = \prod_{i=1}^{n} f_i(y_i|\boldsymbol{\theta}).$$

The prior beliefs for $\boldsymbol{\theta}$ can be represented as a probability density or probability mass function, $\pi(\boldsymbol{\theta})$. Hence, we can summarise our beliefs about $\boldsymbol{\theta}$ using information from the prior and this is subsequently updated by the likelihood, resulting in a posterior distribution, $\pi(\boldsymbol{\theta}|\boldsymbol{y})$. Using Bayes' theorem, the posterior density $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ can be expressed as:

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{\pi(\boldsymbol{\theta})f(\boldsymbol{y}|\boldsymbol{\theta})}{f(\boldsymbol{y})} \tag{2.7}$$

where

$$f(\boldsymbol{y}) = \begin{cases} \int_{\Theta} \pi(\boldsymbol{\theta})f(\boldsymbol{y}|\boldsymbol{\theta})d\boldsymbol{\theta} & \text{if } \boldsymbol{\theta} \text{ is continuous} \\ \\ \sum_{\Theta} \pi(\boldsymbol{\theta})f(\boldsymbol{y}|\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \text{ is discrete} \end{cases}$$

and Θ is the set of possible values of $\boldsymbol{\theta}$. From (2.7), $f(\boldsymbol{y})$ is a normalising constant or marginal likelihood of the data which ensures that the posterior density always integrates to one (if continuous) or sums to one (if discrete). Since this is not a function of $\boldsymbol{\theta}$, then the posterior distribution can be simplified as

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{\theta}) \times f(\boldsymbol{y}|\boldsymbol{\theta}) \tag{2.8}$$

that is

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood}.$$

In many cases, the normalising constant $p(\boldsymbol{y})$ is not available in analytical closed form. Numerical integration or analytic approximation is hence required to solve the difficulty of not having a closed form especially when we have an integral for continuous unknowns. There are several techniques that can be used to solve this problem but the most popular one is known as Markov Chain Monte Carlo (MCMC) which generates samples of the

values of parameters from the posterior distribution in complex models without computing the integral for the normalising constant $f(\boldsymbol{y})$ in Bayes' rule. As this technique is easier and more practical to use, MCMC techniques will be used for analysis in this study.

2.3.3 Prior distributions

In Bayesian analysis, the prior specification for unknown quantities is required to obtain the posterior distribution by combining it with the likelihood of the data. Hence, the choice of prior density for unknown quantities has to be taken into account in Bayesian analysis. The prior distribution should manifest the information about unknown quantities in the form of a probability distribution before looking at the data. Often, prior information is gathered from the opinion or subjective belief from an "expert" in the field in which we are interested. For example, we can obtain prior information about the rainfall or wind speed from a meteorologist who is an expert in the meteorological area. Therefore, elicitation of opinion or information from the expert can be regarded as an essential step in the Bayesian analysis. An expert in the application field is not necessarily an expert in statistics and probability so care is required in elicitation.

The prior elicitation is a process which transforms the information about one or more uncertain quantities gathered from the expert into a probability distribution or a moment. The elicitation process can be illustrated as a facilitator that helps the expert to express his knowledge in probabilistic form (Garthwaite *et al.*, 2005). To help the expert, Kadane & Wolfson (1998) suggested that the elicitation questions should always ask about observable quantities.

It is crucial to brief the expert about the type of questions before starting the procedure. This will help them to be familiar with the process. In most cases, the elicitation procedure will involve with the first and second moments (i.e. means, variances and covariance). People tend to do well in assessing the mean for a symmetric distribution but not for a skewed distribution. For a skewed distribution, assessing the mode or median is a better choice to be more accurate. Elicitation of the variances of unknown quantities is also important to determine a prior distribution. However, most of the time, eliciting the variance is quite difficult for people to interpret accurately. Hence, Garthwaite *et al.* (2005) proposed to use other quantities such as credible intervals to elicit the spread of a distribution. A considerable and comprehensive discussion about the prior elicitation can be found in O'Hagan (1998), Kadane & Wolfson (1998), Garthwaite *et al.* (2005), and O'Hagan *et al.* (2006). An alternative to using expert opinion might be using some kind of "representative" prior or a vague prior.

It is often convenient to use a conjugate prior. Suppose that we have a prior distribu-

tion for θ that has probability density function $\pi(\theta)$. When we have observed data, the likelihood function is given as $L(\theta|y)$. By combining the prior distribution and likelihood, we can obtain the posterior distribution as

$$\pi(\theta|y) \propto \pi(\theta)L(\theta|y).$$

If $\pi(\theta)$ and $\pi(\theta|y)$ belong to the same family of distributions, then $\pi(\theta)$ is said to be a conjugate prior distribution. The form of the conjugate prior relies closely on the form of the likelihood. Hence we will have a different conjugate distribution if we have a different data distribution. Prior beliefs might not always be well represented by a conjugate distribution. However, sometimes a conjugate prior will represent prior beliefs sufficiently closely and this will help to make the calculations easier.

2.3.4 Markov Chain Monte Carlo (MCMC) methods

2.3.4.1 Introduction

In Bayesian analysis, computing the posterior density is relatively straightforward when we use a conjugate prior. However, it may become difficult if the distributions are complex and analytically intractable especially in higher dimensions. Markov chain Monte Carlo (MCMC) is a technique to tackle this problem by drawing samples from the complex posterior distributions without computing the integral in Bayes' rule. The basic idea of MCMC techniques is to construct a Markov chain which has the posterior distribution as its stationary distribution. Then, by generating a realisation of the Markov chain for a sufficiently long number of steps and provided that the chain has converged, the samples will be generated from the posterior distribution. The introduction of MCMC gives an alternative for us to make any inference of interest by generating the whole distribution numerically. For comprehensive theory, developments and applications of MCMC, the reader can refer to Besag *et al.* (1995), Gamerman (1997), Brooks & Gelman (1998), Gamerman & Lopes (2006) and Brooks *et al.* (2011). In the following sections, we will briefly define two fundamental MCMC techniques: the Gibbs sampler and Metropolis-Hastings algorithms, to generate samples from posterior distributions.

2.3.4.2 Metropolis-Hastings Algorithm

The Metropolis algorithm was developed by Metropolis *et al.* (1953) before it was generalised by Hastings (1970) to be the Metropolis-Hastings algorithm. Suppose we are interested in sampling realisations from the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ which has a non-standard form. By introducing a proposal distribution with density $q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$, it makes

the sampling of the posterior density $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ simple or feasible. The Metropolis-Hastings algorithm is given by Algorithm 1.

Algorithm 1: Metropolis-Hastings Algorithm

1. Set the initial state of the chain to $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, ..., \theta_k^{(0)})^T$ and set iteration counter to $j = 1$.
2. Generate a proposal $\boldsymbol{\theta}^*$ from the proposal distribution $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)})$.
3. Evaluate the acceptance probability $\alpha(\boldsymbol{\theta}^{(j-1)}, \boldsymbol{\theta}^*)$ of the proposed move as follows:

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \min\left\{1, \frac{\pi(\boldsymbol{\theta}^*|\boldsymbol{y})q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}|\boldsymbol{y})q(\boldsymbol{\theta}^*|\boldsymbol{\theta})}\right\}$$

.

4. Then set $\boldsymbol{\theta}^{(j)} = \boldsymbol{\theta}^*$ if we accept the proposal with probability $\alpha(\boldsymbol{\theta}^{(j-1)}|\boldsymbol{\theta}^*)$. Otherwise, we reject the proposed value and set $\boldsymbol{\theta}^{(j)} = \boldsymbol{\theta}^{(j-1)}$.
5. Change the counter to $j + 1$ and return to step 2.

In this algorithm, a new value is proposed at every iteration from the proposal distribution. Then, the proposal is accepted or rejected according to the acceptance probability and, if it is accepted, the chain moves or, if it is rejected, it stays at the same position. By generating the values for $\theta^{(1)}, \theta^{(2)}, \ldots$, the above algorithm forms a Markov chain with $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ as the stationary distribution.

In a Metropolis-Hastings algorithm, it is important to choose a suitable proposal distribution so that realisations from the parameters of interest can be generated. There are several types of proposal distributions that are commonly used in Metropolis-Hastings algorithms.

Symmetric chain

A special case of a proposal distribution for the Metropolis-Hastings algorithm is a symmetric proposal distribution with $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}) = q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$, $\forall \boldsymbol{\theta}, \boldsymbol{\theta}^*$. Then, the acceptance probability can be simplified as follows:

$$\alpha(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = \min\left\{1, \frac{\pi(\boldsymbol{\theta}^*|\boldsymbol{y})}{\pi(\boldsymbol{\theta}|\boldsymbol{y})}\right\},$$

and so the acceptance probability does not involve this proposal density at all.

Random walk proposals

It is possible to introduce the proposed value $\boldsymbol{\theta}^*$ from a random walk as follows:

$$\boldsymbol{\theta}^* = \boldsymbol{\theta}^{(j-1)} + \boldsymbol{w}_j$$

where $\boldsymbol{w}_j$ are independent and identically distributed random $p \times 1$ vectors which have density $f(.)$ and are easily sampled. Usually, the distribution of $\boldsymbol{w}_j$ has a mean of $\mathbf{0}$ and is symmetric about its mean. Then the proposal distribution can be set as $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}) = f(|\boldsymbol{\theta}^* - \boldsymbol{\theta}|)$. We can choose any suitable distribution for $f(.)$, typically, a uniform or normal distribution. An important consideration for a random walk proposal is to determine a suitable variance for the chosen distribution since it will affect the acceptance probability and the overall proportion of accepted moves. The chain is said to be too "cold" if the variance for the innovation $\boldsymbol{w}_j$ is too low, making the proposed values mostly accepted. Conversely, the chain is said to be "hot" if the variance for the innovation is too large, so that only a few proposed values will be accepted. Acceptance rates between 20% and 50% are considered acceptable (Besag *et al.*, 1995; Bennett *et al.*, 1996; Gamerman & Lopes, 2006) although Gelman *et al.* (1996) found out that the acceptance rate should be optimally around 25%. It is suggested that to get an acceptance rate within this range, the variance of the innovation should be "tuned" first.

Independence chain

Suppose that the proposal distribution is formed independently of the position of the chain and so $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}) = f(\boldsymbol{\theta}^*)$ for some density $f(.)$. As a result, the acceptance probability is:

$$\alpha(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = \min\left\{1, \frac{\pi(\boldsymbol{\theta}^*|\boldsymbol{y})}{\pi(\boldsymbol{\theta}|\boldsymbol{y})} \Big/ \frac{f(\boldsymbol{\theta}^*)}{f(\boldsymbol{\theta})}\right\}.$$

If $f(.)$ is set to be as close as possible to $\pi(.|\boldsymbol{y})$, this will optimize the acceptance probability of the chain.

Componentwise transitions

In some circumstances, constructing a suitable proposal density $q(.|\boldsymbol{\theta})$ could be difficult. Suppose we have parameters $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)$. The full conditional distribution (FCD) of θ_i is given by

$$\pi(\theta_i|\theta_1, \cdots, \theta_{i-1}, \theta_{i+1}, \cdots, \theta_k, \boldsymbol{y}).$$

This FCD is not only dependent on the data $\boldsymbol{y}$, but also on the current values of other parameters. For many problems of interest, the FCD for a subset of $\boldsymbol{\theta}$ may be

suitable for sampling. Let the FCD for the i^{th} component of $\boldsymbol{\theta}$ be denoted by

$$\pi(\theta_i \mid \theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_k, \boldsymbol{y}) = \pi(\theta_i \mid \theta_{-i}, \boldsymbol{y}); \qquad i = 1, \cdots, k.$$

The algorithm for componentwise transitions is given by Algorithm 2.

Algorithm 2: Metropolis-Hastings: Componentwise Transitions

1. Set the initial state of the chain to $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, ..., \theta_k^{(0)})^T$ and set the iteration counter to $j = 1$.
2. For every iteration j, we obtain a new value of $\boldsymbol{\theta}^{(j)}$ from $\boldsymbol{\theta}^{(j-1)}$ by successive generation from distributions:
   - $\theta_1^{(j)} \sim \pi(\theta_1|\theta_2^{(j-1)}, \theta_3^{(j-1)}, ..., \theta_k^{(j-1)}, \boldsymbol{y})$ using a Metropolis-Hastings step with proposal distribution $q_1(\theta_1^*|\theta_1^{(j-1)})$
   - $\theta_2^{(j)} \sim \pi(\theta_2|\theta_1^{(j)}, \theta_3^{(j-1)}, ..., \theta_k^{(j-1)}, \boldsymbol{y})$ using a Metropolis-Hastings step with proposal distribution $q_2(\theta_2^*|\theta_2^{(j-1)})$

     $\vdots$
   - $\theta_k^{(j)} \sim \pi(\theta_k|\theta_1^{(j)}, \theta_2^{(j)}, ..., \theta_{k-1}^{(j)}, \boldsymbol{y})$ using a Metropolis-Hastings step with proposal distribution $q_k(\theta_k^*|\theta_k^{(j-1)})$.
3. Change the iteration counter from j to $j + 1$ and return to step 2.

This is in fact the original form of the Metropolis algorithm where the Metropolis-Hastings algorithm presented in Algorithm 1 can be regarded as a special case of this algorithm. If the full conditional distribution for the particular component θ_i is available for sampling directly, then it is easy to show that the resulting acceptance probability is one. For this reason, this algorithm can also be called Metropolis-within-Gibbs. When all full conditional distributions are completely known and available for sampling from, then we can obtain an algorithm known as the Gibbs sampler which is presented in the next section.

2.3.4.3 Gibbs Sampler

The Gibbs sampler was originally developed by Geman & Geman (1984) for image processing before Gelfand & Smith (1990) brought this approach to the larger statistical community (Gamerman & Lopes, 2006). Suppose we have a posterior density $\pi(\boldsymbol{\theta}|\boldsymbol{y})$, where $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_k)^T$. To generate realisations from this posterior density, samples are drawn from full conditional distributions, with densities

$$\pi(\theta_i|\theta_1, ..., \theta_{i-1}, \theta_{i+1}..., \theta_k, \boldsymbol{y}) = \pi(\theta_i|.), \ i = 1, 2, ..., k.$$

The algorithm for Gibbs sampler is hence given by Algorithm 3.

Algorithm 3: The Gibbs Sampler

1. Set the initial state of the chain to $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, ..., \theta_k^{(0)})^T$ and set the iteration counter to $j = 1$.

2. For every iteration j, we obtain a new value of $\boldsymbol{\theta}^{(j)}$ from $\boldsymbol{\theta}^{(j-1)}$ by successive sampling from the full conditional distributions as follows:

$$\begin{aligned}
\theta_1^{(j)} &\sim \pi(\theta_1 | \theta_2^{(j-1)}, \theta_3^{(j-1)}, ..., \theta_k^{(j-1)}, \boldsymbol{y}) \\
\theta_2^{(j)} &\sim \pi(\theta_2 | \theta_1^{(j)}, \theta_3^{(j-1)}, ..., \theta_k^{(j-1)}, \boldsymbol{y}) \\
&\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\
\theta_k^{(j)} &\sim \pi(\theta_k | \theta_1^{(j)}, \theta_2^{(j)}, ..., \theta_{k-1}^{(j)}, \boldsymbol{y})
\end{aligned}$$

3. Change the iteration counter from j to $j + 1$ and return to step 2.

2.3.4.4 Analysing MCMC output

By taking a sufficient number of iterations, the MCMC scheme will give a representative sample from the posterior distribution. However the number is only sufficient if convergence has been reached. There is a possibility that the output from a MCMC scheme can lead to inaccuracy and computational inefficiency. Therefore, it is important to monitor the convergence carefully to ensure that the convergence has been reached. We can assess convergence of the output by doing some diagnostic checks. A "burn-in" period is a process where we discard some number of iterations of the samples at the beginning of the MCMC run. Since we are interested only in samples from the posterior distribution, it is appropriate to remove these samples which are obtained before the chain has converged. Next, we can use a trace plot to check the convergence informally where the chain will show the same qualitative behaviour if it has converged. Otherwise, the chain will display a trend over the sample space. However, there are some circumstances where the chain has converged but it was trapped in a local mode rather than exploring the full posterior. Therefore, running two or more chains simultaneously from different starting points can be helpful. The lack of convergence can be detected if the chains fail to overlap after comparing their trace plots. It is also dangerous to look at selected trace plots only because every parameter might have a different outcome. Then, it is crucial for us to check the trace plot for all parameters. In some cases, a bad trace plot for one parameter can undermine the inference of other parameters. Hence, we should treat results with caution

if there is evidence of non-convergence in any of the trace plots. There are also various formal diagnostic checks available such as those recommended by Heidelberger & Welch (1983), Geweke (1992), Raftery & Lewis (1992) and Gelman & Rubin (1992) and these had been thoroughly reviewed by Cowles & Carlin (1996).

Apart from that, we can also use kernel density plots to identify multimodality in the posterior distribution. If multimodality of the marginal posterior is detected, then we may need to take an action by running the MCMC algorithm longer to ensure that the entire sample space is covered adequately.

Samples of the MCMC scheme will be dependent, meaning successive draws are autocorrelated. This can be observed by looking at an autocorrelation plot. When there is little dependence between successive samples, the chain is said to mix well. If there is strong correlation between successive values in the chain, it will take longer to explore the entire region of the parameter space which is described as poor mixing in the chain.

In general, the distribution of the chain $\boldsymbol{\theta}^{(i)}|\boldsymbol{y}$ tends to the posterior distribution $\boldsymbol{\theta}|\boldsymbol{y}$ by increasing the number of iterations until convergence is reached.

2.3.5 Data augmentation

In some circumstances, the direct computation of the posterior density for some models is difficult to handle because it is laborious to compute the likelihood. In some cases, it is possible to simplify this problem by using data augmentation. Data augmentation is a common approach in Bayesian statistics to construct an iterative algorithm for the posterior sampling by introducing unobservable variables, known as auxiliary variables. If these variables were observed, then the computation of the posterior density would become straightforward. This approach was demonstrated by Tanner & Wong (1987) and MCMC methods are well suited to this approach.

Suppose we have observed data, $\boldsymbol{y}$ from a distribution which is conditional on the parameter vector, $\boldsymbol{\theta}$. The idea of this approach is to augment $\boldsymbol{y}$ with the auxiliary data, $\boldsymbol{z}$. Thus, the posterior density is given as follows:

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) = \int \pi(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{z})\pi(\boldsymbol{z}|\boldsymbol{y})d\boldsymbol{z} \tag{2.9}$$

where

$$\pi(\boldsymbol{z}|\boldsymbol{y}) = \int \pi(\boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{y})\pi(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}. \tag{2.10}$$

Tanner & Wong (1987) then introduced an iterative sampling scheme to construct approximations for $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ and $\pi(\boldsymbol{z}|\boldsymbol{y})$ using two steps:

1. draw samples for $\boldsymbol{z}$ using $\pi(\boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{y})$,

2. then, based on $\boldsymbol{z}$, draw samples for $\boldsymbol{\theta}$ using $\pi(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{z})$.

The first step is known as "imputation" step and the later is called "posterior" step. Often, the data augmentation is used to handle missing data using an MCMC scheme. It can also be applied in mixture distributions by introducing a group-membership variable which is unobserved. In this case, $\boldsymbol{z}$ can be expressed as the component of the mixture which generates the observed data. If $\boldsymbol{z}$ is known, then the computational analysis become simpler by considering that $z_i = j$, where $j \in \{1, \cdots, J\}$ is a component of the mixture distribution. This will be discussed in Chapter 3, Section 3.2.

2.3.6 Bayesian inference for time series models

In time series analysis, the Bayesian approach can provide a systematic framework which offers a complete way to analyse data by combining the prior information with the data. Then, Bayesian analysis can be applied to various types of time series models. For instance, a Bayesian approach can be used to analyse the uncertainty of missing values in the sequence of any observations. Kong *et al.* (1994) showed that the posterior distribution for missing data could be generated by applying the data augmentation and a Gibbs sampler to the model. Originally, the Bayesian approach was not very popular in time series analysis since it is analytically intractable especially when the conjugate distributions do not represent the prior distribution of the parameters. However, this difficulty has been resolved in the 1990s due to more advanced computing power and the introduction of MCMC for sampling from Bayesian posterior distributions. Some of the Bayesian methods to compute the posterior distribution have already been discussed in Section 2.3.4.

Traditionally, the DLM approach tends to be favoured by Bayesian statisticians for modelling time series, probably for historical reasons (see Harrison & Stevens (1971) and West & Harrison (1997)). The DLM approach is more easily described in Bayesian terms. For instance, our beliefs about the future can be represented by the expectation of future values and can be updated when we observe data. The Bayesian approach can also be extended to other types of state space model such as DGLM (see West *et al.* (1985)) and HMM (see Scott (2002)). Consider a DGLM where the observational distribution at time t has a density $f_y(y_t|\mu_t, \phi)$, where ϕ is a scale parameter and μ_t is a location parameter. Let $\eta_t = g(\mu_t) = F_t' X_t$ and the evolution is given in (2.6b). Using the ideas of data augmentation, as in Section 2.3.5, the full-data likelihood is

$$f_0(X_0) \prod_{t=1}^{T} f_Y(y_t|\mu_t, \phi) f_X(X_t|X_{t-1}).$$

The system vectors $X_0, X_1, \cdots, X_T$ are unobserved and can be treated as auxiliary data in an MCMC scheme.

However, there is no reason why Bayesians can not use other models in time series such as ARMA models. A considerable amount of early literature has been published on ARMA models within a Bayesian framework, for example, Zellner (1971), Harrison & Stevens (1976), and Monahan (1983). The ARMA models can also be found in some application of microeconomics time series such as unemployment rates using Bayesian approach (de Alba, 1993; Rosenberg & Young, 1999). Thus, there is strong evidence that the Bayesian approach can be applied to any time series models.

2.4 Conclusion

This chapter has briefly discussed the general approaches in time series models and Bayesian inference. We have presented a simple discussion in Section 2.3.6 about the application of Bayesian approach in time series models. There are a few time series methods that we will apply to build a daily rainfall model within the Bayesian framework. It is important to introduce some of the preliminary ideas that we will use in this study. We have also emphasised the use of MCMC techniques to generate posterior samples for the unknown parameters. Next, we will discuss the application of the Bayesian approach in a mixture model.

Chapter 3

Introduction to Bayesian inference in mixture models

3.1 Introduction

Mixture models were introduced by Newcomb (1886) and Pearson (1894) as a tool for statistical modelling. There are two main reasons why we want to use a mixture model as a sampling distribution. Firstly, we believe that there are two or more sub-populations and it is therefore reasonable to represent them as mixture components. For example, the size of starlings might be different for resident birds and migrants. Thus, it is sensible to assume that there are two sub-populations of starlings in the samples. Secondly, a mixture distribution gives greater flexibility in the sampling model and the shape of the distribution in spite of not having the "physical" interpretation for the mixture components.

In this chapter we will review the application of Bayesian inference in mixture models, focusing on finite mixtures. The form of the likelihood for the finite mixture model will be discussed in Section 3.2. Section 3.3 illustrates the construction of Monte Carlo Markov Chain (MCMC) schemes for the finite mixture model with discussions on the label switching problem. We will discuss the special case of mixture models which is also known as a mixed distribution in Section 3.4. The observations in this model type contain both discrete and continuous components for example zero values and positive values. Then, we will apply the developed mixture models to data on ultrasound measurements within the Bayesian framework in Section 3.5. In this section, two types of mixture models will be considered in this application. The construction of the prior distribution and the resulting posterior distribution for both models will also be discussed in this section.

3.2 Finite mixture model

In general, the distribution of Y_i given the parameter θ_j is called a finite mixture model if the probability density function of Y_i is

$$f(y_i|\boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{j=1}^{J} \pi_j f_j(y_i|\boldsymbol{\theta}_j) \tag{3.1}$$

where $\sum_{j=1}^{J} \pi_j = 1$ and $\pi_j \geq 0$ is the probability of each component j membership with J as the number of components. Then, if we have independent observations $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)^T$, the likelihood for the mixture model can be written as:

$$f(\boldsymbol{y}|\boldsymbol{\pi}, \boldsymbol{\theta}) = \prod_{i=1}^{n} \left\{ \sum_{j=1}^{J} \pi_j f_j(y_i|\boldsymbol{\theta}_j) \right\} \tag{3.2}$$

which is known as the observed data likelihood.

For instance, suppose that we have a simple two-component mixture model. Then the density of Y_i is given as:

$$f(y_i|\pi, \theta_1, \theta_2) = \pi f_1(y_i|\theta_1) + (1 - \pi) f_2(y_i|\theta_2)$$

where $f_j(y|\theta_j)$ is the density for component j which depends on parameters θ_j. In this model, the probability for the first component is represented by π and the probability of second component is $(1 - \pi)$, with $0 \leq \pi \leq 1$. Thus, the likelihood of independent observations $\boldsymbol{y}$ is given as:

$$f(\boldsymbol{y}|\pi, \theta_1, \theta_2) = \prod_{i=1}^{n} [\pi f_1(y_i|\theta_1) + (1 - \pi) f_2(y_i|\theta_2)].$$

However, if we have more than two components, the likelihood in (3.2) could be very complicated. One way to simplify inference for this model is by using data augmentation (Tanner & Wong, 1987). Using this approach, we can introduce a group-membership variable for mixture components which is unobserved. Suppose that we have a set of auxiliary variables $\boldsymbol{z} = (z_1, \cdots, z_n)^T$ which corresponds to the unknown component of mixture with $z_i \in \{1, \cdots, J\}$. Given $z_i = j$, the conditional likelihood can be written as

$$f(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\theta}) = \prod_{i=1}^{n} f_{z_i}(y_i|\boldsymbol{\theta}_{z_i}).$$

If $\boldsymbol{z}$ is known, the conditional density for y_i can thus be simplified as $f_{z_i}(y_i|\boldsymbol{\theta}_i)$. Given the

probability of each unknown component of mixture, $\Pr(z_i = j) = \pi_j$, the variables $\boldsymbol{z}$ have a multinomial or categorical distribution. Then, by observing $\boldsymbol{y}$ and $\boldsymbol{z}$, the likelihood is given as

$$f(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{\pi}, \boldsymbol{\theta}) = \prod_{i=1}^{n} \pi_{z_i} f_{z_i}(y_i|\boldsymbol{\theta}_{z_i}).$$

This likelihood is called the complete data likelihood, which is in contrast to the likelihood in (3.2), which only assumes the knowledge of the observed data and not the auxiliary variables. Nonetheless, the marginal likelihood of the observed data, $\boldsymbol{y}$ can be obtained by summing the complete data likelihood over all possible $\boldsymbol{z}$. Here, we can either use a Gibbs sampler or Metropolis-within-Gibbs algorithm to sample the correct marginal posterior distribution of the parameters and the auxiliary variables.

A considerable amount of literature has been published on mixture models. McLachlan & Peel (2001), Jasra *et al.* (2005), Marin *et al.* (2005), and Frühwirth-Schnatter (2006) give a comprehensive review for finite mixture models. Finite mixture models can be used in wide-ranging application areas such as economics, medicine, biology and physics, and this can be further extended to bioinformatics and genetics (Frühwirth-Schnatter, 2006). For further information, see Delmar *et al.* (2004) and Tadesse *et al.* (2005). A fine example on the use of a finite mixture model for the modelling of the time intervals between eruptions at the "Old Faithful" geyser in the Yellowstone National Park, Wyoming, USA is given by Azzalini & Bowman (1990). The data are illustrated in Figure 3.1. From this figure, we

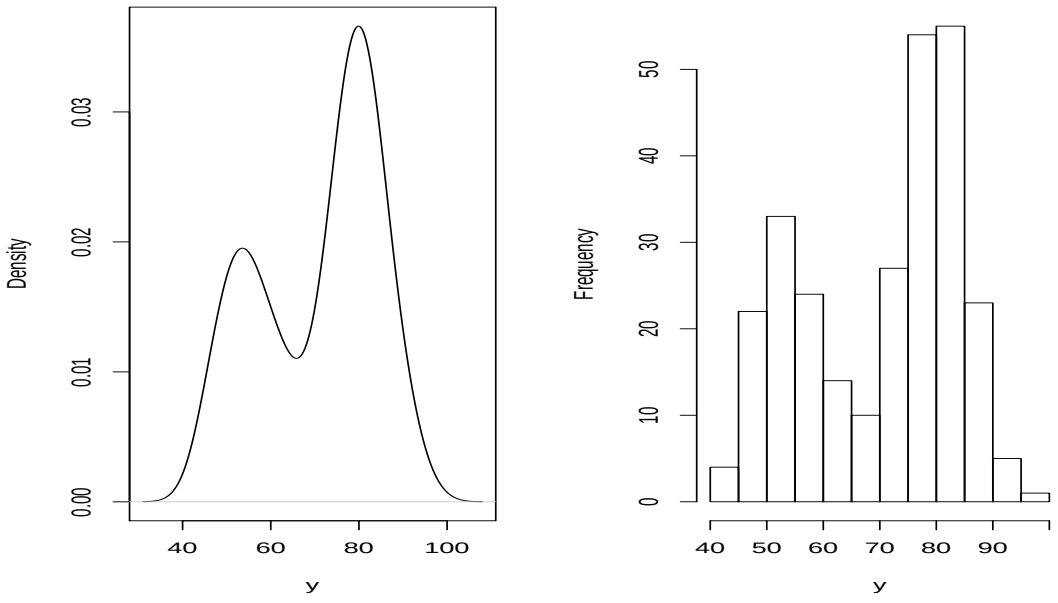


Figure 3.1: Density plot and histogram of mixture model example from "Old Faithful" geyser data

can see clearly that there are two distinct modes. Thus, it is possible to infer that there are two sub-populations in the data, justifying the use of a mixture model for this kind of problem.

3.3 Markov chain Monte Carlo (MCMC) for mixture model

3.3.1 Algorithm

In this section, we will briefly describe the Markov chain Monte Carlo (MCMC) steps for a mixture model. Suppose that we have a joint prior density $\pi(\boldsymbol{\pi}, \boldsymbol{\theta})$ where $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_J)^T$ and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_J)$. In general, the joint density for all quantities is given by:

$$\begin{aligned}\pi(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\theta}) &= f(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{\pi}, \boldsymbol{\theta})\pi(\boldsymbol{\pi}, \boldsymbol{\theta}) \\ &= \left\{\prod_{i=1}^{n} \pi_{z_i} f_{z_i}(y_i|\boldsymbol{\theta}_{z_i})\right\} \pi(\boldsymbol{\pi}, \boldsymbol{\theta}).\end{aligned}$$

To draw posterior samples for $\boldsymbol{\theta}$, $\boldsymbol{\pi}$ and $\boldsymbol{z}$, we can use an MCMC scheme and draw samples from their full conditional distributions which are in turn dependent upon the form of the joint prior for $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$. Consider that we have a case where $\boldsymbol{\theta}_j$ and $\boldsymbol{\pi}$ are independent from each other and $\boldsymbol{\theta}_j$ is independent of $\boldsymbol{\theta}_k$ for $k \neq j$. The joint prior distribution can be written as

$$\pi(\boldsymbol{\pi}, \boldsymbol{\theta}) = \pi(\boldsymbol{\pi}) \prod_{j=1}^{J} \pi(\boldsymbol{\theta}_j).$$

Therefore, an MCMC scheme for a mixture model is given as follows:

1. Sample a new value for $\boldsymbol{\theta}$.

   Let $\boldsymbol{\theta}_{-j} = (\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_{j-1}, \boldsymbol{\theta}_{j+1}, \cdots, \boldsymbol{\theta}_J)$. The full conditional distribution for $\boldsymbol{\theta}_j$ is given as

   $$\begin{aligned}\pi(\boldsymbol{\theta}_j|\boldsymbol{\theta}_{-j}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\pi}) &\propto \pi(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\theta}) \\ &\propto \left\{\prod_{i=1}^{n} \pi_{z_i} f_{z_i}(y_i|\boldsymbol{\theta}_{z_i})\right\} \pi(\boldsymbol{\pi}) \prod_{j=1}^{J} \pi_j(\boldsymbol{\theta}_j) \\ &\propto \pi_j(\boldsymbol{\theta}_j) \prod_{i \in C_j} f_j(y_i|\boldsymbol{\theta}_j)\end{aligned}$$

   where $C_j = \{i : z_i = j\}$ represents the subset of observations belonging to component j. The parameterisation of $\pi(\boldsymbol{\theta}_j|\boldsymbol{\theta}_{-j}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\pi})$ can be simplified as $\pi(\boldsymbol{\theta}_j|\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\pi})$ since the full conditional distribution does not depend on $\boldsymbol{\theta}_k$ where $k \neq j$. Then, the new

value for $\boldsymbol{\theta}$ is obtained by taking an independent realisation from each of $\pi(\boldsymbol{\theta}_j|\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\pi})$ with $j = 1, \cdots, J$.

2. Sample a new value for $\boldsymbol{\pi}$

   The full conditional distribution of $\boldsymbol{\pi}$ can be written as

$$\begin{aligned}\pi(\boldsymbol{\pi}|\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\theta}) &\propto \pi(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\theta}) \\ &\propto \left\{\prod_{i=1}^{n} \pi_{z_i} f_{z_i}(y_i|\boldsymbol{\theta}_{z_i})\right\} \pi(\boldsymbol{\pi}) \prod_{j=1}^{J} \pi(\boldsymbol{\theta}_j) \\ &\propto \pi(\boldsymbol{\pi}) \prod_{j=1}^{J} \pi_j^{n_j}\end{aligned}$$

   where n_j represents the number of observations belonged to component j. Typically, the density of $\pi(\boldsymbol{\pi})$ is that of a Dirichlet distribution and since the full conditional distribution is also multinomial, then this is a conjugate update.

3. Sample a new value for $\boldsymbol{z}$

   Let $\boldsymbol{z}_{-i} = (z_1, \cdots, z_{i-1}, z_{i+1}, \cdots, z_n)$ and then, the full conditional distribution for z_i is a discrete distribution with

$$\begin{aligned}\pi(z_i = j|y_i, \boldsymbol{\pi}, \boldsymbol{\theta}) &\propto \pi(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\theta}) \\ &\propto \left\{\prod_{i=1}^{n} \pi_{z_i} f_{z_i}(y_i|\boldsymbol{\theta}_{z_i})\right\} \pi(\boldsymbol{\pi}) \prod_{j=1}^{J} \pi(\boldsymbol{\theta}_j) \\ &\propto \pi_j f_j(y_i|\boldsymbol{\theta}_j).\end{aligned}$$

   where $\pi(z_i = j|y_i, \boldsymbol{\pi}, \boldsymbol{\theta}) = \pi(z_i = j|\boldsymbol{z}_{-i}, \boldsymbol{y}, \boldsymbol{\pi}, \boldsymbol{\theta})$ since the full conditional distribution does not depend on z_k with $k \neq i$. The new value for $\boldsymbol{z}$ can then be sampled by taking independent realisations from each $\pi(z_i = j|y_i, \boldsymbol{\pi}, \boldsymbol{\theta})$, $j = 1, \cdots, J$ for $i = 1, \cdots, n$.

3.3.2 Label switching

Label switching in mixture models was first described by Redner & Walker (1984). Label switching causes difficulties in MCMC sampling, making the posterior distribution difficult to summarize. This usually occurs in a mixture model because of the unchanged likelihood under relabelling of mixture components (Stephens, 2000). For example, consider J components in (3.2) where each component is set to the same distribution family. Then, the likelihood value for all $J!$ possible permutations will all be the same and the posterior will become multimodal unless we have a strong prior that can distingush between the components of the mixture.

Label switching happens when the sampler starts to jump from one component to another, causing switching from one permutation of the labels to another. One solution to prevent label switching is to restrict the order of the J parameters such as $\theta_1 < \theta_2 < \ldots < \theta_J$. This solution is effective but it has the disadvantage when our model contains more components than we actually need. Hence, we might encounter another problem where label switching occurs between different possibilities. At present, there are several advanced solutions that can be used to prevent label switching in mixture models. For example, see Stephens (2000) and Puolamäki & Kaski (2009).

3.4 Mixed Distribution

In most discussion of time series, the data are either discrete or continuous. However, there is a special case of mixture models where the data are composed of both discrete and continuous components and this is called a mixed distribution. An important case of a mixed distribution occurs when the observations may not only take positive values but also include a lot of zero values. As a result, the observations of the time series are a combination of a discrete component at zero and a continuous component on the positive real line. A distribution of this type commonly has the features of a spike or a discrete probability mass at zero, followed by a bump or ramp for the positive values on the real line as demonstrated in Figure 3.2. In this figure, the area of the ellipse at $y = 0$ represents

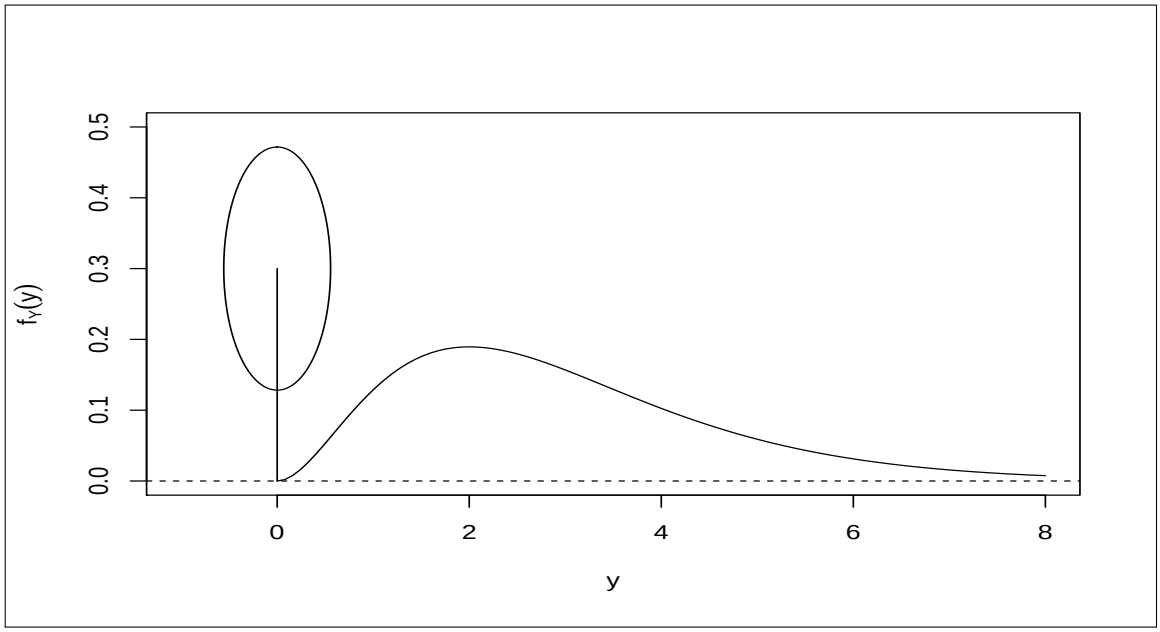


Figure 3.2: The example of the features for zero-continuous distribution

the probability of a zero observation and the total area of this circle and the area under the

density curve is equal to one. Frequently, this happens in many application areas such as meteorology, ecology, economics, and environmental sciences. An example is daily rainfall amounts.

The majority of literature suggests that data of this type can be conceptually modelled using a "two-part framework" (Tooze *et al.*, 2002; Grunwald & Jones, 2000; Stern & Coe, 1984). The model can be separated into two parts that represent two different processes. The first part is the occurrence process which governs the probability of a positive observation whilst the second part corresponds to the amount process which characterizes the amount of each positive observation. To fit the amount process, the choice of distribution is critical to ensure a good fit and an accurate predictive distribution. For rainfall data, Heaps *et al.* (2015) used a log-normal distribution, while Grunwald & Jones (2000) and Stern & Coe (1984) used a gamma distribution. On the other hand, Sanso & Guenni (1999) used a power transformed truncated normal distribution for fitting rainfall amounts.

In general, the likelihood for independent observations from a zero-positive mixed distribution with parameters π and $\boldsymbol{\theta}$ can be represented as follows:

$$L(\pi, \boldsymbol{\theta}|\boldsymbol{y}) = \prod_{i=1}^{n} \left[\pi f(y_i|\boldsymbol{\theta})\right]^{r_i} \left[(1-\pi)\right]^{1-r_i} \tag{3.3}$$

where $f(.)$ is the probability density function for the amount process and $r = 1$ if $y > 0$ and otherwise $r = 0$. The binomial distribution is employed for the occurrence process and this is used for finding the probability for each component. To relax the assumption of independence in the occurrences, a Markov chain is often used to model the sequence of zero and positive components. For example, Grunwald & Jones (2000) and Stern & Coe (1984) employ a Markov chain for the occurrence process when modelling rainfall data.

3.5 Application: Mixture Model for Ultrasound Data

3.5.1 Background

Ultrasound can be used as a means to monitor and characterize fluid systems in research and industry for example in fermentation processes, since ultrasound can be propagated through fluid systems. Normally, the propagation is characterized by the time-of flight and the voltage amplitude attenuation of the wave. However, the existence of large-scale dynamic fluctuations such as the presence of bubbles crossing the ultrasound path in the ultrasonic data is inevitable due to the nature of the system. One way to ensure that the ultrasonic properties of the background fluid are accurately isolated is by removing these

"rogue" measurements.

Cowburn *et al.* (2018) set up an experiment to measure the time-of-flight and voltage of the ultrasound pulse traveling through water containing bubbles. The experiment is set up by housing the ultrasound transducer near to an aluminum vessel which contained the sample fluid. To maintain an ambient temperature of $25°C$ ($\pm 0.5°C$), the vessel needed to be submerged in a heated water bath. The sample fluid was stirred to ensure the horizontal positions of the bubbles are randomized. To produce a broadband ultrasound pulse, a USB-connected pulser-receiver played a role to stimulate the transducer using a narrow Radio Frequency (RF) pulse with a duration of 1 μs. Subsequently, the pulse was repeated at a frequency of 1 kHz (for details, see Cowburn *et al.* (2018)).

The ultrasonic data are collected by recording the measurement of returning ultrasound waves which generate a voltage signal. These returning ultrasound waves occur because of two conditions. Firstly, the echo occurs secondary to the reflection off the container's wall and secondly, from the bubbles in the container. In this experiment, monitoring the peak of the voltage trace and recording its time-of-flight and voltage amplitude are important. Two voltage peaks can be seen at the early and late times of flight, corresponding to the reflections off the container's wall and bubble surfaces.

3.5.2 The model

In this section, we will introduce the mixture model for the ultrasound data. As we believe that there are two distinct types of observations (bubble reflections - rogue measurement and the real reflection off the container's wall), it makes sense to represent each component using a mixture model. For this case, the first component is represented by the ultrasound waves reflected from bubbles and the second component is represented by the ultrasound waves that were being reflected from the container's wall. Figure 3.3 shows a histogram of the ultrasound data. The figure shows that the data are mostly concentrated around 0.13 and the other observations are scattered between 0 and 0.12. Even though we do not have strong evidence from these data, we will still consider fitting a two-component mixture model to these data. We hence propose to fit the data using a two-component lognormal mixture model since the observations are always positive. Furthermore, the lognormal distribution is similar in shape to a gamma distribution. Therefore we also consider applying a two-component gamma mixture to these data. There are five parameters in total for each mixture model: one for the mixing probability and two for each component distribution. We note that the likelihood might not clearly distinguish all possible values of these parameters. Therefore, the prior specification for each parameter is important because this may affect the separation and identification of each component in the mixture.

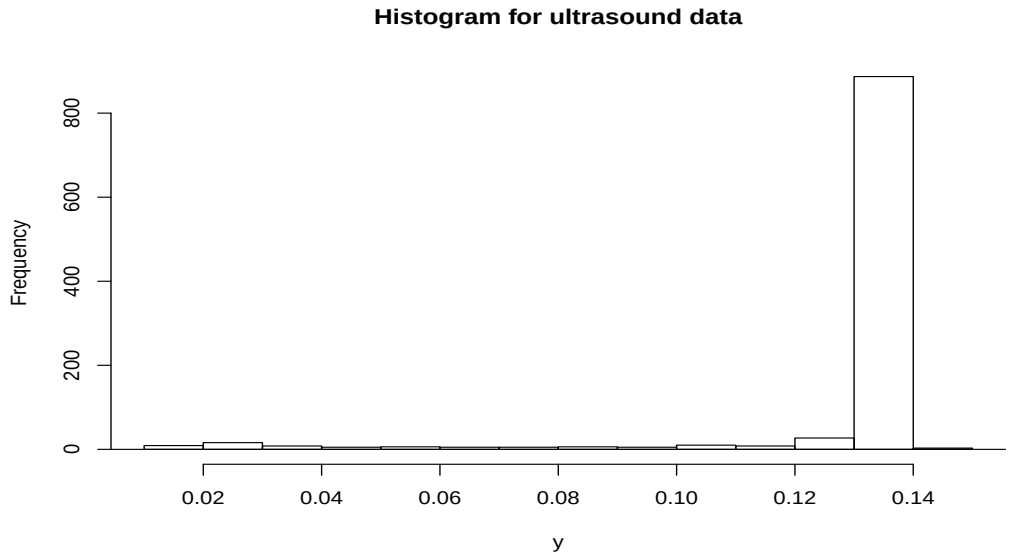


Figure 3.3: Histogram for Ultrasound data

3.5.2.1 Lognormal mixture model

Let $Y_i \in \mathbb{R}^+$ be a random variable for the time-of-flight of the ultrasound data with observed value, y_i. Suppose that Z_i represent the occurrence variable and serve as an indicator function for Y_i where

$$Z_i = \begin{cases} 1, & \text{reflected from bubbles} \\ 2, & \text{reflected from the container's wall.} \end{cases} \tag{3.4}$$

The variable Z_i has conditional probabilities which follows a Bernoulli distribution:

$$Z_i|p \sim Bern(p)$$

and so

$$f(z_i|p) = \begin{cases} p, & (z_i = 1) \\ 1-p, & (z_i = 2) \end{cases} \tag{3.5}$$

where p is a probability for the first component.

In this preliminary analysis, we assume Y_i follows a lognormal distribution:

$$Y_i|\mu_j, \tau_j, z_i = j \sim LN(\mu_j, 1/\tau_j)$$

where the mean and variance of Y_i are given as $\lambda_j = e^{\mu_j + 1/2\tau_j}$ and $V_j = e^{2\mu_j + 1/\tau_j}(e^{1/\tau_j} - 1)$

with $j \in \{1, 2\}$. In this model, the priors for the parameters are given as follows:

$$p \sim Beta(a_p, b_p)$$
$$\mu_j \sim N(\mu_{0,j}, \nu_{0,j})$$
$$\tau_j \sim Ga(g_{\tau_j}, h_{\tau_j})$$

Then, the joint density is given by:

$$\pi(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\mu}, \boldsymbol{\tau}, p) = \prod_{i=1}^{I} f(y_i|\mu_{z_i}, \tau_{z_i}) \times \prod_{j=1}^{2} \pi(\mu_j) \times \prod_{j=1}^{2} \pi(\tau_j) \times \prod_{i=1}^{I} f(z_i|p) \times \pi(p).$$

According to Bayes' rule, the posterior distribution can be written as

$$\pi(\mu_1, \mu_2, \tau_1, \tau_2, p|\boldsymbol{y}) \propto \pi(\mu_1, \mu_2)\pi(\tau_1, \tau_2)\pi(p)L(\boldsymbol{y}|\mu_1, \mu_2, \tau_1, \tau_2, p)$$

where $L(\boldsymbol{y}|\mu_1, \mu_2, \tau_1, \tau_2, p)$ is the likelihood for $\boldsymbol{y}$.

3.5.2.1.1 Prior distributions The prior specification for the parameters is chosen carefully by giving numerical values to $\mu_{0,j}, \nu_{0,j}, g_{\tau_j}, h_{\tau_j}, a_p$ and b_p. Suppose that, for component j, we give μ_j a normal prior and τ_j a gamma prior, where μ_j and τ_j are deemed independent. Suppose that we imagine a large sample $Y_{1j}, \cdots, Y_{nj}$ from component j, and let $X_{ij} = \log Y_{ij}$ where n is large enough so that the sample mean $\bar{X}_j = \frac{1}{n}\sum X_{ij}$ is approximately μ_j and the sample variance $S^2_{xj} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_{ij} - \bar{X}_j\right)^2$ is approximately $\sigma_j^2 = \tau_j^{-1}$. Suppose that our three quartiles for $\bar{X}_j$ are Q_{m1j}, Q_{m2j} and Q_{m3j}. To obtain these values, we can ask the expert as follows:

Q1 *"Please think about the sample mean $\bar{X}_j$ from component j. Can you give a value such that you think that it is equally likely that $\bar{X}_j$ will be less than or greater than this value?"* Let the given value be Q_{m2j}.

Q2 *"Please think about the sample mean $\bar{X}_j$ from component j. Suppose that you were told that $\bar{X}_j$ is less than $\bar{Q}_{m2j}$. Can you give a new value such that you think that it is equally likely that $\bar{X}_j$ will be less than or greater than this value?"* Let the given value be Q_{m1j}.

Q3 *"Still thinking about the sample mean $\bar{X}_j$ from component j. Suppose that you were told that $\bar{X}_j$ is greater than $\bar{Q}_{m2j}$. Can you give a new value such that you think that it is equally likely that $\bar{X}_j$ will be less than or greater than this value.* Let the given value be Q_{m3j}.

Then our prior mean for μ_j is $\mu_{0j} = \frac{1}{2}(Q_{m1j} + Q_{m3j})$ and our prior variance for μ_j is $v_{0j} = [(Q_{m3j} - Q_{m1j})/1.349]^2$.

Suppose that our three quartiles for S^2_{xj} are Q_{S1j}, Q_{S2j} and Q_{S3j} where we can gain these point values using these questions:

Q4 *"Please think about the sample variance S^2_{xj} from component j. Please give a value such that S^2_{xj} is equally likely to be less than or greater than this value."* Let the given value be Q_{S2j}.

Q5 *"Still thinking about the sample variance S^2_{xj} from component j. Suppose that you were told that S^2_{xj} is less than Q_{S2j}. Please give a new value such that it is equally likely that S^2_{xj} is less than or greater than this value."*.Let the given value be Q_{S1j}.

Q6 *"Still thinking about the sample variance S^2_{xj} from component j. Suppose that you were told that S^2_{xj} is above than Q_{S2j}. Please give a new value such that it is equally likely that S^2_{xj} is less than or greater than this value."* Let the given value be Q_{S3j}.

Then our three quartiles for τ_j are Q^{-1}_{S3j}, Q^{-1}_{S2j} and Q^{-1}_{S1j}. The parameter τ_j has a $Ga(g_{\tau j}, h_{\tau j})$ distribution where $g_{\tau j}$ is chosen by iteratively solving $Q^{-1}_{S1j}/Q^{-1}_{S3j} = R_3(g_{\tau j})/R_1(g_{\tau j})$, where $R_q(a)$ is quartile q for a $Ga(a,1)$ distribution, and $h_{\tau j} = R_2(g_{\tau j})/Q^{-1}_{S2j}$.

Using this method, we have reasonable values which can be used for illustration as follows:

$$\begin{array}{lll} Q_{m11} = -3.5 & Q_{m31} = -2.5 & \\ Q_{m12} = -2.05 & Q_{m32} = -1.95 & \\ Q_{S11} = 0.5 & Q_{S21} = 0.8 & Q_{S31} = 1.2 \\ Q_{S12} = 0.0002 & Q_{S22} = 0.0004 & Q_{S32} = 0.0008 \end{array}$$

and hence

$$\begin{array}{ll} \mu_{01} = -3 & v_{01} = 0.55 \\ \mu_{02} = -2 & v_{02} = 0.005 \\ g_{\tau 1} = 2.64 & h_{\tau 1} = 1.85 \\ g_{\tau 2} = 1.22 & h_{\tau 2} = 0.0004. \end{array}$$

Now, suppose that the probability of component 1 is given by p which follows a beta distribution. To elicit the prior for p, we can use these questions to ask the expert, based on the hypothetical future sample method of Winkler (1967):

Q7 *"Please give your assessment of the probability that a randomly chosen observation belongs to component j."* Let the given value be m_0 which is the prior mean for p.

Q8 *"Suppose that you were given a random sample of n_f observations and that x_f of these belonged to component j. With these additional data, what would now be your assessment of the probability that a new observation would belong to component j?"* Let the given value be m_f. Note that actual numerical values are given for n_f and x_f. These are chosen in the light of the value given for m_0, so that x_f/n_f is sufficiently different from m_0 and an initial judgement about the expert's certainty about p. Larger values of n_f are required if the expert has greater prior precision.

Then we have

$$\frac{a_p}{a_p + b_p} = m_0 \qquad \text{and} \qquad \frac{a_p + x_f}{a_p + b_p + n_f} = m_f.$$

Solving these, we obtain

$$a_p = m_0 \left(\frac{n_f m_f - x_f}{m_0 - m_f} \right) \qquad \text{and} \qquad b_p = (1 - m_0) \left(\frac{n_f m_f - x_f}{m_0 - m_f} \right).$$

Using this method, we obtain illustrative values as follows:

$$m_0 = 1/6 = 0.167, \qquad m_f = 1/5 = 0.2$$

with $n_f = 12$ and $x_f = 4$. Hence

$$a_p = 8, \qquad b_p = 40.$$

For further discussion of relevant elicitation methods, see, for example, O'Hagan *et al.* (2006).

3.5.2.1.2 Posterior distributions The joint posterior distribution has been demonstrated in Section 3.5.2.1. To draw the posterior samples for the unknown parameters, an MCMC scheme was applied using the R JAGS package (Plummer, 2012). We also impose the restriction $\mu_1 < \mu_2$ to avoid label switching. Following a burn-in of 5000 iterations, a further 10000 iterations were taken to draw the posterior samples. The computing time that was required to generate 10000 posterior draws was around 40 seconds by using R software on a 2.00GHz Samsung laptop 300V3A model with Intel Core i7-2630QM processor and 12 Gbytes of random-access memory. Based on the analysis, 10000 iterations are sufficient to obtain the realisations from the posterior distribution since the chains converge after the initial burn-in period. The trace plots in Figure 3.4 indicate that the mixing for all parameters is very satisfactory. Figure 3.5 shows the differences between

prior and posterior densities for the unknown parameters. A full summary of the posterior distributions for the unknown parameters is given in Table 3.1.

Parameter	Prior mean	Prior SD	Posterior mean	Posterior SD
μ_1	-3	0.74	-2.7970	0.0736
μ_2	-2	0.07	-1.9884	0.0005
τ_1	1.43	0.88	1.8801	0.2563
τ_2	3050	2761.34	5021.5155	304.4
π	0.17	0.05	0.1086	0.0103

Table 3.1: The prior and posterior means and standard deviations (SD) of the unknown parameters for the lognormal mixture model

3.5.2.2 Gamma Mixture Model

The second mixture model is a two-component gamma mixture which is an alternative to the lognormal mixture. In this model, we assume Y_i follows a gamma distribution:

$$Y_i|\alpha_j, \beta_j, z_i = j \sim Ga(\alpha_j, \beta_j)$$

where $\beta_j = \alpha_j/\lambda_j$ and $\lambda_j = \exp(\mu_j)$. The variables α_j and β_j are the shape and scale parameters for the gamma distribution where λ_j represent the mean of Y_i for $j \in \{1, 2\}$. The priors for the parameters in this model are provided as follows:

$$\mu_j \sim N(\mu_{0,j}, v_{0,j})$$
$$\alpha_j \sim Ga(g_{\alpha_j}, h_{\alpha_j}).$$

The parameterisation for $f(z_i|p)$ is similar to that for the lognormal mixture model. Then, the joint density can be expressed as:

$$\pi(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\mu}, \boldsymbol{\alpha}, p) = \prod_{i=1}^{I} f(y|\mu_{z_i}, \alpha_{z_i}) \times \prod_{j=1}^{2} \pi(\mu_j) \times \prod_{j=1}^{2} \pi(\alpha_j) \times \prod_{i=1}^{I} f(z_i|p) \times \pi(p).$$

Hence, the posterior distribution for the gamma mixture model can thus be written as

$$\pi(\mu_1, \mu_2, \alpha_1, \alpha_2, p|\boldsymbol{y}) \propto \pi(\mu_1, \mu_2)\pi(\alpha_1, \alpha_2)\pi(p)L(\boldsymbol{y}|\mu_1, \mu_2, \alpha_1, \alpha_2, p).$$

3.5.2.2.1 Prior Specification The parameter μ_j here is not the same as in the lognormal mixture model. In the lognormal case it is the mean of the logarithms of the observations and also the logarithm of the median of the observations. In the gamma case, on the other hand, it is the logarithm of the mean of the observations.

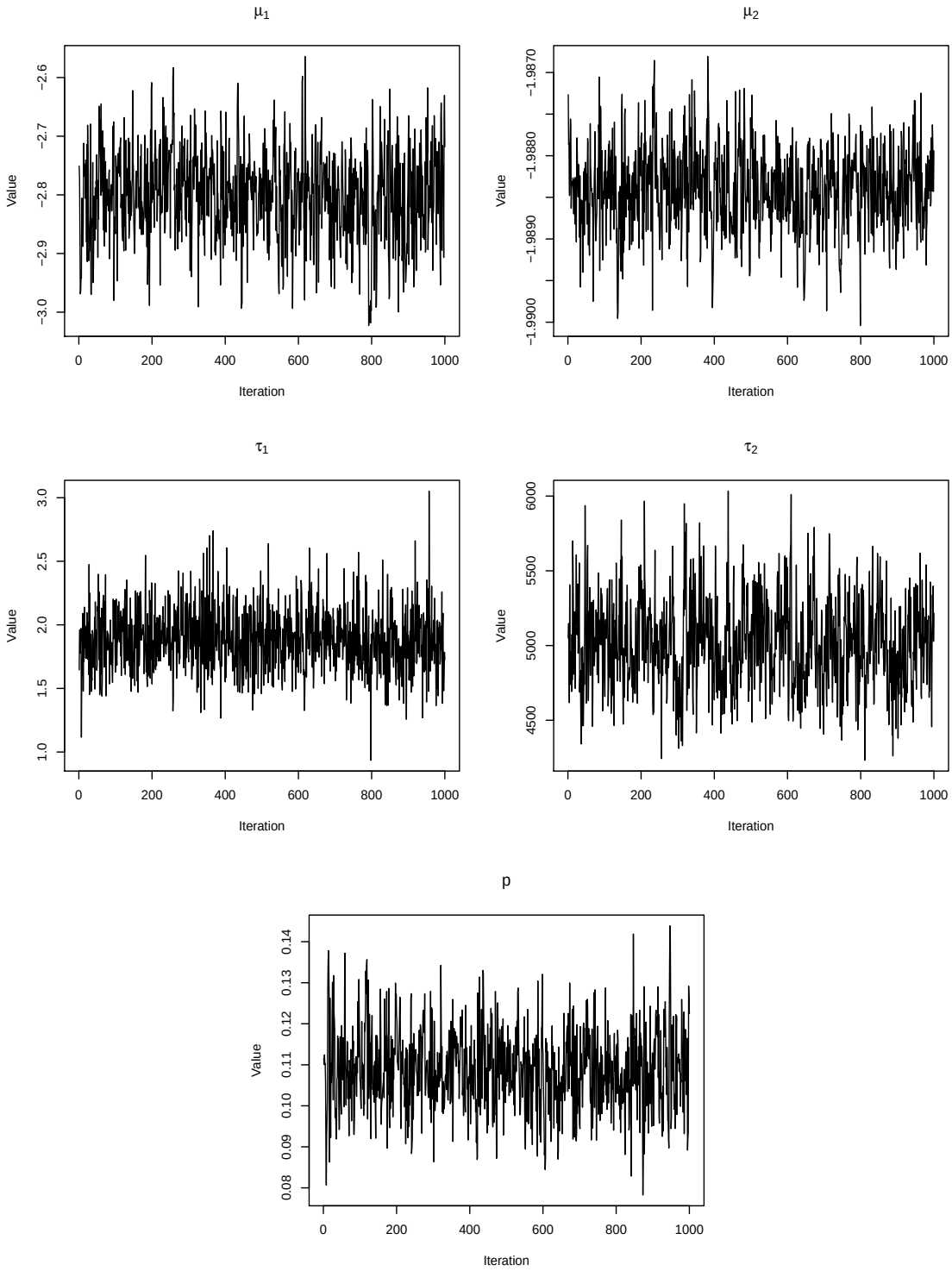


Figure 3.4: The trace plots for the first 1000 iterations of μ_1, μ_2, τ_1, τ_2 and p for the lognormal mixture model

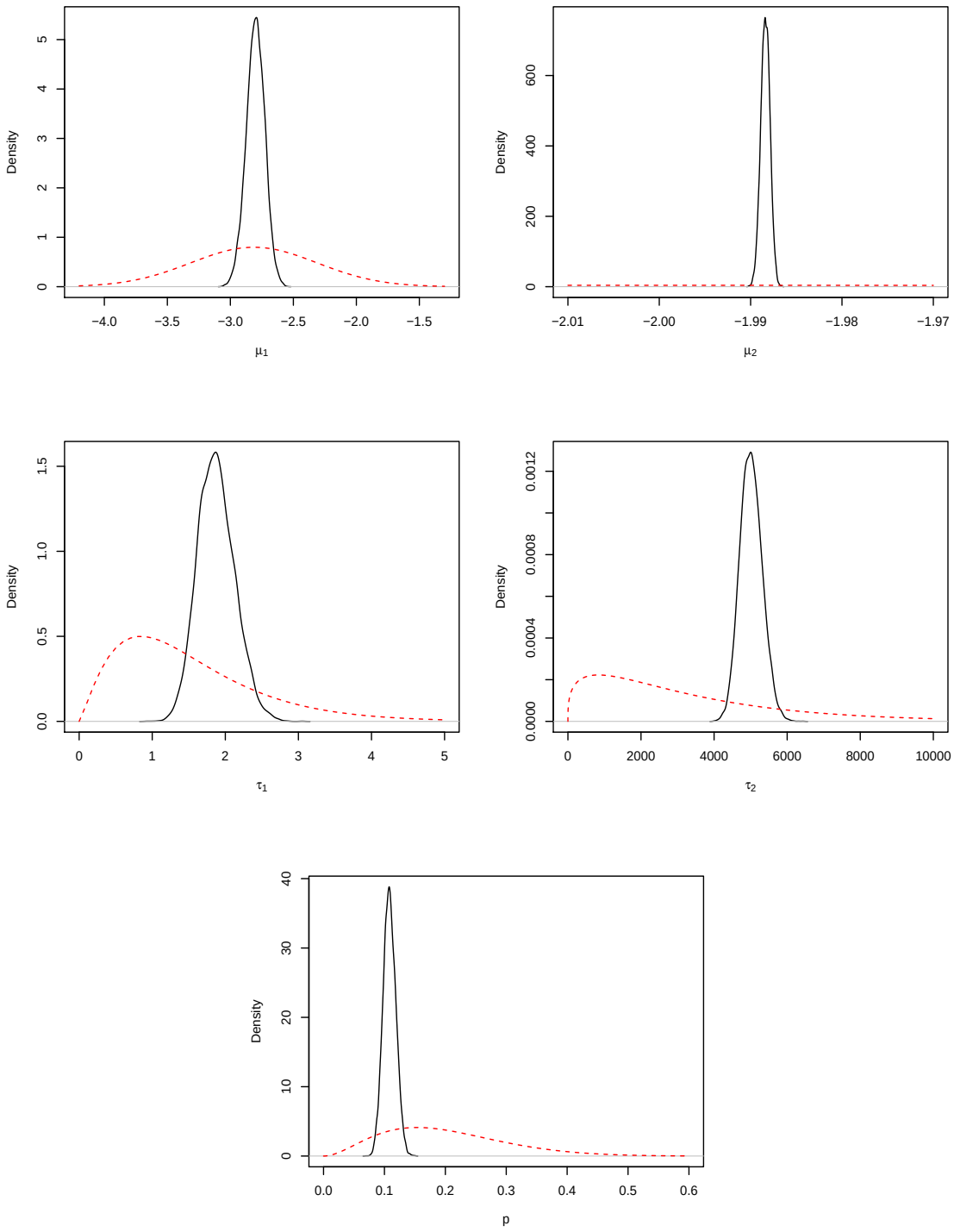


Figure 3.5: The prior (red dashed) and posterior (black solid) densities of μ_1, μ_2, τ_1, τ_2 and p for the lognormal mixture model

Suppose that we imagine a large sample $Y_{1j}, \cdots, Y_{nj}$ from component j, where n is large enough so that the sample mean $\bar{Y}_j = \frac{1}{n}\sum_{i=1}^{n} Y_{ij}$ is approximately $\lambda_j = \alpha_j/\beta_j$ and the sample variance $S^2_{Yj} = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_{ij} - \bar{Y}_j\right)^2$ is approximately $\mathrm{Var}\left(Y_{ij}|\alpha_j, \beta_j\right) = \alpha_j/\beta_j^2$. The sample coefficient of variation $C_j = S_{Yj}/\bar{Y}_j$ is then approximately $\alpha_j^{-1/2}$.

Suppose that we judge that our prior median for $\bar{Y}_j$ is L_j and our prior lower and upper quartiles for $\bar{Y}_j$ are respectively L_j/K_j and K_jL_j. To obtain these point values, we can ask the expert as follows:

Q9 *"Please think about the sample mean $\bar{Y}_j$ from component j. Please give a value such that $\bar{Y}_j$ is equally likely to be less than or greater than this value."* Let the given value be L_j.

Q10 *"Still thinking about the sample mean $\bar{Y}_j$ from component j. Please provide a value K_j such that the events $\bar{Y}_j < L_j/K_j$, $L_j/K_j < \bar{Y}_j < L_j$, $L_j < \bar{Y}_j < K_jL_j$ and $K_jL_j < \bar{Y}_j$ are all equally likely."*

Then, our prior median and lower and upper quartiles for μ_j are $\log L_j$, $\log L_j - \log K_j$ and $\log L_j + \log K_j$, respectively. Our prior mean for μ_j is therefore $\log L_j$ and our prior standard deviation for μ_j is $\log K_j/0.6745$.

Suppose that we judge that our prior quartiles for C_j are Q_{C1j}, Q_{C2j} and Q_{C3j}. We can obtain these point values from the expert by adapting Questions **Q4**, **Q5**, and **Q6** in Section 3.5.2.1.1. Then our prior quartiles for α_j are Q^{-2}_{C3j}, Q^{-2}_{C2j} and Q^{-2}_{C1j}. Then, α_j has a gamma $Ga\left(g_{\alpha j}, h_{\alpha j}\right)$ distribution where $g_{\alpha j}$ is chosen by iteratively solving $\left(Q_{C3j}/Q_{C1j}\right)^2 = R_3\left(g_{\alpha j}\right)/R_1\left(g_{\alpha j}\right)$, where $R_q(a)$ is quartile q for a $Ga(a,1)$ distribution, and $h_{\alpha j} = R_2\left(g_{\alpha j}\right)Q^2_{C2j}$. Using this method we have reasonable values for illustration as follows:

$$L_1 = 0.06 \qquad K_1 = 1.5$$
$$L_2 = 0.135 \qquad K_2 = 1.05$$

$$Q_{C11} = 0.6 \qquad Q_{C21} = 0.7 \qquad Q_{C31} = 0.8$$
$$Q_{C12} = 0.013 \qquad Q_{C22} = 0.015 \qquad Q_{C32} = 0.018$$

and hence

$$\mu_{01} = -2.81 \qquad v_{01} = 0.36$$
$$\mu_{02} = -2 \qquad v_{02} = 0.005$$
$$g_{\alpha 1} = 5.77 \qquad h_{\alpha 1} = 2.67$$
$$g_{\alpha 2} = 4.57 \qquad h_{\alpha 2} = 0.001$$

The prior specification for the probability of component 1 is still the same as in the lognormal mixture model.

3.5.2.2.2 Posterior distribution The form of the joint posterior distribution was given in Section 3.5.2.2. We used a similar approach as for the lognormal mixture model by using an MCMC scheme to draw posterior samples for the unknown parameters. Here, the MCMC chain is run for 5000 iterations as burn-in and then a further 10000 iterations were obtained as the posterior samples. Similar to the lognormal mixture model, the computing time that was used to obtain 10000 posterior draws is around 1.3 minutes using the same software and machine. Based on the trace plots in Figure 3.6, the chain converges very well and the mixing appears to be very satisfactory for all parameters. The prior and posterior densities of the unknown parameters are shown in Figure 3.7. The full summary of the posterior distributions for the unknown parameters is given in Table 3.2. In addition, Figure 3.8 shows the comparison of posterior density of the mean,

Parameter	Prior mean	Prior SD	Posterior mean	Posterior SD
α_1	2.16	0.9	2.4490	0.2951
α_2	4570	2137.76	5099.9498	302.7
μ_1	-2.81	0.6	-2.5733	0.0638
μ_2	-2	0.07	-1.9882	0.0005
π	0.17	0.05	0.1102	0.0105

Table 3.2: The prior and posterior means with standard deviations (SD) of the unknown parameters for the gamma mixture model

λ_j for component 1 and 2 for both models. From this figure, there are some differences between the lognormal and gamma mixture models for component 1 but there are no large differences for component 2. These differences might be attributed to the different shape of the distributions. In component 2, where α_2 is large, there is little difference in shape.

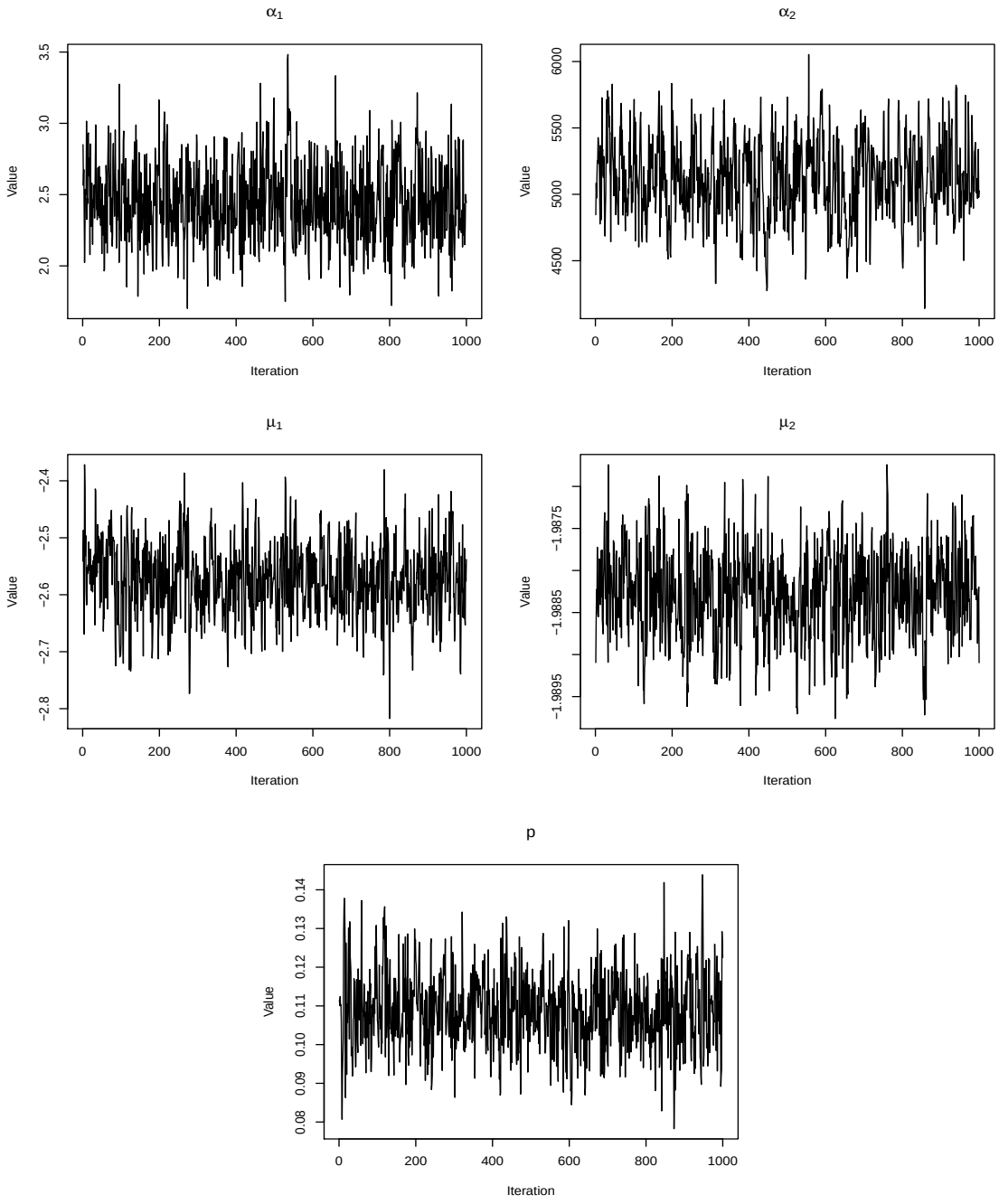


Figure 3.6: The trace plots for the first 1000 iterations of α_1, α_2, μ_1, μ_2 and p for the gamma mixture model

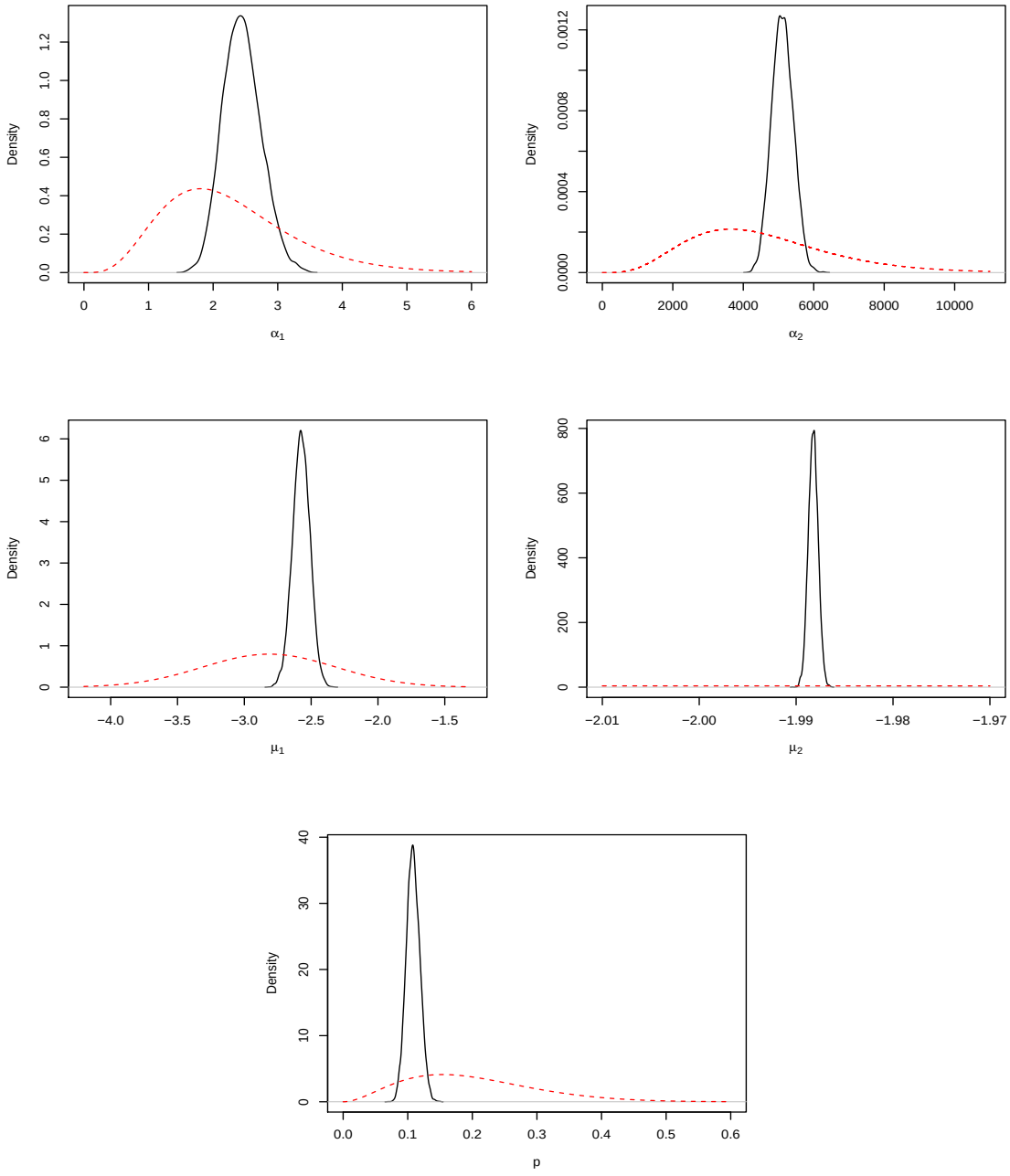


Figure 3.7: The prior (red dashed) and posterior (black solid) densities of α_1, α_2, μ_1, μ_2 and p for the gamma mixture model

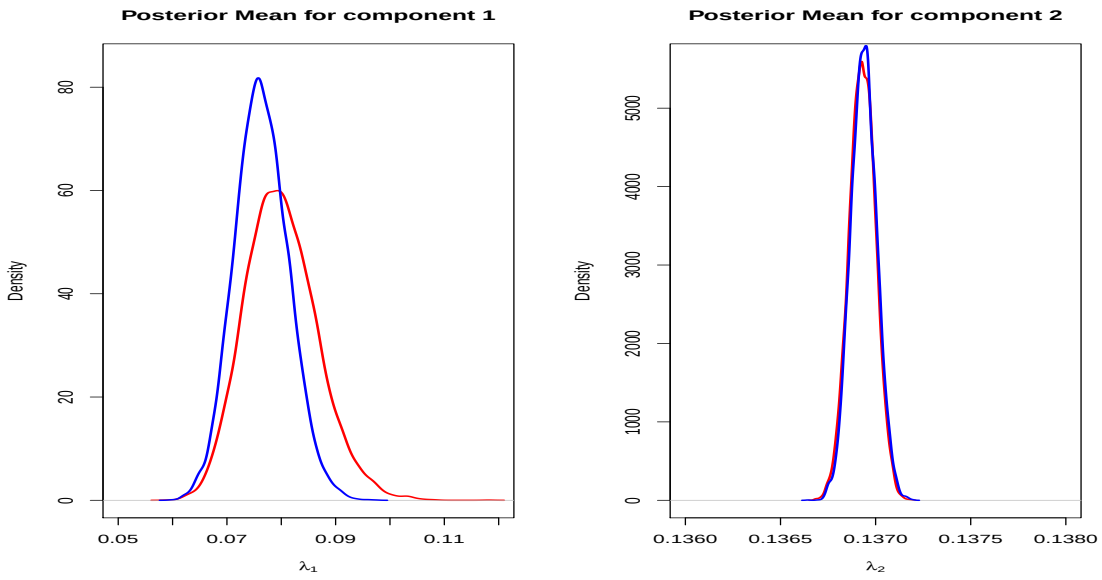


Figure 3.8: The posterior density of λ_j for component 1 and component 2 with lognormal (red) and gamma (blue) models

3.6 Summary

We have described simple mixture models in this chapter with a focus on finite mixture models. This model is usually used when the data are assumed to have two or more subpopulations, or we want to give greater flexibility in the sampling model and the shape of the distribution. We showed how data augmentation can be used to simplify the likelihood and computations. We also introduced the MCMC scheme for the finite mixture models with further discussion on the label switching problem.

In most cases, the data are either discrete or continuous. However, there is a special case in mixture models where the data contain both discrete and continuous components and these data hence follow a mixed distribution. An important special case is when the discrete component of the mixed distribution represents the zero values whereas the continuous component corresponds to the positive values of observations. An example of this case is daily rainfall amounts which we will further expound in the next chapter. Next, we applied two different mixture models to ultrasound data: the lognormal mixture and the gamma mixture. The detailed parameterisation of the models and prior specification for each parameter were discussed in this application. The RJAGS package (Plummer, 2012) was used to run the MCMC scheme and generate posterior samples for the parameters. The posterior summaries for the parameters were then provided.

Chapter 4

Modelling Univariate Daily Rainfall Data

4.1 Introduction

The literature of time series modelling is immense and involves many approaches and parameterisations. In some circumstances, the time series data consist of positive and zero values. This type of data is quite challenging since it requires models for both discrete and continuous components. A distribution of this type is known as a mixed distribution. This frequently materializes in many application areas such as meteorology, ecology, economics and environmental sciences. A well-known example is the daily rainfall whose values are either positive if rain occurs or zero if rain is absent.

This chapter introduces the idea of modelling daily rainfall data within the Bayesian framework. In this chapter, we only fit the daily rainfall data for one site with two different applications before extending it to multiple locations in the next chapter. Section 4.2 outlines the basic model for modelling mixed distributions. In this section, we will give detailed descriptions on how to model the amount and occurrence processes. This method will be subsequently applied to daily rainfall data in Italy in Section 4.3. The prior beliefs will be incorporated with the likelihood of the data to obtain the posterior distribution using MCMC. In addition, we also endeavour to determine and propose the best probability density function (pdf) to represent the amount process. To accomplish this, three pdfs will be considered to fit the daily rainfall data.

Next, in Section 4.4, we will extend the rainfall model developed in Section 4.3 by incorporating the atmospheric circulation pattern, which is also known as the Lamb weather type (LWT), into the British daily rainfall model. The LWT is included in the model by allowing the rainfall occurrence probabilities and the mean of the amount distribution to

be dependent upon the LWT. This will result in a more complex and challenging model construction. Relating the occurrence probabilities and the amounts directly to the LWT is in contrast to the approach of Germain (2010) and Heaps *et al.* (2015) who introduced a latent "weather state" between the LWT and the observed rainfall. In addition Germain (2010) and Heaps *et al.* (2015) considered only winter rainfall and did not include seasonal effects in the model. Seasonal effects will be introduced in Section 4.2.4. MCMC techniques will be used to evaluate the posterior distribution for each model parameter. This will be discussed in detail in each section.

4.2 Modelling Daily Rainfall : The Basic Model

4.2.1 General structure

In this section, we shall describe the construction of a general daily rainfall model for a single site. This work aims to provide the methodology required for modelling daily rainfall within the Bayesian framework before this can be extended to multiple sites. Modelling daily rainfall data is a complex process since the distribution is a mixture of discrete and continuous components. Daily rainfall models commonly distinguish between the discrete component which represents rainfall occurrence and the continuous counterpart which corresponds to the rainfall amount when rain occurs. Hence, the daily rainfall can be modelled separately in two parts: the occurrence process and the amount process. The occurrence process is a process that models the probability of rainfall occurrences which means that the data has a zero value when rain is absent and a positive value when rain occurs. On the other hand, the amount process is a process that models the amount of rainfall which occurs during a rainy day. It is hence important to represent the rainfall amount with an appropriate distribution that only allows non-negative values.

Over the past few decades, most research using mixed distributions has advocated the use of a two-stage approach. Stern & Coe (1984) demonstrated that the utilisation of a two-stage approach in a daily rainfall model is relatively straightforward for hydrology applications. Tooze *et al.* (2002) also employed the same approach for the analysis of medical expenditure in the United States. Both examples have demonstrated that the two-stage approach is an appropriate way to model mixed distributions.

Generalized Linear Models (GLMs) have also been employed to model a daily rainfall. The GLMs are an effective class of probability models which can accommodate various types of data such as climatological and meteorological data. The response variable for the GLM approach generally does not necessarily need to be Gaussian, as long as it comes from one of the members of the exponential family. Therefore, GLMs can handle non-normal

responses using well defined characteristics for modelling the continuous and strictly positive amount process and the discrete component of occurrence process. Basically, the response variable is linked to the linear predictor through a specific link function. The early works by Coe & Stern (1982) and Stern & Coe (1984) have shown that it is quite straightforward to model daily rainfall using GLMs in a hydrology context. They used a simple linear regression with a Fourier series function as the covariate for both amount and occurrence processes. Grunwald & Jones (2000) utilised a similar approach to this in an Australian daily rainfall model by incorporating more complex covariates in the GLM model. Moreover, recent literature in rainfall modelling indicates that the covariates in the GLM may also include sophisticated weather variables such as temperature, wind speed, and atmospheric circulation pattern (Chandler & Wheater, 2002; Furrer & Katz, 2007). The GLM approach can also be extended to include spatio-temporal models. Chandler & Wheater (2002), Yang *et al.* (2005) and Fernandes *et al.* (2009) successfully proposed a GLM-based framework for spatiotemporal structure in daily rainfall models, an idea which we will use in the next chapter. However, the majority of this literature in rainfall modelling does not use the Bayesian approach. Therefore, we want to contribute to this small literature and utilise the ideas from Coe & Stern (1982), Stern & Coe (1984) and Grunwald & Jones (2000) by using a Fourier series function as a covariate for both the amount and occurrence processes within the Bayesian framework. Particular attention is also given to the relationship between these two processes. For the case of British daily rainfall, we will extend the model by incorporating Lamb weather types (LWTs) directly to the amount and occurrence processes.

Let $W_t \in \mathbb{R}^+$ be a random variable for the daily rainfall amount with an observed value, w_t at a single site at time t (measured in days), where $t = 1, \ldots, T$. Suppose that each observation in the daily rainfall data is generated from a random process where the distribution of W_t, given quantities p_t and μ_t (where $0 \leq p_t \leq 1$), can be defined as follows:

$$F_w(w_t|p_t, \mu_t) = \Pr\left(W_t \leq w_t|p_t, \mu_t\right) = \begin{cases} 0, & (w_t < 0) \\ 1 - p_t, & (w_t = 0) \\ p_t F_A(w_t|\mu_t), & (w_t > 0) \end{cases}$$

where $F_A(w_t|\mu_t)$ is the distribution function of the amount distribution. Hence $\Pr(W_t = 0) = 1 - p_t$. Given that $W_t > 0$, the conditional pdf of W_t is $f(w_t|\mu_t)$. Let R_t represent the rainfall occurrence and serve as an indicator function for W_t. Then we have:

$$R_t = \begin{cases} 0, & W_t = 0 \\ 1, & W_t > 0. \end{cases} \tag{4.1}$$

Thus, the random variable W_t can be re-expressed as:

$$\begin{aligned} W_t &= I(W_t > 0)Y_t \\ &= R_t Y_t \end{aligned} \tag{4.2}$$

where $Y_t = g(Z_t)$ is a continuous random variable and $g(.)$ is some monotonic function defining a suitable transformation (e.g. exponential) together with the transformed value, Z_t. Here W_t is the actual rainfall amount whose value could be zero but it is always observed. According to Stern & Coe (1984) and Grunwald & Jones (2000), Y_t can be regarded as the intensity process which can also be viewed as the potential rainfall amount. The value of Y_t is always positive but not always observed.

In order to describe the serial structure of the daily rainfall, it is necessary to consider a model that can accommodate the relationship between the amount and occurrence processes. Stern & Coe (1984) does not explicitly take into account any relationship between the probability of rainfall and the rainfall amount. It is hence one of our primary objectives to relate these two processes in our modelling strategies so that no important information about the rainfall amount and occurrence is lost. Our goal is further emphasized by Tooze *et al.* (2002) who stressed that it is critical to elucidate the relationship between the probability and the level of nonzero observation so that the accuracy and adequacy of analysis might be improved. There are several approaches that we can employ to expound the relationship between the rainfall amount and occurrence. In the first approach, we can fit the occurrence process and then use that information to analyse the amount process as follows:

$$\Pr(R_t = 1) = p_t \quad \text{and} \quad f(w_t|R_t = 1) \tag{4.3}$$

where $f(w_t|R_t = 1)$ is the conditional pdf for the rainfall amount when $r_t = 1$. Alternatively, we can evaluate the model for the potential rainfall amount first, and then, compute the rainfall probability as:

$$f(y_t) \quad \text{and} \quad \Pr(R_t = 1|Y_t = y_t) = \hat{h}(y_t) \tag{4.4}$$

where $\hat{h}(y)$ is some function for the rainfall probability which incorporates the information contained in the amount process. A third possibility is that we can use a property of the distribution, such as the mean, μ, as a covariate of the rainfall probability to link between the amount and occurrence processes.

Heaps *et al.* (2015) demonstrated the first approach by introducing two normal variables, $\{Z_0, Z_1\}$ where

$$W_t = I(Z_{0,t} > 0)\exp(Z_{1,t}) = R_t \exp(Z_{1,t})$$

with $Z_{0,t}$ as the occurrence function and $Z_{1,t}$ as the amount function. In this case, $f(w_t|R_t = 1)$ is a lognormal density and the authors included the $Z_{0,t}$ as the covariate of $Z_{1,t}$ for the log-rainfall amount. On the other hand, Sofia (2007) utilised the third approach by conditioning the occurrence process on the mean of the log potential rainfall amount, ϑ_t using logistic regression for monthly rainfall data. Thus, the probability of rainfall is given by

$$\Pr(W_t > 0|\vartheta_t) = p_t = \frac{\exp(\zeta_0 + \zeta_1\vartheta_t)}{1 + \exp(\zeta_0 + \zeta_1\vartheta_t)}.$$

In this case, R_t and Y_t are conditionally independent given ϑ_t so $f(w_t|R_t = 1, \vartheta_t) = f(y_t|\vartheta_t)$. For example, we can have $Z_t = \log(Y_t)$ and, given ϑ_t, Z_t has a normal distribution with mean ϑ_t and Y_t has a lognormal distribution.

While Germain (2010) and Heaps *et al.* (2015) used the first approach, in this thesis we propose to investigate the other two approaches as alternative models for daily rainfall, within the Bayesian framework. In particular, we will use the third approach, which was used by Sofia (2007), for the Italian daily rainfall and the second approach for the British daily rainfall. We will also investigate the daily rainfall data over the whole year in our model instead of just focusing on the daily rainfall data in the winter as in the case of Germain (2010) and Heaps *et al.* (2015), or monthly data, as in the case of Sofia (2007). Figure 4.1 illustrates the example of a directed acyclic graph (DAG) for the general structure of the univariate daily rainfall model. This model is the first proposal that we will use for the univariate daily rainfall in the context of the Italian data. In this model, $\cdots, Y_{t-1}, Y_t, Y_{t+1}, \cdots$ are assumed to be independent and only depend on μ_t which changes over time. It means that Y_t variables are independent given the model parameters and not just conditionally independent given R_t. The detail of this model will be discussed in Section 4.3.2 and we will also examine the assumption of conditional independence in the amount process, through analysis of residuals in Section 4.3.3.5. There are other possibilities which we might consider, such as direct dependence between $\cdots, Y_{t-1}, Y_t, Y_{t+1}, \cdots$ as well as $\cdots, R_{t-1}, R_t, R_{t+1}, \cdots$ but with a different context of rainfall data.

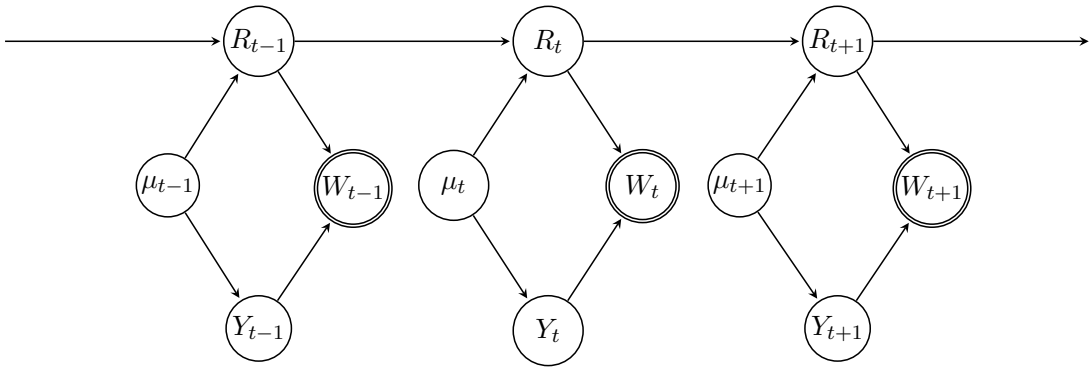


Figure 4.1: A DAG showing the temporal dependence structure of the Italian daily rainfall model

4.2.2 Amount Process

The ideal distribution for the amount process has been a topic of interest in the past few decades, especially in modelling daily rainfall. It is thus vital to choose the right distribution to model the rainfall amount. Several studies have been conducted to identify the best distribution for daily rainfall. However, there is no general consensus on which probability density function is the best to represent rainfall amount (Wilks, 1990, Kedem et al 1990, Cho et al 2004). The most commonly utilised distributions for the amount process are gamma (Katz, 1977; Stern & Coe, 1984; Grunwald & Jones, 2000), lognormal (Tooze *et al.*, 2002; Heaps *et al.*, 2015) and power transformed truncated normal distributions (Sanso & Guenni, 1999). Other distributions that have been employed for the modelling of daily rainfall data are Weibull (Tooze *et al.*, 2002), exponential (Fernandes *et al.*, 2009), and mixed exponential distributions (Wilks, 1999).

For the amount process, $f(y_t|\mu_t, \theta)$ is the pdf for the potential rainfall amount with a range between 0 and ∞. If Y_t and R_t are conditionally independent given μ_t then $f(w_t|R_t = 1, \mu_t)$ is equivalent to $f(y_t|\mu_t, \theta)$ except for dry days, where $W_t = 0$. The distribution for Y_t still remains identical whether the day is wet or dry. However, Y_t is not observed if $W_t = 0$. For example, Y_t could have a lognormal distribution and then, if $Y_t = \exp(Z_t)$, Z_t has a normal distribution. We can then propose to utilise the GLM approach for fitting the mean of the amount distribution as:

$$\begin{aligned} \hat{g}(\mu_t) &= \hat{\eta}_t \\ \hat{\eta}_t &= \boldsymbol{x}'_t \boldsymbol{\eta} \end{aligned} \tag{4.5}$$

where $\hat{g}(.)$ is the link function, $\hat{\eta}_t$ is the linear predictor, $\boldsymbol{\eta}$ is a vector of coefficients and $\boldsymbol{x}_t$ is the covariate vector for observation t that may consist of temporal covariates, seasonal and trend components, and weather variables (e.g. the Lamb weather type). For this process, Y_t and Y_s where $s \neq t$ could be conditionally independent given μ_t, μ_s and the shape of the distribution could also change when one of the parameters varies over time t. For this reason, it is useful to illustrate and monitor these changes using zero-continuous plots (as in Figure 3.2) as the shape could shift from time to time. This will be explained in more detail using the Italian daily rainfall example.

4.2.3 Occurrence Process

The proportion of wet and dry days can be described by modelling the occurrence process. Let $\boldsymbol{R} = \{R_1, R_2, \cdots, R_t\}$ denote the sequence of daily rainfall occurrence with $t = 1, 2, \cdots, T$. The conditional distribution of the occurrence process given p_t is a

Bernoulli distribution:

$$R_t \sim Bern(p_t).$$

In the literature on rainfall models, the vast majority of studies have utilised the first-order Markov chain to model the occurrence probability and this was performed by assuming that the conditional probability of rain occurrence on a specific day only depends on the occurrence of rainfall on the previous day (Gabriel & Neumann, 1962; Coe & Stern, 1982; Stern & Coe, 1984; Suhaila *et al.*, 2011; Grunwald & Jones, 2000). For instance, Gabriel & Neumann (1962) successfully demonstrated that the sequence of daily rainfall occurrence in Tel Aviv, Israel could be modelled by using the first-order Markov chain. The Markov chain can also be extended to higher-order models. However, Jimoh & Webster (1996) found that a first-order Markov chain model is sufficient to predict the daily rainfall occurrence. In our case, it is thus reasonable to use the first-order Markov chain for the occurrence process.

Similar to the amount process, we may also use a GLM approach, considering that the Bernoulli distribution is a part of the exponential family. Following the GLM framework, the probability of rainfall occurrence is governed by the link function. Practically, let

$$\begin{aligned} \hat{h}(p_t) &= \hat{\zeta}_t \\ \hat{\zeta}_t &= \boldsymbol{x}'_t \boldsymbol{\zeta} \end{aligned} \tag{4.6}$$

where $\hat{h}(.)$ is the link function, $\hat{\zeta}_t$ is the linear predictor, $\boldsymbol{\zeta}$ is a vector of coefficients and $\boldsymbol{x}_t$ is the covariate vector. Without loss of generality, this is the same covariate vector as for the amount process. There are several link functions that can be used to relate the observed variable to the linear predictor. Most research has emphasized the use of the logit link function to transform p_t from a $(0, 1)$ scale to a $(-\infty, +\infty)$ scale (Stern & Coe, 1984; Grunwald & Jones, 2000; Lima & Lall, 2009). By using a logistic transformation, the relationship between the linear predictor, $\hat{\zeta}_t$, and transition probabilities of rainfall occurrence, p_t, can be expressed as

$$\hat{\zeta}_t = \log\left(\frac{p_t}{1-p_t}\right) = \text{logit}\left\{p_t\right\} \tag{4.7}$$

with an inverse transformation of

$$p_t = \frac{\exp(\hat{\zeta}_t)}{1+\exp\left(\hat{\zeta}_t\right)}. \tag{4.8}$$

From this relationship, the value of $\hat{\zeta}_t \to \infty$ if $p_t \to 1$ and if $p_t \to 0$ then $\hat{\zeta}_t \to -\infty$. Alternatively, we can employ other link functions such as the probit link function which

is denoted by:

$$\hat{\zeta}_t = \Phi^{-1}\{p_t\},$$

where Φ^{-1} is the inverse of the standard normal distribution function, or the complementary log-log link:

$$\hat{\zeta}_t = \log\left[-\log\{1 - p_t\}\right].$$

Since Heaps *et al.* (2015) used the first approach for linking the occurence and amount processes and used a lognormal amount distribution, in effect they used a probit link. Other authors in the rainfall modelling literature, such as Gabriel & Neumann (1962), Coe & Stern (1982), Stern & Coe (1984), Grunwald & Jones (2000), Sofia (2007) and Fernandes *et al.* (2009), have used a logit link. We have chosen to use a logit link in common with these authors and to provide a contrast to Heaps et al. (2015). Although the shapes of the two link functions are quite similar, the use of the logit link may provide an advantage in terms of computational speed since we do not have to evaluate a normal cumulative distribution function or its inverse.

4.2.4 Seasonal Effect

One of the novelties that are proposed in this study is dealing with seasonal effects, especially when the rainfall data is expected to have a cyclical pattern. In temperate and subpolar regions such as the European countries, daily rainfall occurrence may be affected by seasonal changes where the mean rainfall amount and the frequency of occurrence might be different, for example, during the winter and the summer periods. The changes of the annual rainfall cycle are important in agricultural planning especially for determining the optimal time for planting crops and for estimating the crop yields. Hence, it is reasonable to incorporate seasonal effect to evaluate the nature of rainfall variability for countries in the temperate region. In the Heaps *et al.* (2015) model, only winter rainfall data were used and the model did not incorporate seasonal effects.

The most popular method that is extensively used in meteorological studies for modelling periodic time series is a truncated Fourier series. The earlier work by Jones & Brelsford (1967) and West & Harrison (1997) have highlighted that the periodic structure in time series can be modelled by sinusoidal representations. Furthermore, Stern & Coe (1984) have adopted the Fourier series to evaluate daily rainfall occurrence and amount for agriculture planning. Grunwald & Jones (2000) used the same approach to describe the seasonal pattern of the occurrence and amount processes for the Australian daily rainfall data.

The Fourier series consists of sine and cosine terms of harmonic frequencies with a

period ω and this can be represented as:

$$\hat{F}(t) \quad = \sum_{f=1}^{F} \left[a_f \cos\left(\omega_f t\right) + b_f \sin\left(\omega_f t\right) \right] \tag{4.9}$$

where F denotes the number of sinusoids and a_f and b_f represent the Fourier coefficients with $\omega_f = 2\pi f/L$, the angular frequency where L is the period. For daily data that exhibit an annual cyclic pattern, $L = 365.25$. This Fourier series representation can be further modified by performing further substitutions for a_f and b_f. Let φ_f be a phase of the Fourier function, then a_f and b_f are given by:

$$\begin{aligned} a_f &= A_f \cos(\varphi_f) \\ b_f &= -A_f \sin(\varphi_f). \end{aligned}$$

Then, we can define the amplitude, A and the phase, φ_f as

$$A_f = \sqrt{a_f^2 + b_f^2}$$

and

$$\varphi_f = -\arctan\left(b_f/a_f\right).$$

Using this information, we can reparameterise the Fourier function in equation (4.9) in the following form:

$$\hat{F}(t) \quad = \textstyle\sum_{f=1}^{F} A_f \cos\left(\omega_f t + \varphi_f\right). \tag{4.10}$$

Now, this alternative Fourier series representation is easier for interpretation than the original Fourier series form.

To fit a Fourier function in the model, we need to determine the optimal number of harmonics to be used for this purpose. We can opt for any reasonable number of Fourier harmonics, up to a maximum of 182. However, previous studies have advocated that the first three harmonics are generally adequate to model seasonal effects on rainfall (Liu *et al.*, 2011; Roldán & Woolhiser, 1982; Richardson, 1981). We will therefore use this recommendation when incorporating the seasonal effects into our models. We may employ both Fourier series forms to model the amount and occurrence processes for the daily rainfall model that will also allow for seasonal variability.

4.2.5 Prior and Posterior Distributions

To complete the model, we need to specify the prior distribution for the parameters. Let $\boldsymbol{\theta}$ be the unknown parameter vector. If we assume prior independence between the parameters for the amount and occurrence processes, the joint prior distribution is hence given by:

$$\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}_{amt}) \times \pi(\boldsymbol{\theta}_{occ})$$

where $\pi(\boldsymbol{\theta}_{amt})$ and $\pi(\boldsymbol{\theta}_{occ})$ are the prior distributions for the amount and occurrence processes. Practically, prior selections for the unknown parameters should be based on our prior beliefs, the information gained from the expert or from the previous studies. For instance, normal prior distributions are often assigned to the regression coefficients $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$ since $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$ can take any value within the range of $-\infty$ to ∞. Apart from the requirement that prior specifications should reflect reasonable prior beliefs, inappropriate or careless specifications might also adversely affect, for example, the convergence of MCMC algorithms.

Based on Bayes theorem, the posterior distribution of $\boldsymbol{\theta}$ is proportional to likelihood $\times$ prior. Thus, the posterior distribution of $\boldsymbol{\theta}$ can be written as:

$$\pi(\boldsymbol{\theta}|\boldsymbol{w}, \boldsymbol{r}) \propto f(\boldsymbol{w}|\boldsymbol{\theta}) \times f(\boldsymbol{r}|\boldsymbol{w}, \boldsymbol{\theta}) \times \pi(\boldsymbol{\theta}). \tag{4.11}$$

To obtain samples from $\pi(\boldsymbol{\theta}|\boldsymbol{w}, \boldsymbol{r})$, we can use an MCMC scheme. The full conditional distribution (FCD) for each unknown parameter can be easily derived from Equation (4.11). If the FCD is in a standard form, we can draw samples directly from it. If it does not correspond to any standard density form, we need to use the Metropolis-Hastings (MH) algorithm to obtain samples. Thus, we use a Metropolis-within-Gibbs algorithm.

4.2.5.1 Priors in truncated Fourier representation of seasonality

In this section, we will describe prior distributions for Fourier series coefficients. The prior distribution for Fourier parameters in Equation (4.9) is difficult to interpret. However, we can think more directly about the prior distribution in terms of the phase and amplitude as in Equation (4.10). Suppose that we have $(a_f, b_f)' \sim N_2(\boldsymbol{m}_f, V_f)$ where $\boldsymbol{m}_f = (m_{f,a}, m_{f,b})'$ and

$$V_f = \begin{pmatrix} v_{f,a}^2 & \rho_f v_{f,a} v_{f,b} \\ \rho_f v_{f,a} v_{f,a} & v_{f,b}^2 \end{pmatrix}.$$

We can derive the prior distribution for A_f and φ_f using the fact that $A_f = \sqrt{a_f^2 + b_f^2}$ and $\varphi_f = -\arctan(b_f/a_f) \in [-\pi, \pi)$. The Jacobian of the transformation is given by

$$\left(\frac{\partial(a_f, b_f)}{\partial(A_f, \varphi_f)}\right) = \begin{bmatrix} \frac{\partial a_f}{\partial A_f} & \frac{\partial b_f}{\partial A_f} \\ \frac{\partial a_f}{\partial \varphi_f} & \frac{\partial b_f}{\partial \varphi_f} \end{bmatrix} = \begin{bmatrix} \cos(\varphi_f) & -\sin(\varphi_f) \\ -A_f \sin(\varphi_f) & -A_f \cos(\varphi_f) \end{bmatrix}$$

and hence

$$\left|\det\left(\frac{\partial(a_f, b_f)}{\partial(A_f, \varphi_f)}\right)\right| = \left|-A_f \cos^2(\varphi_f) - A_f \sin^2(\varphi_f)\right| = |-A_f| = A_f.$$

The density for $(A_f, \varphi_f)'$ is therefore

$$\begin{aligned} f(A_f, \varphi_f) =& f(a_f, b_f) \left|\det\left(\frac{\partial(a_f, b_f)}{\partial(A_f, \varphi_f)}\right)\right| \\ & \frac{A_f}{2\pi v_{f,a} v_{f,a} \sqrt{1-\rho_f^2}} \exp\left\{-\frac{Q(A_f, \varphi_f)}{2(1-\rho_f^2)}\right\} \end{aligned}$$

where

$$\begin{aligned} Q(A_f, \varphi_f) =& \frac{(A_f \sin(\varphi_f) - m_{f,a})^2}{v_{f,a}^2} - \frac{2\rho_f (A_f \sin(\varphi_f) - m_{f,a})(A_f \cos(\varphi_f) - m_{f,b})}{v_{f,a} v_{f,b}} \\ & + \frac{(A_f \cos(\varphi_f) - m_{f,b})^2}{v_{f,b}^2} \end{aligned}$$

for $A_f \geq 0$ and $\varphi \in [-\pi, \pi)$. Assuming $m_{f,a} = m_{f,b} = 0$ and $v_{f,a} = v_{f,b} = \tilde{v}_f$, the density simplifies to

$$f(A_f, \varphi_f) = \frac{A_f}{2\pi \tilde{v}_f^2 \sqrt{1-\rho_f^2}} \exp\left\{-\frac{A_f^2}{2\tilde{v}_f^2(1-\rho_f^2)} \left\{1 - \rho_f \sin(2\varphi_f)\right\}.\right\} \tag{4.12}$$

This density can be further simplified by assuming $\rho_f = 0$ which gives

$$f(A_f, \varphi_f) = \frac{A_f}{\tilde{v}_f^2} \exp\left\{-\frac{A_f^2}{2\tilde{v}_f^2}.\right\} \times \frac{1}{2\pi}.$$

In this case, it is clear that A_f and φ_f are independent and

$$A_f \sim Ray(\tilde{v}_f) \text{ and } \varphi \sim U(-\pi, \pi)$$

where $Ray(\tilde{v}_f)$ denotes the Rayleigh distribution with scale parameter, $\tilde{v}_f$.

Provided that we wish to give the phase a uniform prior, this provides a convenient way to convert a prior for the amplitude into a bivariate normal prior for the Fourier coefficients. This may be appropriate for the higher frequencies. In the case of the fundamental frequency, with period one year, and perhaps the second frequency, we may well have prior information about the annual pattern of rainfall which we can use. In this case we can give the phase a more informative prior, for example a von Mises distribution. In this case the distribution of the Fourier coefficients will not be bivariate normal. The von Mises distribution is not a standard distribution in JAGS. While, of course, we need not use JAGS and could use a specially-written sampler, in principle we can also overcome the difficulty by using, for example, the "ones trick" (Lunn *et al.*, 2012, Chapter 9) in JAGS.

However a more direct approach is as follows. Suppose that we wish to use F harmonics. We can choose $2F$ time-points during the year and elicit beliefs about the value of the seasonal term at these times. The times might include, for example the times which are expected to give the maximum and minimum of the seasonal effect. Let the times be $t_1, \ldots, t_{2F}$. Then, using elicitation questions in which other model quantities are fixed as appropriate, we can first elicit marginal means and variances for $\hat{F}_1, \ldots, \hat{F}_{2F}$ where $\hat{F}_j$ is the value of the seasonal effect at time t_j. So

$$\hat{F}_j = \sum_{f=1}^{F} [a_f \cos(\omega_f t_j) + b_f \sin(\omega_f t_j)].$$

We choose a multivariate normal prior distribution for $\hat{\mathbf{F}} = (\hat{F}_1, \ldots, \hat{F}_{2F})'$ and assume that this will adequately represent prior beliefs. It remains to specify covariances between the elements of $\hat{\mathbf{F}}$. In the style of Farrow (2003) we write

$$\hat{F}_j - \mathrm{E}(\hat{F}_j) = k_j\{U_0 + U_1 \cos(\omega t_j) + U_2 \sin(\omega t_j) + E_j\}$$

where $\omega = 2\pi/365.25$, U_0, U_1, U_2 and $E_1, \ldots, E_{2F}$ are all independent, zero-mean random variables and $\mathrm{Var}(U_0) = V_0$, $\mathrm{Var}(U_1) = \mathrm{Var}(U_1) = V_1$ and $\mathrm{Var}(E_j) = V_E$. We constrain $V_0 + V_1 + V_E = 1$. Hence the marginal variance is $\mathrm{Var}(\hat{F}_j) = k_j^2$. The covariances are given by

$$\mathrm{Covar}(\hat{F}_j, \hat{F}_h) = V_0 + V_1(\cos\omega t_j \cos\omega t_h + \sin\omega t_j \sin\omega t_h) = V_0 + V_1 \cos(\omega[t_j - t_h]).$$

By choosing values for V_0, V_1 and V_E, we determine a covariance matrix V_F for $\hat{\mathbf{F}}$.

Now we can write

$$\hat{\mathbf{F}} = T\mathbf{c}$$

where $\mathbf{c} = (a_1, \ldots, a_F, b_1, \ldots, b_F)'$ and T is a $2F \times 2F$ matrix where column f is $(\cos\omega_f t_1, \ldots, \cos\omega_f t_F)'$

and column $f + F$ is $(\sin \omega_f t_1, \ldots, \sin \omega_f t_F)'$. Then $\mathbf{c} = T^{-1}\hat{\mathbf{F}}$ and the covariance matrix of $\mathbf{c}$ is $T^{-1}V_F(T^{-1})'$. Thus we obtain a multivariate normal prior distribution for the set of Fourier coefficients.

4.2.6 Diagnostic checking for mixed-distribution time series

One class of diagnostic checks for the non-standard time series models has been proposed by Smith (1985) to assess whether the fitted model is well calibrated. The author introduced two types of residuals for the diagnostic checking which are the "forecast distribution transformed residuals" (FDTR) and the normal-transformed version called "normal forecast transformed residuals" (NFTR). An important feature of the FDTR is that, when the distribution of the observations is continuous and if the observations are drawn from the forecast distribution, the FDTR are distributed uniformly $U(0, 1)$. If the observations were generated according to our model and we knew the true values of the parameters and used these values to calculate the forecast distribution, the FDTR would then be independent. Since we do not know the values of the parameters and our forecast distribution is a posterior predictive distribution, the FDTR will only be approximately independent.

In the case of rainfall, the distribution of the observations is not continuous since it is a mixture of zeros (discrete) and positive values (continuous). Therefore, some modifications are required to make the Smith (1985) method work. Smith (1985) already discusses in great detail how to deal with the case where the observation distribution is discrete. In our case, we need to do something similar and we therefore choose that, instead of working in terms of the observations on the random variable W, which has a mixed distribution, we define another random variable U, which has a continuous distribution, and work in terms of this. Of course U and W are closely related and, in fact, given the value of U, we can find the value of W. Thus, we propose to define U as follows:

- If $W > 0$ then $U = W$.
- If $W = 0$ then U is drawn from a uniform $U(-1, 0)$ distribution. In fact, we could use any continuous distribution on $(-\infty, 0]$ for this purpose but this uniform distribution is sufficiently convenient.

Hence, we simulate the values of U for the days with zero rainfall and use these as data. We then use the forecast distributions of U, rather than W, and compare the observed values of U with these distributions. The FDTR should then have a uniform $U(0, 1)$ distribution as required. NFTR can be obtained by using

$$V_U = \Phi^{-1}(U)$$

where Φ is the standard normal distribution function. If the fitted model is "good", then the residuals U and V_U should be approximately independent and follow the uniform $U(0,1)$ or standard normal, $N(0,1)$ distributions. If the residuals are unsatisfactory, then we need to revise our model to improve the forecasting system by using the deviations in $\{U\}$ and $\{V\}$ from their expected forms (Smith, 1985).

4.3 Daily Rainfall Model for Urbino, Italy

4.3.1 Data

Figure 4.2: Location of daily rainfall in the city of Urbino, Italy

To apply the developed model, we utilized daily rainfall data from the Urbino rain gauge station in Italy. Figure 4.2 shows the location of Urbino which is in the region of Marche, the central area of Italy. We used daily rainfall data recorded from 1981 until 2007 (27 years), a total of 9861 observations. These data have more zeros than positive values as indicated in Figure 4.3.

From Figure 4.3, we can see that the occurrence of rainfall is quite low, with inconsistent rainfall patterns for each year. On average, the wettest days are usually between around November and January, whilst the driest days took place in July and August. The highest recorded rainfall amount was on the 24^{th} of November 1991 with a precipitation of 113.8 mm. Fitting this type of rainfall data is considered very challenging since we have more zeros than positive values. Hence, this warrants a more flexible modelling approach. This will be discussed in more detail in the next section.

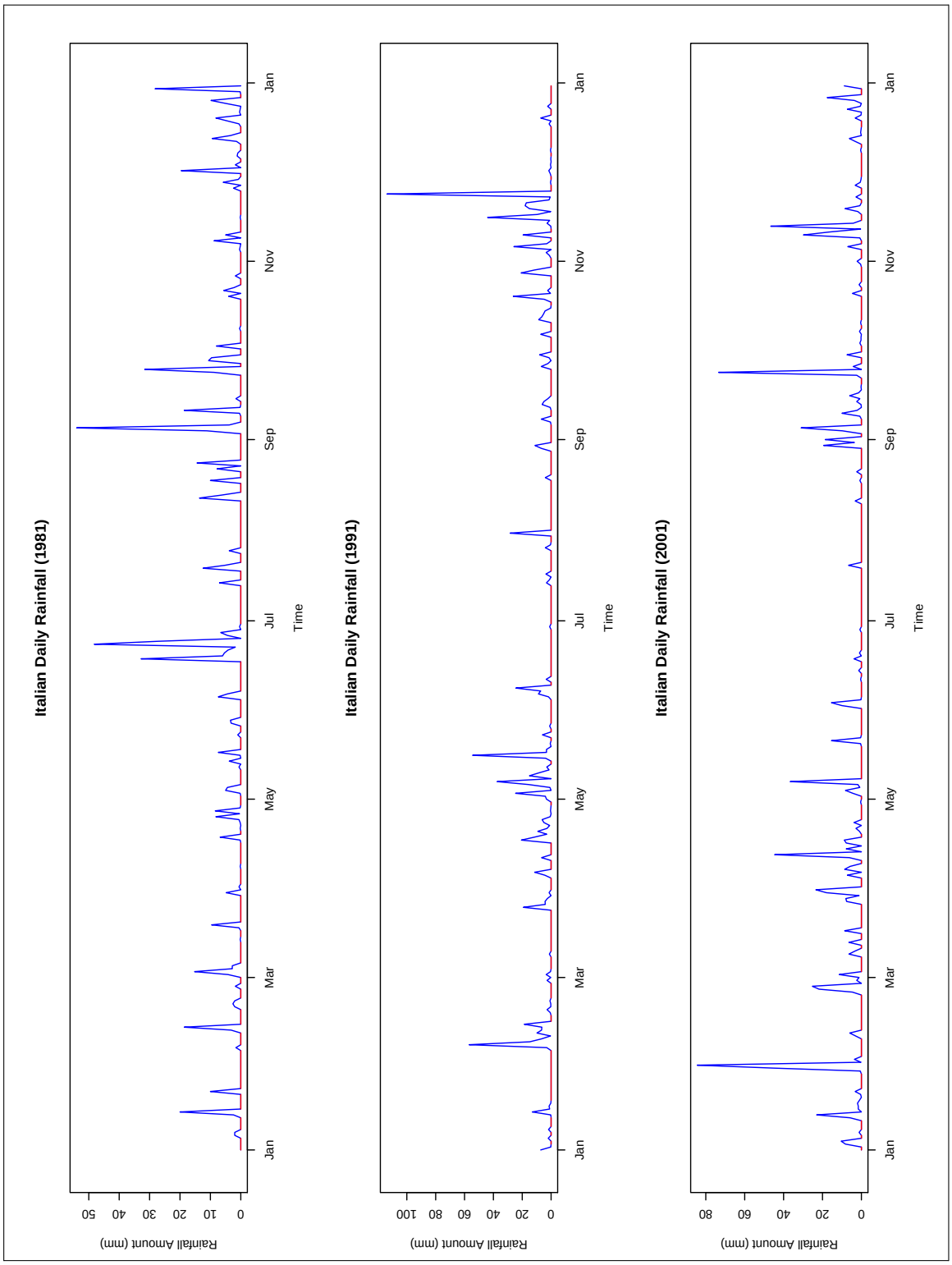


Figure 4.3: Daily rainfall for Urbino rain gauge station in year 1981, 1991 and 2001. Days with zero rainfall are shown in red.

4.3.2 The Model

4.3.2.1 Amount Process

In order to model the amount process, we utilised the ideas from Fernandes *et al.* (2009) and Suhaila *et al.* (2011) by which we chose several distributions so that we can assess and determine the best distribution for the amount process. To accomplish this, we will compare the lognormal and gamma distributions since both distributions have 0 to ∞ support, a desirable property for modelling the amount process. Cho *et al.* (2004) had previously compared the lognormal and gamma distributions for rainfall data and they established that the lognormal distribution is more appropriate for dry regions, whilst a wet region is better modelled using a gamma distribution. However, we will use two different parameterisations for the gamma distribution. For the first parameterisation, we shall use a gamma distribution with fixed shape parameter (α) and variable scale parameter, (β), denoted as "G1" distribution, and in the second parameterisation, a gamma distribution with a variable shape parameter (α) and fixed scale parameter, (β), denoted as "G2" distribution, will be used. The full details of these distributions are given below:

- Lognormal distribution:
  The observed rainfall amount can be defined as

  $$W_t = R_t Y_t$$

  where Y_t is assumed to follow a lognormal distribution. If $Z_t = \log(Y_t)$, then Z_t follows a normal distribution with the probability density function given by:

  $$f(z_t|\vartheta_t, \tau) = \sqrt{\frac{\tau}{2\pi}} \exp\left\{-\frac{\tau}{2}\left(z_t - \vartheta_t\right)^2\right\} \tag{4.13}$$

  where $\mathrm{E}(Z_t) = \vartheta_t$ and $\mathrm{Var}(Z_t) = 1/\tau$. The mean and variance of Y_t are given as $\mu_t = e^{\vartheta_t + 1/2\tau}$ and $V_{\boldsymbol{y}_y} = e^{2\vartheta_t + 1/\tau}(e^{1/\tau} - 1)$.

  If Z_t and Z_s, $s \neq t$, are conditionally independent given ϑ_t, ϑ_s and τ, the likelihood function for ϑ_t and τ is represented by:

  $$L(\vartheta_t, \tau|z_t) = \prod_{t=1}^{T} \sqrt{\frac{\tau}{2\pi}} \exp\left\{-\frac{\tau}{2}\left[z_t - \vartheta_t\right]^2\right\}. \tag{4.14}$$

  This transformation allows us to evaluate the parameter in a much simpler and straightforward way since a Gaussian distribution can be used. Sofia (2007) employed the lognormal distribution to evaluate the monthly rainfall in the Mediterranean region. Heaps *et al.* (2015) also used the lognormal distribution in non-homogeneous

hidden Markov models (NHMMs) for describing the rainfall amount during winter. Figure 4.4 shows different plots of probability density functions for the lognormal

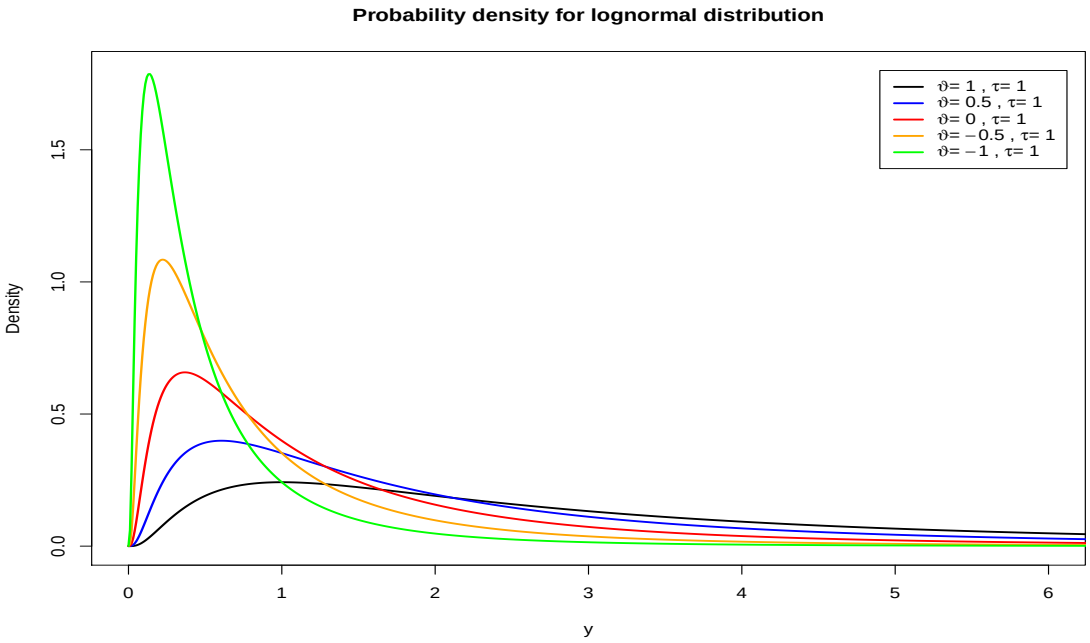


Figure 4.4: The forms of probability density function for lognormal distribution with different ϑ

distribution, each having different mean, ϑ and constant precision, τ, for the logarithm of rainfall amount. As we can clearly observe, the shape of the distribution is skewed to the right and its mode decreases as ϑ decreases. The density is always zero at $y = 0$.

- Gamma distribution:
  Another distribution that is widely used for rainfall amount representation is the gamma distribution. The gamma distribution is a popular choice for rainfall data because it produces a good fit and its shape is approximately identical to the histogram of rainfall data (Ben-Gai *et al.*, 1998). The observed rainfall is hence given by:

$$W_t = R_t Y_t$$

  where Y_t follows a gamma distribution. The G1 distribution parameterisation is represented by:

$$f(y_t|\alpha, \beta_t) = \frac{\beta_t^{\alpha} y_t^{\alpha-1} e^{-\beta_t y_t}}{\Gamma(\alpha)} \tag{4.15}$$

  with constant shape parameter, $\alpha > 0$ and variable scale parameter, $\beta_t = \alpha/\mu_t$. The mean of this distribution is $\mathrm{E}(Y_t) = \mu_t$ and the variance is $\mathrm{Var}(Y_t) = \mu_t^2/\alpha$. If the

observations are conditionally independent over time, the likelihood function based on the G1 distribution for α and μ_t is given by:

$$L(\alpha, \mu_t|y_t) = \prod_{t=1}^{T} \frac{(\alpha/\mu_t)^{\alpha} y_t^{\alpha-1} e^{-(\alpha/\mu_t)y_t}}{\Gamma(\alpha)}. \tag{4.16}$$

In Figure 4.5, the probability density function plots for gamma distributions with

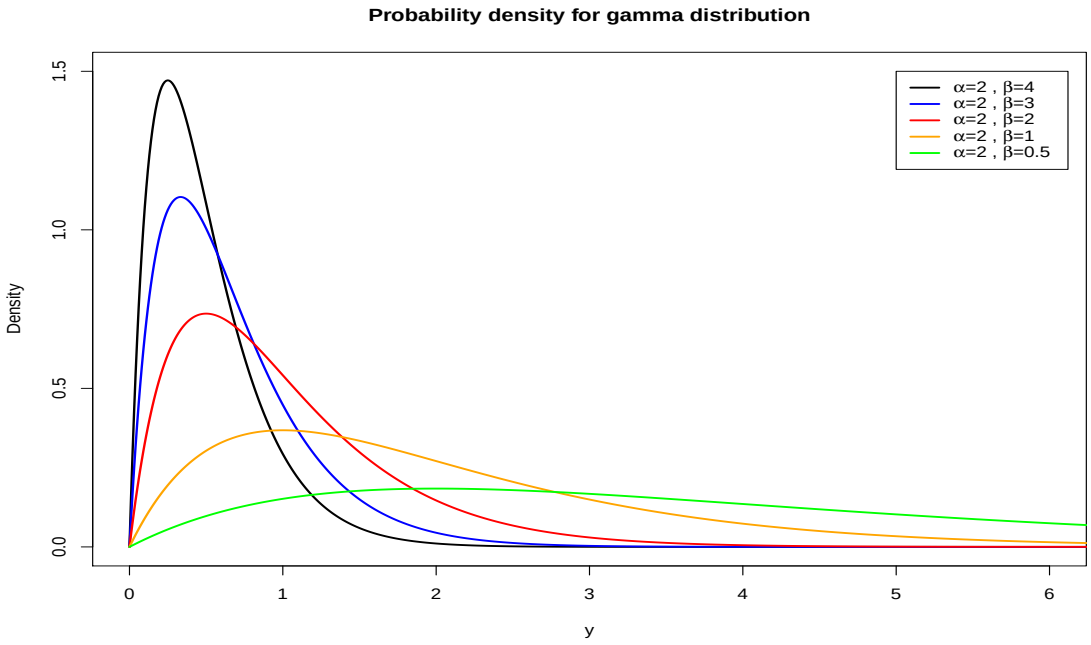


Figure 4.5: The forms of probability density function for gamma distribution with different β

variable β and constant $\alpha = 2$ are shown.

For the G2 distribution, the probability density function is specified as

$$f(y_t|\beta, \alpha_t) = \frac{\beta^{\alpha_t} y_t^{\alpha_t - 1} e^{-\beta y_t}}{\Gamma(\alpha_t)} \tag{4.17}$$

where $\alpha_t = \mu_t\beta$ represents the variable shape parameter and $\beta > 0$ is the fixed scale parameter. The mean for this distribution is analogous to the mean of the G1 distribution but the variance is dissimilar. Since we have fixed β, the variance for this distribution is $\text{Var}(Y_t) = \mu_t/\beta$. If the observations are assumed to be conditionally independent with time, then the likelihood function for β and μ_t is:

$$L(\beta, \mu_t|y_t) = \prod_{t=1}^{T} \frac{\beta^{\mu_t\beta} y_t^{\mu_t\beta - 1} e^{-\beta y_t}}{\Gamma(\mu_t\beta)}. \tag{4.18}$$

Figure 4.6 demonstrates the different probability density function plots when α is allowed to vary and β is fixed at 1. For $\alpha > 1$, $f(0) = 0$. If $\alpha = 1$ then $f(0) = \beta = 1/\mu$. If $\alpha < 1$ then $f(y) \to \infty$ as $y \to 0$. In the case of $\alpha = 1$, it is identical to the exponential distribution. In general, the shape of the gamma distribution

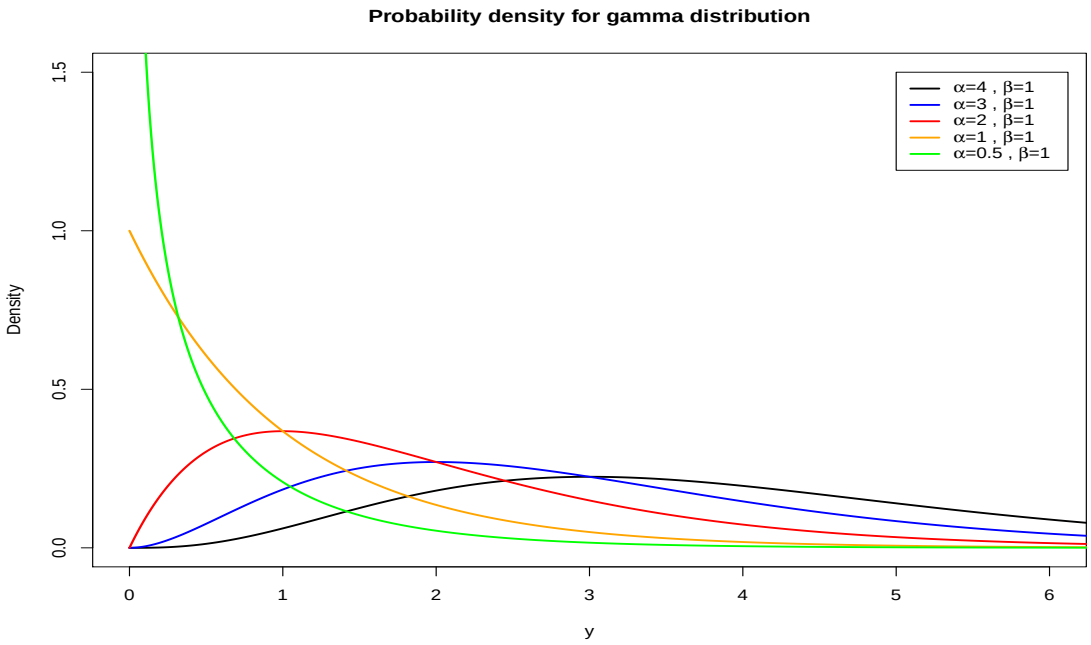


Figure 4.6: The forms of probability density function for gamma distribution with different α

is similar to the lognormal distribution's shape except when $\alpha \leq 1$. An important consideration when choosing between a gamma and lognormal distribution is the behaviour of the density when y is close to zero, in particular whether the density should go to zero. We also need to consider whether and, if so, how the shape of the distribution should change if the mean changes. In particular this is the difference between the G1 and G2 models.

For the rainfall amount, we assume that the mean of the distribution varies over time. To fit the mean, we used the idea from Stern & Coe (1984) and this involves utilising the truncated Fourier series for modelling the variation of daily rainfall over the year. In the case of the lognormal distribution, the mean of the log amount Z_t is given by

$$\vartheta_t = \hat{\eta}_t,$$

while, for G1 and G2 distributions, the mean is:

$$\mu_t = e^{\hat{\eta}_t}$$

where

$$\hat{\eta}_t = \eta_0 + \sum_{f=1}^{F}\left[a_f \cos\left(\frac{2\pi f t}{365.25}\right) + b_f \sin\left(\frac{2\pi f t}{365.25}\right)\right]. \tag{4.19}$$

The parameter η_0 is the common level for the amount of rainfall. We will use three harmonics of the Fourier series as indicated in Section 4.2.4 to represent the seasonal effect of the rainfall amount over the year.

4.3.2.2 Occurrence Process

In the Italian daily rainfall application, we use a first-order Markov chain to model the probability of rainfall. The transition probabilities of the first-order Markov chain model are given by:

$$p_{ij}(t) = \Pr\{r_t = j | r_{t-1} = i\}; \qquad i, j \in \{0, 1\},$$

with the transition matrix:

$$P_t = \begin{array}{c} \\ r_{t-1}=0 \\ r_{t-1}=1 \end{array} \begin{array}{c} \begin{array}{cc} r_t = 0 & r_t = 1 \end{array} \\ \begin{pmatrix} p_{00}(t) & p_{01}(t) \\ p_{10}(t) & p_{11}(t) \end{pmatrix} \end{array}.$$

To represent the rainfall probability, we use a logistic link function to connect the predictors with the rainfall probability $p_{i1}(t)$:

$$\text{logit}\left[p_{i1}(t)\right] = \hat{\zeta}_{i,t}; \qquad i = 0, 1.$$

If the distribution of Y_t is a lognormal distribution, then the parameterisation for $\hat{\zeta}_{i,t}$ is given by

$$\hat{\zeta}_{i,t} = \zeta_0 + \sum_{f=1}^{F}\left[c_f \cos\left(\frac{2\pi f}{365.25}\right) + d_f \sin\left(\frac{2\pi f}{365.25}\right)\right] + \zeta_1 \vartheta_t + \zeta_2\left(i - \frac{1}{2}\right). \tag{4.20}$$

If Y_t follows a gamma distribution, then we have

$$\hat{\zeta}_{i,t} = \zeta_0 + \sum_{f=1}^{F}\left[c_f \cos\left(\frac{2\pi f}{365.25}\right) + d_f \sin\left(\frac{2\pi f}{365.25}\right)\right] + \zeta_1 \log(\mu_t) + \zeta_2\left(i - \frac{1}{2}\right). \tag{4.21}$$

The parameter ζ_0 is a common intercept for $\text{logit}\left[p_{i1}(t)\right]$ and the term i is an indicator function with a value of $i = 1$ if rain occurs on the previous day and $i = 0$ otherwise. We also included ϑ_t and $\log(\mu_t)$ in Equations (4.20) and (4.21) to create a link between the

amount and occurrence processes so that the accuracy of the rainfall probability can be improved. For instance, the rainfall probability increases if μ_t increases. We also employed a Fourier series with three harmonics for rainfall occurrence to allow for the seasonal effect over the year.

The conditional probabilities can be summarised as follows:

$$p_{11}(t) = \frac{\exp\left(\hat{\zeta}_{1,t}\right)}{1+\exp\left(\hat{\zeta}_{1,t}\right)}; \qquad p_{10}(t) = 1 - p_{11}(t),$$
$$p_{01}(t) = \frac{\exp\left(\hat{\zeta}_{0,t}\right)}{1+\exp\left(\hat{\zeta}_{0,t}\right)}; \qquad p_{00}(t) = 1 - p_{01}(t).$$

The unconditional probability can be represented by:

$$\begin{aligned}\hat{p}_t =& p_{11}(t)\hat{p}_{t-1} + p_{01}(t)\left(1-\hat{p}_{t-1}\right)\\ =& p_{01}(t) + \left[p_{11}(t) - p_{01}(t)\right]\hat{p}_{t-1}\\ \hat{p}_t + \left[p_{01}(t) - p_{11}(t)\right]\hat{p}_{t-1} =& p_{01}(t).\end{aligned}$$

We can write this as

$$\ddot{A}\hat{\boldsymbol{P}} = \boldsymbol{p}$$

and hence

$$\hat{\boldsymbol{P}} = \ddot{A}^{-1}\boldsymbol{p} \tag{4.22}$$

with $\hat{P} = (\hat{p}_1, \cdots, \hat{p}_{365})'$, $\boldsymbol{p} = (p_{01}(1), \cdots, p_{01}(365))'$ and

$$\ddot{A} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & \ddot{a}_1 \\ \ddot{a}_2 & 1 & 0 & \cdots & 0 & 0 \\ 0 & \ddot{a}_3 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & \ddot{a}_{365} & 1 \end{bmatrix}$$

where $\ddot{a}_t = p_{01}(t) - p_{11}(t)$.

However, the parameterisation in Equations (4.20) and (4.21) may lead to poor mixing when we run the MCMC algorithm. To overcome this problem, we centered the predictor variables as follows:

- lognormal distribution

$$\hat{\zeta}_{i,t} = \zeta_0 + \sum_{f=1}^{F} \left[c_f \cos\left(\frac{2\pi f}{365.25}\right) + d_f \sin\left(\frac{2\pi f}{365.25}\right) \right] + \zeta_1 \left(\vartheta_t - \zeta_0\right) + \zeta_2 \left(i - \frac{1}{2}\right). \tag{4.23}$$

- gamma distribution

$$\hat{\zeta}_{i,t} = \zeta_0 + \sum_{f=1}^{F} \left[c_f \cos\left(\frac{2\pi f}{365.25}\right) + d_f \sin\left(\frac{2\pi f}{365.25}\right) \right] + \zeta_1 \left(\log(\mu_t) - \zeta_0\right) + \zeta_2 \left(i - \frac{1}{2}\right). \tag{4.24}$$

4.3.2.3 Prior specifications

This section provides the detailed descriptions on the construction of the prior distributions for the unknown parameters in our model. The prior density of the unknown parameters is represented by $\pi(\boldsymbol{\theta})$. Priors will be constructed based on the information acquired from previous studies and personal beliefs.

Suppose that the precision parameter, τ, in the lognormal distribution follows a gamma distribution:

$$\tau \sim Ga(g_\tau, h_\tau)$$

for fixed hyperparameters g_τ and h_τ since the values are strictly positive on the real line. This is a sensible choice because the gamma distribution is semi-conjugate to the likelihood function of the lognormal form. For parameters α in the G1 distribution and β in the G2 distribution, we also assign gamma priors since their values are stricly positive. The priors are therefore:

$$\alpha \sim Ga(g_\alpha, h_\alpha); \qquad \beta \sim Ga(g_\beta, h_\beta)$$

with fixed hyperparameters g_α, h_α, g_β and h_β.

In our model, we use a GLM approach to relate the linear predictor to the response variable using a link function for both the amount and occurrence processes. The set of linear coefficients for the amount process is represented by $\boldsymbol{\eta} = (\eta_0, a_1, a_2, a_3, b_1, b_2, b_3)$ and $\boldsymbol{\zeta} = (\zeta_0, \zeta_1, \zeta_2, c_1, c_2, c_3, d_1, d_2, d_3)$ is the parameter vector for the occurrence process. Since the linear predictor can take any values from $-\infty$ to ∞, a normal prior for each unknown parameter in the linear predictors is appropriate. If we assume prior independence between the parameters, then the priors for all parameters are given below:

- Amount process

$$\eta_0 \sim N(m_{A0}, v_{A0}) \qquad a_1 \sim N(m_{a_1}, v_{a_1})$$
$$a_2 \sim N(m_{a_2}, v_{a_2}) \qquad a_3 \sim N(m_{a_3}, v_{a_3})$$
$$b_1 \sim N(m_{b_1}, v_{b_1}) \qquad b_f \sim N(m_{b_2}, v_{b_2})$$
$$b_3 \sim N(m_{b_3}, v_{b_3})$$

- Occurrence process

$$\zeta_0 \sim N(m_{C0}, v_{C0}) \qquad \zeta_1 \sim N(m_{C1}, v_{C1})$$
$$\zeta_2 \sim N(m_{C2}, v_{C2}) \qquad c_1 \sim N(m_{c_1}, v_{c_1})$$
$$c_2 \sim N(m_{c_2}, v_{c_2}) \qquad c_3 \sim N(m_{c_3}, v_{c_3})$$
$$d_1 \sim N(m_{d_1}, v_{d_1}) \qquad d_2 \sim N(m_{d_2}, v_{d_2})$$
$$d_3 \sim N(m_{d_3}, v_{d_3}).$$

We can relax the prior independence assumption by assigning a multivariate normal prior for $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$:

$$\boldsymbol{\eta} \sim N_7(\bar{\eta}, P_\eta^{-1})$$
$$\boldsymbol{\zeta} \sim N_9(\bar{\zeta}, P_\zeta^{-1})$$

where $\bar{\eta}$ and $\bar{\zeta}$ represent the mean vectors with precisions P_η and P_ζ.

4.3.2.4 Posterior distributions

The joint density of $(\boldsymbol{\theta}, \boldsymbol{y}, \boldsymbol{r})$ is given by:

$$\begin{aligned}\pi(\boldsymbol{\theta}, \boldsymbol{y}, \boldsymbol{r}) &= \pi(\boldsymbol{\theta}) \times f(\boldsymbol{y}|\boldsymbol{\theta}) \times f(\boldsymbol{r}|\boldsymbol{\theta}) \\ &= \pi(\boldsymbol{\theta}_{amt}) \times \pi(\boldsymbol{\theta}_{occ}) \times \prod_{t=1}^{T} f(y_t|\boldsymbol{\theta}_{amt}) \times f(r_1 \mid \boldsymbol{\theta}_{occ}, \boldsymbol{\theta}_{amt}) \prod_{t=2}^{T} f(r_t \mid r_{t-1}, \boldsymbol{\theta}_{occ}, \boldsymbol{\theta}_{amt}).\end{aligned}$$

Note that y_t is only observed when $r_t = 1$. To obtain the posterior distribution, we need to work out the FCD for each unknown parameter from this joint density. The FCDs in this section are provided only as general information and for future work. The full details of the joint posterior and the FCDs for the amount process for these three distributions are given as follows:

1. Lognormal distribution

The FCD for τ is given as follows:

$$\begin{aligned}\pi(\tau|\boldsymbol{z},\boldsymbol{\theta}_{amt}) &\propto \prod_{t=1}^{T} f(z_t|\vartheta_t) \times \pi(\tau) \\ &\propto \prod_{t=1}^{T} \sqrt{\frac{\tau}{2\pi}} \exp\left\{-\frac{\tau}{2}(z_t-\vartheta_t)^2\right\} \times \frac{{h_\tau}^{g_\tau}\tau^{g_\tau-1}e^{-h_\tau\tau}}{\Gamma(g_\tau)} \\ &\propto \tau^{T/2+g_\tau-1} \exp\left\{-\tau\left[\frac{1}{2}\sum_{t=1}^{T}(z_t-\vartheta_t)^2+h_\tau\right]\right\}.\end{aligned}$$

This form corresponds to a gamma density

$$\tau \sim Ga(G_\tau, H_\tau)$$

where

$$\begin{aligned}G_\tau &= T/2 + g_\tau \\ H_\tau &= \frac{1}{2}\sum_{t=1}^{T}(y_t-\vartheta_t)^2 + h_\tau.\end{aligned}$$

Let $\boldsymbol{\vartheta} = \{\vartheta_1, \cdots, \vartheta_T\}'$ represent the mean vector, hence $\sum(z_t-\vartheta_t)^2)$ is $(\boldsymbol{z}-\boldsymbol{\vartheta})'(\boldsymbol{z}-\boldsymbol{\vartheta})$. Then,

$$f(\boldsymbol{z}|\boldsymbol{\theta}_{amt}) \propto \exp\left\{-\frac{\tau}{2}(\boldsymbol{z}-\boldsymbol{\vartheta})'(\boldsymbol{z}-\boldsymbol{\vartheta})\right\}.$$

Although the Markov chain is non-homogeneous, the transition probabilities change slowly and a reasonable approximation to the marginal probabilities for r_1 is

$$\begin{aligned}f(r_1|\boldsymbol{\theta}_{occ},\boldsymbol{\theta}_{amt}) &= \left[\frac{1-p_{01}(1)}{p_{01}(1)+p_{10}(1)}\right]^{r_1}\left[\frac{p_{01}(1)}{p_{01}(1)+p_{10}(1)}\right]^{1-r_1} \\ &= \frac{[p_{01}(1)]^{r_1}\,[p_{10}(1)]^{1-r_1}}{p_{10}(1)+p_{01}(1)},\end{aligned}$$

and hence,

$$f(\boldsymbol{r}|\boldsymbol{\theta}_{occ},\boldsymbol{\theta}_{amt}) = \frac{[p_{01}(1)]^{r_1}\,[p_{10}(1)]^{1-r_1}}{p_{10}(1)+p_{01}(1)}\prod_{t=2}^{T}[p_{i1}(t)]^{r_t}\,[1-p_{i1}(t)]^{1-r_t}.$$

Therefore, the FCD for $\boldsymbol{\eta}$ is given by:

$$\begin{aligned}
\pi(\boldsymbol{\eta}|\boldsymbol{z},\boldsymbol{r}) \propto & f(\boldsymbol{z}|\boldsymbol{\theta}_{amt}) \times f(\boldsymbol{r}|\boldsymbol{\theta}_{occ},\boldsymbol{\theta}_{amt}) \times \pi(\boldsymbol{\eta}) \\
\propto & \exp\left\{-\frac{\tau}{2}\left(\boldsymbol{z}-\boldsymbol{\vartheta}\right)'\left(\boldsymbol{z}-\boldsymbol{\vartheta}\right)\right\} \\
& \times \frac{\left[p_{01}(1)\right]^{r_1}\left[p_{10}(1)\right]^{1-r_1}}{p_{10}(1)+p_{01}(1)} \prod_{t=2}^{T}\left[p_{i1}(t)\right]^{r_t}\left[1-p_{i1}(t)\right]^{1-r_t} \\
& \times \frac{|P_\eta|^{1/2}}{(2\pi)^{2F/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\eta}-\underline{\eta}_0)'P_\eta(\boldsymbol{\eta}-\underline{\eta}_0)\right\} \\
\propto & \frac{\left[p_{01}(1)\right]^{r_1}\left[p_{10}(1)\right]^{1-r_1}}{p_{10}(1)+p_{01}(1)} \prod_{t=2}^{T}\left[p_{i1}(t)\right]^{r_t}\left[1-p_{i1}(t)\right]^{1-r_t} \\
& \exp\left\{-\frac{1}{2}\left[\tau\left(\boldsymbol{z}-\boldsymbol{\vartheta}\right)'\left(\boldsymbol{z}-\boldsymbol{\vartheta}\right)+(\boldsymbol{\eta}-\underline{\eta}_0)'P_\eta(\boldsymbol{\eta}-\underline{\eta}_0)\right]\right\}
\end{aligned}$$

where $i = r_{t-1} \in \{0,1\}$. This distribution does not belong to any standard form.

2. G1 distribution

The FCD of the shape parameter, α for the G1 distribution is given by:

$$\begin{aligned}
\pi(\alpha|\boldsymbol{y}) \propto & \prod_{t=1}^{T} f(y_t|\boldsymbol{\theta}_{amt}) \times \pi(\alpha) \\
\propto & \frac{\prod_{t=1}^{T}(\alpha/\mu_t)^{\alpha} y_t^{\alpha-1} e^{-(\alpha/\mu_t)y_t}}{\Gamma(\alpha)} \times \frac{h_\alpha^{g_\alpha}\alpha^{g_\alpha-1}e^{-h_\alpha\alpha}}{\Gamma(g_\alpha)} \\
\propto & \frac{\alpha^{\alpha T+g_\alpha-1}\prod_{t=1}^{T}E_t^{\alpha}\exp\left\{-\alpha\left[\sum_{t=1}^{T}E_t+h_\alpha\right]\right\}}{\left[\Gamma(\alpha)\right]^T}
\end{aligned}$$

where $E_t = y_t/\mu_t$. This is again not in a standard form.

For the scale parameter, $\boldsymbol{\eta}$, the FCD is derived as follows:

$$\begin{aligned}
\pi(\boldsymbol{\eta}|\boldsymbol{y},\boldsymbol{r}) &\propto \prod_{t=1}^{T} f(y_t|\boldsymbol{\theta}_{amt}) \times \prod_{t=1}^{T} f(r_t|\boldsymbol{\theta}_{occ},\boldsymbol{\theta}_{amt}) \times \pi(\boldsymbol{\eta}) \\
&\propto \frac{\prod_{t=1}^{T}(\alpha/\mu_t)^{\alpha} y_t^{\alpha-1} e^{-(\alpha/\mu_t)y_t}}{\Gamma(\alpha)} \\
&\quad \times \frac{[p_{01}(1)]^{r_1}\,[p_{10}(1)]^{1-r_1}}{p_{10}(1)+p_{01}(1)} \prod_{t=2}^{T} [p_{i1}(t)]^{r_t}\,[1-p_{i1}(t)]^{1-r_t} \\
&\quad \times \frac{|P_\eta|^{1/2}}{(2\pi)^{2F/2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{\eta}-\underline{\eta}_0)' P_\eta (\boldsymbol{\eta}-\underline{\eta}_0) \right\} \\
&\propto \frac{[p_{01}(1)]^{r_1}\,[p_{10}(1)]^{1-r_1}}{p_{10}(1)+p_{01}(1)} \prod_{t=2}^{T} [p_{i1}(t)]^{r_t}\,[1-p_{i1}(t)]^{1-r_t} \\
&\quad \prod_{t=1}^{T} E_t^{\alpha} \exp\left\{ -\left[\alpha \sum_{t=1}^{T} E_t + \frac{1}{2}(\boldsymbol{\eta}-\underline{\eta}_0)' P_\eta (\boldsymbol{\eta}-\underline{\eta}_0) \right] \right\}.
\end{aligned}$$

This FCD for $\boldsymbol{\eta}$ again does not correspond to any standard density form.

3. G2 distribution

The FCD of β can be derived as follows:

$$\begin{aligned}
\pi(\beta|\boldsymbol{y}) &\propto \prod_{t=1}^{T} f(y_t|\boldsymbol{\theta}_{amt}) \times \pi(\beta) \\
&\propto \prod_{t=1}^{T} \frac{\beta^{\beta\mu_t} y_t^{\beta\mu_t-1} e^{-\beta y_t}}{\Gamma(\beta\mu_t)} \times \frac{h_\beta^{g_\beta} \beta^{g_\beta-1} e^{-h_\beta \beta}}{\Gamma(g_\beta)} \\
&\propto \frac{\beta^{\beta\sum_{t=1}^{T}\mu_t + g_\beta - 1} \prod_{t=1}^{T} y_t^{\beta\mu_t-1}}{\Gamma(\beta\mu_t)} \exp\left[-\beta \left(\sum_{t=1}^{T} y_t + h_\beta \right) \right].
\end{aligned}$$

This is not the density of a standard distribution.

For $\boldsymbol{\eta}$, the FCD is obtained as follows:

$$\begin{aligned}
\pi(\boldsymbol{\eta}|\boldsymbol{y},\beta) \propto & \prod_{t=1}^{T} f(y_t|\boldsymbol{\theta}_{amt}) \times \prod_{t=1}^{T} f(r_t|\boldsymbol{\theta}_{occ},\boldsymbol{\theta}_{amt}) \times \pi(\boldsymbol{\eta}) \\
\propto & \prod_{t=1}^{T} \frac{\beta^{\beta\mu_t} y_t^{\beta\mu_t-1} e^{-\beta y_t}}{\Gamma(\beta\mu_t)} \\
& \times \frac{[p_{01}(1)]^{r_1}\,[p_{10}(1)]^{1-r_1}}{p_{10}(1)+p_{01}(1)} \prod_{t=2}^{T} [p_{i1}(t)]^{r_t}\,[1-p_{i1}(t)]^{1-r_t} \\
& \times \frac{|P_\eta|^{1/2}}{(2\pi)^{2F/2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{\eta}-\underline{\eta}_0)' P_\eta (\boldsymbol{\eta}-\underline{\eta}_0) \right\} \\
\propto & \frac{[p_{01}(1)]^{r_1}\,[p_{10}(1)]^{1-r_1}}{p_{10}(1)+p_{01}(1)} \prod_{t=2}^{T} [p_{i1}(t)]^{r_t}\,[1-p_{i1}(t)]^{1-r_t} \\
& \frac{\beta^{\beta\sum_{t=1}^{T}\mu_t+g_\beta-1}\prod_{t=1}^{T} y_t^{\beta\mu_t-1}}{\Gamma(\beta\mu_t)} \exp\left\{ -\left[\beta\sum_{t=1}^{T} y_t + \frac{1}{2}(\boldsymbol{\eta}-\underline{\eta}_0)' P_\eta (\boldsymbol{\eta}-\underline{\eta}_0) \right] \right\}.
\end{aligned}$$

This density also does not correspond to any standard-form density.

The latent values, y_t and z_t are not observed when it is a dry day ($W_t = 0$). It is not strictly necessary to sample the missing potential rainfall values in the Italian model. However it simplifies the computations if we do. Let y_{dry} and z_{dry} be the values that are not observed and thus we can treat them as auxiliary data within the data augmentation framework (Tanner & Wong, 1987). When $r_t = 0$, it is necessary to sample the posterior value of y_{dry} and z_{dry} by considering the FCD as follows:

1. Lognormal distribution

$$\begin{aligned}
\pi(z_{dry}|r=0,\vartheta_t,\tau) \propto & f(z_{dry}|\vartheta_t,\tau) \\
\propto & \sqrt{\frac{\tau}{2\pi}} \exp\left\{ -\frac{\tau}{2}(z_{dry}-\vartheta_t)^2 \right\}.
\end{aligned}$$

2. G1 distribution

$$\begin{aligned}
\pi(y_{dry}|r=0,\alpha,\mu_t) \propto & f(y_{dry}|\alpha,\mu_t) \\
\propto & \frac{(\alpha/\mu_t)^\alpha y_{dry}^{\alpha-1} e^{-(\alpha/\mu_t)y_{dry}}}{\Gamma(\alpha)}.
\end{aligned}$$

3. G2 distribution

$$\pi(y_{dry}|r=0,\beta,\mu_t) \propto \pi(y_{dry}|\beta,\mu_t) \propto \frac{\beta^{\beta\mu_t} y_{dry}^{\beta\mu_t-1} e^{-\beta y_{dry}}}{\Gamma(\beta\mu_t)}.$$

For the occurrence process, the FCD for $\boldsymbol{\zeta}$ can be written as

$$\begin{aligned}
\pi(\boldsymbol{\zeta}|\boldsymbol{r}) &\propto \prod_{t=1}^{T} f(r_t|\boldsymbol{\theta}_{occ},\boldsymbol{\theta}_{amt}) \times \pi(\boldsymbol{\zeta}) \\
&\propto \frac{[p_{01}(1)]^{r_1}\,[p_{10}(1)]^{1-r_1}}{p_{10}(1)+p_{01}(1)} \prod_{t=2}^{T} [p_{i1}(t)]^{r_t}\,[1-p_{i1}(t)]^{1-r_t} \times \frac{|P_\zeta|^{1/2}}{(2\pi)^{2F/2}} \\
&\quad \exp\left\{ -\frac{1}{2}(\boldsymbol{\zeta}-\underline{\zeta}_0)' P_\zeta (\boldsymbol{\zeta}-\underline{\zeta}_0) \right\} \\
&\propto \frac{[p_{01}(1)]^{r_1}\,[p_{10}(1)]^{1-r_1}}{p_{10}(1)+p_{01}(1)} \prod_{t=2}^{T} [p_{i1}(t)]^{r_t}\,[1-p_{i1}(t)]^{1-r_t} \exp\left\{ -\frac{1}{2}(\boldsymbol{\zeta}-\underline{\zeta}_0)' P_\zeta (\boldsymbol{\zeta}-\underline{\zeta}_0) \right\}
\end{aligned}$$

where $i = r_{t-1} \in \{0,1\}$. Again, the FCD does not conform to any density of a standard form.

As we can see, the FCDs of the majority of parameters for the amount and occurrence processes are not of any standard form. Therefore, the implementation of a Metropolis-within-Gibbs scheme is required to generate the posterior samples for each parameter. For example, the detailed steps of the MCMC scheme when we use the lognormal distribution are given as follows:

1. Initialise the iteration counter to $j = 1$ and set the initial state of the chain to $\boldsymbol{\theta}^{(0)} = \left(\tau^{(0)}, \boldsymbol{\eta}^{(0)}, \boldsymbol{\zeta}^{(0)}, z_{dry}^{(0)}\right)$.

2. Obtain a new value $\boldsymbol{\theta}^{(j)}$ from $\boldsymbol{\theta}^{(j-1)}$ by successive generation of values for the following quantities

   - $\tau^{(j)} \sim \pi\left(\tau|\boldsymbol{\eta}^{(j-1)}, \boldsymbol{z}\right)$ using a Gibbs sampler step where

     $$\tau^{(j)} \sim Ga\left(G_\tau, H_\tau\right).$$

   - $\boldsymbol{\eta}^{(j)} \sim \pi\left(\boldsymbol{\eta}|\alpha^{(j)}, \boldsymbol{z}\right)$ using a Metropolis-Hastings step with proposal distribution:

     $$\boldsymbol{\eta}^{*} \sim N_7\left(\boldsymbol{\eta}^{(j-1)}, \Sigma_{\boldsymbol{\eta}^{*}}\right).$$

- $\boldsymbol{\zeta}^{(j)} \sim \pi\left(\boldsymbol{\zeta}|\boldsymbol{r}\right)$ using a Metropolis-Hastings step with proposal distribution:

$$\boldsymbol{\zeta}^{*} \sim N_9\left(\boldsymbol{\zeta}^{(j-1)}, \Sigma_{\boldsymbol{\zeta}^{*}}\right).$$

- $z_{dry}^{(j)} \sim \pi\left(z_{dry}|\tau^{(j)}, \boldsymbol{\eta}^{(j)}, \boldsymbol{z}\right)$.

3. Change the iteration counter from j to $j+1$ and return to step 2.

4.3.3 Application

4.3.3.1 Prior distributions for the amount process

For the Italian daily rainfall application, the prior specification was constructed using suitable distributions based on the information acquired from previous studies and personal beliefs. The form of the prior here is a recommendation but the choice of hyperparameters is an illustrative example only. However, it is recommended that, in practical applications, prior information should be obtained from experts on rainfall since they are the ones who have a great deal of knowledge about rainfall. While it was not possible, within the scope of this project, to conduct experiments to test elicitation methods, we provide here elicitation questions about observable quantities which could help the experts to specify judgements for prior distributions. By providing these questions we show that it is possible to construct an elicitation scheme based on judgements about observable quantities. We adopt a similar approach with the other examples later in this thesis.

We begin by considering the prior distribution for the parameters of the amount distribution at a particular time of year. We will consider the seasonal effect later, in Section 4.3.3.3.

- Lognormal distribution

  Suppose that we imagine a large number of wet days, $Y_1, \cdots, Y_{T_w}$ for a certain time, for example, in March. Let $Z_j = \log Y_j$ and we have T_w which is large enough so that the sample mean $\bar{Z} = \frac{1}{T_w}\sum_{j=1}^{T_w} Z_j$ is approximately ϑ_t and the sample variance $S_z^2 = \frac{1}{T_w - 1}\sum_{t=1}^{T_w}\left(Z_t - \bar{Z}\right)^2$ is approximately $\sigma^2 = 1/\tau$. To assess the uncertainty for $\bar{Z}$, we can ask the expert as follows:

  Q11 *"Please give a value such that you think that it is equally likely that $\bar{Z}$ is less than or greater than the value."* Let the given value be $Q_{z,2}$.

  Q12 *"Suppose you were told that $\bar{Z}$ is less than $Q_{z,2}$. Please give a new value such that you think that it is equally likely that $\bar{Z}$ is less than or greater than this value."* Let the given value be $Q_{z,1}$.

Q13 *"Suppose you were told that $\bar{Z}$ is greater $Q_{z,2}$. Please give a new value such that you think that it is equally likely that $\bar{Z}$ is less than or greater than this value."* Let the given value be $Q_{z,3}$.

Then $Q_{z,1}$ and $Q_{z,3}$ are our lower and upper quartiles for $\bar{Z}$. Hence, the prior mean for $\hat{\eta}_t$ is $m_{A0} = \frac{1}{2}(Q_{z,1}+Q_{z,3})$ and the prior variance for $\hat{\eta}_t$ is $v_{A0} = [(Q_{z,3}-Q_{z,1})/1.349]^2$.

Suppose that our three quartiles for S_Z^2 are $Q_{S,1}$, $Q_{S,2}$ and $Q_{S,3}$ where we can obtain these values by asking the expert as follows:

Q14 *"Please now consider the sample variance, S_Z^2. Can you give a value $Q_{S,2}$ such that S_Z^2 is equally likely to be less than or greater than $Q_{S,2}$."*

Q15 *"Suppose you were told that S_Z^2 will be less than $Q_{S,2}$. Please give a value $Q_{S,1}$ such that it is equally that S_Z^2 is less than or greater than $Q_{S,1}$."*

Q16 *"Suppose you were told that S_Z^2 will be greater than $Q_{S,2}$. Please give a value $Q_{S,3}$ such that it is equally that S_Z^2 is less than or greater than $Q_{S,3}$."*

Then our three quartiles for τ are $Q_{S,3}^{-1}$, $Q_{S,2}^{-1}$ and $Q_{S,1}^{-1}$. Let τ follow a $Ga(g_\tau, h_\tau)$ distribution. Then we solve $Q_{S,1}^{-1}/Q_{S,3}^{-1} = \tilde{A}_3(g_\tau)/\tilde{A}_1(g_\tau)$ interatively to find the hyperparameter g_τ, where $\tilde{A}_q(a)$ is quartile q for a $Ga(a,1)$ distribution and $h_\tau = \tilde{A}_2(g_\tau)/Q_{S,2}^{-1}$. Using this method, we have chosen suitable numbers for illustration as follows:

$$Q_{z,1} = 0.5, \qquad Q_{z,3} = 1.5,$$
$$Q_{S,1} = 1.25, \qquad Q_{S,2} = 2, \qquad Q_{S,3} = 2.5,$$

and hence

$$m_{A0} = 1, \qquad v_{A0} = 0.55,$$
$$g_\tau = 4.08, \qquad h_\tau = 7.51.$$

- G1 and G2 distributions

Suppose that we imagine a large number of wet days, $Y_1, \cdots, Y_{T_w}$ in, for example, March where T_w is large enough so that the sample mean $\bar{Y} = \frac{1}{T_w}\sum_{j=1}^{T_w} Y_j$ is approximately $\mu = \alpha/\beta$ and the sample variance $S_Y^2 = \frac{1}{T_w-1}\sum_{j=1}^{T_w}\left(Y_j - \bar{Y}\right)^2$ is approximately $\mathrm{Var}(\mathrm{Y_t}|\alpha,\beta) = \alpha/\beta^2$. The sample coefficient of variation $C_Y = S_Y/\bar{Y}$ is then approximately $\alpha^{-1/2}$. Furthermore, the sample mean divided by the sample variance, $D_Y = \bar{Y}/S_Y^2$, is approximately β. In the case of the G1 distribution, we need to assess prior beliefs about the mean $\hat{\eta}_t$ and the shape parameter α. In the case of the G2 distribution, we need prior beliefs for the mean $\hat{\eta}_t$ and the scale parameter β.

To assess the values of the parameters of the prior distribution for the mean rainfall amount, we can ask a sequence of questions to the expert as follows:

Q17 *"Please think about the sample mean $\bar{Y}$. Please give a value such that $\bar{Y}$ is equally likely to be less than or greater than this value."* Let the given value be L.

Q18 *"Please provide a value K such that the events $\bar{Y} < L/K$, $L/K < \bar{Y} < L$, $L < \bar{Y} < KL$ and $KL < \bar{Y}$ are all equally likely."*

Then L is our median for $\bar{Y}$ and the lower and upper quartiles for $\bar{Y}$ are respectively L/K and KL. Hence, our prior median and lower and upper quartiles for $\hat{\eta}_t$ are $\log L$, $\log L - \log K$ and $\log L + \log K$, respectively. Therefore, the prior mean for $\hat{\eta}_t$ is $m_{A0} = \log L$ and the prior variance for $\hat{\eta}_t$ is $v_{A0} = (\log K/0.6745)^2$.

For the G1 distribution, we need to assess beliefs about C_Y. We can use a direct approach to ask the expert as follows:

Q19 *"Please give a value $Q_{C,2}$ such that C_Y is equally likely to be less than or greater than $Q_{C,2}$."* The given value, $Q_{C,2}$ is the median for C_Y.

Q20 *"Suppose you were told that C_Y will be less than $Q_{C,2}$. Please give a number $Q_{C,1}$ such that it is equally that C_Y is less than or greater than $Q_{C,1}$."* The given value, $Q_{C,1}$ is the lower quartile for C_Y.

Q21 *"Suppose you were told that C_Y is greater $Q_{C,2}$. Please give a number $Q_{C,3}$ such that it is equally that C_Y is less than or greater than $Q_{C,3}$."* The given value, $Q_{C,3}$ is the upper quartile for C_Y.

Then our prior quartiles for α are $Q_{C,3}^{-2}$, $Q_{C,2}^{-2}$ and $Q_{C,1}^{-2}$. We suppose that α follows a gamma $Ga(g_\alpha, h_\alpha)$ distribution and then g_α is chosen iteratively solving $(Q_{C,3}/Q_{C,1})^2 = \tilde{A}_3(g_\alpha)/\tilde{A}_1(g_\alpha)$ where $\tilde{A}_q(a)$ is quartile q for a $Ga(a,1)$ distributions, and $h_\alpha = \tilde{A}_2(g_\alpha)Q_{C,2}^2$.

For the G2 distribution, we need to assess beliefs about D_Y. We can use a direct approach to ask the expert as follows:

Q20 *"Please give a value $Q_{D,2}$ such that D_Y is equally likely to be less than or greater than $Q_{D,2}$."* The given value, $Q_{D,2}$ is the median for D_Y.

Q21 *"Suppose you were told that D_Y will be less than $Q_{D,2}$. Please give a number $Q_{D,1}$ such that it is equally that D_Y is less than or greater than $Q_{D,1}$."* The given value, $Q_{D,1}$ is the lower quartile for D_Y.

Q19 *"Suppose you were told that D_Y is greater $Q_{D,2}$. Please give a number $Q_{D,3}$ such that it is equally that D_Y is less than or greater than $Q_{D,3}$."* The given value, $Q_{D,3}$ is the upper quartile for D_Y.

Then our prior quartiles for β are $Q_{D,1}$, $Q_{D,2}$ and $Q_{D,3}$. We suppose that β follows a gamma $Ga(g_\beta, h_\beta)$ distribution and then g_β is chosen iteratively solving $Q_{D,3}/Q_{D,1} = \tilde{A}_3(g_\beta)/\tilde{A}_1(g_\beta)$ where $\tilde{A}_q(a)$ is quartile q for a $Ga(a,1)$ distributions, and $h_\beta = \tilde{A}_2(g_\beta)Q_{D,2}$.

Suppose that we choose a day of the year when we believe that the seasonal effects in (4.19) will cancel out. That is, it is a time of the year when mean rainfall amounts will be expected to be close to the annual average. Then our distribution for $\hat{\eta}_t$ on this day is, in effect, our distribution for η_0. Using the elicitation procedure above we obtain, for illustration, $\eta_0 \sim \mathrm{N}(1.95, 1.06)$. That is $m_{A0} = 1.95$ and $v_{A0} = 1.06$. We will return to the question of beliefs about the seasonal effects, and therefore the Fourier coefficients in (4.19) in Section 4.3.3.3.

For the G1 distribution, we obtain $\alpha \sim \mathrm{Ga}(4.31, 6.75)$. That is $g_\alpha = 4.31$ and $h_\alpha = 6.75$.

For the G2 distribution, we obtain $\beta \sim \mathrm{Ga}(3.87, 44.33)$. That is $g_\beta = 3.87$ and $h_\beta = 44.33$.

4.3.3.2 Prior distributions for the occurrence process

The prior specification of the unknown parameters for the occurrence process is more arduous. However, we can start by considering rainfall occurence on a day of the year when we believe that the seasonal effects in (4.23) and (4.24) will cancel out. That is, it is a time of year when rainfall occurence probabilities will be expected to be close to the annual average. (An alternative would be to consider a randomly chosen day of the year). In this case the Fourier series terms in (4.23) and (4.24) can be disregarded (at least approximately) . We can then have a simple regression model for the occurrence process as

$$\mathrm{logit}(p) = \log\left(\frac{p}{1-p}\right) = \zeta_0 + \zeta_1\psi_1 + \zeta_2\psi_2$$

where $\psi_1 = \hat{\eta}_t - \eta_0$ and $\psi_2 = r_{t-1} - 0.5$.

To determine the prior distribution for unknown parameters, ζ_0, ζ_1 and ζ_2, we need to consult the expert to elicit beliefs about rainfall probabilities in three situations in order to obtain three simultaneous equations. We can specify $r_{t-1} = 1$, that is it rained on the preceding day, or $r_{t-1} = 0$, that it it was dry on the preceding day. We can also specify different values for ψ_1. Given a specified value for r_{t-1} and ψ_1, we can use questions similar to **Q7** and **Q8** in Section 3.5.2.1.1. For example:

Q22 *"Please give your assessment of the probability p_1 that it will rain on the selected day, given that it rained on the preceding day and that $\psi_1 = A_\psi$."* Let the given

value be m_1 which is the prior mean for p_1.

Q23 *"Suppose that you were given a random sample of n_f observations from similar days and that on x_f of these it rained. With these additional data, what would now be your assessment of p_1?"* Let the given value be m_f. Note that actual numerical values are given for n_f and x_f. These are chosen in the light of the value given for m_0, so that x_f/n_f is sufficiently different from m_0 and an initial judgement about the expert's certainty about p. Larger values of n_f are required if the expert has greater prior precision.

As in 3.5.2.1.1, the answers can be used to choose a distribution for p_1. Ideally we would use a logit-normal distribution but, for elicitation purposes, it is reasonable to use a beta distribution as an approximation. The quartiles of this distribution can then be transformed to quartiles of $\text{logit}(p_1)$. We give $\text{logit}(p_1)$ a normal distribution with mean given by the average of the first and third quartiles and standard deviation given by the difference between the third and first quartiles divided by 1.349.

Using Question **Q22**, we have chosen three different (A_ψ, r_{t-1}) pairs, $(2, 0)$, $(0.1, 1)$ and $(-1, 1)$, for illustration. This leads to the following three equations.

$$\begin{aligned} K_{p1} &= \text{logit}(p_1) = \zeta_0 + 2\zeta_1 - 0.5\zeta_2 \\ K_{p2} &= \text{logit}(p_2) = \zeta_0 + 0.1\zeta_1 + 0.5\zeta_2 \\ K_{p3} &= \text{logit}(p_3) = \zeta_0 - \zeta_1 + 0.5\zeta_2 \end{aligned}$$

where

$$\begin{aligned} &\Pr(p_1 < 0.407) = \Pr(p_1 > 0.489) = 0.25 \\ &\Pr(p_2 < 0.459) = \Pr(p_2 > 0.541) = 0.25 \\ &\Pr(p_3 < 0.249) = \Pr(p_3 > 0.332) = 0.25. \end{aligned}$$

Suppose that $K_{p1} \sim N(m_{K_{p1}}, \sigma^2_{K_{p1}})$. Then from the specification for p_1 we have

$$\begin{aligned} &\Pr(K_{p1} < \text{logit}\, 0.407 = -0.375) = 0.25 \\ &\Pr(K_{p1} > \text{logit}\, 0.489 = -0.045) = 0.25 \end{aligned}$$

and so, from the properties of the normal distribution, we have that

$$m_{K_{p1}} - 0.674\sigma_{K_{p1}} = -0.375, \qquad m_{K_{p1}} + 0.674\sigma_{K_{p1}} = 0.045.$$

This leads to $m_{K_{p1}} = (-0.375 - 0.045)/2 = -0.210$ and $\sigma^2_{K_{p1}} = [(0.375 - 0.045)/1.35]^2 =$

0.060. Similar calculations are given to K_{p2} and K_{p3} where we have

$$K_{p1} \sim N(-0.21, 0.06)$$
$$K_{p2} \sim N(0, 0.06)$$
$$K_{p3} \sim N(-0.9, 0.09).$$

This gives marginal means and variances for K_{p1}, K_{p2} and K_{p3}. We might wish to assess nonzero covariances between them. This can be done by asking questions about the effect of a hypothetical future sample in one scenario on beliefs in the other scenarios. For illustration here, however, we treat K_{p1}, K_{p2} and K_{p2} as mutually independent.

Let $\zeta = (\zeta_0, \zeta_1, \zeta_2)'$ and $\mathbf{K}_p = (K_{p1}, K_{p2}, K_{p2})'$. Hence $\mathbf{K}_p = M_{K\zeta}\zeta$ where

$$M_{K\zeta} = \begin{pmatrix} 1 & 2 & -0.5 \\ 1 & 0.1 & 0.5 \\ 1 & -1 & 0.5 \end{pmatrix}.$$

By solving the linear equations, we obtain

$$\begin{pmatrix} \mathrm{E}(\zeta_0) \\ \mathrm{E}(\zeta_1) \\ \mathrm{E}(\zeta_2) \end{pmatrix} = M_{K,\zeta}^{-1}\mathrm{E}(\mathbf{K}_\mathrm{p}) = \begin{pmatrix} -0.96 \\ 0.82 \\ 1.76 \end{pmatrix}$$

and

$$\begin{pmatrix} \mathrm{Var}(\zeta_0) & \mathrm{Cov}(\zeta_0, \zeta_1) & \mathrm{Cov}(\zeta_0, \zeta_2) \\ \mathrm{Cov}(\zeta_0, \zeta_1) & \mathrm{Var}(\zeta_1) & \mathrm{Cov}(\zeta_1, \zeta_2) \\ \mathrm{Cov}(\zeta_0, \zeta_2) & \mathrm{Cov}(\zeta_2, \zeta_2) & \mathrm{Var}(\zeta_2) \end{pmatrix} = M_{K,\zeta}^{-1}\mathrm{Var}(\mathbf{K}_\mathrm{p})(\mathrm{M}_{\mathrm{K},\zeta}^{-1})'$$
$$= \begin{pmatrix} 0.11 & -0.10 & -0.25 \\ -0.10 & 0.12 & 0.29 \\ -0.25 & 0.29 & 0.77 \end{pmatrix}.$$

4.3.3.3 Prior distributions for the Fourier series coefficients

Provided that we are willing, at each frequency f, to give the amplitude A_f and phase φ_f independent prior distributions with the distribution of A_f being Rayleigh and the distribution of φ_f being uniform, then we can construct the prior distributions for the Fourier coefficients based on the discussion in Section 4.2.5.1. At each frequency, this only requires elicitation of the parameter, $\tilde{v}_f$, of the Rayleigh distribution. This can be done by eliciting the mean of the amplitude, that is the mean of the difference between the

maximum and minimum in the annual cycle. Then

$$\tilde{v}_f = \sqrt{\frac{2}{\pi}}\ \mathrm{E}(\mathrm{A_f}).$$

Following Section 4.2.5.1, we then see that the Fourier coefficients, a_f and b_f (or c_f and d_f), have independent and identical normal distributions with

$$a_f \sim \mathrm{N}(0, \tilde{v}_f^2).$$

The amplitude would typically be expected to become smaller as the frequency increases. Therefore, for example, we can then set

$$\tilde{v}_f^2 = \frac{2}{\pi f}$$

for $f = 1, 2, 3$. The prior elicitations for c_f and d_f are performed similarly as for a_f and b_f. The full prior specifications of the Fourier coefficients for the amount and occurrence processes are then given as follows:

$$a_f \sim N\left(0, \frac{2}{\pi f}\right), \qquad b_f \sim N\left(0, \frac{2}{\pi f}\right),$$
$$c_f \sim N\left(0, \frac{2}{\pi f}\right), \qquad d_f \sim N\left(0, \frac{2}{\pi f}\right),$$

for $f = 1, 2, 3$.

4.3.3.4 Fitting the model

To generate the posterior sample, the RJAGS package (Plummer, 2012) was used and this was implemented in R software (R Development Core Team, 2008). We will use the data augmentation technique as elaborated in Section 2.3.5 to simplify the calculation where the potential rainfall amount is entered as "NA" for any day when rain was absent. The MCMC algorithm was run for 1000 iterations and these were discarded as a burn-in. Then, the subsequent 20000 iterations were preserved as the posterior samples. The computing time that was required to obtain 20000 posterior samples was around 5 to 6 hours for all models by using R software on a 3.40GHz Ergo Desktop AS4 All-in-One with Intel Core i7-3770 processor and 8 Gbytes of random-access memory. Based on our analysis, 20000 iterations are deemed sufficient to obtain realisations from the posterior distribution since the chain has sufficiently converged after the initial burn-in period. Convergence was assessed by the visual inspection of trace plots which indicates that the chain has converged very well, and that there is good mixing for all parameters in both

processes. The trace and density plots for some parameters of the lognormal, G1 and G2 distributions are shown in Figure 4.7. By combining the prior beliefs and the data, the uncertainty associated with the unknown parameters has been reduced. For example, there is a reduction in the variability of parameter τ when comparison is made between the prior and posterior distributions. The full summaries of the posterior means and standard deviations of the unknown parameters for both processes are presented in Tables 4.1 and 4.2.

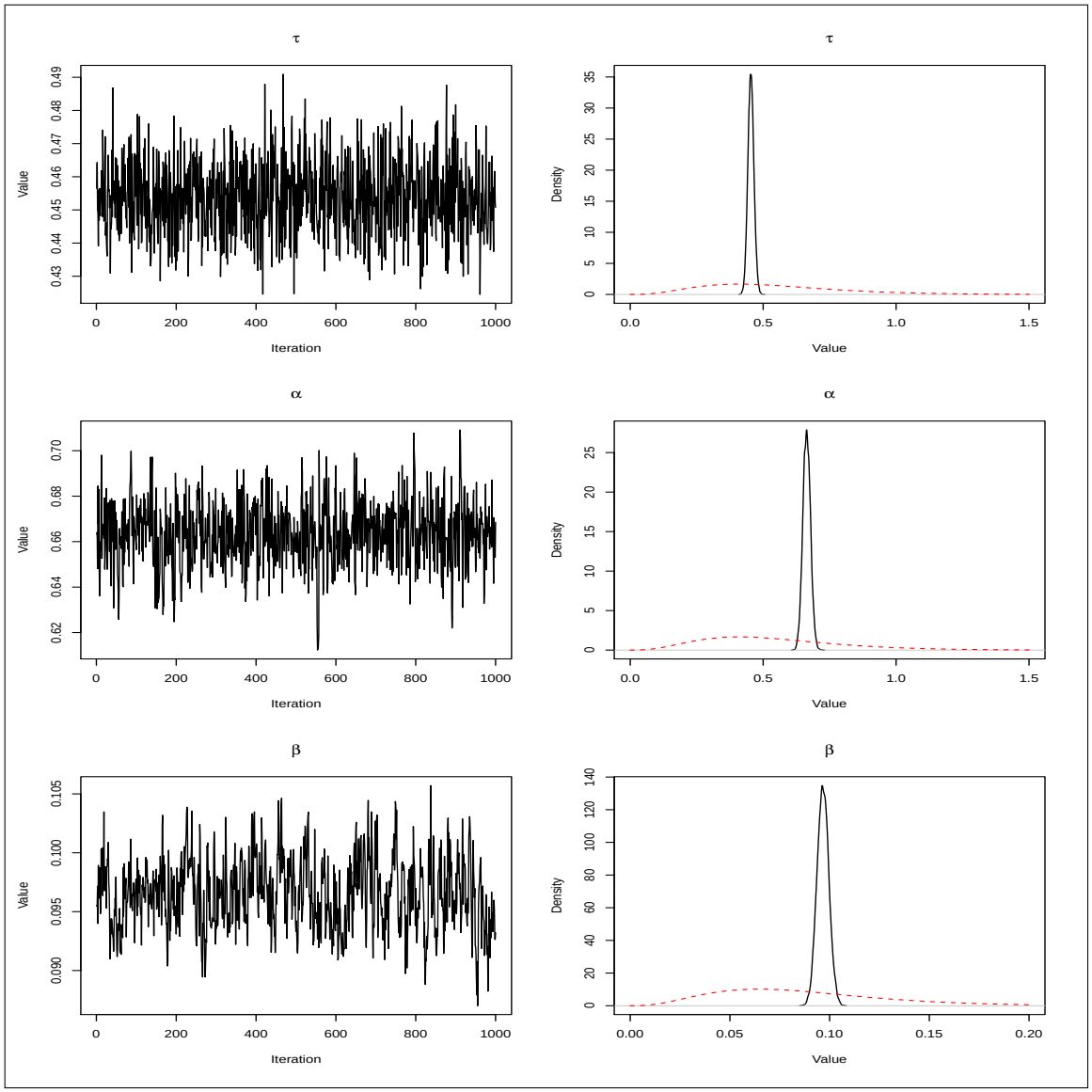


Figure 4.7: The trace (for the first 1000 iterations) and density plots for parameters τ, α and β

The posterior mean for the mean of the potential rainfall amounts μ_t is shown in Figure 4.8 for the three different amount distributions. Obvious differences can be seen

Lognormal distribution (LN)				
Parameter	Prior mean	Prior SD	Posterior mean	Posterior SD
τ	0.54	0.27	0.45	0.01
η_0	1	0.74	1.04	0.03
a_1	0	0.80	-0.21	0.04
a_2	0	0.56	-0.07	0.04
a_3	0	0.46	0.05	0.04
b_1	0	0.80	-0.13	0.04
b_2	0	0.56	0.13	0.04
b_3	0	0.46	-0.08	0.04
Gamma 1 distribution (G1)				
Parameter	Prior mean	Prior SD	Posterior mean	Posterior SD
α	0.64	0.31	0.66	0.01
η_0	1.95	1.03	1.94	0.02
a_1	0	0.80	-0.12	0.03
a_2	0	0.56	-0.03	0.03
a_3	0	0.46	0.01	0.03
b_1	0	0.80	-0.16	0.03
b_2	0	0.56	0.05	0.03
b_3	0	0.46	-0.09	0.03
Gamma 1 distribution (G2)				
Parameter	Prior mean	Prior SD	Posterior mean	Posterior SD
β	0.09	0.04	0.10	0.003
η_0	1.95	1.03	1.94	0.02
a_1	0	0.80	-0.10	0.02
a_2	0	0.56	-0.04	0.02
a_3	0	0.46	0.03	0.02
b_1	0	0.80	-0.06	0.02
b_2	0	0.56	0.07	0.02
b_3	0	0.46	-0.04	0.02

Table 4.1: The prior and posterior means with standard deviations (SD) of the unknown parameters for three different amount distributions

between the gamma and lognormal distributions. The fitted values for the G1 and G2 distributions indicate that they are close to each other. In contrast, the fitted values for the lognormal distribution are consistently above those of the G1 and G2 distributions. We observe that the fitted values produce smooth plots which indicate the seasonal variabilities over the year. One may note, for example, that the maximum level of potential rainfall amounts for all three different distributions is between July and October, while the lowest level is roughly between November and February for the lognormal and G2 distributions. However, for the G1 distribution, the lowest level of potential rainfall amounts is possibly between January and March. Figure 4.9 illustrates the posterior mean of the medians potential rainfall amount from three different distributions. Once again, we can see obvious

Occurrence process								
Parameter	Prior mean	Prior SD	Posterior mean			Posterior SD		
			LN	G1	G2	LN	G1	G2
ζ_0	-0.96	0.33	-0.57	-0.57	-0.57	0.02	0.02	0.02
ζ_1	0.82	0.35	0.52	0.55	0.54	0.30	0.30	0.30
ζ_2	1.76	0.88	1.45	1.45	1.45	0.05	0.04	0.05
c_1	0	0.80	0.39	0.35	0.34	0.07	0.05	0.05
c_2	0	0.56	-0.11	-0.13	-0.13	0.05	0.04	0.04
c_3	0	0.46	0.01	0.03	0.02	0.04	0.04	0.04
d_1	0	0.80	0.06	0.09	0.04	0.05	0.06	0.04
d_2	0	0.56	-0.31	-0.27	-0.28	0.06	0.04	0.04
d_3	0	0.46	0.06	0.06	0.04	0.05	0.05	0.04

Table 4.2: The prior and posterior means with standard deviations (SD) of the unknown parameters for occurrence process

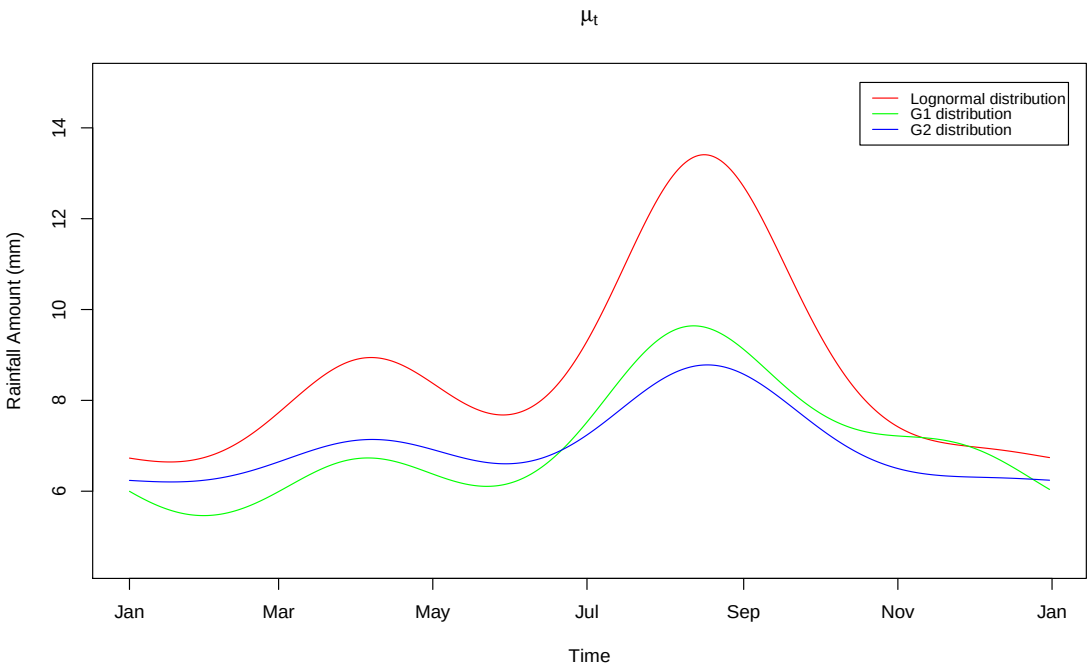


Figure 4.8: The posterior mean of the mean potential rainfall amount from three different distributions

differences between the gamma and lognormal distributions where the medians of the G1 and G2 distributions are now higher than the lognormal distribution. Hence, we can conclude that the reason why the results for the lognormal distribution are different is because of the difference in the shape of the distribution.

The posterior means for the transition probabilities p_{01} and p_{11} of the occurrence

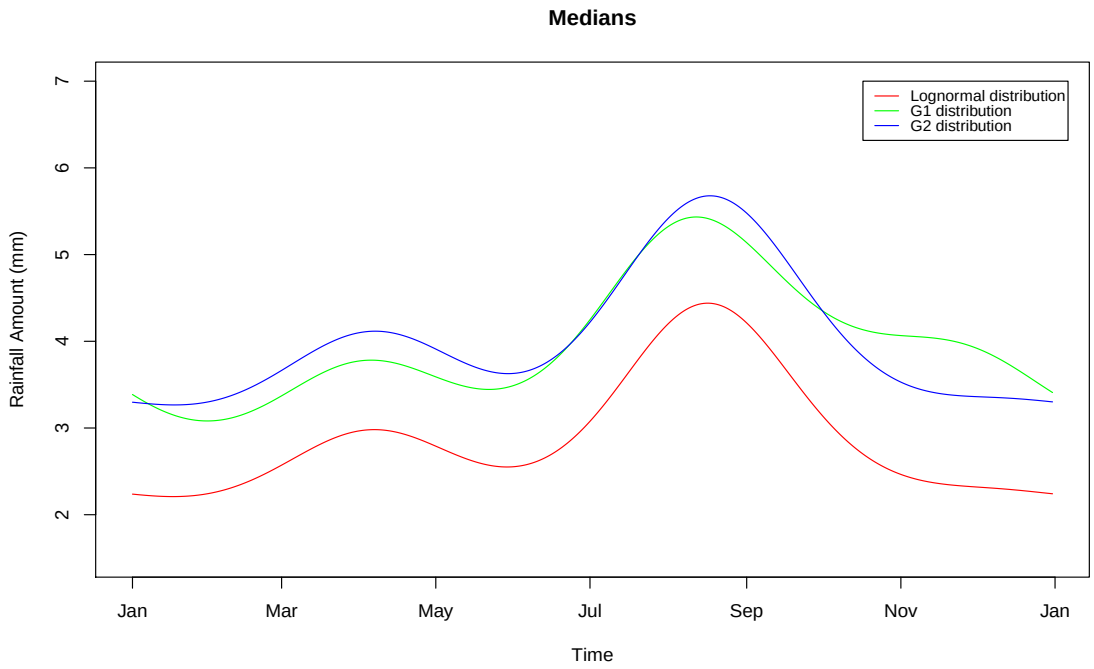


Figure 4.9: The posterior mean of the median potential rainfall amount from three different distributions

process are distinctly different (Figure 4.10) and these are similar for all three distributions of the amount process. We notice that if rain was absent on the previous day, the rainfall probability for the present day is only between 0.2 and 0.3. In contrast, typically the rainfall probability will be more than 0.5 if rain occurred on the previous day. Both plots are strongly indicative of the relative dependence of the rainfall probability on the previous day's rainfall occurrence. The fitted values also produce smooth plots for both conditional probabilities over the year where the minimum rainfall probability is between June and September. Surprisingly, this is in contradiction to the results obtained for the amount process since the rainfall amounts in this period are higher than in the other period. This can be explained by the fact that when rain occurs, the duration and intensity of rain (heavy rain) are higher than for precipitation in the other periods. In addition, the unconditional rainfall probabilities can also be calculated using Equation (4.22). Figure 4.11 shows the unconditional rainfall probabilities over the year where these probabilities are mostly below 0.5. This indicates that the rain was absent on most days at the Urbino rain gauge station and this explains why our data have many zeros.

We also generated the posterior values for the predictive distribution of Y over the year for the three different amount distributions. We can use these predictive values to compute the posterior predictive means for the monthly average amount of rainfall

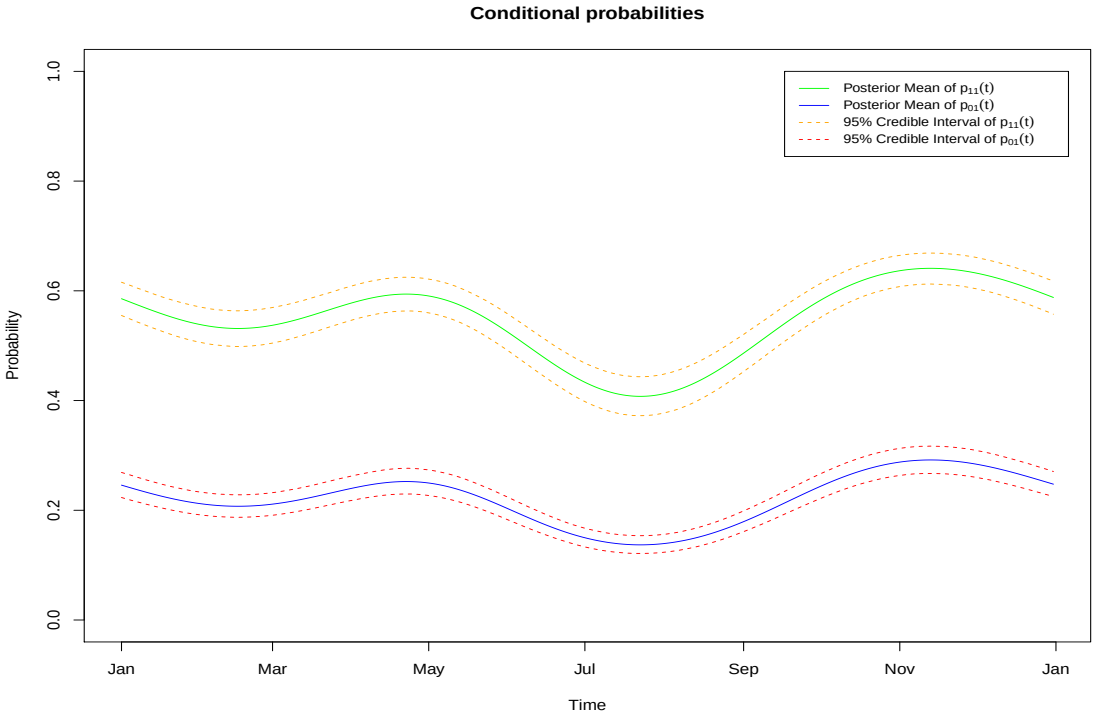


Figure 4.10: The posterior mean of conditional probabilities, p_{01} and p_{11}

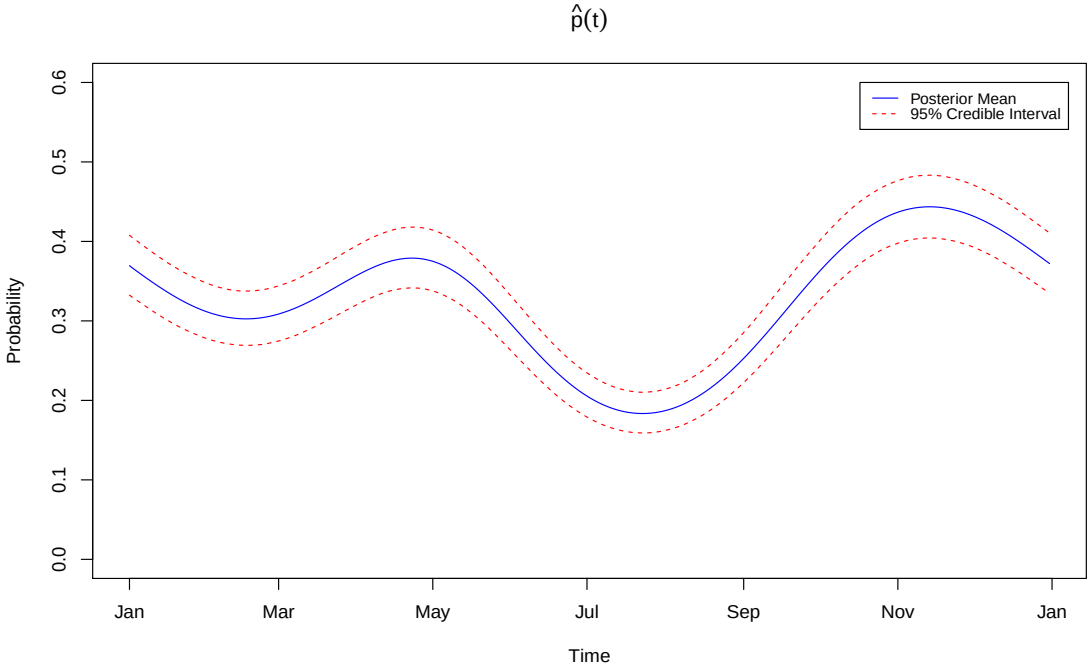


Figure 4.11: The posterior mean of unconditional probabilities

as depicted in Figure 4.12. The monthly average amount is obtained by computing the average of posterior predictive means for every month. Then, we may use this quantity to compare the posterior predictive mean with the observed data. We may conclude that both the posterior predictive mean and the observed data have similar patterns over the year especially for the G1 and G2 distributions which very closely resemble the observed data. However, the predictive means of the monthly rainfall amounts when the lognormal distribution is used seem always greater than those obtained by the G1 and G2 distributions. A possible explanation is that the shape parameter for the gamma distribution is less than 1 so the density does not go to zero as the rainfall amount goes to zero but, in the case of the lognormal distribution, the density always goes to zero as the amount goes to zero. The shape of the implied gamma distribution is therefore different from any shape which can be achieved with a lognormal distribution, especially close to the origin. So it is harder for the lognormal distribution to provide a good fit to that part of the distribution and this will affect the posterior distribution of the parameter values. On average, the lowest level of monthly rainfall amounts is in July where the highest is around November for the G1 and G2 distributions and in April for the lognormal distribution.

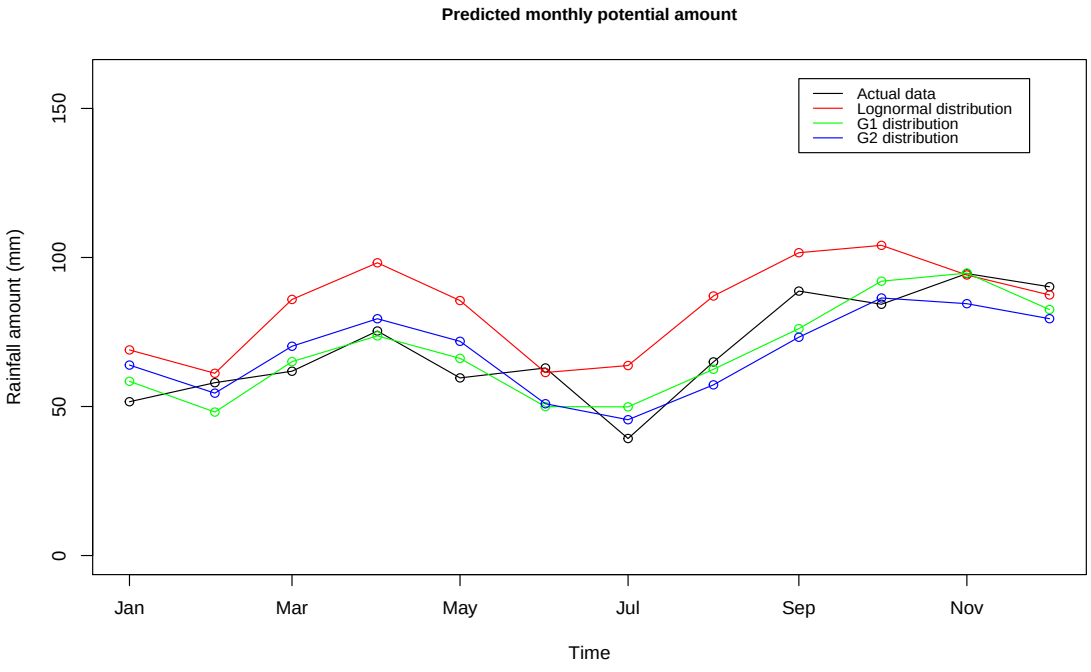


Figure 4.12: Posterior mean for predictive distribution of monthly potential rainfall amounts

4.3.3.5 Residuals

In this section, we will re-examine the assumption we have made in this model where Y_t variables are independent given model parameters. To re-examine this assumption, we need to do some transformation for y_t. Let $u_t = G_t(y_t)$ where $G_t(.)$ is the cumulative distribution function (cdf) of Y_t. Then we transform u_t to a normally distributed $\hat{d}_t = \Phi^{-1}(u_t)$ where $\Phi(.)$ is the standard normal cdf and $\Phi^{-1}(.)$ is its inverse. Then we can plot the autocorrelation function (ACF) and partial autocorrelation function (PACF) for these residuals. Based on Figure 4.13, the ACF and PACF plots show some evidence of

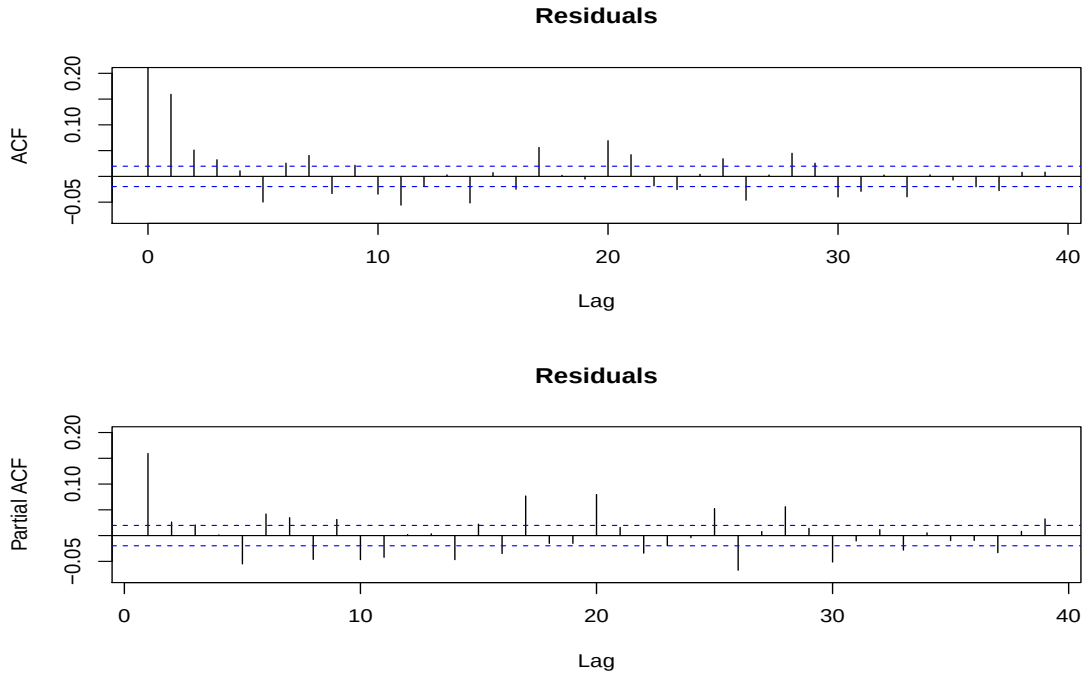


Figure 4.13: The ACF and PACF plots for residuals

autocorrelation. Hence, the assumption of conditional independence between Y_t and Y_s where $s \neq t$, given μ_t, μ_s, for this model is not valid. Therefore, we will describe a new model which allows for temporal correlation in the amount process in Section 4.4.

4.3.3.6 Zero/positive distribution

Another goal of this study is to investigate the changes in the shape and scale of the probability density function for the mixed distribution. As discussed in Chapter 3, Section 3.4, it is difficult to construct a graph for the mixed distribution which contains both positive and zero components. Nonetheless, we can visualise the mixed distribution by representing the pdf for the non-zero rainfall amounts with a density curve and the "lol-

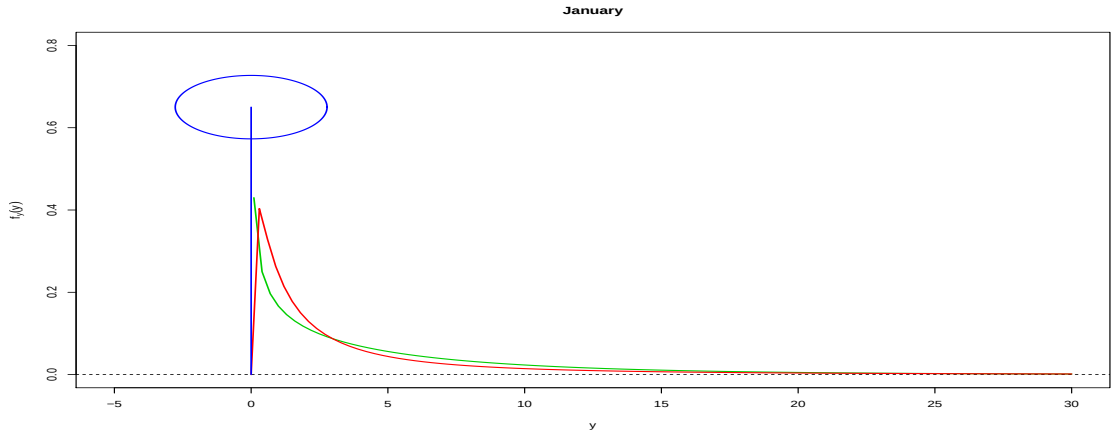
lipop" component at zero for the zero-rainfall probability as shown in Figure 3.2. These two components are then linked together and the area of the ellipse is equal to the probability of zero rainfall. For illustrative purposes, we will plot the graphs on a monthly basis to monitor the changes in the rainfall pattern.

Figures 4.14 until 4.17 represent the changes in the zero-positive plots from January until December for two different amount distributions. It is evident from the zero-positive plots that the G1 and G2 distributions are very similar to each other. Hence, we will portray them as a single gamma distribution and this will be compared with the lognormal distribution. With respect to the gamma distribution, the distribution become more "L" shaped when the shape parameter is less than 1 as $f_y(y)$ approaches ∞. In contrast, $f_y(y)$ approaches 0 when $y = 0$ for the lognormal distribution. Overall, there are no significant changes in the shape of the distributions for the amount process from month to month. The only significant change is in the scale parameter, where a relatively decreasing magnitude of the rainfall amount was detected in August and September. We also observed that the sizes of the ellipses are bigger in July and September due to the lack of rainfall occurrence.

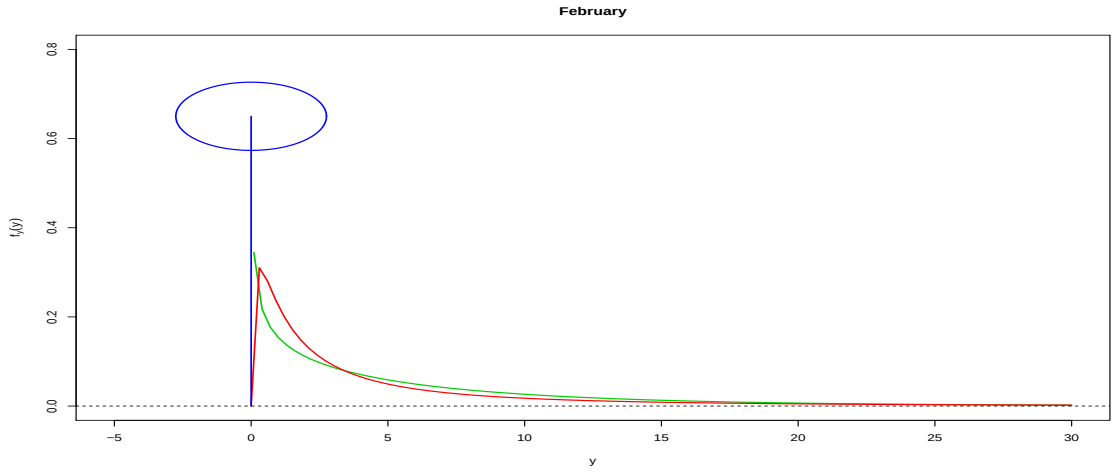
It is also important to note that the "lollipop" plots are based on the ellipse for $\Pr(R_t = 0)$ and the curve for $\Pr(R_t = 1)f(Y_t|R_t = 1) = \Pr(R_t = 1)f(W_t|R_t = 1)$. In this model, R_t and Y_t are conditionally independent given μ_t and this means that, given μ_t, $f(Y_t|\mu_t, R_t = 1)$ is the specified gamma pdf. However, this would not work if we allow $\Pr(R_t = 1)$ to be dependent on Y_t rather than on μ_t. In this case, $f(Y_t|R_t = 1)$ and $f(W_t|R_t = 1)$ would still be equal but they would not be the specified gamma distribution. In fact, they would not be a gamma distribution at all.

4.3.4 Conclusion

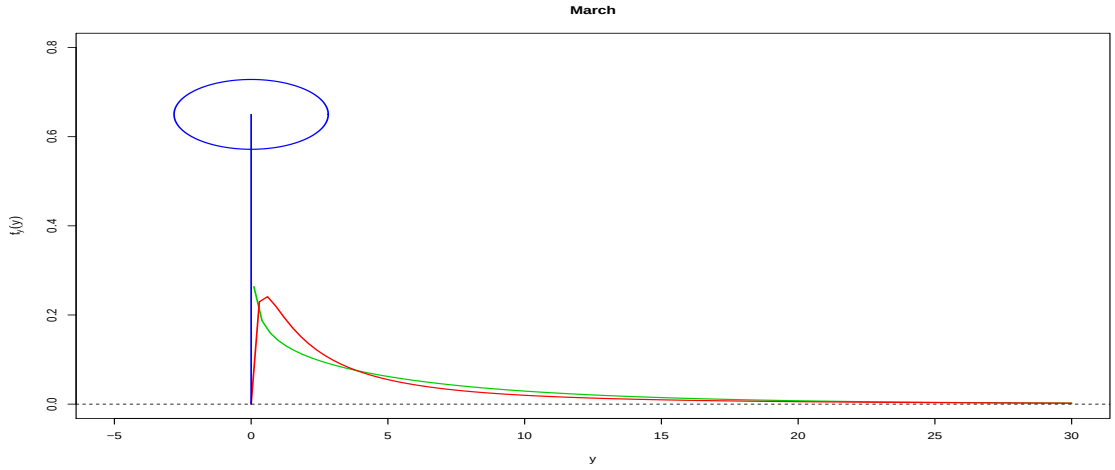
From these results, we can conclude that the gamma distribution is a better choice for the modelling of the amount process in the Italian rainfall data as shown in Figure 4.12. Our findings are corroborated by several results from previous studies which verify that the gamma distribution fits the data slightly better than other distributions. According to Fernandes *et al.* (2009), they discovered that the gamma distribution is a preferable alternative in terms of its predictive properties to either the exponential or lognormal distributions. Bruno *et al.* (2014) also established that the gamma distribution is superior to the log-normal distribution in terms of modelling spatiotemporal rainfall data at San Pietro Capofiume, Brazil. There is also evidence from our study that the shape parameter does not noticeably change over the year. In contrast, the scale parameter shows conspicuous shifts when comparisons were made from month to month. This might suggest that the parameterisation used in the G1 distribution is more appropriate to represent the Italian daily rainfall than that in G2 distribution.



(a)

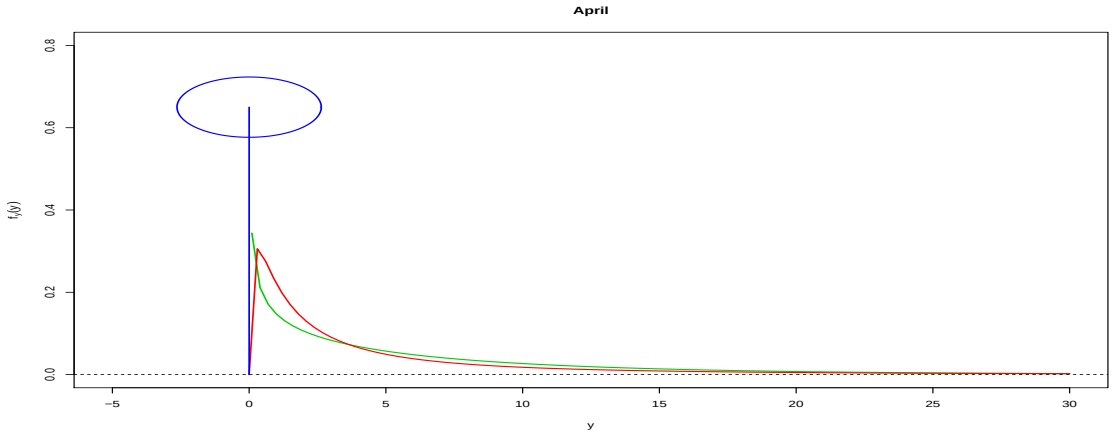


(b)

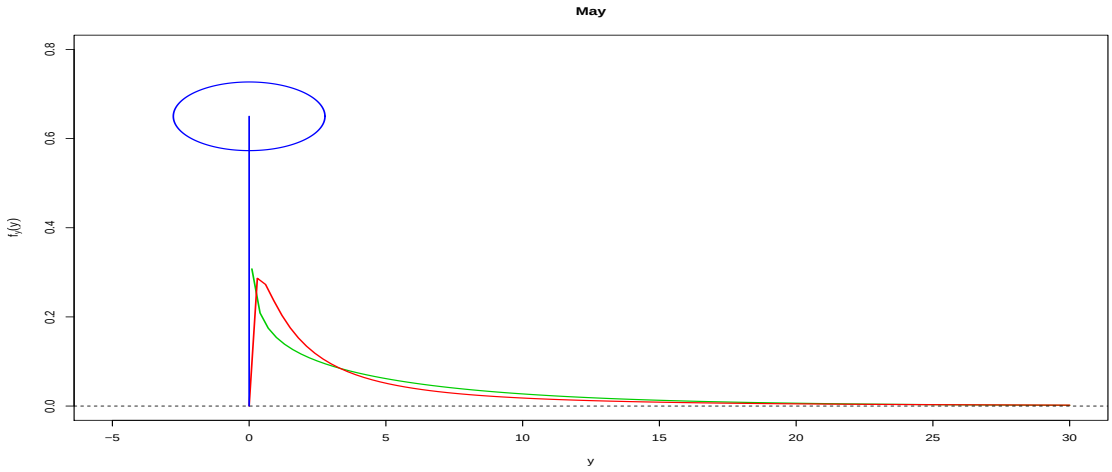


(c)

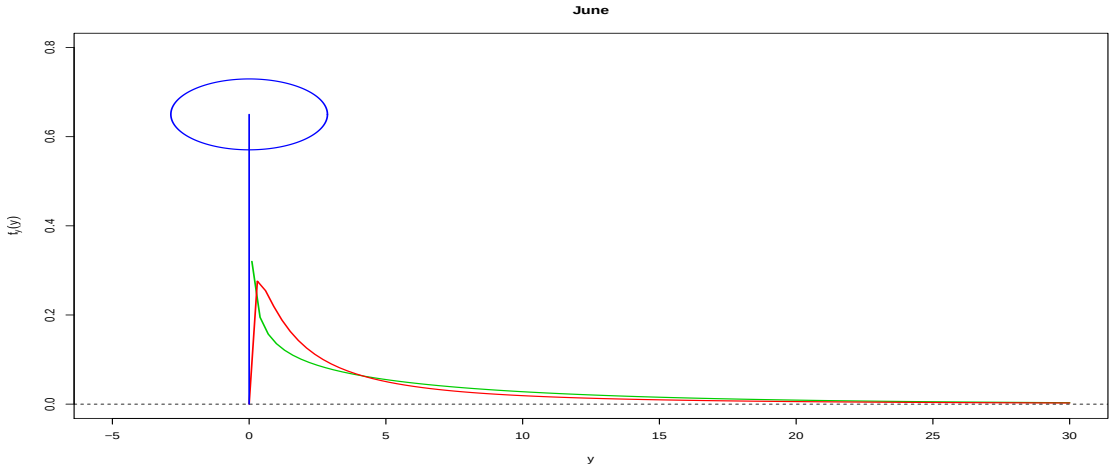
Figure 4.14: Zero-positive plots for gamma (green) and lognormal (red) distributions from January until March



(a)

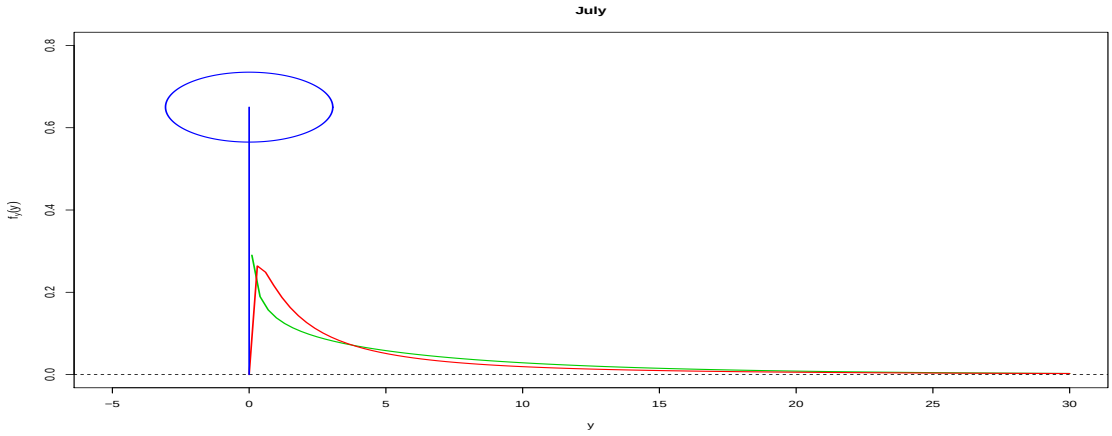


(b)

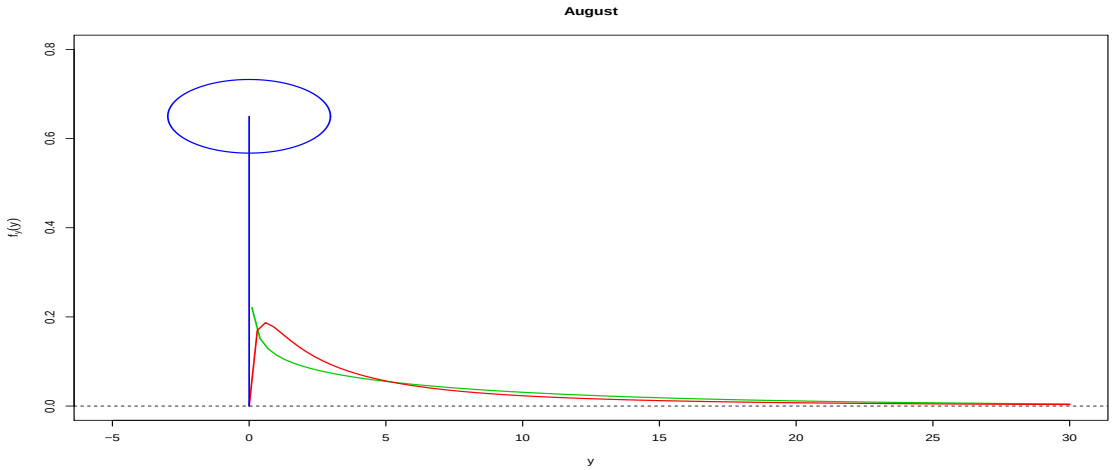


(c)

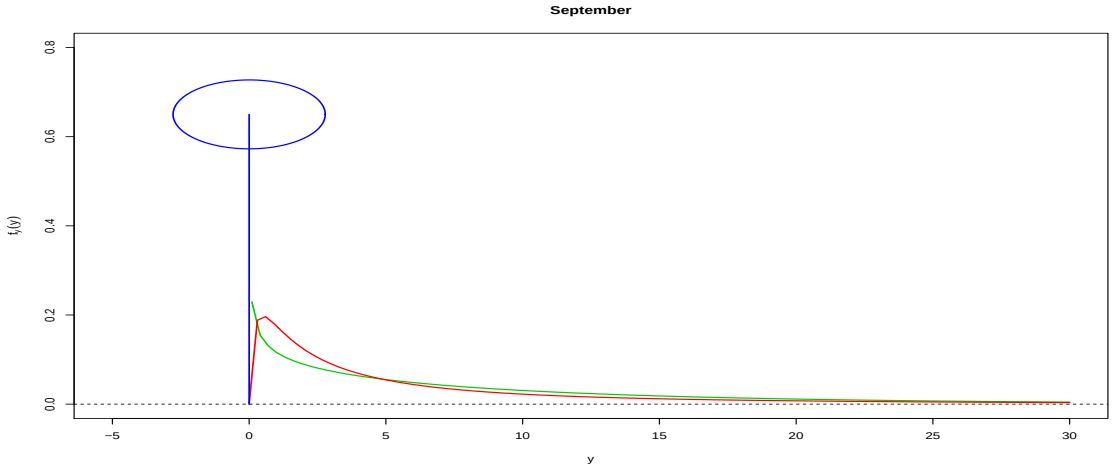
Figure 4.15: Zero-positive plots for gamma (green) and lognormal (red) distributions from April until June



(a)

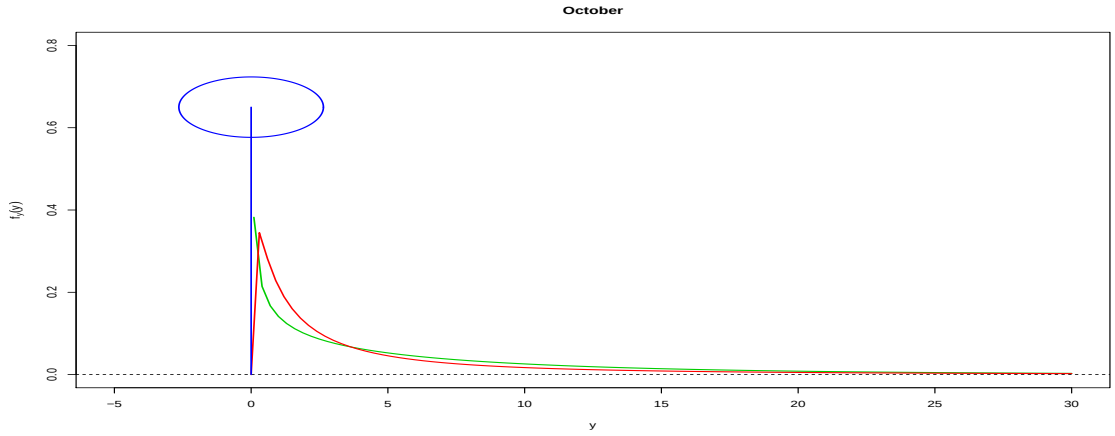


(b)

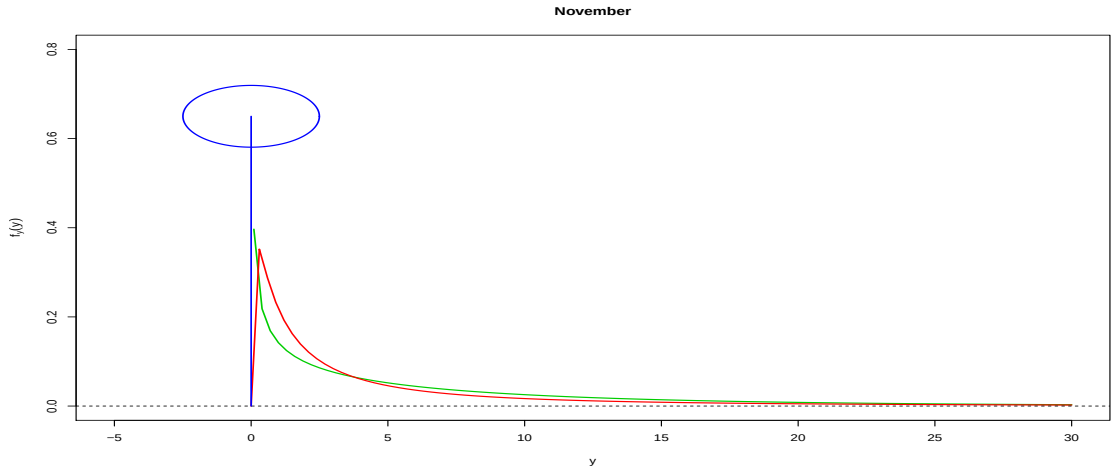


(c)

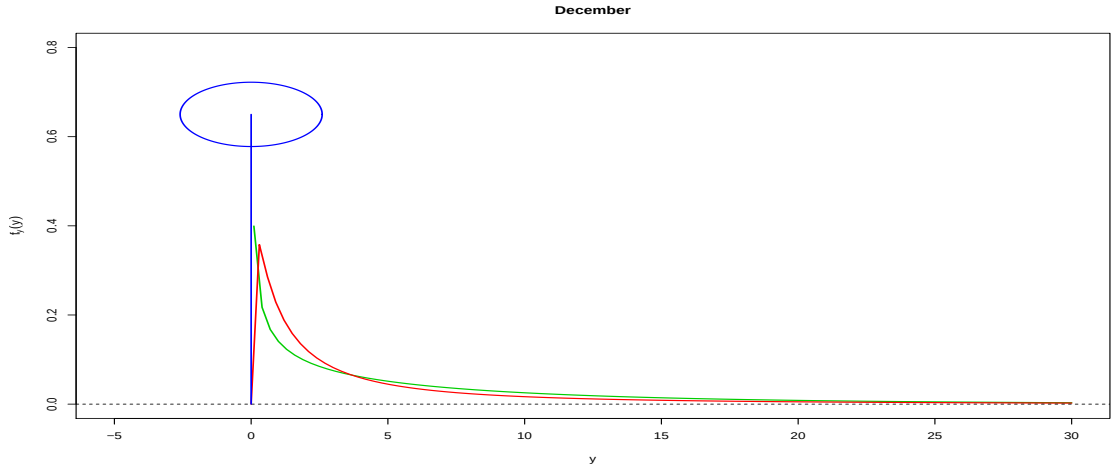
Figure 4.16: Zero-positive plots for gamma (green) and lognormal (red) distributions from July until September



(a)



(b)



(c)

Figure 4.17: Zero-positive plots for gamma (green) and lognormal (red) distributions from October until December

To find an appropriate model for the occurrence and amount processes in daily rainfall data is the primary objective of this study. For the amount process, we have investigated and compared three different distributions to identify which distribution has the best fit to our data. Our findings show that the G1 and G2 distributions fitted the data slightly better than the lognormal distribution, as evident from the more accurate posterior mean daily rainfall and posterior predictive values. The reasons for these findings have already been discussed in Section 4.3.3.4.

Our next objective is to extend our model by incorporating the atmospheric circulation patterns known as the Lamb weather types (LWTs). The LWTs will be incorporated into the models for both the amount and occurrence processes via the mean of amount distribution and rainfall probabilities.

4.4 Application to daily rainfall in Britain

In the previous section, we successfully demonstrated the application of our approach to Italian daily rainfall by making the occurrence process dependent on the information from the amount process. For the British daily rainfall application, we will extend the previous model by incorporating the atmospheric circulation patterns. There is a common problem, especially in the hydrology and agriculture fields, which concerns how to downscale synoptic atmospheric variables to the small-scale daily precipitation patterns. One of the solutions that we can use to handle this problem is by using statistical downscaling. Statistical downscaling provides a solution by using a weather state model where each day is assigned to one of a small number of weather states based on the atmospheric circulation patterns and typically these weather states are observable. Precipitation is then modelled within each weather state which is generally assumed to evolve according to some temporal process. The weather state model is a part of the statistical downscaling model which was introduced by Hay *et al.* (1991).

The classification of atmospheric circulation patterns can be represented by weather types (WTs) (Yarnal, 1993; Barry & Chorley, 2003; Jones *et al.*, 2013). Notable examples of WTs include the Grosswetterlagen types for Central Europe (Hess & Brezowsky, 1969) and the Lamb weather types (LWTs) for the British Isles (Lamb, 1972). In this application, the LWTs will be incorporated into the British daily rainfall model. Heaps *et al.* (2015) and Germain (2010) used the objective LWT scheme for characterising atmospheric circulation patterns by combining them with the weather state process using non-homogeneous hidden Markov models (NHMMs) for UK winter rainfall. We propose to build an alternative model to Heaps *et al.* (2015) and Germain (2010) that allows the rainfall amounts and occurrences to be conditional upon the values of some synoptic atmospheric variables. However, instead of using weather state models, we will apply LWTs directly to our model because we believe that LWTs have a critical role in determining rainfall amounts and occurrences over the British Isles. In addition, we are also interested to investigate the relationship between the LWTs and local daily rainfall. For the first phase, we will only focus on a single site before extending the model to multiple sites around the United Kingdom in Chapter 5.

4.4.1 Data

For this application, we choose daily rainfall data from Darlington South Park weather station located at Darlington in County Durham, North East England. The exact location of this station is at $54°30'54.0''N$ latitude and $1°33'29.4''W$ longitude as depicted in Figure 4.18. The location is close to the River Tees that represents the natural boundary between

Durham and Yorkshire. The data cover the period from January 1966 to December 1985 with a total of 7305 observations, which include five leap years. The datasets were obtained from the United Kingdom (UK) Meteorological Office integrated data archive system (MIDAS)(Met Office, 2017).

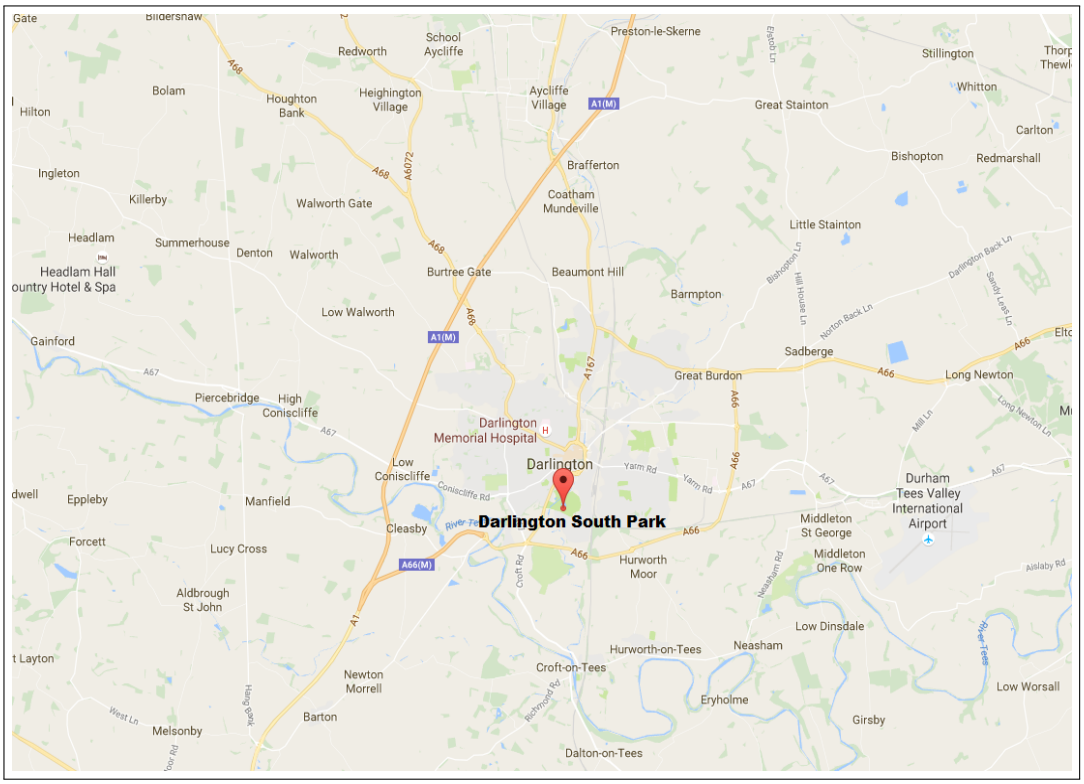


Figure 4.18: Location of Darlington South Park weather station

4.4.2 Atmospheric data: Lamb weather types

In UK climatology, the atmospheric circulation patterns are commonly represented by the objective Lamb Weather Types (LWTs) which are a popular choice to represent the atmospheric variables. The original objective of LWTs as established by Lamb (1972), was to characterise weather types using a subjective classification scheme based on daily synoptic charts and atmospheric flows over the British Isles. Subsequently, Jenkinson & Collinson (1977) created an objective scheme to identify the daily pattern of atmospheric circulation in terms of the LWTs using daily grid-point mean sea level pressure data. These data give enough information to evaluate the dominant direction and speed of the flow including its vorticity which are usually associated with cyclonic or anticyclonic weather patterns. By using this information, the direction and synoptic type of the surface flow over the British Isles can be categorised on any particular day as a specific LWT.

Using current state-of-the-art technology, Jones *et al.* (2013) reanalysed the objective LWT scheme to produce a more consistent weather classification system. This system is based on the new extended reanalysis (20CR) developed by Compo *et al.* (2011) and the National Centers for Environmental Prediction (NCEP) reanalyses by Kalnay *et al.* (1996), which have been used to classify the weather types over the British Isles since 1880.

In general, the Jenkinson classification scheme (Jenkinson & Collinson, 1977) has 27 objective LWTs which are composed of eight main directional types: North-East (NE), East (E), South-East (SE), South (S), South-West (SW), West (W), North-West (NW) and North (N); three non-directional types: anticyclonic (A), cyclonic (C) and unclassifiable (U); and 16 hybrid types, which are the combination of the eight directional types with the anticyclonic and cyclonic non-directional types. The full classification scheme of the objective LWTs is presented in Table 4.3.

Label	**Objective LWT**	**Label**	**Objective LWT**	**Label**	**Objective LWT**
1	A	27	U	18	C
2	ANE	10	NE	19	CNE
3	AE	11	E	20	CE
4	ASE	12	SE	21	CSE
5	AS	13	S	22	CS
6	ASW	14	SW	23	CSW
7	AW	15	W	24	CW
8	ANW	16	NW	25	CNW
9	AN	17	N	26	CN

Table 4.3: Labelling of the objective Lamb weather types

The LWTs can be labelled according to the direction of air flow and the vorticity. For example, if the direction of air flow comes from the west and the vorticity is close to zero, this is categorised as westerly type (W). If the vorticity is strongly positive or negative, this LWT is classified as either pure cyclonic or anticyclonic, respectively. The LWTs are classified as hybrid types if the vorticity is moderately positive or negative, in combination with any direction of air flow. For instance, if the direction of air flow comes from the west and the vorticity is positive, then it will be classified as cyclonic westerly (CW) or anticyclonic westerly (AW) LWT. Finally, the LWT is deemed unclassifiable if the atmospheric circulation pattern is too complicated to be classified as of any other type. Figure 4.19 shows the frequencies of LWTs occurrence from 1966 until 1985 where the highest frequency is purely anticyclonic (type 1) followed by purely cyclonic (type 18) and purely westerly (type 15).

The proportion of wet days and the mean daily rainfall amount on wet days by LWT for Darlington South Park Station are shown in Figure 4.20. Clear patterns can be seen

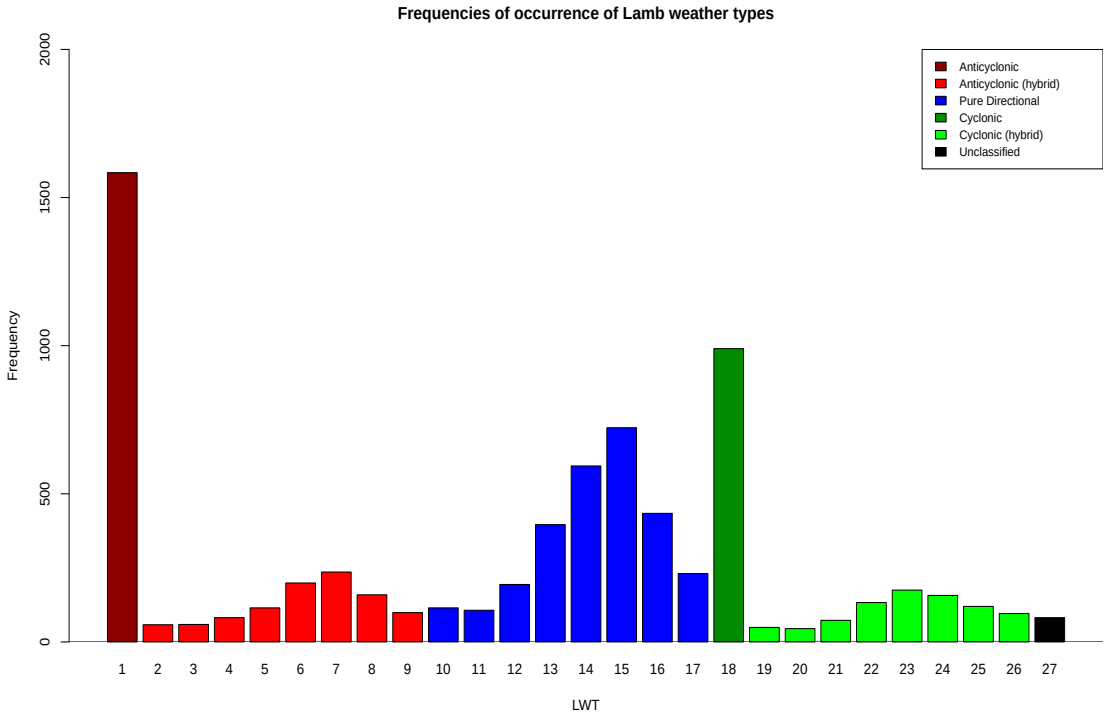


Figure 4.19: Frequencies of Lamb Weather types occurrence from 1966 until 1985

in the proportion of wet days, with lower proportions being associated with anticyclonic types (1-9) and higher proportions with cyclonic types (18-26). Similarly, based on the mean daily rainfall amounts on wet days across LWTs, the pattern shows that high (low) level rainfall amounts are typically associated with cyclonic (anticyclonic) types.

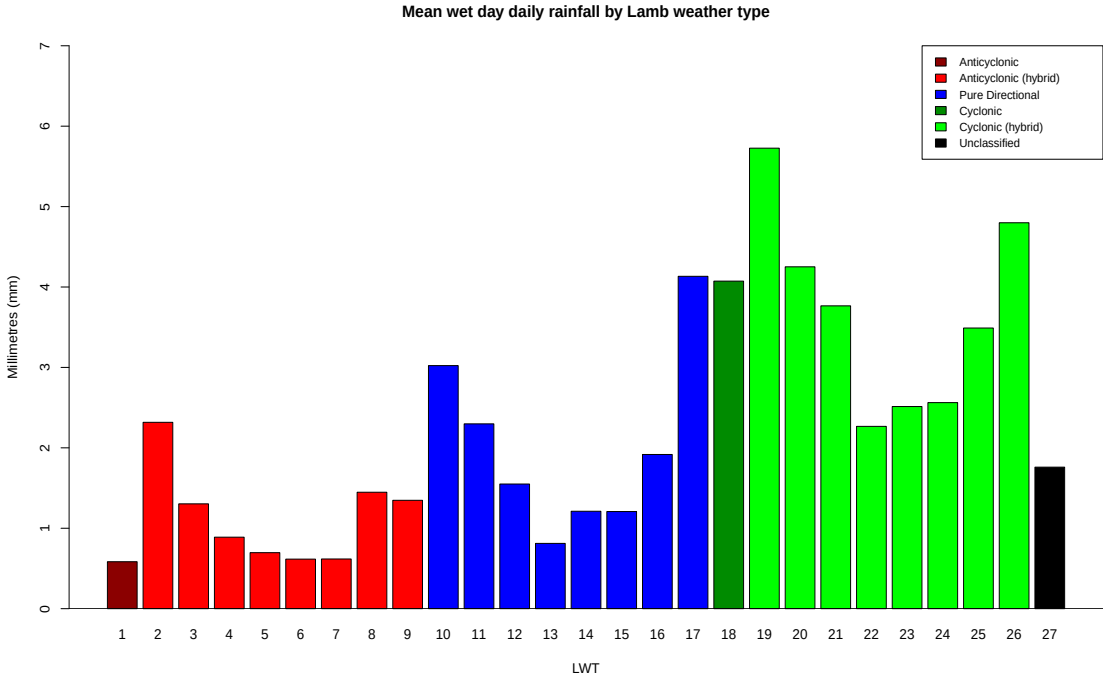
4.4.3 Modelling Lamb weather types

In this section, we propose to build a suitable stochastic model for LWTs. First-order homogeneous and nonhomogeneous Markov chain will be used to describe the pattern of LWTs.

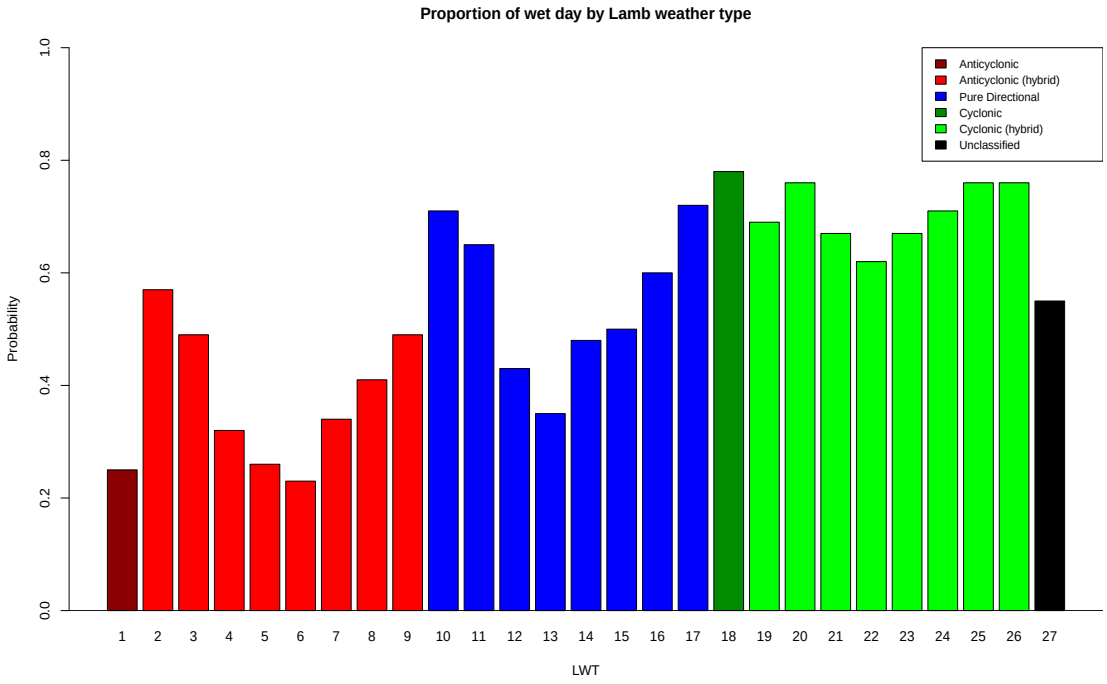
4.4.3.1 Homogeneous Markov chain for Lamb weather types

Let $\boldsymbol{X} = X_1, \cdots, X_T$ be the random sequence of LWTs at each time, t for all sites within the British Isles. By using a homogeneous Markov chain, the conditional probabilities for LWTs are given as

$$\Pr(X_t = l | X_{t-1} = k) = q_{kl}$$



(a)



(b)

Figure 4.20: (a) Mean wet day daily rainfall amounts and (b) proportion of wet days by the Lamb weather types for Darlington South Park Station

where state X_t is dependent on X_{t-1} and q_{kl} is the transition probability from state k to state l. Let Q represent the transition matrix for LWTs as:

$$Q = \begin{pmatrix} q_{1,1} & \cdots & q_{1,27} \\ \vdots & \ddots & \vdots \\ q_{27,1} & \cdots & q_{27,27} \end{pmatrix}$$

where $\sum_{l=1}^{27} q_{kl} = 1$ and $0 \leq q_{kl} \leq 1$ with the vector $\boldsymbol{q}_k$ representing row k of matrix Q. Then we can obtain a probability vector $\boldsymbol{\pi} = (\pi_1, \pi_2, \cdots, \pi_{27})$ satisfying:

$$\boldsymbol{\pi} Q = \boldsymbol{\pi}$$

which gives the marginal probabilities when the process is stationary. Subject to certain conditions (see Stewart (2009)), the state probabilities converge to this stationary distribution as time increases. We can rearrange this equation as:

$$\boldsymbol{\pi}(Q - I) = \mathbf{0} \tag{4.25}$$

$$\boldsymbol{\pi}\hat{Q} = \mathbf{0} \tag{4.26}$$

where I is the identity matrix, $\mathbf{0}$ is a zero vector and $\hat{Q} = Q - I$. By using the information given by

$$\pi_1 + \pi_2 + \cdots + \pi_{27} = 1,$$

we can replace the last column of $\hat{Q}$ and $\mathbf{0}$ by 1, so that

$$\tilde{Q} = \begin{pmatrix} \tilde{q}_{1,1} & \cdots & 1 \\ \tilde{q}_{2,1} & \ddots & 1 \\ \vdots & \vdots & \vdots \\ \tilde{q}_{27,1} & \cdots & 1 \end{pmatrix}$$

and $\tilde{\mathbf{0}} = (0, 0, 0, \cdots, 1)$. Hence,

$$\boldsymbol{\pi} = \tilde{\mathbf{0}}\tilde{Q}^{-1}. \tag{4.27}$$

Now, let N_{kl} represent the observed number of transitions from state k to state l:

$$\boldsymbol{N} = \begin{pmatrix} N_{11} & \cdots & N_{1,27} \\ \vdots & \ddots & \vdots \\ N_{27,1} & \cdots & N_{27,27} \end{pmatrix}$$

with $\boldsymbol{N}_k = (N_{k,1}, \cdots, N_{k,27})$ with $\hat{n}_k = \sum_{l=1}^{27} N_{kl}$. Then, $\boldsymbol{N}_k$ follows a multinomial distribution:

$$\boldsymbol{N}_k \sim M_{27}(\hat{n}_k, \boldsymbol{q}_k). \tag{4.28}$$

Let the prior specification for the transition probabilities, q_{kl}, be defined as follows. Let

$$q_{kl} = \frac{e^{\Upsilon_{kl}}}{\sum_{l=1}^{27} e^{\Upsilon_{kl}}}.$$

To avoid a non-identiafiability problem, we set $\Upsilon_{kk} = 0$ for all $k = 1, \cdots, 27$. For Υ_{kl} where $k \neq l$, we define it as:

$$\Upsilon_{kl} = \Upsilon_{0l} + \delta_{kl}$$

where $\delta_{kl} \sim N(0, \upsilon_\epsilon)$. To induce general positive correlations between the non-diagonal transition probabilities and allow borrowing of strength for less-frequent LWTs, we will use a hierarchical prior as follows:

$$\begin{aligned}
\Upsilon_{0l} &= \hat{\Upsilon}_0 + \gamma_l \\
\hat{\Upsilon}_0 &\sim N(m_{\hat{\Upsilon}_0}, \upsilon_{\hat{\Upsilon}_0}) \\
\gamma_l &\sim N(0, \upsilon_\gamma).
\end{aligned}$$

Table 4.4 shows an example of posterior means for LWT marginal probabilities using 7305 observations of LWTs from 1966 until 1985. The posterior samples were obtained using an MCMC algorithm with 2000 burn-in iterations and subsequent 5000 iterations to obtain the posterior samples. The computing time that was required to obtain 5000 posterior samples was around 2 minutes by using R software on a 2.00GHz Samsung laptop 300V3A model with Intel Core i7-2630QM processor and 12 Gbytes of random-access memory. From the table, we can see clearly that type 1 LWT has the highest probability with 0.211, where type 2 and 3 LWTs have the lowest probabilities which are 0.08. This result is consistent with the earlier inference in Figure 4.19.

4.4.3.2 Nonhomogeneous Markov chain of Lamb weather types

Initially, the LWTs are assumed to follow a homogeneous Markov chain. However, this may not be the case since the frequencies of the different LWTs may depend on the time of the year resulting in varied LWTs probabilities for every time t. Let the transition probabilities for LWTs be given as follows:

$$\Pr(X_t = l | X_{t-1} = k) = q_{kl}(t)$$

Parameter	Mean	Parameter	Mean
π_1	0.211	π_{15}	0.095
π_2	0.008	π_{16}	0.060
π_3	0.008	π_{17}	0.033
π_4	0.012	π_{18}	0.135
π_5	0.016	π_{19}	0.007
π_6	0.027	π_{20}	0.007
π_7	0.032	π_{21}	0.010
π_8	0.022	π_{22}	0.019
π_9	0.015	π_{23}	0.025
π_{10}	0.015	π_{24}	0.022
π_{11}	0.014	π_{25}	0.018
π_{12}	0.026	π_{26}	0.015
π_{13}	0.053	π_{27}	0.012
π_{14}	0.079		

Table 4.4: Posterior means of the LWTs probabilties

with a transition matrix of:

$$Q(t) = \begin{pmatrix} q_{1,1}(t) & \cdots & q_{1,27}(t) \\ \vdots & \ddots & \vdots \\ q_{27,1}(t) & \cdots & q_{27,27}(t) \end{pmatrix}$$

where $\sum_{l=1}^{27} q_{kl}(t) = 1$ and $0 \leq q_{kl}(t) \leq 1$. Since the transition probabilities depend on time, the process is called a non-homogeneous Markov chain. Given that $X_{t-1} = k$, the state X_t follows a categorical distribution:

$$X_t \sim Cat(q_{kl}(t))$$

where the transition probabilities $q_{kl}(t)$ are defined to be:

$$q_{kl}(t) = \frac{e^{\Upsilon_{kl}(t)}}{\sum_{l=1}^{27} e^{\Upsilon_{kl}(t)}}.$$

Similar to the previous homogeneous process, zero values are assigned to $\Upsilon_{kk}(t)$ for all $k = 1, \cdots, 27$ so that the non-identifiability problem can be avoided. In general, the $\Upsilon_{kl}(t)$ for $k \neq l$ can be written as

$$\Upsilon_{kl}(t) = \Upsilon_{0l} + \sum_{f=1}^{F} \left[\ddot{a}_f^l \cos\left(\frac{2\pi f t}{365.25}\right) + \ddot{b}_f^l \sin\left(\frac{2\pi f t}{365.25}\right) \right] + \delta_{kl} \tag{4.29}$$

where

$$\Upsilon_{kk}(t) = 0$$

and $\delta_{kl} \sim N(0, \upsilon_\epsilon)$. The truncated Fourier series is included in this model to allow for seasonal variability in the transition probabilities. For the intercept term Υ_{0l}, we will use a hierarchical prior to allow borrowing of strength for the less frequent LWTs:

$$\begin{aligned}
\Upsilon_{0l} &= \hat{\Upsilon}_0 + \gamma_l \\
\hat{\Upsilon}_0 &\sim N(m_{\hat{\Upsilon}_0}, \upsilon_{\hat{\Upsilon}_0}) \\
\gamma_l &\sim N(0, \upsilon_\gamma).
\end{aligned}$$

Figures 4.21, 4.22, 4.23 and 4.24 show an example of posterior means for transition probabilities $q_{1l}(t)$ over the year when $X_t = l$ is conditional on $X_{t-1} = k = 1$ (type 1 LWT) for $l = 1, 2, \cdots, 27$. The posterior samples were obtained by an MCMC algorithm with 1000 burn-in iterations and subsequent 5000 iterations to acquire the posterior samples. The computing time that was required to generate 5000 posterior draws was around 38 hours by using R software on a 2.00GHz Samsung laptop 300V3A model with Intel Core i7-2630QM processor and 12 Gbytes of random-access memory. From our analysis, 5000 iterations are considered sufficient to draw samples from the posterior distribution since the mixing is good for all parameters. As we can observe from these figures, the transition probability $q_{11}(t)$ (anticyclonic) is always above 0.45 over the year with the maximum and minimum probabilities in May and December, respectively . On the other hand, the transition probabilities $q_{1,l}(t)$ for $X_t = l = 2, \cdots, 27$, are always below 0.1. We can also observe that there is a different seasonal pattern for every LWT.

4.4.4 The rainfall model

In this section, we will focus on constructing a British daily rainfall model which is different from the Italian rainfall model in Section 4.3. The main objective here is to provide an alternative model to that of Heaps *et al.* (2015) and Germain (2010) by incorporating the Lamb weather types (LWTs) into the amount and occurrence processes to capture and explore the essential relationship that exists between the atmospheric circulation pattern and local daily rainfall. For preliminary investigation, we will focus only on a single site before we extend it to multiple sites in Chapter 5.

Let $f(y_t|\mu_t)$ represent the probability density function for the amount process with a range of $(0, \infty)$ where y_t is the potential rainfall amount. This approach was applied to the Italian rainfall data, and the results indicated that the gamma distribution was slightly superior to the lognormal distribution with respect to model fit and posterior

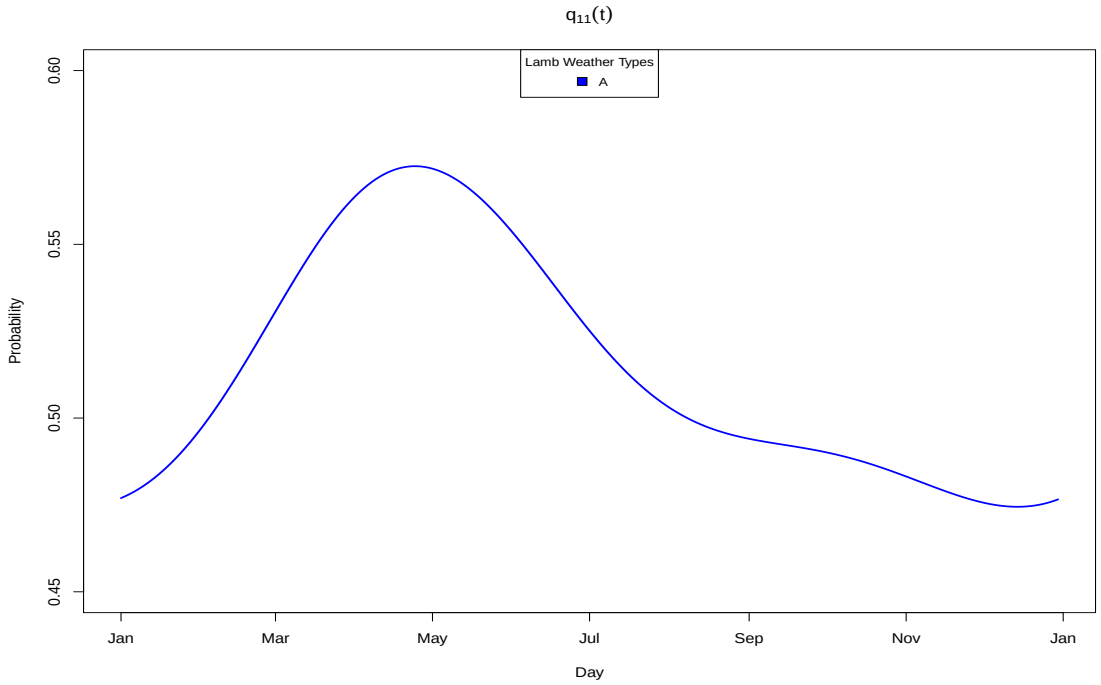


Figure 4.21: The posterior mean of transition probabilities, q_{11}. For details, see Table 4.3.

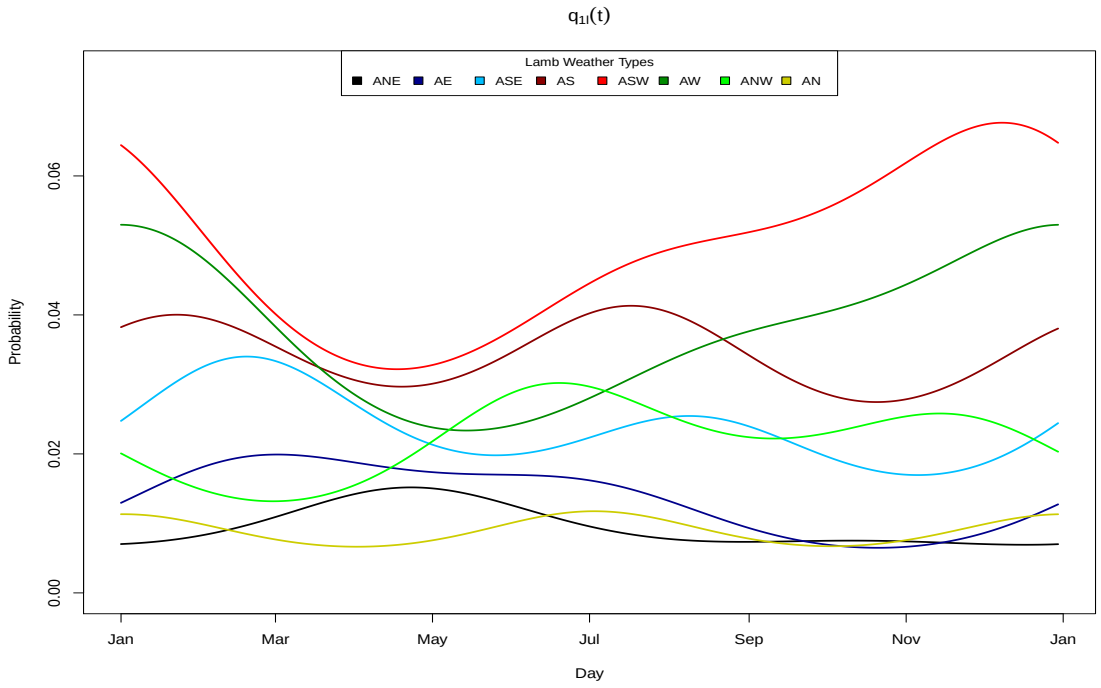


Figure 4.22: The posterior mean of transition probabilities, q_{1l} where $l = 2, \cdots, 9$. For details, see Table 4.3.

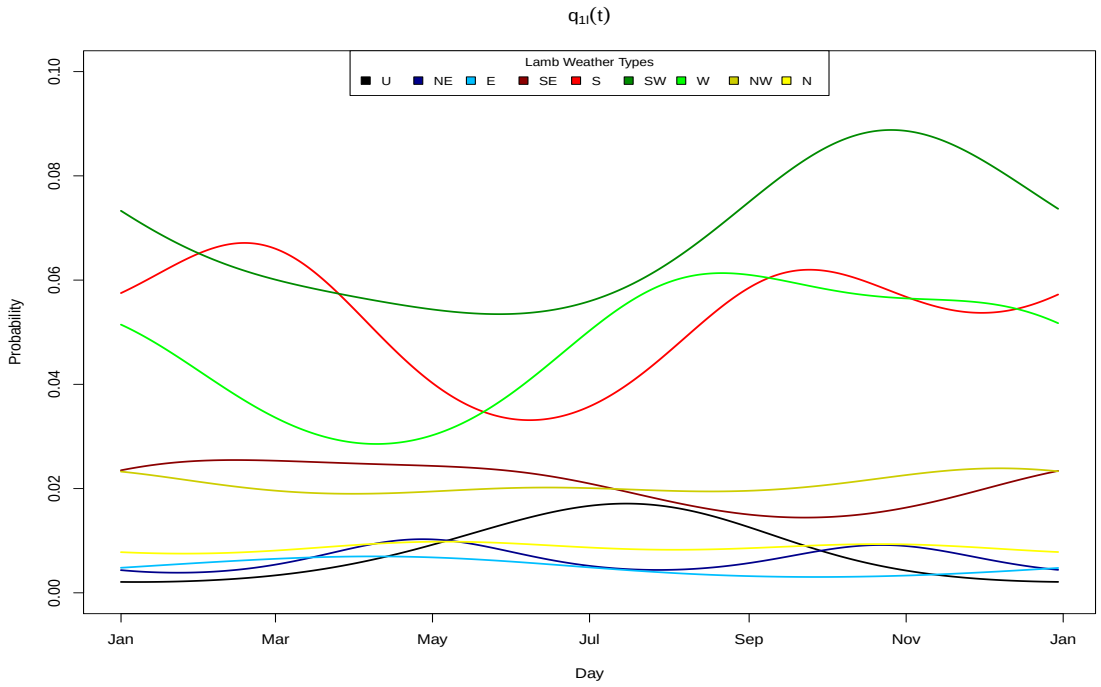


Figure 4.23: The posterior mean of transition probabilities, q_{1l} where $l = 10, \cdots, 17, 27$. For details, see Table 4.3.

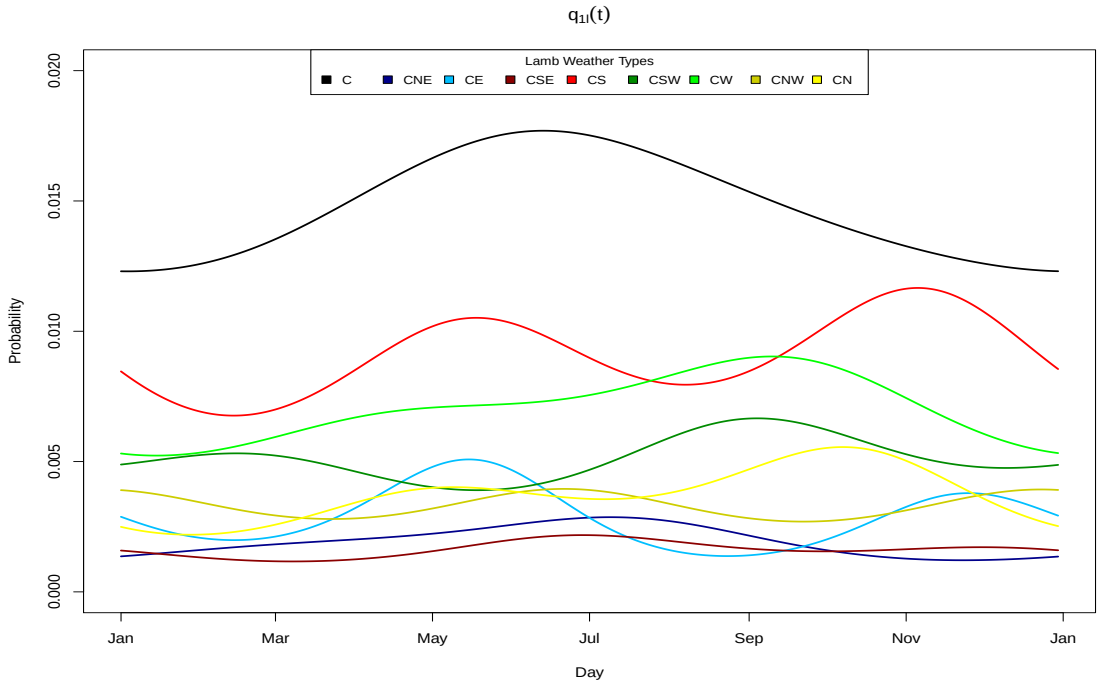


Figure 4.24: The posterior mean of transition probabilities, q_{1l} where $l = 18, \cdots, 26$. For details, see Table 4.3.

predictive performance. From the previous investigation in Section 4.3.3.6, we found that the shape parameter did not show significant changes from month to month, while the scale parameter did show some changes over time. Therefore, we propose to use the G1 distribution to describe the potential rainfall amount for the British daily rainfall as follows:

$$Y_t|\alpha, \mu_t \sim Ga\left(\alpha, \frac{\alpha}{\mu_t}\right)$$

where α is the shape parameter, and μ_t is the mean of the gamma distribution. For simplification, we will make the parameter α constant over time. In the case of the amount process, the LWT will be incorporated through the parameter μ_t. Specifically, let the parameter μ_t be modelled using the log link function:

$$\log(\mu_t|x_t = k) = \eta_0^k + \sum_{f=1}^{F}\left[a_f^k \cos\left(\frac{2\pi f t}{365.25}\right) + b_f^k \sin\left(\frac{2\pi f t}{365.25}\right)\right] + \epsilon_t \tag{4.30}$$

where ϵ_t is a random effect at time t. The Fourier series is added to the amount process to capture seasonal variability over the year using three harmonics as indicated in Section 4.2.4. The parameter μ_t is said to be conditional on the LWT by making the intercept, η_0^k, and Fourier coefficients dependent on the LWT at time t. By formulating the η_0^k and Fourier coefficients to vary between the LWT, we gain the facility to model the behaviour of each LWT. As indicated in Figure 4.20a, we can observe that certain LWTs are associated with greater amounts of rainfall than others. Moreover, some of the LWTs might be associated with greater rainfall at some times of year and less at others. Then, the intercept parameter, η_0^k plays a prominent role in controlling the level of potential rainfall amounts while Fourier coefficients play a role in accommodating changes in the rainfall amounts associated with a LWT over the year. In this model, we introduce a random effect in the linear regression of $\log(\mu_t)$ which is assumed to follow a first-order autoregressive model given by

$$\epsilon_t = \phi_\epsilon \epsilon_{t-1} + \omega_\epsilon$$

where $\omega_\epsilon \sim N(0, 1/\tau)$. Specifically, this random effect is assumed to follow a normal distribution:

$$\epsilon_t|\epsilon_{t-1} \sim N\left(\phi_\epsilon \epsilon_{t-1}, \frac{1}{\tau}\right)$$

and the distribution of the random effect at time $t = 1$ is given by

$$\epsilon_1 \sim N\left(0, \frac{1}{\tau\left(1 - \phi_\epsilon^2\right)}\right)$$

with $|\phi_\epsilon| < 1$ to ensure stationarity. Then, the mean of the distribution is said to be conditionally dependent on the previous random effect, ϵ_{t-1}. Therefore, the distribution

of μ_t is a lognormal distribution. Note that the marginal distribution of Y_t given other parameters obtained by integrating out μ_t from the joint density of Y_t and μ_t is now not a gamma distribution.

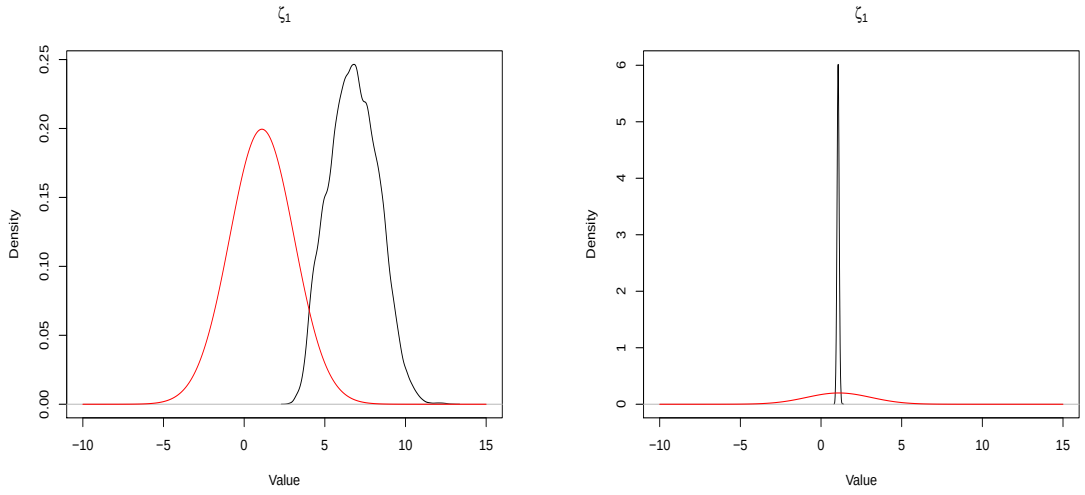
The rainfall occurrence can be modelled using a logit link function and a linear regression to relate the rainfall probabilities to the potential rainfall amount:

$$\text{logit}\left(p_1(t)|x_t = k\right) = \zeta_0^k + \sum_{f=1}^{F}\left[c_f^k \cos\left(\frac{2\pi f}{365.25}\right) + d_f^k \sin\left(\frac{2\pi f}{365.25}\right)\right] + \zeta_1 \log(y_t) \quad (4.31)$$

where $p_0(t) = 1 - p_1(t)$ represents the probability of zero rainfall. This parameterisation is in contrast to the Italian daily rainfall application where we used a nonhomogeneous first-order Markov chain to generate the probability of rainfall. As in the amount process, ζ_0^k is regarded as an intercept parameter for the probability of rainfall with a Fourier series to allow for seasonal variability where the information of the LWT is incorporated through these parameters. Three harmonics of the Fourier series will be used to represent the seasonal effect of the rainfall amount over the year as suggested in Section 4.2.4.

There are various ways we can use to build the relationship between the amount and occurrence processes. For example, we can use the mean of the amount distribution, μ_t, as in Sofia (2007) and the previous Italian model to link between the amount and occurrence processes. In this model, we use a different approach by incorporating the potential rainfall amount, y_t, into Equation (4.31) to build the relationship between the amount and occurrence processes. Based on the preliminary analysis, we found that the parameters in the occurrence process are well identified when we use y_t rather than μ_t. For instance, Figure 4.25 shows that the posterior density of ζ_1 is well identified when we used y_t instead of μ_t. Hence, we decided to use the potential rainfall amount, y_t as a bridge to link between the amount and occurrence processes. The resulting structure is shown in the DAG in Figure 4.26. This figure illustrates the detail of temporal dependence structure for the univariate model in the context of the British data which is different from the previous univariate model.

The parameterisations used for both the amount and occurrence processes are reasonable to model British daily rainfall. However, the model contains a very large number of parameters since we have many parameters for every single LWT. Therefore, we decided to modify our model to avoid any difficulties caused by this. The model for both the amount and occurrence process can be simplified by allowing only the intercept parameters η_0^k and ζ_0^k to depend on the LWT while the Fourier coefficients are constant over LWT. Thus, the new models for rainfall amounts and occurrences are given by



(a) (b)

Figure 4.25: The posterior (black) and prior (red) densities plots for ζ_1 when linking the occurrence process through (a) μ_t and (b) y_t.

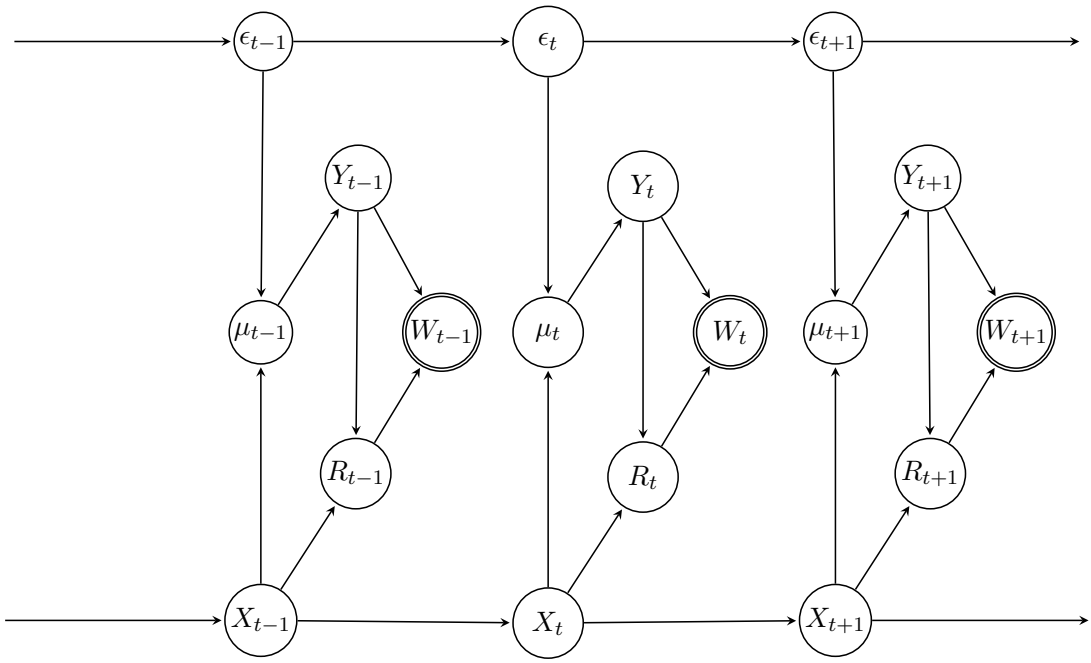


Figure 4.26: A DAG showing the temporal dependence structure for a single site in the British daily rainfall model

- Amount process

$$\log(\mu_t|x_t = k) = \eta_0^k + \sum_{f=1}^{F}\left[a_f \cos\left(\frac{2\pi f t}{365.25}\right) + b_f \sin\left(\frac{2\pi f t}{365.25}\right)\right] + \epsilon_t \qquad (4.32)$$

- Occurrence process

$$\text{logit}\left(p_1(t)|x_t = k\right) = \zeta_0^k + \sum_{f=1}^{F}\left[c_f \cos\left(\frac{2\pi f t}{365.25}\right) + d_f \sin\left(\frac{2\pi f t}{365.25}\right)\right] + \zeta_1 \log(y_t). \tag{4.33}$$

4.4.4.1 Prior Specifications

This section provides a suggestion for the structure of the prior specifications for the unknown parameters in our model. Normally, the prior distribution is chosen based on the information acquired from previous studies or personal beliefs. Illustrative values for the hyperparameters in the prior specifications are given in Section 4.4.5.1.

The shape parameter, α is given a gamma prior distribution:

$$\alpha \sim Ga(g_\alpha, h_\alpha).$$

A gamma distribution is a sensible prior choice for α since it has only a positive support. The prior distribution for the linear parameters is a normal distribution to allow the parameters to take the values from $-\infty$ to ∞. Then, the prior distribution of the linear parameters for the amount and occurrence processes can be written as follows:

- Amount process

$$\begin{aligned}
\eta_0^k &\sim N\left(m_{\eta_0}, v_{\eta_0}\right) \\
a_f &\sim N\left(m_{a,f}, v_{a,f}\right) \\
b_f &\sim N\left(m_{b,f}, v_{b,f}\right)
\end{aligned}$$

- Occurrence process

$$\begin{aligned}
\zeta_0^k &\sim N\left(m_{\zeta_0}, v_{\zeta_0}\right) \\
\zeta_1 &\sim N\left(m_{\zeta_1}, v_{\zeta_1}\right) \\
c_f &\sim N\left(m_{c,f}, v_{c,f}\right) \\
d_f &\sim N\left(m_{d,f}, v_{d,f}\right)
\end{aligned}$$

where $k \in \{1, 2, \cdots, 27\}$ and $f \in \{1, 2, 3\}$. The Fourier coefficients can be combined as a vector for each process as $\ddot{\boldsymbol{\eta}} = (a_1, a_2, a_3, b_1, b_2, b_3)'$ and $\ddot{\boldsymbol{\zeta}} = (c_1, c_2, c_3, d_1, d_2, d_3)'$, respec-

tively. Then, the prior distributions for $\ddot{\boldsymbol{\eta}}$ and $\ddot{\boldsymbol{\zeta}}$ can be multivariate normal distributions:

$$\ddot{\boldsymbol{\eta}} \sim N_F\left(\underline{\ddot{\eta}}_0, P_{\ddot{\boldsymbol{\eta}}}^{-1}\right)$$
$$\ddot{\boldsymbol{\zeta}} \sim N_F\left(\underline{\ddot{\zeta}}_0, P_{\ddot{\boldsymbol{\zeta}}}^{-1}\right)$$

where $\underline{\ddot{\eta}}_0$ and $\underline{\ddot{\zeta}}_0$ are the mean vectors with precision matrices $P_{\ddot{\boldsymbol{\eta}}}$ and $P_{\ddot{\boldsymbol{\zeta}}}$.

We also need to specify the prior distributions for the random effect parameters ϕ_ϵ and τ. The prior distribution for ϕ_ϵ must respect the stationarity condition $|\phi_\epsilon| < 1$. Any distribution for ϕ_ϵ including the prior must have zero probability outside this range. For example, we can use a uniform prior as follows

$$\phi_\epsilon \sim U(-1, 1).$$

Alternatively, we can introduce a new variable, v which follows a beta distribution:

$$v \sim Beta(a_v, b_v).$$

The beta distribution is defined on the support $(0, 1)$, so we can obtain the prior value for ϕ_ϵ by making:

$$\phi_\epsilon = 2v - 1. \tag{4.34}$$

Therefore, the value for ϕ_ϵ can always be in the range $(-1, 1)$. Similar to parameter α, a suitable prior distribution for τ is a gamma distribution:

$$\tau \sim Ga(g_\tau, h_\tau),$$

because a variance or precision can never be negative.

4.4.4.2 Posterior distributions

Let $\boldsymbol{\theta}$ represent the collection of unknown parameters. Then, the joint density of $(\boldsymbol{\theta}, \boldsymbol{y}, \boldsymbol{r})$ can be given as follows:

$$\pi(\boldsymbol{\theta}, \boldsymbol{y}, \boldsymbol{r}) = \pi(\boldsymbol{\theta}_{amt}) \times \pi(\boldsymbol{\theta}_{occ}) \times f(\boldsymbol{y}|\boldsymbol{\theta}_{amt}) \times f(\boldsymbol{r}|\boldsymbol{y}, \boldsymbol{\theta}_{occ}) \tag{4.35}$$

where $\boldsymbol{\theta}_{amt}$ and $\boldsymbol{\theta}_{occ}$ are the collections of parameters for the amount and occurrence processes, repectively. From this joint density, we can work out the FCD for each parameter. The FCDs in this section are provided as general information and for future work. The comprehensive description of the FCD for each parameter is given as follows:

- Parameter α

  The FCD for the shape parameter, α is given by:

$$\begin{aligned}\pi(\alpha|\eta_0^k, \ddot{\boldsymbol{\eta}}, v, \tau, \boldsymbol{y}) &\propto \prod_{t=1}^{T} f(y_t|\boldsymbol{\theta}_{amt}, \epsilon_t) \times \pi(\alpha) \\ &\propto \prod_{t=1}^{T} \frac{(\alpha/\mu_t)^\alpha y_t^{\alpha-1} e^{-(\alpha/\mu_t)y_t}}{\Gamma(\alpha)} \times \frac{h_\alpha^{g_\alpha} \alpha^{g_\alpha - 1} e^{-h_\alpha \alpha}}{\Gamma(g_\alpha)} \\ &\propto \frac{(\alpha)^{\alpha T + g_\alpha - 1}}{[\Gamma(\alpha)]^T} \prod_{t=1}^{T} E_t^\alpha \exp\left\{-\alpha\left(\sum_{t=1}^{T} E_t + h_\alpha\right)\right\}\end{aligned}$$

  where $E_t = y_t/\mu_t$. This form does not correspond to any standard distribution.

- Parameter η_0^k

  Let $\eta_0^{k*} = \left(\eta_0^1, \cdots, \eta_0^{k-1}, \eta_0^{k+1}, \cdots, \eta_0^K\right)$. With this prior, the FCD for η_0^k is

$$\begin{aligned}\pi(\eta_0^k|\eta_0^{k*}, \alpha, \ddot{\boldsymbol{\eta}}, v, \tau, \boldsymbol{y}) &\propto \prod_{t\in T_k} f(y_t|\boldsymbol{\theta}_{amt}, \epsilon_t) \times \pi(\eta_0^k) \\ &\propto \prod_{t\in T_k} \frac{(\alpha/\mu_t)^\alpha y_t^{\alpha-1} e^{-(\alpha/\mu_t)y_t}}{\Gamma(\alpha)} \times \sqrt{\frac{1}{2\pi v_{\eta_0}}} \exp\left\{-\frac{1}{2v_{\eta_0}}(\eta_0^k - m_{\eta_0})^2\right\} \\ &\propto \prod_{t\in T_k} E_t^\alpha \exp\left\{-\left[\alpha \sum_{t\in T_k} E_t + \frac{1}{2v_{\eta_0}}\left((\eta_0^k)^2 - 2\eta_0^k m_{\eta_0}\right)\right]\right\}\end{aligned}$$

  where $T_k = \{t : x_t = k\}$ is the set of observations for days where the LWT is k. Notice that this FCD does not depend on η_0^l, $l \neq k$ and so $\pi(\eta_0^k|\eta_0^{k*}, \alpha, \ddot{\boldsymbol{\eta}}, \phi_\epsilon, \tau, \boldsymbol{y}) = \pi(\eta_0^k|\alpha, \ddot{\boldsymbol{\eta}}, \phi_\epsilon, \tau, \boldsymbol{y})$. The posterior samples can be obtained by taking an independent realisation from each of $\pi(\eta_0^k|\alpha, \ddot{\boldsymbol{\eta}}, \phi_\epsilon, \tau, \boldsymbol{y})$, where $k = 1, \cdots, K$. However, this FCD does not belong to any standard form.

- Parameter $\ddot{\boldsymbol{\eta}}$

  The FCD for $\ddot{\boldsymbol{\eta}}$ is

$$\begin{aligned}\pi(\ddot{\boldsymbol{\eta}}|\alpha,\eta_0^k,v,\tau,\boldsymbol{y}) &\propto \prod_{t=1}^{T} f(y_t|\boldsymbol{\theta}_{amt},\epsilon_t)\times\pi(\ddot{\boldsymbol{\eta}})\\ &\propto \prod_{t=1}^{T}\frac{(\alpha/\mu_t)^{\alpha}y_t^{\alpha-1}e^{-(\alpha/\mu_t)y_t}}{\Gamma(\alpha)}\times\frac{|P_{\ddot{\boldsymbol{\eta}}}|^{1/2}}{(2\pi)^{2F}}\exp\left\{-\frac{1}{2}(\ddot{\boldsymbol{\eta}}-\underline{\ddot{\eta}}_0)'P_{\ddot{\boldsymbol{\eta}}}(\ddot{\boldsymbol{\eta}}-\underline{\ddot{\eta}}_0)\right\}\\ &\propto \prod_{t=1}^{T}E_t^{\alpha}\exp\left\{-\left[\alpha\sum_{t=1}^{T}E_t+\frac{1}{2}(\ddot{\boldsymbol{\eta}}-\underline{\ddot{\eta}}_0)'P_{\ddot{\boldsymbol{\eta}}}(\ddot{\boldsymbol{\eta}}-\underline{\ddot{\eta}}_0)\right]\right\}.\end{aligned}$$

  This is again not in a standard form of any distribution.

- Parameter v

  Now, we need to define the FCD for the parameter v as follows:

$$\begin{aligned}\pi(v|\boldsymbol{\epsilon},\tau) \propto& f(\epsilon_1|\tau,\phi_\epsilon)\prod_{t=2}^{T}f(\epsilon_t|\epsilon_{t-1},\tau,\phi_\epsilon)\times\pi(v)\\ \propto&\sqrt{\frac{\tau\left(1-\phi_\epsilon^2\right)}{2\pi}}\exp\left\{-\frac{\tau\left(1-\phi_\epsilon^2\right)}{2}\epsilon_1^2\right\}\times\prod_{t=2}^{T}\sqrt{\frac{\tau}{2\pi}}\exp\left\{-\frac{\tau}{2}\left(\epsilon_t-\phi_\epsilon\epsilon_{t-1}\right)^2\right\}\\ &\times\frac{\Gamma(a_v+b_v)}{\Gamma(a_v)\Gamma(b_v)}v^{a_v-1}(1-v)^{b_v-1}\\ \propto& v^{a_v-1}(1-v)^{b_v-1}\sqrt{(1-\phi_\epsilon^2)}\exp\left\{-\frac{\tau}{2}\left[\left(1-\phi_\epsilon^2\right)\epsilon_1^2+\sum_{t=2}^{T}\left(\epsilon_t-\phi_\epsilon\epsilon_{t-1}\right)^2\right]\right\}\end{aligned}$$

  which also does not correspond to any standard distribution.

- Parameter τ

  The FCD for the parameter τ is given by

$$\begin{aligned}\pi(\tau|\boldsymbol{\epsilon},v) \propto& f(\epsilon_1|\tau,\phi_\epsilon)\prod_{t=2}^{T}f(\epsilon_t|\epsilon_{t-1},\tau,\phi_\epsilon)\times\pi(\tau)\\ \propto&\sqrt{\frac{\tau\left(1-\phi_\epsilon^2\right)}{2\pi}}\exp\left\{-\frac{\tau\left(1-\phi_\epsilon^2\right)}{2}\epsilon_1^2\right\}\times\prod_{t=2}^{T}\sqrt{\frac{\tau}{2\pi}}\exp\left\{-\frac{\tau}{2}\left(\epsilon_t-\phi_\epsilon\epsilon_{t-1}\right)^2\right\}\\ &\times\frac{h_\tau^{g_\tau}\tau^{g_\tau-1}e^{-h_\tau\tau}}{\Gamma(g_\tau)}\\ \propto&\tau^{g_\tau+\frac{T}{2}-1}\exp\left\{-\frac{\tau}{2}\left[\left(1-\phi_\epsilon^2\right)\epsilon_1^2+\sum_{t=2}^{T}\left(\epsilon_t-\phi_\epsilon\epsilon_{t-1}\right)^2+2h_\tau\right]\right\}.\end{aligned}$$

This is a gamma distribution $Ga(G_\tau, H_\tau)$ with

$$G_\tau = g_\tau + \frac{T}{2}$$

and

$$H_\tau = h_\tau + \frac{1}{2}\left[\left(1-\phi_\epsilon^2\right)\epsilon_1^2 + \sum_{t=2}^{T}\left(\epsilon_t - \phi_\epsilon\epsilon_{t-1}\right)^2\right].$$

- Random effect ϵ_t

To obtain the sampled values for the random effect ϵ_t, the FCD for ϵ_t is given by:

$$\begin{aligned}
\pi(\epsilon_t|\tau,\phi_\epsilon,y_t) \propto & f(y_t|\boldsymbol{\theta}_{amt},\epsilon_t)\times\pi(\epsilon_t|\epsilon_{t-1})\times\pi(\epsilon_{t+1}|\epsilon_t)\\
\propto & \frac{(\alpha/\mu_t)^\alpha y_t^{\alpha-1}e^{-(\alpha/\mu_t)y_t}}{\Gamma(\alpha)}\times\sqrt{\frac{\tau}{2\pi}}\exp\left\{-\frac{\tau}{2}\left(\epsilon_t-\phi_\epsilon\epsilon_{t-1}\right)^2\right\}\\
& \times\sqrt{\frac{\tau}{2\pi}}\exp\left\{-\frac{\tau}{2}\left(\epsilon_{t+1}-\phi_\epsilon\epsilon_t\right)^2\right\}\\
\propto & E_t^\alpha\exp\left\{-\left[\alpha E_t+\frac{\tau}{2}\left[\left(\epsilon_t-\phi_\epsilon\epsilon_{t-1}\right)^2+\left(\epsilon_{t+1}-\phi_\epsilon\epsilon_t\right)^2\right]\right]\right\}
\end{aligned}$$

for $1 < t < T-1$. For $t = 1$, the FCD is given by:

$$\begin{aligned}
\pi(\epsilon_1|\tau,\phi_\epsilon,y_1) \propto & f(y_1|\boldsymbol{\theta}_{amt},\epsilon_1)\times\pi(\epsilon_1)\times\pi(\epsilon_2|\epsilon_1)\\
\propto & \frac{(\alpha/\mu_1)^\alpha y_1^{\alpha-1}e^{-(\alpha/\mu_1)y_1}}{\Gamma(\alpha)}\times\sqrt{\frac{\tau\left(1-\phi_\epsilon^2\right)}{2\pi}}\exp\left\{-\frac{\tau\left(1-\phi_\epsilon^2\right)}{2}\epsilon_1^2\right\}\\
& \times\sqrt{\frac{\tau}{2\pi}}\exp\left\{-\frac{\tau}{2}\left(\epsilon_2-\phi_\epsilon\epsilon_1\right)^2\right\}\\
\propto & E_1^\alpha\exp\left\{-\left[\alpha E_1+\frac{\tau}{2}\left[\left(1-\phi_\epsilon^2\right)\epsilon_1^2+\left(\epsilon_2-\phi_\epsilon\epsilon_1\right)^2\right]\right]\right\}
\end{aligned}$$

and the FCD for ϵ_T is

$$\begin{aligned}
\pi(\epsilon_T|\tau,\phi_\epsilon,y_T) \propto & f(y_T|\boldsymbol{\theta}_{amt},\epsilon_T)\times\pi(\epsilon_T|\epsilon_{T-1})\\
& \frac{(\alpha/\mu_T)^\alpha y_T^{\alpha-1}e^{-(\alpha/\mu_T)y_T}}{\Gamma(\alpha)}\times\sqrt{\frac{\tau}{2\pi}}\exp\left\{-\frac{\tau}{2}\left(\epsilon_T-\phi_\epsilon\epsilon_{T-1}\right)^2\right\}\\
\propto & E_T^\alpha\exp\left\{-\left[\alpha E_T+\frac{\tau}{2}\left[\left(\epsilon_T-\phi_\epsilon\epsilon_{T-1}\right)^2\right]\right]\right\}.
\end{aligned}$$

- Parameter ζ_0

  Let $\zeta_0^{k*} = \left(\zeta_0^1, \cdots, \zeta_0^{k-1}, \zeta_0^{k+1}, \cdots, \zeta_0^K\right)$. Therefore, the FCD for ζ_0^k is given by

$$\begin{aligned}\pi(\zeta_0^k|\zeta_0^{k*}, \zeta_1, \ddot{\boldsymbol{\zeta}}, \boldsymbol{r}) &\propto \prod_{t\in T_k} f(r_t|y_t, \boldsymbol{\theta}_{occ}) \times \pi(\zeta_0^k) \\ &\propto \prod_{t\in T_k} [p_1(t)]^{r_t} [1-p_1(t)]^{(1-r_t)} \times \sqrt{\frac{1}{2\pi v_{\zeta_0}}} \exp\left\{-\frac{1}{2v_{\zeta_0}}(\zeta_0^k - m_{\zeta_0})^2\right\} \\ &\propto \prod_{t\in T_k} [p_1(t)]^{r_t} [1-p_1(t)]^{(1-r_t)} \exp\left\{-\frac{1}{2v_{\zeta_0}}\left(\zeta_0^k\right)^2 - 2\zeta_0^k m_{\zeta_0}\right\}.\end{aligned}$$

  This distribution is not in a standard form. Notice that this FCD does not depend on ζ_0^l, $l \neq k$ and so $\pi(\zeta_0^k|\zeta_0^{k*}, \zeta_1, \ddot{\boldsymbol{\zeta}}, \boldsymbol{r}) = \pi(\zeta_0^k|, \zeta_1, \ddot{\boldsymbol{\zeta}}, \boldsymbol{r})$. We can sample the posterior value by taking an independent realisation from each of $\pi(\zeta_0^k|, \zeta_1, \ddot{\boldsymbol{\zeta}}, \boldsymbol{r})$, where $k = 1, \cdots, 27$.

- Parameter ζ_1

  Next, the FCD for ζ_1 is

$$\begin{aligned}\pi(\zeta_1|\zeta_0^k\ddot{\boldsymbol{\zeta}}, \boldsymbol{r}) &\propto \prod_{t=1}^{T} f(r_t|y_t, \boldsymbol{\theta}_{occ}) \times \pi(\zeta_1) \\ &\propto \prod_{t=1}^{T} [p_1(t)]^{r_t} [1-p_1(t)]^{(1-r_t)} \times \sqrt{\frac{1}{2\pi v_{\zeta_1}}} \exp\left\{-\frac{1}{2v_{\zeta_1}}(\zeta_1 - m_{\zeta_1})^2\right\} \\ &\propto \prod_{t=1}^{T} [p_1(t)]^{r_t} [1-p_1(t)]^{(1-r_t)} \exp\left\{-\frac{1}{2v_{\zeta_1}}\left(\zeta_1^k\right)^2 - 2\zeta_1^k m_{\zeta_1}\right\}\end{aligned}$$

  where the form indicates that it does not belong to any standard distribution.

- Parameter $\ddot{\boldsymbol{\zeta}}$

  For parameter $\ddot{\boldsymbol{\zeta}}$, the FCD can be obtained as follows:

$$\begin{aligned}\pi(\ddot{\boldsymbol{\zeta}}|\zeta_0^k, \zeta_1, \boldsymbol{z}) &\propto \prod_{t=1}^{T} f(r_t|y_t, \boldsymbol{\theta}_{occ}) \times \pi(\ddot{\boldsymbol{\zeta}}) \\ &\propto \prod_{t=1}^{T} [p_1(t)]^{r_t} [1-p_1(t)]^{(1-r_t)} \times \frac{|P_{\ddot{\boldsymbol{\zeta}}}|^{1/2}}{(2\pi)^3} \exp\left\{-\frac{1}{2}(\ddot{\boldsymbol{\zeta}} - \ddot{\underline{\zeta}}_0)' P_{\ddot{\boldsymbol{\zeta}}} (\ddot{\boldsymbol{\zeta}} - \ddot{\underline{\zeta}}_0)\right\} \\ &\propto \prod_{t=1}^{T} [p_1(t)]^{r_t} [1-p_1(t)]^{(1-r_t)} \exp\left\{-\frac{1}{2}(\ddot{\boldsymbol{\zeta}} - \ddot{\underline{\zeta}}_0)' P_{\ddot{\boldsymbol{\zeta}}} (\ddot{\boldsymbol{\zeta}} - \ddot{\underline{\zeta}}_0)\right\}.\end{aligned}$$

Once again, this distribution does not belong to any standard form.

- Latent y-values: $\boldsymbol{y}_{dry}$

  The latent value, y_t is not observed when it is a dry day ($W_t = 0$) which can be defined as y_{dry}. The values of y_{dry} are treated as auxiliary data in a data augmentation process (Tanner & Wong, 1987). Then, it is necessary to sample the posterior value of y_{dry} when $r_t = 0$ by considering the FCD as follows:

$$\begin{aligned}\pi(y_{dry}|r_t = 0, \alpha, \mu_t) \propto & f(y_{dry}|\alpha, \mu_t) \times \Pr(r_t = 0|y_{dry}) \\ \propto & \frac{(\alpha/\mu_t)^{\alpha} y_{dry}^{\alpha-1} e^{-(\alpha/\mu_t)y_{dry}}}{\Gamma(\alpha)} \times p_0(t)\end{aligned}$$

  where $p_0(t) = 1 - p_1(t)$ represents the probability of a dry day ($r_t = 0$) which also depends on y_{dry}.

As we can see, the distributions of most parameters are not in a standard form. Hence, we need to use a Metropolis-within-Gibbs scheme to generate posterior samples for each parameter. The detailed steps of the MCMC scheme for drawing posterior samples are given as follow:

1. Initialise the iteration counter to j=1. Set the initial state of the chain to $\boldsymbol{\theta}^{(0)} = \left(\alpha^{(0)}, \eta_0^{1(0)}, \cdots, \eta_0^{27(0)}, \ddot{\boldsymbol{\eta}}^{(0)}, v^{(0)}, \tau^{(0)}, \epsilon_t^{(0)}, \zeta_0^{1(0)}, \cdots, \zeta_0^{27(0)}, \zeta_1^{(0)}, \ddot{\boldsymbol{\zeta}}^{(0)}, y_{dry}\right)'$.

2. Obtain a new value $\boldsymbol{\theta}^{(j)}$ from $\boldsymbol{\theta}^{(j-1)}$ by successive generation of values

   - $\alpha^{(j)} \sim \pi\Big(\alpha\Big|\eta_0^{1(j-1)}, \cdots, \eta_{j-1}^{27(j-1)}, \ddot{\boldsymbol{\eta}}^{(j-1)}, v^{(j-1)}, \tau^{(j-1)}, \epsilon_t^{(t-1)}, \zeta_0^{1(j-1)}, \cdots, \zeta_0^{27(j-1)}, \zeta_1^{(j-1)}, \ddot{\boldsymbol{\zeta}}^{(j-1)}, \boldsymbol{y}, \boldsymbol{x}\Big)$ using a Metropolis-Hastings step with proposal distribution:

$$\alpha^* \sim Ga\left(h_{\alpha^*}\alpha^{(j-1)}, h_{\alpha^*}\right).$$

   - $\eta_0^{1(j)} \sim \pi\Big(\eta_0^1\Big|\alpha^{(j)}, \eta_0^{2(j-1)}, \cdots, \eta_0^{27(j-1)}, \ddot{\boldsymbol{\eta}}^{(j-1)}, v^{(j-1)}, \tau^{(j-1)}, \epsilon_t^{(t-1)}, \zeta_0^{1(j-1)}, \cdots, \zeta_0^{27(j-1)}, \zeta_1^{(j-1)}, \ddot{\boldsymbol{\zeta}}^{(j-1)}, \boldsymbol{y}, \boldsymbol{x}\Big)$ using a Metropolis-Hastings step with proposal distribution:

$$\eta_0^{1^*} \sim N\left(\eta_0^{1(j-1)}, \Sigma_{\eta_0^*}\right)$$

     with a suitable variance, $\Sigma_{\eta_0^*}$.

     $\vdots$

- $\eta_0^{27(j)} \sim \pi\Big(\eta_0^{27}\Big|\alpha^{(j)}, \eta_0^{1(j)}, \cdots, \eta_0^{26(j)}, \ddot{\boldsymbol{\eta}}^{(j-1)}, v^{(j-1)}, \tau^{(j-1)}, \epsilon_t^{(j-1)}, \zeta_0^{1(j-1)}, \cdots, \zeta_0^{27(j-1)}, \zeta_1^{(j-1)}, \ddot{\boldsymbol{\zeta}}^{(j-1)}, \boldsymbol{y}, \boldsymbol{x}\Big)$ using a Metropolis-Hastings step with proposal distribution:

$$\eta_0^{27^*} \sim N\left(\eta_0^{27(j-1)}, \Sigma_{\eta_0^*}\right).$$

- $\ddot{\boldsymbol{\eta}}^{(j)} \sim \pi\Big(\ddot{\boldsymbol{\eta}}\Big|\alpha^{(j)}, \eta_0^{1(j)}, \cdots, \eta_0^{27(j)}, v^{(j-1)}, \tau^{(j-1)}, \epsilon_t^{(j-1)}, \zeta_0^{1(j-1)}, \cdots, \zeta_0^{27(j-1)}, \zeta_1^{(j-1)}, \ddot{\boldsymbol{\zeta}}^{(j-1)}, \boldsymbol{y}, \boldsymbol{x}\Big)$ using a Metropolis-Hastings step with proposal distribution:

$$\ddot{\boldsymbol{\eta}}^* \sim N_6\left(\ddot{\boldsymbol{\eta}}^{(j-1)}, \Sigma_{\ddot{\boldsymbol{\eta}}^*}\right).$$

- $v^{(j)} \sim \pi\Big(v\Big|\alpha^{(j)}, \eta_0^{1(j)}, \cdots, \eta_0^{27(j)}, \ddot{\boldsymbol{\eta}}^{(j)}, \tau^{(j-1)}, \epsilon_t^{(j-1)}, \zeta_0^{1(j-1)}, \cdots, \zeta_0^{27(j-1)}, \zeta_1^{(j-1)}, \ddot{\boldsymbol{\zeta}}^{(j-1)}, \boldsymbol{y}, \boldsymbol{x}\Big)$ using a Metropolis-Hastings step with proposal distribution:

$$v^* \sim Beta\left(b_{v^*} v^{(j-1)}, b_{v^*}\left(1 - v^{(j-1)}\right)\right).$$

- $\tau^{(j)} \sim \pi\Big(\tau\Big|\alpha^{(j)}, \eta_0^{1(j)}, \cdots, \eta_0^{27(j)}, \ddot{\boldsymbol{\eta}}^{(j)}, v^{(j)}, \epsilon_t^{(j-1)}, \zeta_0^{1(j-1)}, \cdots, \zeta_0^{27(j-1)}, \zeta_1^{(j-1)}, \ddot{\boldsymbol{\zeta}}^{(j-1)}, \boldsymbol{y}, \boldsymbol{x}\Big)$ using a Gibbs sampler step where

$$\tau^* \sim Ga\left(G_\tau, H_\tau\right).$$

- $\epsilon_t^{(j)} \sim \pi\left(\epsilon_t\Big|\alpha^{(j)}, \eta_0^{1(j)}, \cdots, \eta_0^{27(j)}, \ddot{\boldsymbol{\eta}}^{(j)}, v^{(j)}, \tau^j, \zeta_0^{1(j-1)}, \cdots, \zeta_0^{27(j-1)}, \zeta_1^{(j-1)}, \ddot{\boldsymbol{\zeta}}^{(j-1)}, \boldsymbol{y}\right)$ using a Metropolis-Hastings step with proposal distribution:

$$\epsilon_t^* \sim N\left(\epsilon_t^{(j-1)}, \sigma_{\epsilon^*}^2\right)$$

with a suitable variance, $\sigma_{\epsilon^*}^2$. For example $\sigma_{\epsilon^*}^2 = [\tau(1-\phi_\epsilon^2)]^{-1}$.

- $\zeta_0^{1(j)} \sim \pi\Big(\zeta_0^1\Big|\alpha^{(j)}, \eta_0^{1(j)}, \cdots, \eta_0^{27(j)}, \ddot{\boldsymbol{\eta}}^{(j)}, v^{(j)}, \tau^{(j)}, \epsilon_t^{(j)}, \zeta_0^{2(j-1)}, \cdots, \zeta_0^{27(j-1)}, \zeta_1^{(j-1)}, \ddot{\boldsymbol{\zeta}}^{(j-1)}, \boldsymbol{y}, \boldsymbol{x}\Big)$ using a Metropolis-Hastings step with proposal distribution:

$$\zeta_0^{1^*} \sim N\left(\zeta_0^{1(j-1)}, \Sigma_{\zeta_0^*}\right).$$

$\vdots$

- $\zeta_0^{27(j)} \sim \pi\Big(\zeta_0^{27}\Big|\alpha^{(j)}, \eta_0^{1(j)}, \cdots, \eta_0^{27(j)}, \ddot{\boldsymbol{\eta}}^{(j)}, v^{(j)}, \tau^{(j)}, \epsilon_t^{(j)}, \zeta_0^{1(j)}, \cdots, \zeta_0^{26(j)}, \zeta_1^{(j-1)}, \ddot{\boldsymbol{\zeta}}^{(j-1)}, \boldsymbol{y}, \boldsymbol{x}\Big)$ using a Metropolis-Hastings step with proposal distribution:

$$\zeta_0^{27^*} \sim N\left(\zeta_0^{27(j-1)}, \Sigma_{\zeta_0^*}\right).$$

- $\zeta_1^{(j)} \sim \pi\Big(\zeta_1 \Big| \alpha^{(j)}, \eta_0^{1(j)}, \cdots, \eta_0^{27(j)}, \ddot{\boldsymbol{\eta}}^{(j)}, v^{(j)}, \tau^{(j)}, \epsilon_t^{(j)}, \zeta_0^{1(j)}, \cdots, \zeta_0^{27(j)}, \ddot{\boldsymbol{\zeta}}^{(j-1)}, \boldsymbol{y}, \boldsymbol{x}\Big)$ using a Metropolis-Hastings step with proposal distribution:

$$\zeta_1^* \sim N\left(\zeta_1^{(j-1)}, \Sigma_{\zeta_1^*}\right).$$

- $\ddot{\boldsymbol{\zeta}}^{(j)} \sim \pi\Big(\ddot{\boldsymbol{\zeta}} \Big| \alpha^{(j)}, \zeta_0^{1(j)}, \cdots, \zeta_0^{27(j)}, \ddot{\boldsymbol{\eta}}^{(j)}, v^{(j)}, \tau^{(j)}, \epsilon_t^{(j)}, \zeta_0^{1(j)}, \cdots, \zeta_0^{27(j)}, \zeta_1^{(j)}, \boldsymbol{y}, \boldsymbol{x}\Big)$ using a Metropolis-Hastings step with proposal distribution:

$$\ddot{\boldsymbol{\zeta}}^* \sim N_6\left(\ddot{\boldsymbol{\zeta}}^{(j-1)}, \Sigma_{\ddot{\boldsymbol{\zeta}}^*}\right).$$

- $y_{dry}^{(j)} \sim \pi\Big(y_{dry} \Big| \alpha^{(j)}, \zeta_0^{1(j)}, \cdots, \zeta_0^{27(j)}, \ddot{\boldsymbol{\eta}}^{(j)}, v^{(j)}, \tau^{(j)}, \epsilon_t^{(j)}, \zeta_0^{1(j)}, \cdots, \zeta_0^{27(j)}, \zeta_1^{(j)}, \ddot{\boldsymbol{\zeta}}^{(j)}, \boldsymbol{y}, \boldsymbol{x}\Big)$

3. Change counter j to $j+1$, and return to step 2.

4.4.5 Application

4.4.5.1 Prior distribution

The prior specifications for the British daily rainfall model were constructed using suitable distributions based on the information gained from previous studies and personal beliefs. The chosen prior structure for this model and the methods of elicitation are intended as recommendations. The actual choice of hyperparameters here is for illustrative purposes only.

In the British daily rainfall model we use the G1 distribution for the rainfall amount. However the situation is more complicated for two reasons. The first is that we do not observe samples from the distribution of Y_t. We only observe samples from the distribution of W_t. The second is the presence of the random effect ϵ_t in (4.32). Pragmatically for elicitation purposes, we propose to ignore the first problem and elcit beliefs about observed rainfall amounts and to approximate the distribution of the rainfall amount Y_t with a lognormal distribution. So, approximately, $Z_t = \log Y_t$ has a normal distribution.

Suppose that we write

$$Z_t = \eta_t^\star + e_t + \epsilon_t$$

where $e_t \sim \mathrm{N}(0, v_e)$, $\epsilon_t \sim \mathrm{N}(0, v_\epsilon)$ and $v_\epsilon = [\tau(1-\phi_\epsilon^2)]^{-1}$. Then the expectation of Y_t,

given model parameters and the LWT, is

$$\mathrm{E}(Y_t \mid x_t = k, \hat{\eta}_t) = \exp\{\eta_t^\star + (v_e + v_\epsilon)/2\}.$$

So we write

$$\log\{\mathrm{E}(\mathrm{Y_t} \mid \mathrm{x_t} = \mathrm{k}, \hat{\eta}_\mathrm{t})\} = \hat{\eta}_\mathrm{t} = \eta_\mathrm{t}^\star + (\mathrm{v_e} + \mathrm{v}_\epsilon)/2$$

where, from (4.32),

$$\eta_t^\star = \eta_0^k + \sum_{f=1}^{F} \left[a_f \cos\left(\frac{2\pi f t}{365.25}\right) + b_f \sin\left(\frac{2\pi f t}{365.25}\right) \right].$$

To find the marginal prior distribution of $\eta_t^\star = \mathrm{E}(\mathrm{Z_t})$ we can use the first part of the method for the lognormal distribution in Section 4.3.3.1, and questions **Q11**, **Q12** and **Q13**.

To elicit beliefs about v_e, τ and ϕ_ϵ, we need to consider rainfall on, at least, three wet days, $t, t-k_1$ and $t-k_2$, where, for example, $k_1 = 1$ and $k_2 = 3$. Given the model parameters and LWT, the log rainfall amounts, Z_t, Z_{t-k_1} and Z_{t-k_2} on these days have approximately a multivariate normal distribution. Al-Awadhi & Garthwaite (1998, 2001) describe a method for eliciting prior distributions for the parameters of multivariate normal distributions. However, in our case, the multivariate normal distributions has a specific, time-series, structure which does not correspond well with this method. We propose instead that the expert is asked to imagine a large number of replications of sequences of wet days at a similar time of year and under the same atmospheric conditions. The variance of Z_t is $v_e + v_\epsilon$. The covariance of Z_t and Z_{t-k} is $\phi_\epsilon^k v_\epsilon$. The corresponding correlation is

$$\mathrm{Corr}(\mathrm{Z_t}, \mathrm{Z_{t-k}}) = \frac{\phi_\epsilon^\mathrm{k} \mathrm{v}_\epsilon}{\mathrm{v_e} + \mathrm{v}_\epsilon}.$$

The expert's quartiles for the corresponding sample covariances from a large number of replications can be elicited. Alternatively, the correlation may be expressed in terms of the reduction in conditional variance of Z_t given knowledge of Z_{t-k}. We then give the parameters distributions as follows:

$$\begin{aligned} \tau &\sim \mathrm{Ga}(g_\tau, h_\tau), \\ \alpha = [\exp(v_e) - 1]^{-1} &\sim \mathrm{Ga}(g_\alpha, h_\alpha), \\ v = (\phi_\epsilon + 1)/2 &\sim \mathrm{Beta}(a_v, b_v). \end{aligned} \tag{4.36}$$

The hyperparameters $g_\tau, h_\tau, g_\alpha, h_\alpha, a_v$ and b_v are varied iteratively until the specified quartiles for the variance and covariances are reasonably matched by the implied quartiles

which are numerically computed. The relationship $\alpha = [\exp(v_e) - 1]^{-1}$ in (4.36) arises from considering the case where $\epsilon_t = 0$, in which case

$$\begin{aligned} \mathrm{E}(\mathrm{Y_t} \mid \epsilon_\mathrm{t} = 0) &= \exp\{\eta_t^\star + v_e/2\} = \alpha/\beta \qquad \text{and} \\ \mathrm{Var}(\mathrm{Y_t} \mid \epsilon_\mathrm{t} = 0) &= \exp\{2\eta_t^\star + v_e\}(\exp\{v_e\} - 1) = \alpha/\beta^2. \end{aligned}$$

For illustration, we have chosen carefully suitable values for the hyperparameters of the unknown parameters in the amount process using the above methods as follows:

$$\begin{aligned} m_{A0} &= 0.69, \quad v_{A0} = 4.22, \\ g_\alpha &= 4.31, \qquad h_\alpha = 5.78, \\ g_\tau &= 3.69, \qquad h_\tau = 1.68, \\ a_v &= 3.7, \qquad b_v = 1.9. \end{aligned}$$

As in Section 4.3.3.3, we can elicit beliefs about rainfall amounts at several different times of year in order to develop a prior distribution for the Fourier coefficients. In our illustration we use the same values as used there.

The prior specification for the unknown parameters of the occurrence process is more cumbersome. However we can use a method similar to that in Section 4.3.3.2. If we consider rainfall occurence on a day of the year when we believe that the seasonal effects in (4.33) will cancel out, we might then represent the occurrence process using a simple logistic regression:

$$\mathrm{logit}(p) = \log\left(\frac{p}{1-p}\right) = \zeta_0 + \zeta_1 \log(y).$$

The second term on the right hand side has two effects which are observable, at least in theory. Firstly it causes a change in the distribution of observed rainfall amounts but direct prior beliefs about such a subtle change are unlikely to be reliably elicited. Secondly, because Y_t and Y_{t-1} are autocorrelated, it causes dependence between successive rainfall occurrences. However this relationship is somewhat complicated.

Our approach is as follows. We substitute $\log Y_t = \eta_t^\star + e_t + \epsilon_t$. Then, as a first step, we consider different values of $\eta_t^\star$ with e_t and ϵ_t assumed to be zero. We use a similar approach as in Questions **Q22** and **Q23** in Section 4.3.3.2 to elicit a prior for unknown parameters in the occurrence process in this situation. In this case, though, we do not condition on it having rained on the preceding day. It is then possible to refine the elicitation by using an iterative numerical procedure considering the effect of the distribution of e_t and ϵ_t. We

have chosen carefully two different scenarios for illustration as follows:

$$K_{p1} = \text{logit}(p_1) = \zeta_0 + \log(0.2)\zeta_1,$$
$$K_{p2} = \text{logit}(p_2) = \zeta_0 + \log(0.6)\zeta_1$$

where we assess that

$$\Pr(p_1 < 0.139) = \Pr(p_1 > 0.432) = 0.25,$$
$$\Pr(p_2 < 0.335) = \Pr(p_2 > 0.806) = 0.25.$$

Suppose that $K_{p1} \sim N(m_{K_{p1}}, \sigma^2_{K_{p1}})$. From the specification for p_1 we have

$$\Pr(K_{p1} < \text{logit}\, 0.139 = -1.824) = 0.25,$$
$$\Pr(K_{p1} > \text{logit}\, 0.432 = -0.274) = 0.25$$

and using the properties of the normal distribution, we can get

$$m_{K_{p1}} - 0.674\sigma_{K_{p1}} = -1.824, \qquad m_{K_{p1}} + 0.674\sigma_{K_{p1}} = -0.274$$

where $m_{K_{p1}} = (-1.824 - 0.274)/2 = -1.049$ and $\sigma^2_{K_{p1}} = [(1.824 - 0.274)/2 \times 0.674]^2 = 1.33$. Similar calculations are given to K_{p2} where we have

$$K_{p1} \sim N(-1.05, 1.33)$$
$$K_{p2} \sim N(0.37, 2.44).$$

Suppose that K_{p1} and K_{p2} is independent and therefore this can be summarised as follows:

$$\text{E}(\zeta_0 - 1.6094\zeta_1) = -1.05$$
$$\text{E}(\zeta_0 - 0.5108\zeta_1) = 0.37$$

and

$$\text{Var}(\zeta_0 - 1.6094\zeta_1) = 1.33$$
$$\text{Var}(\zeta_0 - 0.5108\zeta_1) = 2.44$$
$$\text{Cov}(\zeta_0 - 1.6094\zeta_1, \zeta_0 - 0.5108\zeta_1) = 0.$$

By solving the linear equations above, we obtain

$$\begin{pmatrix} \mathrm{E}(\zeta_0) \\ \mathrm{E}(\zeta_1) \end{pmatrix} = \begin{pmatrix} 1.03 \\ 1.29 \end{pmatrix}$$

and

$$\begin{pmatrix} \mathrm{Var}(\zeta_0) & \mathrm{Cov}(\zeta_0, \zeta_1) \\ \mathrm{Cov}(\zeta_0, \zeta_1) & \mathrm{Var}(\zeta_1) \end{pmatrix} = \begin{pmatrix} 5.52 & 3.82 \\ 3.82 & 3.12 \end{pmatrix}.$$

We can use the method of Section 4.3.3.3 to determine a prior distribution for the Fourier coefficients.

4.4.5.2 Fitting the model

The Rjags package (Plummer, 2012) is used to compute the posterior samples for this model. The MCMC algorithm was run for 10000 iterations to allow for burn-in, and an additional 150000 iterations were obtained as the posterior samples. The computing time that was used to obtain 150000 posterior draws was around 38 hours by using R software on a 3.40GHz Ergo Desktop AS4 All-in-One with Intel Core i7-3770 processor and 8 Gbytes of random-access memory. The trace plots can be used as a visual diagnostic to monitor the convergence of the sampler. In this model, the trace plots for all parameters were considered before we make any inferences or conclusions. Figure 4.27 displays the example of trace plots and posterior density of the parameters for the Anticyclonic type (type 1) with some other parameters. An issue that has to be considered when using MCMC methods is that the sample values from the stationary distribution are not independent. The trace plots in Figure 4.27, for the parameters α and τ indicate that these parameters have a relatively high autocorrelation, which suggests mixing is slow. The chain moves more slowly around the distribution, and so more samples are required. Thus, 150000 iterations were considered necessary to draw posterior samples. Poor mixing can affect posterior summary statistics, and obviously we want the sampled numerical summaries to be as close to the true posterior values as possible. We can observe that the samples for the other parameters are not as strongly autocorrelated and the mixing is good in the same figure. The uncertainty associated with most of the unknown parameters has been reduced after we combined the prior beliefs with the data. For example, there is a reduction in the variability of parameter ζ_1 when a comparison is made between the prior and posterior distributions. The posterior mean of parameter ζ_1 is shifted to the left of its prior mean. The full summaries of posterior means and standard deviations of the unknown parameters for the amount and occurrence processes are given in Table 4.5 and 4.6.

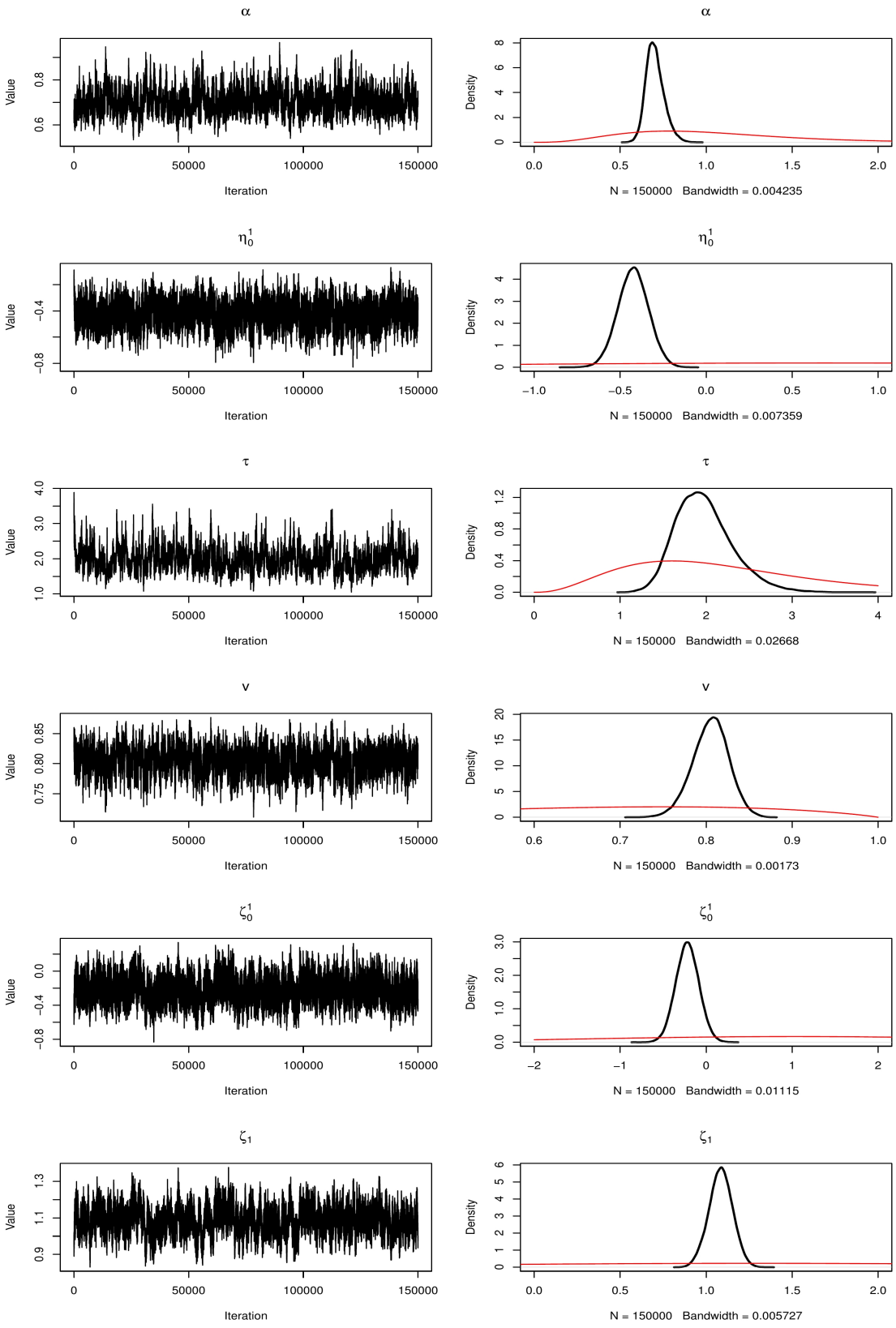


Figure 4.27: The trace and density plots for parameters α, η_0^1, τ, v, ζ_0^1 and ζ_1

Amount process				
Parameter	Prior mean	Prior SD	Posterior mean	Posterior SD
α	0.75	0.36	0.717	0.056
η_0^1	0.69	2.05	-0.411	0.085
η_0^2	0.69	2.05	0.776	0.236
η_0^3	0.69	2.05	0.125	0.253
η_0^4	0.69	2.05	-0.229	0.241
η_0^5	0.69	2.05	-0.756	0.226
η_0^6	0.69	2.05	-0.735	0.188
η_0^7	0.69	2.05	-0.395	0.147
η_0^8	0.69	2.05	0.392	0.164
η_0^9	0.69	2.05	0.305	0.186
η_0^{10}	0.69	2.05	0.846	0.163
η_0^{11}	0.69	2.05	0.531	0.179
η_0^{12}	0.69	2.05	-0.045	0.159
η_0^{13}	0.69	2.05	-0.452	0.128
η_0^{14}	0.69	2.05	-0.030	0.093
η_0^{15}	0.69	2.05	0.074	0.081
η_0^{16}	0.69	2.05	0.410	0.093
η_0^{17}	0.69	2.05	1.140	0.118
η_0^{18}	0.69	2.05	1.116	0.066
η_0^{19}	0.69	2.05	1.354	0.244
η_0^{20}	0.69	2.05	1.093	0.250
η_0^{21}	0.69	2.05	0.875	0.210
η_0^{22}	0.69	2.05	0.448	0.158
η_0^{23}	0.69	2.05	0.541	0.137
η_0^{24}	0.69	2.05	0.773	0.136
η_0^{25}	0.69	2.05	1.006	0.152
η_0^{26}	0.69	2.05	1.258	0.170
η_0^{27}	0.69	2.05	0.339	0.203
a_1	0	0.80	0.033	0.044
a_2	0	0.56	0.049	0.042
a_3	0	0.46	0.011	0.042
b_1	0	0.80	-0.106	0.041
b_2	0	0.56	-0.001	0.042
b_3	0	0.46	0.011	0.042
τ	2.2	1.14	1.967	0.328
v	0.66	0.18	0.803	0.021

Table 4.5: The prior and posterior means with standard deviations (SD) of the unknown parameters for the amount process

Occurrence process				
Parameter	Prior mean	Prior SD	Posterior mean	Posterior SD
ζ_0^1	1.03	2.35	-0.256	0.131
ζ_0^2	1.03	2.35	0.414	0.401
ζ_0^3	1.03	2.35	0.717	0.410
ζ_0^4	1.03	2.35	-0.150	0.359
ζ_0^5	1.03	2.35	0.077	0.337
ζ_0^6	1.03	2.35	-0.220	0.272
ζ_0^7	1.03	2.35	0.181	0.228
ζ_0^8	1.03	2.35	0.047	0.253
ζ_0^9	1.03	2.35	0.516	0.304
ζ_0^{10}	1.03	2.35	1.423	0.324
ζ_0^{11}	1.03	2.35	1.170	0.314
ζ_0^{12}	1.03	2.35	0.337	0.247
ζ_0^{13}	1.03	2.35	0.157	0.186
ζ_0^{14}	1.03	2.35	0.530	0.151
ζ_0^{15}	1.03	2.35	0.651	0.145
ζ_0^{16}	1.03	2.35	1.029	0.179
ζ_0^{17}	1.03	2.35	1.222	0.239
ζ_0^{18}	1.03	2.35	1.291	0.137
ζ_0^{19}	1.03	2.35	0.659	0.490
ζ_0^{20}	1.03	2.35	1.247	0.521
ζ_0^{21}	1.03	2.35	1.011	0.395
ζ_0^{22}	1.03	2.35	0.938	0.285
ζ_0^{23}	1.03	2.35	1.175	0.262
ζ_0^{24}	1.03	2.35	1.262	0.269
ζ_0^{25}	1.03	2.35	1.378	0.312
ζ_0^{26}	1.03	2.35	1.179	0.358
ζ_0^{27}	1.03	2.35	0.906	0.348
c_1	0	0.80	0.369	0.054
c_2	0	0.56	-0.034	0.053
c_3	0	0.46	0.005	0.053
d_1	0	0.80	0.107	0.053
d_2	0	0.56	0.104	0.053
d_3	0	0.46	-0.037	0.053
ζ_1	1.29	1.77	1.060	0.068

Table 4.6: The prior and posterior means with standard deviations (SD) of the unknown parameters for the occurrence process

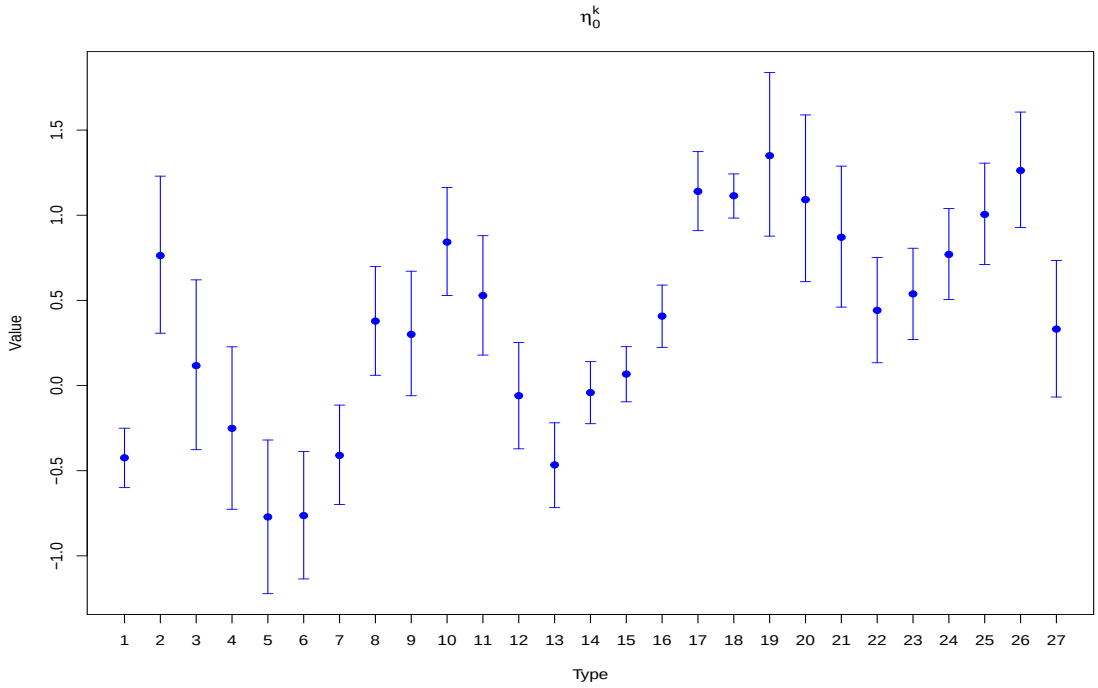


Figure 4.28: The posterior mean with 95% credible intervals for η_0^k where $k \in \{1, \cdots, 27\}$

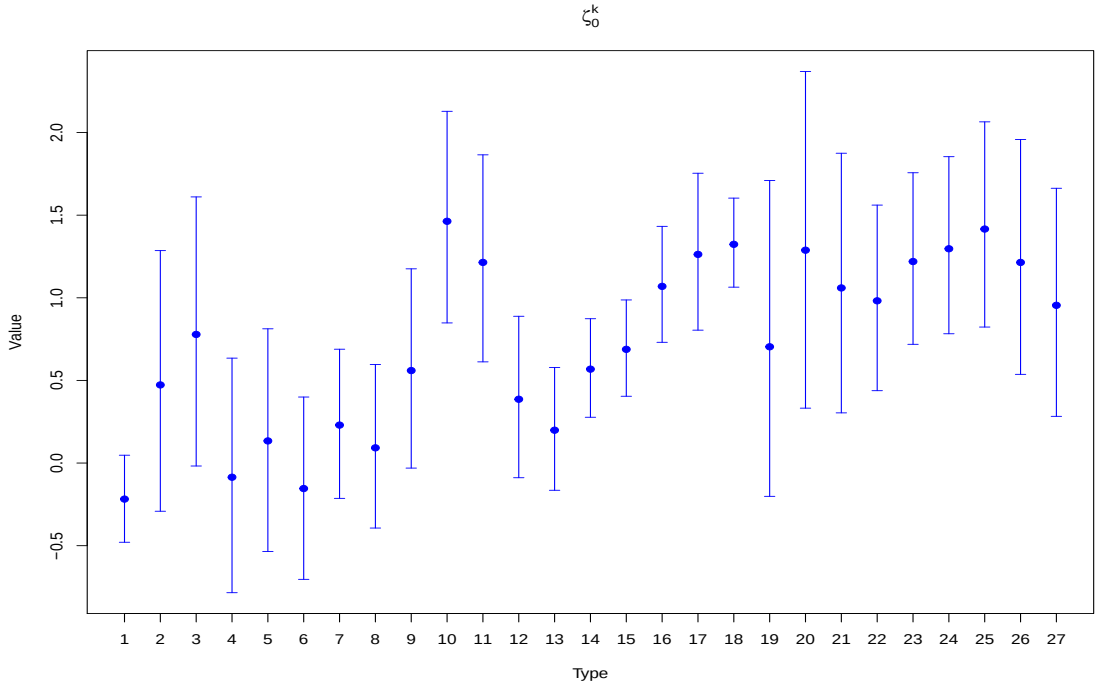


Figure 4.29: The posterior mean with 95% credible intervals for ζ_0^k where $k \in \{1, \cdots, 27\}$

Figure 4.28 shows the posterior mean and 95% credible intervals of the parameter η_0^k for every type $k \in \{1, \cdots, 27\}$. In this figure, the cyclonic types (18-26) have larger posterior means of η_0^k with the means being greater than 0.5. This finding is consistent with the early inference as shown in Figure 4.20a. The LWTs thus play an important role in influencing the quantity of rainfall. For example, the cyclonic type might produce a greater rainfall amount than other types over a year. Also, the previous random effect at time $t-1$ is also vital in determining the level of the rainfall amount at time t. It is highly possible that the rainfall amount will increase if the rainfall does occur on the previous day. Figure 4.29 shows that the posterior mean of ζ_0^k varies between types $k \in \{1, \cdots, 27\}$, and thus the LWT will determine the rainfall probability at time t. This can be illustrated by referring to the Figure 4.20b where the cyclonic types are associated with the greatest levels of rainfall probability which is also consistent with the early inference.

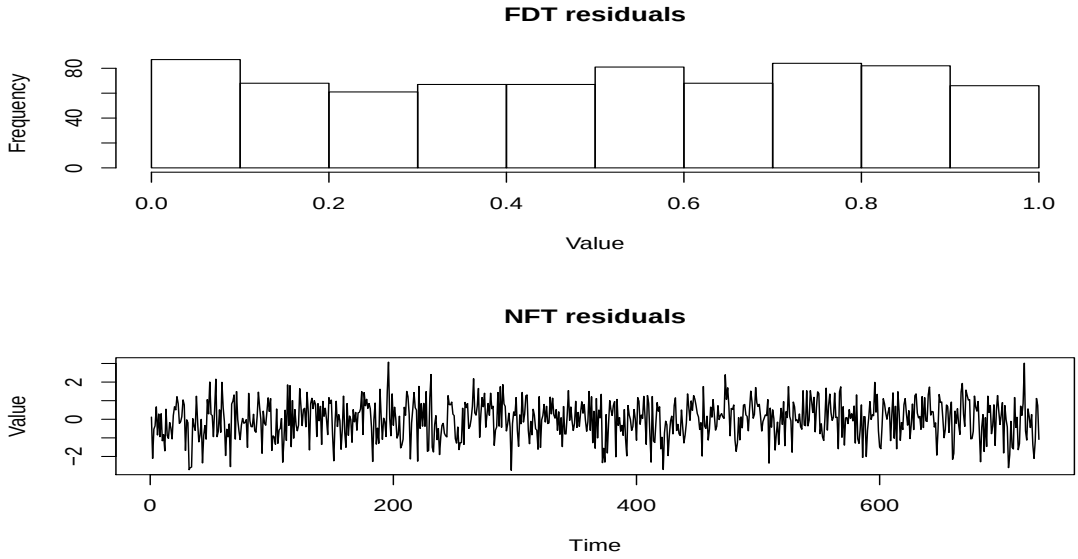


Figure 4.30: The FDTR and NFTR plots for British daily rainfall

Section 4.2.6 describes the diagnostic checking methods used for the mixed distribution based on Smith (1985). Therefore, we will use this approach to assess the adequacy of our model by generating the posterior mean of the FDTR and NFTR from year 1967 until 1968. The graphs of these residuals are shown in Figure 4.30. In these figures, we can observe that these residuals conform to the respective series of independent uniform and independent normal observations. We can further examine the model by looking at the quantile-quantile (QQ), ACF and PACF plots for the NFTR in Figure 4.31. The linearity of the points in the QQ plot indicates that the NFTR are normally distributed. The ACF and PACF plots demonstrate that the residuals are not highly correlated. The evidence from the plots suggests that the model fits the data very well.

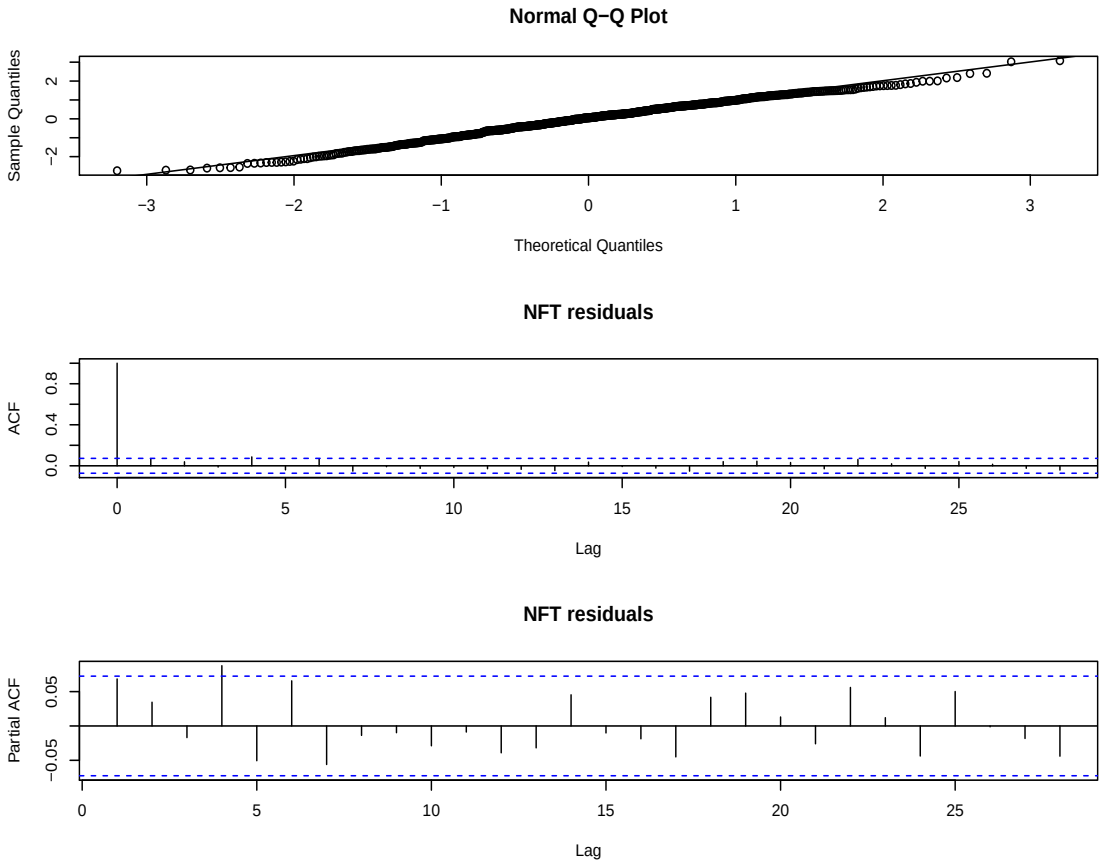


Figure 4.31: QQ, ACF and PACF plots for British daily rainfall

4.4.6 Conclusion

A daily rainfall model was presented in this application by using British daily rainfall. The objective of this model is to capture the relationship that exists between the atmospheric circulation pattern (LWTs) and daily rainfall. This model can be considered more complicated compared to the Italian daily rainfall application since it involves more variables and parameters. Based on the MCMC results, we can consider that the proposed model fits the data very well. The amount and occurrence processes indicate that they are highly dependent on the characteristics of the LWT. Thus, the rainfall amounts and probabilities vary according to the type of the LWT occurring at time t.

In the next chapter, we will describe the extension of the British daily rainfall model to include spatiotemporal effects based on multiple locations. In the model, inter-site spatial correlations will be included and this will hence be our main focus. Moreover, we will also investigate the effects of spatial correlations on the occurrence and amount processes at different sites.

4.5 Summary

In this chapter, we have described the construction of a general daily rainfall model for a single site within the Bayesian framework. Modelling daily rainfall can be quite challenging since there is a mixture of discrete and continuous components in the data. In the general model, we have emphasised the relationship between the amount and occurrence processes using a two-stage approach as shown in Section 4.2. We have also described methods that we can use to construct a model for the amount and occurrence processes in Sections 4.2.2 and 4.2.3. An important novelty in this study is dealing with seasonal effects as the data obtained spans the whole year. We have introduced two different types of Fourier series in Section 4.2.4 to take into account the seasonal variability over a year. Then, we applied this general model to the Italian and British datasets.

For the Italian daily rainfall application in Section 4.3, we have proposed to use three different distributions for the amount process: lognormal distribution, G1 distribution and G2 distribution (for details, see Section 4.3.2.1). From these three distributions, it was found that both the G1 and G2 distributions fitted the data slightly better compared to a lognormal distribution based on the posterior mean for the mean of the potential rainfall amount and posterior predictive values. Since both G1 and G2 distributions are gamma distributions, we can observe that the gamma distribution is a better choice for the amount process modelling for the Italian rainfall data. Furthermore, we have discovered that the shape parameter does not change noticeably over the year while the scale parameter has clear-cut shifts as comparisons were made from month to month as shown in Section 4.3.3.6. For the occurrence process, we have used a first-order Markov chain to fit the rainfall occurrence where the rainfall occurrence at current time t is dependent on that of the previous time $t-1$. The detailed parameterisation for the occurrence process can be found in Section 4.3.2.2.

We have also built a model for the British daily rainfall which is different from the Italian application by incorporating the atmospheric circulation patterns in the model. In the British model, we have chosen the LWTs to represent the atmospheric circulation patterns. The LWT is incorporated into the model through the amount and occurrence processes. Our aim for this model is to build an alternative model from the previous work by Heaps *et al.* (2015) and Germain (2010). These authors used the LWT to characterise atmospheric circulation patterns by combining it with the weather state process using a NHMM approach for the UK winter rainfall. This model is entirely different from our model where we applied the LWT directly to the model instead of using hidden weather states and a NHMM approach. We believe that the LWT plays a critical role in determining rainfall amounts and occurrences over the British Isles as we can see in

Section 4.4.2. Furthermore, we used data from all times of year to generate the British daily rainfall model where the Fourier series is used to allow for seasonal variability in the data. To build a model for the amount process, we have used a gamma distribution where we introduced a random effect at time t which is conditionally dependent on the previous time $t-1$. For the occurrence process, we proposed a simple parameterisation to calculate the probability of rainfall occurrence. As we can see in Section 4.4.4, the parameterisation for the amount and occurrence processes of the British application is in contrast with the Italian application. The MCMC results have indicated that the model fits the data very well as shown in Section 4.4.5.2. It also shows that the level of the potential rainfall amount and the probability of rainfall occurrence are relatively dependent on the type of LWT. In the next chapter, we will describe the extension of this British daily rainfall model with multiple sites.

Chapter 5

Spatiotemporal Model for Daily Rainfall Data

5.1 Introduction

Following the introduction of the daily rainfall model in Chapter 4, we will now introduce a spatiotemporal model which is the extended version of the previous univariate model. The purpose of this chapter is hence to present an alternative spatiotemporal model to the previous models developed by Heaps *et al.* (2015) and Germain (2010) who used non-homogeneous hidden Markov models to fit a model for the winter daily rainfall data in the United Kingdom (UK). In their modelling framework, the atmospheric circulation patterns were linked to the latent weather states to model their influence on the daily rainfall pattern. Instead of using hidden weather states, we propose a model in which we can directly link the atmospheric circulation patterns to the rainfall occurrences and amounts via the conditional model parameters. We also aim to capture the relationship that exists between the sites and investigate how this relationship can affect the daily rainfall pattern at neighboring sites.

Section 5.2 describes the general information about the spatiotemporal model. The exploration of the daily rainfall dataset and the spatial characteristics of each location will be subsequently discussed in Section 5.3. Next, we introduce the spatiotemporal model for UK daily rainfall in Section 5.4. Similar to the univariate model, the Lamb weather types (LWTs) will be incorporated into the model via the amount and occurrence processes. This will lead to a more complex model that requires extensive computational time. The detailed prior and posterior distributions for the model are described in Sections 5.5 and 5.6, respectively.

5.2 Description of the spatiotemporal model

Let $\boldsymbol{W}_t = (W_t(1), W_t(2), \cdots, W_t(S))'$ be a random vector for the observed rainfall amounts which occur at time, t ($t = 1, \cdots, T$) and at sites $s = 1, \cdots, S$. The rainfall amount, $W_t(s)$, can either be zero ($W_t(s) = 0$) if no rain was recorded or positive ($W_t(s) > 0$) if rain was recorded. The notation for rainfall occurrence is thus given by:

$$R_t(s) = \begin{cases} 1 & \text{if it rains, } W_t(s) > 0 \\ 0 & \text{if it does not rain, } W_t(s) = 0, \end{cases}$$

where $\boldsymbol{R}_t = (R_t(1), R_t(2), \cdots, R_t(S))'$ is an indicator vector of rainfall occurrence at time t for sites $s = 1, \cdots, S$. The observed rainfall at site s on day t can therefore be defined as:

$$W_t(s) = R_t(s)Y_t(s) \tag{5.1}$$

where $Y_t(s) = g\left(Z_t(s)\right)$ is a continuous random variable representing the potential rainfall amount at time t and site s, and $g(.)$ is some monotonic function defining a suitable transformation (e.g. exponential) together with a transformed value, $Z_t(s)$. The value of $Y_t(s)$ is always positive but not always observed. Thus, $W_t(s) = Y_t(s)$ if and only if $W_t(s) > 0$. Otherwise $Y_t(s)$ does not play any role. The probability of rainfall occurrence is accordingly given by $\Pr(W_t(s) > 0) = p_1(s,t)$ and as a result, $\Pr(W_t(s) = 0) = p_0(s,t) = 1 - p_1(s,t)$. Let the collections of the observed values of $\boldsymbol{W}_t$, $\boldsymbol{Y}_t$ and $\boldsymbol{R}_t$ be $\boldsymbol{w}$, $\boldsymbol{y}$ and $\boldsymbol{r}$.

In UK climatology, the LWTs have been used extensively for characterising the atmospheric circulation patterns, making them a natural choice of atmospheric variable for downscaling models. Heaps *et al.* (2015) suggested using the objective LWTs for UK climatology since they have been used extensively for many years. The sequence of LWTs is designated by $\boldsymbol{x} = (x_1, x_2, \cdots, x_t)$ where $x_t = k$ with $k \in \{1, 2, \cdots, 27\}$. In this model, a categorical covariate will be used to represent LWTs as a weather state. Hence, X_t represents the weather state on day t and this indicates the rainfall patterns on each day. The detailed descriptions of LWTs have already been elaborated in Chapter 4, Section 4.4.2.

5.3 Data

In this chapter we will use a dataset to illustrate the methods. The dataset consists of daily rainfall observations (measured in millimeters) at a network of 5 sites in the UK. The data are available for 365 days per year and 366 days for each leap year for a whole period of 20 years (7305 days in total). The five selected sites are Balmoral, Darlington South Park, Knighton Victoria Road, Ardleigh Hull Farm, and Bastreet. The maximum

and minimum distances between the sites in this network are 724.8 km and 221.8 km, respectively. A spatial plot of the network of the measurement sites (i.e. the 5 weather stations) is given in Figure 5.1.

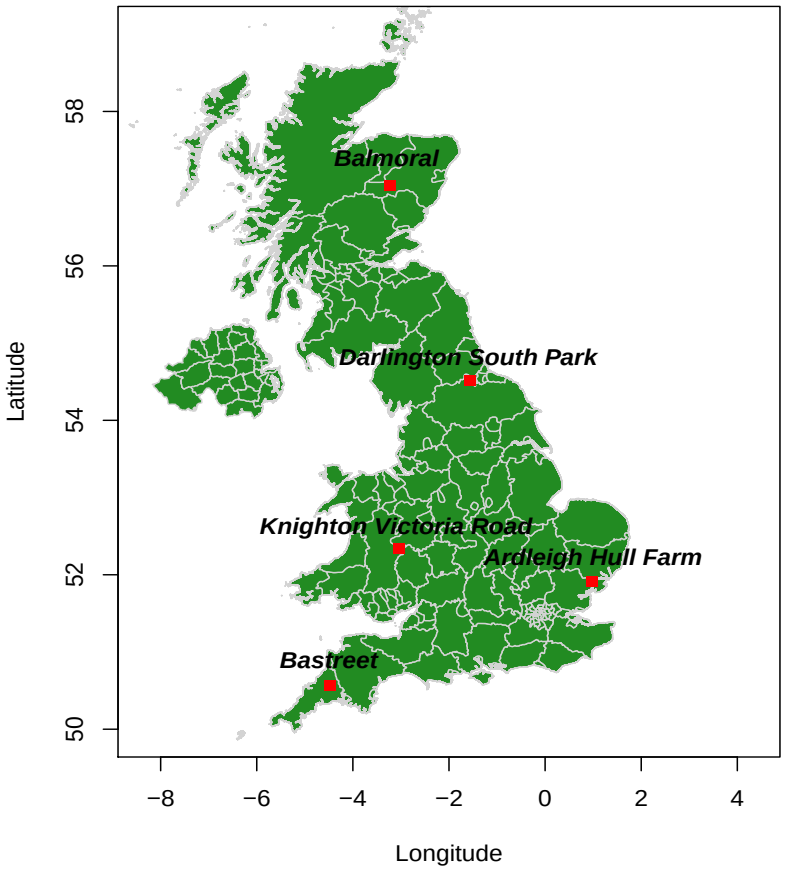


Figure 5.1: The locations of the United Kingdom weather stations chosen as the measurement sites for daily rainfall

As explained in Chapter 4, rainfall data contain zeros and positive values. A dry day is defined as a day in which the rain does not occur or the rainfall amount is insufficient to meet the standard operational definition of a positive rainfall. For this purpose, we use the operational definition adopted by the American Meteorological Society which is that a recorded rainfall amount of less than 0.01 inches or 0.2 mm within a 24-hour period is classified as zero rainfall (Glickman, 2000; Germain, 2010). This cut-off will hence be used to classify a day as either dry or wet as we scrutinise the British rainfall data.

Table 5.1 shows summaries of the altitude, the proportion of wet days, the mean and maximum daily rainfall amounts for each site. Contingent upon this information, the highest proportion of wet days were recorded at Bastreet (66.50%) whilst Darlington South Park had the smallest proportion (49.39%) of wet days. The mean daily rainfall amount

	Site	**Altitude (m)**	**Proportion wet days (%)**	**Mean daily rainfall (mm)**	**Max rainfall amount (mm)**
1	Balmoral	276.82	60.82	2.33	74.7
2	Darlington South Park	44.20	49.39	1.84	48.8
3	Knighton Victoria Road	179.58	65.43	2.75	51.8
4	Ardleigh Hull Farm	34.09	49.50	1.57	47.1
5	Bastreet	270.60	66.50	4.65	114.3

Table 5.1: Summary of data from five sites within the UK from 1966 until 1985

ranges from 1.5 mm to 5.0 mm with the highest mean daily rainfall amount recorded at Bastreet. In addition, the daily rainfall amount may also reach up to nearly 115 mm for a particular site such as Bastreet which recorded a maximal daily rainfall amount of 114.3 mm over the 20-year observation period.

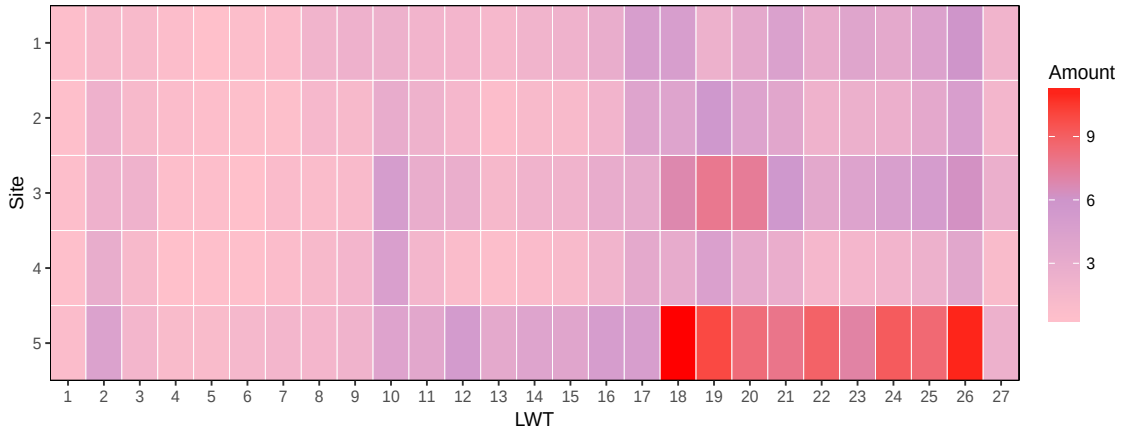
Figure 5.2 depicts the mean daily rainfall amounts on the wet days and the proportions of wet days classified according to the LWTs for all sites. We can observe that the high amounts of daily rainfall in Figure 5.2a are associated with the cyclonic types whereas the low daily rainfall amounts are related to the anticyclonic types. A similar pattern can also be evidently seen for the proportions of wet days (Figure 5.2b), where a higher probability of rain occurrence is associated with the cyclonic types whilst the anticyclonic types are associated with a lower probability of rainfall occurrence.

We can also investigate the spatial structure of the daily rainfall pattern to evaluate whether similar rainfall characteristics are observed between sites. For this endeavour, we take heed of the recommendations made by Germain (2010). Hence, we used the Spearman's rank correlation and log odds ratio to assess the spatial autocorrelations in the rainfall amounts and occurrences. Suppose that we have two sites, i and j, the odds ratio can then be defined as

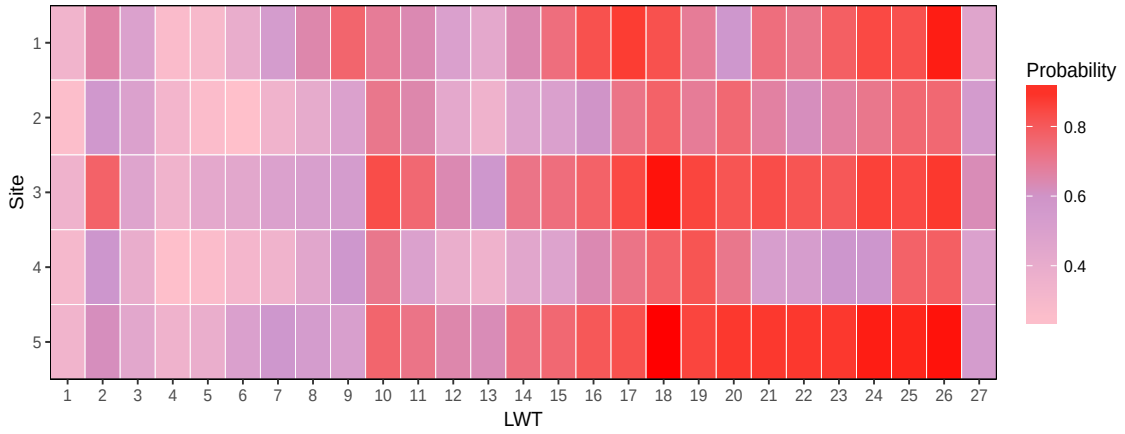
$$\frac{n_{11}}{n_{01}} \Big/ \frac{n_{10}}{n_{00}} = \frac{n_{00}n_{11}}{n_{01}n_{10}} \tag{5.2}$$

where

- n_{00}= the number of dry days at both sites
- n_{11}= the number of wet days at both sites
- n_{01}= the number of dry days at site i and wet days at site j
- n_{10}= the number of wet days at site i and dry days at site j.



(a)



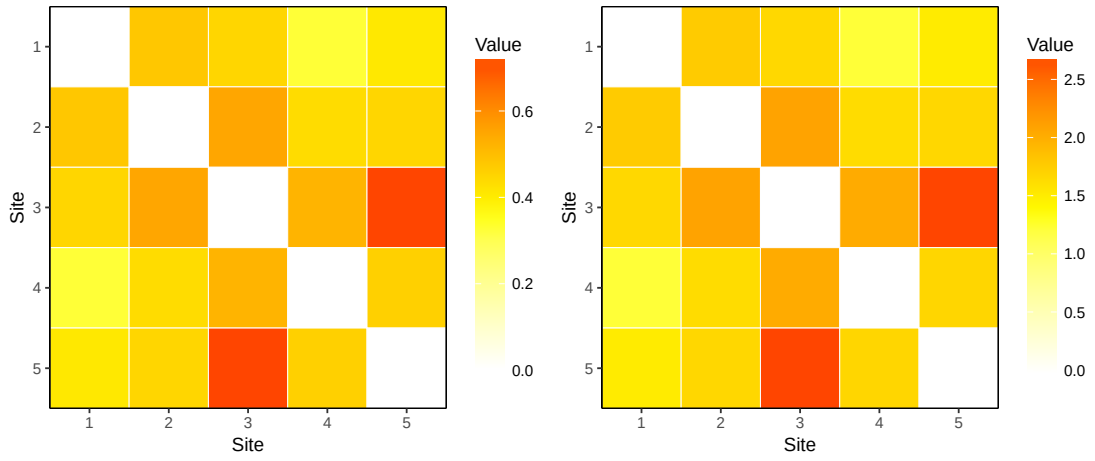
(b)

Figure 5.2: (a) The mean daily rainfall on wet days and (b) the proportion of wet days by the Lamb weather types for five sites within the UK. See Table 4.3 for details.

By taking logarithm on both sides of the (5.2), the log odds ratio is now expressible as:

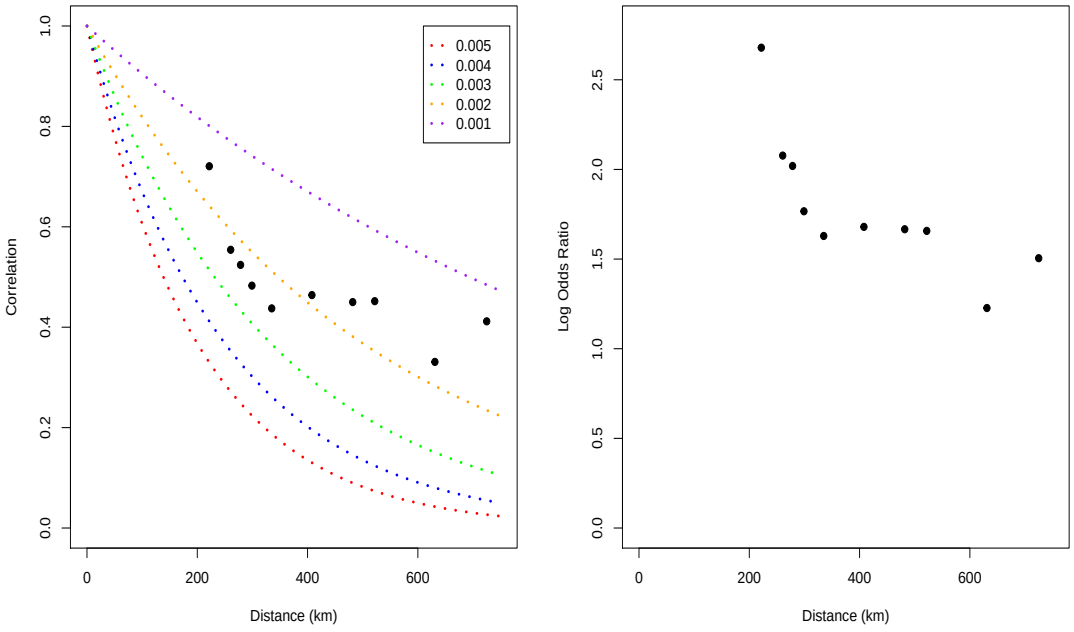
$$\log\left(\frac{n_{00}n_{11}}{n_{01}n_{10}}\right).$$

The correlations between two sites, presented in Figure 5.3a using the Spearman's rank correlation coefficients, show that one of them exceeds 0.6. Figure 5.3b also indicates a similar pattern for the log odds ratio in which the highest correlations are associated with the highest log odds ratio. In addition, we can observe that the correlation and the log odds ratio decrease with distance as shown in Figure 5.4a and 5.4b, respectively. This relationship may suggest that the rainfall patterns are also governed by the distances between the sites. Therefore, a pair of sites might have similar rainfall characteristics if they are in the neighbourhood of each other. Exponential correlation functions, with



(a) (b)

Figure 5.3: (a) Spearman's rank correlation coefficients between non-zero rainfall amounts and (b) log odds ratios for rainfall occurrence for all pairs of sites, UK network



(a) (b)

Figure 5.4: (a) Correlations and (b) log odds ratios against distance, UK network

different values of the decay parameter, are also superimposed on Figure 5.4a. Apart from one point, the point at the greatest distance, the fit to an exponential function with decay parameter around 0.002 seems reasonably good. There is no suggestion that the limit of the correlation as the distance tends to zero is less than 1 so there is no obvious need to include a "nugget" effect.

5.4 The model

5.4.1 Model structure

When building a spatiotemporal model for rainfall, it is natural to provide a more general parameterisation for the model before we modify it to simplify the process. In our case, the daily rainfall patterns assume a mixed distribution. It is therefore crucial to consider the relationship between the amount and occurrence processes. Following the discussion in Chapter 4, we opted for the same approach developed by Sofia (2007) whose conceptual framework relies on making the occurrence process dependent upon the amount process. Consequently, this method is contrary to the model implemented by Heaps *et al.* (2015) and Germain (2010). A similar parameterisation has been applied to a single site in North East England. In this section, we will introduce and develop a spatiotemporal model for daily rainfall which is an extension of the previous univariate daily rainfall model in Chapter 4, Section 4.4.

The resulting model structure is shown in the DAG in Figure 5.5. The structure is essentially the same as that for the univariate British rainfall model in Section 4.4 except that now the nodes represent vector variables.

5.4.2 Modelling the rainfall amount

Following the univariate rainfall model elaborated in Chapter 4, Section 4.4, we will develop a spatiotemporal model for daily rainfall using the concept of the potential daily rainfall amount. Firstly, the potential daily rainfall amount at site s on day t is assumed to follow a gamma distribution:

$$Y_t(s)|x_t = k, \alpha, \mu_t(s) \sim Ga\left(\alpha, \frac{\alpha}{\mu_t(s)}\right) \tag{5.3}$$

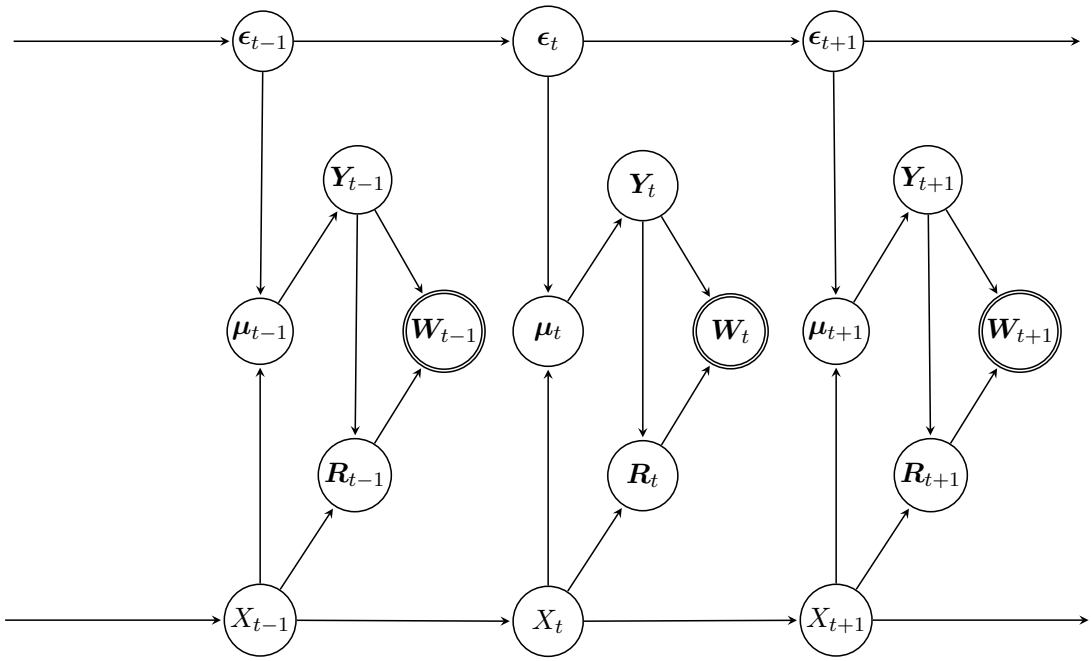


Figure 5.5: A DAG showing the temporal dependence structure of the spatiotemporal model

with a constant shape parameter, α, and variable scale parameter, $\alpha/\mu_t(s)$. We can subsequently employ the GLM approach to model its mean:

$$\log(\mu_t(s)|x_t = k) = \eta_0^k(s) + \sum_{f=1}^{F}\left[a_f^k(s)\cos\left(\frac{2\pi ft}{365.25}\right) + b_f^k(s)\sin\left(\frac{2\pi ft}{365.25}\right)\right] + \epsilon_t(s) \quad (5.4)$$

for $k \in \{1, \cdots, 27\}$. The truncated Fourier series is used in (5.4) to allow for seasonal variation in the rainfall amount. The key assumption here is that $\log(\mu_t(s))$ depends on the spatiotemporal random effect, $\epsilon_t(s)$, which accounts for the correlations between times, t and sites, s. Thus, the distribution of $\mu_t(s)$ is a log-normal distribution. By obtaining the joint density density of $Y_t(s)$ and $\mu_t(s)$ and subsequently integrating out $\mu_t(s)$, we shall then recognize that the marginal distribution of $Y_t(s)$ given other parameters is not a gamma distribution.

There are several approaches that can be used to model $\boldsymbol{\epsilon}_t = (\epsilon_t(1), \epsilon_t(2), \cdots, \epsilon_t(S))'$. For example, Fernandes *et al.* (2009) recommended using a zero mean Gaussian process (GP):

$$\boldsymbol{\epsilon}_t \sim GP(0, \frac{1}{\tau}\rho\left(\|s_i - s_j\|; \phi\right)$$

for $i \neq j$ with a correlation function $\rho\left(.; \phi\right)$ where $\|.\|$ denotes the Eulidean distance. However, this approach does not take into consideration the dependency between time points. Velarde *et al.* (2004) employed conditional autoregressive (CAR) models to represent the

spatiotemporal effects. This approach, introduced by Besag (1974), has been routinely used for mapping disease (Waller & Carlin (2010), Reich *et al.* (2006), and Knorr-Held & Besag (1998)). In our model, we assume each $\boldsymbol{\epsilon}_t$ is dependent on the previous value, $\boldsymbol{\epsilon}_{t-1}$, in a vector autoregression allowing for the temporal dependency. More specifically, suppose $\boldsymbol{\epsilon}_t$ follows a multivariate normal distribution:

$$\boldsymbol{\epsilon}_t|\boldsymbol{\epsilon}_{t-1} \sim N_S\left(\phi_\epsilon \boldsymbol{\epsilon}_{t-1}, \frac{1}{\tau}\boldsymbol{\Sigma}\right) \tag{5.5}$$

for $t > 1$. For $t = 1$, the random effect is given by:

$$\boldsymbol{\epsilon}_1 \sim N_S\left(0, \frac{1}{\tau\left(1-\phi_\epsilon^2\right)}\boldsymbol{\Sigma}\right). \tag{5.6}$$

The marginal mean of $\boldsymbol{\epsilon}_t$ is zero with parameters ϕ_ϵ and τ where $\boldsymbol{\Sigma}$ is the spatial correlation matrix. Banerjee *et al.* (2004) recommended some possible forms for the spatial covariance function and the most popular choice is the exponential covariance function which is given by:

$$\Sigma_{ij} = \exp(-\phi\|s_i - s_j\|) \tag{5.7}$$

with $i \neq j$, $\phi > 0$ and $\|s_i - s_j\|$ is the distance between site s_i and s_j. If $i = j$, the distance is equal to zero. According to Sanso & Guenni (1999), this is the standard correlation function in hydrological applications which ensures that Σ_{ij} is always symmetric and positive definite. Another popular choice for the spatial covariance function is the Matérn correlation function which was originally introduced by Matérn (1986).

This model also incorporates LWTs, x_t, by applying them directly to the parameters of the linear function as shown in (5.4). This modelling strategy is in contrast to the methods used by Heaps *et al.* (2015) and Germain (2010) who introduced the LWTs through the hidden Markov models. Since many parameters are dependent on the LWTs and site, there is a strong possibility of poor identification of some parameters. Therefore, we simplify the parameterisation in (5.4) by allowing only $\eta_0^k(s)$ to depend on the LWTs while the Fourier coefficients are constant over LWTs and sites. Thus, the new parameterisation for $\log\left(\mu_t(s)\right)$ is :

$$\log(\mu_t(s)|x_t = k) = \eta_0^k(s) + \sum_{f=1}^{F}\left[a_f \cos\left(\frac{2\pi f t}{365.25}\right) + b_f \sin\left(\frac{2\pi f t}{365.25}\right)\right] + \epsilon_t(s).$$

5.4.3 Modelling the rainfall occurrence

The rainfall occurrence can only take the values of either 0 or 1, therefore the appropriate probability mass function (pmf) for the occurrence process is a Bernoulli distribution:

$$R_t(s)|x_t = k, y_t(s) \sim Bern\left(p_1(s,t)\right) \tag{5.8}$$

where $p_1(s,t)$ is the rainfall probability at site s on day t. There are several approaches that we can use to model the probability of rainfall which have already been discussed in Chapter 4, Section 4.2.3. The most common approach is a logit link function. An alternative approach is a probit link function which has been used by Heaps *et al.* (2015), Germain (2010) and Thompson *et al.* (2007). We then propose to apply a logit link function in this occurrence process to model the rainfall probability as follows:

$$\begin{aligned}\text{logit}\left(p_1(s,t)|x_t = k\right) =& \zeta_0^k(s) + \sum_{f=1}^{F}\left[c_f^k(s)\cos\left(\frac{2\pi ft}{365.25}\right) + d_f^k(s)\sin\left(\frac{2\pi ft}{365.25}\right)\right] \\ &+ \zeta_1^k(s)\log\left(y_t(s)\right)\end{aligned} \tag{5.9}$$

for $k \in \{1, \cdots, 27\}$. Similar to the amount process, the truncated Fourier series is used to allow for seasonal variation in the rainfall probability. Note that $\text{logit}\left(p_1(s,t)|x_t = k\right)$ is conditioned upon $\log\left(y_t(s)\right)$ to create a link between the amount and occurrence processes. This parameterisation is in contrast to that of Heaps *et al.* (2015) and Germain (2010) who used the rainfall occurrence as a covariate to construct a model for the log-rainfall amounts. We do not include the spatiotemporal effect directly in the occurrence process since the spatiotemporal effect has already been represented by $y_t(s)$.

Similar to the amount process, the original idea is to apply the LWTs directly to all parameters in the logit function (5.9). However, this may lead to poor identification of some parameters. Therefore, we simplify the parameterisation by applying the LWTs only to $\zeta_0^k(s)$ while the Fourier coefficients and ζ_1 are constant over LWTs and sites. Thus, the new parameterisation for the occurrence process is

$$\text{logit}\left(p_1(s,t)|x_t = k\right) = \zeta_0^k(s) + \sum_{f=1}^{F}\left[c_f\cos\left(\frac{2\pi ft}{365.25}\right) + d_f\sin\left(\frac{2\pi ft}{365.25}\right)\right] + \zeta_1\log\left(y_t(s)\right).$$

5.5 Prior distribution

5.5.1 Prior distribution for the amount process

Since the shape parameter, α can only take positive values, it is assigned a gamma distribution:

$$\alpha \sim Ga(g_\alpha, h_\alpha).$$

For the intercept term, $\eta_0^k(s)$, we assign a common prior mean, $m_{\bar{\eta}0}$, for all s and k. In order to generate an appropriate prior covariance structure for the collection of intercept terms, we represent the deviations from the prior mean in terms of a number of zero-mean random variables, $\eta_{01}(k)$, $\eta_{0,2}(s)$ and $\eta_{0,3}(k,s)$, which are mutually independent, in the style of Farrow (2003). Thus we write

$$\eta_0^k(s) = m_{\bar{\eta}0} + \eta_{01}(k) + \eta_{0,2}(s) + \eta_{0,3}(k,s).$$

In the cases of $\eta_{01}(k)$ and $\eta_{0,2}(s)$, we introduce hierarchical structures. By conditioning $\eta_{01}(k)$ on $\check{\eta}_{01}$, we have

$$\eta_{01}(k) \mid \check{\eta}_{01} \sim \mathrm{N}(\check{\eta}_{01}, v_{\eta_1})$$

where

$$\check{\eta}_{01} \sim \mathrm{N}(0, v_{\eta_0})$$

and the variance, covariance and correlation are given by:

$$\begin{aligned} \mathrm{Var}[\eta_{01}(k)] &= v_{\eta_0} + v_{\eta_1}, \\ \mathrm{Covar}[\eta_{01}(k),\ \eta_{0,1}(k')] &= v_{\eta_0}, \\ \mathrm{Corr}[\eta_{01}(k),\ \eta_{0,1}(k')] &= \frac{v_{\eta_0}}{v_{\eta_0} + v_{\eta_1}}. \end{aligned}$$

Similarly, for $\eta_{02}(s)$, we have

$$\eta_{02}(s) \mid \check{\eta}_{02} \sim \mathrm{N}(\check{\eta}_{02}, \ddot{v}_{\eta_1})$$

where

$$\check{\eta}_{02} \sim \mathrm{N}(0, \ddot{v}_{\eta_0})$$

and

$$\begin{aligned} \text{Var}[\eta_{02}(s)] &= \ddot{v}_{\eta_0} + \ddot{v}_{\eta_1}, \\ \text{Covar}[\eta_{02}(s),\ \eta_{0,2}(s)] &= \ddot{v}_{\eta_0}, \\ \text{Corr}[\eta_{02}(s),\ \eta_{0,2}(s)] &= \frac{\ddot{v}_{\eta_0}}{\ddot{v}_{\eta_0} + \ddot{v}_{\eta_1}}. \end{aligned}$$

The terms $\eta_{0,3}(k,s)$ are mutually independent and

$$\eta_{03}(k,s) \sim \text{N}(0,\ \tau_{\eta_3}^{-1}).$$

In the terminology of Farrow (2003), $\eta_{01}(k)$ and $\eta_{02}(s)$ are common uncertainty factors and the terms $\eta_{03}(k,s)$ are specific uncertainty factors.

Let $\ddot{\boldsymbol{\eta}} = (a_1, a_2, b_1, b_2)'$ represent the collection of Fourier coefficients for the amount process. Then, the prior distribution for $\ddot{\boldsymbol{\eta}}$ is given by:

$$\ddot{\boldsymbol{\eta}} \sim N_4\left(\boldsymbol{m}_{\ddot{\boldsymbol{\eta}}}, P_{\ddot{\boldsymbol{\eta}}}^{-1}\right)$$

where $\boldsymbol{m}_{\ddot{\boldsymbol{\eta}}}$ is the mean vector with precision matrix, $P_{\ddot{\boldsymbol{\eta}}}$. For the parameters τ and ϕ, gamma distributions are assigned as their priors:

$$\tau \sim Ga(g_\tau, h_\tau)$$

and

$$\phi \sim Ga(g_\phi, h_\phi).$$

A gamma distribution is a sensible prior choice for τ and ϕ since it can only assume positive values. The prior distribution for ϕ_ϵ must adhere to the stationary condition $|\phi_\epsilon| < 1$. Suppose that a new variable, v, follows a beta distribution:

$$v \sim Beta(a_v, b_v).$$

Since the beta distribution has a range between 0 and 1, we can then fix:

$$\phi_\epsilon = 2v - 1$$

so that the value for ϕ_ϵ is between -1 and 1.

5.5.2 Prior distribution for the occurrence process

The prior distribution for the parameters in the occurrence process is similar to that for the amount process. Let the intercept term $\zeta_0^k(s)$ is given by:

$$\zeta_0^k(s) = m_{\bar{\zeta 0}} + \zeta_{01}(k) + \zeta_{02}(s) + \zeta_{03}(k,s)$$

where $m_{\bar{\zeta 0}}$ is a common prior mean for all s and k. Similar to the amount process, $\zeta_{01}(k)$, $\zeta_{02}(s)$ and $\zeta_{03}(k,s)$ are given to generate an approriate prior covariance structure for the collection of intercept terms.

In the case of $\zeta_{01}(k)$ and $\zeta_{02}(s)$, we introduce hierarchical structures. By conditioning $\zeta_{01}(k)$ upon $\check{\zeta}_{01}$, the distribution of $\zeta_{01}(k)$ is:

$$\zeta_{01}(k)|\check{\zeta}_{01} \sim N(\check{\zeta}_{01}, \upsilon_{\zeta_1})$$

where

$$\check{\zeta}_{01} \sim N(0, \upsilon_{\zeta_0})$$

and the variance, covariance and correlation are given by:

$$\begin{aligned}
\text{Var}[\zeta_{01}(k)] =& \upsilon_{\zeta_0} + \upsilon_{\zeta_1} \\
\text{Covar}[\zeta_{01}(k), \zeta_{01}(k')] =& \upsilon_{\zeta_0} \\
\text{Corr}[\zeta_{01}(k), \zeta_{01}(k')] =& \frac{\upsilon_{\zeta_0}}{\upsilon_{\zeta_0} + \upsilon_{\zeta_1}}.
\end{aligned}$$

Similarly, for $\zeta_{02}(s)$, we have:

$$\zeta_{02}(s)|\check{\zeta}_{02} \sim N(\check{\zeta}_{02}, \ddot{\nu}_{\zeta_1})$$

where

$$\check{\zeta}_{02} \sim N(0, \ddot{\nu}_{\zeta_0})$$

and

$$\begin{aligned}
\text{Var}[\zeta_{02}(s)] =& \ddot{\nu}_{\zeta_0} + \ddot{\nu}_{\zeta_1} \\
\text{Covar}[\zeta_{02}(s), \zeta_{02}(s')] =& \ddot{\nu}_{\zeta_0} \\
\text{Corr}[\zeta_{02}(s), \zeta_{02}(s')] =& \frac{\ddot{\nu}_{\zeta_0}}{\ddot{\nu}_{\zeta_0} + \ddot{\nu}_{\zeta_1}}.
\end{aligned}$$

The terms $\zeta_{03}(k,s)$ are mutually independent and:

$$\zeta_{03}(k,s) \sim N\left(0, 1/\tau_{\zeta_3}^{-1}\right).$$

By following the terminology of Farrow (2003), $\zeta_{01}(k)$ and $\zeta_{02}(s)$ are common uncertainty factors and the terms $\zeta_{03}(k,s)$ are specific uncertainty factors. A normally-distributed prior is also assigned for ζ_1:

$$\zeta_1 \sim N\left(m_{\zeta_1}, 1/\tau_{\zeta_1}\right)$$

since ζ_1 can take any values from $-\infty$ to ∞.

Let the set of Fourier coefficients for the occurrence process be represented by $\ddot{\boldsymbol{\zeta}} = (c_1, c_2, d_1, d_2)'$. Then, the prior distribution for the Fourier coefficients is defined in a similar manner as for the amount process:

$$\ddot{\boldsymbol{\zeta}} \sim N_4\left(\boldsymbol{m}_{\ddot{\boldsymbol{\zeta}}}, P_{\ddot{\boldsymbol{\zeta}}}^{-1}\right)$$

where $\boldsymbol{m}_{\ddot{\boldsymbol{\zeta}}}$ and $P_{\ddot{\boldsymbol{\zeta}}}$ are the mean vector and the precision matrix, respectively.

5.6 Posterior distribution

Let $\boldsymbol{\theta}$ be the collection of unknown parameters. The joint density of $(\boldsymbol{\theta}, \boldsymbol{Y}, \boldsymbol{R})$ is then given by:

$$\pi(\boldsymbol{\theta}, \boldsymbol{y}, \boldsymbol{r}) = \pi(\boldsymbol{\theta}_{amt}) \times \pi(\boldsymbol{\theta}_{occ}) \times f(\boldsymbol{y}|\boldsymbol{\theta}_{amt}) \times f(\boldsymbol{r}|\boldsymbol{y}, \boldsymbol{\theta}_{occ}) \tag{5.10}$$

where $\boldsymbol{\theta}_{amt}$ and $\boldsymbol{\theta}_{occ}$ are the collections of parameters for the amount and occurrence processes, respectively. To evaluate the posterior distribution, we use a Metropolis-within-Gibbs algorithm. We use data augmentation and treat the unobserved potential rainfall amounts $Y_t(s)$ and the random effects $\epsilon_t(s)$ as auxiliary data. We therefore need to find the full conditional distributions of the model parameters and the auxiliary data.

5.6.1 The full conditional distributions for the amount process

From (5.10), we can work out the full conditional distribution (FCD) for each parameter in the amount process. The FCDs in this section are provided only as general information and for future work. Then, the FCD for each parameter are derived as follows:

- Parameter α

  The FCD for the shape parameter, α, is:

$$\begin{aligned}\pi\left(\alpha|\eta_0^k(s),\ddot{\boldsymbol{\eta}},v,\tau,\boldsymbol{y}\right) &\propto \prod_{t=1}^{T}\prod_{s=1}^{S} f\left(y_t(s)|\boldsymbol{\theta}_{amt},\epsilon_t(s)\right)\times\pi\left(\alpha\right)\\ &\propto \prod_{t=1}^{T}\prod_{s=1}^{S}\left[\frac{\alpha}{\mu_t(s)}\right]^{\alpha}\frac{y_t(s)^{\alpha-1}}{\Gamma(\alpha)}\exp\left[-\frac{\alpha y_t(s)}{\mu_t(s)}\right]\times\frac{h_\alpha^{g_\alpha}\alpha^{g_\alpha-1}e^{-h_\alpha\alpha}}{\Gamma(g_\alpha)}\\ &\propto \frac{(\alpha)^{\alpha ST+g_\alpha-1}}{[\Gamma(\alpha)]^{ST}}\prod_{t=1}^{T}\prod_{s=1}^{S}[E_t(s)]^{\alpha}\exp\left\{-\alpha\left(\sum_{t=1}^{T}\sum_{s=1}^{S}E_t(s)+h_\alpha\right)\right\}.\end{aligned}$$

  where $E_t(s)=y_t(s)/\mu_t(s)$.

- Parameter $\eta_0^k(s)$

  We have $\eta_0^k(s) = m_{\bar{\eta}0}+\eta_{01}(k)+\eta_{0,2}(s)+\eta_{0,3}(k,s)$. The FCD of $\eta_0^k(s)$ is represented by:

$$\begin{aligned}&\pi\left(\eta_0^k(s)\middle|\ddot{\boldsymbol{\eta}},v,\tau,\boldsymbol{y},\eta_{01}(k),\eta_{02}(s),\boldsymbol{\epsilon}\right)\\ &\qquad\propto\prod_{t\in T_k}f\left(y_t(s)|\boldsymbol{\theta}_{amt},\epsilon_t(s)\right)\times\pi\left(\eta_0^k(s)|\eta_{01}(k),\eta_{02}(s)\right)\\ &\qquad\propto\prod_{t\in T_k}\left[\frac{\alpha}{\mu_t(s)}\right]^{\alpha}\frac{y_t(s)^{\alpha-1}}{\Gamma\left(\alpha\right)}\exp\left[-\frac{\alpha y_t(s)}{\mu_t(s)}\right]\\ &\qquad\quad\times\sqrt{\frac{\tau_{\bar{\eta}_0}}{2\pi}}\exp\left\{-\frac{\tau_{\bar{\eta}_0}}{2}\left(\eta_0^k(s)-\eta_{01}(k)-\eta_{02}(s)-m_{\bar{\eta}_0}\right)^2\right\}\\ &\qquad\propto\prod_{t\in T_k}[E_t(s)]^{\alpha}\exp\left\{-\left[\alpha\sum_{t\in T_k}E_t(s)\right.\right.\\ &\qquad\quad\left.\left.+\frac{\tau_{\bar{\eta}_0}}{2}\left(\left\{\eta_0^k(s)\right\}^2-2\eta_0^k(s)\left\{\eta_{01}(k)+\eta_{02}(s)+m_{\bar{\eta}_0}\right\}\right)\right]\right\}\end{aligned}$$

  where $T_k=\{t:x_t=k\}$ is the set of observations that belongs to type $k=1,\cdots,27$.

The FCD of $\eta_{01}(k)$ is given by:

$$\begin{aligned}
\pi\Big(\eta_{01}(k)\Big|\eta_0^k(1),\cdots,\eta_0^k(S),&\eta_{02}(1),\cdots,\eta_{02}(S),\check{\eta}_{01}\Big) \\
&\propto \prod_{s=1}^{S} f\left(\eta_0^k(s)|\eta_{01}(k),\eta_{02}(s)\right)\times\pi\left(\eta_{01}(k)|\check{\eta}_{01}\right) \\
&\propto \prod_{s=1}^{S}\sqrt{\frac{\tau_{\bar{\eta}_0}}{2\pi}}\exp\left\{-\frac{\tau_{\bar{\eta}_0}}{2}\left(\eta_0^k(s)-\eta_{01}(k)-\eta_{02}(s)-m_{\bar{\eta}_0}\right)^2\right\} \\
&\quad\times\sqrt{\frac{1}{2\pi\upsilon_{\eta_1}}}\exp\left\{-\frac{1}{2\upsilon_{\eta_1}}\left(\eta_{01}(k)-\check{\eta}_{01}\right)^2\right\} \\
&\propto \exp\left\{-\frac{D_{\eta_{01}}}{2}\left(\eta_{01}(k)-B_{\eta_{01}}\right)^2\right\}
\end{aligned}$$

where

$$D_{\eta_{01}} = S\tau_{\bar{\eta}_0}+\frac{1}{\upsilon_{\eta_1}}$$

and

$$B_{\eta_{01}} = \frac{1}{D_{\eta_{01}}}\left\{\tau_{\bar{\eta}_0}\left[\sum_{s=1}^{S}\eta_0^k(s)-\sum_{s=1}^{S}\eta_{02}(s)-Sm_{\bar{\eta}_0}\right]+\frac{\check{\eta}_{01}}{\upsilon_{\eta_1}}\right\}.$$

This form corresponds to a normal density, $N(B_{\eta_{01}},1/D_{\eta_{01}})$.

The FCD of $\check{\eta}_{01}$ is derived as follows:

$$\begin{aligned}
\pi\Big(\check{\eta}_{01}\Big|\eta_{01}(1),\cdots,\eta_{01}(27)\Big) &\propto \prod_{k=1}^{27} f\left(\eta_{01}(k)|\check{\eta}_{01}\right)\times\pi(\check{\eta}_{01}) \\
&\propto \prod_{k=1}^{27}\sqrt{\frac{1}{2\pi\upsilon_{\eta_1}}}\exp\left\{-\frac{1}{2\upsilon_{\eta_1}}\left(\eta_{01}(k)-\check{\eta}_{01}\right)^2\right\} \\
&\quad\times\sqrt{\frac{1}{2\pi\upsilon_{\eta_0}}}\exp\left\{-\frac{1}{2\upsilon_{\eta_0}}(\check{\eta}_{01})^2\right\} \\
&\propto \exp\left\{\frac{\tilde{D}_{\check{\eta}_{01}}}{2}\left(\check{\eta}_{01}-\tilde{B}_{\check{\eta}_{01}}\right)^2\right\}
\end{aligned}$$

where

$$\tilde{D}_{\check{\eta}_{01}} = \frac{K}{\upsilon_{\eta_1}}+\frac{1}{\upsilon_{\eta_0}}$$

and

$$\tilde{B}_{\check{\eta}_{01}} = \frac{\sum_{k=1}^{27}\eta_{01}(k)}{\tilde{D}_{\check{\eta}_{01}}\upsilon_{\eta_0}}.$$

This form also corresponds to a normal density, $N(\tilde{B}_{\check{\eta}_{01}},1/\tilde{D}_{\check{\eta}_{01}})$.

The FCD of $\eta_{02}(s)$ is represented by:

$$\begin{aligned}
\pi\left(\eta_{02}(s)\middle|\eta_0^1(s),\cdots,\eta_0^{27}(s),\eta_{01}(1),\cdots,\eta_{01}(27),\check{\eta}_{02}\right) & \\
&\propto \prod_{k=1}^{27} f\left(\eta_0^k(s)|\eta_{01}(k),\eta_{02}(s)\right)\times\pi\left(\eta_{02}(s)|\check{\eta}_{02}\right) \\
&\propto \prod_{k=1}^{27}\sqrt{\frac{\tau_{\bar{\eta}_0}}{2\pi}}\exp\left\{-\frac{\tau_{\bar{\eta}_0}}{2}\left(\eta_0^k(s)-\eta_{01}(k)-\eta_{02}(s)-m_{\bar{\eta}_0}\right)^2\right\} \\
&\quad\times\sqrt{\frac{1}{2\pi\nu_{\eta_1}}}\exp\left\{-\frac{1}{2\nu_{\eta_1}}\left(\eta_{02}(s)-\check{\eta}_{02}\right)^2\right\} \\
&\propto \exp\left\{-\frac{D_{\eta_{02}}}{2}\left(\eta_{02}(s)-B_{\eta_{02}}\right)^2\right\}
\end{aligned}$$

where

$$D_{\eta_{02}} = S\tau_{\bar{\eta}_0}+\frac{1}{\nu_{\eta_1}}$$

and

$$B_{\eta_{02}} = \frac{1}{D_{\eta_{02}}}\left\{\tau_{\bar{\eta}_0}\left[\sum_{k=1}^{27}\eta_0^k(s)-\sum_{k=1}^{27}\eta_{01}(k)-Km_{\bar{\eta}_0}\right]+\frac{\check{\eta}_{02}}{\nu_{\eta_1}}\right\}.$$

This is a normal distribution.

For $\check{\eta}_{02}$, the FCD is:

$$\begin{aligned}
\pi\left(\check{\eta}_{02}\middle|\eta_{02}(1),\cdots,\eta_{02}(S)\right) &\propto \prod_{s=1}^{S} f\left(\eta_{02}(s)|\check{\eta}_{02}\right)\times\pi(\check{\eta}_{02}) \\
&\propto \prod_{s=1}^{S}\sqrt{\frac{1}{2\pi\nu_{\eta_1}}}\exp\left\{-\frac{1}{2\nu_{\eta_1}}\left(\eta_{02}(s)-\check{\eta}_{02}\right)^2\right\} \\
&\quad\times\sqrt{\frac{1}{2\pi\nu_{\eta_0}}}\exp\left\{-\frac{1}{2\nu_{\eta_0}}\check{\eta}_{02}\right\} \\
&\propto \exp\left\{\frac{\tilde{D}_{\check{\eta}_{02}}}{2}\left(\check{\eta}_{01}-\tilde{B}_{\check{\eta}_{02}}\right)^2\right\}
\end{aligned}$$

where

$$\tilde{D}_{\check{\eta}_{02}} = \frac{S}{\nu_{\eta_1}}+\frac{1}{\nu_{\eta_0}}$$

and

$$\tilde{B}_{\check{\eta}_{02}} = \frac{\sum_{s=1}^{S}\eta_{02}(s)}{\tilde{D}_{\check{\eta}}\nu_{\eta_0}}.$$

Again, this is the density form of a normal distribution.

- Parameter $\ddot{\boldsymbol{\eta}}$

The FCD for $\ddot{\boldsymbol{\eta}}$ is given by:

$$\begin{aligned}
\pi\left(\ddot{\boldsymbol{\eta}}|\eta_0^k(s), v, \tau, \boldsymbol{y}\right) &\propto \prod_{t=1}^{T}\prod_{s=1}^{S} f\left(y_t(s)|\boldsymbol{\theta}_{amt}, \epsilon_t(s)\right) \times \pi\left(\ddot{\boldsymbol{\eta}}\right) \\
&\propto \prod_{t=1}^{T}\prod_{s=1}^{S}\left[\frac{\alpha}{\mu_t(s)}\right]^{\alpha}\frac{y_t(s)^{\alpha-1}}{\Gamma\left(\alpha\right)}\exp\left[-\frac{\alpha y_t(s)}{\mu_t(s)}\right] \\
&\quad \times (2\pi)^{-2F}|P_{\ddot{\boldsymbol{\eta}}}|^{1/2}\exp\left\{-\frac{1}{2}(\ddot{\boldsymbol{\eta}}-\boldsymbol{m}_{\ddot{\boldsymbol{\eta}}})^T P_{\ddot{\boldsymbol{\eta}}}(\ddot{\boldsymbol{\eta}}-\boldsymbol{m}_{\ddot{\boldsymbol{\eta}}})\right\} \\
&\propto \prod_{t=1}^{T}\prod_{s=1}^{S}\left[E_t(s)\right]^{\alpha}\exp\left\{-\left[\alpha\sum_{t=1}^{T}\sum_{s=1}^{S}E_t(s)+\frac{1}{2}(\ddot{\boldsymbol{\eta}}-\boldsymbol{m}_{\ddot{\boldsymbol{\eta}}})^T P_{\ddot{\boldsymbol{\eta}}}(\ddot{\boldsymbol{\eta}}-\boldsymbol{m}_{\ddot{\boldsymbol{\eta}}})\right]\right\}.
\end{aligned}$$

- Parameter v

The FCD for v can be derived as follows:

$$\begin{aligned}
\pi\left(v|\boldsymbol{\epsilon}, \phi, \tau\right) &\propto f\left(\boldsymbol{\epsilon}_1|v, \phi, \tau\right)\prod_{t=2}^{T} f\left(\boldsymbol{\epsilon}_t|\boldsymbol{\epsilon}_{t-1}, v, \phi, \tau\right) \times \pi\left(v\right) \\
&\propto \sqrt{\frac{\tau\left(1-\phi_\epsilon^2\right)}{(2\pi)^S|\Sigma|}}\exp\left\{-\frac{\tau\left(1-\phi_\epsilon^2\right)}{2}\left(\boldsymbol{\epsilon}_1\right)'\Sigma^{-1}\left(\boldsymbol{\epsilon}_1\right)\right\} \\
&\quad \times \prod_{t=2}^{T}\sqrt{\frac{\tau}{(2\pi)^S|\Sigma|}}\exp\left\{-\frac{\tau}{2}\left(\boldsymbol{\epsilon}_t-\phi_\epsilon\boldsymbol{\epsilon}_{t-1}\right)'\Sigma^{-1}\left(\boldsymbol{\epsilon}_t-\phi_\epsilon\boldsymbol{\epsilon}_{t-1}\right)\right\} \\
&\quad \times \frac{\Gamma(a_v+b_v)}{\Gamma(a_v)\Gamma(b_v)}v^{a_v-1}(1-v)^{b_v-1} \\
&\propto v^{a_v-1}(1-v)^{b_v-1}\sqrt{\left(1-\phi_\epsilon^2\right)}\exp\left\{-\frac{\tau}{2}\left[\left(1-\phi_\epsilon^2\right)\left(\boldsymbol{\epsilon}_1\right)'\Sigma^{-1}\left(\boldsymbol{\epsilon}_1\right)\right.\right. \\
&\quad \left.\left.+\sum_{t=1}^{T}\left(\boldsymbol{\epsilon}_t-\phi_\epsilon\boldsymbol{\epsilon}_{t-1}\right)'\Sigma^{-1}\left(\boldsymbol{\epsilon}_t-\phi_\epsilon\boldsymbol{\epsilon}_{t-1}\right)\right]\right\}.
\end{aligned}$$

- Parameter ϕ

For ϕ, the FCD can be expressed using the following:

$$\begin{aligned}
\pi\left(\phi|\boldsymbol{\epsilon}, v, \tau\right) \propto & f\left(\boldsymbol{\epsilon}_1|v, \phi, \tau\right) \prod_{t=2}^{T} f\left(\boldsymbol{\epsilon}_t|\boldsymbol{\epsilon}_{t-1}, v, \phi, \tau\right) \times \pi\left(\phi\right) \\
\propto & \sqrt{\frac{\tau\left(1-\phi_\epsilon^2\right)}{(2\pi)^S|\Sigma|}} \exp\left\{-\frac{\tau\left(1-\phi_\epsilon^2\right)}{2}\left(\boldsymbol{\epsilon}_1\right)'\Sigma^{-1}\left(\boldsymbol{\epsilon}_1\right)\right\} \\
& \times \prod_{t=2}^{T} \sqrt{\frac{\tau}{(2\pi)^S|\Sigma|}} \exp\left\{-\frac{\tau}{2}(\boldsymbol{\epsilon}_t - \phi_\epsilon\boldsymbol{\epsilon}_{t-1})'\Sigma^{-1}(\boldsymbol{\epsilon}_t - \phi_\epsilon\boldsymbol{\epsilon}_{t-1})\right\} \\
& \times \frac{h_\phi^{g_\phi}\phi^{g_\phi-1}\exp\left[-h_\phi\phi\right]}{\Gamma(g_\phi)} \\
\propto & \phi^{g_\phi-1}\exp\left\{-\left[\frac{\tau}{2}\left[\left(1-\phi_\epsilon^2\right)\left(\boldsymbol{\epsilon}_1\right)'\Sigma^{-1}\left(\boldsymbol{\epsilon}_1\right) + \right.\right.\right. \\
& \left.\left.\left.\sum_{t=2}^{T}(\boldsymbol{\epsilon}_t - \phi_\epsilon\boldsymbol{\epsilon}_{t-1})'\Sigma^{-1}(\boldsymbol{\epsilon}_t - \phi_\epsilon\boldsymbol{\epsilon}_{t-1})\right] + h_\phi\phi\right]\right\}.
\end{aligned}$$

- Parameter τ

The full conditional distribution for τ is:

$$\begin{aligned}
\pi\left(\tau|\boldsymbol{\epsilon}, v, \phi\right) \propto & f\left(\boldsymbol{\epsilon}_1|v, \phi, \tau\right) \prod_{t=2}^{T} f\left(\boldsymbol{\epsilon}_t|\boldsymbol{\epsilon}_{t-1}, v, \phi, \tau\right) \times \pi\left(\tau\right) \\
\propto & \sqrt{\frac{\tau\left(1-\phi_\epsilon^2\right)}{(2\pi)^S|\Sigma|}} \exp\left\{-\frac{\tau\left(1-\phi_\epsilon^2\right)}{2}\left(\boldsymbol{\epsilon}_1\right)'\Sigma^{-1}\left(\boldsymbol{\epsilon}_1\right)\right\} \\
& \times \prod_{t=2}^{T} \sqrt{\frac{\tau}{(2\pi)^S|\Sigma|}} \exp\left\{-\frac{\tau}{2}(\boldsymbol{\epsilon}_t - \phi_\epsilon\boldsymbol{\epsilon}_{t-1})'\Sigma^{-1}(\boldsymbol{\epsilon}_t - \phi_\epsilon\boldsymbol{\epsilon}_{t-1})\right\} \\
& \times \frac{h_\tau^{g_\tau}\tau^{g_\tau-1}\exp\left[-h_\tau\tau\right]}{\Gamma(g_\tau)} \\
\propto & \tau^{g_\tau+\frac{T}{2}-1}\left\{-\tau\left[\frac{1}{2}\left(1-\phi_\epsilon^2\right)\left(\boldsymbol{\epsilon}_1\right)'\Sigma^{-1}\left(\boldsymbol{\epsilon}_1\right)\right.\right. \\
& \left.\left. + \frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{\epsilon}_t - \phi_\epsilon\boldsymbol{\epsilon}_{t-1})'\Sigma^{-1}(\boldsymbol{\epsilon}_t - \phi_\epsilon\boldsymbol{\epsilon}_{t-1}) + h_\tau\right]\right\}.
\end{aligned}$$

This form corresponds to a gamma density, $Ga(G_\tau, H_\tau)$ where

$$G_\tau = g_\tau + \frac{T}{2}$$

and

$$H_\tau = \left[\frac{1}{2}\left(1-\phi_\epsilon^2\right)(\boldsymbol{\epsilon}_1)'\,\Sigma^{-1}(\boldsymbol{\epsilon}_1) + \frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{\epsilon}_t-\phi_\epsilon\boldsymbol{\epsilon}_{t-1})'\Sigma^{-1}(\boldsymbol{\epsilon}_t-\phi_\epsilon\boldsymbol{\epsilon}_{t-1}) + h_\tau\right].$$

- Random effect $\boldsymbol{\epsilon}_t$

To obtain the sampled values for the random effect $\boldsymbol{\epsilon}_t$, the FCD for $\boldsymbol{\epsilon}_t$ is worked out as follows:

$$\begin{aligned}
\pi(\boldsymbol{\epsilon}_t|\phi_\epsilon,\phi,\tau,\boldsymbol{y}_t) &\propto \prod_{s=1}^{S} f(y_t(s)|\boldsymbol{\theta}_{amt},\boldsymbol{\epsilon}_t)\times\pi(\boldsymbol{\epsilon}_t|\boldsymbol{\epsilon}_{t-1})\times\pi(\boldsymbol{\epsilon}_{t+1}|\boldsymbol{\epsilon}_t)\\
&\propto \prod_{s=1}^{S}\left[\frac{\alpha}{\mu_t(s)}\right]^\alpha \frac{y_t(s)^{\alpha-1}}{\Gamma(\alpha)}\exp\left[-\frac{\alpha y_t(s)}{\mu_t(s)}\right]\\
&\quad\times\sqrt{\frac{\tau}{(2\pi)^S|\Sigma|}}\exp\left\{-\frac{\tau}{2}(\boldsymbol{\epsilon}_t-\phi_\epsilon\boldsymbol{\epsilon}_{t-1})'\Sigma^{-1}(\boldsymbol{\epsilon}_t-\phi_\epsilon\boldsymbol{\epsilon}_{t-1})\right\}\\
&\quad\times\sqrt{\frac{\tau}{(2\pi)^S|\Sigma|}}\exp\left\{-\frac{\tau}{2}(\boldsymbol{\epsilon}_{t+1}-\phi_\epsilon\boldsymbol{\epsilon}_t)'\Sigma^{-1}(\boldsymbol{\epsilon}_{t+1}-\phi_\epsilon\boldsymbol{\epsilon}_t)\right\}\\
&\propto \prod_{s=1}^{S}[E_t(s)]^\alpha\exp\left\{-\left[\alpha\sum_{s=1}^{S}E_t(s)+\frac{\tau}{2}\left[Q_t+Q_{t+1}\right]\right]\right\}
\end{aligned}$$

where

$$Q_t = (\boldsymbol{\epsilon}_t-\phi_\epsilon\boldsymbol{\epsilon}_{t-1})'\,\Sigma^{-1}(\boldsymbol{\epsilon}_t-\phi_\epsilon\boldsymbol{\epsilon}_{t-1})$$

for $1<t<T-1$. For $t=1$, the FCD is given by:

$$\begin{aligned}
\pi(\boldsymbol{\epsilon}_1|\phi_\epsilon,\phi,\tau,\boldsymbol{y}_1) &\propto \prod_{s=1}^{S} f(y_1(s)|\boldsymbol{\theta}_{amt},\boldsymbol{\epsilon}_1)\times\pi(\boldsymbol{\epsilon}_1)\times\pi(\boldsymbol{\epsilon}_2|\boldsymbol{\epsilon}_1)\\
&\propto \prod_{s=1}^{S}\left[\frac{\alpha}{\mu_1(s)}\right]^\alpha \frac{y_1(s)^{\alpha-1}}{\Gamma(\alpha)}\exp\left[-\frac{\alpha y_1(s)}{\mu_1(s)}\right]\\
&\quad\times\sqrt{\frac{\tau(1-\phi_\epsilon^2)}{(2\pi)^S|\Sigma|}}\exp\left\{-\frac{\tau(1-\phi_\epsilon^2)}{2}(\boldsymbol{\epsilon}_1)'\Sigma^{-1}(\boldsymbol{\epsilon}_1)\right\}\\
&\quad\times\sqrt{\frac{\tau}{(2\pi)^S|\Sigma|}}\exp\left\{-\frac{\tau}{2}(\boldsymbol{\epsilon}_2-\phi_\epsilon\boldsymbol{\epsilon}_1)'\Sigma^{-1}(\boldsymbol{\epsilon}_2-\phi_\epsilon\boldsymbol{\epsilon}_1)\right\}\\
&\propto \prod_{s=1}^{S}[E_1(s)]^\alpha\exp\left\{-\left[\alpha\sum_{s=1}^{S}E_1(s)+\frac{\tau}{2}\Big[\left(1-\phi_\epsilon^2\right)(\boldsymbol{\epsilon}_1)'\Sigma^{-1}(\boldsymbol{\epsilon}_1)\right.\right.\\
&\quad\left.\left.+(\boldsymbol{\epsilon}_2-\phi_\epsilon\boldsymbol{\epsilon}_1)'\Sigma^{-1}(\boldsymbol{\epsilon}_2-\phi_\epsilon\boldsymbol{\epsilon}_1)\Big]\right]\right\}
\end{aligned}$$

and the FCD for $\boldsymbol{\epsilon}_T$ is

$$\begin{aligned}\pi\left(\boldsymbol{\epsilon}_T|\phi_\epsilon,\phi,\tau,\boldsymbol{y}_T\right) &\propto \prod_{s=1}^{S} f\left(y_T(s)|\boldsymbol{\theta}_{amt},\boldsymbol{\epsilon}_T\right)\times\pi\left(\boldsymbol{\epsilon}_T|\boldsymbol{\epsilon}_{T-1}\right)\\ &\propto \prod_{s=1}^{S}\left[\frac{\alpha}{\mu_T(s)}\right]^{\alpha}\frac{y_T(s)^{\alpha-1}}{\Gamma\left(\alpha\right)}\exp\left[-\frac{\alpha y_T(s)}{\mu_T(s)}\right]\\ &\quad\times\sqrt{\frac{\tau}{(2\pi)^S|\Sigma|}}\exp\left\{-\frac{\tau}{2}\left(\boldsymbol{\epsilon}_T-\phi_\epsilon\boldsymbol{\epsilon}_{T-1}\right)'\Sigma^{-1}\left(\boldsymbol{\epsilon}_T-\phi_\epsilon\boldsymbol{\epsilon}_{T-1}\right)\right\}\\ &\propto \prod_{s=1}^{S}\left[E_T(s)\right]^{\alpha}\exp\left\{-\left[\alpha\sum_{s=1}^{S}E_T(s)+\frac{\tau}{2}Q_T\right]\right\}\end{aligned}$$

where

$$Q_T=\left(\boldsymbol{\epsilon}_T-\phi_\epsilon\boldsymbol{\epsilon}_{T-1}\right)'\Sigma^{-1}\left(\boldsymbol{\epsilon}_T-\phi_\epsilon\boldsymbol{\epsilon}_{T-1}\right).$$

- Latent y-values: $\boldsymbol{y}_{dry}$

  When $W_t(s)=0$, the latent value, $y_t(s)$, is not observed and we can hence regard it as y_{dry}. In this case, the values of y_{dry} are treated as auxiliary data within the data augmentation framework. Therefore, the FCD for y_{dry} can be written as:

$$\begin{aligned}\pi(y_{dry}|r_t(s)=0,\alpha,\mu_t(s)) &\propto f(y_{dry}|\alpha,\mu_t(s))\times\Pr(R_t(s)=0|y_{dry},\boldsymbol{\theta}_{occ})\\ &\propto\frac{(\alpha/\mu_t(s))^{\alpha}y_{dry}^{\alpha-1}e^{-(\alpha/\mu_t(s))y_{dry}}}{\Gamma(\alpha)}\times p_0(s,t)\end{aligned}$$

  for every $r_t(s)=0$ where $p_0(t,s)=1-p_1(t,s)$ is the probability of a dry day which depends on $y_t(s)$.

The FCDs of the majority of parameters in the amount process are not in standard forms. Therefore, we have to utilise a Metropolis-within-Gibbs scheme to obtain the posterior samples for each parameter.

5.6.2 Full conditional distributions for the occurrence process

Suppose that we want to obtain posterior samples for each parameter in the occurrence process. Therefore, we need to derive the FCD for each parameter from equation (5.10). Similar to the amount process, the FCDs are provided as general information and for future work. The FCD for each parameter is given as follows:

- Parameter $\zeta_0^k(s)$

We have $\zeta_0^k(s) = m_{\bar{\zeta}0} + \zeta_{01}(k) + \zeta_{02}(s) + \zeta_{03}(k,s)$. The FCD of $\zeta_0^k(s)$ is represented by:

$$\begin{aligned}
&\pi\left(\zeta_0^k(s)\middle|\ddot{\boldsymbol{\zeta}}, \boldsymbol{r}, \zeta_{01}(k), \zeta_{02}(s), \boldsymbol{y}\right) \\
&\qquad \propto \prod_{t\in T_k} f\left(r_t(s)|y_t(s), \boldsymbol{\theta}_{occ}\right) \times \pi\left(\zeta_0^k(s)|\zeta_{01}(k), \zeta_{02}(s)\right) \\
&\qquad \propto \prod_{t\in T_k} \left[p_1(s,t)\right]^{r_t(s)} \left[1-p_1(s,t)\right]^{1-r_t(s)} \\
&\qquad\quad \times \sqrt{\frac{\tau_{\bar{\zeta}_0}}{2\pi}} \exp\left\{-\frac{\tau_{\bar{\zeta}_0}}{2}\left(\zeta_0^k(s) - \zeta_{01}(k) - \zeta_{02}(s) - m_{\bar{\zeta}_0}\right)^2\right\} \\
&\qquad \propto \prod_{t\in T_k} \left[p_1(s,t)\right]^{r_t(s)} \left[1-p_1(s,t)\right]^{1-r_t(s)} \\
&\qquad\quad \exp\left\{-\frac{\tau_{\bar{\zeta}_0}}{2}\Big(\left\{\zeta_0^k(s)\right\}^2 - 2\zeta_0^k(s)\left\{\zeta_{01}(k) + \zeta_{02}(s) + m_{\bar{\zeta}_0}\right\}\Big)\right\}
\end{aligned}$$

where $T_k = \{t : x_t = k\}$ is the set of observations that belongs to type $k = 1, \cdots, 27$.

The FCD of $\zeta_{01}(k)$ is given by:

$$\begin{aligned}
&\pi\left(\zeta_{01}(k)\middle|\zeta_0^k(1), \cdots, \zeta_0^k(S), \zeta_{02}(1), \cdots, \zeta_{02}(S), \check{\zeta}_{01}\right) \\
&\qquad \propto \prod_{s=1}^{S} f\left(\zeta_0^k(s)|\zeta_{01}(k), \zeta_{02}(s)\right) \times \pi\left(\zeta_{01}(k)|\check{\zeta}_{01}\right) \\
&\qquad \propto \prod_{s=1}^{S} \sqrt{\frac{\tau_{\bar{\zeta}_0}}{2\pi}} \exp\left\{-\frac{\tau_{\bar{\zeta}_0}}{2}\left(\zeta_0^k(s) - \zeta_{01}(k) - \zeta_{02}(s) - m_{\bar{\zeta}_0}\right)^2\right\} \\
&\qquad\quad \times \sqrt{\frac{1}{2\pi\upsilon_{\zeta_1}}} \exp\left\{-\frac{1}{2\upsilon_{\zeta_1}}\left(\zeta_{01}(k) - \check{\zeta}_{01}\right)^2\right\} \\
&\qquad \propto \exp\left\{-\frac{D_{\zeta_{01}}}{2}\left(\zeta_{01}(k) - B_{\zeta_{01}}\right)^2\right\}
\end{aligned}$$

where

$$D_{\zeta_{01}} = S\tau_{\bar{\zeta}_0} + \frac{1}{\upsilon_{\zeta_1}}$$

and

$$B_{\zeta_{01}} = \frac{1}{D_{\zeta_{01}}}\left\{\tau_{\bar{\zeta}_0}\left[\sum_{s=1}^{S}\zeta_0^k(s) - \sum_{s=1}^{S}\zeta_{02}(s) - Sm_{\bar{\zeta}_0}\right] + \frac{\check{\zeta}_{01}}{\upsilon_{\zeta_1}}\right\}.$$

This form corresponds to a normal density, $N(B_{\zeta_{01}}, 1/D_{\zeta_{01}})$.

The FCD of $\check{\zeta}_{01}$ is derived as follows:

$$\begin{aligned}\pi\left(\check{\zeta}_{01}\middle|\zeta_{01}(1),\cdots,\zeta_{01}(27)\right) &\propto \prod_{k=1}^{27} f\left(\zeta_{01}(k)|\check{\zeta}_{01}\right)\times\pi(\check{\zeta}_{01})\\ &\propto \prod_{k=1}^{27}\sqrt{\frac{1}{2\pi\upsilon_{\zeta_1}}}\exp\left\{-\frac{1}{2\upsilon_{\zeta_1}}\left(\zeta_{01}(k)-\check{\zeta}_{01}\right)^2\right\}\\ &\quad\times\sqrt{\frac{1}{2\pi\upsilon_{\zeta_0}}}\exp\left\{-\frac{1}{2\upsilon_{\zeta_0}}\check{\zeta}_{01}\right\}\\ &\propto \exp\left\{\frac{\tilde{D}_{\check{\zeta}_{01}}}{2}\left(\check{\zeta}_{01}-\tilde{B}_{\check{\zeta}_{01}}\right)^2\right\}\end{aligned}$$

where

$$\tilde{D}_{\check{\zeta}_{01}}=\frac{K}{\upsilon_{\zeta_1}}+\frac{1}{\upsilon_{\zeta_0}}$$

and

$$\tilde{B}_{\check{\zeta}_{01}}=\frac{\sum_{k=1}^{27}\zeta_{01}(k)}{\tilde{D}_{\check{\zeta}_{01}}\upsilon_{\zeta_0}}.$$

This form also corresponds to a normal density, $N(\tilde{B}_{\check{\zeta}_{01}},1/\tilde{D}_{\check{\zeta}_{01}})$.

The FCD of $\zeta_{02}(s)$ is represented by:

$$\begin{aligned}&\pi\left(\zeta_{02}(s)\middle|\zeta_0^1(s),\cdots,\zeta_0^{27}(s),\zeta_{01}(1),\cdots,\zeta_{01}(27),\check{\zeta}_{02}\right)\\ &\qquad\propto\prod_{k=1}^{27} f\left(\zeta_0^k(s)|\zeta_{01}(k),\zeta_{02}(s)\right)\times\pi\left(\zeta_{02}(s)|\check{\zeta}_{02}\right)\\ &\qquad\propto\prod_{k=1}^{27}\sqrt{\frac{\tau_{\bar{\zeta}_0}}{2\pi}}\exp\left\{-\frac{\tau_{\bar{\zeta}_0}}{2}\left(\zeta_0^k(s)-\zeta_{01}(k)-\zeta_{02}(s)-m_{\bar{\zeta}_0}\right)^2\right\}\\ &\qquad\quad\times\sqrt{\frac{1}{2\pi\nu_{\zeta_1}}}\exp\left\{-\frac{1}{2\nu_{\zeta_1}}\left(\zeta_{02}(s)-\check{\zeta}_{02}\right)^2\right\}\\ &\qquad\propto\exp\left\{-\frac{D_{\zeta_{02}}}{2}\left(\zeta_{02}(s)-B_{\zeta_{02}}\right)^2\right\}\end{aligned}$$

where

$$D_{\zeta_{02}}=S\tau_{\bar{\zeta}_0}+\frac{1}{\nu_{\zeta_1}}$$

and

$$B_{\zeta_{02}}=\frac{1}{D_{\zeta_{02}}}\left\{\tau_{\bar{\zeta}_0}\left[\sum_{k=1}^{27}\zeta_0^k(s)-\sum_{k=1}^{27}\zeta_{01}(k)-Km_{\bar{\zeta}_0}\right]+\frac{\check{\zeta}_{02}}{\nu_{\zeta_1}}\right\}.$$

This is a normal distribution.

For $\check{\zeta}_{02}$, the FCD is:

$$\begin{aligned}\pi\left(\check{\zeta}_{02}\middle|\zeta_{02}(1),\cdots,\zeta_{02}(S)\right) &\propto \prod_{s=1}^{S} f\left(\zeta_{02}(s)|\check{\zeta}_{02}\right)\times\pi(\check{\zeta}_{02})\\ &\propto \prod_{s=1}^{S}\sqrt{\frac{1}{2\pi\nu_{\zeta_1}}}\exp\left\{-\frac{1}{2\nu_{\zeta_1}}\left(\zeta_{02}(s)-\check{\zeta}_{02}\right)^2\right\}\\ &\quad\times\sqrt{\frac{1}{2\pi\nu_{\zeta_0}}}\exp\left\{-\frac{1}{2\nu_{\zeta_0}}(\check{\zeta}_{02})^2\right\}\\ &\propto \exp\left\{\frac{\tilde{D}_{\check{\zeta}_{02}}}{2}\left(\check{\zeta}_{01}-\tilde{B}_{\check{\zeta}_{02}}\right)^2\right\}\end{aligned}$$

where

$$\tilde{D}_{\check{\zeta}_{02}}=\frac{S}{\nu_{\zeta_1}}+\frac{1}{\nu_{\zeta_0}}$$

and

$$\tilde{B}_{\check{\zeta}_{02}}=\frac{\sum_{s=1}^{S}\zeta_{02}(s)}{\tilde{D}_{\check{\zeta}}\nu_{\zeta_0}}.$$

Again, this is the density form of a normal distribution.

- Parameter $\ddot{\boldsymbol{\zeta}}$

The FCD for $\ddot{\boldsymbol{\zeta}}$ can be derived as follows:

$$\begin{aligned}\pi\left(\ddot{\boldsymbol{\zeta}}|\zeta_0^k(s),\zeta_1,\boldsymbol{r},\boldsymbol{y}\right) &\propto \prod_{t=1}^{T}\prod_{s=1}^{S} f\left(r_t(s)|y_t(s),\boldsymbol{\theta}_{occ}\right)\times\pi\left(\ddot{\boldsymbol{\zeta}}\right)\\ &\propto \prod_{t=1}^{T}\prod_{s=1}^{S}\left[p_1(s,t)\right]^{r_t(s)}\left[1-p_1(s,t)\right]^{1-r_t(s)}\\ &\quad\times(2\pi)^{-4}|P_{\ddot{\boldsymbol{\zeta}}}|^{1/2}\exp\left\{-\frac{1}{2}(\ddot{\boldsymbol{\zeta}}-\boldsymbol{m}_{\ddot{\boldsymbol{\zeta}}})'P_{\ddot{\boldsymbol{\zeta}}}(\ddot{\boldsymbol{\zeta}}-\boldsymbol{m}_{\ddot{\boldsymbol{\zeta}}})\right\}\\ &\propto \prod_{t=1}^{T}\prod_{s=1}^{S}\left[p_1(s,t)\right]^{r_t(s)}\left[1-p_1(s,t)\right]^{1-r_t(s)}\\ &\quad\exp\left\{-\frac{1}{2}(\ddot{\boldsymbol{\zeta}}-\boldsymbol{m}_{\ddot{\boldsymbol{\zeta}}})'P_{\ddot{\boldsymbol{\zeta}}}(\ddot{\boldsymbol{\zeta}}-\boldsymbol{m}_{\ddot{\boldsymbol{\zeta}}})\right\}.\end{aligned}$$

- Parameter $\zeta_1(s)$

  For ζ_1, the FCD is given by:

$$\begin{aligned}
\pi\left(\zeta_1|\zeta_0^k(s),\ddot{\boldsymbol{\zeta}},\boldsymbol{r},\boldsymbol{y}\right) &\propto \prod_{t=1}^{T}\prod_{s=1}^{S} f\left(r_t(s)|y_t(s),\boldsymbol{\theta}_{occ}\right)\times\pi\left(\zeta_1\right)\\
&\propto \prod_{t=1}^{T}\prod_{s=1}^{S}\left[p_1(s,t)\right]^{r_t(s)}\left[1-p_1(s,t)\right]^{1-r_t(s)}\\
&\times\sqrt{\frac{\tau_{\zeta_1}}{2\pi}}\exp\left\{-\frac{\tau_{\zeta_1}}{2}\left(\zeta_1-m_{\zeta_1}\right)^2\right\}\\
&\propto \prod_{t=1}^{T}\prod_{s=1}^{S}\left[p_1(s,t)\right]^{r_t(s)}\left[1-p_1(s,t)\right]^{1-r_t(s)}\\
&\exp\left\{-\frac{\tau_{\zeta_1}}{2}\left(\zeta_1^2-2\zeta_1 m_{\zeta_1}\right)\right\}.
\end{aligned}$$

Similar to the amount process, the FCDs of the majority of parameters for the occurrence process are not of any standard form. Therefore, we require a Metropolis-within-Gibbs scheme to sample the posterior values for each parameter.

5.7 Application

5.7.1 Prior distributions

Similar to the previous models, the prior distributions for the spatiotemporal model were elicited using the information gained from previous studies and personal beliefs. The prior construction for a single site in the UK has already been discussed in Chapter 4, Section 4.4.5.1. Therefore, we will use the same argument for the spatiotemporal model to specify the priors since the sites are within the same network.

For the additional parameter ϕ, which governs the spatial correlation, we follow the idea of Banerjee *et al.* (2004) by considering the "effective range" $\hat{D}$ as follows:

$$\begin{aligned}
& e^{-\phi\hat{D}} = 0.05\\
\text{so}\quad & -\phi\hat{D} = \log(0.05)\\
\text{so}\quad & \phi = -\frac{\log(0.05)}{\hat{D}}
\end{aligned}$$

where $\hat{D}$ is the interlocation distance at which the between-site correlation has fallen to 0.05. We can use a direct approach to elicit prior quartiles for $\hat{D}$ and then convert these to prior quartiles for ϕ. Let our prior quartiles for ϕ be $Q_{\phi,1}$, $Q_{\phi,2}$ and $Q_{\phi,3}$, respectively. We

give ϕ a $\text{Ga}(g_\phi, h_\phi)$ distribution. Since h_ϕ is a scale parameter, $Q_{\phi,3}/Q_{\phi,1}$ depends only on g_ϕ where g_ϕ can be determined using numerical iteration to solve $Q_{\phi,3}/Q_{\phi,1} = R_3(a)/R_1(a)$ where $R_q(a)$ is quartile q for a $\text{Ga}(a, 1)$ distribution. Once g_ϕ is obtained, then we can find the value for h_ϕ using $R_2(g_\phi)/Q_{\phi,2}$. Suppose that the prior median for the effective range is 750 km, then the prior median of ϕ is $Q_{\phi,2} = 0.004$. We have also chosen carefully a suitable value of the lower and upper quartiles for illustration where we have $Q_{\phi,1} = 0.001$ and $Q_{\phi,3} = 0.0045$, respectively. Using this method, we have $g_\phi = 1.07$ and $h_\phi = 190.59$.

In this spatiotemporal model, we employed a Fourier series with just two harmonics for the amount process. The prior elicitation for the Fourier coefficients is still the same as in Section 4.3.3.3. The full prior specifications for the unknown parameters of the amount process can be summarised as follows:

$$\begin{aligned} \alpha &\sim \text{Ga}(4.31, 5.78); \qquad & \upsilon &\sim \text{Beta}(3.7, 1.9); \\ \tau &\sim \text{Ga}(3.69, 1.69); \qquad & \phi &\sim \text{N}(1.07, 190.59); \end{aligned}$$

and

$$\ddot{\boldsymbol{\eta}} \sim \text{N}_4\left(\underset{\sim}{\ddot{\eta}}_0, P_{\ddot{\boldsymbol{\eta}}}^{-1}\right)$$

where

$$\underset{\sim}{\ddot{\eta}}_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, P_{\ddot{\boldsymbol{\eta}}}^{-1} = \begin{pmatrix} \frac{2}{\pi} & 0 & 0 & 0 \\ 0 & \frac{1}{\pi} & 0 & 0 \\ 0 & 0 & \frac{2}{\pi} & 0 \\ 0 & 0 & 0 & \frac{1}{\pi} \end{pmatrix}$$

for $k \in \{1, \cdots, 27\}$ and $s \in \{1, \cdots, 5\}$.

Based on the previous univariate British model, we have chosen the prior mean of η_0 to be 0.69. Then, we will use the same argument to give a value for $m_{\bar{\eta}0}$. In the case of $\eta_{01}(k)$, $\eta_{02}(s)$ and $\eta_{03}(k, s)$, we have chosen a suitable values for the variances of the uncertainty factors as follows:

$$\begin{aligned} \eta_{01}(k)|\check{\eta}_{01} &\sim N(\check{\eta}_{01}, 1); \qquad & \check{\eta}_{01} &\sim N(0, 1) \\ \eta_{02}(s)|\check{\eta}_{02} &\sim N(\check{\eta}_{02}, 1); \qquad & \check{\eta}_{02} &\sim N(0, 1) \\ \eta_{03}(k, s) &\sim N(0, 4). \end{aligned}$$

This gives $\eta_0^k(s)$ a marginal prior variance of 8. Let $k' \neq k$ and $s' \neq s$. The covariance between $\eta_0^k(s)$ and $\eta_0^{k'}(s)$ is 3, giving a correlation 0f 0.375. Similarly the covariance between $\eta_0^k(s)$ and $\eta_0^k(s')$ is 3, giving a correlation 0f 0.375. The covariance between $\eta_0^k(s)$ and $\eta_0^{k'}(s')$ is 2, giving a correlation of 0.250. Thus the prior correlations are fairly weak.

For the occurrence process, we also use the same argument as for the British univariate

model and the amount process to elicit the prior for the unknown parameters. We will use two harmonics of the Fourier series to represent the seasonal effect of the rainfall occurrence over the year. Hence, the prior specifications for the unknown parameters for the occurrence process are given by the following:

$$\zeta_1 \sim N(1.29, 3.12)$$

and

$$\ddot{\boldsymbol{\zeta}} \sim N_4\left(\ddot{\underline{\zeta}}_0, P_{\ddot{\boldsymbol{\zeta}}}^{-1}\right)$$

where

$$\ddot{\underline{\zeta}}_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, P_{\ddot{\boldsymbol{\zeta}}}^{-1} = \begin{pmatrix} \frac{2}{\pi} & 0 & 0 & 0 \\ 0 & \frac{1}{\pi} & 0 & 0 \\ 0 & 0 & \frac{2}{\pi} & 0 \\ 0 & 0 & 0 & \frac{1}{\pi} \end{pmatrix}$$

for $k \in \{1, \cdots, 27\}$ and $s \in \{1, \cdots, 5\}$.

The values for $m_{\bar{\zeta}0}$ and the variances of $\zeta_{01}(k)$, $\zeta_{02}(s)$ and $\zeta_{03}(k, s)$ are chosen as follows:

$$\begin{aligned} \zeta_{01}(k)|\check{\zeta}_{01} &\sim N(\check{\zeta}_{01}, 1); & \check{\zeta}_{01} &\sim N(0, 1) \\ \zeta_{02}(s)|\check{\zeta}_{02} &\sim N(\check{\zeta}_{02}, 1); & \check{\zeta}_{02} &\sim N(0, 1) \\ \zeta_{03}(k, s) &\sim N(0, 5.52); & m_{\bar{\zeta}0} &= 1.03. \end{aligned}$$

This gives $\zeta_0^k(s)$ a marginal prior variance of 9.52. Let $k' \neq k$ and $s' \neq s$. The covariance between $\zeta_0^k(s)$ and $\zeta_0^{k'}(s)$ is 3, giving a correlation 0f 0.315. Similarly the covariance between $\zeta_0^k(s)$ and $\zeta_0^k(s')$ is 3, giving a correlation 0f 0.315. The covariance between $\zeta_0^k(s)$ and $\zeta_0^{k'}(s')$ is 2, giving a correlation of 0.210. Thus the prior correlations are again fairly weak.

5.7.2 Fitting the model

The Rjags package (Plummer, 2012) is used to compute the posterior samples for this model. The MCMC algorithm was run for 50000 iterations to allow for burn-in, and an additional 150000 iterations were drawn as the posterior samples. The computing time that was required to generate 150000 posterior draws was around 240 hours by using R software on a 2.00GHz Samsung laptop 300V3A model with Intel Core i7-2630QM processor and 12 Gbytes of random-access memory. The trace plots can be used as a visual diagnostic method to monitor the convergence of the sampler. Before we make any inferences or conclusions, it is essential to look at the trace plots for parameters.

Figure 5.6 displays the example of trace plots and posterior densities for α, τ and ζ_1. Based on the trace plots, the mixing appears very satisfactory for the majority of the unknown parameters. However, some of the unknown parameters have a relatively high autocorrelation, which suggests mixing is a little slower. This can be seen in Figure 5.6 for the parameters α and τ which suggest that the effective sample size (ESS) of the parameters is likely to be extremely low. The ESS is the number of effectively independent samples obtained from the posterior distribution. The issue of low posterior sample size for α and τ happens because both α and τ contribute to the variance of the rainfall amount and therefore they are, to some extent, confounded. As a result, they are strongly negatively correlated in the posterior. This might also suggest that a large number of iterations are required to obtain a good representation of the posterior distribution.

The uncertainty associated with most of the unknown parameters has been reduced after we combined the prior beliefs with the data. For example, there is a reduction in the variability of parameter ζ_1 when a comparison is made between the prior and posterior distributions. The posterior mean of parameter ζ_1 is shifted to the right of its prior mean. The full summaries of posterior means and standard deviations of the unknown parameters for the amount and occurrence processes at Balmoral weather station are given in Tables 5.2 and 5.3 (see Appendices A.1 and A.2 for the other weather stations).

In this spatiotemporal model, it is also important to observe the parameter ϕ since it controls the rate of decay of the correlation as the distance between sites increases. From Table 5.2, we can observe that the posterior value for ϕ is centred around 0.0012. This suggests that the correlation falls to 0.5 after about 578 km. This implies high spatial correlation even at large distances. Figure 5.4 suggests that the correlation does not decrease quickly with increasing distance. It may also be that the correlation does not tend to 1 as the distance tends to zero because of a "nugget effect". For this reason, it might be better to use a different spatial correlation function. A lot of the correlation may be explained by the LWT, leaving little for the spatial effect to explain. The LWT plays an important role in determining the rainfall amount and probability at time t. This is supported by looking at Figures 5.7 and 5.8. These plots show the posterior means and 95% credible intervals of the intercepts parameters $\eta_0^k(s)$ and $\zeta_0^k(s)$ for all weather stations. It can be seen that the value for the LWT of type 1 (Anticylconic) is quite small for both processes despite being the most commonly-occurred LWT. The previous random effect at time $t-1$ also plays a crucial role in determining the level of rainfall at time t. It is therefore highly possible that the rainfall amount will increase if rainfall did occur on the previous day.

For model checking, we chose two weather stations that have the maximum and minimum mean of daily rainfall amounts and these are Bastreet and Ardleigh Hull Farm

Amount process				
Parameter	Prior mean	Prior SD	Posterior mean	Posterior SD
α	0.75	0.36	1.153	0.037
$\eta_0^1(1)$	0.69	2.83	-0.927	0.075
$\eta_0^2(1)$	0.69	2.83	-0.194	0.267
$\eta_0^3(1)$	0.69	2.83	-0.389	0.298
$\eta_0^4(1)$	0.69	2.83	-1.065	0.289
$\eta_0^5(1)$	0.69	2.83	-1.523	0.239
$\eta_0^6(1)$	0.69	2.83	-1.256	0.167
$\eta_0^7(1)$	0.69	2.83	-0.656	0.140
$\eta_0^8(1)$	0.69	2.83	0.183	0.161
$\eta_0^9(1)$	0.69	2.83	0.645	0.190
$\eta_0^{10}(1)$	0.69	2.83	0.266	0.190
$\eta_0^{11}(1)$	0.69	2.83	-0.048	0.203
$\eta_0^{12}(1)$	0.69	2.83	-0.647	0.168
$\eta_0^{13}(1)$	0.69	2.83	-1.184	0.122
$\eta_0^{14}(1)$	0.69	2.83	-0.468	0.088
$\eta_0^{15}(1)$	0.69	2.83	0.009	0.082
$\eta_0^{16}(1)$	0.69	2.83	0.407	0.093
$\eta_0^{17}(1)$	0.69	2.83	0.907	0.117
$\eta_0^{18}(1)$	0.69	2.83	0.575	0.065
$\eta_0^{19}(1)$	0.69	2.83	0.131	0.278
$\eta_0^{20}(1)$	0.69	2.83	0.047	0.319
$\eta_0^{21}(1)$	0.69	2.83	0.137	0.233
$\eta_0^{22}(1)$	0.69	2.83	-0.200	0.170
$\eta_0^{23}(1)$	0.69	2.83	0.136	0.144
$\eta_0^{24}(1)$	0.69	2.83	0.547	0.143
$\eta_0^{25}(1)$	0.69	2.83	0.814	0.170
$\eta_0^{26}(1)$	0.69	2.83	1.121	0.180
$\eta_0^{27}(1)$	0.69	2.83	-0.0003	0.241
a_1	0	0.8	0.420	0.035
a_2	0	0.56	0.055	0.039
b_1	0	0.8	-0.095	0.043
b_2	0	0.56	0.004	0.031
τ	2.18	1.14	0.479	0.013
v	0.66	0.18	0.750	0.007
ϕ	0.006	0.005	0.001	4.79E-05

Table 5.2: The prior and posterior means with standard deviations (SDs) of the unknown parameters of the amount process for Balmoral weather station

Occurrence process				
Parameter	Prior mean	Prior sd	Posterior mean	Posterior sd
$\zeta_0^1(1)$	1.03	3.09	1.422	0.117
$\zeta_0^2(1)$	1.03	3.09	2.707	0.492
$\zeta_0^3(1)$	1.03	3.09	1.835	0.491
$\zeta_0^4(1)$	1.03	3.09	0.584	0.468
$\zeta_0^5(1)$	1.03	3.09	1.601	0.397
$\zeta_0^6(1)$	1.03	3.09	1.940	0.275
$\zeta_0^7(1)$	1.03	3.09	2.160	0.230
$\zeta_0^8(1)$	1.03	3.09	2.224	0.295
$\zeta_0^9(1)$	1.03	3.09	2.707	0.416
$\zeta_0^{10}(1)$	1.03	3.09	1.821	0.360
$\zeta_0^{11}(1)$	1.03	3.09	1.847	0.352
$\zeta_0^{12}(1)$	1.03	3.09	1.414	0.277
$\zeta_0^{13}(1)$	1.03	3.09	1.346	0.200
$\zeta_0^{14}(1)$	1.03	3.09	2.116	0.161
$\zeta_0^{15}(1)$	1.03	3.09	2.609	0.157
$\zeta_0^{16}(1)$	1.03	3.09	2.830	0.220
$\zeta_0^{17}(1)$	1.03	3.09	2.929	0.340
$\zeta_0^{18}(1)$	1.03	3.09	2.259	0.148
$\zeta_0^{19}(1)$	1.03	3.09	2.072	0.519
$\zeta_0^{20}(1)$	1.03	3.09	0.708	0.547
$\zeta_0^{21}(1)$	1.03	3.09	2.002	0.499
$\zeta_0^{22}(1)$	1.03	3.09	1.866	0.331
$\zeta_0^{23}(1)$	1.03	3.09	2.546	0.314
$\zeta_0^{24}(1)$	1.03	3.09	2.783	0.365
$\zeta_0^{25}(1)$	1.03	3.09	1.951	0.390
$\zeta_0^{26}(1)$	1.03	3.09	2.901	0.597
$\zeta_0^{27}(1)$	1.03	3.09	0.612	0.406
c_1	0	0.8	0.529	0.031
c_2	0	0.56	-0.092	0.031
d_1	0	0.8	0.014	0.031
d_2	0	0.56	-0.050	0.031
ζ_1	1.29	1.77	1.512	0.036

Table 5.3: The prior and posterior means with standard deviations (SDs) of the unknown parameters of the occurrence process for Balmoral weather station

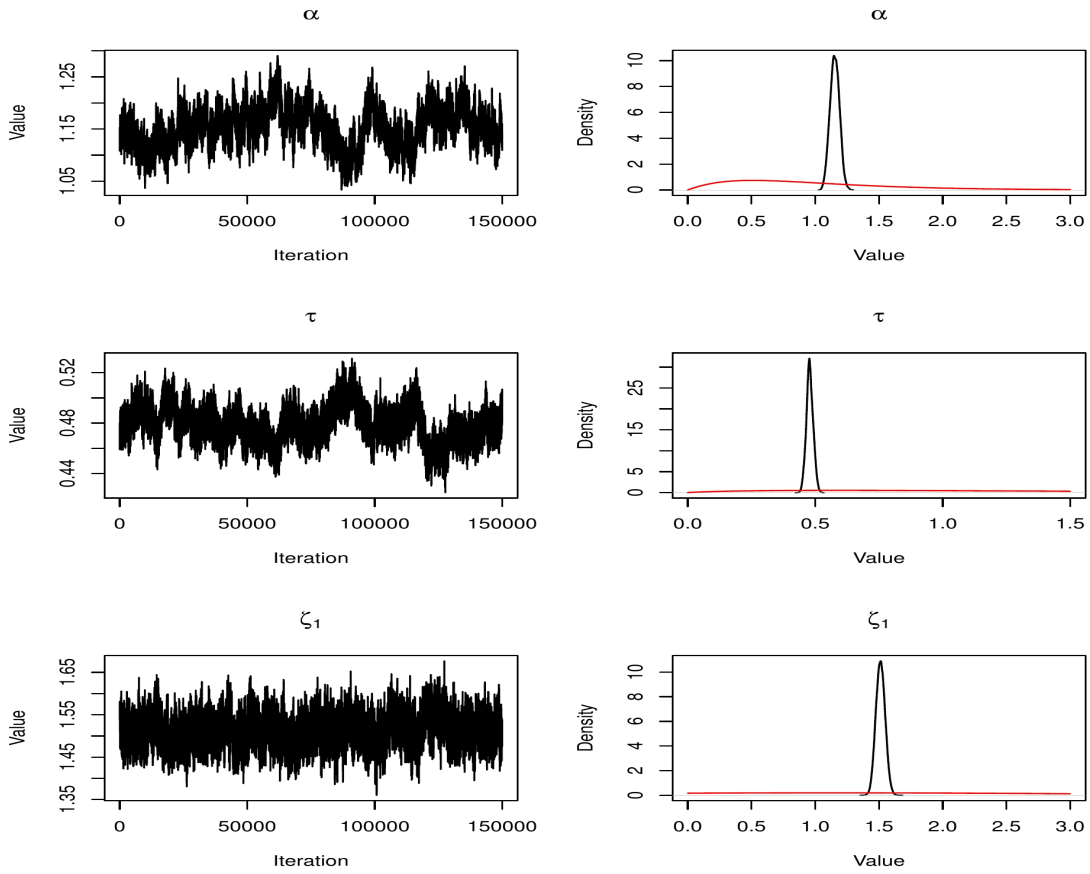


Figure 5.6: The trace and density plots for parameters α, ϕ and τ

weather stations. We used the same diagnostic checking methods as in Section 4.2.6 to assess the adequacy of our model. The graphs of these residuals for both sites are shown in Figure 5.9 and 5.11. The residuals for both sites indicate that they conform to the respective series of independent uniform and normal distributions. We also assessed the quantile-quantile (QQ), autocorrelation function (ACF), and partial autocorrelation function (PACF) plots for the NFTR and these are shown in Figures 5.10 and 5.12. The QQ plots for both sites indicate that the NFTR are normally distributed. The ACF and PACF plots also demonstrate that these residuals are independent to one another. From these plots, we can see that this spatiotemporal model fits the data very well.

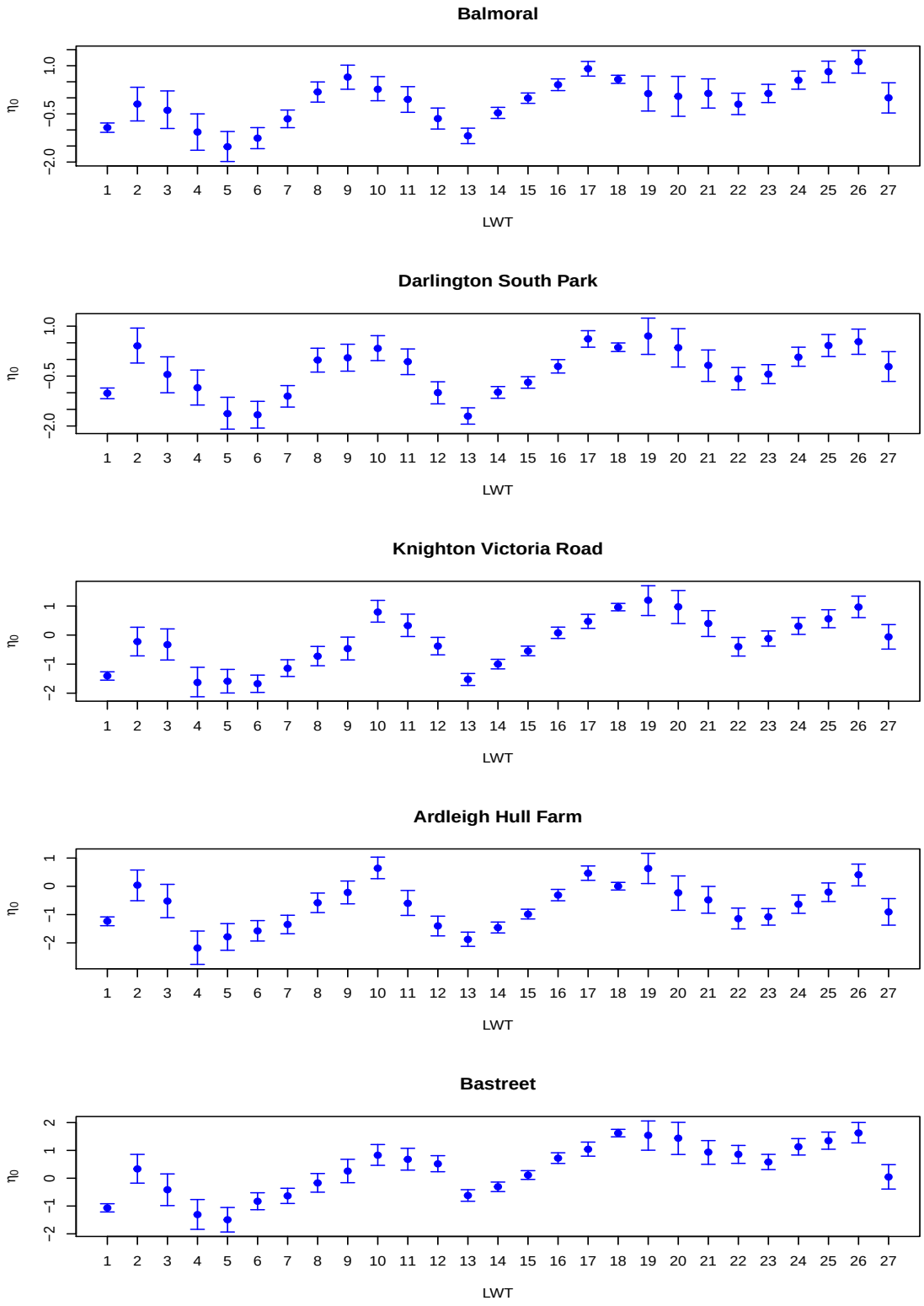


Figure 5.7: The posterior means of $\eta_0^k(s)$ with 95% credible intervals for all weather stations where $s \in \{1, \cdots, 5\}$ and $k \in \{1, \cdots, 27\}$. See Table 4.3 for details of LWTs.

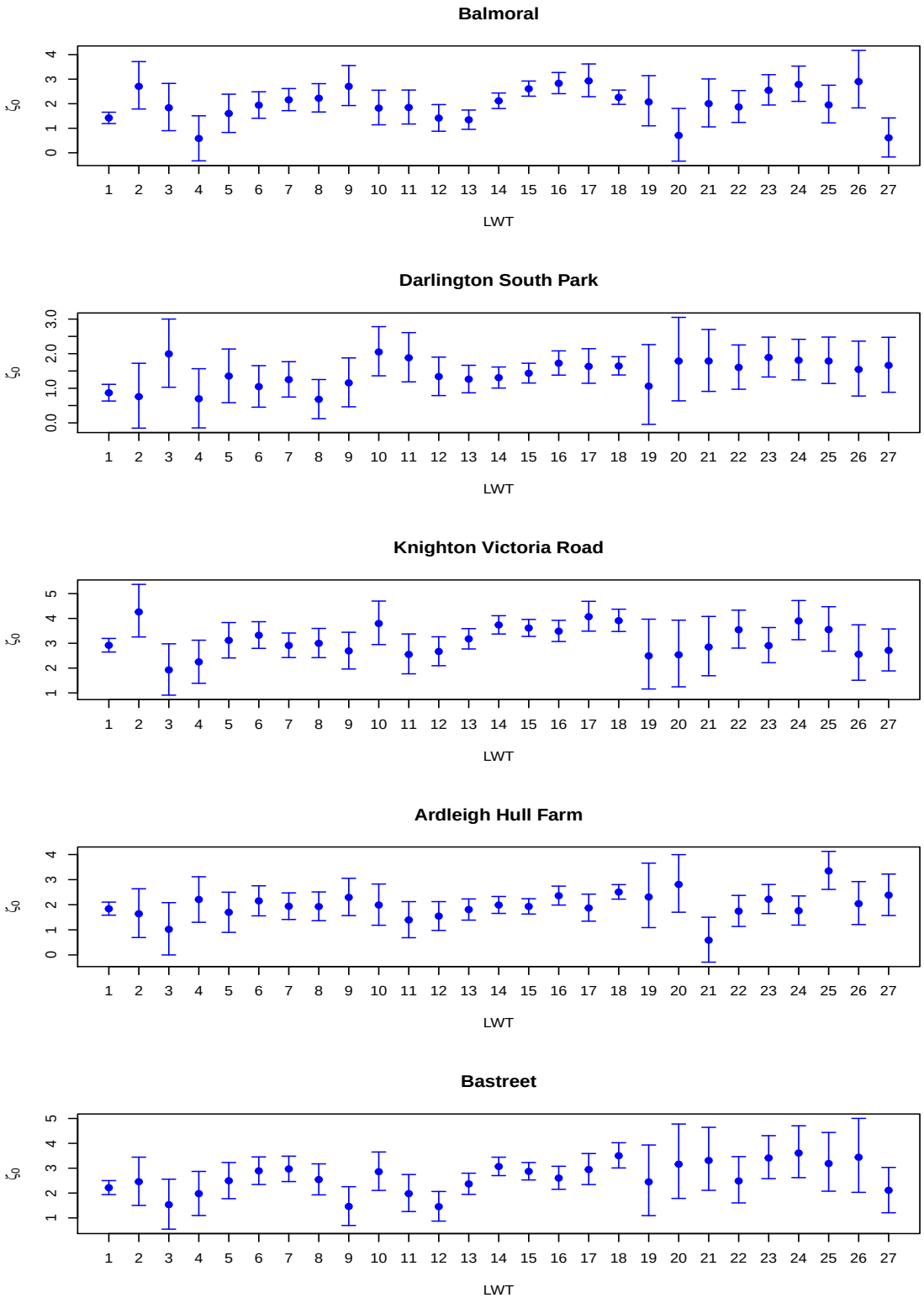


Figure 5.8: The posterior means of $\zeta_0^k(s)$ with 95% credible intervals for all weather stations where $s \in \{1, \cdots, 5\}$ and $k \in \{1, \cdots, 27\}$. See Table 4.3 for details of LWTs.

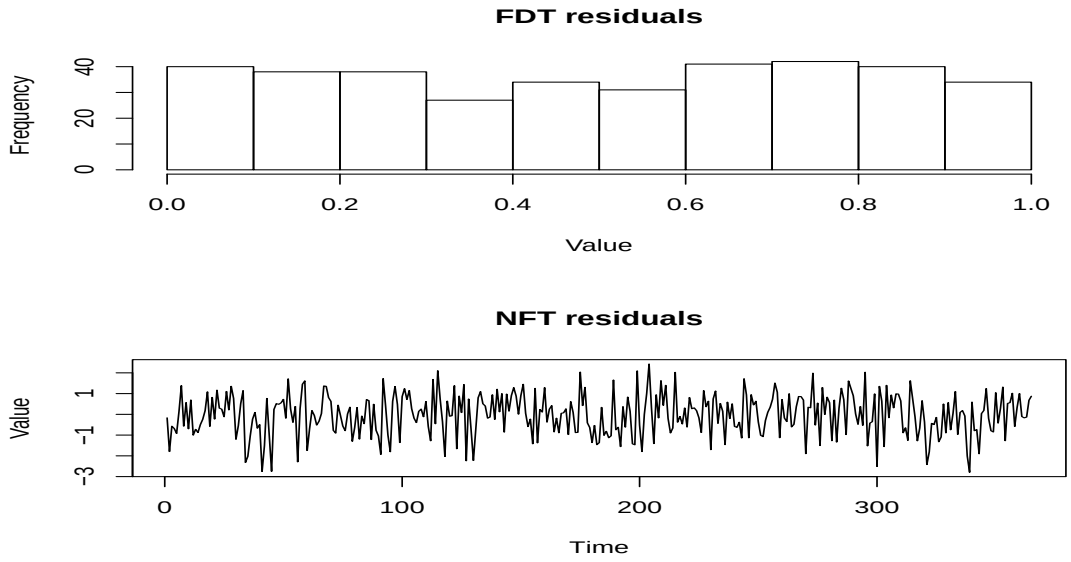


Figure 5.9: The FDTR and NFTR plots for Ardleigh Hull Farm weather station

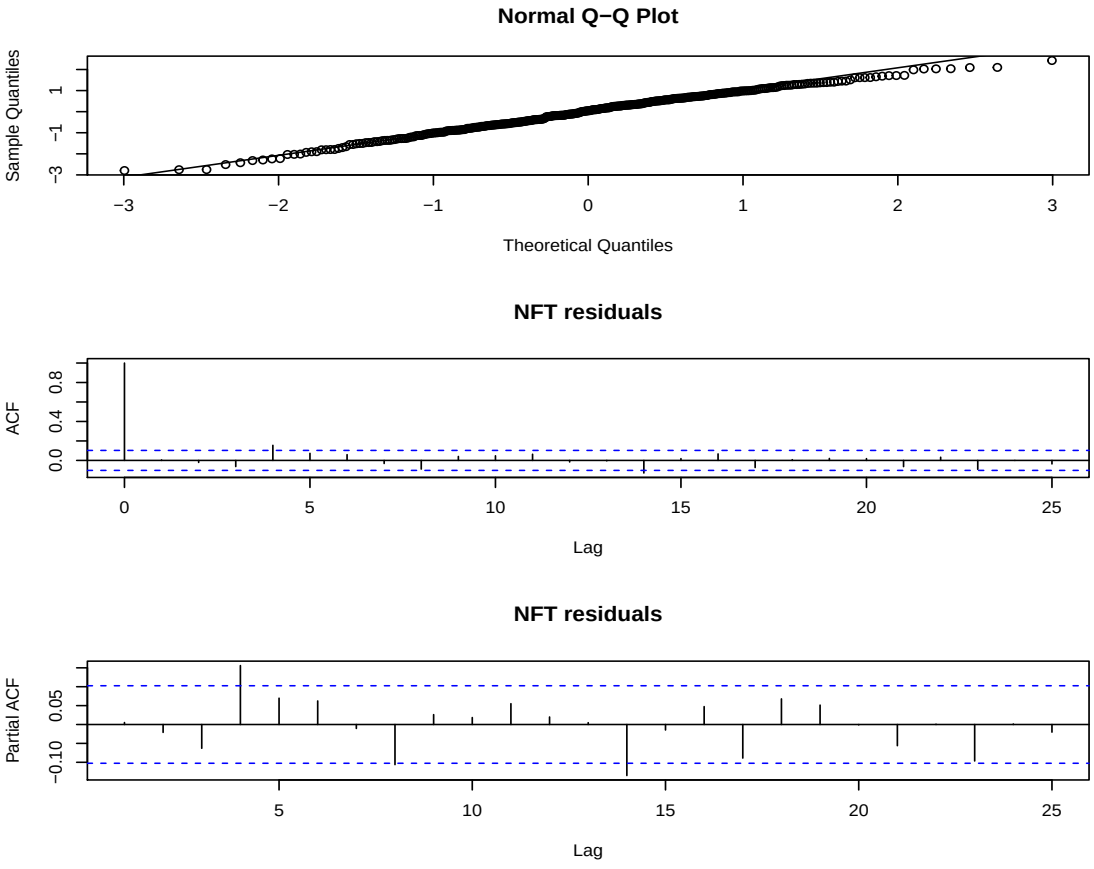


Figure 5.10: QQ, ACF and PACF plots for Ardleigh Hull Farm weather station

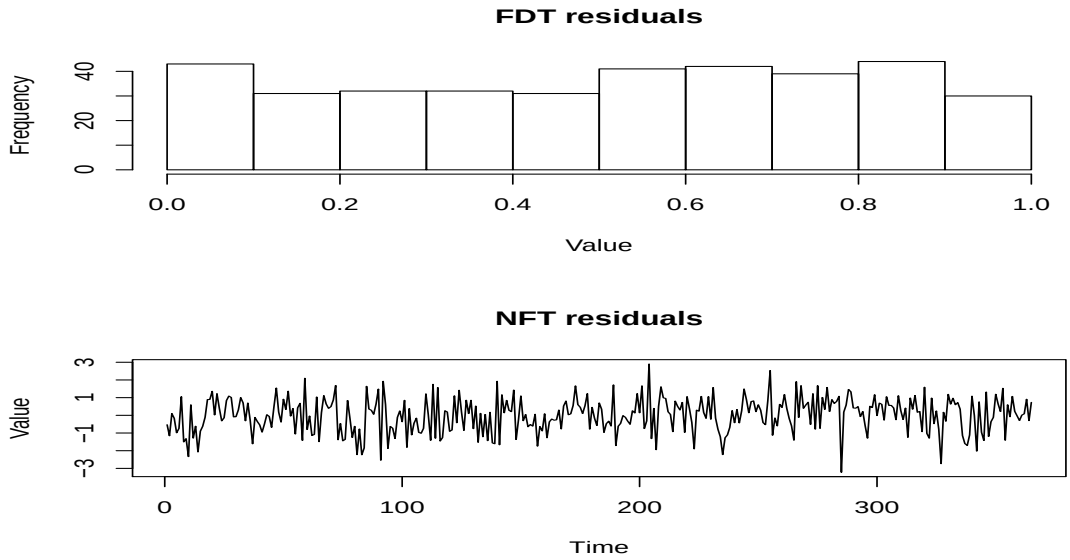


Figure 5.11: The FDTR and NFTR plots for Bastreet weather station

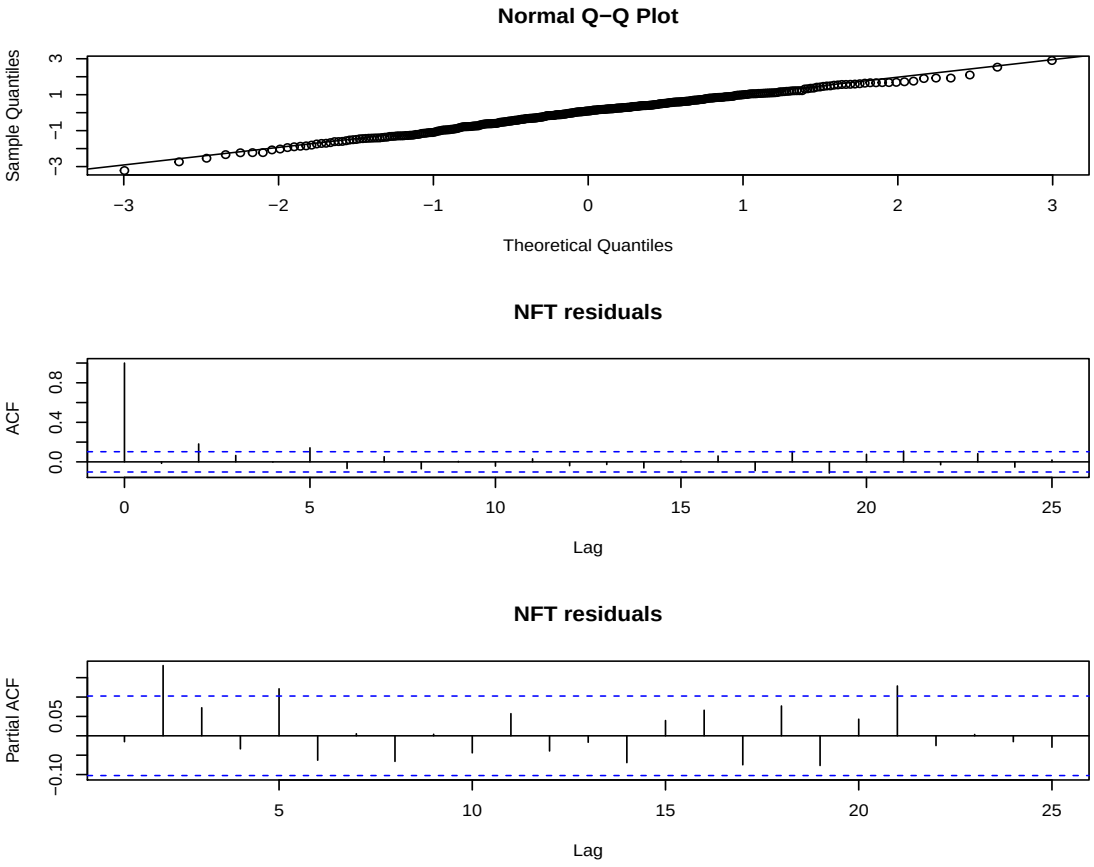


Figure 5.12: QQ, ACF and PACF plots for Bastreet weather station

5.8 Summary

The spatiotemporal model was presented in this chapter and applied to British daily rainfall data at multiple sites within the UK network. This model is the extended version of the univariate model in Chapter 4, Section 4.4 and also an alternative model to the previous works by Heaps *et al.* (2015) and Germain (2010). The objective of this model is to capture the relationship that exists between the sites and how this relationship may affect the daily rainfall for each site, especially when two sites are close to each other. In addition, the LWT is directly incorporated into the model through the amount and occurrence processes since the LWT plays a crucial role in determining rainfall amounts and occurrence over the UK network.

This model is more challenging to develop than the previous univariate model in Chapter 4, Section 4.4 because it involves many variables and parameters to fit the data. It also requires intensive computation to obtain the results. Based on the MCMC results in Section 5.7.2, we can consider that the proposed model fits the data very well. This is strongly supported by the diagnostic check for the FDTR and NFTR which show that they are independently distributed and respectively follow uniform and normal distributions. We also found high spatial correlation between sites, even at large distances in this model. From our discussion, we may consider using different types of spatial correlation function to capture the spatial variability between sites. The amount and occurrence processes indicate that they are still highly dependent on the LWT ocurring at time t.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

This thesis is concerned with the development of Bayesian approaches to the modelling and analysis of univariate and multivariate time series data, especially when the observational distribution is a mixture distribution. Particular attention is given to the special case of a mixture distribution which has a degenerate component at zero and another component with positive values and this is known as a mixed distribution. The main focus of this research is to investigate and develop both univariate and multivariate models in the case of a mixed distribution, focusing on daily rainfall data. Developing and modifying the work of Heaps *et al.* (2015) and Germain (2010) on spatiotemporal models for daily rainfall at multiple sites is also an important feature of this research. Instead of using hidden weather states, we applied the Lamb weather types (LWTs) directly to the model. We also used the daily rainfall dataset for the whole year in modelling the daily rainfall with seasonal effects. This is in contrast with Heaps *et al.* (2015) and Germain (2010) who only considered winter rainfall data and excluded the seasonal effects from the model.

In Chapter 2, we presented the general concepts of time series models and Bayesian inference. We also briefly discussed the application of Bayesian inference in time series models. In Chapter 3, we introduced the basic ideas of Bayesian inference in mixture models, focusing on finite mixture models. Then, we presented the MCMC scheme for finite mixture models with some discussion on the label switching problem. The case of a mixed distribution was also discussed in this chapter and this was further expounded in Chapters 4 and 5 using an example of daily rainfall. We also applied two different mixture models to ultrasound data: a lognormal mixture and gamma mixture to illustrate the application of Bayesian inference in mixture models. The detailed parameterisation of the models and prior elicitation for each parameter were discussed in this chapter. We

ran an MCMC scheme using the RJAGS package to generate posterior samples for the parameters. The posterior summaries for both models were discussed and in general, the results indicated that there is no large difference between the models.

In Chapter 4, we presented a model for a univariate time series with a mixed distribution applied to daily rainfall data. Firstly, we introduced the general model for daily rainfall data in which we divided the model into two distinct processes: the amount and occurrence processes. The amount process was used to model the amount of daily rainfall, whilst the occurrence process was used to model the probability of rainfall occurrence. In the daily rainfall model, we emphasised the relationship between the amount and occurrence process so that no important information was lost. An important novelty in this study is the way we deal with the seasonal effects for the daily rainfall. To achieve this, we introduced two different types of truncated Fourier series to take into account the seasonal variability over a year. We also constructed suitable prior distributions for the Fourier coefficients. The diagnostic checking procedure for the mixed distribution was also introduced in this chapter to assess the adequacy of the model.

The general model was then applied to Italian and British daily rainfall datasets. For the Italian daily rainfall application, we considered three different models for the amount process. We found that the gamma distribution provided a slightly better fit compared to the lognormal distribution based on the posterior mean of mean potential rainfall amount and the posterior predictive values. Therefore, we concluded that the gamma distribution is a better choice for modelling the amount process for the Italian daily rainfall data. We compared two different ways to model the effect of covariates on the gamma distribution. The covariate effect was applied to the scale parameter for the G1 model, whilst the effect was applied to the shape parameter for the G2 model. Analysis of the Italian rainfall data suggested that the shape parameter tended not to change and those effects were seen in the scale parameter. We therefore concluded that the G1 model provided a better fit to the observed behaviour. For the occurrence process, we used a first-order Markov chain to represent the probability of rainfall occurrence. In this parameterisation, we used the mean of the amount distribution as a covariate to create a link between the amount and occurrence processes.

For the British daily rainfall dataset, we developed and modified the previous work by Heaps *et al.* (2015) and Germain (2010) by incorporating atmospheric circulation patterns directly into the model. The atmospheric circulation patterns for the United Kingdom are represented by the LWTs. Based on the exploratory analysis of the British dataset, we found a strong relationship between the LWTs and the daily rainfall data where a lower proportion of wet days was associated with anticyclonic types and a higher proportion tended to be associated with cyclonic types. A similar relationship was found with the

mean daily rainfall amounts on wet days, where the high rainfall amounts are associated with cyclonic types, whilst the low level rainfall amounts are typically associated with anticyclonic types. We also developed two models for the LWTs using homogeneous and nonhomogenous first-order Markov chains. After that, we built a univariate model for British daily rainfall where the LWT is directly incorporated into the model through the amount and occurrence processes. We also introduced a random effect at each time t which is conditionally dependent on the previous time $t-1$ to allow for temporal variability in the amount process parameterisation. The potential daily amounts were then fitted using a gamma distribution. To build the relationship between the amount and occurrence processes, we used the potential daily amount y_t as a covariate in the parameterisation of the occurrence process. From the MCMC results and diagnostic checking, we found that the model fitted the data very well. It indicated that the potential rainfall amount and the probability of rainfall occurrence are strongly associated with the LWTs.

In Chapter 5, we presented a spatiotemporal model for British daily rainfall using the data from multiple sites within the UK network. This is an extended version of the previous univariate model discussed in Chapter 4 and also an alternative model to the model of Heaps *et al.* (2015) and Germain (2010). We also investigated the spatial characteristics of the dataset and we found a decreasing dependence between the rainfall amounts and occurrences as the between-site distance increases. Therefore, it is important to add the spatial effect into the model to capture spatial variability between sites. The computation of posterior distributions for the multivariate model is more challenging than the univariate model since it involves a large number of parameters. It also requires longer computational time to generate posterior samples for the parameters. The adequacy of the model was assessed using similar diagnostic checking methods as in Chapter 4. Based on this assessment, we found that the model provides a good fit to the data. The decay parameter also suggested that the spatial correlation falls to 0.5 when the distance between sites is more than 578 km and this indicates high correlations between sites, even for the sites that are separated by large distances. We also found that the amount and occurrence processes are highly dependent on the LWT occurring at time t, a similar characteristic to the British univariate model.

6.2 Review of objectives

The objectives of the research were stated in the beginning of this thesis as follows:

1. To investigate Bayesian time series modelling in mixed or mixture distribution applications.

2. To develop novel approaches to modelling dependence, both on covariates and between realisations, and how this affects both the between-component distribution (*eg* the occurence process in the rainfall case) and the conditional within-component distribution (*eg* for the amount process in the rainfall case).

3. To investigate the computation of posterior distributions in univariate and multivariate models for mixture and mixed distributions.

4. To apply the developed methods to a number of practical problems and assess the strengths and weaknesses of each approach.

We have fulfilled the first objective in Chapter 3 by presenting Bayesian inference in mixture models. This chapter provided guidance on the general ideas and concepts of the analysis of mixture and mixed distributions within the Bayesian framework. For example, we presented the MCMC scheme for mixture models to sample the posterior distribution of parameters and auxiliary variables. We also addressed several interesting ideas and issues that were relevant to mixture models such as the label switching problem. This investigation helps us to increase our understanding of Bayesian analysis in mixture models in terms of methodology and applications.

The development of Bayesian approaches to the modelling and analysis of univariate and multivariate time series data is the central theme of this thesis. Throughout Chapters 4 and 5, we have developed univariate and multivariate models for mixed distribution using daily rainfall data. This fulfills the second objective of this thesis. In these chapters, we give particular attention to modelling dependences, both on covariates and between realisations. For instance, we have shown how the Lamb weather types (LWTs) can be directly incorporated into the model for the British daily rainfall. We found that the level of potential rainfall amount and the probability of rainfall occurrence depend on the LWTs. We also emphasised the relationship between the amount and occurrence processes for all models by making the occurrence process dependent on the amount process.

For the computation of posterior distributions, we described the general idea in Chapter 3 of an MCMC scheme to sample the posterior distribution of the parameters of mixture models. Thus, we use a similar technique to compute the posterior distributions for the univariate and multivariate models for daily rainfall in Chapters 4 and 5. Both univariate and multivariate models are quite complicated. However, we found that the computation of posterior distributions for the multivariate model is more challenging since it involves a large number of parameters which required a high computational cost to fit the model. This investigation satisfies the third objective.

Finally, the developed models are applied to the datasets in Chapters 3, 4 and 5 and this accomplishes the fourth objective. We used the ultrasound data in Chapter 3 to

illustrate the application of mixture models. We also used the Italian and British daily rainfall data to fit the univariate model in Chapter 4. The British univariate model is then extended to the multivariate model in Chapter 5 by adding multiple sites within the United Kingdom network. We have also discussed the strengths and pitfalls of the approaches in each chapter. For instance, the British daily rainfall model fitted the data very well but it required a large number of iterations when running the MCMC algorithm to obtain the posterior samples.

6.3 Future work

There are several possible areas throughout this thesis where we can suggest further work using similar concepts and methodologies. One of the possible areas is further investigation of daily rainfall using different datasets from tropical regions such as Malaysia, Indonesia and Brazil. The tropical regions have a high average precipitation over a year which is in contrast with the temperate and subpolar regions (e.g. Italy and United Kingdom). They have also a clear pattern of wet and dry seasons over a year. For instance, in Malaysia, the wet season usually occurs between November and March and the dry season always occurs from May to September. The rainfall patterns for the tropical regions are usually influenced by the El Nino Southern Oscilation (ENSO) which is a different type of atmospheric circulation pattern. Therefore, we would endeavour to apply and extend the methodologies in this thesis to daily rainfall data for the tropical regions. This will improve our understanding of the rainfall impacts on flood and drought occurrences at some specific locations.

We also recommend incorporating other physical variables such as temperature, humidity and wind speed into our daily rainfall model since there are strong correlations between precipitation and these variables. Hence, the addition of these physical variables may improve the predictive accuracy and adequacy of the model. Looking at Figure 5.4a, there may be some suggestion that the spatial correlation, as the distance becomes greater, is not tending to zero. Therefore a correlation function where the limit of the correlation as the distance tends to infinity is greater than zero might be considered. For example, instead of (5.7) we could have

$$\Sigma_{ij} = \tilde{\alpha} + (1 - \tilde{\alpha}) \exp(-\phi||s_i - s_j||)$$

with the additional parameter $\tilde{\alpha}$ where $0 \leq \tilde{\alpha} < 1$. We can also use different types of spatial correlation functions such as the Matérn correlation function to capture the spatial variability between sites. The Matérn correlation function is a popular correlation function that is increasingly used in spatial statistical modelling. The Matérn correlation

function can be defined as follows:

$$\Sigma_{ij} = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\sqrt{\nu}\phi\|s_i - s_j\|\right)^{\nu} K_{\nu} \left(2\sqrt{\nu}\phi\|s_i - s_j\|\right)$$

where K_ν is a modified Bessel function, $\nu > 0$ and ϕ are smoothness and scale parameters, respectively. We can then compare the exponential and Matérn correlation functions in terms of their suitability for modelling the spatial effect. The computing time for the spatiotemporal model to obtain posterior samples is quite long. Hence, it is recommended to do future research on the computational methods to reduce computing time.

Appendix A

Appendix to Chapter 5

A.1 The full summaries of posterior mean and standard deviation of the unknown parameters for the amount process

Amount process				
Parameter	Prior mean	Prior SD	Posterior mean	Posterior SD
$\eta_0^1(2)$	0.69	2.83	-1.014	0.082
$\eta_0^2(2)$	0.69	2.83	0.409	0.271
$\eta_0^3(2)$	0.69	2.83	-0.451	0.275
$\eta_0^4(2)$	0.69	2.83	-0.849	0.268
$\eta_0^5(2)$	0.69	2.83	-1.628	0.245
$\eta_0^6(2)$	0.69	2.83	-1.663	0.203
$\eta_0^7(2)$	0.69	2.83	-1.102	0.164
$\eta_0^8(2)$	0.69	2.83	-0.018	0.181
$\eta_0^9(2)$	0.69	2.83	0.051	0.207
$\eta_0^{10}(2)$	0.69	2.83	0.331	0.192
$\eta_0^{11}(2)$	0.69	2.83	-0.067	0.197
$\eta_0^{12}(2)$	0.69	2.83	-0.997	0.168
$\eta_0^{13}(2)$	0.69	2.83	-1.704	0.124
$\eta_0^{14}(2)$	0.69	2.83	-0.984	0.091
$\eta_0^{15}(2)$	0.69	2.83	-0.688	0.088
$\eta_0^{16}(2)$	0.69	2.83	-0.209	0.102
$\eta_0^{17}(2)$	0.69	2.83	0.617	0.127
$\eta_0^{18}(2)$	0.69	2.83	0.362	0.066
$\eta_0^{19}(2)$	0.69	2.83	0.705	0.278
$\eta_0^{20}(2)$	0.69	2.83	0.354	0.295
$\eta_0^{21}(2)$	0.69	2.83	-0.177	0.241
$\eta_0^{22}(2)$	0.69	2.83	-0.580	0.172

Amount process				
Parameter	Prior mean	Prior SD	Posterior mean	Posterior SD
$\eta_0^{23}(2)$	0.69	2.83	-0.443	0.146
$\eta_0^{24}(2)$	0.69	2.83	0.069	0.146
$\eta_0^{25}(2)$	0.69	2.83	0.419	0.169
$\eta_0^{26}(2)$	0.69	2.83	0.534	0.194
$\eta_0^{27}(2)$	0.69	2.83	-0.216	0.228
$\eta_0^{1}(3)$	0.69	2.83	-1.403	0.074
$\eta_0^{2}(3)$	0.69	2.83	-0.225	0.253
$\eta_0^{3}(3)$	0.69	2.83	-0.330	0.274
$\eta_0^{4}(3)$	0.69	2.83	-1.632	0.258
$\eta_0^{5}(3)$	0.69	2.83	-1.590	0.208
$\eta_0^{6}(3)$	0.69	2.83	-1.673	0.152
$\eta_0^{7}(3)$	0.69	2.83	-1.144	0.147
$\eta_0^{8}(3)$	0.69	2.83	-0.727	0.170
$\eta_0^{9}(3)$	0.69	2.83	-0.465	0.199
$\eta_0^{10}(3)$	0.69	2.83	0.793	0.190
$\eta_0^{11}(3)$	0.69	2.83	0.325	0.198
$\eta_0^{12}(3)$	0.69	2.83	-0.379	0.154
$\eta_0^{13}(3)$	0.69	2.83	-1.526	0.105
$\eta_0^{14}(3)$	0.69	2.83	-0.999	0.085
$\eta_0^{15}(3)$	0.69	2.83	-0.550	0.085
$\eta_0^{16}(3)$	0.69	2.83	0.076	0.098
$\eta_0^{17}(3)$	0.69	2.83	0.473	0.125
$\eta_0^{18}(3)$	0.69	2.83	0.958	0.065
$\eta_0^{19}(3)$	0.69	2.83	1.199	0.261
$\eta_0^{20}(3)$	0.69	2.83	0.973	0.289
$\eta_0^{21}(3)$	0.69	2.83	0.400	0.225
$\eta_0^{22}(3)$	0.69	2.83	-0.398	0.163
$\eta_0^{23}(3)$	0.69	2.83	-0.121	0.135
$\eta_0^{24}(3)$	0.69	2.83	0.308	0.145
$\eta_0^{25}(3)$	0.69	2.83	0.560	0.159
$\eta_0^{26}(3)$	0.69	2.83	0.964	0.190
$\eta_0^{27}(3)$	0.69	2.83	-0.061	0.215
$\eta_0^{1}(4)$	0.69	2.83	-1.233	0.080
$\eta_0^{2}(4)$	0.69	2.83	0.042	0.279
$\eta_0^{3}(4)$	0.69	2.83	-0.523	0.300
$\eta_0^{4}(4)$	0.69	2.83	-2.183	0.302
$\eta_0^{5}(4)$	0.69	2.83	-1.786	0.239
$\eta_0^{6}(4)$	0.69	2.83	-1.575	0.185
$\eta_0^{7}(4)$	0.69	2.83	-1.350	0.168
$\eta_0^{8}(4)$	0.69	2.83	-0.582	0.176
$\eta_0^{9}(4)$	0.69	2.83	-0.217	0.205
$\eta_0^{10}(4)$	0.69	2.83	0.639	0.193
$\eta_0^{11}(4)$	0.69	2.83	-0.601	0.225
$\eta_0^{12}(4)$	0.69	2.83	-1.400	0.178
$\eta_0^{13}(4)$	0.69	2.83	-1.877	0.129

Amount process				
Parameter	Prior mean	Prior SD	Posterior mean	Posterior SD
$\eta_0^{14}(4)$	0.69	2.83	-1.460	0.098
$\eta_0^{15}(4)$	0.69	2.83	-0.986	0.089
$\eta_0^{16}(4)$	0.69	2.83	-0.311	0.102
$\eta_0^{17}(4)$	0.69	2.83	0.463	0.130
$\eta_0^{18}(4)$	0.69	2.83	0.005	0.069
$\eta_0^{19}(4)$	0.69	2.83	0.629	0.273
$\eta_0^{20}(4)$	0.69	2.83	-0.228	0.309
$\eta_0^{21}(4)$	0.69	2.83	-0.481	0.241
$\eta_0^{22}(4)$	0.69	2.83	-1.144	0.188
$\eta_0^{23}(4)$	0.69	2.83	-1.079	0.150
$\eta_0^{24}(4)$	0.69	2.83	-0.635	0.166
$\eta_0^{25}(4)$	0.69	2.83	-0.203	0.168
$\eta_0^{26}(4)$	0.69	2.83	0.408	0.195
$\eta_0^{27}(4)$	0.69	2.83	-0.906	0.239
$\eta_0^{1}(5)$	0.69	2.83	-1.068	0.076
$\eta_0^{2}(5)$	0.69	2.83	0.344	0.265
$\eta_0^{3}(5)$	0.69	2.83	-0.414	0.291
$\eta_0^{4}(5)$	0.69	2.83	-1.307	0.274
$\eta_0^{5}(5)$	0.69	2.83	-1.493	0.225
$\eta_0^{6}(5)$	0.69	2.83	-0.830	0.155
$\eta_0^{7}(5)$	0.69	2.83	-0.635	0.139
$\eta_0^{8}(5)$	0.69	2.83	-0.170	0.171
$\eta_0^{9}(5)$	0.69	2.83	0.258	0.216
$\eta_0^{10}(5)$	0.69	2.83	0.827	0.191
$\eta_0^{11}(5)$	0.69	2.83	0.680	0.200
$\eta_0^{12}(5)$	0.69	2.83	0.517	0.146
$\eta_0^{13}(5)$	0.69	2.83	-0.621	0.106
$\eta_0^{14}(5)$	0.69	2.83	-0.306	0.088
$\eta_0^{15}(5)$	0.69	2.83	0.113	0.080
$\eta_0^{16}(5)$	0.69	2.83	0.721	0.097
$\eta_0^{17}(5)$	0.69	2.83	1.041	0.128
$\eta_0^{18}(5)$	0.69	2.83	1.620	0.068
$\eta_0^{19}(5)$	0.69	2.83	1.542	0.269
$\eta_0^{20}(5)$	0.69	2.83	1.439	0.293
$\eta_0^{21}(5)$	0.69	2.83	0.938	0.217
$\eta_0^{22}(5)$	0.69	2.83	0.858	0.164
$\eta_0^{23}(5)$	0.69	2.83	0.582	0.140
$\eta_0^{24}(5)$	0.69	2.83	1.130	0.149
$\eta_0^{25}(5)$	0.69	2.83	1.348	0.157
$\eta_0^{26}(5)$	0.69	2.83	1.628	0.188
$\eta_0^{27}(5)$	0.69	2.83	0.045	0.224

Table A.1: The prior and posterior means with standard deviations (SDs) of the unknown parameters of the amount process

A.2 The full summaries of posterior mean and standard deviation of unknown parameters for the occurrence process

Occurrence process				
Parameter	Prior mean	Prior SD	Posterior mean	Posterior SD
$\zeta_0^1(2)$	1.03	3.09	0.869	0.123
$\zeta_0^2(2)$	1.03	3.09	0.759	0.476
$\zeta_0^3(2)$	1.03	3.09	1.993	0.503
$\zeta_0^4(2)$	1.03	3.09	0.697	0.436
$\zeta_0^5(2)$	1.03	3.09	1.353	0.394
$\zeta_0^6(2)$	1.03	3.09	1.048	0.307
$\zeta_0^7(2)$	1.03	3.09	1.252	0.261
$\zeta_0^8(2)$	1.03	3.09	0.681	0.288
$\zeta_0^9(2)$	1.03	3.09	1.156	0.361
$\zeta_0^{10}(2)$	1.03	3.09	2.050	0.363
$\zeta_0^{11}(2)$	1.03	3.09	1.881	0.363
$\zeta_0^{12}(2)$	1.03	3.09	1.339	0.283
$\zeta_0^{13}(2)$	1.03	3.09	1.262	0.203
$\zeta_0^{14}(2)$	1.03	3.09	1.307	0.155
$\zeta_0^{15}(2)$	1.03	3.09	1.435	0.146
$\zeta_0^{16}(2)$	1.03	3.09	1.725	0.178
$\zeta_0^{17}(2)$	1.03	3.09	1.633	0.255
$\zeta_0^{18}(2)$	1.03	3.09	1.644	0.135
$\zeta_0^{19}(2)$	1.03	3.09	1.062	0.587
$\zeta_0^{20}(2)$	1.03	3.09	1.787	0.615
$\zeta_0^{21}(2)$	1.03	3.09	1.787	0.456
$\zeta_0^{22}(2)$	1.03	3.09	1.603	0.327
$\zeta_0^{23}(2)$	1.03	3.09	1.892	0.294
$\zeta_0^{24}(2)$	1.03	3.09	1.811	0.299
$\zeta_0^{25}(2)$	1.03	3.09	1.787	0.342
$\zeta_0^{26}(2)$	1.03	3.09	1.546	0.403
$\zeta_0^{27}(2)$	1.03	3.09	1.662	0.404
$\zeta_0^1(3)$	1.03	3.09	2.918	0.140
$\zeta_0^2(3)$	1.03	3.09	4.263	0.540
$\zeta_0^3(3)$	1.03	3.09	1.927	0.527
$\zeta_0^4(3)$	1.03	3.09	2.250	0.443
$\zeta_0^5(3)$	1.03	3.09	3.118	0.365
$\zeta_0^6(3)$	1.03	3.09	3.323	0.274
$\zeta_0^7(3)$	1.03	3.09	2.914	0.251
$\zeta_0^8(3)$	1.03	3.09	2.998	0.300
$\zeta_0^9(3)$	1.03	3.09	2.692	0.377
$\zeta_0^{10}(3)$	1.03	3.09	3.795	0.447

Occurrence process				
Parameter	Prior mean	Prior SD	Posterior mean	Posterior SD
$\zeta_0^{11}(3)$	1.03	3.09	2.550	0.408
$\zeta_0^{12}(3)$	1.03	3.09	2.670	0.297
$\zeta_0^{13}(3)$	1.03	3.09	3.176	0.209
$\zeta_0^{14}(3)$	1.03	3.09	3.737	0.188
$\zeta_0^{15}(3)$	1.03	3.09	3.613	0.175
$\zeta_0^{16}(3)$	1.03	3.09	3.487	0.218
$\zeta_0^{17}(3)$	1.03	3.09	4.070	0.306
$\zeta_0^{18}(3)$	1.03	3.09	3.912	0.229
$\zeta_0^{19}(3)$	1.03	3.09	2.493	0.718
$\zeta_0^{20}(3)$	1.03	3.09	2.536	0.686
$\zeta_0^{21}(3)$	1.03	3.09	2.849	0.607
$\zeta_0^{22}(3)$	1.03	3.09	3.545	0.389
$\zeta_0^{23}(3)$	1.03	3.09	2.906	0.360
$\zeta_0^{24}(3)$	1.03	3.09	3.901	0.402
$\zeta_0^{25}(3)$	1.03	3.09	3.553	0.456
$\zeta_0^{26}(3)$	1.03	3.09	2.556	0.569
$\zeta_0^{27}(3)$	1.03	3.09	2.714	0.430
$\zeta_0^{1}(4)$	1.03	3.09	1.839	0.133
$\zeta_0^{2}(4)$	1.03	3.09	1.639	0.495
$\zeta_0^{3}(4)$	1.03	3.09	1.019	0.532
$\zeta_0^{4}(4)$	1.03	3.09	2.209	0.462
$\zeta_0^{5}(4)$	1.03	3.09	1.699	0.409
$\zeta_0^{6}(4)$	1.03	3.09	2.154	0.306
$\zeta_0^{7}(4)$	1.03	3.09	1.939	0.270
$\zeta_0^{8}(4)$	1.03	3.09	1.927	0.291
$\zeta_0^{9}(4)$	1.03	3.09	2.293	0.379
$\zeta_0^{10}(4)$	1.03	3.09	1.989	0.418
$\zeta_0^{11}(4)$	1.03	3.09	1.394	0.367
$\zeta_0^{12}(4)$	1.03	3.09	1.547	0.293
$\zeta_0^{13}(4)$	1.03	3.09	1.808	0.215
$\zeta_0^{14}(4)$	1.03	3.09	1.988	0.172
$\zeta_0^{15}(4)$	1.03	3.09	1.931	0.156
$\zeta_0^{16}(4)$	1.03	3.09	2.358	0.193
$\zeta_0^{17}(4)$	1.03	3.09	1.870	0.275
$\zeta_0^{18}(4)$	1.03	3.09	2.506	0.148
$\zeta_0^{19}(4)$	1.03	3.09	2.308	0.654
$\zeta_0^{20}(4)$	1.03	3.09	2.809	0.586
$\zeta_0^{21}(4)$	1.03	3.09	0.587	0.456
$\zeta_0^{22}(4)$	1.03	3.09	1.745	0.315
$\zeta_0^{23}(4)$	1.03	3.09	2.220	0.294
$\zeta_0^{24}(4)$	1.03	3.09	1.760	0.296
$\zeta_0^{25}(4)$	1.03	3.09	3.346	0.386

Occurrence process				
Parameter	Prior mean	Prior SD	Posterior mean	Posterior SD
$\zeta_0^{26}(4)$	1.03	3.09	2.040	0.436
$\zeta_0^{27}(4)$	1.03	3.09	2.382	0.420
$\zeta_0^{1}(5)$	1.03	3.09	2.216	0.144
$\zeta_0^{2}(5)$	1.03	3.09	2.457	0.495
$\zeta_0^{3}(5)$	1.03	3.09	1.534	0.513
$\zeta_0^{4}(5)$	1.03	3.09	1.975	0.452
$\zeta_0^{5}(5)$	1.03	3.09	2.496	0.371
$\zeta_0^{6}(5)$	1.03	3.09	2.893	0.283
$\zeta_0^{7}(5)$	1.03	3.09	2.969	0.261
$\zeta_0^{8}(5)$	1.03	3.09	2.543	0.318
$\zeta_0^{9}(5)$	1.03	3.09	1.462	0.398
$\zeta_0^{10}(5)$	1.03	3.09	2.859	0.394
$\zeta_0^{11}(5)$	1.03	3.09	1.976	0.380
$\zeta_0^{12}(5)$	1.03	3.09	1.454	0.304
$\zeta_0^{13}(5)$	1.03	3.09	2.369	0.217
$\zeta_0^{14}(5)$	1.03	3.09	3.069	0.189
$\zeta_0^{15}(5)$	1.03	3.09	2.871	0.179
$\zeta_0^{16}(5)$	1.03	3.09	2.602	0.236
$\zeta_0^{17}(5)$	1.03	3.09	2.947	0.318
$\zeta_0^{18}(5)$	1.03	3.09	3.503	0.259
$\zeta_0^{19}(5)$	1.03	3.09	2.449	0.726
$\zeta_0^{20}(5)$	1.03	3.09	3.162	0.761
$\zeta_0^{21}(5)$	1.03	3.09	3.311	0.646
$\zeta_0^{22}(5)$	1.03	3.09	2.486	0.476
$\zeta_0^{23}(5)$	1.03	3.09	3.411	0.439
$\zeta_0^{24}(5)$	1.03	3.09	3.609	0.534
$\zeta_0^{25}(5)$	1.03	3.09	3.187	0.602
$\zeta_0^{26}(5)$	1.03	3.09	3.440	0.760
$\zeta_0^{27}(1)$	1.03	3.09	2.113	0.464

Table A.2: The prior and posterior means with standard deviations (SDs) of the unknown parameters of the occurrence process

References

Al-Awadhi, S. A. & Garthwaite, P. H. 1998 An elicitation method for multivariate normal distributions. *Communications in Statistics – Theory and Methods* **27**, 1123–1142.

Al-Awadhi, S. A. & Garthwaite, P. H. 2001 Prior distribution assessment for a multivariate normal distribution: An experimental study. *Journal of Applied Statistics* **28**, 5–23.

de Alba, E. 1993 Constrained forecasting in autoregressive time series models: A Bayesian analysis. *International Journal of Forecasting* **9** (1), 95–108.

Azzalini, A. & Bowman, A. W. 1990 A look at some data on the old faithful geyser. *Applied Statistics* **39** (3), 357.

Banerjee, S., Carlin, B. & Gelfand, A. 2004 *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press.

Barry, R. G. & Chorley, R. J. 2003 *Atmosphere, weather, and climate*, 8th edn. London : Routledge.

Ben-Gai, T., Bitan, A., Manes, A., Alpert, P. & Rubin, S. 1998 Spatial and temporal changes in rainfall frequency distribution patterns in Israel. *Theoretical and Applied Climatology* **61** (3), 177–190.

Bennett, J., Racine-Poon, A. & Wakefiled, J. 1996 MCMC for non-linear hierarchical models. In *Markov chain Monte Carlo in practice* (ed. W. R. Gilks, S. Richardson & D. J. Spiegelhalter), pp. 339–357. Chapman & Hall/CRC.

Besag, J. 1974 Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)* **36** (2), 192–236.

Besag, J., Green, P., Higdon, D. & Mengersen, K. 1995 Bayesian computation in stochastic systems. *Statistical Science* **10**, 3–66.

Box, G. E. P. & Jenkins, G. M. 1976 *Time Series Analysis: Forecasting and Control*, revised edn. Holden-Day.

Brooks, S., Gelman, A., Jones, G. & Meng, X. 2011 *Handbook of Markov Chain Monte Carlo*. CRC Press.

Brooks, S. P. & Gelman, A. 1998 General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* **7** (4), 434–455.

Bruno, F., Cocchi, D., Greco, F. & Scardovi, E. 2014 Spatial reconstruction of rainfall fields from rain gauge and radar data. *Stochastic Environmental Research and Risk Assessment* **28** (5), 1235–1245.

Cappé, O., Moulines, E. & Ryden, T. 2005 *Inference in hidden Markov models*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Chandler, R. E. & Wheater, H. S. 2002 Analysis of rainfall variability using generalized linear models: A case study from the West of Ireland. *Water Resources Research* **38** (10), 10–1–10–11.

Chatfield, C. 2003 *The analysis of time series : an introduction*. Boca Raton, FL: Chapman & Hall/CRC.

Cho, H.-K., Bowman, K. P. & North, G. R. 2004 A comparison of gamma and lognormal distributions for characterizing satellite rain rates from the tropical rainfall measuring mission. *Journal of Applied Meteorology* **43** (11), 1586–1597.

Christensen, A. L., Lundbye-Christensen, S., Overvad, K., Rasmussen, L. H. & Dethlefsen, C. 2012 Modeling gradually changing seasonal variation in count data using state space models: a cohort study of hospitalization rates of stroke in atrial fibrillation patients in Denmark from 1977 to 2011. *BMC Medical Research Methodology* **12** (1), 174.

Coe, R. & Stern, R. D. 1982 Fitting models to daily rainfall data. *Journal of Applied Meteorology* **21** (7), 1024–1031.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, ., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D. & Worley, S. J. 2011 The twentieth century reanalysis project. *Quarterly Journal of the Royal Meteorological Society* **137** (654), 1–28.

Cowburn, G. J., Parker, N. G. & Farrow, M. 2018 Bayesian inference for fluctuating ultrasonic pulse-echo measurements. In preparation.

Cowles, M. K. & Carlin, B. P. 1996 Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association* **91** (434), 883–904.

Delmar, P., Robin, S. & Daudin, J. J. 2004 VarMixt: efficient variance modelling for the differential analysis of replicated gene expression data. *Bioinformatics* **21** (4), 502–508.

Durbin, J. & Koopman, S. J. 2000 Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Society B* **62**, 3–56.

Farrow, M. 2003 Practical building of subjective covariance structures for large complicated systems. *The Statistician* **52**, 553–573.

Fernandes, M. V., Schmidt, A. M. & Migon, H. S. 2009 Modelling zero-inflated spatio-temporal processes. *Statistical Modelling* **9** (1), 3–25.

Frühwirth-Schnatter, S. 2006 *Finite mixture and Markov switching models*. Springer.

Furrer, E. & Katz, R. 2007 Generalized linear modeling approach to stochastic weather generators. *Climate Research* **34**, 129–144.

Gabriel, K. R. & Neumann, J. 1962 A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quarterly Journal of the Royal Meteorological Society* **88** (375), 90–95.

Gamerman, D. 1997 Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing* **7** (1), 57–68.

Gamerman, D. & Lopes, H. F. 2006 *Markov chain Monte Carlo : stochastic simulation for Bayesian inference*. Boca Raton: Taylor & Francis.

Garthwaite, P. H., Kadane, J. B. & O'Hagan, A. 2005 Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association* **100** (470), 680–701.

Gelfand, A. E. & Smith, A. F. M. 1990 Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85** (410), 398–409.

Gelman, A., Roberts, G. & Gilks, W. 1996 Efficient Metropolis jumping rules. In *Bayesian Statistics* (ed. J. M. Bernado *et al.*), pp. 599–607. OUP.

Gelman, A. & Rubin, D. B. 1992 Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7** (4), 457–472.

Geman, S. & Geman, D. 1984 Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Recognition* **6**, 721–741.

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

Geweke, J. 1992 Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *IN BAYESIAN STATISTICS*, pp. 169–193. University Press.

Glickman, T. S. 2000 *Glossary of meteorology*. American Meteorological Society.

Grunwald, G. K. & Jones, R. H. 2000 Markov models for time series with mixed distribution. *Environmetrics* **11** (3), 327–339.

Harrison, P. J. & Stevens, C. F. 1971 A Bayesian approach to short-term forecasting. *Journal of the Operational Research Society* **22** (4), 341–362.

Harrison, P. J. & Stevens, C. F. 1976 Bayesian forecasting. *Journal of the Royal Statistical Society. Series B (Methodological)* **38** (3), 205–247.

Hastings, W. K. 1970 Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** (1), 97–109.

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

Heaps, S. E., Boys, R. J. & Farrow, M. 2015 Bayesian modelling of rainfall data by using non-homogeneous hidden Markov models and latent Gaussian variables. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **71**, 543–568.

Heidelberger, P. & Welch, P. D. 1983 Simulation run length control in the presence of an initial transient. *Operations Research* **31** (6), 1109–1144.

Hess, P. & Brezowsky, H. 1969 *Katalog der Grosswetterlagen Europas*. D.W.

Jasra, A., Holmes, C. C. & Stephens, D. A. 2005 Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statist. Sci.* **20** (1), 50–67.

Jenkinson, A. F. & Collinson, B. P. 1977 An initial climatology of gales over the North Sea. Synoptic climatology branch memorandum 62. Meteorological Office, Bracknell, UK.

Jimoh, O. & Webster, P. 1996 The optimum order of a Markov chain model for daily rainfall in Nigeria. *Journal of Hydrology* **185** (1), 45 – 69.

Jones, P. D., Harpham, C. & Briffa, K. R. 2013 Lamb weather types derived from reanalysis products. *International Journal of Climatology* **33** (5), 1129–1139.

Jones, R. H. & Brelsford, W. M. 1967 Time series with periodic structure. *Biometrika* **54** (3-4), 403–408.

Kadane, J. & Wolfson, L. J. 1998 Experiences in elicitation. *Journal of the Royal Statistical Society: Series D (The Statistician)* **47** (1), 3–19.

Kalman, R. E. & Bucy, R. S. 1961 New results in linear filtering and prediction theory. *Journal of Basic Engineering* **83** (1), 95.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R. & Joseph, D. 1996 The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society* **77** (3), 437–471.

Katz, R. W. 1977 Precipitation as a chain-dependent process. *Journal of Applied Meteorology* **16** (7), 671–676.

Kendall, D. G. 1953 Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. *The Annals of Mathematical Statistics* **24** (3), 338–354.

Kitagawa, G. 1998 A self-organizing state-space model. *Journal of the American Statistical Association* **93** (443), 1203.

Knorr-Held, L. & Besag, J. 1998 Modelling risk from a disease in time and space. *Statistics in Medicine* **17** (18), 2045–2060.

Kong, A., Liu, J. S. & Wong, W. H. 1994 Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association* **89** (425), 278.

Lamb, H. H. 1972 *British Isles weather types and a register of the daily sequence of circulation patterns 1861-1971*. H.M. Stationery Off.

Lima, C. H. R. & Lall, U. 2009 Hierarchical Bayesian modeling of multisite daily rainfall occurrence: Rainy season onset, peak, and end. *Water Resources Research* **45** (7).

Liu, Y., Zhang, W., Shao, Y. & Zhang, K. 2011 A comparison of four precipitation distribution models used in daily stochastic models. *Advances in Atmospheric Sciences* **28** (4), 809–820.

Lunn, D., Jackson, C., Best, N., Thomas, A. & Spiegelhalter, D. 2012 *The BUGS Book: A Practical Introduction to Bayesian Analysis*. London: Chapman and Hall.

Marin, J.-M., Mengersen, K. & Robert, C. P. 2005 Bayesian modelling and inference on mixtures of distributions. *Handbook of Statistics 25, (D. Dey and C. R. Rao, eds.)* **volume 25 (to appear)**, 15840–15845.

Masson, P. R. 2001 Exchange rate regime transitions. *Journal of Development Economics* **64** (2), 571–586.

Matérn, B. 1986 *Spatial Variation*, 2nd edn. Berlin: Springer-Verlag.

McCullagh, P. & Nelder, J. A. 1989 *Generalized linear models (Second edition)*. London: Chapman & Hall.

McLachlan, G. & Peel, D. 2001 *Finite mixture models*. Wiley.

Met Office 2017 Meteorological Office land surface stations observations 2017. `http://data.ceda.ac.uk/badc/ukmo-surface/data/united_kingdom/`.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. 1953 Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* **21** (6), 1087–1092.

Monahan, J. F. 1983 Fully Bayesian analysis of ARMA time series models. *Journal of Econometrics* **21** (3), 307–331.

Nelder, J. A. & Wedderburn, R. W. M. 1972 Generalized linear models. *Journal of the Royal Statistical Society, Series A, General* **135**, 370–384.

Newcomb, S. 1886 A generalized theory of the combination of observations so as to obtain the best result. *American Journal of Mathematics* **8** (4), 343.

O'Hagan, A. 1998 Eliciting expert beliefs in substantial practical applications. *Journal of the Royal Statistical Society: Series D (The Statistician)* **47** (1), 21–35.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E. & Rakow, T. 2006 *Uncertain Judgements: Eliciting Expert Probabilities*. Chichester: John Wiley.

Pearson, K. 1894 Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **185** (0), 71–110.

Petris, G., Petrone, S. & Campagnoli, P. 2009 *Dynamic linear models with R*. Dordrecht New York: Springer-Verlag.

Plummer, M. 2012 *Bayesian Graphical Models using MCMC*, 3rd edn. CRAN.

Puolamäki, K. & Kaski, S. 2009 Bayesian solutions to the label switching problem. *Lecture Notes in Computer Science* **5772 LCNS**, 381–392, cited By 1.

R Development Core Team 2008 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Raftery, A. E. & Lewis, S. 1992 How many iterations in the Gibbs sampler? In *Bayesian Statistics 4*, pp. 763–773. Oxford University Press.

Redner, R. A. & Walker, H. F. 1984 Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.* **26** (2), 195–239.

Reich, B. J., Hodges, J. S. & Zadnik, V. 2006 Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics* **62** (4), 1197–1206.

Richardson, C. W. 1981 Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research* **17** (1), 182–190.

Roldán, J. & Woolhiser, D. A. 1982 Stochastic daily precipitation models: 1. A comparison of occurrence processes. *Water Resources Research* **18** (5), 1451–1459.

Rosenberg, M. A. & Young, V. R. 1999 A Bayesian approach to understanding time series data. *North American Actuarial Journal* **3** (2), 130–143.

Sanso, B. & Guenni, L. 1999 A stochastic model for tropical rainfall at a single location. *Journal of Hydrology* **214** (1-4), 64–73.

Scott, S. L. 2002 Bayesian methods for hidden Markov models. *Journal of the American Statistical Association* **97** (457), 337–351.

Smith, J. Q. 1985 Diagnostic checks of non-standard time series models. *Journal of Forecasting* **4** (3), 283–291.

Sofia, S. 2007 Nonlinear time-series models for Mediterranean rainfall data with zeroes. PhD thesis, University of Sunderland, UK.

Stephens, M. 2000 Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62** (4), 795–809.

Stern, R. D. & Coe, R. 1984 A model fitting analysis of daily rainfall data. *Journal of the Royal Statistical Society. Series A (General)* **147** (1), 1.

Stewart, W. 2009 *Probability, Markov chains, queues, and simulation : the mathematical basis of performance modeling*. Princeton, N.J: Princeton University Press.

Suhaila, J., Ching-Yee, K., Fadhilah, Y. & Hui-Mean, F. 2011 Introducing the mixed distribution in fitting rainfall data. *Open Journal of Modern Hydrology* **01** (02), 11–22.

Tadesse, M. G., Sha, N. & Vannucci, M. 2005 Bayesian variable selection in clustering high-dimensional data. *Journal of the American Statistical Association* **100** (470), 602–617.

Tanner, M. A. & Wong, W. H. 1987 The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82** (398), 528.

Thompson, C. S., Thomson, P. J. & Zheng, X. 2007 Fitting a multisite daily rainfall model to New Zealand data. *Journal of Hydrology* **340** (1), 25 – 39.

Tooze, J. A., Grunwald, G. K. & Jones, R. H. 2002 Analysis of repeated measures data with clumping at zero. *Statistical Methods in Medical Research* **11** (4), 341–355.

Velarde, L. G. C., Migon, H. S. & Pereira, B. d. B. 2004 Space–time modeling of rainfall data. *Environmetrics* **15** (6), 561–576.

Waller, L. A. & Carlin, B. P. 2010 Disease mapping. *Chapman Hall CRC Handb Mod Stat Methods* **2010**, 217–243.

West, M. & Harrison, J. 1997 *Bayesian forecasting and dynamic methods*. Springer-Verlag.

West, M., Harrison, P. J. & Migon, H. S. 1985 Dynamic generalised linear models and Bayesian forecasting (with discussion). *Journal of the American Statistical Association* **80**, 73–97.

Wilks, D. 1999 Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. *Agricultural and Forest Meteorology* **93** (3), 153–169.

Winkler, R. L. 1967 The assessment of prior distributipons in Bayesian analysis. *Journal of the American Statistical Association* **62**, 776–880.

Yang, C., Chandler, R. E., Isham, V. S. & Wheater, H. S. 2005 Spatial-temporal rainfall simulation using generalized linear models. *Water Resources Research* **41** (11), w11415.

Yarnal, B. 1993 *Synoptic climatology in environmental analysis : a primer*. Belhaven Press London ; Boca Raton, Fla.

Zellner, A. 1971 *An introduction to Bayesian inference in econometrics*. John Wiley.

Zucchini, W. 2009 *Hidden Markov models for time series : an introduction using R*. Boca Raton: CRC Press/Taylor & Francis Group.