A time series model for precipitation based on disaggregation and lognormal point processes

Nikolaus Sebastian Ruf

Vom Fachbereich Mathematik der Universität Kaiserslautern
zur Verleihung des akademischen Grades Doktor der
Naturwissenschaften (Doctor rerum naturalium, Dr. rer. nat.)
genehmigte Dissertation

17. Juni 2008

Gutacher: Prof. Dr. Jürgen Franke
Prof. Dr. Claudia Czado

D 386

ii

Abstract

In this thesis, we investigate a statistical model for precipitation time series recorded at a single site. The sequence of observations consists of rainfall amounts aggregated over time periods of fixed duration. As the properties of this sequence depend strongly on the length of the observation intervals, we follow the approach of Rodriguez-Iturbe et. al. [37] and use an underlying model for rainfall intensity in continuous time. In this idealized representation, rainfall occurs in clusters of rectangular cells, and each observations is treated as the sum of cell contributions during a given time period. Unlike the previous work, we use a multivariate lognormal distribution for the temporal structure of the cells and clusters.

After formulating the model, we develop a Markov-Chain Monte-Carlo algorithm for fitting it to a given data set. A particular problem we have to deal with is the need to estimate the unobserved intensity process alongside the parameter of interest. The performance of the algorithm is tested on artificial data sets generated from the model.

Acknowledgements

Contents

List of Figures

List of Tables

Chapter 1

Introduction

If one asked a meteorologist what he thought of a statistical model for rainfall at a single site, he would probably answer that he did not need such a thing for two reasons. Firstly, single site models are not very useful for predicting the weather, even in relatively close proximity and for short periods. Weather is a highly variable process, and measurements tend to be less accurate the more extreme the events one tries to observe. Prediction requires a "big picture" approach, taking into account multiple sites or spatial information like satellite data. And secondly, there seems to be a bias against purely statistical models in meteorology. Black-box models that do not rely on physics are not popular.

So why do we deal with this topic? The original motivation was to find a compact description of rainfall over a small area for use in rainwater runoff studies. In this context, it is useful if one can generate precipitation time series of arbitrary length that capture key features of the real process like the distribution of wet and dry periods, or of the rainfall amounts during a single storm. While we briefly discuss prediction and extensions to multi-site modelling in the last chapter, the main purpose of our model is the simulation of localized events.

Before we talk about rainfall models, we give a brief review of the statistical concepts used in this thesis in Chapter 2. The focus is on Markov-Chain Monte-Carlo (MCMC) methods, in particular the Metropolis-Hastings algorithm. We provide an overview of its theoretical background and discuss some implementation and diagnostic issues.

Chapter 3 introduces the rainfall model we want to analyze in the remainder of the thesis. It is based on the work of Rodriguez-Iturbe et. al. [37] The general idea is to describe rainfall as a sequence of rectangular pulses or cells, which results in a step function for the intensity over time. The actual observations are treated as the aggregated contributions (integrals) of these cells for each time interval. As the cell lengths and origins "live" in continuous time, the resulting model is independent of the time scale of the observations. For a more realistic description of actual rainfall, cells are only permitted to occur during certain periods called storms, which leads to clustering. The origins of the storms and of the cells within each storm form point processes, which are taken to be Poisson in the original work.

The novel part of our approach is that we are interested in adding an explicit correlation

structure for the durations and lags of the cells and storms. As this is awkward to specify for the exponential distribution, we use a multivariate lognormal model instead. One downside of this choice is that the point processes for the origins are no longer Poisson, so we have to deal with Monte-Carlo estimation of the associated weights.

Fitting the rainfall model to a given sequence of observations is difficult, as a simple parametric description is available only for the unobserved storm process. We are dealing with a hierarchical model where the dependencies between adjacent layers of the process are straightforward, but the relationship between the top layer (observations) and the bottom layer (parameter) is not. Before we can estimate the parameter of interest, we need a "disaggregation" step to reconstruct the storm process. We show how this can be accomplished using the Metropolis-Hastings algorithm in Chapter 4. The central idea is to use reversible jump MCMC as introduced by Green [22] to transform the unobserved process in accordance with the parameters and observations.

In Chapter 5, we provide numerical results for artificial data generated from the model. It turns out that the most general case is effectively overparametrized for a storm process that is not directly observable. Thus, we investigate several variants of the process with restrictions on the parameters that make them identifiable. As attempts to fit the model to real precipitation data were unsatisfactory, we briefly discussion the problems we encountered. In the final Chapter 6, we treat possible changes to the model or algorithm, both to enable fitting to real data and to extend the purpose of the model.

Chapter 2

Basic concepts

This chapter introduces notation and concepts necessary for understanding the rainfall model and estimation algorithm. In particular, we give a brief review of Markov chain theory on general state spaces and the Metropolis-Hastings algorithm.

2.1 Notation

Table 2.1 lists common symbols and expressions that are used throughout this thesis.

For the probabilistic notation, we use a tilde '$\sim$' to indicate the distribution of a random variable X, while P_X refers to the associated probability measure. The density or weights are denoted by p_X. For example, let X be a standard normal random variable on the measure space $(\mathbb{R}, \mathcal{L}(\mathbb{R}))$. We may write

$$X \sim \mathrm{N}(0,1) \qquad P_X((-\infty, 0]) = 0.5 \qquad p_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \qquad (2.1)$$

In case of a multivariate distribution $P_{X,Y}(x, y)$, we write $P_{X,Y}(x, dy)$ to denote the y-differential as $dP_{X,Y}(x, y)$ could be confusing. A bullet '$\bullet$' is sometimes used as an abbreviation for the argument. For example, $P_X(\bullet)$ stresses that we are interested in the probability measure but not a specific value for X.

2.2 Bayesian inference

We use a Bayesian framework for parametric models to fit the rainfall model developed in Chapter 3. A thorough introduction to Bayesian statistics can be found e.g. in the book by Bernardo and Smith [6], but the main idea is to treat the parameter of interest as a particular realization of a random variable:

Table 2.1: Some common expressions and abbreviations

Term	Definition
$\underline{0}$	a matrix or vector of zeros
$\mathbb{1}_S$	characteristic function of set S
$\mathrm{Corr}(X, Y)$	correlation of random variables X and Y
$\mathrm{Cov}(X, Y)$	covariance of random variables X and Y
$\mathrm{E}(X)$	expectation of random variable X
I_n	identity matrix in $\mathbb{R}^{n \times n}$
$\mathcal{L}(\mathbb{R}^n)$	Lebesgue-measurable subsets of $\mathbb{R}^n$
MCMC	Markov-Chain Monte-Carlo
MH	Metropolis-Hastings (algorithm)
$\mathbb{S}_+^n$	cone of symmetric, positive definite matrices in $\mathbb{R}^{n \times n}$
$\mathrm{Std}(X)$	standard deviation of random variable X
$\mathrm{Var}(X)$	variance of random variable X
$\overline{X}_N$	mean of $X_1, \ldots, X_N$
i.i.d.	independent and identically distributed
$\log(x)$	for $x \in \mathbb{R}_+^n$ is the vector of logarithms of the components of x
$\mathrm{tr}(A)$	trace of matrix A (sum of diagonal entries)
$|A|$	determinant of matrix A
$|S|$	cardinality of set S

Definition 2.1 *Let $(\Omega, \mathcal{A})$ and $(\Theta, \mathcal{B})$ be measure spaces with σ-algebras $\mathcal{A}$ and $\mathcal{B}$. We call Ω the* **observation space** *and Θ the* **parameter space**.

An **observation** *X is a random variable with values in Ω and a probability distribution $P_X(\bullet|\theta)$ — the* **model** *— depending on the* **unknown parameter** *$\theta \in \Theta$.*

Given a **prior distribution** *P_θ, the purpose of* **Bayesian inference** *is to determine key features (moments, quantiles, etc.) of the* **posterior distribution** *$P_\theta(\bullet|X)$.*

In this context, Bayes's theorem implies

$$p_\theta(\theta|X) = \frac{p_{X,\theta}(X,\theta)}{\int_\Theta P_{X,\theta}(X, d\theta)} \propto p_X(X|\theta)p_\theta(\theta) \tag{2.2}$$

This can be read as a statement about how to modify beliefs about the parameter once observations are taken into account. The usual problem when dealing with expressions of the type (2.2) is that the numerator can be easily evaluated at each point (X, θ), while the denominator requires considerable numerical effort. A key advantage of the Metropolis-Hastings algorithm we use for parameter estimation is that the posterior distribution needs to be known only up to a scaling constant. Thus, it is customary to drop the normalization and specify distributions only in terms of proportionality to a function of the parameters.

Remark 2.1 *Bayesian inference depends on two important choices. Even if the model is correct in the sense that it can adequately represent the data, the prior can still introduce a major bias. This is not a problem for the model discussed in this thesis, as choosing priors with a large variance results in a negligible distortion (see Section 2.3.4).*

2.3 Probability distributions

As the way certain distributions are parametrized in the stochastic literature is not unique, we want to review the ones used in this thesis briefly to avoid confusion. The expressions for the densities were taken from [6]. In all cases, the underlying measure space is $\mathbb{R}$ respectively $\mathbb{R}^n$ with the Lebesgue-measurable sets.

2.3.1 Uniform distribution

If $S \subset \mathbb{R}^n$ has positive and finite Lebesgue-measure, we say that $X \in \mathbb{R}^n$ has a uniform distribution on S, written $X \sim U(S)$, iff its density is $p_X(x) = \frac{\mathbb{1}_S(x)}{\int_S dy}$.

2.3.2 Normal and lognormal distributions

The random variable $X \in \mathbb{R}$ has a normal distribution with mean $\mu \in \mathbb{R}$ and precision $\lambda > 0$ (inverse variance) iff its density is

$$p_X(x) = \sqrt{\frac{\lambda}{2\pi}} \exp\left(-\frac{\lambda}{2}(x-\mu)^2\right) \tag{2.3}$$

In this case, we write $X \sim \mathrm{N}(\mu, \lambda)$. The density of the standard normal distribution with mean $\mu = 0$ and precision $\lambda = 1$ is denoted by

$$\phi(x) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{2.4}$$

Extending this to the multivariate case, $X \in \mathbb{R}^n$ is normally distributed with mean vector $\mu \in \mathbb{R}^n$ and precision matrix $\Lambda \in \mathbb{S}^n_+$, or $X \sim \mathrm{N}_n(\mu, \Lambda)$, iff its density at $x \in \mathbb{R}^n$ is

$$p_X(x) = \left(\frac{|\Lambda|}{(2\pi)^n}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Lambda (x-\mu)\right) \tag{2.5}$$

If we partition

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \qquad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad \Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12}^T & \Lambda_{22} \end{pmatrix} \tag{2.6}$$

with consistent dimensions, the marginal distribution for X_1 is

$$X_1 \sim \mathrm{N}(\mu_1, \Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{12}^T) \tag{2.7}$$

and the conditional distribution of X_2 given X_1 is

$$X_2 | X_1 \sim \mathrm{N}(\mu_2 - \Lambda_{22}^{-1}\Lambda_{12}^T(X_1 - \mu_1), \Lambda_{22}) \tag{2.8}$$

Note that in terms of precision, the matrix for the marginal distribution changes, while the precision matrix of the conditional distribution is the associated submatrix. This observation is reversed if we parametrize the distribution in terms of its covariance matrix.

We call $X > 0$ lognormally distributed with parameters $\mu \in \mathbb{R}$, $\lambda > 0$ iff $\log(X) \sim \mathrm{N}(\mu, \lambda)$. The transformation formula yields the density

$$p_X(x) = \sqrt{\frac{\lambda}{2\pi}} \frac{1}{x} \exp\left(-\frac{\lambda}{2}(\log(x) - \mu)^2\right) \tag{2.9}$$

The expectation and variance of a lognormal random variable are

$$\mathrm{E}(X) = \exp\left(\mu + \frac{1}{2\lambda}\right) \qquad \mathrm{Var}(X) = \exp\left(2\mu + \frac{1}{\lambda}\right)\left(\exp\left(\frac{1}{\lambda}\right) - 1\right) \tag{2.10}$$

Consequently, we call $X \in \mathbb{R}^n_+$ a multivariate lognormal random variable with parameters $\mu \in \mathbb{R}^n$ and $\Lambda \in \mathbb{S}^n_+$ iff $\log(X) \sim \mathrm{N}_n(\mu, \Lambda)$. The resulting density at $x \in \mathbb{R}^n$ is

$$p_X(x) = \left(\frac{|\Lambda|}{(2\pi)^n}\right)^{\frac{1}{2}} \frac{1}{\prod_{i=1}^n x_i} \exp\left(-\frac{1}{2}(\log(x) - \mu)^T \Lambda (\log(x) - \mu)\right) \tag{2.11}$$

2.3.3 Gamma and Wishart distributions

We say that $X > 0$ has a gamma distribution with parameters $\alpha, \beta > 0$ iff it has density

$$p_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x) \tag{2.12}$$

and write $X \sim \mathrm{Gamma}(\alpha, \beta)$. Its expectation and variance are

$$\mathrm{E}(X) = \frac{\alpha}{\beta} \qquad\qquad \mathrm{Var}(X) = \frac{\alpha}{\beta^2} \tag{2.13}$$

For our purpose, the correct multivariate extension of the gamma distribution is the Wishart distribution. We say that $X \in \mathbb{S}^n_+$ is a Wishart random variable with parameters $\alpha > \frac{n-1}{2}$ and $B \in \mathbb{S}^n_+$, or $X \sim \mathrm{Wishart}_n(\alpha, B)$, iff the corresponding density at $A \in \mathbb{S}^n_+$ is

$$p_X(A) = \frac{|B|^\alpha}{\Gamma_n(\alpha)} |A|^{\alpha - \frac{k+1}{2}} \exp(-\mathrm{tr}(BA)) \tag{2.14}$$

The density is integrated over the $\frac{n(n+1)}{2}$ distinct entries of A. To normalize, we need the generalized gamma function

$$\Gamma_n(\alpha) := \pi^{\frac{k(k-1)}{4}} \prod_{i=1}^n \Gamma\left(\frac{2\alpha + 1 - i}{2}\right) \tag{2.15}$$

The expectation of Λ is

$$\mathrm{E}(\Lambda) = \alpha B^{-1} \tag{2.16}$$

To sample from a Wishart distribution, we use the following relation found e.g. in [6]:

Proposition 2.1 *Let $k > n$ and $X_1, \ldots, X_k$ be i.i.d. $\mathrm{N}_n(\mu, \Lambda)$. Then, the random variable*

$$X := \sum_{i=1}^k (X_i - \overline{X}_k)(X_i - \overline{X}_k)^T \tag{2.17}$$

is independent of $\overline{X}_k$ and has the distribution $X \sim \mathrm{Wishart}_n\left(\frac{k-1}{2}, \frac{\Lambda}{2}\right)$.

2.3.4 Normal-Wishart conjugate priors

For a multivariate normal model, using a multivariate normal prior for the mean and a Wishart prior for the precision results in posterior distributions of the same type:

Proposition 2.2 *Let $\mathbf{X} = \{X_1, \ldots, X_k\}$ where the X_i are i.i.d. $\mathrm{N}_n(\mu, \Lambda)$. If we take as prior distributions for the parameters*

$$\mu \sim \mathrm{N}_n(\nu, L) \qquad\qquad \Lambda \sim \mathrm{Wishart}_n(\alpha, B) \tag{2.18}$$

their posterior distributions become

$$\mu|\mathbf{X}, \Lambda \sim \mathrm{N}_n \left((L + k\Lambda)^{-1} \left(L\nu + \Lambda \sum_{i=1}^{k} X_i \right), L + k\Lambda \right)$$

(2.19)

$$\Lambda|\mathbf{X}, \mu \sim \mathrm{Wishart}_n \left(\alpha + \frac{k}{2}, B + \frac{1}{2} \sum_{i=1}^{k} (X_i - \mu)(X_i - \mu)^T \right)$$

Proof   The proof is elementary, but we include it here to demonstrate the usual technique for deriving posteriors:

$$p_\mu(x|\mathbf{X}, \Lambda) \propto p_\mu(x) p_{\mathbf{X}}(\bullet|\mu = x, \Lambda)$$

(2.20)

$$\propto \exp\left(-\frac{1}{2} \left((x - \nu)^T L(x - \nu) + \sum_{i=1}^{k} (X_i - x)^T \Lambda (X_i - x) \right) \right)$$

$$\Rightarrow \mu|\mathbf{X}, \Lambda \sim \mathrm{N}_n \left((L + k\Lambda)^{-1} \left(L\nu + \Lambda \sum_{i=1}^{k} X_i \right), L + k\Lambda \right)$$

$$p_\Lambda(A|\mathbf{X}, \mu) \propto p_\Lambda(A) p_{\mathbf{X}}(\bullet|\mu, \Lambda = A)$$

(2.21)

$$\propto |A|^{\alpha + \frac{k}{2} - \frac{n+1}{2}} \exp\left(-\mathrm{tr}(BA) - \frac{1}{2} \sum_{i=1}^{k} (X_i - \mu)^T A(X_i - \mu) \right)$$

$$\Rightarrow \Lambda|\mathbf{X}, \mu \sim \mathrm{Wishart}_n \left(\alpha + \frac{k}{2}, B + \frac{1}{2} \sum_{i=1}^{k} (X_i - \mu)(X_i - \mu)^T \right)$$

To identify the second density, we use the fact that

$$\sum_{i=1}^{k} (X_i - \mu)^T A(X_i - \mu) = \mathrm{tr}\left(\sum_{i=1}^{k} (X_i - \mu)(X_i - \mu)^T A \right)$$

(2.22)

∎

Remark 2.2 *A prior distribution that results in a posterior of the same general shape is called a **conjugate prior**. This notion can be formalized to apply to all exponential families for which sufficient statistics of fixed dimension exist [6]. However, since the normal-Wishart conjugate prior is the only one we need, we include no general results here.*

We want to look at some of the properties of the one-dimensional case, which is known as the normal-gamma conjugate prior. Let $\mathbf{X} = \{X_1, \dots, X_k\}$ be i.i.d. $\mathrm{N}(\mu, \lambda)$ and assume as priors for the parameters

$$\mu \sim \mathrm{N}(0, r) \qquad\qquad \lambda \sim \mathrm{Gamma}(s, s)$$

(2.23)

The values $r, s > 0$ are arbitrary choices for the prior precisions. By the above proposition, the resulting posterior distributions are

$$\mu | \mathbf{X}, \lambda \sim N \left(\frac{\lambda \sum_{i=1}^{k} X_i}{r + k\lambda}, r + k\lambda \right)$$

(2.24)

$$\lambda | \mathbf{X}, \mu \sim \text{Gamma} \left(s + \frac{k}{2}, s + \frac{1}{2} \sum_{i=1}^{k} (X_i - \mu)^2 \right)$$

If r is small relative to k, we have

$$\text{E}(\mu | \mathbf{X}, \lambda) \approx \overline{X}_k \qquad \text{Var}(\mu | \mathbf{X}, \lambda) \approx \frac{1}{k\lambda} \qquad (2.25)$$

Thus, μ converges in probability to the maximum likelihood estimator and is asymptotically unbiased. A similar result holds for λ if s is small:

$$\text{E}(\lambda | \mathbf{X}, \mu) \approx \frac{k}{\sum_{i=1}^{k} (X_i - \mu)^2} \qquad \text{Var}(\lambda | \mathbf{X}, \mu) \approx \frac{k}{\left(\sum_{i=1}^{k} (X_i - \mu)^2 \right)^2} \qquad (2.26)$$

Here, we get convergence in probability to the inverse of the maximum likelihood estimator for the variance.

2.4 Markov chains on general state spaces

This section summarizes the key concepts necessary for discussing MCMC methods. Markov chains on general state spaces have been treated in detail by Revuz [35] and Nummelin [32]. This section is based for the most part on the more recent book by Meyn and Tweedie [30], as well as an article by Tierney [41] that deals with application to MCMC sampling. A more in-depth discussion of these methods can be found in [40], also by Tierney. The central limit theorem is due to Kipnis and Varadhan [26], as suggested by Geyer [17].

2.4.1 Markov chains with invariant distributions

In the following, let $\{X_n\}_{n \in \mathbb{N}}$ be a stochastic process with values in the measure space $(\Omega, \mathcal{A})$ with countably generated σ-algebra $\mathcal{A}$. The whole process is measurable w.r.t. the space $(\Omega^{\mathbb{N}}, \mathcal{A}^{\mathbb{N}})$, where $\mathcal{A}^{\mathbb{N}}$ denotes the joint σ-algebra (not just the Cartesian product).

Definition 2.2 *We call $\{X_n\}$ a* **Markov chain** *iff for all $n \geq 1$*

$$P_{X_n}(\bullet | X_{n-1}, \dots, X_0) = P_{X_n}(\bullet | X_{n-1}) \qquad (2.27)$$

The **initial distribution** P_{X_0} *is arbitrary.*

A function Q on $\Omega \times \mathcal{A}$ is called a **Markov transition kernel** *iff for all $x \in \Omega$ and all $A \in \mathcal{A}$*

$$Q(x, \bullet) : \mathcal{A} \to [0, 1] \text{ is a probability distribution on } \mathcal{A}$$

$$\text{(2.28)}$$

$$Q(\bullet, A) : \Omega \to \mathbb{R}_0^+ \text{ is } \mathcal{A}\text{-measurable}$$

A Markov chain is fully characterized by its initial distribution and a sequence of transition kernels $\{Q_n\}$ such that for all $A \in \mathcal{A}$ holds $P_{X_n}(A|X_{n-1}) = Q_n(X_{n-1}, A)$. We are especially interested in the case where the transition probabilities do not depend on time:

Definition 2.3 *A Markov chain $\{X_n\}$ is called* **time-homogeneous** *iff there exists a Markov transition kernel Q such that for all $n \geq 1$*

$$P_{X_n}(A|X_{n-1}) = Q(X_{n-1}, A) \tag{2.29}$$

For a time-homogeneous chain, we define the **iterated transition kernel** *Q^n via induction as*

$$Q^1(x, A) := Q(x, A) \qquad Q^n(x, A) := \int_{y \in \Omega} Q(y, A) Q^{n-1}(x, dy) \tag{2.30}$$

Note that $Q^n(x, A) = P_{X_n}(A|X_0 = x)$ is simply the n-step transition probability.

Our goal is to use Markov chains for Monte Carlo integration, i.e. to estimate functionals of a distribution of interest by their sample averages. A necessary condition for this to be meaningful is that all X_n have the same marginal distribution, at least asymptotically:

Definition 2.4 *Let Q be a Markov transition kernel and π any probability distribution. We say that Q* **has invariant distribution** *π iff for all $A \in \mathcal{A}$ holds*

$$\pi(A) = \int_\Omega Q(x, A) d\pi(x) \tag{2.31}$$

If $\{X_n\}$ is a Markov chain with transition kernels $\{Q_n\}$ and π is an invariant distribution for each Q_n, we say that $\{X_n\}$ **has invariant distribution** *π.*

If $\{X_n\}$ has invariant distribution π and initial distribution $P_{X_0} = \pi$, the marginal distribution is also π. I.e., we can shown by induction over $n \in \mathbb{N}$ that for all $A \in \mathcal{A}$ holds

$$P_{X_n}(A) = \int_\Omega P_{X_n}(A|X_{n-1} = x) dP_{X_{n-1}}(x) = \int_\Omega Q_n(x, A) d\pi(x) = \pi(A) \tag{2.32}$$

The Metropolis-Hastings algorithm we use for sampling is based on the following sufficient condition for the existence of an invariant distribution:

Proposition 2.3 *Let $\{X_n\}$ be a Markov Chain with transition kernels $\{Q_n\}$, and π a probability measure on $(\Omega, \mathcal{A})$. If the* **detailed balance equation**

$$Q_n(x, dy) d\pi(x) = Q_n(y, dx) d\pi(y) \tag{2.33}$$

is satisfied for all $n \geq 1$ and all $x, y \in \Omega$, π is an invariant distribution for $\{X_n\}$.

This is easy to show using (2.33) and the fact that $Q_n(y, \bullet)$ is a probability distribution:

$$\int_{x \in \Omega} Q_n(x, A) d\pi(x) = \int_{x \in \Omega} \int_{y \in A} Q_n(x, dy) d\pi(x) \tag{2.34}$$

$$= \int_{y \in A} \int_{x \in \Omega} Q_n(y, dx) d\pi(y) = \pi(A)$$

2.4.2 Ergodicity and asymptotic behavior

Results that rely on $P_{X_0} = \pi$ are inadequate for our purpose, as we turn to MCMC methods precisely because we are unable to sample from a given distribution π. Fortunately, it can be shown that the distribution of a Markov chain converges to its invariant distribution (if it exists) regardless of P_{X_0} under certain regularity conditions.

For the remainder of this section, let $\{X_n\}$ be a time-homogeneous Markov chain with transition kernel Q. This is the case for which most convergence results have been established in the relevant literature. It is also sufficient for our needs — although our algorithm does not generate a time-homogeneous chain, all estimators are calculated from a sub-chain with this property.

The usual condition used to establish a central limit theorem for Markov chain is ergodicity, which means that the chain. . .

- has positive probability to reach all "sets of interest" (irreducibility).

- does reach any "set of interest" infinitely often for all starting points (Harris recurrence).

- does not exhibit cyclic behavior (aperiodicity).

These notions can be formalized as follows:

Definition 2.5   *1. The* **first return time** *of $\{X_n\}$ on $A \in \mathcal{A}$ is defined as*

$$\tau_A := \min\{n \geq 1 : X_n \in \mathbf{A}\} \tag{2.35}$$

By convention, we set $\tau_A = \infty$ if the chain never returns to the set A.

*2. Let φ be any measure on the space $(\Omega, \mathcal{A})$. We say that $\{X_n\}$ is φ-**irreducible** iff for all $x \in \Omega$ and all $A \in \mathcal{A}$*

$$\varphi(A) > 0 \Rightarrow P(\tau_A < \infty | X_0 = x) > 0 \tag{2.36}$$

A Markov chain is called **irreducible** *iff it is φ-irreducible for some measure φ.*

3. We call a probability measure ψ a **maximal irreducibility distribution for** *$\{X_n\}$ iff any irreducibility measure φ is absolutely continuous w.r.t. ψ.*

*4. The chain $\{X_n\}$ is called ψ-**irreducible** iff it is irreducible and ψ is a maximum irreducibility distribution.*

Definition 2.6   *1. For $A \in \mathcal{A}$, the event that the chain visits A infinitely often is*

$$R_A := \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{X_n \in A\} \tag{2.37}$$

*2. A ψ-irreducible chain $\{X_n\}$ is called **recurrent** iff there exists a set $H \in \mathcal{A}$ with $\psi(H) = 1$ and for all $A \in \mathcal{A}$ with $\psi(A) > 0$ holds*

$$P(R_A|X_0 = x) > 0, \quad x \in \Omega \qquad P(R_A|X_0 = x) = 1, \quad x \in H \tag{2.38}$$

*3. A recurrent chain is called **Harris recurrent** iff for all $A \in \mathcal{A}$ with $\psi(A) > 0$ and all $x \in \Omega$ holds $P(R_A|X_0 = x) = 1$.*

Definition 2.7   *1. A collection of sets $A_1, \ldots, A_d \in \mathcal{A}$ is called a **cycle of length** d iff for all $i \in \{2, \ldots, d\}$ holds*

$$P_{X_n}(A_1|X_{n-1} \in A_d) = 1 \qquad P_{X_n}(A_i|X_{n-1} \in A_{i-1}) = 1 \tag{2.39}$$

*2. The **period of** $\{X_n\}$ is the length of the largest cycle associated with the chain.*

*3. We call the chain $\{X_n\}$ **aperiodic** iff it has period 1.*

Definition 2.8 *We call $\{X_n\}$ ergodic iff it is irreducible, Harris recurrent, and aperiodic.*

The usual approach in Markov chain theory is to show that ergodicity implies the existence of an invariant distribution and a law of large numbers for averages of functionals of the chain. Our situation is slightly different, as the chain we use for inference is guaranteed to have an invariant distribution by construction. As a consequence, recurrence becomes redundant. This can be seen from the following theorem found e.g. in [41]:

Theorem 2.4 *Let $\{X_n\}$ be an irreducible, time-homogeneous Markov chain with invariant distribution π. Then, π is the unique invariant distribution of the chain, it is a maximal irreducibility distribution, and $\{X_n\}$ is recurrent.*

While the theorem does not guarantee Harris recurrence, this is a mere technicality, as the following results show (see e.g. [30]):

Definition 2.9 *Let $A \in \mathcal{A}$.*

*1. The **number of returns** of $\{\mathbf{X}_n\}$ to A is*

$$\eta_A := \sum_{n=1}^{\infty} \mathbb{1}_A(X_n) \tag{2.40}$$

2. We call A **transient** *iff* $E(\eta_A|X_0 = x) < \infty$ *holds for all* $x \in \Omega$.

3. We call A **absorbing** *iff* $Q(x, A) = 1$ *for all* $x \in A$.

Theorem 2.5 *Let* $\{X_n\}$ *be recurrent with invariant distribution* π. *Then, there exist disjoint sets* $H, N \in \mathcal{A}$ *such that* $\Omega = H \cup N$ *and*

- N *is transient with* $\pi(N) = 0$.

- H *is absorbing and has the Harris property* $P_{\{X_n\}}(R_H|X_0 = x) = 1$ *for all* $x \in H$.

The existence of H already follows from Definition 2.6. The important point is the absorbing property. Assume that we generate a sample path from a recurrent chain such that P_{X_0} is absolutely continuous w.r.t. π. Then, Theorem 2.5 guarantees that path remains inside the Harris set H with probability 1 and all results for ergodic chains apply.

Ergodicity yields a strong law of large numbers and a convergence result for the transition probabilities:

Theorem 2.6 *Let* $\{X_n\}$ *be an irreducible, time-homogeneous Markov chain with transition kernel* Q *and invariant distribution* π.

1. *Any measurable function* f *on* $(\Omega, \mathcal{A})$ *with* $\int_\Omega |f(x)|d\pi(x) < \infty$ *satisfies*

$$P\left(\frac{1}{N}\sum_{n=1}^{N} f(X_n) \to \int_\Omega f(y)d\pi(y) \,\middle|\, X_0 = x\right) = 1 \tag{2.41}$$

for π-*almost all* x.

2. *If* $\{X_n\}$ *is aperiodic, the iterated transition kernels converge to* π *in total variation norm, i.e.*

$$\lim_{N\to\infty}\left(\sup_{A\in\mathcal{A}} |Q^n(x, A) - \pi(A)|\right) = 0 \tag{2.42}$$

for π-*almost all* x.

A proof from first principles can be found in the paper by Athreya, Doss, and Sethuraman [1].

Note that convergence of the kernel also implies convergence of the chain's marginal distribution. To see this, let $\epsilon > 0$. By (2.42), there exists an index N_ϵ such that for all $n \geq N_\epsilon$ holds $\sup_{A\in\mathcal{A}} |Q^n(x, A) - \pi(A)| < \epsilon$. Thus

$$\sup_{A\in\mathcal{A}} |P_{X_n}(A) - \pi(A)| = \sup_{A\in\mathcal{A}}\left|\int_\Omega Q^n(x, A) - \pi(A)\, dP_{X_0}(x)\right| \tag{2.43}$$

$$\leq \int_\Omega \sup_{A\in\mathcal{A}} |Q^n(x, A) - \pi(A)|\, dP_{X_0}(x) \leq \epsilon$$

2.4.3　The Kipnis-Varadhan Central Limit Theorem

There are several ways to strengthen the conditions further to obtain a central limit theorem. The version we want to use applies directly only to a chain that has reached a stable state:

Definition 2.10 *The process $\{X_n\}$ is called (strictly) stationary iff the vectors $(X_{i_1}, \ldots, X_{i_J})$ and $(X_{i_1+k}, \ldots, X_{i_J+k})$ have the same distribution for all finite sets $I = \{i_1, \ldots, i_J\} \subset \mathbb{N}$ and all time shifts $k \in \mathbb{N}$.*

A time-homogeneous Markov chain with invariant distribution π and $P_{X_0} = \pi$ is an example of a stationary process. This follows from (2.32) and the Markov chain's conditional independence of the past (2.27).

The following CLT is a special case of the results given by Kipnis and Varadhan in [26]:

Theorem 2.7 *Let $\{X_n\}$ be a time-homogeneous, stationary, and ergodic Markov chain with transition kernel Q and invariant distribution π satisfying the detailed balance equation (2.33). Let f be any measurable function mapping $(\Omega, \mathcal{A})$ into $(\mathbb{R}, \mathcal{L}(\mathbb{R}))$ such that*

$$\mu := \int_\Omega f(x)\, d\pi(x) \qquad\qquad s := \int_\Omega f(x)^2 d\pi(x) \qquad (2.44)$$

exist. Then, if

$$\sigma^2 := \lim_{N \to \infty} \mathrm{E}\left(\frac{1}{N} \left(\sum_{n=0}^{N-1} (f(X_n) - \mu) \right)^2 \right) < \infty \qquad (2.45)$$

the estimator

$$\overline{\mu}_N := \frac{1}{N} \sum_{n=0}^{N-1} f(X_n) \qquad (2.46)$$

converges in distribution as

$$\sqrt{N}\, \frac{\overline{\mu}_N - \mu}{\sigma} \to \mathrm{N}(0,1) \qquad (2.47)$$

Under the conditions of the theorem, we can also write

$$\sigma^2 = r_0 + 2 \sum_{n=1}^{\infty} r_n < \infty \qquad (2.48)$$

where the r_n are the autocovariances of the stationary sequence $\{f(X_n)\}$, i.e.

$$r_n := \mathrm{Cov}(f(X_0), f(X_n)) \qquad (2.49)$$

Note that σ^2 is the spectral density of the stationary process at zero, which may be estimated using a window estimator (see e.g. the book by Priestly [34]). However, we can exploit the special structure of the process as shown by Geyer in [17]:

Theorem 2.8 *Under the conditions of Theorem (2.7) and with r_n defined as in (2.49), the sequence $\{S_m\}$ given for $m \geq 0$ by*

$$S_m := r_{2m} + r_{2m+1} \tag{2.50}$$

is positive, strictly decreasing, and strictly convex in m.

This leads to the following scheme for estimating σ^2:

Definition 2.11 *Let $\{X_n\}$ and f satisfy the conditions of Theorem (2.7). For $N \in \mathbb{N}$ and $0 \leq n \leq N - 1$, we estimate the autocovariances by*

$$\hat{r}_n^N := \frac{1}{N} \sum_{k=0}^{N-n-1} (f(X_k) - \overline{\mu}_N)(f(X_{k+n}) - \overline{\mu}_N) \tag{2.51}$$

and the sequence $\{S_m\}$ by

$$\hat{S}_m^N := \hat{r}_{2m}^N + \hat{r}_{2m+1}^N \tag{2.52}$$

Let $M_N \in \{0, \ldots, \left[\frac{N-1}{2}\right]\}$ be the maximal index such that the sequence $\hat{S}_1^N$ to $\hat{S}_{M_N}^N$ is positive and strictly decreasing. The **initial monotone sequence estimator** *for the variance is*

$$\hat{\sigma}_N^2 := \hat{r}_0 + 2 \sum_{n=1}^{2M_N+1} \hat{r}_n^N = \hat{r}_0 + 2\hat{r}_1 + \sum_{m=1}^{M_N} \hat{S}_m^N \tag{2.53}$$

As shown in [17], this is a consistent overestimate and thus suitable for calculating confidence intervals:

Theorem 2.9 *For $\hat{\sigma}_N^2$ as in (2.53) and almost all sample paths of the chain holds*

$$\lim_{N \to \infty} \inf \hat{\sigma}_N^2 \geq \sigma^2 \tag{2.54}$$

Proof: We repeat the proof from [17] here, as we need a similar argument later:

By Theorem (2.6), we know that for all n and m holds

$$\hat{r}_n^N \overset{a.s.}{\to} r_n \qquad\qquad \hat{S}_m^N \overset{a.s.}{\to} S_m \tag{2.55}$$

As the true S_n are positive and strictly decreasing, this implies

$$M_N \overset{a.s.}{\to} \infty \tag{2.56}$$

Let $\epsilon > 0$. By Equation (2.48), there exists an $M_0 > 0$ such that

$$\left| \sigma^2 - r_0 - 2r_1 - 2 \sum_{m=1}^{M_0} S_n \right| < \frac{\epsilon}{2} \tag{2.57}$$

Furthermore, (2.55) implies that there exists an $N_0 \geq 0$ such that for all $N \geq N_0$ holds

$$\left| r_0 + 2r_1 + 2 \sum_{m=1}^{M_0} S_m - \hat{r}_0^N - 2\hat{r}_1^N - 2 \sum_{m=1}^{M_0} \hat{S}_m^N \right| < \frac{\epsilon}{2} \tag{2.58}$$

with probability 1. Thus, we get for sufficiently large N and almost all sample paths of the chain that

$$\hat{\sigma}_N^2 \geq \hat{r}_0^N + 2\hat{r}_1^N + 2 \sum_{m=1}^{M_0} \hat{S}_m^N \geq \sigma_N^2 - \epsilon \tag{2.59}$$

which implies (2.54). ∎

The above result is still true if we exploit the convexity of $\{S_m\}$ and base estimation on the largest convex minorant of $\{\hat{S}_M^N\}$. However, Geyer reports that the extra computational effort yields little improvement.

For a detailed proof why Theorem 2.7 applies to our situation, we need to relate the convergence of probability measures to the convergence of the associated integrals:

Proposition 2.10 *Let P and $\{P_n\}_{n \in \mathbb{N}}$ be probability measures on $(\Omega, \mathcal{A})$ such that the P_n are absolutely continuous w.r.t. P and converge in total variation norm, i.e.*

$$\lim_{n \to \infty} \left(\sup_{A \in \mathcal{A}} |P_n(A) - P(A)| \right) = 0 \tag{2.60}$$

Let $b > 0$ and define

$$\mathcal{F}_b := \{f : \Omega \to \mathbb{R} : \ f \text{ is } P\text{- and } P_n\text{-integrable for all } n, \text{ and satisfies } 0 \leq f \leq b \ P\text{-a.e.}\} \tag{2.61}$$

For all $\epsilon > 0$ exists a $N_\epsilon \geq 0$ such that for all $n \geq N_\epsilon$ and all $f \in \mathcal{F}_b$ holds

$$\left| \int_\Omega f(x) \, dP_n(x) - \int_\Omega f(x) \, dP(x) \right| < \epsilon \tag{2.62}$$

I.e., the P_n-integrals of non-negative, bounded functions converge to the P-integral uniformly w.r.t. the bound b.

Proof: Let $b > 0$, $\epsilon > 0$, and $f \in \mathcal{F}_b$. By the construction of the Lebesgue-Stieltjes integral, we can choose a sequence $\{s_n\}$ of measurable step functions which approximate f from below in the following sense:

There exist disjoint sets $A_{n,1}, \ldots, A_{n,j(n)} \in \mathcal{A}$, and constants $c_{n,1}, \ldots, c_{n,j(n)} \geq 0$ such that

$$\bigcup_{i=1}^{j(n)} A_{n,i} = \Omega \qquad\qquad s_n(x) = \sum_{i=1}^{j(n)} c_{n,i} \mathbb{1}_{A_{n,i}}(x) \tag{2.63}$$

Furthermore, $s_n \leq f$ a.e. and

$$\left| \int_\Omega s_n(x) \, dP_n(x) - \int_\Omega f(x) \, dP_n(x) \right| < \frac{\epsilon}{3} \tag{2.64}$$

W.l.o.g. each s_n is also a good approximation of f w.r.t. the P-integral:

$$\left| \int_\Omega s_n(x) \, dP(x) - \int_\Omega f(x) \, dP(x) \right| < \frac{\epsilon}{3} \tag{2.65}$$

(If this is not the case, choose a step function s satisfying $0 \le s \le f$ and (2.65). The pointwise maximum of s and s_n is a measurable step function with the desired properties.)

By (2.60) exists a N_ϵ such that for all $n \ge N_\epsilon$ and all $A \in \mathcal{A}$ holds $|P_n(A) - P(A)| < \frac{\epsilon}{6b}$. We define

$$J_0^n := \{i : P_n(A_{n,i}) \ge P(A_{n,i})\} \qquad J_1^n := \{i : P_n(A_{n,i}) < P(A_{n,i})\} \tag{2.66}$$

and observe that for all $n \ge N_\epsilon$

$$\left| \int_\Omega s_n(x) \, dP_n(x) - \int_\Omega s_n(x) \, dP(x) \right| \tag{2.67}$$

$$\le \sum_{i \in J_0^n} c_{n,i}(P_n(A_{n,i}) - P(A_{n,i})) + \sum_{i \in J_1^n} c_{n,i}(P(A_{n,i}) - P_n(A_{n,i}))$$

$$\le b \left| P_n \left(\bigcup_{i \in J_0^n} A_{n,i} \right) - P \left(\bigcup_{i \in J_0^n} A_{n,i} \right) \right| + b \left| P \left(\bigcup_{i \in J_1^n} A_{n,i} \right) - P_n \left(\bigcup_{i \in J_1^n} A_{n,i} \right) \right| \le \frac{\epsilon}{3}$$

Using (2.64), (2.65), and (2.67), we obtain for all $n \ge N_\epsilon$

$$\left| \int_\Omega f(x) \, dP_n(x) - \int_\Omega f(x) \, dP(x) \right| \le \left| \int_\Omega f(x) \, dP_n(x) - \int_\Omega s_n(x) \, dP_n(x) \right| \tag{2.68}$$

$$+ \left| \int_\Omega s_n(x) \, dP_n(x) - \int_\Omega s_n(x) \, dP(x) \right|$$

$$+ \left| \int_\Omega s_n(x) \, dP(x) - \int_\Omega f(x) \, dP(x) \right| \le \epsilon$$

∎

The previous proposition is sufficient to extend Theorem (2.7) to Markov chains with an arbitrary starting distribution:

Corollary 2.11 *Let $\{\tilde{X}_n\}$ be a time-homogeneous, ergodic Markov chain with marginal distributions $\{P_n\}$ and invariant distribution π satisfying the detailed balance equation (2.33). Let f be any measurable function mapping $(\Omega, \mathcal{A})$ into $(\mathbb{R}, \mathcal{L}(\mathbb{R}))$ such that*

$$\mu := \int_\Omega f(x) \, d\pi(x) \qquad\qquad s := \int_\Omega f(x)^2 \, d\pi(x) \tag{2.69}$$

$$\mu_n := \int_\Omega f(x) \, dP_n(x) \qquad\qquad s_n := \int_\Omega f(x)^2 \, dP_n(x)$$

exist for all n. Furthermore, let $\{X_n\}$ be a Markov chain satisfying the conditions of Theorem 2.7 with the same invariant distribution π. If σ^2 defined as in (2.45) is finite, the estimator

$$\tilde{\mu}_N := \frac{1}{N} \sum_{n=0}^{N-1} f(\tilde{X}_n) \tag{2.70}$$

converges in distribution as

$$\sqrt{N}\frac{\tilde{\mu}_N - \mu}{\sigma} \to \mathrm{N}(0,1) \tag{2.71}$$

Proof: We can assume w.l.o.g. that $\mu = 0$. As the variance σ^2 is the same as in Theorem (2.7), we need to show for

$$\gamma_N : \Omega^N \to \mathbb{R} \qquad\qquad (x_0, \ldots, x_{N-1}) \mapsto \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} f(x_n) \tag{2.72}$$

that $\sqrt{N}\tilde{\mu}_N = \gamma_N(\tilde{X}_0, \ldots, \tilde{X}_{N-1})$ has the same asymptotic distribution as $\gamma_N(X_0, \ldots, X_{N-1})$. For this purpose, define truncated averages for $M < N$ as

$$\gamma_{M,N} : \Omega^{M-N} \to \mathbb{R}, \qquad\qquad (x_M, \ldots, x_{N-1}) \mapsto \frac{1}{\sqrt{N}} \sum_{n=M}^{N-1} f(x_n) \tag{2.73}$$

Note that for fixed $M \geq 1$ holds

$$\lim_{N \to \infty} \mathrm{E}((\gamma_N(\tilde{X}_0, \ldots, \tilde{X}_{N-1}) - \gamma_{M,N}(\tilde{X}_M, \ldots, \tilde{X}_{N-1})^2) \tag{2.74}$$

$$= \lim_{N \to \infty} \frac{1}{N} \mathrm{E}\left(\left(\sum_{n=0}^{M-1} f(\tilde{X}_n) \right)^2 \right) = 0$$

I.e., the expressions converge in mean square and therefore also in distribution.

Let $\epsilon > 0$. As P_n converges to π, we can apply Proposition (2.10): there exist an M_ϵ such that for all $n \geq M_\epsilon$ and all functions f which satisfy $0 \leq f \leq 1$ and are integrable w.r.t. P_n and π holds

$$\left| \int_\Omega f(x) \, dP_n(x) - \int_\Omega f(x) \, d\pi(x) \right| < \epsilon \tag{2.75}$$

Let $B \in \mathcal{L}(\mathbb{R})$. The previous equation implies for all $N > M_\epsilon$ that

$$|P(\gamma_{M_\epsilon,N}(X_{M_\epsilon}, \ldots, X_{N-1}) \in B) - P(\gamma_{M_\epsilon,N}(\tilde{X}_{M_\epsilon}, \ldots, \tilde{X}_{N-1}) \in B)| \tag{2.76}$$

$$= \left| \int_\Omega P((X_{M_\epsilon}, \ldots, X_{N-1}) \in \gamma_{M_\epsilon,N}^{-1}(B) | X_{M_\epsilon} = x) \, dP_{X_{M_\epsilon}}(x) \right.$$

$$\left. - \int_\Omega P((\tilde{X}_{M_\epsilon}, \ldots, \tilde{X}_{N-1}) \in \gamma_{M_\epsilon,N}^{-1}(B) | \tilde{X}_{M_\epsilon} = x) \, d\pi(x) \right| < \epsilon$$

as the time-homogeneous Markov chain structure of $\{X_n\}$ and $\{\tilde{X}_n\}$ ensures that the conditional probabilities are identical and depend only on the lag $N - M_\epsilon$.

As we can write

$$|P(\gamma_N(X_0, \ldots, X_{N-1}) \in B) - P(\gamma_N(\tilde{X}_0, \ldots, \tilde{X}_{N-1}) \in B)| \tag{2.77}$$

$$\leq |P(\gamma_N(X_0, \ldots, X_{N-1}) \in B) - P(\gamma_{M_\epsilon,N}(X_0, \ldots, X_{N-1}) \in B)|$$

$$+ |P(\gamma_{M_\epsilon,N}(X_0, \ldots, X_{N-1}) \in B) - P(\gamma_{M_\epsilon,N}(\tilde{X}_0, \ldots, \tilde{X}_{N-1}) \in B)|$$

$$+ |P(\gamma_{M_\epsilon,N}(\tilde{X}_0, \ldots, \tilde{X}_{N-1}) \in B) - P(\gamma_N(\tilde{X}_0, \ldots, \tilde{X}_{N-1}) \in B)|$$

equations (2.74) and (2.76) are sufficient to establish convergence.   ∎

Remark 2.3 *Note that we can estimate σ^2 as in Definition (2.11) if we replace X_n by $\tilde{X}_n$. The proof of Theorem (2.9) relies only on the law of large numbers, Theorem (2.7), which is also satisfied by $\{\tilde{X}_n\}$.*

2.5 The Metropolis-Hastings algorithm

In the previous section, we considered conditions under which a Markov chain can be used for Monte Carlo estimation. Generating a suitable chain is possible via the Metropolis-Hastings (MH) algorithm. This sampling scheme was introduced by Metropolis et. al. [29] and later adapted for a more general context by Hastings [24]. The algorithm relies on the detailed balance equation (2.33) to ensure that the chain has the desired invariant distribution.

Algorithm 2.1 *Let π be the distribution on the probability space $(\Omega, \mathcal{A})$ from which we want to sample, let $Q(x, A)$ be any transition kernel on the same space, and let f be a π-measurable function such that*

$$\mu := \int_\Omega f(x)d\pi(x) \tag{2.78}$$

exists. Given a starting value X_0 and a desired sample size N, we construct a Markov chain $\{X_n\}$ with invariant distribution π as follows:

1. *Sample a proposal $\tilde{X}_{n+1}$ from $Q(X_n, \bullet)$.*

2. *Set $X_{n+1} = \tilde{X}_{n+1}$ with probability $\min(1, \alpha(X_n, \tilde{X}_{n+1}))$ and $X_{n+1} = X_n$ else, where*

$$\alpha(x, y) := \frac{d\pi(y)Q(y, dx)}{d\pi(x)Q(x, dy)} \tag{2.79}$$

3. *Repeat 1. and 2. $N_0 + N$ times, until the last N values of the chain resemble a stationary process.*

Estimate μ as

$$\hat{\mu} := \frac{1}{N} \sum_{n=N_0+1}^{N_0+N} f(X_n) \tag{2.80}$$

Remark 2.4   *1. The sequence of N_0 initial values which are discarded is called the **burn-in period** of the sampler.*

2. *If π and $Q(x, \bullet)$ have density or weights p respectively $q(x, \bullet)$, the quotient α can be expressed in the more usual form*

$$\alpha(X, \tilde{X}) = \frac{p(\tilde{X})q(\tilde{X}, X)}{p(X)q(X, \tilde{X})} \tag{2.81}$$

To employ the algorithm, we need to be able to evaluate the density or weights at each point. However, as α is a quotient, we do not need to worry about proper scaling. I.e., if $\pi = P_\theta(\bullet|X)$, equation (2.2) implies that we have to deal only with the prior and model distributions.

The algorithm generates an ergodic chain with invariant distribution π under minimal conditions. The following are sufficient and easily ensured in practice:

Proposition 2.12 *Let $\{X_n\}$ be a Markov chain generated by the Metropolis-Hastings algorithm 2.1 with starting point X_0 drawn from P_{X_0}. If $Q(x, \bullet)$ has a positive density w.r.t. π for all $x \in \Omega$, the chain is π-irreducible and aperiodic with invariant distribution π.*

These points are discussed e.g. in [41] or [40]. The proof is included here to illustrate how the quotient rule enforces detailed balance regardless of the proposal distribution:

Proof: To see that π is the invariant distribution of the chain, we verify that it satisfies the detailed balance equation (2.33), which is sufficient by Proposition 2.3. The transition kernel of the chain for any value X_n and set $A \in \mathcal{A}$ can be written as

$$P_{X_{n+1}}(A|X_n) = \int_{y \in A} \min(1, \alpha(X_n, y))Q(X_n, dy) \tag{2.82}$$
$$+ \mathbb{1}_{\{X_n\} \cap A}(X_{n+1}) \left(1 - \int_{y \in \Omega} \min(1, \alpha(X_n, y))Q(X_n, dy)\right)$$

The first term is the probability to propose a candidate in A and accept it. The second term is the probability to reject any proposal and stay at X_n, provided $X_n \in A$. This can be written in differential form using a δ-distribution for the point-mass:

$$dP_{X_{n+1}}(y|X_n = x) = \min(1, \alpha(x, y))Q(x, dy) \tag{2.83}$$
$$+ \delta_x(y) \left(1 - \int_{y \in \Omega} \min(1, \alpha(x, y))Q(x, dy)\right)$$

Note that the definition of α implies

$$\min(1, \alpha(x, y))Q(x, dy)d\pi(x) = \min(1, \alpha(y, x))Q(y, dx)d\pi(y) \tag{2.84}$$

as either $\min(1, \alpha(x, y))$ or $\min(1, \alpha(y, x))$ (or both) have value 1. Thus, multiplying (2.83) by $\pi(x)$ yields

$$dP_{X_{n+1}}(y|X_n = x)d\pi(x) = \min(1, \alpha(x, y))Q(x, dy)d\pi(x) \tag{2.85}$$
$$+ \delta_x(y)d\pi(x)\left(1 - \int_{y \in \Omega} \min(1, \alpha(x, y))Q(x, dy)\right)$$
$$= \min(1, \alpha(y, x))Q(y, dx)d\pi(y)$$
$$+ \delta_y(x)d\pi(y)\left(1 - \int_{x \in \Omega} \min(1, \alpha(y, x))Q(y, dx)\right)$$
$$= dP_{X_{n+1}}(x|X_n = y)d\pi(y)$$

which is the detailed balance equation (2.33) as required.

Let $x \in \Omega$ and $A \in \mathcal{A}$ with $\pi(A) > 0$. The chain is π-irreducible as

$$L(x, A) > P(X_1 \in A | X_0 = x) > \int_{y \in A} \min(1, \alpha(x, y)) Q(x, dy) > 0 \qquad (2.86)$$

Similarly, we get for all n that $P_{X_{n+1}}(A | X_n = x) > 0$, which implies aperiodicity. $\blacksquare$

Under the above conditions, Theorem 2.4 guarantees that the chain is also recurrent.

2.5.1 Choosing a proposal distribution

The shape of the proposal distribution Q has a major impact on the speed of convergence of the Markov chain to a stationary state. Two situations in particular are problematic:

- The proposals have a low likelihood w.r.t. π, resulting in a high rejection rate and a chain that remains at the same value for many iterations.

- The changes to the state of the chain are very small. While this will often result in a high acceptance rate, the chain still needs a long time to explore the entire parameter space. This is especially problematic if the starting value is unlikely w.r.t. π.

Unfortunately, there is a trade-off between the two cases, as bigger changes tend to have a lower acceptance probability and vice versa.

2.5.1.1 Blocking and the Gibbs sampler

Blocking is a basic technique for improving the performance of the MH-algorithm. It is discussed in detail in the relevant literature, e.g. in [40], in the article by Gilks et. al. [19] in [18], or in the book by Chen et. al. [7] where it is called grouping. The general idea is to generate new states of a multivariate distribution by sequential updating of blocks (groups) of components conditional on all others:

Example 2.1 *Assume that we want to sample from $\pi = P_\theta(\bullet | Z)$ in the Bayesian inference problem (2.2), and that the parameters can be partitioned as $\theta = (\theta_1, \theta_2)$. We can decompose the densities or weights at the point $x = (x_1, x_2)$ as*

$$p_\theta(x|Z) \propto p_{\theta_1}(x_1 | Z, \theta_2 = x_2) p_{\theta_2}(x_2 | Z) \qquad (2.87)$$

If we draw a candidate $\tilde{\theta}_1$ from some proposal density $q_1(\theta_1, \bullet | Z, \theta_2)$, the acceptance probability is governed by

$$\alpha(\theta_1, \tilde{\theta}_1) := \frac{p_{\theta_1}(\tilde{\theta}_1 | Z, \theta_2) q_1(\tilde{\theta}_1, \theta_1 | Z, \theta_2)}{p_{\theta_1}(\theta_1 | Z, \theta_2) q_1(\theta_1, \tilde{\theta}_1 | Z, \theta_2)} \qquad (2.88)$$

As terms that do not depend on θ_1 cancel, the rate of acceptance will often be higher than for a simultaneous update of both components.

Using the MH-acceptance probability for each update results in a chain which preserves detailed balance, although it is no longer necessarily time-homogeneous or aperiodic. However, the sub-chain we get if we take only the elements of the chain after each complete update cycle will be. Note that the block updates do not need to follow a fixed schedule; random or mixed schemes are also admissible.

Choosing how to block components is often suggested by the model. For example, in a multivariate normal model, the means can be treated as one block and the precision matrix as another. Ideally, the conditional distribution is of a known form, i.e. one from which we can sample directly. In this case, we can choose $q_1(\theta_1, \tilde{\theta}_1|Z, \theta_2) = p_{\theta_1}(\tilde{\theta}_1|Z, \theta_2)$ and achieve an acceptance probability of 1. A scheme of this type is called a Gibbs sampler, a name proposed by Geman and Geman when they applied it to Gibbs distributions on a lattice [16].

2.5.1.2   Other methods for improving convergence

The algorithm developed in Chapter 4 uses Gibbs sampling for the means of the lognormal distribution driving our rainfall model. Unfortunately, the model structure does not permit such a sampler for the precisions except in simplified cases, as the conjugate prior properties of Section 2.3.4 are destroyed. However, we can choose a factor of the posterior density that is proportional to a Wishart density and use it to define a proposal distribution. This leads to an MH-algorithm where the proposal density cancels part of the likelihood in the quotient α. The approach works as long as the posterior distribution is characterized closely by its Wishart part, since the acceptance rate will be close to 1. Unfortunately, this is not always true, and the algorithm can get stuck at the same value for many iterations.

It should be noted that a Gibbs sampler can also converge slowly, although it never gets stuck. While the blocks are updated using their true posterior distribution, this happens conditional on the current state of the model. And if part of this is misspecified, the sampler can remain in an unlikely region (w.r.t. the stationary distribution) of the parameter space for a long time.

A common technique for improving a slow sampler is reparametrization (see e.g. the article by Gilks and Roberts [20] in [18]). The basic idea is to transform the parameters in order to minimize their posterior correlation. Orthogonal search directions allow the sampler to explore the parameter space more quickly. Since de-correlation requires a linear transform, it is easy to implement for normally distributed components — the posterior stays normal. It is less clear how to handle distributions which are not invariant under linear transform. Another drawback is that one has to know the posterior correlation at least approximately, or reparametrization can worsen the situation by introducing additional dependencies.

We tested reparametrization for the normally distributed components of our model, using a correlation matrix estimated from previous iterations of the algorithm. This did not yield any significant improvement, as the algorithm's main problem seems to be identifying precisions based on a Wishart prior. And while it would be possible to specify a linear transform for the log-precisions (to maintain non-negativity), dependencies between mean and precision

parameters result in complicated posteriors. Thus, we do not include reparametrization in the algorithm presented in Chapter 4.

Remark 2.5 *Blocking and reparametrization are special cases of multi-grid Monte Carlo methods, which treat the parameter space in a coordinate-free manner. Given a (sufficiently rich) class of transforms on this space, a new proposal for θ is generated by applying a randomly chosen transform to the current state. A treatment of multi-grid Monte Carlo can be found in the article by Liu and Sabatti [28] or in [7].*

To prevent the precision sampler from getting stuck, we decided to use alternating proposal distributions for the diagonal entries of the precision matrix. Our algorithm switches randomly between a sampler based on density approximation and a normal random walk on the log-precisions. The latter is centered on the current value and has a small variance. Thus, many proposals stay close to the current value, resulting in a high acceptance rate and keeping the sampler in motion. For details of the implementation, see Section 4.3.8.

2.5.2 Convergence diagnostics

Working with an MCMC sampler, we need a way to determine the length of the burn-in period and the total length of a run. In particular, we are interested in whether the Markov chain has reached its stationary state, and how well it is mixing. Some models permit mathematical analysis to make such assessments in advance, but given the complexity of our model, we only want to look at techniques that are universally applicable. All of them focus on the monitoring of output from the sampler, either the parameter θ itself or some other statistic of the Markov chain.

The first analysis step is to look at plots of the parameter values for each iteration of the sampler and their autocorrelation functions. This will give a first indication of how many values should be discarded, and whether it is advisable to thin the chain. But there are problematic cases. For example, a process with several spontaneous shifts in the mean and slowly decaying sample autocorrelations could be non-stationary or just slowly mixing. In fact, if we monitor multiple parameters, some plots may appear stationary, while others clearly indicate that the chain has not converged. Gelman and Rubin provide strong arguments that inference from a single run is not sufficient to establish convergence in practice [15]. They give an example with multiple parallel chains where each individual process appears stationary but has a distinct mean from the rest. The lesson is that one should consider multiple summaries taken from several runs when assessing convergence if possible. Unfortunately, our algorithm for fitting the rainfall model requires substantial computation time so that we can provide only few results for parallel runs.

We employ two different methods for analyzing convergence. The first is based on confidence intervals for θ, which we can construct using Corollary 2.11. If we have $2N$ parameter samples available, let I_1 be a $\frac{\gamma}{2}$-confidence interval derived from the first N values, and I_2 a $\frac{\gamma}{2}$ interval

based on samples $N + 1$ to $2N$. Under the hypothesis that the process is stationary, we know that

$$P(I_1 \cap I_2 = \emptyset) \leq P(\theta \notin I_1 \vee \theta \notin I_2) \leq \gamma \tag{2.89}$$

I.e., we have a level-γ test for non-stationarity. We can use it to compare parts of a single sequence or from two independent sequences. Unfortunately, it is based on the variance estimator $\hat{\sigma}_N^2$ of Definition 2.11. Thus, the level is by no means guaranteed for a slow mixing sampler.

The second method is developed by Gelman and Rubin for analyzing the output from multiple runs in [14]. We only employ the simplified version given by Gelman in [13] which does not correct for the uncertainty introduced by estimating the posterior means and variances:

Definition 2.12 *Consider K sequences of estimators for the parameter $\theta \in \mathbb{R}^d$, each of length L. Denote them by $\{\theta^{k,l}\}_l$, where $k \in \{1, \ldots, K\}$ and $l \in \{1, \ldots, L\}$.*

The **between-sequence variance** *for the i-th component of θ is*

$$B^i := \frac{L}{K-1} \sum_{k=1}^{K} (\overline{(\theta_i^{k,l})}_L - \overline{(\theta_i^{k,l})}_{K,L})^2 \tag{2.90}$$

Here, the first average is taken over the L values of each sequence, whereas the second is take over all values and sequences.

The **within-sequence variance** *is*

$$W^i := \frac{1}{K(L-1)} \sum_{k=1}^{K} \sum_{l=1}^{L} (\theta_i^{k,l} - \overline{(\theta_i^{k,l})}_L)^2 \tag{2.91}$$

Together, these yield the estimated **potential scale reduction**

$$\mathrm{PSR}_i := \sqrt{\frac{(L-1)W^i + B^i}{LW^i}} \tag{2.92}$$

The values B^i and W^i are unbiased estimates for the posterior variance of θ_i under stationarity. If we generate starting points from a distribution which is overdispersed compared to the posterior, B^i is an overestimate. On the other hand, W^i is an underestimate, as no single chain will have explored the entire parameter space after finitely many iterations. Therefore, the posterior scale reduction is an upper bound on the relative reduction of the standard deviation for $\overline{(\theta_i^{k,l})}_{K,L}$ that could be attained if the chain was allowed to run longer. Gelman suggests that a potential scale reduction of 1.2 for all quantities of interest should be taken to indicate convergence.

The book by Chen et. al. [7] contains an overview of several tools for analyzing the convergence behavior of a chain, including the posterior scale reduction criterion.

Chapter 3

The lognormal rainfall model

The rainfall model introduced in this section is an attempt to capture the behavior of precipitation data at a single point (rain-gauge readings or radar-data for a small area) over time. The observations consist of the amount of rainfall recorded during periods of fixed length. This results in a distribution with a continuous component on $\mathbb{R}_{>0}$ and a point mass at 0.

Unfortunately, many properties of the data depend strongly on the length of the observation interval. For example, observations recorded at 10 minute intervals have a high probability to be 0, while the point mass may be negligible for monthly data. We follow the approach of Rodriguez-Iturbe et. al. [37] and try to address this problem by disaggregation. The rainfall intensity is treated as a continuous function over time, while the observations are the aggregated amounts per time interval. Thus, the intensity process is formally independent of the length of the observation periods. Of course, the intensities are estimated from the observations, so the precision of the estimator depends on the time scale.

It is well known that the duration and intensity of rainstorms are negatively correlated. See e.g. the article by Bacchi et. al. [2] for a model focusing on this property. We also want to be able to incorporate short-range dependencies of this kind. On the other hand, there seem to be no meaningful long-term dependencies except for seasonal effects. These are taken into account only in so far that model is fitted to data from a single season.

3.1   The rainfall process

The model treats rainfall intensity as a step function. Rain is assumed to occur in cells of constant intensity. These may overlap to produce heavier precipitation. Cells are grouped into storms, i.e. periods during which rainfall may occur. The following definition of a rainfall process is based on the model proposed by Rodriguez-Iturbe et. al. in [37] and [38]:

Definition 3.1 *A **rainfall process** Π is characterized by*

- *The time scale $\tau > 0$ [min].*

- *The number $T \geq 1$ of observation periods of length τ.*

- *The number $M \geq 1$ of storms.*

- *The duration $X_1^i > 0$ [min] of storm i, $i \in \{1, \ldots, M\}$.*

- *The lag $X_2^i > 0$ [min]* **between the origins** *of storms i and $i + 1$, $i \in \{1, \ldots, M - 1\}$. Storm M also has an associated lag X_2^M, but we require that it places the origin of the (unobserved) successor after the end of the observation period.*

- *The number $M^i \geq 1$ of cells in storm i, $i \in \{1, \ldots, M\}$.*

- *The duration $X_1^{i,j} > 0$ [min] of cell j in storm i, $j \in \{1, \ldots, M^i\}$.*

- *The lag $X_2^{i,j} > 0$ [min]* **between the origins** *of cells j and $j + 1$ in storm i, $j \in \{1, \ldots, M^i - 1\}$. The last cell of each storm is assigned a lag X_2^{i,M^i} that would place its (non-existent) successor after the end of the storm.*

- *The depth (intensity) $D^{i,j} > 0$ [mm/min] of cell j in storm i, $j \in \{1, \ldots, M^i\}$.*

From this, we derive the **storm and cell origins**

$$O^i := \sum_{l=1}^{i-1} X_2^l \qquad\qquad O^{i,j} := O^i + \sum_{l=1}^{j-1} X_2^{i,l} \qquad (3.1)$$

and the **rainfall aggregates** *for $t \in \{1, \ldots, t\}$*

$$Y_t := f_t(\{X^i\}, \{X^{i,j}\}, \{D^{i,j}\}) := \sum_{i=1}^{M} \sum_{j=1}^{M^i} |[O^{i,j}, O^{i,j} + X_1^{i,j}] \cap [(t-1)\tau, t\tau]| D^{i,j} \qquad (3.2)$$

We write

$$X^i := \begin{pmatrix} X_1^i \\ X_2^i \end{pmatrix} \qquad X^{i,j} := \begin{pmatrix} X_1^{i,j} \\ X_2^{i,j} \end{pmatrix} \qquad \mathbf{X} := \{(X^i, \{X^{i,j}\})\} \qquad (3.3)$$

$$\mathbf{M} := \{M^i\} \qquad \mathbf{D} := \{D^{i,j}\} \qquad \mathbf{O} := \{(O^i, \{O^{i,j}\})\} \qquad \mathbf{Y} := \{Y_t\}$$

and identify the rainfall process by $\Pi = (\mathbf{Y}, \mathbf{D}, \mathbf{O}, \mathbf{X}, \mathbf{M}, M, T, \tau)$. It is also convenient to denote the total number of cells by

$$\overline{M} := \sum_{i=1}^{M} M^i \qquad (3.4)$$

A rainfall process is said to be **consistent** *iff*

$$O^M \leq T\tau < O^M + X_2^M \qquad (3.5)$$

$$\forall i: \ O^{i,M^i} \leq O^i + X_1^i < O^{i,M^i} + X_2^{i,M^i}$$

Figure 3.1: Storm process and aggregation
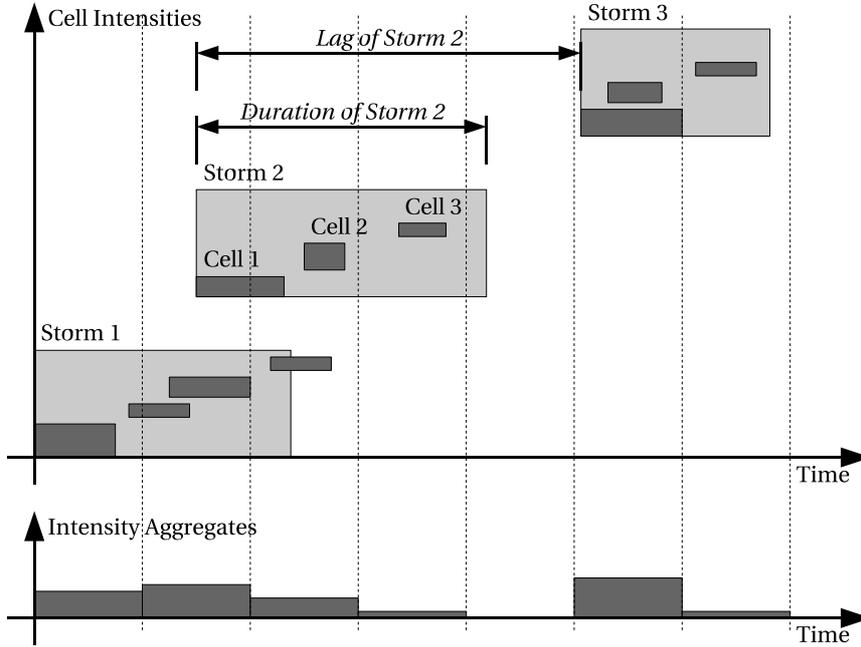
Remark 3.1 *1. The first storm is assumed to start at time $t = 0$, while the first cell of each storm starts at the storm's origin. Thus, the storm origins and the cell origins within each storm form a non-general point process.*

2. The consistency condition ensures that M and M^i actually are the correct number of storms and cells. Prospective successors are placed beyond the end of the observation period or storm. In terms of simulating a storm process, the condition actually defines M and M^i:

To generate a single storm, we first sample the duration X_1^1 and then generate a sequence of cell lags $X_2^{1,1}$, $X_2^{1,2}$, etc. until we reach an index j_0 satisfying

$$O^{1,j_0} \leq O^1 + X_1^1 < O^{1,j_0} + X_2^{1,j_0} \tag{3.6}$$

By setting M^1 equal to j_0, we get a consistent cell process for storm 1.

Figure 3.1 shows a storm process and the resulting aggregates. For a consistent process, the storms mark periods where cells may start, whereas the cells contribute the actual precipitation. A cell may last longer than the storm it belongs to, and cells and storms may freely overlap. Nesting cell processes within the storm process leads to a clustering of rain cells, which appears to describe the behavior of actual rainfall reasonably well [38].

A wide range of precipitation models can be fitted within the framework defined above. In [37], the components of **X** and **D** are treated as independent exponentials. The major advantage of this model is that the distribution of the storm and cell origins is easily tractable,

as the number of events in a given period is Poisson distributed. Unfortunately, there exists no multivariate extension of the exponential distribution that allows for a straightforward specification of a covariance structure. Some suggestions how to model correlation in this framework are given in [2], in [38], or in the technical report by Granville and Smith [21]. The latter is of particular interest to us, since they also use MCMC methods for estimation.

We want to use a different approach and specify a multivariate lognormal distribution for the elements of $\mathbf{X}$. More specifically, we assume for storm i and cell j

$$\log \begin{pmatrix} X_1^i \\ X_2^i \\ X_1^{i,j} \\ X_2^{i,j} \end{pmatrix} \Bigg| \mu, \Lambda \sim \mathrm{N}_4(\mu, \Lambda) \tag{3.7}$$

with mean $\mu = (\mu_1, \ldots, \mu_4)^T \in \mathbb{R}^4$ and $\Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12}^T & \Lambda_{22} \end{pmatrix} \in \mathbb{S}_+^4$, partitioned into submatrices of size 2×2. The storm quantities X^i are treated as i.i.d. with distribution

$$\log \begin{pmatrix} X_1^i \\ X_2^i \end{pmatrix} \Bigg| \mu, \Lambda \sim \mathrm{N}_2(\check{\mu}, \check{\Lambda}_{11}) \tag{3.8}$$

The parameters describing the marginal distribution of the first two components can be derived from (2.7):

$$\check{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad\qquad \check{\Lambda}_{11} = \Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{12}^T \tag{3.9}$$

The cell quantities are i.i.d. conditional on the duration and lag of their storm:

$$\log \begin{pmatrix} X_1^{i,j} \\ X_2^{i,j} \end{pmatrix} \Bigg| X^i, \mu, \Lambda \sim \mathrm{N}_2(\check{\mu}^i, \Lambda_{22}) \tag{3.10}$$

According to (2.8), the conditional mean is

$$\check{\mu}^i = \begin{pmatrix} \mu_3 \\ \mu_4 \end{pmatrix} - \Lambda_{22}^{-1}\Lambda_{12}^T \begin{pmatrix} \log(X_1^i) - \mu_1 \\ \log(X_2^i) - \mu_2 \end{pmatrix} \tag{3.11}$$

Cell depths are not included in the lognormal framework, but sampled from a uniform distribution independent of all other quantities:

$$D^{i,j} \sim \mathrm{U}(0, 2\iota) \tag{3.12}$$

where ι is some intensity scale [mm/min].

The reason for treating cell depth independently of the temporal structure is identifiability. As long as a rain cell is contained in a single observation period, its contribution to that period's total is equal to its duration times its depth. However, the product of two lognormals is again

lognormally distributed. I.e., a joint lognormal model could not distinguish between depth and duration in this case, and would use five parameters (two means, three (co)precisions) to describe a quantity that can be fully characterized by two. While these considerations suggest a constant intensity, variable cell depths result in lower rejection rates for storm structure transforms in the MCMC algorithm and lead to better mixing for the sampler. The uniform distribution was chosen for ease of use.

Remark 3.2 *There is some evidence in real data sets that the rain rate averaged over a sufficiently large area is approximately lognormally distributed conditional on the presence of rain. A theoretical justification for the lognormality of rain rate can be found in the works of Kedem and Chiu [25] as well as Pavlopoulos and Kedem [33].*

Our model does not reproduce this exactly, as the rainfall amount in a period is equal to the sum of several lognormals multiplied with uniforms. Nevertheless, we expect to obtain a distribution with a similar extremal behavior.

We denote the actually observed rainfall amounts by Z_t, $t \in \{1, \ldots, T\}$, and treat them as a distorted version of the aggregates Y_t. The error distribution is assumed to be unsystematic, i.e. $E(Z_t | Y_t) = Y_t$. It has to account for both the errors of measurement and the errors introduced by the simplifying assumptions of the model. No attempt is made to distinguish between the two. The Z_t are assumed to be 0 iff $Y_t = 0$ and independent lognormals with unknown precision scale $\zeta > 0$ else:

$$\log(Z_t) | Y_t, Y_t > 0 \sim \mathrm{N}\left(\log(Y_t) - \frac{1}{2p(\zeta, \tau)}, p(\zeta, \tau)\right) \tag{3.13}$$

The term $p(\zeta, \tau)$ is defined as

$$p(\zeta, \tau) := \left(\log\left(\frac{\tau}{\zeta} + 1\right)\right)^{-1} \tag{3.14}$$

which yields

$$\mathrm{E}(Z_t | Y_t) = Y_t \qquad\qquad \mathrm{Var}(Z_t | Y_t) = Y_t^2 \frac{\tau}{\zeta} \tag{3.15}$$

As it is easier to handle, the estimation algorithm consistently parametrizes the model in terms of $p(\zeta, \tau)$ and retrieves ζ from the output.

Remark 3.3 *For rain gauge readings, the error of measurement is higher during periods of intense precipitation. The distribution of Z_t was chosen to take this into account, i.e. the standard deviation is proportional to the true rainfall amounts. In other words, the relative standard deviation is constant for $Y_t > 0$:*

$$\frac{\mathrm{Std}(Z_t | Y_t)}{Y_t} = \sqrt{\frac{\tau}{\zeta}} \tag{3.16}$$

Note also that the value of ζ does not depend on the time scale τ. For example, if we transform the observations to a timescale of 2τ by adding pairs of observations, we get

$$\mathrm{Var}\left(\frac{Z_{2t-1}}{Y_{2t-1}} + \frac{Z_{2t}}{Y_{2t}} \,\middle|\, Y_{2t-1}, Y_{2t}, Y_{2t-1} > 0, Y_{2t} > 0\right) = \frac{2\tau}{\zeta} \tag{3.17}$$

Table 3.1: Model variants

Model	Description	$\dim(\theta)$
I	The full model as per Definition 3.2.	15
	(parameters $\mu \in \mathbb{R}^4$, $\Lambda \in \mathbb{S}_+^4$, $\zeta \in \mathbb{R}$)	
II	Λ is treated as a diagonal matrix.	9
	(parameters $\mu \in \mathbb{R}^4$, $\lambda \in \mathbb{R}_+^4$, $\zeta \in \mathbb{R}$)	
III	Means μ_1 and μ_4 are matched via typical number of cells K.	11
	(parameters $\mu_R \in \mathbb{R}^3$, $\Lambda_R \in \mathbb{S}_+^3$, $\Lambda_{4,4} \in \mathbb{R}_+$, $\zeta \in \mathbb{R}$)	
IV	As model III, with diagonal Λ.	8
	(parameters $\mu_R \in \mathbb{R}^3$, $\lambda \in \mathbb{R}_+^4$, $\zeta \in \mathbb{R}$)	
V	Cell durations are uncorrelated with the other quantities.	12
	(parameters $\mu \in \mathbb{R}^4$, $\Lambda_R \in \mathbb{S}_+^3$, $\Lambda_{3,3} \in \mathbb{R}^+$, $\zeta \in \mathbb{R}$)	
VI	Only storm duration and cell lags have non-zero correlation.	10
	(parameters $\mu \in \mathbb{R}^4$, $\Lambda_{1,1}$ to $\Lambda_{4,4}$, $\Lambda_{1,4}$, $\zeta \in \mathbb{R}$)	

As real world data frequently contain missing observations, we include that possibility in our model, denoting the set of unknown measurements by $U \subset \{1, \dots, T\}$. These need to be estimated alongside the rainfall process and its governing parameters.

We can now formulate the complete model as

Definition 3.2 *The **lognormal rainfall model** comprises*

-   *a consistent rainfall process $\Pi = (\mathbf{Y}, \mathbf{O}, \mathbf{D}, \mathbf{X}, \mathbf{M}, M, T, \tau)$ with average cell intensity ι.*

-   *observations $\mathbf{Z} := \{Z_t\}$, where $\{Z_t\}_{t \in U}$ are missing values.*

-   *parameters $\theta := (\mu, \Lambda, p(\zeta, \tau))$.*

The interaction of these quantities is determined by Definition 3.1, as well as Equations (3.7), (3.12), and (3.13).

Fitting this model to a data set requires that we estimate $(\mathbf{D}, \mathbf{X}, \mathbf{M}, M)$ alongside the parameters θ. We show how to do this using MCMC methods in Chapter 4.

3.2   Simplifying the model

The relatively high dimension of the parameter makes it difficult to obtain a good fit via the MH-algorithm in reasonable time. Thus, we want to look at some simplifications that reduce

the dimension of the parameter while preserving the general framework. These variations are identified by Roman numerals, with model I being the full model (see Table 3.1):

3.2.1 Model II: Uncorrelated quantities

An obvious way to simplify the model is to assume uncorrelated storm and cell quantities $\mathbf{X}$, i.e. to treat Λ as a diagonal matrix. In this context, we write $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_4)$. Since we choose the lognormal model precisely because we want to detect correlations, this reduction has no practical value. However, it is useful for testing the estimation algorithm, as it simplifies many calculations.

3.2.2 Models III and IV: Matched means

Numerical trials show that the algorithm is unable to accurately identify the parameters associated with the cell process if the average duration of the cells is less than τ. Depending on the starting point, we may get parameter sets favoring a few long or many short cells. This is especially problematic as the computational effort increases with the number of cells in the estimated process.

We can address this by introducing additional dependencies among the parameters. If possible, the number of cells per storm should remain reasonably small. Our approach is to fix a **typical number of cells per storm** K and treat the mean parameter for the cell lag as a function of those for storm duration. Let

$$\sigma_1^2 := \text{Var}(\log(X_1^i)|\theta) = (\Lambda^{-1})_{1,1} \qquad \sigma_4^2 := \text{Var}(\log(X_2^{i,j})|\theta) = (\Lambda^{-1})_{4,4} \qquad (3.18)$$

The number K is considered to be 'typical' in the sense that we match the means

$$\text{E}(X_1^i|\theta) = \text{E}\left(\sum_{j=1}^{K} X_2^{i,j} \,\middle|\, \theta\right) \qquad (3.19)$$

If the duration of a storm and the sum of K cell lags have the same mean, we expect that 'most' storms have close to K cells. Note that we do not condition $X_2^{i,j}$ on X^i, as we use the additional assumption that the lags are uncorrelated with the other elements of $\mathbf{X}$. In particular, variations in rainfall intensity between storms are attributed uniquely to cell duration.

Matching the means yields the equation

$$\exp\left(\mu_1 + \frac{\sigma_1^2}{2}\right) = K \exp\left(\mu_4 + \frac{\sigma_4^2}{2}\right) \qquad (3.20)$$

which is solved by

$$\mu_4(\mu_1, \sigma_1^2, \sigma_4^2) := \mu_1 + \frac{\sigma_1^2 - \sigma_4^2}{2} - \log(K) \qquad (3.21)$$

The model is now parametrized in terms of $\mu_R \in \mathbb{R}^3$, $\Lambda_R \in \mathbb{S}_+^3$, and $\Lambda_{4,4} > 0$ which satisfy

$$\mu = \begin{pmatrix} \mu_R \\ \mu_4(\mu_1, \sigma_1^2, \sigma_4^2) \end{pmatrix} \qquad\qquad \Lambda = \begin{pmatrix} \Lambda_R & \underline{0} \\ \underline{0} & \Lambda_{4,4} \end{pmatrix} \tag{3.22}$$

This approach reduces the dimension of the parameter vector $\theta_R := (\mu_R, \Lambda_R, p(\zeta, \tau))$ from 15 to 11. However, with K, we incur one additional parameter which has to be specified in advance.

Like model II, model IV is merely a test case where we use mean matching in conjunction with diagonal Λ.

Remark 3.4 *While it is possible to match both the means and variances of X_1^i and $\sum_{j=1}^K X_2^{i,j}$, this approach did not perform well in numerical trials.*

As for choosing a typical number of cells, we use $K = 10$ to generate artificial data with a reasonably small total number of cells. Deciding on the correct value for real data is part of the problem of choosing a starting parameter we treat in Chapter 5.

3.2.3 Model V: Uncorrelated cell durations

The matched means approach is attractive from a computational point of view as it keeps the number of cells small, but it may not be a good representation of the real process. A more natural way to model a higher rainfall intensity for short storms would be to increase the frequency of cells instead of their duration. The difference between the two descriptions is negligible only if the average cell duration is much shorter than τ. Thus, we also want to investigate the properties of an alternative model, where the cell durations are independent of the remaining quantities, while the lags are not. The precision matrix becomes

$$\Lambda = \begin{pmatrix} \Lambda_{1,1} & \Lambda_{1,2} & 0 & \Lambda_{1,4} \\ \Lambda_{1,2} & \Lambda_{2,2} & 0 & \Lambda_{2,4} \\ 0 & 0 & \Lambda_{3,3} & 0 \\ \Lambda_{1,4} & \Lambda_{2,4} & 0 & \Lambda_{4,4} \end{pmatrix} \tag{3.23}$$

In the context of model IV, we also write

$$\Lambda_R = \begin{pmatrix} \Lambda_{1,1} & \Lambda_{1,2} & \Lambda_{1,4} \\ \Lambda_{1,2} & \Lambda_{2,2} & \Lambda_{2,4} \\ \Lambda_{1,4} & \Lambda_{2,4} & \Lambda_{4,4} \end{pmatrix} \qquad\qquad \mu_R = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_4 \end{pmatrix} \tag{3.24}$$

No additional dependencies are introduced for the means, which makes this model easier to implement than mean matching, but it requires 12 parameters instead of 11.

3.2.4 Model VI: Limited correlation

The last model we consider uses but a single non-zero correlation parameter for storm duration and cell lags:

$$\Lambda = \begin{pmatrix} \Lambda_{1,1} & 0 & 0 & \Lambda_{1,4} \\ 0 & \Lambda_{2,2} & 0 & 0 \\ 0 & 0 & \Lambda_{3,3} & 0 \\ \Lambda_{1,4} & 0 & 0 & \Lambda_{4,4} \end{pmatrix} \tag{3.25}$$

For model V, we write

$$\Lambda_R = \begin{pmatrix} \Lambda_{1,1} & \Lambda_{1,4} \\ \Lambda_{1,4} & \Lambda_{4,4} \end{pmatrix} \qquad\qquad \mu_R = \begin{pmatrix} \mu_1 \\ \mu_4 \end{pmatrix} \tag{3.26}$$

With 10 parameters, this is the smallest variant of the model which is non-trivial in the sense that it can reproduce the effect of a higher rainfall intensity for short storms.

3.3 Prior distributions

To carry out estimation within the Bayesian framework of the MH algorithm, we need to specify suitable prior distributions for the components of θ. Since we are working with the lognormal distribution, we use the normal-Wishart conjugate prior of Proposition 2.2 for the mean and precision of the logarithmic quantities.

In the full model I, we choose independent components for μ:

$$\mu_k \sim \mathrm{N}(0, r) \tag{3.27}$$

The precision matrix Λ characterizing $\mathbf{X}$ is distributed as

$$\Lambda \sim \mathrm{Wishart}_4(2.5, \sqrt{s}I_4) \tag{3.28}$$

The reason for this choice of parameters is that $\alpha = 2.5$ eliminates the determinant of Λ from the density (see 2.14), while $\beta = \sqrt{s}I_4$ means that the precision of the diagonal elements is proportional to s. In the one-dimensional case, the resulting gamma distribution is usually parametrized via $\alpha = \beta = s$, which yields expectation 1 and precision s. Unfortunately, this choice cannot be extended to our situation, as the four-dimensional Wishart distribution is only defined for $\alpha > 1.5$.

Finally, the error precision $p(\zeta, \tau)$ for the observations is assumed to satisfy

$$p(\zeta, \tau) \sim \mathrm{Gamma}(s, s) \tag{3.29}$$

The prior precisions r and s need to be chosen sufficiently small to allow for a wide range of likely parameters.

3.3.1   Priors for the simplified models

The simplified models use the same distribution (3.27) for the means as model I, except that μ_4 in models II and IV is given by the deterministic relationship (3.21) and does not require a separate prior.

In models II and IV where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_4)$, the precisions are are taken to be independently gamma-distributed as

$$\lambda_k \sim \mathrm{Gamma}(s, s) \tag{3.30}$$

Models III, V, and VI also use this prior for the precision of those components which are uncorrelated with the rest, while the reduced precision matrix Λ_R is Wishart:

$$\Lambda_R \sim \mathrm{Wishart}_l \left(\frac{l+1}{2}, \sqrt{s}I_l \right) \tag{3.31}$$

where $l = 3$ for model III or V, and $l = 2$ for model VI.

3.4   The full probability model

Taken together, Definition 3.2 and the prior distributions of the previous section define a probabilistic model suitable for Bayesian inference. It can be decomposed as

$$
\begin{aligned}
&p_{\mathbf{Z},\mathbf{Y},\mathbf{X},\mathbf{M},M,\theta}(\bullet) \\
&= p_{\mathbf{Z}}(\bullet|\mathbf{Y},\theta)\mathbb{1}_{f(\mathbf{D},\mathbf{X})}(\mathbf{Y})p_{\mathbf{D}}(\bullet|\mathbf{M},M)p_{\mathbf{X}}(\bullet|\mathbf{M},M,\theta)p_{\mathbf{M}}(\bullet|M,\theta)p_M(\bullet|\theta)p_\theta(\bullet)
\end{aligned} \tag{3.32}
$$

The factors are determined by our previous choice of distributions. For the components with density, we obtain

$$
\begin{aligned}
p_{\mathbf{Z}}(\{z_t\}|\mathbf{Y} = \{y_t\},\theta) &= \left(\prod_{t:y_t=0} \mathbb{1}_{\{0\}}(z_t) \right) \left(\prod_{t:y_t>0} \frac{1}{z_t} \left(\frac{p(\zeta,\tau)}{2\pi} \right)^{\frac{1}{2}} \right) \\
&\times \exp\left(-\frac{p(\zeta,\tau)}{2} \sum_{t:y_t>0} \left(\log(z_t) - \log(y_t) + \frac{1}{2p(\zeta,\tau)} \right)^2 \right)
\end{aligned} \tag{3.33}
$$

$$
p_{\mathbf{D}}(\{d^{i,j}\}|\mathbf{M},M) = \prod_{i=1}^{M}\prod_{j=1}^{M^i} \frac{\mathbb{1}_{[0,2\iota]}(d^{i,j})}{2\iota} \tag{3.34}
$$

$$p_{\mathbf{X}}(\{(x^i, \{x^{i,j}\})\}|\mathbf{M}, M, \theta) = \prod_{i=1}^{M} p_{X^i}(x^i|\theta) \prod_{j=1}^{M^i} p_{X^{i,j}}(x^{i,j}|X^i = x^i, \theta) \tag{3.35}$$

$$= \left(\frac{|\check{\Lambda}_{11}|}{(2\pi)^2}\right)^{\frac{M}{2}} \left(\frac{|\Lambda_{22}|}{(2\pi)^2}\right)^{\frac{\overline{M}}{2}} \left(\prod_{i=1}^{M} \frac{1}{x_1^i x_2^i}\right) \left(\prod_{i=1}^{M} \prod_{j=1}^{M^i} \frac{1}{x_1^{i,j} x_2^{i,j}}\right)$$

$$\times \exp\left(-\frac{1}{2} \sum_{i=1}^{M} (\log(x^i) - \check{\mu})^T \check{\Lambda}_{11} (\log(x^i) - \check{\mu})\right)$$

$$\times \exp\left(-\frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M^i} (\log(x^{i,j}) - \check{\mu}^i)^T \Lambda_{22} (\log(x^{i,j}) - \check{\mu}^i)\right)$$

$$p_\theta(x, A, q) = p_\mu(x) p_\Lambda(A) p_{p(\zeta,\tau)}(q) \tag{3.36}$$

$$= \frac{r^2 s^{5+s} q^{s-1}}{(2\pi)^2 \Gamma_4(2.5) \Gamma(s)} \exp\left(-\frac{r}{2} \sum_{k=1}^{4} x_i^2 - \sqrt{s}\,\mathrm{tr}(A) - sq\right)$$

These expressions correspond to the choice of (log)normal distributions for the Z_t in (3.13), $(X^i, X^{i,j})$ in (3.7), and μ in (3.27), the choice of a uniform distribution for the $D^{i,j}$ in (3.12), and the choice of Wishart resp. Gamma distributions for Λ in (3.28) and $p(\zeta, \tau)$ in (3.29). The general forms and relevant properties of the appropriate distributional families are summarized in Section 2.3.

Unfortunately, there is no closed-form solution for the weights of $\mathbf{M}$ and M. We show how they can be estimated by Monte-Carlo simulation in Section 4.4 from relationships based on the consistency condition (3.5):

$$p_M(m|\theta) = P\left(\sum_{i=1}^{m-1} X_2^i \le T\tau < \sum_{i=1}^{m} X_2^i \,\middle|\, \theta\right) \tag{3.37}$$

$$p_{\mathbf{M}}(\{m^i\}|M, \theta) = \prod_{i=1}^{M} P\left(\sum_{j=1}^{m^i-1} X_2^{i,j} \le X_1^i < \sum_{i=1}^{m^i} X_2^{i,j} \,\middle|\, \theta\right)$$

Finally, the indicator function $\mathbb{1}_{f(\mathbf{X},\mathbf{D})}(\mathbf{Y})$ enforces the deterministic relation $Y_t = f_t(\mathbf{X}, \mathbf{D})$ in the sense of (3.2).

For the reduced models, the expressions in (3.35) and (3.36) simplify in a canonical manner, as one or more entries of Λ are fixed to 0. Likewise, the matched means used in models III and IV are easily included in the model likelihood: the mean vector reduces to three dimensions and μ_4 is replaced by its representation in (3.21). Note that the latter change means that the posterior distributions of μ and Λ conditional on $\mathbf{X}$ become more complicated, but this is discussed in Chapter 4.

A representation of the model as a directed acyclic graph (DAG) is shown in Figure 3.2. It includes the fixed parameters ι, r, and s. The (redundant) origins $\mathbf{O}$ and the number of periods T are missing, as their role is obvious. The graph can be read as follows:

- squares represent known quantities, including those that have to be chosen in advance.

- the bold square represents the observations.

- circles represent unknown quantities.

- the dashed box contains the parameters of interest.

- shadowed boxes represent multiple, independent copies of the same structure.

- bold lines represent stochastic dependencies.

- thin lines represent deterministic dependencies.

A motivation for expressing stochastic beliefs in directed graphs can be found in the article by Lauritzen and Spiegelhalter [27], while [18] contains several examples used for developing MCMC models. A DAG allows us to quickly identify conditional dependencies, e.g. it can be seen from Figure 3.2 that $\mathbf{Z}$ is conditionally independent of $\mathbf{X}$ given $\mathbf{Y}$.

Figure 3.2: The model as a directed acyclic graph
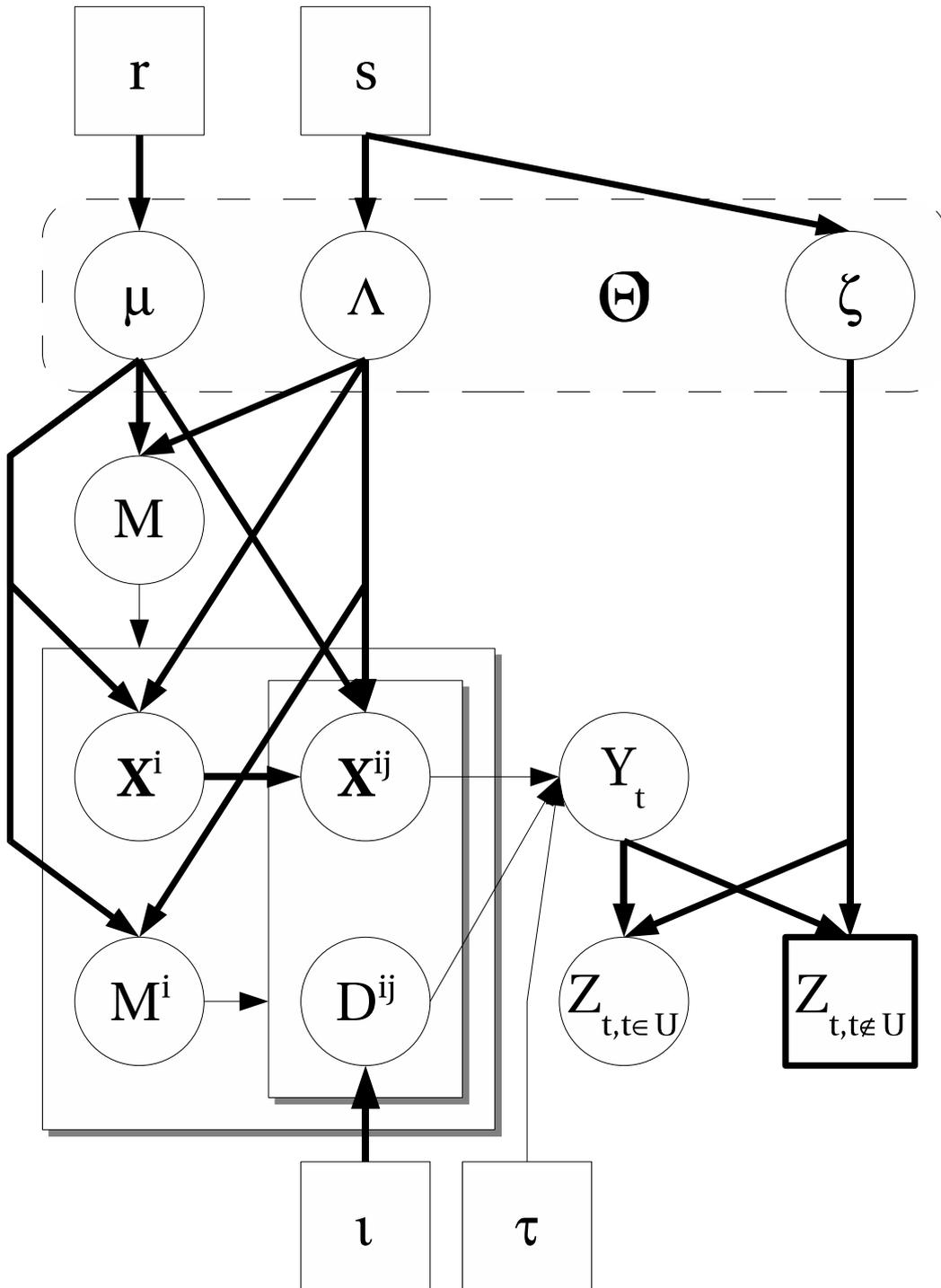
Chapter 4

Parameter estimation using Markov-Chain Monte-Carlo

In this chapter, we formulate an MCMC-algorithm for estimating the parameter θ and the unknown quantities $(\mathbf{D}, \mathbf{X}, \mathbf{M}, M)$ of the lognormal rainfall model introduced in Chapter 3. This involves choosing a starting point, specifying an update scheme, and deciding on how to sample from the conditional distributions that arise.

4.1   Starting values

The algorithm needs starting values for θ, M, $\mathbf{M}$, $\mathbf{X}$, and $\mathbf{D}$. At least in theory, θ can be chosen arbitrarily, although a gross misspecification will result in slow convergence or prevent convergence due to numerical issues. In Chapter 5, we propose a heuristic for a reasonable choice based on properties of the data set.

The requirements for the remaining quantities are more strict, as they need to be consistent with the observations. We use the following scheme to generate a process that "explains" $\mathbf{Z}$:

- Each uninterrupted sequence of non-zero observations forms a single storm. A storm starts at the beginning of the first time period it covers and stops at the end of the last period. The lag of a storm to its successor is equal to the difference of origins, or to the difference to $T\tau + 1$ for the last storm.

  To bring the process in line with the heuristic for choosing the first θ, later versions of the algorithm allow gaps of zero precipitation within the initial storms. A storm is terminated only by a dry period exceeding a certain length.

- Unknown observations are assigned a single cell. Without this 'seeding', the algorithm can take a long time before it is able to place storms in the unobserved period.

- If time period t has non-zero precipitation, it is assigned $\lceil \frac{2Z_t}{\iota\tau} \rceil$ cells. This is the average number of cells of length $\frac{\tau}{2}$ required to generate the observed precipitation. For computational reasons, this number can be capped to avoid excessively many cells in the initial process.

- Cell depths are drawn from a $U(0, 2\iota)$ distribution per (3.12). Cell origins are uniformly distributed in the first half of the observation period they belong to and end at a time point uniformly distributed on the second half. This yields an average cell length of $\frac{\tau}{2}$.

  Exception: The first cell of each storm starts at its origin and has a duration drawn from $U(0, \tau)$. This is necessary to satisfy the assumption that $O^i = O^{i,1}$ for all storms i.

- All cells belonging to a storm are sorted in ascending order of their origins, which determines their lags. The last cell of each storm gets a lag that would place the next cell 1 minute after the storm ends.

Remark 4.1   *1. The model assumes that the first observation period has non-zero precipitation, so any initial zero observations need to be dropped.*

2. The cell durations are randomized to avoid a degenerate process with sample variance 0. The same could happen for the storm durations if all wet periods have the same length, but this is highly unlikely.

The process generated in this manner is consistent but tends to be highly atypical for the lognormal model. To obtain a more regular state, we fix θ during the first few iterations of the algorithm and only transform the storm structure. Without this measure, the starting parameter would have virtually no impact on the process.

4.2   General outline of the estimation algorithm

The storm process and observations are estimated using an MH-type algorithm as introduced in Section 2.5. During each iteration of the scheme, we perform two major update steps:

1. Sample the parameters θ conditional on the storm process and observations.

2. Transform the storm process (and resample missing observations) conditional on θ and the observations.

These steps are further divided into subsamplers operating on suitable blocks of unknown values, several of which admit a Gibbs sampler (see Section 2.5.1.1). The decomposition of the full model likelihood into conditional distributions (3.32) plays a major role in identifying such blocks and calculating the acceptance probability for the MH-proposals. Estimators for θ and $\mathbf{Y}$ are obtained by averaging over the values in the chain after a suitable burn-in period.

4.3 Parameter updates

The update schemes for μ and Λ depend on the variant of the model we want to use. Fixing some or all of the correlations to 0 simplifies the calculations considerably, whereas the matched means of Section 3.2.2 introduce additional terms. The sampler for ζ stays the same in all cases, as it depends only on $\mathbf{Y}$ and $\mathbf{Z}$.

4.3.1 Model II: Sampling μ and diagonal Λ

We start with Model II, as it is the most simple. For $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_4)$, all parameter updates can be done via Gibbs sampling, as we are dealing with ordinary normal-gamma priors (see Section 2.3.4).

The posterior distributions for the parameters describing storm behavior ($k \in \{1, 2\}$) are

$$\mu_k | \{X_k^i\}, M, \lambda_k \sim \mathrm{N} \left(\frac{\lambda_k \sum_{i=1}^M \log(X_k^i)}{r + M\lambda_k}, r + M\lambda_k \right)$$

(4.1)

$$\lambda_k | \{X_k^i\}, M, \mu_k \sim \mathrm{Gamma} \left(s + \frac{M}{2}, s + \frac{1}{2} \sum_{i=1}^M (\log(X_k^i) - \mu_k)^2 \right)$$

For the cell parameters ($k \in \{3, 4\}$), we get

$$\mu_k | \{X_k^{i,j}\}, \mathbf{M}, M, \lambda_k \sim \mathrm{N} \left(\frac{\lambda_k \sum_{i=1}^M \sum_{j=1}^{M^i} \log(X_{k-2}^{i,j})}{r + \overline{M}\lambda_k}, r + \overline{M}\lambda_k \right)$$

(4.2)

$$\lambda_k | \{X_k^{i,j}\}, \mathbf{M}, M, \mu_k \sim \mathrm{Gamma} \left(s + \frac{\overline{M}}{2}, s + \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^{M^i} (\log(X_{k-2}^{i,j}) - \mu_k)^2 \right)$$

One iteration of the Gibbs sampler for μ and Λ simply consists of drawing a new value for each component from these distributions.

4.3.2 Model I: Sampling μ and Λ in the full model

For the full model, the posterior distributions become more involved. The mean vector μ is still normally distributed and permits a Gibbs sampler, but the precision matrix Λ is only approximately Wishart. This is due to the imbalance in the model, where each vector of storm quantities X^i affects the distribution of several cell quantities $X^{i,j}$.

To express the distributions that arise, it is convenient to introduce the following expressions:

$$\xi^{i,j} := \begin{pmatrix} \log(X^i) \\ \log(X^{i,j}) \end{pmatrix} \qquad \check{\xi}^i := \begin{pmatrix} \xi_1^{i,1} \\ \xi_2^{i,1} \end{pmatrix} \qquad \hat{\xi}^{i,j} := \begin{pmatrix} \xi_3^{i,j} \\ \xi_4^{i,j} \end{pmatrix}$$

$$\delta^{i,j} := \xi^{i,j} - \mu \qquad \check{\delta}^i := \begin{pmatrix} \delta_1^{i,1} \\ \delta_2^{i,1} \end{pmatrix} \qquad \hat{\delta}^{i,j} := \begin{pmatrix} \delta_3^{i,j} \\ \delta_4^{i,j} \end{pmatrix} \qquad (4.3)$$

$$\hat{\delta}^i := \begin{pmatrix} \delta_1^{i,1} \\ \delta_2^{i,1} \\ 0 \\ 0 \end{pmatrix} \qquad \hat{\delta}^{i,j} := \begin{pmatrix} 0 \\ 0 \\ \delta_3^{i,j} \\ \delta_4^{i,j} \end{pmatrix}$$

$$Q_1 := \begin{pmatrix} \check{\Lambda}_{11} & \underline{0} \\ \underline{0} & \underline{0} \end{pmatrix} \in \mathbb{R}^{4\times 4} \qquad Q_2 := Q_3^T \Lambda_{22} Q_3 = \begin{pmatrix} \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{12}^T & \Lambda_{12} \\ \Lambda_{12}^T & \Lambda_{22} \end{pmatrix} \qquad (4.4)$$

$$Q_3 := (\Lambda_{22}^{-1}\Lambda_{12}^T, I_2)$$

Note that

$$Q_1 = \Lambda - Q_2 \qquad\qquad Q_3(\xi^{i,j} - \mu) = \check{\xi}^{i,j} - \check{\mu}^i \qquad (4.5)$$

In particular, Q_3 multiplied with the difference vector $\delta^{i,j}$ yields the difference for the cell quantities w.r.t. the conditional mean $\check{\mu}^i$ as per (3.11). This allows us to express the posterior density for μ as

$$p_\mu(x|\mathbf{X}, \mathbf{M}, M, \Lambda) \qquad\qquad\qquad (4.6)$$
$$\propto p_\mu(x) p_{\{X^i\}}(\bullet|M, \mu = x, \Lambda) p_{\{X^{i,j}\}}(\bullet|\{X^i\}, \mathbf{M}, M, \mu = x, \Lambda)$$
$$\propto \exp\left(-\frac{1}{2}\left(rx^T x + \sum_{i=1}^{M}(\xi^{i,1} - x)^T Q_1(\xi^{i,1} - x) + \sum_{i=1}^{M}\sum_{j=1}^{M^i}(\xi^{i,j} - x)^T Q_2(\xi^{i,j} - x)\right)\right)$$

which yields the distribution

$$\mu|\mathbf{X}, \mathbf{M}, M, \Lambda \sim \mathrm{N}_4\left(Q^{-1}\left(Q_1\sum_{i=1}^{M}\xi^{i,1} + Q_2\sum_{i=1}^{M}\sum_{j=1}^{M^i}\xi^{i,j}\right), Q\right) \qquad (4.7)$$

with precision matrix

$$Q := rI_4 + MQ_1 + \overline{M}Q_2 \qquad\qquad (4.8)$$

As this is a normal distribution, we can use a Gibbs sampler to update μ in a single step. Unfortunately, the same is no longer true for Λ:

$$p_\Lambda(A|\mathbf{X}, \mathbf{M}, M, \mu) \propto p_\Lambda(A) p_{\{X^i\}}(\bullet|M, \mu, \Lambda = A) p_{\{X^{i,j}\}}(\bullet|\{X^i\}, \mathbf{M}, M, \mu, \Lambda = A) \qquad (4.9)$$

This is not a Wishart distribution, as A undergoes some non-linear transforms — see the definitions of $\check{\Lambda}_{11}$ in (3.9) and Q_2 in (4.4). Instead of updating the entire matrix at once, it is more convenient to sample the components Λ_{11}, Λ_{22}, and Λ_{12} conditional on everything else.

If we consider $\check{\Lambda}_{11}$ instead of Λ_{11}, we get

$$p_{\check{\Lambda}_{11}}(A|\mathbf{X}, M, \mu) = p_{\Lambda_{11}}(A + \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{12}^T|\mathbf{X}, M, \mu) \tag{4.10}$$

$$\propto p_{\Lambda}\left(\begin{pmatrix} A + \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{12}^T & \Lambda_{12} \\ \Lambda_{12}^T & \Lambda_{22} \end{pmatrix}\middle| \mathbf{X}, M, \mu\right)$$

$$\propto |A|^{\frac{M}{2}}\exp\left(-\operatorname{tr}(\sqrt{s}A) - \frac{1}{2}\sum_{i=1}^{M}(\check{\delta}^i)^T A \check{\delta}^i\right)$$

To be precise, we would have to add an indicator function that is 1 iff the full matrix Λ is positive definite. We leave this out for the sake of simplicity, but have to keep in mind that we are dealing with a truncated Wishart density. I.e., we get a Gibbs sampler where the new value is generated by rejection sampling from

$$\check{\Lambda}_{11}|\mathbf{X}, M, \mu \sim \operatorname{Wishart}_2\left(\frac{M+3}{2}, \sqrt{s}I_2 + \frac{1}{2}\sum_{i=1}^{M}\check{\delta}^i(\check{\delta}^i)^T\right) \tag{4.11}$$

Proposals are accepted as long as they preserve positive definiteness, and the new Λ_{11} can be obtained from $\check{\Lambda}_{11}$ via (3.9).

For Λ_{22}, the posterior density is

$$p_{\Lambda_{22}}(A|\mathbf{X}, \mathbf{M}, M, \mu, \Lambda_{11}, \Lambda_{12}) \tag{4.12}$$

$$\propto p_{\Lambda}\left(\begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12}^T & A \end{pmatrix}\middle| \mathbf{X}, \mathbf{M}, M, \mu, \Lambda_{11}, \Lambda_{12}\right)$$

$$\propto |\Lambda_{11} - \Lambda_{12}A^{-1}\Lambda_{12}^T|^{\frac{M}{2}}|A|^{\frac{\overline{M}}{2}}\exp\left(-\operatorname{tr}(\sqrt{s}A) + \frac{1}{2}\sum_{i=1}^{M}(\check{\delta}^i)^T\Lambda_{12}A^{-1}\Lambda_{12}^T\check{\delta}^i\right)$$

$$\times \exp\left(-\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M^i}(\check{\xi}^{i,j} - \check{\mu}^i)^T A(\check{\xi}^{i,j} - \check{\mu}^i)\right)$$

$$= |\Lambda_{11} - \Lambda_{12}A^{-1}\Lambda_{12}^T|^{\frac{M}{2}}|A|^{\frac{\overline{M}}{2}}\exp\left(-\operatorname{tr}(\sqrt{s}A) + \frac{1}{2}\sum_{i=1}^{M}(\check{\delta}^i)^T\Lambda_{12}A^{-1}\Lambda_{12}^T\check{\delta}^i\right)$$

$$\times \exp\left(-\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M^i}(\check{\delta}^{i,j} + A^{-1}\Lambda_{12}^T\check{\delta}^i)^T A(\check{\delta}^{i,j} + A^{-1}\Lambda_{12}^T\check{\delta}^i)\right)$$

$$\propto |\Lambda_{11} - \Lambda_{12}A^{-1}\Lambda_{12}^T|^{\frac{M}{2}}|A|^{\frac{\overline{M}}{2}}\exp\left(-\operatorname{tr}(\sqrt{s}A) - \frac{1}{2}\sum_{i=1}^{M}(M^i - 1)(\check{\delta}^i)^T\Lambda_{12}A^{-1}\Lambda_{12}^T\check{\delta}^i\right)$$

$$\times \exp\left(-\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M^i}(\check{\delta}^{i,j})^T A \check{\delta}^{i,j}\right)$$

The form of the density suggests drawing a proposal $\tilde{\Lambda}_{22}$ from

$$\tilde{\Lambda}_{22}|\mathbf{X}, \mathbf{M}, M, \mu \sim \operatorname{Wishart}_2\left(\frac{\overline{M}+3}{2}, \sqrt{s}I_2 + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M^i}\check{\delta}^{i,j}(\check{\delta}^{i,j})^T\right) \tag{4.13}$$

For this move, the acceptance probability is governed by

$$\alpha(\Lambda_{22}, \tilde{\Lambda}_{22}) := \left(\frac{|\Lambda_{11} - \Lambda_{12}\tilde{\Lambda}_{22}^{-1}\Lambda_{12}^T|}{|\check{\Lambda}_{11}|} \right)^{\frac{M}{2}} \tag{4.14}$$

$$\times \exp\left(\frac{1}{2} \sum_{i=1}^{M} (M^i - 1)(\Lambda_{12}^T \check{\delta}^i)^T (\Lambda_{22}^{-1} - \tilde{\Lambda}_{22}^{-1}) \Lambda_{12}^T \check{\delta}^i \right)$$

As for $\check{\Lambda}_{11}$, the proposal can be rejected outright if the resulting Λ is no longer positive definite. The posterior density for Λ_{12} is

$$p_{\Lambda_{12}}(A|\mathbf{X}, \mathbf{M}, M, \mu, \Lambda_{11}, \Lambda_{22}) \tag{4.15}$$

$$\propto p_\Lambda\left(\begin{pmatrix} \Lambda_{11} & A \\ A^T & \Lambda_{22} \end{pmatrix} \middle| \mathbf{X}, \mathbf{M}, M, \mu, \Lambda_{11}, \Lambda_{22} \right)$$

$$\propto |\Lambda_{11} - A\Lambda_{22}^{-1}A^T|^{\frac{M}{2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{M} (M^i - 1)(\check{\delta}^i)^T A\Lambda_{22}^{-1} A^T \check{\delta}^i - \sum_{i=1}^{M} \sum_{j=1}^{M^i} (\check{\delta}^i)^T A\check{\delta}^{i,j} \right)$$

It is easier to find a suitable proposal distribution if the entries of Λ_{12} are sampled individually. If at least one storm has two or more cells, the posterior for $\Lambda_{1,3}$ is

$$p_{\Lambda_{1,3}}(x|\mathbf{X}, \mathbf{M}, M, \mu, \Lambda_{11}, \Lambda_{22}, \Lambda_{1,4}, \Lambda_{2,3}, \Lambda_{2,4}) \tag{4.16}$$

$$\propto p_{\Lambda_{12}}\left(\begin{pmatrix} x & \Lambda_{1,4} \\ \Lambda_{2,3} & \Lambda_{2,4} \end{pmatrix} \middle| \mathbf{X}, \mathbf{M}, M, \mu, \Lambda_{11}, \Lambda_{22}, \Lambda_{1,4}, \Lambda_{2,3}, \Lambda_{2,4} \right)$$

$$\propto \left| \Lambda_{11} - \begin{pmatrix} x & \Lambda_{1,4} \\ \Lambda_{2,3} & \Lambda_{2,4} \end{pmatrix} \Lambda_{22}^{-1} \begin{pmatrix} x & \Lambda_{2,3} \\ \Lambda_{1,4} & \Lambda_{2,4} \end{pmatrix} \right|^{\frac{M}{2}} \exp\left(-\frac{\pi_{1,3}}{2}(x - \nu_{1,3})^2 \right)$$

using

$$\pi_{1,3} := \frac{\Lambda_{4,4}}{|\Lambda_{22}|} \sum_{i=1}^{M} (M^i - 1)(\delta_1^{i,1})^2 \tag{4.17}$$

$$\nu_{1,3} := \frac{1}{\pi_{1,3}} \sum_{i=1}^{M} \left(\frac{(M^i - 1)}{|\Lambda_{22}|} (\Lambda_{1,4}\Lambda_{3,4}(\delta_1^{i,1})^2 + (\Lambda_{2,4}\Lambda_{3,4} - \Lambda_{2,3}\Lambda_{4,4})\delta_1^{i,1}\delta_2^{i,1}) - \delta_1^{i,1} \sum_{j=1}^{M^i} \delta_3^{i,j} \right)$$

Thus, we can draw a proposal $\tilde{\Lambda}_{1,3}$ from

$$\tilde{\Lambda}_{1,3}|\mathbf{X}, \mathbf{M}, M, \mu, \Lambda_{22}, \Lambda_{1,4}, \Lambda_{2,3}, \Lambda_{2,4} \sim N(\nu_{1,3}, \pi_{1,3}) \tag{4.18}$$

which has an acceptance probability determined by

$$\alpha(\Lambda_{1,3}, \tilde{\Lambda}_{1,3}) := \left(\frac{\left| \Lambda_{11} - \begin{pmatrix} \tilde{\Lambda}_{1,3} & \Lambda_{1,4} \\ \Lambda_{2,3} & \Lambda_{2,4} \end{pmatrix} \Lambda_{22}^{-1} \begin{pmatrix} \tilde{\Lambda}_{1,3} & \Lambda_{2,3} \\ \Lambda_{1,4} & \Lambda_{2,4} \end{pmatrix} \right|}{|\check{\Lambda}_{11}|} \right)^{\frac{M}{2}} \tag{4.19}$$

The remaining parameters are sampled in a similar manner. The means and variances for the normal proposals belonging to $\Lambda_{1,4}$, $\Lambda_{2,3}$, and $\Lambda_{2,4}$ are

$$\pi_{1,4} := \frac{\Lambda_{3,3}}{|\Lambda_{22}|} \sum_{i=1}^{M} (M^i - 1)(\delta_1^{i,1})^2 \tag{4.20}$$

$$\nu_{1,4} := \frac{1}{\pi_{1,4}} \sum_{i=1}^{M} \left(\frac{(M^i - 1)}{|\Lambda_{22}|} (\Lambda_{1,3}\Lambda_{3,4}(\delta_1^{i,1})^2 + (\Lambda_{2,3}\Lambda_{3,4} - \Lambda_{2,4}\Lambda_{3,3})\delta_1^{i,1}\delta_2^{i,1}) - \delta_1^{i,1} \sum_{j=1}^{M^i} \delta_4^{i,j} \right)$$

$$\pi_{2,3} := \frac{\Lambda_{4,4}}{|\Lambda_{22}|} \sum_{i=1}^{M} (M^i - 1)(\delta_2^{i,1})^2$$

$$\nu_{2,3} := \frac{1}{\pi_{2,3}} \sum_{i=1}^{M} \left(\frac{(M^i - 1)}{|\Lambda_{22}|} (\Lambda_{2,4}\Lambda_{3,4}(\delta_2^{i,1})^2 + (\Lambda_{1,4}\Lambda_{3,4} - \Lambda_{1,3}\Lambda_{4,4})\delta_1^{i,1}\delta_2^{i,1}) - \delta_2^{i,1} \sum_{j=1}^{M^i} \delta_3^{i,j} \right)$$

$$\pi_{2,4} := \frac{\Lambda_{3,3}}{|\Lambda_{22}|} \sum_{i=1}^{M} (M^i - 1)(\delta_2^{i,1})^2$$

$$\nu_{2,4} := \frac{1}{\pi_{2,4}} \sum_{i=1}^{M} \left(\frac{(M^i - 1)}{|\Lambda_{22}|} (\Lambda_{2,3}\Lambda_{3,4}(\delta_2^{i,1})^2 + (\Lambda_{1,3}\Lambda_{3,4} - \Lambda_{1,4}\Lambda_{3,3})\delta_1^{i,1}\delta_2^{i,1}) - \delta_2^{i,1} \sum_{j=1}^{M^i} \delta_4^{i,j} \right)$$

The acceptance probabilities for these samplers look similar to (4.19), except that the appropriate entry of the matrix Λ_{12} has to be replaced by its proposal in the numerator of α. As before, a proposal is only admissible if Λ stays positive definite.

Remark 4.2 *In case all the M^i are equal to 1, the posterior does not resemble a normal distribution. We do not examine this special case further, as our implementation of the algorithm does not allow storms with a single cell (see Section 5.1.2).*

One iteration for updating μ and Λ under model I consists of drawing μ and $\check{\Lambda}_{11}$ from their posterior distributions, as well as generating proposals and checking acceptance for the remaining components of Λ.

4.3.3 Model IV: Sampling μ and diagonal Λ for matched means

Again, we look at the simple variant with diagonal Λ before the full version. If μ_4 is treated as a function of the other parameters as per (3.21), the components of the remaining mean vector μ_R still have a normal posterior distribution. However, the posterior for two of the precisions is no longer gamma.

Consider first

$$p_{\mu_1}(x|\mathbf{X}, \mathbf{M}, M, \lambda_1) \tag{4.21}$$
$$\propto p_{\mu_1}(x) p_{\{X_1^i\}}(\bullet|M, \mu_1 = x, \lambda_1) p_{\{X_2^{i,j}\}}(\bullet|\mathbf{M}, M, \mu_1 = x, \lambda_1, \lambda_4)$$
$$\propto \exp\left(-\frac{1}{2}\left(rx^2 + \lambda_1 \sum_{i=1}^{M}(\log(X_1^i) - x)^2\right)\right)$$
$$\times \exp\left(-\frac{\lambda_4}{2} \sum_{i=1}^{M}\sum_{j=1}^{M^i}\left(V^{i,j}(\sigma_1^2, \sigma_4^2) - x\right)^2\right)$$

using

$$V^{i,j}(\sigma_1^2, \sigma_4^2) := \log(X_2^{i,j}) + \frac{\sigma_4^2 - \sigma_1^2}{2} + \log(K) \qquad \sigma_1^2 = \frac{1}{\lambda_1} \qquad \sigma_4^2 = \frac{1}{\lambda_4} \tag{4.22}$$

This posterior is normal, so we can use a Gibbs sampler that draws from

$$\mu_1|\mathbf{X}, \mathbf{M}, M, \lambda_1 \tag{4.23}$$
$$\sim \mathrm{N}\left(\frac{\lambda_1 \sum_{i=1}^{M}\log(X_1^i) + \lambda_4 \sum_{i=1}^{M}\sum_{j=1}^{M^i} V^{i,j}(\sigma_1^2, \sigma_4^2)}{r + M\lambda_1 + \overline{M}\lambda_4}, r + M\lambda_1 + \overline{M}\lambda_4\right)$$

The means μ_2 and μ_3 are not affected by the change in the model, they can be generated as shown for model II in Section 4.3.1.

For λ_1, we obtain

$$p_{\lambda_1}(x|\mathbf{X}, \mathbf{M}, M, \mu_1) \tag{4.24}$$
$$\propto p_{\lambda_1}(x) p_{\{X_1^i\}}(\bullet|M, \mu_1, \lambda_1 = x) p_{\{X_2^{i,j}\}}(\bullet|\mathbf{M}, M, \mu_1, \lambda_1 = x, \lambda_4)$$
$$\propto x^{s+\frac{M}{2}-1} \exp\left(-sx - \frac{x}{2} \sum_{i=1}^{M}(\log(X_1^i) - \mu_1)^2\right)$$
$$\times \exp\left(-\frac{\lambda_4}{2} \sum_{i=1}^{M}\sum_{j=1}^{M^i}(V^{i,j}(x^{-1}, \sigma_4^2) - \mu_1)^2\right)$$

This is not a gamma distribution, but we can draw a proposal $\tilde{\lambda}_1$ from

$$\tilde{\lambda}_1|\mathbf{X}, M, \mu_1 \sim \mathrm{Gamma}\left(s + \frac{M}{2}, s + \frac{1}{2} \sum_{i=1}^{M}(\log(X_1^i) - \mu_1)^2\right) \tag{4.25}$$

The acceptance probability depends on

$$\alpha(\lambda_1, \tilde{\lambda}_1) := \exp\left(\frac{\lambda_4}{2} \sum_{i=1}^{M}\sum_{j=1}^{M^i}((V^{i,j}(\lambda_1^{-1}, \sigma_4^2) - \mu_1)^2 - (V^{i,j}(\tilde{\lambda}_1^{-1}, \sigma_4^2) - \mu_1)^2)\right) \tag{4.26}$$

While λ_2 and λ_3 can be sampled as for model II, the posterior distribution for λ_4 is

$$p_{\lambda_4}(x|\mathbf{X}, \mathbf{M}, M, \mu_1, \lambda_1) \tag{4.27}$$

$$\propto p_{\lambda_4}(x)p_{\{X_2^{i,j}\}}(\bullet|\mathbf{M}, M, \mu_1, \lambda_1, \lambda_4 = x)$$

$$\propto x^{s+\frac{\overline{M}}{2}-1} \exp\left(-sx - \frac{x}{2}\sum_{i=1}^{M}\sum_{j=1}^{M^i}\left(\log(X_2^{i,j}) - \mu_1 - \frac{1}{2\lambda_1} + \log(K)\right)^2 - \frac{\overline{M}}{8x}\right)$$

Again, this suggests a gamma proposal drawn from

$$\tilde{\lambda}_4|\mathbf{X}, \mathbf{M}, M, \mu_1, \lambda_1 \tag{4.28}$$

$$\sim \text{Gamma}\left(s + \frac{\overline{M}}{2}, s + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M^i}\left(\log(X_2^{i,j}) - \mu_1 - \frac{1}{2\lambda_1} + \log(K)\right)^2\right)$$

with an acceptance probability determined by

$$\alpha(\lambda_4, \tilde{\lambda}_4) := \exp\left(\frac{\overline{M}}{8\lambda_4} - \frac{\overline{M}}{8\tilde{\lambda}_4}\right) \tag{4.29}$$

4.3.4 Model III: Sampling μ and arbitrary Λ for matched means

As in the previous section, treating μ_4 as a function of the other parameters does not change the fact that the remaining mean vector μ_R has a normal posterior. And not surprisingly, the precisions are no longer exactly Wishart.

First, we define some auxiliary expressions similar to those used for model I:

$$\Lambda_R =: \begin{pmatrix} \Lambda_{11} & \Lambda_{R,12} \\ \Lambda_{R,12}^T & \Lambda_{3,3} \end{pmatrix} \qquad \check{\Lambda}_{R,11} := \Lambda_{11} - \frac{\Lambda_{R,12}\Lambda_{R,12}^T}{\Lambda_{3,3}} \tag{4.30}$$

$$\xi_R^{i,j} := \log\begin{pmatrix} X_1^i \\ X_2^i \\ X_1^{i,j} \end{pmatrix} \qquad \hat{V}^{i,j}(\sigma_1^2, \sigma_4^2) := \begin{pmatrix} V^{i,j}(\sigma_1^2, \sigma_4^2) \\ 0 \\ 0 \end{pmatrix}$$

$$Q_{1,R} := \begin{pmatrix} \check{\Lambda}_{R,11} & \underline{0} \\ \underline{0} & 0 \end{pmatrix} \in \mathbb{R}^{3\times 3} \qquad Q_{2,R} := \Lambda_{3,3}Q_{3,R}^T Q_{3,R} = \begin{pmatrix} \frac{\Lambda_{R,12}\Lambda_{R,12}^T}{\Lambda_{3,3}} & \Lambda_{R,12} \\ \Lambda_{R,12}^T & \Lambda_{3,3} \end{pmatrix} \tag{4.31}$$

$$Q_{3,R} := \begin{pmatrix} \frac{\Lambda_{R,12}^T}{\Lambda_{3,3}} & 1 \end{pmatrix} \qquad Q_{4,R} := \text{diag}(\Lambda_{4,4}, 0, 0)$$

Note also that we can calculate the covariance σ_1^2 from

$$\sigma_1^2 = (\check{\Lambda}_{R,11}^{-1})_{1,1} = \frac{(\check{\Lambda}_{R,11})_{2,2}}{(\check{\Lambda}_{R,11})_{1,1}(\check{\Lambda}_{R,11})_{2,2} - (\check{\Lambda}_{R,11})_{1,2}^2} \tag{4.32}$$

By (2.8), the conditional distribution of $\log(X_1)^{i,j}$ given X^i is

$$\log(X_1)^{i,j}|X^i, \mu, \Lambda \sim \mathrm{N}\left(\mu_3 - \frac{\Lambda_{R,12}^T}{\Lambda_{3,3}}\left(\begin{array}{c}\log(X_1^i) - \mu_1 \\ \log(X_2^i) - \mu_2\end{array}\right), \Lambda_{3,3}\right) \tag{4.33}$$

This allows us to express the posterior density for μ_R as

$$p_{\mu_R}(x|\mathbf{X}, \mathbf{M}, M, \Lambda_R) \tag{4.34}$$

$$\propto p_{\mu_R}(x)p_{\{X^i\}}(\bullet|M, \mu_R = x, \Lambda_R)p_{\{X^{i,j}\}}(\bullet|\{X^i\}, \mathbf{M}, M, \mu_R = x, \Lambda_R)$$

$$\propto \exp\left(-\frac{1}{2}\left(rx^Tx + \sum_{i=1}^{M}(\xi_R^{i,1} - x)^TQ_{1,R}(\xi_R^{i,1} - x)\right)\right)$$

$$\times \exp\left(-\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M_i}(\xi_R^{i,j} - x)^TQ_{2,R}(\xi_R^{i,j} - x)\right)$$

$$\times \exp\left(-\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M^i}(\hat{V}^{i,j}(\sigma_1^2, \sigma_4^2) - x)^TQ_{4,R}(\hat{V}^{i,j}(\sigma_1^2, \sigma_4^2) - x)\right)$$

The resulting posterior distribution is normal with

$$\mu_R|\mathbf{X}, \mathbf{M}, M, \Lambda_R \tag{4.35}$$

$$\sim \mathrm{N}_3\left(Q_R^{-1}\left(\sum_{i=1}^{M}Q_{1,R}\xi_R^{i,1} + \sum_{i=1}^{M}\sum_{j=1}^{M^i}(Q_{2,R}\xi_R^{i,j} + Q_{4,R}\hat{V}^{i,j}(\sigma_1^2, \sigma_4^2))\right), Q_R\right)$$

and precision matrix

$$Q_R := rI_3 + MQ_{1,R} + \overline{M}(Q_{2,R} + Q_{4,R}) \tag{4.36}$$

This permits a single-step Gibbs update for μ_R.

For Λ, we follow decomposition approach similar to the one used in Section 4.3.2:

$$p_{\Lambda_R}(A|\mathbf{X}, \mathbf{M}, M, \mu_R) \tag{4.37}$$

$$\propto p_{\Lambda_R}(A)p_{\{X^i\}}(\bullet|M, \Lambda_R = A, \mu_R)$$

$$\times p_{\{X_1^{i,j}\}}(\bullet|\{X^i\}, \mathbf{M}, M, \Lambda_R = A, \mu_R)p_{\{X_2^{i,j}\}}(\bullet|\mathbf{M}, M, \Lambda_R = A, \mu_R)$$

As before, we update $\check{\Lambda}_{R,11}$ instead of Λ_{11}:

$$p_{\check{\Lambda}_{R,11}}(A|\mathbf{X}, \mathbf{M}, M, \mu, \Lambda_{R,12}, \Lambda_{3,3}, \Lambda_{4,4}) \tag{4.38}$$

$$\propto p_{\Lambda_R}\left(\left(\begin{array}{cc}A + \frac{\Lambda_{R,12}\Lambda_{R,12}^T}{\Lambda_{3,3}} & \Lambda_{R,12} \\ \Lambda_{R,12}^T & \Lambda_{3,3}\end{array}\right)\middle|\mathbf{X}, \mathbf{M}, M, \mu, \Lambda_{R,12}, \Lambda_{3,3}, \Lambda_{4,4}\right)$$

$$\propto |A|^{\frac{M}{2}}\exp\left(-\mathrm{tr}(\sqrt{s}A) - \frac{1}{2}\sum_{i=1}^{M}(\check{\delta}^i)^TA\check{\delta}^i\right)$$

$$\times \exp\left(-\frac{\Lambda_{4,4}}{2}\sum_{i=1}^{M}\sum_{j=1}^{M^i}(V^{i,j}(\sigma_1^2, \sigma_4^2) - \mu_1)^2\right)$$

If we draw a proposal $\tilde{\tilde{\Lambda}}_{R,11}$ from

$$\tilde{\tilde{\Lambda}}_{R,11}|\mathbf{X}, M, \mu \sim \text{Wishart}_2\left(\frac{M+3}{2}, \sqrt{s}I_2 + \frac{1}{2}\sum_{i=1}^{M}\check{\delta}^i(\check{\delta}^i)^T\right) \quad (4.39)$$

the acceptance probability is determined by

$$\alpha(\check{\Lambda}_{R,11}, \tilde{\tilde{\Lambda}}_{R,11}) := \exp\left(\frac{\Lambda_{4,4}}{2}\sum_{i=1}^{M}\sum_{j=1}^{M^i}((V^{i,j}(\sigma_1^2, \sigma_4^2) - \mu_1)^2 - (V^{i,j}(\tilde{\sigma}_1^2, \sigma_4^2) - \mu_1)^2)\right) \quad (4.40)$$

The posterior density for $\Lambda_{3,3}$ is

$$p_{\Lambda_{3,3}}(x|\mathbf{X}, \mathbf{M}, M, \mu, \Lambda_{11}, \Lambda_{R,12}) \quad (4.41)$$

$$\propto p_{\Lambda_R}\left(\begin{pmatrix}\Lambda_{R,11} & \Lambda_{R,12}\\ \Lambda_{R,12}^T & x\end{pmatrix}\bigg|\mathbf{X}, \mathbf{M}, M, \mu, \Lambda_{11}, \Lambda_{R,12}\right)$$

$$\propto \left|\Lambda_{R,11} - \frac{\Lambda_{R,12}\Lambda_{R,12}^T}{x}\right|^{\frac{M}{2}}|x|^{\frac{\overline{M}}{2}}\exp\left(-\sqrt{s}x - \frac{1}{2x}\sum_{i=1}^{M}(M^i - 1)(\Lambda_{R,12}^T\check{\delta}^i)^2\right)$$

$$\times \exp\left(-\frac{x}{2}\sum_{i=1}^{M}\sum_{j=1}^{M^i}(\delta_3^{i,j})^2\right)$$

If we draw a proposal $\tilde{\Lambda}_{3,3}$ from

$$\tilde{\Lambda}_{3,3}|\mathbf{X}, \mathbf{M}, M, \mu \sim \text{Gamma}\left(\frac{\overline{M}+2}{2}, \sqrt{s} + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M^i}(\delta_3^{i,j})^2\right) \quad (4.42)$$

the acceptance probability depends on

$$\alpha(\Lambda_{3,3}, \tilde{\Lambda}_{3,3}) := \left(\frac{\left|\Lambda_{R,11} - \frac{\Lambda_{R,12}\Lambda_{R,12}^T}{\tilde{\Lambda}_{3,3}}\right|}{|\check{\Lambda}_{R,11}|}\right)^{\frac{M}{2}} \quad (4.43)$$

$$\times \exp\left(\left(\frac{1}{2\Lambda_{3,3}} - \frac{1}{2\tilde{\Lambda}_{3,3}}\right)\sum_{i=1}^{M}(M^i - 1)(\Lambda_{R,12}^T\check{\delta}^i)^2\right)$$

The components of $\Lambda_{R,12}$ are updated individually. The posterior densities are

$$p_{\Lambda_{1,3}}(x|\mathbf{X}, \mathbf{M}, M, \mu, \Lambda_{11}, \Lambda_{2,3}, \Lambda_{3,3}) \quad (4.44)$$

$$\propto \left|\Lambda_{R,11} - \frac{1}{\Lambda_{3,3}}\begin{pmatrix}x\\ \Lambda_{2,3}\end{pmatrix}\begin{pmatrix}x & \Lambda_{2,3}\end{pmatrix}\right|^{\frac{M}{2}}\exp\left(-\frac{\pi_{R,1,3}}{2}(x - \nu_{R,1,3})^2\right)$$

$$p_{\Lambda_{2,3}}(x|\mathbf{X}, \mathbf{M}, M, \mu, \Lambda_{11}, \Lambda_{1,3}, \Lambda_{3,3}) \quad (4.45)$$

$$\propto \left|\Lambda_{R,11} - \frac{1}{\Lambda_{3,3}}\begin{pmatrix}\Lambda_{1,3}\\ x\end{pmatrix}\begin{pmatrix}\Lambda_{1,3} & x\end{pmatrix}\right|^{\frac{M}{2}}\exp\left(-\frac{\pi_{R,2,3}}{2}(x - \nu_{R,2,3})^2\right)$$

with parameters

$$\pi_{R,1,3} := \frac{1}{\Lambda_{3,3}} \sum_{i=1}^{M} (M^i - 1)(\delta_1^{i,1})^2 \tag{4.46}$$

$$\nu_{R,1,3} := -\frac{1}{\pi_{R,1,3}} \sum_{i=1}^{M} \left(\frac{(M^i - 1)\Lambda_{2,3}\delta_1^{i,1}\delta_2^{i,1}}{\Lambda_{3,3}} + \delta_1^{i,1} \sum_{j=1}^{M^i} \delta_3^{i,j} \right)$$

$$\pi_{R,2,3} := \frac{1}{\Lambda_{3,3}} \sum_{i=1}^{M} (M^i - 1)(\delta_2^{i,1})^2$$

$$\nu_{R,2,3} := -\frac{1}{\pi_{R,2,3}} \sum_{i=1}^{M} \left(\frac{(M^i - 1)\Lambda_{1,3}\delta_1^{i,1}\delta_2^{i,1}}{\Lambda_{3,3}} + \delta_2^{i,1} \sum_{j=1}^{M^i} \delta_3^{i,j} \right)$$

If we draw a proposal $\tilde{\Lambda}_{k,3}$ for $k \in \{1, 2\}$ from

$$\Lambda_{k,3} | \mathbf{X}, \mathbf{M}, M, \mu, \Lambda_{3-k,3}, \Lambda_{3,3} \sim \mathrm{N}(\nu_{R,k,3}, \pi_{R,k,3}) \tag{4.47}$$

the acceptance probability is determined by

$$\alpha(\Lambda_{k,3}, \tilde{\Lambda}_{k,3}) := \left(\frac{\left| \Lambda_{R,11} - \frac{\tilde{\Lambda}_{R,12}\tilde{\Lambda}_{R,12}^T}{\Lambda_{3,3}} \right|}{|\check{\Lambda}_{R,11}|} \right)^{\frac{M}{2}} \tag{4.48}$$

As for model I, the case where all storms have but a single cell does not concern us.

The posterior distribution for $\Lambda_{4,4}$ is

$$p_{\Lambda_{4,4}}(x | \mathbf{X}, \mathbf{M}, M, \mu, \Lambda_R) \tag{4.49}$$
$$\propto p_{\Lambda_{4,4}}(x) p_{\{X_2^{i,j}\}}(\bullet | \mathbf{M}, M, \mu, \Lambda_R, \Lambda_{4,4} = x)$$
$$\propto x^{s + \frac{\overline{M}}{2} - 1} \exp\left(-sx - \frac{x}{2} \sum_{i=1}^{M} \sum_{j=1}^{M^i} \left(\log(X_2^{i,j}) - \mu_1 - \frac{\sigma_1^2}{2} + \log(K) \right)^2 - \frac{\overline{M}}{8x} \right)$$

We can generate a proposal $\tilde{\Lambda}_{4,4}$ from a gamma distribution

$$\tilde{\Lambda}_{4,4} | \mathbf{X}, \mathbf{M}, M, \mu, \Lambda_R \tag{4.50}$$
$$\sim \mathrm{Gamma}\left(s + \frac{\overline{M}}{2}, s + \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M^i} \left(\log(X_2^{i,j}) - \mu_1 - \frac{\sigma_1^2}{2} + \log(K) \right)^2 \right)$$

and accept the new value with a probability depending on

$$\alpha(\Lambda_{4,4}, \tilde{\Lambda}_{4,4}) := \exp\left(\frac{\overline{M}}{8\Lambda_{4,4}} - \frac{\overline{M}}{8\tilde{\Lambda}_{4,4}} \right) \tag{4.51}$$

As in Section 4.3.2, we need to test the new Λ for positive definiteness after each component update, rejecting any proposals that destroy the property.

4.3.5 Model V: Sampling μ and Λ for uncorrelated cell duration

This variant is a straightforward reduction of model I, where three of the coprecisions are fixed to 0. As above, we begin by defining some auxiliary expressions:

$$\Lambda_R =: \begin{pmatrix} \Lambda_{11} & \Lambda_{R,12} \\ \Lambda_{R,12}^T & \Lambda_{4,4} \end{pmatrix} \qquad \check{\Lambda}_{R,11} := \Lambda_{11} - \frac{\Lambda_{R,12}\Lambda_{R,12}^T}{\Lambda_{4,4}} \qquad \xi_R^{i,j} := \log \begin{pmatrix} X_1^i \\ X_2^i \\ X_2^{i,j} \end{pmatrix} \qquad (4.52)$$

$$Q_{1,R} := \begin{pmatrix} \check{\Lambda}_{R,11} & \underline{0} \\ \underline{0} & 0 \end{pmatrix} \in \mathbb{R}^{3\times 3} \qquad\qquad Q_{2,R} = \begin{pmatrix} \frac{\Lambda_{R,12}\Lambda_{R,12}^T}{\Lambda_{4,4}} & \Lambda_{R,12} \\ \Lambda_{R,12}^T & \Lambda_{4,4} \end{pmatrix} \qquad (4.53)$$

The conditional density for μ_R can be derived as in (4.6):

$$p_{\mu_R}(x|\mathbf{X}, \mathbf{M}, M, \Lambda_R) \propto \exp\left(-\frac{1}{2}\left(rx^T x + \sum_{i=1}^{M}(\xi_R^{i,1} - x)^T Q_{1,R}(\xi_R^{i,1} - x)\right)\right) \qquad (4.54)$$

$$\times \exp\left(-\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M_i}(\xi_R^{i,j} - x)^T Q_{2,R}(\xi_R^{i,j} - x)\right)$$

The resulting posterior distribution is once again normal

$$\mu_R|\mathbf{X}, \mathbf{M}, M, \Lambda_R \sim \mathrm{N}\left(Q_R^{-1}\left(Q_{1,R}\sum_{i=1}^{M}\xi_R^{i,1} + Q_{2,R}\sum_{i=1}^{M}\sum_{j=1}^{M^i}\xi_R^{i,j}\right), Q_R\right) \qquad (4.55)$$

with

$$Q_R := rI_3 + MQ_{1,R} + \overline{M}Q_{2,R} \qquad (4.56)$$

This yields a Gibbs sampler for μ_R, while the remaining mean parameter μ_3 can be sampled from (4.2) as for model II.

To update Λ_R, we use the same decomposition as before. First, we sample $\check{\Lambda}_{R,11}$ from its Wishart posterior (cp. (4.10)):

$$p_{\check{\Lambda}_{R,11}}(A|\mathbf{X}, M, \mu) \propto |A|^{\frac{M}{2}}\exp\left(-\mathrm{tr}(\sqrt{s}A) - \frac{1}{2}\sum_{i=1}^{M}(\check{\delta}^i)^T A\check{\delta}^i\right) \qquad (4.57)$$

Thus, we can draw a new submatrix from

$$\check{\Lambda}_{R,11}|\mathbf{X}, M, \mu \sim \mathrm{Wishart}_2\left(\frac{M+3}{2}, \sqrt{s}I_2 + \frac{1}{2}\sum_{i=1}^{M}\check{\delta}^i(\check{\delta}^i)^T\right) \qquad (4.58)$$

and accept if the modified Λ_R remains positive definite.

The posterior density for $\Lambda_{4,4}$ can be found from (4.12):

$$p_{\Lambda_{4,4}}(x|\mathbf{X}, \mathbf{M}, M, \mu, \Lambda_{11}, \Lambda_{12}) \tag{4.59}$$

$$\propto \left| \Lambda_{11} - \frac{\Lambda_{R,12}\Lambda_{R,12}^T}{x} \right|^{\frac{M}{2}} x^{\frac{\overline{M}}{2}} \exp\left(-\sqrt{s}x - \frac{1}{2x}\sum_{i=1}^{M}(M^i - 1)(\check{\delta}^i)^T \Lambda_{R,12}\Lambda_{R,12}^T \check{\delta}^i \right)$$

$$\times \exp\left(-\frac{x}{2}\sum_{i=1}^{M}\sum_{j=1}^{M^i}(\delta_2^{i,j})^2 \right)$$

If we draw a proposal $\tilde{\Lambda}_{4,4}$ from

$$\tilde{\Lambda}_{4,4}|\mathbf{X}, \mathbf{M}, M, \mu \sim \text{Gamma}\left(\frac{\overline{M}+2}{2}, \sqrt{s} + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M^i}(\delta_2^{i,j})^2 \right) \tag{4.60}$$

the acceptance probability for the MH-sampler is given by

$$\alpha(\Lambda_{4,4}, \tilde{\Lambda}_{4,4}) = \left(\frac{\left| \Lambda_{11} - \frac{\Lambda_{R,12}\Lambda_{R,12}^T}{\tilde{\Lambda}_{4,4}} \right|}{|\check{\Lambda}_{11}|} \right)^{\frac{M}{2}} \exp\left(\frac{\tilde{\Lambda}_{4,4} - \Lambda_{4,4}}{2\Lambda_{4,4}\tilde{\Lambda}_{4,4}}\sum_{i=1}^{M}(M^i - 1)(\Lambda_{R,12}^T\check{\delta}^i)^2 \right) \tag{4.61}$$

As usual, the proposal needs be rejected if the resulting matrix Λ is not positive definite. The parameter $\Lambda_{3,3}$ can be sampled from (4.2) as given for model II.

Finally, proposals for the coprecisions can drawn from a normal distribution similar to (4.18):

$$\tilde{\Lambda}_{k,4}|\mathbf{X}, \mathbf{M}, M, \mu, \Lambda_{3-k,4}, \Lambda_{4,4} \sim \text{N}(\nu_{R,k}, \pi_{R,k}) \tag{4.62}$$

The parameters are defined as

$$\pi_{R,k} = \frac{1}{\Lambda_{4,4}}\sum_{i=1}^{M}(M^i - 1)(\delta_k^{i,1})^2 \tag{4.63}$$

$$\nu_{R,k} = -\frac{1}{\pi_{R,k}}\sum_{i=1}^{M}\left(\frac{(M^i - 1)\Lambda_{(3-k),4}}{\Lambda_{4,4}}\delta_1^{i,1}\delta_2^{i,1} + \delta_k^{i,1}\sum_{j=1}^{M^i}\delta_4^{i,j} \right)$$

The resulting MH-sampler has an acceptance probability of the same form as (4.19), provided the matrix remains positive definite.

4.3.6 Model VI: Sampling μ and Λ with a single correlation coefficient

This model is even simpler than number V, as it includes only a single non-zero coprecision value. Our auxiliary expressions become

$$\check{\Lambda}_{1,1} := \Lambda_{1,1} - \frac{\Lambda_{1,4}^2}{\Lambda_{4,4}} \qquad\qquad \xi_R^{i,j} := \log\left(\frac{X_1^i}{X_2^{i,j}} \right) \tag{4.64}$$

$$Q_{1,R} := \begin{pmatrix} \check{\Lambda}_{1,1} & 0 \\ 0 & 0 \end{pmatrix} \qquad\qquad Q_{2,R} = \begin{pmatrix} \frac{\Lambda_{1,4}^2}{\Lambda_{4,4}} & \Lambda_{1,4} \\ \Lambda_{1,4} & \Lambda_{4,4} \end{pmatrix} \qquad (4.65)$$

The conditional density for μ_R is

$$p_{\mu_R}(x|\mathbf{X}, \mathbf{M}, M, \Lambda_R) \propto \exp\left(-\frac{1}{2}\left(rx^Tx + \sum_{i=1}^{M}(\xi_R^{i,1} - x)^T Q_{1,R}(\xi_R^{i,1} - x)\right)\right) \qquad (4.66)$$

$$\times \exp\left(-\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M_i}(\xi_R^{i,j} - x)^T Q_{2,R}(\xi_R^{i,j} - x)\right)$$

Again, μ_R is sampled from its normal posterior

$$\mu_R|\mathbf{X}, \mathbf{M}, M, \Lambda_R \sim \mathrm{N}\left(Q_R^{-1}\left(Q_{1,R}\sum_{i=1}^{M}\xi_R^{i,1} + Q_{2,R}\sum_{i=1}^{M}\sum_{j=1}^{M^i}\xi_R^{i,j}\right), Q_R\right) \qquad (4.67)$$

with

$$Q_R := rI_2 + MQ_{1,R} + \overline{M}Q_{2,R} \qquad (4.68)$$

The remaining mean parameters μ_2 and μ_3 can be sampled from (4.1) respectively (4.2) as for model II.

The precision sampler uses the same decomposition as before. We sample $\check{\Lambda}_{1,1}$ from its Wishart posterior:

$$p_{\check{\Lambda}_{1,1}}(x|\mathbf{X}, M, \mu) \propto x^{\frac{M}{2}}\exp\left(-\sqrt{s}x - \frac{x}{2}\sum_{i=1}^{M}(\check{\delta}_1^i)^2\right) \qquad (4.69)$$

Thus, we can draw a new marginal precision from

$$\check{\Lambda}_{1,1}|\mathbf{X}, M, \mu \sim \mathrm{Gamma}\left(\frac{M+2}{2}, \sqrt{s} + \frac{1}{2}\sum_{i=1}^{M}(\check{\delta}_1^i)^2\right) \qquad (4.70)$$

and accept if the modified Λ_R remains positive definite.

The posterior density for $\Lambda_{4,4}$ is

$$p_{\Lambda_{4,4}}(x|\mathbf{X}, \mathbf{M}, M, \mu, \Lambda_{1,1}, \Lambda_{1,4}) \qquad (4.71)$$

$$\propto \left|\Lambda_{1,1} - \frac{\Lambda_{1,4}^2}{x}\right|^{\frac{M}{2}} x^{\frac{\overline{M}}{2}}\exp\left(-\sqrt{s}x - \frac{\Lambda_{1,4}^2}{2x}\sum_{i=1}^{M}(M^i - 1)(\check{\delta}_1^i)^2 - \frac{x}{2}\sum_{i=1}^{M}\sum_{j=1}^{M^i}(\delta_2^{i,j})^2\right)$$

If we draw a proposal $\tilde{\Lambda}_{4,4}$ from

$$\tilde{\Lambda}_{4,4}|\mathbf{X}, \mathbf{M}, M, \mu, \Lambda_{1,1}, \Lambda_{1,4} \sim \mathrm{Gamma}\left(\frac{\overline{M}+2}{2}, \sqrt{s} + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M^i}(\delta_2^{i,j})^2\right) \qquad (4.72)$$

the acceptance probability for the MH-sampler depends on

$$\alpha(\Lambda_{4,4}, \tilde{\Lambda}_{4,4}) = \left(\frac{\left| \Lambda_{1,1} - \frac{\Lambda_{1,4}^2}{\tilde{\Lambda}_{4,4}} \right|}{|\check{\Lambda}_{1,1}|} \right)^{\frac{M}{2}} \exp \left(\frac{\Lambda_{1,4}^2 (\tilde{\Lambda}_{4,4} - \Lambda_{4,4})}{2\Lambda_{4,4}\tilde{\Lambda}_{4,4}} \sum_{i=1}^{M} (M^i - 1)(\check{\delta}_1^i)^2 \right) \tag{4.73}$$

As usual, the proposal has to be rejected if the resulting Λ_R is not positive definite.

The parameters $\Lambda_{2,2}$ and $\Lambda_{3,3}$ can be sampled from the same distributions as for model II, i.e. (4.1) and (4.2).

Finally, proposals for the coprecision can drawn from a normal distribution similar to (4.18):

$$\tilde{\Lambda}_{1,4} | \mathbf{X}, \mathbf{M}, M, \mu, \Lambda_{4,4} \sim \mathrm{N}(\nu_{R,1}, \pi_{R,1}) \tag{4.74}$$

The parameters are defined as

$$\pi_{R,1} = \frac{1}{\Lambda_{4,4}} \sum_{i=1}^{M} (M^i - 1)(\delta_1^{i,1})^2 \tag{4.75}$$

$$\nu_{R,1} = -\frac{1}{\pi_{R,1}} \sum_{i=1}^{M} \delta_1^{i,1} \sum_{j=1}^{M^i} \delta_4^{i,j}$$

The resulting MH-sampler has an acceptance probability of the same form as (4.19), provided the precision matrix remains positive definite.

4.3.7 Sampling ζ

As ζ is the scale parameter defining the precision $p(\zeta, \tau)$ of the observation errors, it is convenient to perform the update in terms of $p(\zeta, \tau)$.

If we denote the index set of non-zero observations by

$$G := \{t \in \{1, \dots, T\} : Z_t > 0\} \tag{4.76}$$

equations (3.13) and (3.29) almost yield a Gamma posterior distribution for $p(\zeta, \tau)$:

$$p_{p(\zeta,\tau)}(x | \mathbf{Y}, \mathbf{Z}) \propto p_{p(\zeta,\tau)}(x) p_{\mathbf{Z}}(\bullet | \mathbf{Y}, p(\zeta, \tau) = x) \tag{4.77}$$

$$\propto x^{s + \frac{|G|}{2} - 1} \exp \left(-sx - \frac{x}{2} \sum_{t \in G} (\log(Z_t) - \log(Y_t))^2 - \frac{|G|}{8x} \right)$$

We can sample from this distribution using a MH-step with proposals $\tilde{p}(\zeta, \tau)$ drawn from

$$\tilde{p}(\zeta, \tau) | \mathbf{Z}, \mathbf{Y} \sim \mathrm{Gamma} \left(s + \frac{|G|}{2}, s + \frac{1}{2} \sum_{t \in G} (\log(Z_t) - \log(Y_t))^2 \right) \tag{4.78}$$

and acceptance probability determined by

$$\alpha(p(\zeta,\tau),\tilde{p}(\zeta,\tau)) := \exp\left(\frac{|G|}{8p(\zeta,\tau)} - \frac{|G|}{8\tilde{p}(\zeta,\tau)}\right) \tag{4.79}$$

If the proposal is accepted, solving (3.14) gives us the associated ζ as

$$\zeta = \frac{\tau}{\exp(p(\zeta,\tau)^{-1}) - 1} \tag{4.80}$$

4.3.8 Random walk updates for the precisions

Numerical trials show that the precision updates that do not use Gibbs sampling can get stuck at the same value for many iterations. This happens if the true posterior density deviates too much from the gamma or Wishart distributions used for generating proposals. To improve mixing, we alternate between the sampler developed above and random walk updates for the diagonal entries of Λ. During each iteration, either scheme is chosen with a 50% probability.

To ensure that the precisions remain strictly positive, we sample proposals from a lognormal distribution that is equivalent to a normal random walk on the log-parameters. The variance of the Gaussian increments is taken to be equal to the posterior variance, estimated from the 500 previous iterations of the algorithm (if available). This ensures a reasonably high acceptance probability. In fact, if the sampler is in a state where most proposals get rejected, the estimated variance decreases, leading to smaller changes and a lower rejection rate.

4.3.8.1 Model I: Precision updates for arbitrary Λ

The proposal distributions for the diagonal entries of Λ are all of the same form

$$\log(\tilde{\Lambda}_{k,k})|\Lambda_{k,k} \sim \mathrm{N}(\log(\Lambda_{k,k}), L_k) \tag{4.81}$$

where L_k is a suitable precision parameter (see above). The corresponding α's can be calculated using (4.9):

$$\alpha(\Lambda_{1,1}, \tilde{\Lambda}_{1,1}) := \left(\frac{\left|\begin{pmatrix} \tilde{\Lambda}_{1,1} & \Lambda_{1,2} \\ \Lambda_{1,2} & \Lambda_{2,2} \end{pmatrix} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{12}^{T}\right|}{|\check{\Lambda}_{11}|}\right)^{\frac{M}{2}} \frac{\tilde{\Lambda}_{1,1}}{\Lambda_{1,1}}$$

$$\times \exp\left((\Lambda_{1,1} - \tilde{\Lambda}_{1,1})\left(\sqrt{s} + \frac{1}{2}\sum_{i=1}^{M}(\delta_1^{i,1})^2\right)\right) \tag{4.82}$$

$$\alpha(\Lambda_{2,2}, \tilde{\Lambda}_{2,2}) := \left(\frac{\left| \begin{pmatrix} \Lambda_{1,1} & \Lambda_{1,2} \\ \Lambda_{1,2} & \tilde{\Lambda}_{2,2} \end{pmatrix} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{12}^{T} \right|}{|\check{\Lambda}_{11}|} \right)^{\frac{M}{2}} \frac{\tilde{\Lambda}_{2,2}}{\Lambda_{2,2}}$$

$$\times \exp\left((\Lambda_{2,2} - \tilde{\Lambda}_{2,2}) \left(\sqrt{s} + \frac{1}{2}\sum_{i=1}^{M}(\delta_2^{i,1})^2 \right) \right)$$

(4.83)

$$\alpha(\Lambda_{3,3}, \tilde{\Lambda}_{3,3}) := \left(\frac{\left| \Lambda_{11} - \Lambda_{12}\begin{pmatrix} \tilde{\Lambda}_{3,3} & \Lambda_{3,4} \\ \Lambda_{3,4} & \Lambda_{4,4} \end{pmatrix}^{-1}\Lambda_{12}^{T} \right|}{|\check{\Lambda}_{11}|} \right)^{\frac{M}{2}} \left(\frac{\left| \begin{pmatrix} \tilde{\Lambda}_{3,3} & \Lambda_{3,4} \\ \Lambda_{3,4} & \Lambda_{4,4} \end{pmatrix} \right|}{|\Lambda_{22}|} \right)^{\frac{\overline{M}}{2}} \frac{\tilde{\Lambda}_{3,3}}{\Lambda_{3,3}}$$

$$\times \exp\left(\frac{1}{2}\sum_{i=1}^{M}(M^i - 1)(\check{\delta}^i)^T\Lambda_{12}\left(\Lambda_{22}^{-1} - \begin{pmatrix} \tilde{\Lambda}_{3,3} & \Lambda_{3,4} \\ \Lambda_{3,4} & \Lambda_{4,4} \end{pmatrix}^{-1} \right)\Lambda_{12}^{T}\check{\delta}^i \right)$$

$$\times \exp\left((\Lambda_{3,3} - \tilde{\Lambda}_{3,3}) \left(\sqrt{s} + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M^i}(\delta_3^{i,j})^2 \right) \right)$$

(4.84)

$$\alpha(\Lambda_{4,4}, \tilde{\Lambda}_{4,4}) := \left(\frac{\left| \Lambda_{11} - \Lambda_{12}\begin{pmatrix} \Lambda_{3,3} & \Lambda_{3,4} \\ \Lambda_{3,4} & \tilde{\Lambda}_{4,4} \end{pmatrix}^{-1}\Lambda_{12}^{T} \right|}{|\check{\Lambda}_{11}|} \right)^{\frac{M}{2}} \left(\frac{\left| \begin{pmatrix} \Lambda_{3,3} & \Lambda_{3,4} \\ \Lambda_{3,4} & \tilde{\Lambda}_{4,4} \end{pmatrix} \right|}{|\Lambda_{22}|} \right)^{\frac{\overline{M}}{2}} \frac{\tilde{\Lambda}_{4,4}}{\Lambda_{4,4}}$$

$$\times \exp\left(\frac{1}{2}\sum_{i=1}^{M}(M^i - 1)(\check{\delta}^i)^T\Lambda_{12}\left(\Lambda_{22}^{-1} - \begin{pmatrix} \Lambda_{3,3} & \Lambda_{3,4} \\ \Lambda_{3,4} & \tilde{\Lambda}_{4,4} \end{pmatrix}^{-1} \right)\Lambda_{12}^{T}\check{\delta}^i \right)$$

$$\times \exp\left((\Lambda_{4,4} - \tilde{\Lambda}_{4,4}) \left(\sqrt{s} + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M^i}(\delta_4^{i,j})^2 \right) \right)$$

(4.85)

Note that the alphas do not depend on the value of L_k, as the lognormal proposal densities cancel out in the quotient, except for the term $\frac{\tilde{\Lambda}_{k,k}}{\Lambda_{k,k}}$. Of course, a large variance leads to a greater number of unlikely proposals and a higher rejection rate.

As for the samplers based on the Wishart distribution, any proposal needs to be rejected if the resulting Λ is not positive definite.

4.3.8.2 Model IV: Precision updates for diagonal Λ with matched means

As λ_2 and λ_3 permit a Gibbs sampler, only the other two precision parameters use the random-walk update. Proposals are drawn from

$$\log(\tilde{\lambda}_1)|\lambda_1 \sim \mathrm{N}(\log(\lambda_1), L_1) \qquad\qquad \log(\tilde{\lambda}_4)|\lambda_4 \sim \mathrm{N}(\log(\lambda_4), L_4) \tag{4.86}$$

with suitable precisions. The acceptance probabilities can be derived from (4.24) and (4.27). They depend on

$$\alpha(\lambda_1, \tilde{\lambda}_1) := \left(\frac{\tilde{\lambda}_1}{\lambda_1}\right)^{s+\frac{M}{2}} \exp\left((\lambda_1 - \tilde{\lambda}_1)\left(s + \frac{1}{2}\sum_{i=1}^{M}(\log(X_1^i) - \mu_1)^2\right)\right) \tag{4.87}$$

$$\times \exp\left(\frac{\lambda_4}{2}\left(\sum_{i=1}^{M}\sum_{j=1}^{M^i}(V^{i,j}(\lambda_1^{-1}, \sigma_4^2) - \mu_1)^2 - \sum_{i=1}^{M}\sum_{j=1}^{M^i}(V^{i,j}(\tilde{\lambda}_1^{-1}, \sigma_4^2) - \mu_1)^2\right)\right)$$

$$\alpha(\lambda_4, \tilde{\lambda}_4) := \left(\frac{\tilde{\lambda}_4}{\lambda_4}\right)^{s+\frac{\overline{M}}{2}} \exp\left(\frac{\overline{M}}{8\lambda_4} - \frac{\overline{M}}{8\tilde{\lambda}_4}\right) \tag{4.88}$$

$$\times \exp\left((\lambda_4 - \tilde{\lambda}_4)\left(s + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M^i}\left(\log(X_2^{i,j}) - \mu_1 - \frac{1}{2\lambda_1} + \log(K)\right)^2\right)\right)$$

Since Λ is a diagonal matrix, positive definiteness is assured.

4.3.8.3 Model III: Precision updates for arbitrary Λ with matched means

As for the full model without mean matching, we use (4.81) as our proposal distribution. The acceptance probabilities can be derived from (4.37) and (4.49). They depend on

$$\alpha(\Lambda_{1,1}, \tilde{\Lambda}_{1,1}) := \left(\frac{\left|\begin{pmatrix}\tilde{\Lambda}_{1,1} & \Lambda_{1,2} \\ \Lambda_{1,2} & \Lambda_{2,2}\end{pmatrix} - \frac{\Lambda_{R,12}\Lambda_{R,12}^T}{\Lambda_{3,3}}\right|}{|\check{\Lambda}_{R,11}|}\right)^{\frac{M}{2}} \frac{\tilde{\Lambda}_{1,1}}{\Lambda_{1,1}} \tag{4.89}$$

$$\times \exp\left((\Lambda_{1,1} - \tilde{\Lambda}_{1,1})\left(\sqrt{s} + \frac{1}{2}\sum_{i=1}^{M}(\delta_1^{i,1})^2\right)\right)$$

$$\times \exp\left(\frac{\Lambda_{4,4}}{2}\sum_{i=1}^{M}\sum_{j=1}^{M^i}((V^{i,j}(\sigma_1^2, \sigma_4^2) - \mu_1)^2 - (V^{i,j}(\tilde{\sigma}_1^2, \sigma_4^2) - \mu_1)^2)\right)$$

$$\alpha(\Lambda_{2,2}, \tilde{\Lambda}_{2,2}) := \left(\frac{\left| \begin{pmatrix} \Lambda_{1,1} & \Lambda_{1,2} \\ \Lambda_{1,2} & \tilde{\Lambda}_{2,2} \end{pmatrix} - \frac{\Lambda_{R,12}\Lambda_{R,12}^T}{\Lambda_{3,3}} \right|}{|\check{\Lambda}_{R,11}|} \right)^{\frac{M}{2}} \frac{\tilde{\Lambda}_{2,2}}{\Lambda_{2,2}} \tag{4.90}$$

$$\times \exp \left((\Lambda_{2,2} - \tilde{\Lambda}_{2,2}) \left(\sqrt{s} + \frac{1}{2} \sum_{i=1}^{M} (\delta_2^{i,1})^2 \right) \right)$$

$$\times \exp \left(\frac{\Lambda_{4,4}}{2} \sum_{i=1}^{M} \sum_{j=1}^{M^i} ((V^{i,j}(\sigma_1^2, \sigma_4^2) - \mu_1)^2 - (V^{i,j}(\tilde{\sigma}_1^2, \sigma_4^2) - \mu_1)^2) \right)$$

$$\alpha(\Lambda_{3,3}, \tilde{\Lambda}_{3,3}) := \left(\frac{\left| \Lambda_{11} - \frac{\Lambda_{R,12}\Lambda_{R,12}^T}{\tilde{\Lambda}_{3,3}} \right|}{|\check{\Lambda}_{R,11}|} \right)^{\frac{M}{2}} \left(\frac{\tilde{\Lambda}_{3,3}}{\Lambda_{3,3}} \right)^{\frac{\overline{M}}{2}+1} \tag{4.91}$$

$$\times \exp \left(\left(\frac{1}{2\Lambda_{3,3}} - \frac{1}{2\tilde{\Lambda}_{3,3}} \right) \sum_{i=1}^{M} (M^i - 1)(\Lambda_{R,12}^T \check{\delta}^i)^2 \right)$$

$$\times \exp \left((\Lambda_{3,3} - \tilde{\Lambda}_{3,3}) \left(\sqrt{s} + \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M^i} (\delta_3^{i,j})^2 \right) \right)$$

$$\alpha(\Lambda_{4,4}, \tilde{\Lambda}_{4,4}) := \left(\frac{\tilde{\Lambda}_{4,4}}{\Lambda_{4,4}} \right)^{s+\frac{\overline{M}}{2}} \exp \left(\frac{\overline{M}}{8\Lambda_{4,4}} - \frac{\overline{M}}{8\tilde{\Lambda}_{4,4}} \right) \tag{4.92}$$

$$\times \exp \left((\Lambda_{4,4} - \tilde{\Lambda}_{4,4}) \left(s + \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M^i} \left(\log(X_2^{i,j}) - \mu_1 - \frac{\sigma_1^2}{2} + \log(K) \right)^2 \right) \right)$$

As usual, any proposal has to be rejected if the resulting Λ is no longer positive definite.

4.3.8.4   Model V: Precision updates for uncorrelated cell durations

For model V, we can use the same lognormal proposals as for the full model, except that $\Lambda_{3,3}$ can be sampled from its gamma posterior distribution. If we draw proposals from (4.81) for $k \in \{1, 2, 4\}$, the acceptance probabilities for the MH-samples can be derived as in (4.82), (4.83), and (4.85):

$$\alpha(\Lambda_{1,1}, \tilde{\Lambda}_{1,1}) := \left(\frac{\left| \begin{pmatrix} \tilde{\Lambda}_{1,1} & \Lambda_{1,2} \\ \Lambda_{1,2} & \Lambda_{2,2} \end{pmatrix} - \frac{\Lambda_{R,12}\Lambda_{R,12}^T}{\Lambda_{4,4}} \right|}{|\check{\Lambda}_{11}|} \right)^{\frac{M}{2}} \frac{\tilde{\Lambda}_{1,1}}{\Lambda_{1,1}} \tag{4.93}$$

$$\times \exp \left((\Lambda_{1,1} - \tilde{\Lambda}_{1,1}) \left(\sqrt{s} + \frac{1}{2} \sum_{i=1}^{M} (\delta_1^{i,1})^2 \right) \right)$$

$$\alpha(\Lambda_{2,2}, \tilde{\Lambda}_{2,2}) := \left(\frac{\left| \begin{pmatrix} \Lambda_{1,1} & \Lambda_{1,2} \\ \Lambda_{1,2} & \tilde{\Lambda}_{2,2} \end{pmatrix} - \frac{\Lambda_{R,12}\Lambda_{R,12}^T}{\Lambda_{4,4}} \right|}{|\check{\Lambda}_{11}|} \right)^{\frac{M}{2}} \frac{\tilde{\Lambda}_{2,2}}{\Lambda_{2,2}} \tag{4.94}$$

$$\times \exp\left((\Lambda_{2,2} - \tilde{\Lambda}_{2,2}) \left(\sqrt{s} + \frac{1}{2} \sum_{i=1}^{M} (\delta_2^{i,1})^2 \right) \right)$$

$$\alpha(\Lambda_{4,4}, \tilde{\Lambda}_{4,4}) := \left(\frac{\left| \Lambda_{11} - \frac{\Lambda_{R,12}\Lambda_{R,12}^T}{\tilde{\Lambda}_{4,4}} \right|}{|\check{\Lambda}_{11}|} \right)^{\frac{M}{2}} \left(\frac{\tilde{\Lambda}_{4,4}}{\Lambda_{4,4}} \right)^{\frac{\overline{M}}{2}+1} \tag{4.95}$$

$$\times \exp\left(\left(\frac{1}{2\Lambda_{4,4}} - \frac{1}{2\tilde{\Lambda}_{4,4}} \right) \sum_{i=1}^{M} (M^i - 1)(\Lambda_{R,12}^T \check{\delta}^i)^2 \right)$$

$$\times \exp\left((\Lambda_{4,4} - \tilde{\Lambda}_{4,4}) \left(\sqrt{s} + \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M^i} (\delta_4^{i,j})^2 \right) \right)$$

Once again, we need to reject any proposal that leads to Λ not being positive definite.

4.3.8.5 Model VI: Precision updates for limited correlation

Model VI uses the same lognormal proposals as the full model, but only for $\Lambda_{1,1}$ and $\Lambda_{4,4}$, as the other two precisions permit a Gibbs sampler. If we draw proposals from (4.81) for $k \in \{1, 4\}$, the acceptance probabilities can be derived as in (4.82) and (4.85):

$$\alpha(\Lambda_{1,1}, \tilde{\Lambda}_{1,1}) := \left(\frac{\tilde{\Lambda}_{1,1} - \frac{\Lambda_{1,4}^2}{\Lambda_{4,4}}}{\check{\Lambda}_{1,1}} \right)^{\frac{M}{2}} \frac{\tilde{\Lambda}_{1,1}}{\Lambda_{1,1}} \tag{4.96}$$

$$\times \exp\left((\Lambda_{1,1} - \tilde{\Lambda}_{1,1}) \left(\sqrt{s} + \frac{1}{2} \sum_{i=1}^{M} (\delta_1^{i,1})^2 \right) \right)$$

$$\alpha(\Lambda_{4,4}, \tilde{\Lambda}_{4,4}) := \left(\frac{\Lambda_{1,1} - \frac{\Lambda_{1,4}^2}{\tilde{\Lambda}_{4,4}}}{\check{\Lambda}_{11}} \right)^{\frac{M}{2}} \left(\frac{\tilde{\Lambda}_{4,4}}{\Lambda_{4,4}} \right)^{\frac{\overline{M}}{2}+1} \tag{4.97}$$

$$\times \exp\left(\left(\frac{1}{2\Lambda_{4,4}} - \frac{1}{2\tilde{\Lambda}_{4,4}} \right) \Lambda_{1,4}^2 \sum_{i=1}^{M} (M^i - 1)(\check{\delta}_1^i)^2 \right)$$

$$\times \exp\left((\Lambda_{4,4} - \tilde{\Lambda}_{4,4}) \left(\sqrt{s} + \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M^i} (\delta_4^{i,j})^2 \right) \right)$$

As before, we need to check whether Λ remains positive definite and reject any proposal that destroys the property.

4.3.8.6   The random-walk sampler for ζ

As in Section 4.3.7, we update the precision $p(\zeta, \tau)$ instead of the scale parameter ζ. If we draw a proposal from

$$\log(p(\tilde{\zeta}, \tau)) | p(\zeta, \tau) \sim \mathrm{N}(\log(p(\zeta, \tau)), L_\zeta) \tag{4.98}$$

with suitable L_ζ, the acceptance probability for the move depends on

$$\alpha(p(\zeta, \tau), p(\tilde{\zeta}, \tau)) =: \left(\frac{p(\tilde{\zeta}, \tau)}{p(\zeta, \tau)} \right)^{s + \frac{|G|}{2}} \exp\left(\frac{|G|}{8 p(\zeta, \tau)} - \frac{|G|}{8 p(\tilde{\zeta}, \tau)} \right) \tag{4.99}$$

$$\times \exp\left((p(\zeta, \tau) - p(\tilde{\zeta}, \tau)) \left(s + \frac{1}{2} \sum_{t \in G} (\log(Z_t) - \log(Y_t))^2 \right) \right)$$

4.4   Weight estimators

One drawback of the lognormal point-process model is the lack of analytic solutions for the distribution of $\mathbf{M}$ and M. Exponentially distributed lags would yield Poisson weights which have a simple explicit form, but we have to resort to a smoothed Monte-Carlo estimator.

To estimate the distribution of the number of storms M, we draw S independent samples of lognormals $\{X_2^{i,(1)}\}, \ldots, \{X_2^{i,(S)}\}$ from the marginal distribution of the storm lags

$$\log(X_2^{i,(s)}) | \theta \sim \mathrm{N}\left(\mu_2, (\check{\Lambda}_{11})_{1,1} - \frac{(\check{\Lambda}_{11})_{1,2}^2}{(\check{\Lambda}_{11})_{2,2}} \right) \tag{4.100}$$

Then, we count how many times exactly m storms occur during the observation period:

$$\kappa_m := \left| \left\{ s \in \{1, \ldots, S\} : \sum_{i=1}^{m-1} X_2^{i,(s)} \leq T\tau < \sum_{i=1}^{m} X_2^{i(s)} \right\} \right| \tag{4.101}$$

Due to the consistency condition for the storm lags (3.5), the relative frequencies $\frac{\kappa_m}{S}$ are estimators for the weights of the distribution of M given in (3.37).

The empirical mode of the weight distribution is denoted by

$$m_+ := \arg\max_{m \geq 1} \kappa_m \tag{4.102}$$

For some initial bandwidth $W \geq 1$ and a scaling constant $L \geq 1$, we estimate weights by local averaging. The bandwidth increases linearly with the distance to the empirical mode:

$$\hat{q}_M(m | \theta) := \frac{\sum_{\max(1, m - d_m)}^{m + d_m} \kappa_m}{S(1 + 2d_m)} \tag{4.103}$$

$$d_m := W + \left\lfloor \frac{|m - m_+|}{L} \right\rfloor \tag{4.104}$$

We divide all values by the full range $(1 + 2d_m)$, even if we are near 1 where fewer counts are available for averaging. Dividing by the actual number of counts can result in an artificial mode at 1, which may seriously distort the behavior of the algorithm.

As the weights $\hat{q}$ do in general not sum up to 1, we rescale to get estimators

$$\hat{p}_M(m|\theta) := \frac{\hat{q}_M(m|\theta)}{\sum_{j=1}^{\infty} \hat{q}_M(j|\theta)} \tag{4.105}$$

A similar scheme can be used to obtain estimators for the probability weights of cell counts per storm. The major difference is that the time horizon is drawn from the distribution of the storm durations, and that the cell lags have to be sampled conditional on this duration.

Remark 4.3 *Local averaging with variable bandwidth is necessary for computational reasons, as the samples are highly dispersed. This is especially problematic in the case of cell counts, where the length of the observation period is also variable. Thus, a small overall bandwidth yields an estimated weight of 0 for many values in the tails of the distribution. On the other hand, a large bandwidth adapted to the tails underestimates the weights near the mode(s) due to oversmoothing.*

4.5 Storm structure transforms

The quantities $\Psi := (\mathbf{Y}, \mathbf{D}, \mathbf{X}, \mathbf{M}, M)$ and the unknown observations $\{Z_t\}_{t \in U}$ are updated using a reversible jump MH-scheme. One iteration of this method works as follows:

1. Choose a transform from a specified list of operations on Ψ.

2. Apply the transform to the current state to generate a proposal process.

3. Accept the change with a probability given by the usual quotient of likelihoods and proposal distributions.

Reversible-jump MCMC was introduced by Green [22]. Some applications are discussed e.g. in [7] or in the book by Frühwirth-Schnatter [12], while Granville and Smith apply it to rainfall disaggregation [21] using an extended version of the model found in [38].

The critical issue is of course the choice of transforms, which must satisfy several conditions:

- Any conceivable state of the model can be reached by a finite sequence of transforms.

- The process needs to remain consistent (as per Definition 3.1) at all times.

- Each transform needs to be reversible (by itself or another transform), so the probabilities to move from one state to its proposed successor and vice versa are non-zero.

Furthermore, the change per transform should be relatively small, so that the proposed moves have a reasonably high acceptance probability.

We use six operations to manipulate the storms and cells:

1. Add a cell to an existing storm.

2. Remove a cell from an existing storm.

3. Split one storm into two adjacent ones.

4. Merge two adjacent storms into one.

5. Update the duration of a storm.

6. Update the depth and duration of a cell.

Remark 4.4 *Operation 1 is reversible by 2 and vice versa. The same is true for 3 and 4. Operation 5 and 6 can each reverse themselves. The first four transforms already allow us to generate any number of storms and cells in arbitrary positions. The last two updates are included for numerical reasons:*

The algorithm is able to identify the position of the storms much more quickly than that of the cells. Thus, mergers and splits become rare events after a few iterations and another transform is necessary to allow duration to respond to changes in θ.

On the other hand, cell transforms usually have lower acceptance probabilities than changes to the storm structure, as they have to "explain" the observations $\mathbf{Z}$ via the aggregates. During the initial iterations of the algorithm, the average intensity of the cell process tends to be less than its expected value ι. Operation 6 improves the fit of the cell process to the observations and thus the rate of convergence of the average towards ι.

In our implementation of the algorithm, operations 1 to 5 are performed in a random order for several thousand iterations:

- Operation 1 and 2 are each chosen with probability $p_1 > 0$.

- Operation 3 and 4 are each chosen with probability $p_2 > 0$.

- Operation 5 is chosen with probability $p_3 > 0$.

Obviously, we need to require $2p_1 + 2p_2 + p_3 = 1$. Pairs of reverse transforms get the same p_i so it cancels when calculating the quotient for the acceptance probability. As the cell process is more difficult to fit, p_1 for the cell transforms should be the largest of the three. For the same reason, operation 6 is not included in the random scheme. Instead, it is applied several times per cell in a separate update step.

The remainder of this section is a detailed description of the six transforms. All quantities associated with a proposed process $\tilde{\Psi}$ are marked with a tilde. To express the acceptance probabilities for a move from one state to the next, we use the index set

$$\tilde{T} := \{t \in \{1, \ldots, T\} : \tilde{Y}_t \neq Y_t\} \tag{4.106}$$

of observation periods where the aggregates change due to the transform.

The acceptance probabilities for the moves are given for the full model. If using one of the reduced models II–VI, the expressions can often be simplified considerably. But as this is mainly an issue of how to implement the algorithm efficiently, we do not give model-specific versions of the α's here.

4.5.1 Add a cell

This operation adds a new cell to an existing storm without changing the absolute position of other cells. Select a storm index i_0 and an index $j_0 - 1$ for the predecessor of the new cell from the Laplace distributions on $\{1, \ldots, M\}$ and $\{1, \ldots, M^{i_0}\}$ respectively. The proposal $\tilde{\Psi}$ is constructed from Ψ in the following manner:

- $\tilde{M} := M$, $\tilde{M}^{i_0} := M^{i_0} + 1$, and for $i \neq i_0$: $\tilde{M}^i := M^i$.

- For all i: $\tilde{X}^i := X^i$.

- For $i \neq i_0$ and all j: $\tilde{X}^{i,j} := X^{i,j}$, $\tilde{D}^{i,j} := D^{i,j}$.

- For $j < j_0 - 1$: $\tilde{X}^{i_0,j} := X^{i_0,j}$, $\tilde{D}^{i_0,j} := D^{i_0,j}$.

- Sample $\tilde{X}^{i_0,j_0}$ from (3.10) conditional on X^{i_0}, and sample $\tilde{D}^{i_0,j_0}$ from (3.12).

- $\tilde{X}^{i_0,j_0-1} := \begin{pmatrix} X_1^{i_0,j_0-1} \\ X_2^{i_0,j_0-1} - \tilde{X}_2^{i_0,j_0} \end{pmatrix}$, $\tilde{D}^{i_0,j_0-1} := D^{i_0,j_0-1}$.

- For $j > j_0$: $\tilde{X}^{i_0,j} := X^{i_0,j-1}$, $\tilde{D}^{i_0,j} := D^{i_0,j-1}$.

- For all t: $\tilde{Y}_t := Y_t + |[\tilde{O}^{i_0,j_0}, \tilde{O}^{i_0,j_0} + \tilde{X}_1^{i_0,j_0}] \cap [(t-1)\tau, t\tau]| \tilde{D}^{i_0,j_0}$.

New values $\tilde{Z}_t$ for $t \in \tilde{T} \cap U$ are sampled from (3.13).

The move can be rejected outright if $\tilde{X}_2^{i_0,j_0-1} < 0$ or if cell j_0 starts after the end of storm i_0, as these would be inconsistent with the model. Else, the transition probability is determined

by (3.32) and the probability of adding/subtracting the chosen cell:

$$
\begin{aligned}
\alpha_1(\Psi, \mathbf{Z}, \tilde{\Psi}, \tilde{\mathbf{Z}}) &= \frac{p_\mathbf{Z}(\tilde{\mathbf{Z}}|\tilde{\mathbf{Y}}, \theta)p_\mathbf{D}(\tilde{\mathbf{D}})p_\mathbf{X}(\tilde{\mathbf{X}}|\tilde{\mathbf{M}}, \tilde{M}, \theta)p_\mathbf{M}(\tilde{\mathbf{M}}|\tilde{M}, \theta)}{p_\mathbf{Z}(\mathbf{Z}|\mathbf{Y}, \theta)p_\mathbf{D}(\mathbf{D})p_\mathbf{X}(\mathbf{X}|\mathbf{M}, M, \theta)p_\mathbf{M}(\mathbf{M}|M, \theta)} \\[2mm]
&\quad \times \frac{\frac{p_1 \prod_{t \in \tilde{T} \cap U} p_{Z_t}(Z_t|Y_t, \theta)}{\tilde{M}(\tilde{M}^{i_0} - 1)}}{\frac{p_1 p_{X^{i,j}}(\tilde{X}^{i_0, j_0}|X^{i_0}, \theta) \prod_{t \in \tilde{T} \cap U} p_{Z_t}(\tilde{Z}_t|\tilde{Y}_t, \theta)}{2\iota M M^{i_0}}} \\[2mm]
&= \left(\prod_{t \in \tilde{T} \setminus U} \frac{p_{Z_t}(Z_t|\tilde{Y}_t, \theta)}{p_{Z_t}(Z_t|Y_t, \theta)} \right) \frac{p_{X^{i,j}}(\tilde{X}^{i_0, j_0 - 1}|X^{i_0}, \theta)p_{M^i}(\tilde{M}^{i_0}|\theta)}{p_{X^{i,j}}(X^{i_0, j_0 - 1}|X^{i_0}, \theta)p_{M^i}(M^{i_0}|\theta)}
\end{aligned}
\tag{4.107}
$$

The densities for Z_t and $X^{i,j}$ can be derived explicitly from the distributions in (3.13) and (3.10), and we have seen how to get weights for M^i in Section 4.4.

Remark 4.5 *Note that a cell may never be added before the first cell of a storm. This is consistent with the next operation, as the first cell of a storm can never be removed.*

4.5.2   Remove a cell

This transform removes an existing cell without changing the absolute position of other cells. Select a storm index i_0 and an index $j_0 - 1$ for the predecessor of the cell to be removed from the Laplace distributions on $\{1, \ldots, M\}$ and $\{1, \ldots, M^{i_0} - 1\}$ respectively. The move can be rejected outright if $M^{i_0} = 1$, as no cell can have fewer than 1 cell. Construct $\tilde{\Psi}$ from Ψ as follows:

- $\tilde{M} := M$, $\tilde{M}^{i_0} := M^{i_0} - 1$, and for $i \neq i_0$: $\tilde{M}^i := M^i$.

- For all i: $\tilde{X}^i := X^i$.

- For $i \neq i_0$ and all j: $\tilde{X}^{i,j} := X^{i,j}$, $\tilde{D}^{i,j} := D^{i,j}$.

- For $j < j_0 - 1$: $\tilde{X}^{i_0,j} := X^{i_0,j}$, $\tilde{D}^{i_0,j} := D^{i_0,j}$.

- $\tilde{X}^{i_0, j_0 - 1} := \begin{pmatrix} X_1^{i_0, j_0 - 1} \\ X_2^{i_0, j_0 - 1} + X_2^{i_0, j_0} \end{pmatrix}$, $\tilde{D}^{i_0, j_0 - 1} := D^{i_0, j_0 - 1}$.

- For $j \geq j_0$: $\tilde{X}^{i_0,j} := X^{i_0,j+1}$, $\tilde{D}^{i_0,j} := D^{i_0,j+1}$.

- For all t: $\tilde{Y}_t := Y_t - |[O^{i_0,j_0}, O^{i_0,j_0} + X_1^{i_0,j_0}] \cap [(t-1)\tau, t\tau]|D^{i_0,j_0}$.

Additionally, new values $\tilde{Z}_t$ have to be sampled from (3.13) for $t \in \tilde{T} \cap U$.

The transition probability is determined by

$$
\alpha_2(\Pi, \mathbf{Z}, \tilde{\Pi}, \tilde{\mathbf{Z}}) = \alpha_1(\tilde{\Pi}, \tilde{\mathbf{Z}}, \Pi, \mathbf{Z})^{-1}
\tag{4.108}
$$

Remark 4.6 *The first cell of a storm can never be removed, as we require each storm to start with a cell (see Remark 3.1).*

4.5.3 Split one storm into two adjacent ones

This operation splits one storm into two adjacent ones without affecting the position of other storms or of any of the cells. Select a storm index i_0 from the Laplace distribution on $\{1, \ldots, M\}$. The cells of this storm will be split among the new storms i_0 and $i_0 + 1$. First, define

$$k_1 := \begin{cases} \max\{j : j \in \{2, \ldots, M^{i_0}\} \wedge O^{i_0,j} < O^{i_0+1}\} & \text{if } i_0 < M \\ M^{i_0} & \text{if } i_0 = M \end{cases} \tag{4.109}$$

If no such k_1 exists, reject the move as no split into adjacent storms is possible. Else, sample $U \sim \mathrm{U}(0, O^{i_0,k_1} - O^{i_0})$ and take as first cell for the new storm $i_0 + 1$ the one with index

$$j_1 := \arg\min_j \{O^{i_0,j} : O^{i_0,j} - O^{i_0} > U\} \in \{2, \ldots, k_1\} \tag{4.110}$$

The probability of choosing j_1 is

$$p(j_1 = j) = \frac{X_2^{i_0,j-1}}{O^{i_0,k_1} - O^{i_0}} \tag{4.111}$$

I.e., it is more likely that the storm is split at a position where the cells are far apart.

The last cell m_0 for the new storm i_0 is selected from $\{j_1 - 1, j_1 + 1, \ldots, M^{i_0}\}$ with geometrically decreasing probability. Thus, if

$$g : \{0, \ldots, M^{i_0} - j_1\} \to \{j_1 - 1, j_1 + 1, \ldots, M^{i_0}\} \tag{4.112}$$

$$j \mapsto \begin{cases} j_1 - 1 & \text{if } j = 0 \\ j_1 + j & \text{if } j > 0 \end{cases}$$

the weights for choosing m_0 are

$$P(m_0 = g(j)) = \frac{\left(\frac{1}{2}\right)^{j+1}}{1 - \left(\frac{1}{2}\right)^{M^{i_0} - j_1 + 1}} \tag{4.113}$$

This way, splits that result in little overlap between the new storms are proposed more often. As we expect clustering of wet periods to be a evident in the data, these should yield a higher acceptance rate.

If $m_0 > j_1 + 1$, we need to distribute the cells with intermediate indices between the two new storms. Sample $\beta \in \{0,1\}^{m_0 - j_1 - 1}$ with independent Bernoulli components $\beta_j \sim \mathrm{B}\left(1, \frac{1}{2}\right)$. Partition the cells of storm i_0 into two new storms as follows:

$$S_0 := \{1, \ldots, j_1 - 1\} \cup \{j_1 + l : \beta_l = 0\} \cup \{m_0\} \tag{4.114}$$

$$S_1 := \{j_1\} \cup \{j_1 + l : \beta_l = 1\} \cup \{m_0 + 1, \ldots, M^{i_0}\}$$

and define counts

$$h_l : \{1, \ldots, |S_l|\} \to S_l, \quad l \in \{0, 1\} \tag{4.115}$$

such that $O^{i_0,h_l(1)} < O^{i_0,h_l(2)} < \cdots < O^{i_0,h_l(|S_l|)}$. The last cell of the new storm $i_0 + 1$ has the index $m_1 := \max(S1)$. We can now define the proposal process $\tilde{\Psi}$ via

- $\tilde{M} := M + 1$ and

$$\tilde{M}^i := \begin{cases} M^i & \text{if } 1 \le i < i_0 \\ |S_0| & \text{if } i = i_0 \\ |S_1| & \text{if } i = i_0 + 1 \\ M^{i-1} & \text{if } i_0 + 1 < i \le \tilde{M} \end{cases} \tag{4.116}$$

- For $i < i_0$ and all j: $\tilde{X}^i := X^i$, $\tilde{X}^{i,j} := X^{i,j}$, $\tilde{D}^{i,j} = D^{i,j}$.

- For $i > i_0 + 1$ and all j: $\tilde{X}^i := X^{i-1}$, $\tilde{X}^{i,j} := X^{i-1,j}$, $\tilde{D}^{i,j} = D^{i-1,j}$.

- If $m_0 < m_1$, sample $X_2^{i_0,\tilde{M}^{i_0}}$ from its marginal distribution

$$\log(X_2^{i_0,\tilde{M}^{i_0}})|\theta \sim \mathrm{N}(\mu_4, (\Lambda^{-1})_{4,4}^{-1}) \tag{4.117}$$

and draw $V \sim \mathrm{U}(0, X_2^{i_0,\tilde{M}^{i_0}})$. Now define

$$\begin{aligned} \tilde{X}_1^{i_0} &:= O^{i_0,m_0} + V - O^{i_0} & \tilde{X}_1^{i_0+1} &:= O^{i_0} + X_1^{i_0} - O^{i_0,j_1} \\ \tilde{X}_1^{i_0,\tilde{M}^{i_0}} &:= X_1^{i_0,m_0} & \tilde{X}_1^{i_0+1,\tilde{M}^{i_0+1}} &:= X^{i_0+1,m_1} \\ \tilde{D}^{i_0,\tilde{M}^{i_0}} &:= D^{i_0,m_0} & \tilde{D}^{i_0+1,\tilde{M}^{i_0+1}} &:= D^{i_0,m_1} \end{aligned} \tag{4.118}$$

- If $m_0 > m_1$, sample $X_2^{i_0+1,\tilde{M}^{i_0+1}}$ from its marginal distribution (same as (4.117)), draw $V \sim \mathrm{U}(0, X_2^{i_0+1,\tilde{M}^{i_0+1}})$, and set

$$\begin{aligned} \tilde{X}_1^{i_0} &:= X_1^{i_0} & \tilde{X}_1^{i_0+1} &:= O^{i_0,m_1} + V - O^{i_0,j_1} \\ \tilde{X}^{i_0,\tilde{M}^{i_0}} &:= X^{i_0,m_0} & \tilde{X}_1^{i_0+1,\tilde{M}^{i_0+1}} &:= X_1^{i_0+1,m_1} \\ \tilde{D}^{i_0,\tilde{M}^{i_0}} &:= D^{i_0,m_0} & \tilde{D}^{i_0+1,\tilde{M}^{i_0+1}} &:= D^{i_0,m_1} \end{aligned} \tag{4.119}$$

- Set $\tilde{X}_2^{i_0} := O^{i_0,j_1} - O^{i_0}$ and $\tilde{X}_2^{i_0+1} := O^{i_0} + X_1^{i_0} - O^{i_0,j_1}$.

- For $j < \tilde{M}^{i_0}$:

$$\tilde{X}^{i_0,j} := \begin{pmatrix} X_1^{i_0,h_0(j)} \\ O^{i_0,h_0(j+1)} - O^{i_0,h_0(j)} \end{pmatrix} \qquad \tilde{D}^{i_0,j} := D^{i_0,h_0(j)} \tag{4.120}$$

- For $j < \tilde{M}^{i_0+1}$:

$$\tilde{X}^{i_0+1,j} := \begin{pmatrix} X_1^{i_0,h_1(j)} \\ O^{i_0,h_1(j+1)} - O^{i_0,h_1(j)} \end{pmatrix} \qquad \tilde{D}^{i_0+1,j} := D^{i_0,h_1(j)} \tag{4.121}$$

It is convenient to define

$$L := \begin{cases} X_2^{i_0,\tilde{M}^{i_0}} & \text{if } m_0 < m_1 \\ X_2^{i_0+1,M^{\tilde{i_0}+1}} & \text{if } m_0 > m_1 \end{cases} \tag{4.122}$$

The acceptance probability is given by

$$
\alpha_3(\Pi, \tilde{\Pi}) = \frac{p_{\mathbf{X}}(\tilde{\mathbf{X}}|\tilde{\mathbf{M}}, \tilde{M}, \theta) p_{\mathbf{M}}(\tilde{\mathbf{M}}|\tilde{M}, \theta) p_M(\tilde{M}|\theta)}{p_{\mathbf{X}}(\mathbf{X}|\mathbf{M}, M, \theta) p_{\mathbf{M}}(\mathbf{M}|M, \theta) p_M(M|\theta)} \tag{4.123}
$$

$$
\times \frac{p_3}{(\tilde{M}-1)} \frac{M(O^{i_0,k_1} - O^{i_0})\left(1 - \left(\frac{1}{2}\right)^{M^{i_0}-j_1+1}\right) L}{p_3 X_2^{i_0,j_1-1} \left(\frac{1}{2}\right)^{g^{-1}(m_0)+1+\max(0,m_0-j_1-1)} p(L|\theta)}
$$

$$
= \frac{\left(\prod_{j=1}^{\tilde{M}^{i_0}} p_{X^{i,j}}(\tilde{X}^{i_0,j}|\tilde{X}^{i_0}, \theta)\right)\left(\prod_{j=1}^{\tilde{M}^{i_0+1}} p_{X^{i,j}}(\tilde{X}^{i_0+1,j}|\tilde{X}^{i_0+1}, \theta)\right)}{\prod_{j=1}^{M^{i_0}} p_{X^{i,j}}(X^{i_0,j}|X^{i_0}, \theta)}
$$

$$
\times \frac{p_{X^i}(\tilde{X}^{i_0}|\theta) p_{X^i}(\tilde{X}^{i_0+1}|\theta)}{p_{X^i}(X^{i_0}|\theta)} \frac{p_{M^i}(\tilde{M}^{i_0}|\theta) p_{M^i}(\tilde{M}^{i_0+1}|\theta)}{p_{M^i}(M^{i_0}|\theta)}
$$

$$
\times \frac{p_M(\tilde{M}|\theta)}{p_M(M|\theta)} \frac{(O^{i_0,k_1} - O^{i_0})\left(1 - \left(\frac{1}{2}\right)^{M^{i_0}-j_1+1}\right) L}{X_2^{i_0,j_1-1} \left(\frac{1}{2}\right)^{g^{-1}(m_0)+1+\max(0,m_0-j_1-1)} p(L|\theta)}
$$

Remark 4.7   *1. The construction in (4.118) and (4.119) ensures the consistency condition (3.5). Note that each storm ends before the time where the lag of its last cell would place its successor.*

2. It is possible to drop storm M^{i_0+1} from the fitted process in case it starts after time T.

4.5.4   Merge two adjacent storms into one

This transform merges two adjacent storms into a single one without affecting the position of other storms or of any of the cells. Select a storm index i_0 from the Laplace distribution on $\{1, \ldots, M-1\}$. This storm will be merged with its successor. If we define a counting function for the cell indices

$$
p : \{1, \ldots, M^{i_0} + M^{i_0+1}\} \to \{(i_0, 1), \ldots (i_0, M^{i_0}), (i_0+1, 1), \ldots, (i_0+1, M^{i_0+1})\} \tag{4.124}
$$

satisfying $O^{p(1)} < O^{p(2)} < \cdots < O^{p(M^{i_0}+M^{i_0+1})}$, we can construct $\tilde{\Pi}$ from Π as follows:

- $\tilde{M} := M - 1$ and

$$
\tilde{M}^i := \begin{cases} M^i & \text{if } 1 \leq i < i_0 \\ M^{i_0} + M^{i_0+1} & \text{if } i = i_0 \\ M^{i+1} & \text{if } i_0 + 1 \leq i \leq \tilde{M} \end{cases} \tag{4.125}
$$

- For $i < i_0$ and all j: $\tilde{X}^i := X^i$, $\tilde{X}^{i,j} := X^{i,j}$, $\tilde{D}^{i,j} = D^{i,j}$.

- For $i > i_0$ and all j: $\tilde{X}^i := X^{i+1}$, $\tilde{X}^{i,j} := X^{i+1,j}$, $\tilde{D}^{i,j} = D^{i+1,j}$.

- $\tilde{X}^{i_0} := \begin{pmatrix} \max(X_1^{i_0}, O^{i_0+1} + X_1^{i_0+1} - O^{i_0}) \\ X_2^{i_0} + X_2^{i_0+1} \end{pmatrix}$

- For $j < \tilde{M}^{i_0}$ define

$$\tilde{X}^{i_0,j} := \begin{pmatrix} X_1^{p(j)} \\ O^{p(j+1)} - O^{p(j)} \end{pmatrix} \qquad\qquad \tilde{D}^{i_0,j} := D^{p(j)} \tag{4.126}$$

as well as $\tilde{X}^{i_0,\tilde{M}^{i_0}} := X^{p(\tilde{M}^{i_0})}$ and $\tilde{D}^{i_0,\tilde{M}^{i_0}} := D^{p(\tilde{M}^{i_0})}$.

The move can only be rejected outright if there is but a single storm left. Otherwise, the acceptance probability is given by

$$\alpha_4(\Pi,\tilde{\Pi}) = \alpha_3(\tilde{\Pi},\Pi)^{-1} \tag{4.127}$$

4.5.5 Update the duration of a storm

This operation changes the duration of as storm without affecting any of the other storms or cells. Select a storm index i_0 from the Laplace distribution on $\{1,\dots,M\}$ and sample $\tilde{X}_1^{i_0}$ from the marginal distribution of the storm's duration conditional on its lag:

$$\log(\tilde{X}_1^{i_0})|X_2^{i_0},\theta \sim \mathrm{N}\left(\mu_1 - \frac{[\check{\Lambda}_{11}]_{1,2}}{[\check{\Lambda}_{11}]_{1,1}}(\log(X_2^{i_0}) - \mu_2), [\check{\Lambda}_{11}]_{1,1}\right) \tag{4.128}$$

If $\tilde{X}_1^{i_0} < O^{i_0,M^{i_0}} - O^{i_0}$, the move is infeasible as the storm can no longer contain all its cells.. Else, construct $\tilde{\Psi}$ from Ψ by replacing $X_1^{i_0}$ with $\tilde{X}_1^{i_0}$. Also, sample L from

$$\log(L)|\theta \sim \mathrm{N}(\mu_4, (\Lambda^{-1})_{4,4}^{-1}) \tag{4.129}$$

and set the lag of the last cell to

$$\tilde{X}_2^{i_0,M^{i_0}} := O^{i_0} + \tilde{X}_1^{i_0} + L - O^{i_0,M^{i_0}} \tag{4.130}$$

The acceptance probability is

$$\alpha_5(\Pi,\tilde{\Pi}) = \frac{p_L(O^{i_0,M^{i_0}} + X^{i_0,M^{i_0}} - O^{i_0} - X_1^{i_0}|\theta)}{p_L(\tilde{O}^{i_0,M^{i_0}} + \tilde{X}_2^{i_0,M^{i_0}} - \tilde{O}^{i_0} - \tilde{X}_1^{i_0}|\theta)} \prod_{j=1}^{M^{i_0}} \frac{p_{X^{i,j}}(\tilde{X}^{i_0,j}|\tilde{X}^{i_0},\theta)}{p_{X^{i,j}}(X^{i_0,j}|X^{i_0},\theta)} \tag{4.131}$$

Remark 4.8 *1. To satisfy the consistency condition (3.5), the lag of the last cell in storm i_0 needs to extend beyond the end of the storm. But the exact distribution of the cell lag conditional on survival for at least $O^{i_0} + \tilde{X}_1^{i_0} - O^{i_0,M^{i_0}}$ minutes can only be evaluated numerically. Thus, (4.130) uses an approximation assuming 'lack of memory' in the distribution. Of course, this would only be correct for exponentially distributed cell lags, but the MH-algorithm permits any proposal.*

2. Note that the terms of the product over j in (4.131) cancel for $j < M^{i_0}$ if the X^i and $X^{i,j}$ are independent.

4.5.6 Updating cell depth and duration

This transform changes the depth and duration of a cell without affecting other cells or storms. The conditional distribution of the duration for cell j in storm i can be derived from (3.10):

$$\log(X_1^{i,j})|X_2^{i,j}, X^i, \theta \sim \mathrm{N}\left(\breve{\mu}_1^i - \frac{\Lambda_{3,4}}{\Lambda_{3,3}}(\log(X_2^{i,j}) - \breve{\mu}_2^i), \Lambda_{3,3}\right) \tag{4.132}$$

To update cell j in storm i, we draw a new duration $\tilde{X}_1^{i,j}$ from the above, as well as a depth proposal from the prior distribution $\tilde{D}^{i,j} \sim \mathrm{U}(0, 2\iota)$. We also need to generate new values $\tilde{Z}_T$ for the unknown observations in $\tilde{T} \cap U$. The new aggregates $\tilde{Y}_t$ become

$$\begin{aligned}
\tilde{Y}_t := Y_t &- |[O^{i,j}, O^{i,j} + X_1^{i,j}] \cap [(t-1)\tau, t\tau]|D^{i,j} \\
&+ |[O^{i,j}, O^{i,j} + \tilde{X}_1^{i,j}] \cap [(t-1)\tau, t\tau]|\tilde{D}^{i,j}
\end{aligned} \tag{4.133}$$

The acceptance probability for the new process $\tilde{\Psi}$ arising from a single cell update is

$$\begin{aligned}
\alpha_6(\Psi, \mathbf{Z}, \tilde{\Psi}, \tilde{\mathbf{Z}}) &= \frac{p_{\mathbf{Z}}(\tilde{\mathbf{Z}}|\tilde{\mathbf{Y}}, \theta)p_{\mathbf{D}}(\tilde{\mathbf{D}})p_{X_1^{i,j}}(\tilde{X}_1^{i,j}|X_2^{i,j}, X^i, \theta)}{p_{\mathbf{Z}}(\mathbf{Z}|\mathbf{Y}, \theta)p_{\mathbf{D}}(\mathbf{D})p_{X_1^{i,j}}(X_1^{i,j}|X_2^{i,j}, X^i, \theta)} \\[2mm]
&\quad\times \frac{2\iota p_{X_1^{i,j}}(X_1^{i,j}|X_2^{i,j}, X^i, \theta)\prod_{t\in\tilde{T}\cap U} p_{Z_t}(Z_t|Y_t, \theta)}{2\iota p_{X_1^{i,j}}(\tilde{X}_1^{i,j}|X_2^{i,j}, X^i, \theta)\prod_{t\in\tilde{T}\cap U} p_{Z_t}(\tilde{Z}_t|\tilde{Y}_t, \theta)} \\[2mm]
&= \prod_{t\in\tilde{T}\backslash U} \frac{p_{Z_t}(\tilde{Z}_t|\tilde{Y}_t, \theta)}{p_{Z_t}(Z_t|Y_t, \theta)}
\end{aligned} \tag{4.134}$$

Chapter 5

Implementation and results

We tested the algorithm on artificial instances generated from the lognormal model of Chapter 3. While we did perform some trials with real precipitation records, the results were unsatisfactory, so this section focuses mainly on the artificial data. In particular, we try to assess the speed of convergence and compare the fitted model with the input data.

Most of the figures and tables belonging to this chapter can be found in Appendices A and B.

5.1   Implementation issues

Before analyzing the results, we want to discuss some of the problems and questions that arose during implementation.

The main algorithm was programmed in C++, using the NEWRAN02B library written by Robert Davies [9] to generate normal, uniform, and gamma distributed random numbers. The figures and summary statistics were generated with MATLAB.

5.1.1   Choice of the starting parameter θ

We have seen in Chapter 2 that the asymptotic behavior of the MH-algorithm does not depend on the starting point. However, a bad choice of the parameter increases the time until the Markov chain reaches a stationary state. In worst case, it can actually prevent convergence due to numerical issues. E.g., a badly conditioned precision matrix can be numerically singular. Also, preliminary tests showed that the algorithm has problems dealing with a low starting precision for the storm durations and cell lags. In this situation, it tends to fit a few very long storms to the data instead of assigning a storm to each cell cluster.

Our heuristic for generating the starting parameters is given below. Any mention of λ_k refers to the marginal precisions for the storm and cell quantities, which are not identical to $\Lambda_{k,k}$ in case of non-zero correlation (see (2.7)).

Observation error: For the artificial data sets, we use the true precision scale ζ. The trials with real data use $\zeta = 6000$, which is equivalent to a 10% relative standard deviation for hourly observations according to (3.16).

Storm parameters: The storm parameters can be estimated from the observations by a simple clustering heuristic. The idea is to select a **gap length** $l \in \mathbb{N}$ and treat two wet periods as belonging to the same storm if they are separated by at most l dry periods. Specifically, we define an index set

$$S^l := \{t : Y_t > 0 \wedge \forall s \in \{1, \ldots, l+1\} : Y_{t+s} = 0\} =: \{s_1^l, \ldots, s_{M_0}^l\} \qquad (5.1)$$

of precipitation events that precede a gap of length $(l+1)\tau$ or larger. For this purpose, assume that observations beyond time $T\tau$ have a value of 0. The s_m^l identify the times when storms end. In particular, M_0 is the initial number of storms.

We can assume w.l.o.g. that the s_i^l are sorted in ascending order and define a second sequence of time points as

$$t_1^l := 1 \qquad\qquad t_i^l := \min\{t > s_{i-1} : Y_t > 0\} \qquad\qquad t_{M_0+1}^l := T + 1 \qquad (5.2)$$

for $i \in \{2, \ldots, M_0\}$. These identify the periods when storms start. The durations and lags are now fitted as

$$(\hat{X}_1^i)^l := (s_i^l + 1 - t_i^l)\tau \qquad\qquad\qquad (\hat{X}_2^i)^l := (t_{i+1}^l - t_i^l)\tau \qquad (5.3)$$

This allows us to estimate $\hat{\mu}_k^l$ and $\hat{\lambda}_k^l$ as the sample mean and inverse sample variance of $\{\log((\hat{X}_k^i)^l)\}$ for $k \in \{1, 2\}$. Note: we actually subtract $\frac{\tau}{2}$ from the durations and lags before calculating the estimators to avoid overestimating the means.

To find a suitable gap length, we can either plot the resulting storms or the estimators $\hat{\lambda}_k^l$ for different values of l. The chosen value l_0 should yield relatively high precisions that remain constant close to l_0. Use $\hat{\mu}_k^{l_0}$ and $\hat{\lambda}_k^{l_0}$ as starting values for μ_k and λ_k.

Cell parameters: For the artificial data, we use $\lambda_3 = \lambda_4 = 1$, calculate μ_4 from the mean matching equation (3.21) using the true value of K, and set $\mu_3 = \mu_4$. This is a very rough estimate, but it appears to be sufficiently close for using the algorithm.

For the real data, we use $\lambda_3 = \lambda_1$ and $\lambda_4 = \lambda_2$, as we are unsure of the correct shape parameter. This choice assumes self-similarity for the storm and cell processes, i.e. the durations or lags for both would have the same relative deviation at different scales. We calculate μ_4 based on (3.21) with $K = 100$ cells and choose μ_3 such that the average cells duration is 10 minutes. This choice means that the average cell is longer than the average cell lag in all cases.

Correlation: For models I to IV, the starting parameter treats all quantities as uncorrelated, i.e. the initial Λ is a diagonal matrix. As the algorithm has problems identifying a correlation between storm duration and cell lag, we introduced a correlation of 0.5 for models V and VI, as well as for the fit of model VI to the real data.

Figure 5.1: Gap length and starting precisions
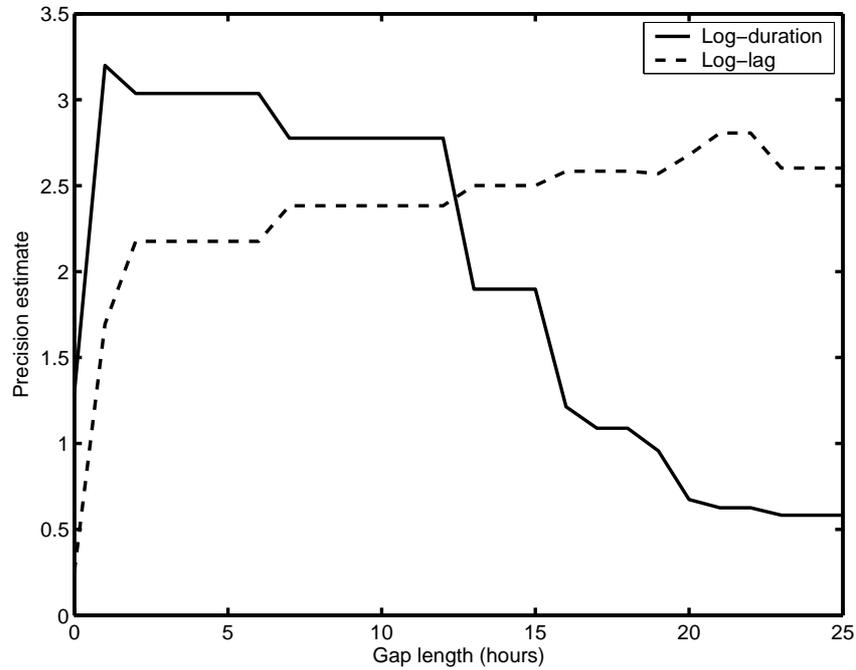
Remark 5.1 *The gap heuristic attempts to identify storms by forming clusters of wet periods. The assumption is that short sequences of dry periods ($Y_t = 0$) occur within storms, while sequences of at least l dry periods lie between storms. This gives rise to the definition of the s_i^l and t_i^l. We can also use them to estimate the position of the storm origins as*

$$(\hat{O}^i)^l := (t_i^l - 1)\tau \tag{5.4}$$

If we assume that clustering of wet periods is evident in the data, there should be a range of choices of l where the 'short' dry spells are all shorter than l periods and all 'long' ones are longer. In particular, the s_i^l and t_i^l, as well as $\hat{\mu}_k^l$ and $\hat{\lambda}_k^l$ remain constant for all values of l in a certain range. By choosing l_0 from this interval, the heuristic should be able to identify the storms reasonably well.

Figure 5.1 shows the precision estimates derived from the heuristic for one of the artificial data sets. The gap length is increased between 0 and 24 hours in 1 hour steps, although τ was actually equal to 10 minutes. In the example, we chose a gap of 4 hours — the middle point of the 2–6 hour interval where both estimators are constant. In Figure 5.2, we see clusters resulting from this choice for part of the data set.

Figure 5.2: Initial storms with 4 hour gaps
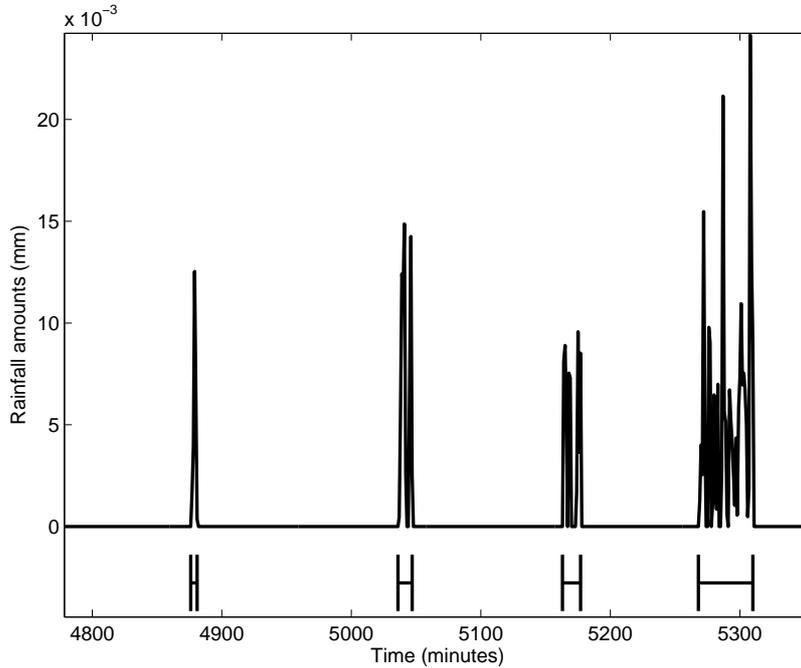
5.1.2   Other considerations

To start the algorithm, we actually need to specify more parameters than just the initial value
for θ. Here is a list of the most important ones:

Cell intensity: While we know the true ι for the artificial data, we need to choose a rea-
sonable value for the real precipitation records. This is critical, as our first test runs
did show that the behavior of the algorithm is very sensitive to misspecification. We
estimated ι using trial and error, with the following heuristic as a guideline:

If $\{Y_t\}$ is an aggregate process with average cell intensity ι and $\{\hat{Y}_t\}$ is a process with
the same θ and $\iota = 1$, the fact that the uniform distribution on the interval $[0, \iota]$ is scale
invariant implies

$$\mathrm{E}(Y_t|\theta) = \iota\,\mathrm{E}(\hat{Y}_t|\theta) \qquad\qquad \mathrm{Var}(Y_t|\theta) = \iota^2\,\mathrm{Var}(\hat{Y}_t|\theta) \qquad\qquad (5.5)$$

These equations allows us to obtain a rough a-priori estimate for the mean intensity
by comparing a process generated from the starting parameter θ and $\iota = 1$ with the
observations. We actually used the sample mean and variance for the observations
Z conditional on rain, but they satisfy (5.5) asymptotically for large sample sizes. If
the quotient of the two expectations is not close to the square-root of the quotient of
variances, the starting parameter is clearly far from the true θ and should be modified.

Prior precisions: The precision parameters r and s for the prior distribution of μ and Λ (see Section 3.3) need to be specified. We used $r = s = 0.01$ for the trials with artificial data and I to IV. To ensure that the problems the algorithm has with identifying correlation do not stem from an overly restrictive prior, we lowered the values to 0.001 for models V and VI, including the runs with real data.

Transition probabilities: As we have discussed in Section 4.5, we want the algorithm to attempt more structural transforms on the cells than on the storms. We used the probabilities $p_1 = p_2 = 35\%$ (add or subtract a cell) and $p_3 = p_4 = p_5 = 10\%$ (split or merge a storm, update its duration).

Number of iterations: A necessary condition for fast convergence of the main algorithm is that each subsampler converges quickly to its stationary limit conditional on the current state of the process. This can be improved by an intelligent choice of the proposal distribution, but ultimately, it means that the samplers need to perform sufficiently many iterations to be able to explore the parameter space. On the other hand, we do not want too spend too much computation time on sampling values conditional on parameters that are 'far from the truth'.

Keeping the above in mind, we have found that the following iteration counts for the subsamplers are sufficient for one iteration of the main algorithm:

- The parameter θ is updated 5,000 times.
- The weights p_M and p_{M^i} are estimated from 100,000 samples each. Additional smoothing is required (see Section 4.4).
- The storm and cell transforms 1–5 are repeated 200,000 times with a random transform chosen at each instant.
- Each cell has its duration and length resampled 100 times using transform 6. We increased this to 1000 samples for models V and VI, including the real data, as this improves the ability of the algorithm to identify the correlation between storm duration and cell lags.

Burn-in period: The test runs show that the burn-in period should be at least 200K iterations. We could not reach this in all cases, as the algorithm exceeded its allocated processing time on some combinations of processes and starting parameters.

Another important question is whether to impose bounds on the parameters to avoid numerically instable states. We found this necessary for the precision matrix Λ when dealing with model V or VI. More precisely, we chose to limit the marginal variances and not the diagonal entries of Λ, as a bound on these would still permit arbitrarily large variances if the matrix was close to singular. All marginal variances were bounded below by 0.1 to prevent a degenerate process with quantities which are practically constant. The upper bounds for the artificial data were set to 2.0 in all cases (the true value is $\frac{1}{3}$, see below). In case of the real data, we tested several sets of bounds but none of them yielded a stable starting configuration.

The bounds do not affect the theoretical validity of the algorithm, as they are equivalent to truncating the range of the priors. We want to fit a model where cells are clustered into distinct storms, so we can exclude any parameters that describe a different behavior.

Another limitation we impose in the final version of the program is a lower bound of 2 for the storm and cell counts M and M^i. Any proposed transform that would result in such a process is rejected. This guarantees that the normal proposal distribution we use for the sub-diagonal entries of Λ is always correct (see Section 4.3) and allows us to calculate sample variances for diagnostic output. The effect on the estimates is negligible as long as the data set exhibits clustering of wet periods, i.e. many storms with several cells each are necessary to explain the observations.

5.2 Generating artificial data

We generated several artificial data sets by simulating processes according to the lognormal rainfall model. This way, we may assess how well the algorithm performs if the model assumptions are true. Our test instances were constructed as follows:

- The observation period for each set is taken to be $T = 150,000$ minutes or approximately 104 days. This is slightly more than a season's worth of data.

- To test how sensitive the algorithm is with regards to changes in the length of the observation period, we generated two processes for each combination of parameters in the test. One data set has 15,000 observations in 10 minute intervals, the other has 2,500 hourly observations.

- We chose an error scale $\zeta = 1,000$, which yields a relative standard deviation of 10% for observation intervals of length $\tau = 10$ minutes by (3.16). As initial attempts showed that the parameter is hard to estimate correctly, all tests were repeated with ζ fixed to its true value. This assumption can be justified in practice if the reliability of the measurements is known.

- As we had no prior knowledge of likely values for Λ, we chose a multiple of identity for testing the simplified model with diagonal Λ. Setting $\lambda_k = 3$ seems to produce a reasonable range of values. The 99% quantiles of the resulting lognormal distributions are larger by a factor of 14.7 than their 1% quantiles.

- For the full model, we introduce a correlation of -0.5 between storm duration and cell durations, as we expect an effect of this kind in real data. The marginal variances are still equal to $\frac{1}{3}$ as above:

$$\Lambda = \begin{pmatrix} 4 & 0 & 2 & 0 \\ 0 & 3 & 0 & 0 \\ 2 & 0 & 4 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \qquad \Lambda^{-1} = \frac{1}{3} \begin{pmatrix} 1 & 0 & -\frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 \\ -\frac{1}{2} & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad (5.6)$$

Table 5.1: Parameters used to generate artifical data

Quantity	Log-mean*	Value	Expectation	Quantiles			
				1%	5%	95%	99%
Storm duration	μ_1	$\log(180) - \frac{1}{6}$	180 min	39.8	58.9	393.8	583.7
Storm lag	μ_2	$\log(2880) - \frac{1}{6}$	48 h	10.6	15.7	105.0	155.7
Cell duration, short	μ_3	$\log(10) - \frac{1}{6}$	10 min	2.2	3.3	21.9	32.4
", long		$\log(30) - \frac{1}{6}$	30 min	6.6	9.8	65.6	97.3
Cell lag	μ_4	$\log(18) - \frac{1}{6}$	18 min	4.0	5.9	39.4	58.4

* The marginal log-variance is $\frac{1}{3}$ in each case.

The same data sets used for fitting model I were also used for model III, to allow for a direct comparison of the performance. Models V and VI were instead tested with a precision matrix introducing positive correlation between storm duration and cell lag:

$$\Lambda = \begin{pmatrix} 4 & 0 & 0 & -2 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ -2 & 0 & 0 & 4 \end{pmatrix} \qquad \Lambda^{-1} = \frac{1}{3}\begin{pmatrix} 1 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & 1 \end{pmatrix} \tag{5.7}$$

- The means for the logarithmic storm duration and lag were chosen to give somewhat plausible numbers (3 hours and 48 hours respectively). As we want cell clustering to be evident in the data sets, the storm durations are significantly shorter than the lags.

- We use two different choices for μ_3, resulting in average cell lengths of 10 or 30 minutes. This allows us to study how accurate the parameter estimates is if the unobservable quantities in $\mathbf{X}$ are of similar size or markedly smaller than the observation period.

- The parameter μ_4 is chosen such that it satisfies the assumption of matched means with an average number of $K = 10$ cells per storm (see (3.21)). Thus, we can test the models with and without matching on the same data sets.

Some properties of the chosen parameters μ and Λ are summarized in Table 5.1. We use the following scheme to classify the data sets and possible model fits:

1. The parameters used to generate the process are denoted by

   (A) no correlation (models II and IV), short cells.
   (B) no correlation (models II and IV), long cells.
   (C) correlation between storm and cell quantities (models I and III), short cells.
   (D) correlation between storm and cell quantities (models I and III), long cells.

(E) correlation between storm quantities and cell lag (models V and VI), short cells.

(F) correlation between storm quantities and cell lag (models V and VI), long cells.

2. For the length of the observation period, we use

(1) 10 minute intervals.

(2) 1 hour intervals.

3. The constraints on the fitted model are given as

(a) no mean matching (models I, II, and V), variable ζ.

(b) no mean matching (models I, II, and V), fixed ζ.

(c) mean matching (models III and IV), variable ζ.

(d) mean matching (models III and IV), fixed ζ.

(e) no mean matching, limited correlation (model VI), fixed ζ.

For example, 'A1a' refers to the test run using the data set with uncorrelated quantities, short cells, and 10 minute intervals. The fitted model uses diagonal Λ, no mean matching, and treats ζ as an unknown parameter. Also, recall that the models with mean matching fix $\Lambda_{1,4} = \Lambda_{2,4} = 0$, i.e. the correlation structure for classes C and D depends on the model.

Table B.1 shows the 30 combinations that were actually tested. If an entry is marked 'alt', we initiated a second run with a different starting parameter. The θ for the first run was generated by the heuristic in Section 5.1.1. For the second trial, we multiplied these values by 0.9, which already constitutes a drastic misspecification of the storm durations and lags. We want to use the repeat runs to assess the impact of the starting parameter on the behavior of the algorithm. Also, the posterior scale reduction criterion for convergence requires output from multiple parallel runs (see Section 2.5.2).

5.3 Numerical results

In this section, we discuss the results of our numerical trials. As multivariate precision matrices are difficult interpret, we invert the estimated Λ and present the results in terms of the marginal variances σ_k^2 and correlations $\rho_{k,l}$. For the latter, indices are assigned in a manner consistent with previous notation, e.g. $\text{Corr}(X_1^i, X_2^i) = \rho_{1,2}$ and $\text{Corr}(X_1^i, X_1^{i,j}) = \rho_{1,3}$.

When dealing with the question of convergence for an MCMC algorithm, a common recommendation is to consider other summary statistics of the chain than just the parameter of interest. We include the total number of cells $\overline{M}$ in our analysis, as well as the normalized intensity average for the cells:

$$\hat{D} := \frac{1}{\iota \overline{M}} \sum_{i=1}^{M} \sum_{j=2}^{M^i} D^{i,j} \tag{5.8}$$

The expectation of this quantity is close to 1 under the model. If $\hat{D}$ for the fitted process deviates considerably from 1, we know that distribution of the cell lags and/or durations has not been correctly identified yet. In this case, the transforms for adding or removing cells will result in a large proportion of cells whose mean intensity deviates considerably from their supposed average of ι.

5.3.1 Results for artificial data

Most of the test runs were made on computers with Athlon 2.2 GHz processors. Computation was limited to 300 hours of CPU time or 500,000 iterations of the main sampler, including burn-in. As shown in Table B.1, most of the A–D runs reached more than 450,000 iterations. Note that simplifying the model and fixing ζ both speed up computation considerably, while the additional cell depth updates for models V and VI (case E–F) take more time. Where possible, we use 300,000 burn-in iterations for the final estimator. To assess whether this is sufficient, we compare estimates based on iterations 150,001–300,000 with those for 300,001–450,000. These are denoted by '150K' and '300K' respectively. If a data set has fewer samples available, the number of burn-in iterations is reduced as indicated in the table.

As the values of the Markov Chain are highly correlated, we calculate estimators from a thinned chain using 1 out of every 100 iterations. This is not necessary if we consider the central limit theorem (2.11), which compensates for high correlation, but the resulting empirical distribution is a better approximation of the posterior density.

The first step in evaluating the performance of the algorithm is the graphical analysis of the output from the sampler. We consider trace plots for the parameters, the corresponding autocorrelation functions (ACF), and kernel density estimates for the posterior density. These figures take up a lot of space, so we limit the analysis to a few representative instances:

(A1a) Model II, short cells, variable ζ, diagonal Λ, 10 min intervals.
   (Figures A.1–A.4)

(C1d) Model III, short cells, fixed ζ, mean matching, restricted Λ, 10 min intervals.
   (Figures A.5–A.10)

(C2d) Model III, short cells, fixed ζ, mean matching, restricted Λ, 1 h intervals.
   (Figures A.11–A.14)

(D1d) Model III, long cells, fixed ζ, mean matching, restricted Λ, 10 min intervals.
   (Figures A.15–A.18)

(F2b) Model V, long cells, fixed ζ, restricted Λ, 1 h intervals
   (Figures A.19–A.24)

Figures for the full model I are not included, as the test runs show that it is clearly overparametrized (see the discussion of the tabulated results below). We focus on models III and V, as these appear to be the most useful for practical purposes.

A first observation that applies to all instances and parameters is that the samplers do not as a rule converge to the 'true values' indicated in the trace plots. However, these are merely the values used for generating the sample processes. As we are dealing with normal-Wishart priors, even knowledge of the underlying process $\mathbf{X}$ would only yield convergence to a (slightly biased) maximum likelihood estimator. And for short samples, this can deviate markedly from the true parameter. Given that we do not observe $\mathbf{X}$ itself, the best result we can expect is that the sampler frequently assumes values in the neighborhood of the 'true value', which is indeed the case for most instances.

As for the different parameters controlling the model, we can draw the following conclusions from the graphical analysis:

- The sampler for the precision scale ζ has a high variance even for the simple case A1a, and its is unclear whether it has reached a stationary state after 150K iterations (Figure A.1). As the other parameters for this instance appear to converge much faster, it seems that the exact value of ζ does have a negligible effect on the behavior of the algorithm. The slow convergence would thus be due to a relatively flat likelihood in terms of the variance scale. For this reason, the other instances we consider for the graphical analysis use a fixed value for ζ.

- The trace plots, ACFs, and kernel density estimates for the posterior distribution of the μ_k show that the sampler converges very fast for the simple model II with no correlation (Figure A.2), and reasonably quickly for all instances of model III (Figures A.5, A.11, and A.15). For the latter, 150K burn-in iterations appear to be insufficient, in particular for the C2d case where we can clearly see that the sampler is stuck in a local maximum for some time. Model V shows an upward trend for μ_3 and μ_4 towards the end of the run (Figure A.19), indicating that even 300K burn-in iterations may not be enough at the coarse 1 h time scale.

  All parameter values are reasonably close to the 'true value', except in the case of C2d. For the latter, the local maximum found early in the run is close, but the algorithm eventually settles into a quite distinct state. Again, this seems to be a problem with the 1 h time scale: a less structured process appears more likely as fewer details are visible at this resolution.

  The density estimates support the conclusions from the trace plots in all cases. From the ACF plots, we see that thinning with a step size of 100 yields effectively independent means for the storm durations and lags in the simple case of model II, but the means for the cell quantities still exhibit a moderately high correlation. The same can be observed for the other instances, except that for model III, the coupling of μ_1 and μ_4 introduces non-zero correlation for μ_1 as well. Note that the extremely high correlation for C2d is due to the process getting stuck in a local maximum for a while, which is an instationarity (at least numerically).

- The results for the variances σ_k^2 are very similar to those for the means (Figures A.3, A.6, A.12, A.16, A.20). In particular, 150K burn-in iterations do not appear sufficient

for the instances with correlated quantities.

- For models III and V, where correlation between some storm and cell quantities is permitted, the samplers give good qualitative results. Even for the 1 h time scale, the samplers are near zero where no correlation exists and have the right sign otherwise. Only the absolute magnitude of the correlation is usually underestimated. For D1d, the sampler for $\rho_{1,3}$ appears to converge to the true values of -0.5, but the remaining two correlation parameters are farther from 0 than for the other instances. Of course, this may be due to the fact that input data has sample correlations that are further from zero than for the other instances.

- No trace plots or ACFs are included for the processes C1d and F2b using alternate starting parameters, but we compare the kernel density estimates after 300K burn-in iterations for the two sets of starting parameters. In case of C1d, a marked difference exists between the original and alternative starting parameters for μ_1 and all cell parameters (Figure A.8). We shall see later that this is likely due to the fact that the alternative starting parameter fixes the precision scale to $\zeta = 900$, whereas the original run has $\zeta = 1000$. However, we cannot rule out at this point that the sampler has not convergence after 300K burn-in iterations. At least in case of F2b, there are hardly and differences between the original and alternative sampler (Figure A.22).

- The remaining plots do not deal with the parameters themselves, but with their impact on the model (Figures A.4, A.9, A.10, A.14, A.18, A.23, and A.24). If we consider the normalized intensity average $\hat{D}$, we see that the fitted density estimate has a peak close to 1 for A1a, C2d, and D1d, while C1c and F2b have peaks slightly to the left, meaning that the algorithm prefers cells with below average contribution for the model. This is another indicator of possible lack of convergence for these instances.

  We also compare the average number of cells for the process reconstructed by the algorithm with the number for the original process on which the estimation is based. This is done both in terms of a kernel density estimate for output from the sampler, as well as for 3 independent resamples using the fitted parameter. In all cases, the mode of the fitted distribution is far from the true value, although the latter is not completely unlikely. As the resampled values themselves show a high variability, we cannot draw a definite conclusion here.

  Finally, we look at the behavior of the observations for the original and fitted parameters. To do this, the duration of wet and dry periods in the original data is compared to that of three resamples using the estimated θ. The same is done for the rainfall intensity conditional on rain. Comparisons are made using kernel density estimates and boxplots for the log-quantities.

  For the simple model II, the distributions of the average duration of wet periods are very similar for the original process and the resamples, and completely indistinguishable for the amounts. The durations of the dry periods are quite distinct at first glance, as the resamples have a bimodal density, whereas the original is unimodal. However, the

left mode is located exactly at the log of the observation scale of 10 minutes, i.e. the apparently major difference is (mostly) a problem of limited resolution:

There are two kinds of dry periods in the model — short gaps between cells in an ongoing storm, and longer gaps between storms. This should result in a bimodal density, provided enough of the shorter gaps are large enough to contain an entire observation interval. However, the original parameter for A1a results in a probability for this event that is too small to yield a distinct second peak, whereas the estimated parameter does.

For the two C1d processes, the originals and resamples cannot be distinguished at all, which may mean that we do not have lack of convergence after all. The remaining processes result in fits that are almost as good, although the distribution for the duration of dry periods varies the most (due to the 'instability' of the left peak).

From the graphical analysis alone, we can conclude the following:

- The true value of ζ is difficult to estimate, but its impact on the behavior of the other parameters appears to be relatively small.

- 150K burn-in iterations appear insufficient for instances with correlated quantities, 300K iterations should be used.

- For the 1 h observation periods, the true parameters cannot be reconstructed exactly, as we lose some information in the aggregation process.

- The algorithm provides an estimator that describes the observable quantities like the length of wet and dry periods or the aggregated rainfall amounts very well. This is true even for hourly observations, where the estimators do not resemble the 'true values'.

Results for all test runs are available in tabulated form in Tables B.2–B.47. Each table shows the true values of the parameters that were used to generate a sample, as well as the starting parameters assigned by the heuristic of Section 5.1.1. The resulting estimators are presented both in the form of a median and credible interval (10% and 90% quantile of the posterior distribution), and as a mean and 95% confidence interval based on the central limit theorem (2.11). For those instances where a second run was made using an alternative starting parameter, we also determine the potential scale reduction (2.92) to assess convergence.

As the data for the remaining runs mostly supports the findings from the graphical analysis, we keep the discussion brief and merely highlight some particular points of interest:

- Even though the estimated values for ζ differ drastically from the true precision scale, the difference in the results for the instances with variable and fixed ζ is comparatively small. Considering this, as well as the decrease in computation time that results from a fixed value (Table B.1), it seems reasonable to exclude ζ from estimation.

- As stated above, the full model I is effectively overparametrized. This can be seen from Tables B.2 and B.3 showing the results for the C1a and C1b runs. Both samplers choose a correlation structure that is completely different from the one used to generate the artificial observations. On the other hand, the estimators for models III and V with restricted precision matrix Λ are all able to determine the sign of the single non-zero correlation correctly. Those with long (30 min average) cells on a 10 min time scale even yield the correct absolute value (Tables B.16 and B.17), while the others tend to underestimate it.

  There appears to be a tendency in the D1 and all E models to assign a positive correlation $\rho_{1,2}$ between storm duration and lag. But as noted above, this may actually be true for the sample correlation of the artificial data sets being used.

- By far the best reconstruction of the true parameter is found for the data sets with long cells and 10 min observation intervals (Tables B.16, B.17, B.38, B.39, and B.46). In all other cases, the resolution seems to be too limited to reconstruct the cell parameters exactly. The coarse grid means that certain processes appear more likely, even though they do not replicate the fine structure belonging to the original parameters. But as we have seen in the graphical analysis, they can still yield a good description of the observations at that time scale.

- A direct comparison of parallel runs using alternative starting parameters shows no appreciable difference for the A1c, C1c, E2b, and F1b instances (Tables B.22, B.10, B.37, and B.40), neither in terms of credible (confidence) intervals, nor in terms of potential scale reduction. However, differences are evident for A1d, C1d, E1b, and F2b (Tables B.25, B.13, B.34, and B.43). For the first two, the distinguishing factor is that the 'c' runs use a variable value for ζ, whereas it is fixed to 1,000 resp. 900 in the 'd' runs. This supports our claim that we do not have lack of convergence after 300K iterations, but that the process merely prefers slightly different parameters conditional on specific values of ζ. For E1b, the effect may be the same, but made worse by the short length of the available run. For the F2b data, the only apparent cause is the 1 h time scale, although the E2b run was not affected in the same manner.

5.3.2 Real precipitation data

After ensuring that the algorithm can identify parameters for a process conforming to the model, the next step was to test it on real precipitation data. For this purpose, we used precipitation data recorded hourly at two meteorological stations in Rheinland-Pfalz during the years 2003 and 2004. As the model does not include seasonal variations, the first step in preparing the data was to visually identify periods where the rainfall distribution appears to be homogeneous. For each individual station, this is the case for May to July and for October to December, although there is a marked difference between stations.

The next issue was finding a suitable starting parameter, as well as an average cell intensity ι and – for models III and IV – a typical number of cells K. Unfortunately, we did not find a

satisfactory solution to this problem. Experimenting with several choices for θ, ι, and possibly K showed that the process fitted by the algorithm depends strongly on the choice of ι. For small values, the number of storms becomes very large as almost each interval of non-zero precipitation is assigned its own storm. Conversely, a larger ι results in very few storms of extreme length. Either way, no meaningful clustering occurs. Also, there appears to be no middle ground — the algorithm falls from one extreme into the other as the average cell intensity rises above a certain threshold.

All trials using real data resulted in a degenerate state of the model or lack of convergence of the sampler. We do not want to present any of the results here, but merely point out some of the difficulties and likely causes:

- The time scale of the available data is too coarse. Hourly observations do not permit identification of a correlation structure that enables clustering. E.g., we are unable to distinguish between an intense 5 minute shower and a moderate 45 minute rainstorm. Recall that we already saw that the algorithm underestimates the correlation between storm duration and cell lag for the artificial data if the cells are not longer than the observation interval.

- The storm durations and lags have a very high variance, which means that clusters are difficult to identify. This is amplified by the previous problem, as the algorithm is unable to introduce sufficient variation in the frequency with which cells occur in storms of widely different length.

- The lognormal distribution may not be a good approximation for the true distribution of the storm and cell lags. At the very least, it yields an extremely flat likelihood as the variance increases, which means that the algorithm may converge slowly or even fail to converge for numerical reasons.

- As ι is so critical, it should be estimated along with μ and Λ.

Unfortunately, we are as yet unable to provide satisfactory solutions to these problems, although some suggestions regarding the last two issues are made in Chapter 6.

In principle, the model is comprehensive enough to reproduce the features of precipitation we are interested in. Figure A.25 shows the distribution of wet and dry periods, distribution of rainfall amounts, and sample observations for real data and some artificial processes. As the latter have not been fitted to the former, the quantities are not closely matched, but the processes do appear to be of the same kind. Note that the observations have been rescaled for better comparison (the maxima are equal to 1).

5.4   Summary

We have seen that the full model specified in Chapter 3 is effectively overparametrized, at least for the resolution and amount of data we have available. This suggests simplifying

the model in various ways as discussed in Section 3.2, as well as treating the non-critical parameter ζ as a constant.

With the above restrictions, the algorithm appears to reach stationary state on artificial data generated from the correct model in at most 300K iterations. The estimated parameters are sensitive to the time scale and cannot reproduce fine structures on a scale smaller than the observation interval. Actually recovering the input parameters requires that the cells are on average longer (e.g. 3 times) than this period. However, the parameters always describe the process well in terms of the observations **Z**.

Unfortunately, the algorithm does not work on real precipitation records. In its present form, it is not suitable for the statistical analysis of hourly rainfall data.

Chapter 6

Further considerations

In the last chapter, we look at how the ideas presented in this thesis could be developed further. We suggest changes to the model and briefly address the question of prediction.

6.1 Changes to the model

The modifications discussed below fall into two categories. The first are changes to the distributions that might enable fitting parameters to real precipitation data, while the second are suggestions for extending the scope of the model.

6.1.1 Variant distributions for X

The framework for precipitation time series presented in Chapter 3 can support a variety of probabilistic models. We settled on the lognormal distribution to describe the process X due to its flexibility and explicit correlation structure. One downside of this choice is the resulting point processes do not possess an analytic solution for the distribution of events per time interval. While we have shown how to approximate these weights in Section 4.4, the Monte-Carlo scheme requires substantial computation time.

A second problem is that the distribution of extreme lags is likely wrong. The lognormal density can be fitted closely to most unimodal, non-negative distributions near the mode, but its asymptotic behavior is different from e.g. the exponential or gamma case. This might even be a contributing factor to our inability to properly fit the lognormal model to real data. If the exponential distribution is a better model for the lags, its heavier tails would lead us to overestimate the variance of log-lags in the lognormal approximation.

While there is no multivariate extension of the exponential distribution that treats correlation explicitly, one could try to combine the advantages of the Gaussian and the exponential model

using a copula approach. We decompose the joint distribution for the components of $\mathbf{X}$ as

$$P_{\log(X^i),\log(X^{i,j})}(x_1, x_2, x_3, x_4 | \theta, R) \tag{6.1}$$
$$= C\left(P_{\log(X_1^i)}(x_1|\theta), P_{\log(X_2^i)}(x_2|\theta), P_{\log(X_1^{i,j})}(x_3|\theta), P_{\log(X_2^{i,j})}(x_4|\theta) \Big| R \right)$$

Here, C is the 4-variate Gaussian copula with correlation matrix R, and the marginals are arbitrary distributions characterized by some parameter vector θ. For exponential lags, the log-quantities need to follow a distribution of the Gumbel type. The cell and storm durations can still be treated as lognormal if desired. An introduction to copulas can be found e.g. in the book by Nelsen [31], while the Gumbel distribution plays a role in extreme value statistics and life testing (see e.g. Barlow and Proschan [3]). A more recent treatment of both topics in the context of financial statistics can be found in the book by Franke et. al. [11]

With exponential lags, the storm origins are distributed according to a Poisson process, which is much easier to handle than the lognormal point process. The conditional distribution of the cell origins will of course not follow any classical model.

6.1.2   Other models for cell depth

Cell depth is another area where a different distribution could lead to improvements. We argued in Chapter 3 that the depths should be left out of the lognormal framework to avoid overparametrization. Treating them as constant is not feasible numerically, as it enforces an almost deterministic relationship between cell lengths and observations. Instead, we chose the uniform distribution on $[0, 2\iota]$ for ease of use. Two points in its favor are that it depends on a single parameter and that it is invariant under multiplication. The latter motivates the heuristic (5.5) for estimating ι. An alternative with the same two properties is the exponential distribution with parameter $\frac{1}{\iota}$. Also, since it is not bounded above, it could make it easier for the algorithm to fit cells to events with high precipitation.

More importantly, as mentioned in Chapter 5, the behavior of the algorithm depends critically on the right choice of ι. This suggests that ι should be treated as part of the estimation problem, replacing the less critical ζ as part of θ.

6.1.3   Inclusion of covariates

So far, we have fitted the rainfall model to observed precipitation without taking into account other meteorological data. But ambient temperature, atmospheric pressure, humidity, wind speed, etc. are often recorded alongside rainfall and will be dependent to some degree. Other interesting candidates for covariates are periodic functions of the time or date to capture daily or seasonal variations in the climate.

Assume that we are given d exogenous variables recorded as a vector $V_t \in \mathbb{R}^d$ for each time period t. A straightforward way to include them in the lognormal model is to express the mean μ in terms of the covariates via linear regression. One question that arises immediately

is how to link the rainfall process to the discrete records $\mathbf{V} := \{V_t\}$. Fitting an unobserved temperature process, a wind process, etc. similarly to the rainfall process $\mathbf{X}$ seems impractical. A simpler approach would be to base the mean for each storm and cell on the value of V_t at its origin. Let

$$t(i) := \left\lceil \frac{O^i}{\tau} \right\rceil \qquad\qquad t(i,j) := \left\lceil \frac{O^{i,j}}{\tau} \right\rceil \tag{6.2}$$

This allows us to define local means as

$$\check{\mu}_B^i := B_1 V_{t(i)} \qquad\qquad \check{\mu}_B^{i,j} := B_2 V_{t(i,j)} - \Lambda_{22}^{-1}\Lambda_{12}^T(\log(X^i) - \check{\mu}_B^i) \tag{6.3}$$

where $B_1, B_2 \in \mathbb{R}^{2\times d}$ are matrices of (unknown) regression coefficients. The distribution of X^i is now taken to be

$$\log(X^i)|\mathbf{V}, \theta \sim \mathrm{N}_2(\check{\mu}_B^i, \check{\Lambda}_{11}) \tag{6.4}$$

instead of (3.8). The distribution of $X^{i,j}$ conditional on X^i, which was previously given by (3.10), is replaced by

$$\log(X^{i,j})|X^i, \mathbf{V}, \theta \sim \mathrm{N}_2(\check{\mu}_B^{i,j}, \Lambda_{22}) \tag{6.5}$$

Note that we can assume w.l.o.g. that the first entry of each V_t equals 1 to include a constant term in the regression model. If we use normal priors for the entries of B_1 and B_2, the posterior distribution for μ (4.6) factorizes into normal posteriors for the coefficients. As this approach requires four parameters for each covariate, we need to limit ourselves to the most essential ones. A technique like principal components analysis should be applied to reduce the dimension of the exogenous component to a manageable level. Methods of this type are discussed e.g. in the book by Draper and Smith [10].

Given the difficulties we had with estimating Λ in the model without covariates, it does not seem practical to give them a direct influence on the distribution of the precision matrix. As for the different distributions discussed in the previous section, the regression approach can still be used w.r.t. an appropriate parameter.

6.1.4 Extension to multiple sites

Our original motivation for looking at the time-series approach to precipitation was to create a tool for simulating rainwater runoff in urban planning. For meteorological analysis on a larger scale, single site models are not adequate, although they can be used as building blocks for a spatial model. In this section, we only consider locales which are both small enough to be treated as planar and uniform enough to have no systematic variations in precipitation (there are no mountains, etc.).

Given observations $\{Z_t^l\}_t$ at multiple locations $a_l \in \mathbb{R}^2$, $l \in \{1, \ldots, S\}$, we want to predict the rainfall intensity at an unobserved site or estimate the total amount over a larger area. To do this, we need a model for spatial dependence. Usually, this is done by requiring spatial stationarity or **homogeneity** of the process. This is essentially the same as stationarity (cp.

Definition 2.10), except that the joint distribution of the observations at different sites remains the same under spatial shifts. In practice, we require only **weak homogeneity**, i.e. the means, variances, and correlations for any set of points in the plane are invariant under a shift in location. A special class of homogeneous spatial processes are **isotropic** processes, where the joint distribution for observations at two sites depends only on their distance. An introduction to spatial statistics can be found in in the book by Ripley [36].

While isotropy is a very restrictive assumption, it can be used for modeling together with a coordinate transform. The isotropic model is than fitted on the basis of the transformed locations. An example of this approach can be found in the thesis by Gründer [23], where a model for pollution measurements is developed that decomposes the observations into a long-term average and fluctuations that are isotropic under an elliptical transform. The long-term effects are estimated using a regression surface approach, a technique known as **kriging** in geostatistics. For rainfall modeling on a small scale, the situation is different in so far that we can assume a constant baseline of 0 (no precipitation) and are only interested in finding a homogeneous model for the short-term changes. This, too, can be fitted with kriging methods if additional constraints on the regression functionals are observed. A spatio-temporal model of this type is developed and applied to rainfall estimation by Stroud et. al. in [39]. It uses kernel-based regression in a state-space framework.

A simple way to combine the approach of [39] with the local rainfall models we have studied in this thesis is the following: Let K_R be the two-dimensional Gaussian kernel with correlation matrix R, and let $\{Y^l\}$ be independent aggregate processes in the sense of definition (3.2), which we associate with the a_l. The rainfall intensity at any other point $a_0 \in \mathbb{R}^2$ and time t is taken to be

$$\hat{Y}_t(a_0) := \sum_{l=1}^{L} Y_t^l K_R(a_0 - a_l) \tag{6.6}$$

This expression converges to a homogeneous spatial process in a_0 as the number of kernels in a bounded neighborhood of a_0 goes to infinity. It is only isotropic if the kernel is spherical, i.e if R is a multiple of identity. Note that in general $\hat{Y}_t(a_l) \neq Y_t^l$, as the rainfall observed at a_l is actually the aggregated contribution from all l processes. Accordingly, the observations $\{Z_t^l\}$ made at the sites a_l are not distorted versions of the $\{Y_t^l\}$ but of the $\{\hat{Y}_t(a_l)\}$. We also need to relax the assumption that $Z_t = 0 \iff Y_t = 0$ we made in Section 3.1 to $\sum_{l=1}^{L} Z_t^l = 0 \iff \sum_{l=1}^{L} Y_t^l = 0$. This accounts for the situation where only some stations register rainfall. In this case, any recorded zeros need to be treated as very small observations.

To fit the joint model, we need to consider all sites simultaneously, and estimate R alongside the other unknown parameters. More involved models are possible, e.g. we could associate a kernel with each individual cell and allow its scale to change depending on cell depth.

The kernel interpolation is somewhat unsatisfactory, as it arbitrarily allocates precipitation to the observation sites. In doing so, they gain a special status that is not justifiable outside the model. In fact, the average cell frequency may depend greatly on the number and proximity of the sites. Rainfall in an area with many gauges will be distributed evenly among them, while an isolated station is "responsible" for most of the precipitation recorded there.

Similar to how the disaggregation approach uses a representation independent of the time scale, we can attempt to "dislocate" rainfall in space. The cells are no longer tied to a specific location but allowed to occur anywhere, or even to move. Of course, this increases the complexity of the model considerably, as we have to specify the location and behavior of the cells. A discussion of spatio-temporal models and their relation to single-site point-process models can be found in an article by Cox and Isham [8].

6.2 Prediction

Once the model is fitted to a data set, we can use it to predict future precipitation. Besides being an important question in meteorology, this can be used to validate the model — or to show its limitations. We did not study the accuracy of such forecasts, but we want to give at least an idea of how they can be obtained. As the algorithm can handle unknown observations, we can use it to estimate future rainfall simply by increasing T. But this approach is computationally expensive and only advisable for short time periods. The long term behavior of the model can be determined much more efficiently by numerical integration.

Once we have identified the model parameters, it is not necessary to restrict ourselves to discrete observations recorded at a certain time scale τ. Instead, we consider prediction in terms of the expected rainfall intensity over time. This can easily be converted to observation forecasts at any desirable scale by integration.

Definition 6.1 *For the rainfall model of Definition 3.1, the* **intensity function** *is defined as*

$$I(x) := \sum_{i=1}^{M} \sum_{j=1}^{M^i} D^{i,j} \mathbb{1}_{[O^{i,j}, O^{i,j}+X_1^{i,j}]}(x) \tag{6.7}$$

where $x \in [0, T\tau]$.

The function $I(x)$ is the total depth of all cells which are 'active' at time x. Note that for $t \in \{1, \ldots, T\}$ holds

$$Y_t = \int_{(t-1)\tau}^{t\tau} I(x) \, dx \tag{6.8}$$

As the individual storms in the model are independent, forecasting can be split into two distinct problems. The first is to estimate the future impact of storms originating during the observation period. We refer to this as the short term forecast. By contrast, the long term forecast tries to assess the effect of storms that start after the last recorded observation.

The two problems can be solved almost separately. The only connection is that the long term forecast depends on the distribution of the origin of the first storm starting after time $T\tau$ conditional on the observations. I.e., the long term forecast depends on the storm lags during the observation period, but not on the storm durations or cell quantities.

6.2.1   Short term forecasts

With the short term forecast, we want to determine

- the effect of known storms on the expected intensity beyond the time $T\tau$.

- the distribution of the origin of the first storm to start after time $T\tau$.

Our estimation algorithm can be modified to handle both tasks. To assess the effect of known storms, we need to run it with fixed parameter θ and record the contribution of storms beyond time $T\tau$. This should not be done in terms of aggregates, but as an average intensity function on a sufficiently fine grid to mesh with the long term forecast. We also need to decide on a time horizon for the records after which the impact of existing storms is negligible. But this is not difficult, as we can derive an upper bound for the storm duration from our estimate for θ.

To obtain the distribution of the origin for the first storm outside the observation period, we need to track its value as an additional variable. The algorithm we have developed so far does not support this, as it keeps the origin fixed at time $T\tau + 1$. But a separate update step for the lag of storm M, similar to the transforms of Section 4.5, is easily included:

6.2.1.1   Update the lag of the last storm

This operation changes the lag of the last storm, without affecting anything else. Sample $\tilde{X}_2^M$ from its distribution conditional on the duration of the storm:

$$\log(\tilde{X}_2^M)|X_1^M, \theta \sim \mathrm{N}\left(\mu_2 - \frac{[\check{\Lambda}_{11}]_{1,2}}{[\check{\Lambda}_{11}]_{2,2}}(\log(X_1^M) - \mu_1), [\check{\Lambda}_{11}]_{2,2}\right) \tag{6.9}$$

The move can be rejected outright if $O^M + X_2^M \leq T\tau$, as this would contradict the fact that storm M is the last. Else, the acceptance probability for the move is given by

$$\alpha_7(\Pi, \tilde{\Pi}) = \prod_{j=1}^{M^M} \frac{p_{X^{i,j}}(\tilde{X}^{M,j}|\tilde{X}^M, \theta)}{p_{X^{i,j}}(X^{M,j}|X^M, \theta)} \tag{6.10}$$

This expression reduces to 1 (a Gibbs sampler) in the simplified model with diagonal Λ.

6.2.2   Long term forecast

To analyze the long term behavior of the model, we need to modify its representation. So far, we have worked with a finite number M of storms occurring in the time interval $[0, T\tau]$. Now, we are interested in the behavior of storms beyond that period. Thus, we assume in the following that the set $\{X^i\}$ is an infinite family of i.i.d. random variables.

The cell quantities $\{X^{i,j}\}$ for each storm i are treated in the same manner. I.e., each storm has an infinite number of cells, but only those whose origins lie in $[O^i, O^i + X_1^i]$ contribute

to precipitation. This modification is not strictly necessary, but it simplifies calculations as we do not have to condition the distribution of the cell lags on the number of cells.

Using the new representation, the intensity function becomes

$$I(x) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} D^{i,j} \mathbb{1}_{[O^{i,j}, O^{i,j} + X_1^{i,j}]}(x) \mathbb{1}_{[O^i, O^i + X_1^i]}(O^{i,j}) \tag{6.11}$$

$$= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} D^{i,j} \mathbb{1}_{[\max(0, x - O^i - X_1^{i,j}), \min(X_1^i, x - O^i)]} \left(\sum_{l=1}^{j-1} X_2^{i,l} \right)$$

The reformulation uses equation (3.1), which relates lags and origins. As the storm origins converge to infinity with probability 1, the sum has only finitely many non-zero terms. To show that its expectation exists, we use the relation

$$0 \leq I(x) \leq 2\iota \sum_{i=1}^{\infty} \mathbb{1}_{[O^i, \infty)}(x) \left(\sum_{j=1}^{\infty} \mathbb{1}_{[0, O^i + X_1^i]}(O^{i,j}) \right) \tag{6.12}$$

The $D^{i,j}$ are bounded above by 2ι as they are uniformly distributed on $[0, 2\iota]$, and the indicator functions are replaced by indicators for supersets. Note that

$$E(\mathbb{1}_{[O^i, \infty)}(x)|\theta) = P\left(\sum_{l=1}^{i-1} X_2^l \leq x \,\middle|\, \theta \right) \tag{6.13}$$

$$\leq P(\forall l \in \{1, \ldots, i-1\} : X_2^l \leq x|\theta)$$

$$= P(X_2^1 < x|\theta)^{i-1} =: p_0^{i-1}$$

$$E(\mathbb{1}_{[0, O^i + X_1^i]}(O^{i,j})|\theta) = P\left(O^i + \sum_{l=1}^{j-1} X_2^{i,j} \leq O^i + X_1^i \,\middle|\, \theta \right) \tag{6.14}$$

$$= P\left(\sum_{l=1}^{j-1} X_2^{1,j} \leq X_1^1 \,\middle|\, \theta \right)$$

$$\leq P(X_2^{1,1} \leq X_1^1|\theta)^{j-1} =: p_1^{j-1}$$

Thus, the expectation of the infinite sums in (6.12) can be bounded above by geometric series and we get

$$0 \leq E(I(x)|\theta) \leq 2\iota \sum_{i=1}^{\infty} E(\mathbb{1}_{[O^i, \infty)}(x)|\theta) \left(\sum_{j=1}^{\infty} E(\mathbb{1}_{[0, O^i + X_1^i]}(O^{i,j})|\theta) \right) \tag{6.15}$$

$$\leq \frac{2\iota}{(1 - p_0)(1 - p_1)} < \infty$$

To calculate the expectation numerically, we use the alternate formulation given in (6.11). If we define

$$f(a, b, c, d) := \mathbb{1}_{[\max(0, a-b), \min(c, a)]}(d) \tag{6.16}$$

the expectation can be written as

$$E(I(x)|\theta) = \iota \sum_{i=1}^{\infty} \int_0^{\infty} \sum_{j=1}^{\infty} E\left(f\left(x - O^i, X_1^{1,j}, X_1^1, \sum_{l=1}^{j-1} X_2^{1,j} \right) \middle| O^i, \theta \right) dP(O^i|\theta) \quad (6.17)$$

The quantities for storm i can be replaced with those for storm 1 as they are independent of O^i. This allows us to estimate f without referring to the distribution of O^i. Given a grid

$$\Gamma := \{nh : n \in \{0, \ldots, N\}\} \quad (6.18)$$

with grid size $h > 0$ and $N + 1$ points, we calculate the expected intensity in three steps:

1. Estimate the function

$$g(y) := \sum_{j=1}^{\infty} E\left(f\left(y, X_1^{1,j}, X_1^1, \sum_{l=1}^{j-1} X_2^{1,j} \right) \middle| \theta \right) \quad (6.19)$$

   for each $y \in \Gamma$ by Monte Carlo integration. I.e., we draw samples $(X_1^1)^k$, $\{(X_1^{1,j})^k\}_j$, and $\{(X_2^{1,j})^k\}_j$ with $k \in \{1, \ldots, K\}$ for sufficiently large K and $j \in \{1, \ldots, j_k\}$ where j_k satisfies

$$\sum_{l=1}^{j_k-1} (X_2^{1,l})^k \le (X_1^1)^k < \sum_{l=1}^{j_k} (X_2^{1,l})^k \quad (6.20)$$

   Using the index sets

$$M_y^k := \left\{ j \in \{1, \ldots, j_k\} : f\left(y, (X_1^{1,j})^k, (X_1^1)^k, \sum_{l=1}^{j-1} (X_2^{1,l})^k \right) = 1 \right\} \quad (6.21)$$

   we can estimate $g(y)$ on Γ as

$$\hat{g}(y) := \frac{1}{K} \sum_{k=1}^{K} |M_y^k| \quad (6.22)$$

2. Calculate the distribution of each O^i on Γ as the $i - 1$-fold convolution of the lognormal lag density with itself. Do this until a cutoff point i_0 where the total mass remaining on Γ is negligible.

3. Estimate the expected intensity for each $x \in \Gamma$ by

$$E(I(x)|\theta) \approx \iota \sum_{i=1}^{i_0} \int_0^{Nh} \hat{g}(x - O^i) \, dP(O^i|\theta) \quad (6.23)$$

   where the integrals are again solved numerically on the grid. As $P(O^1|\theta)$ puts all mass on the origin, the first integral is equal to $\hat{g}(x)$.

This procedure yields the expected intensity given that the first storm starts at $t = 0$. If we replace the distribution of O^1 by the conditional distribution for O^{M+1} derived from the short term forecast, we can use the scheme for a general long term forecasts.

Figure 6.1: Long term prediction for artificial data

(a) Average number of cells during a single storm

(b) Average intensity
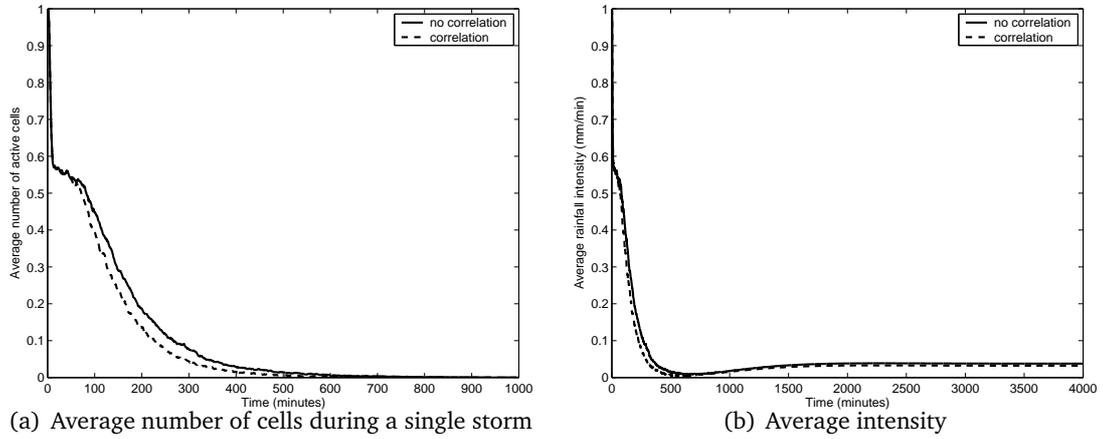
Figure 6.2: Artificial data — Effects of storm lag precision
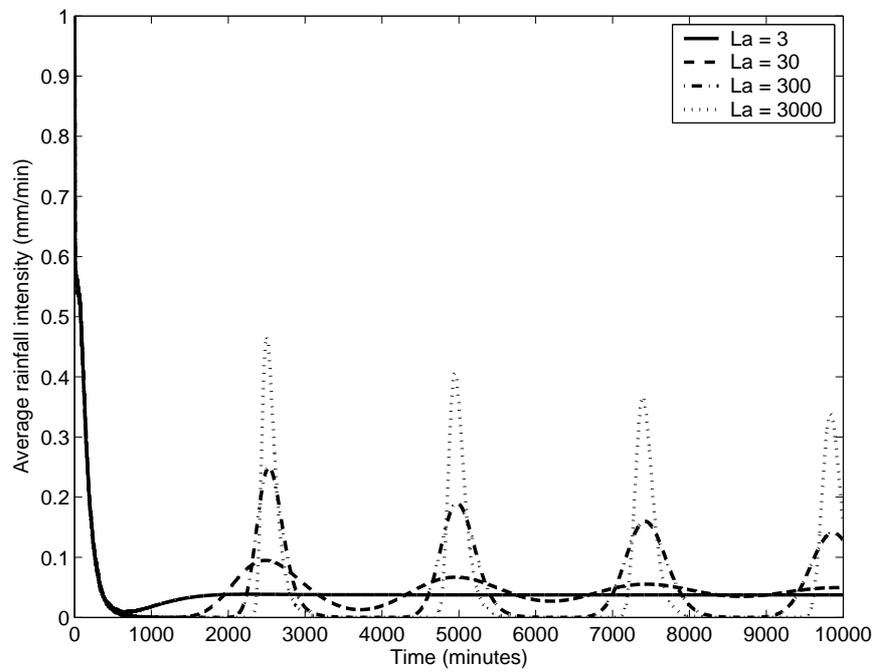
6.2.3   Some numerical results

The scheme for the long term forecast was implemented in MATLAB. The results in this section were calculated on a grid with $h = 1$ minute and using $K = 10,000$ iterations to estimate g. Computation only takes a few minutes on a regular PC.

First, we want to look at the function g itself. This has a nice interpretation — it is the expected number of cells that contribute to precipitation at time y for a storm that starts at time 0. If we multiply it with the intensity scale ι, we get an average precipitation curve for storms under the model. The left plot in Figure 6.1 shows $\hat{g}$ for the parameters used to generate the artificial data sets A and C of Section 5.2. As set C assumes a negative correlation between the duration of storms and cells, we expect fewer active cells on average as the distance from the origin increases. This is evident in the picture.

In the right plot of Figure 6.1, we see the mean intensity $E(I(x))$ for the same two choices of the parameter θ. As we assume $O^1 = 0$, the curve initially resembles $g(x)$, but as time progresses, the contribution of other storms becomes dominant. Judging from (6.23), we would expect to see a series of decreasing peaks at regular intervals. This is due to the fact that the iterated convolution density of the storm lags has an ever increasing variance and 'smears' copies of g across the positive real line. However, the variance in the example is so large that $E(I(x))$ resembles a constant very quickly. For comparison, Figure 6.2 shows the expected intensity for lag precision $\Lambda_{2,2} \in \{3, 30, 300, 3000\}$.

Appendix A

Numerical results — Figures

Figure A.1: Error precision scale ζ for trial A1a (Model II)

Short cells, variable ζ, diagonal Λ, 10 min intervals
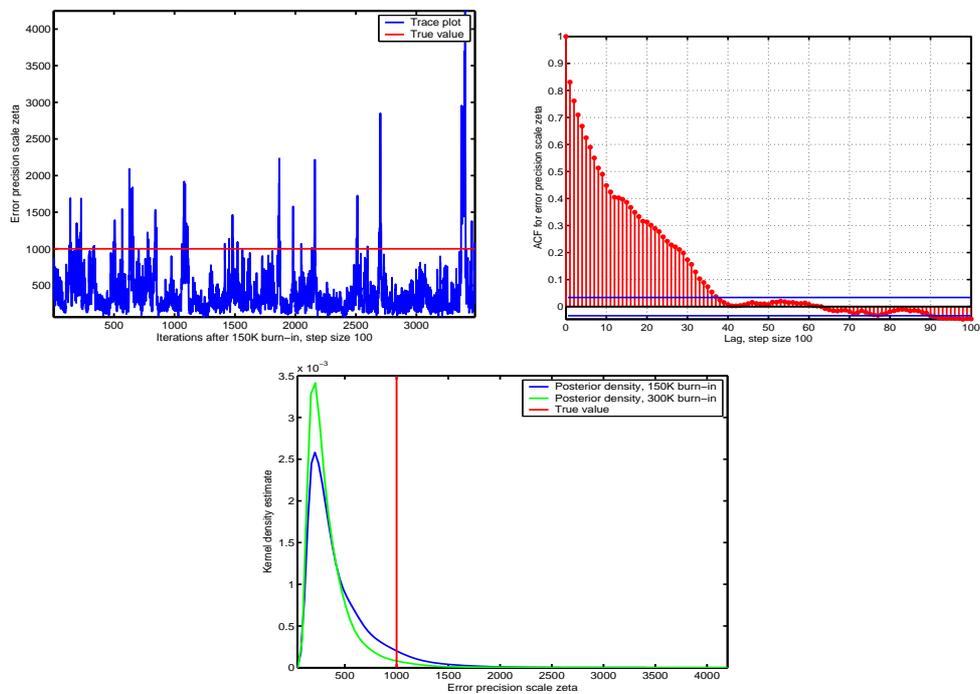
Figure A.2: Mean parameters μ_k for trial A1a (Model II)

Short cells, variable ζ, diagonal Λ, 10 min intervals
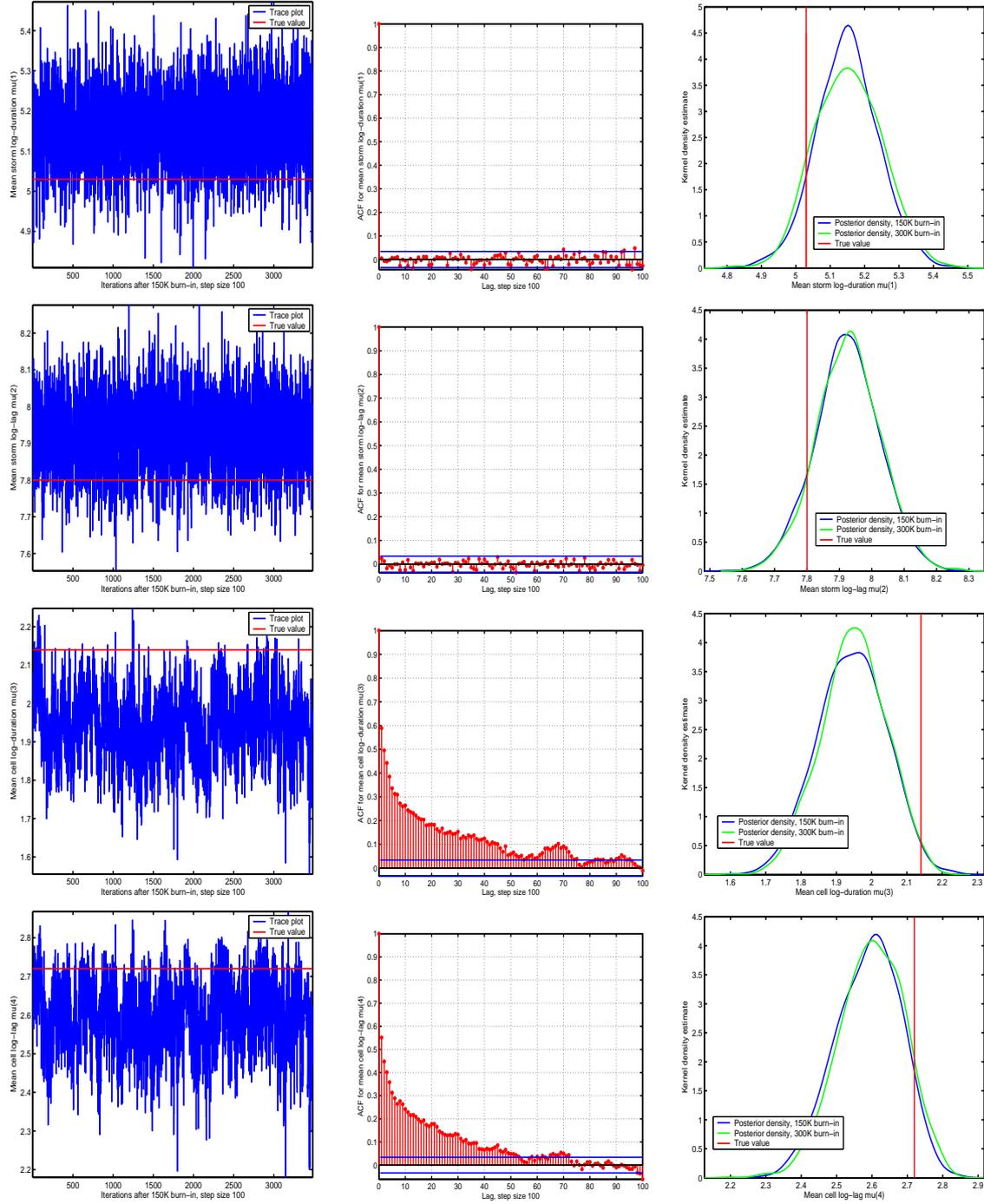
Figure A.3: Variances σ_k^2 for trial A1a (Model II)

Short cells, variable ζ, diagonal Λ, 10 min intervals
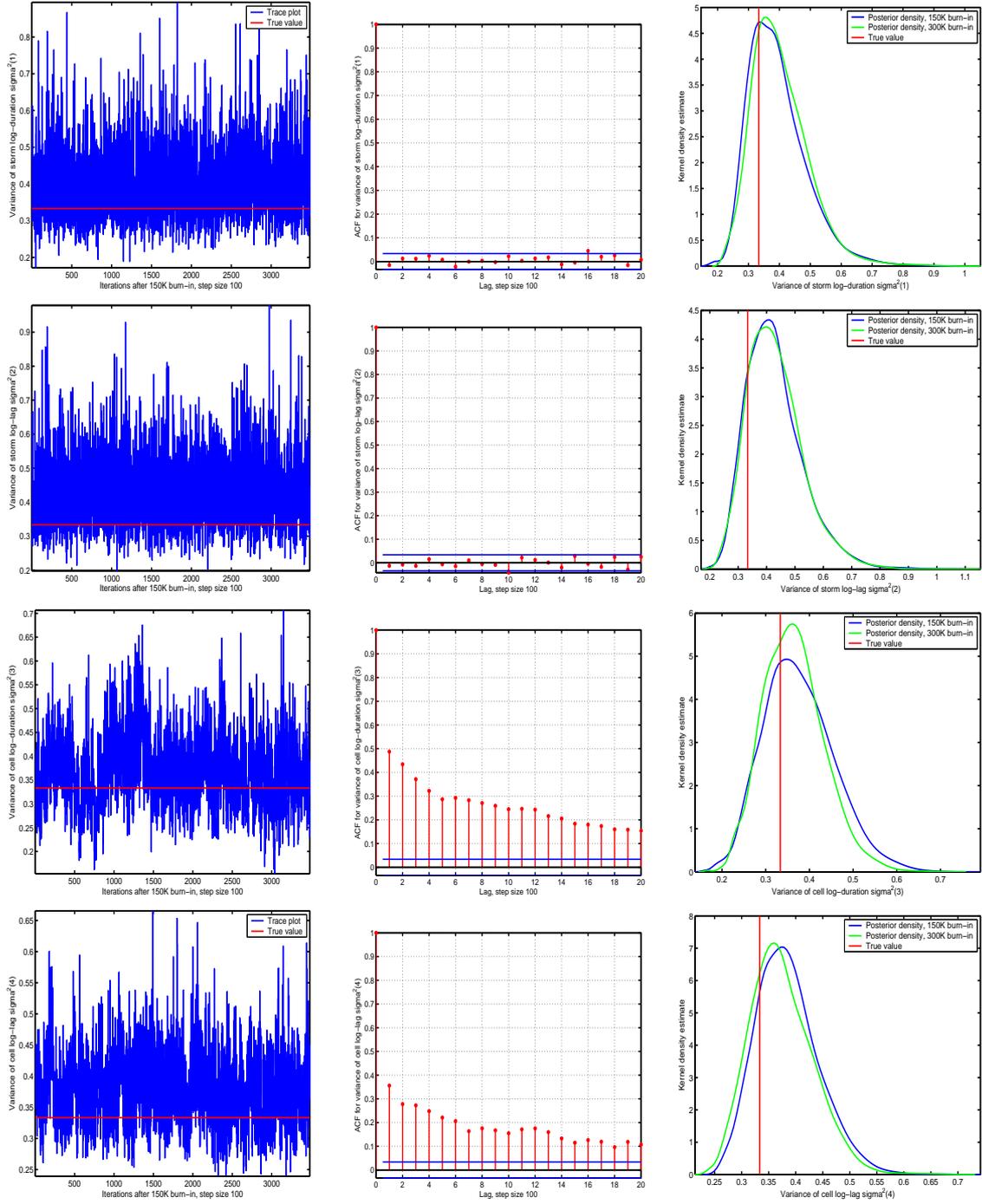
Figure A.4: Rainfall characteristics for trial A1a (Model II)

Short cells, variable ζ, diagonal Λ, 10 min intervals
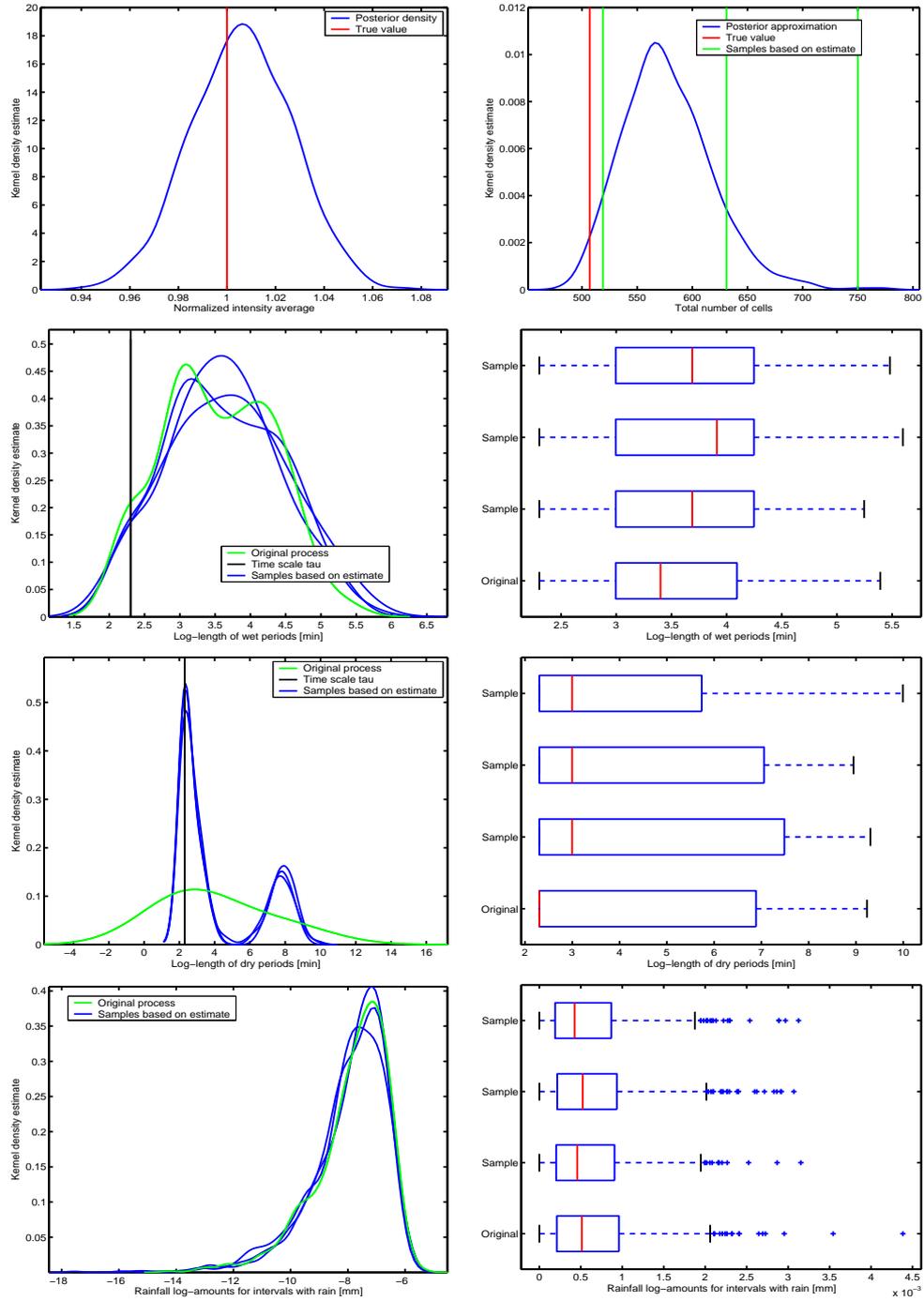
Figure A.5: Mean parameters μ_k for trial C1d (Model III)

Short cells, fixed ζ, mean matching, restricted Λ, 10 min intervals
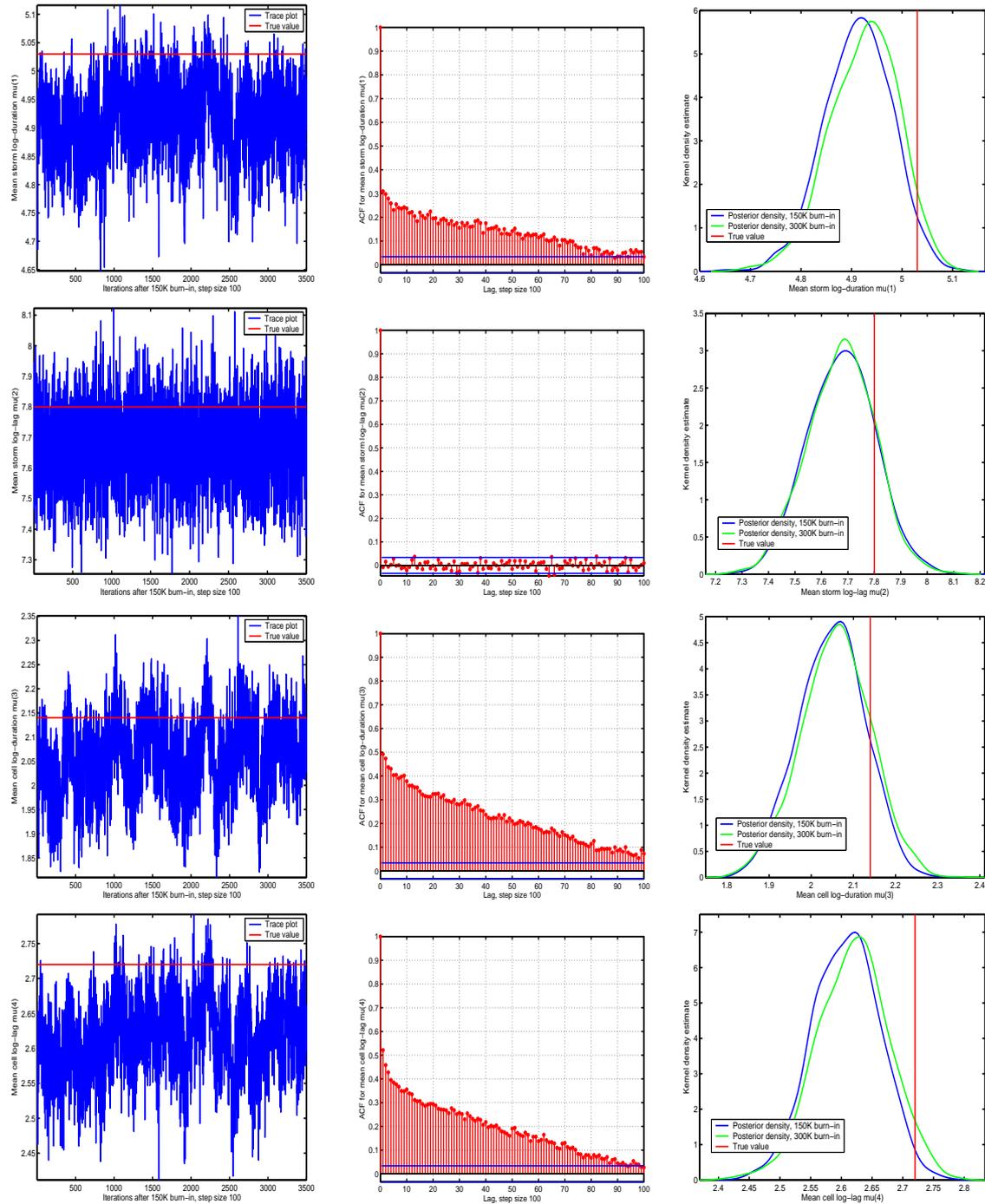
Figure A.6: Variances σ_k^2 for trial C1d (Model III)

Short cells, fixed ζ, mean matching, restricted Λ, 10 min intervals
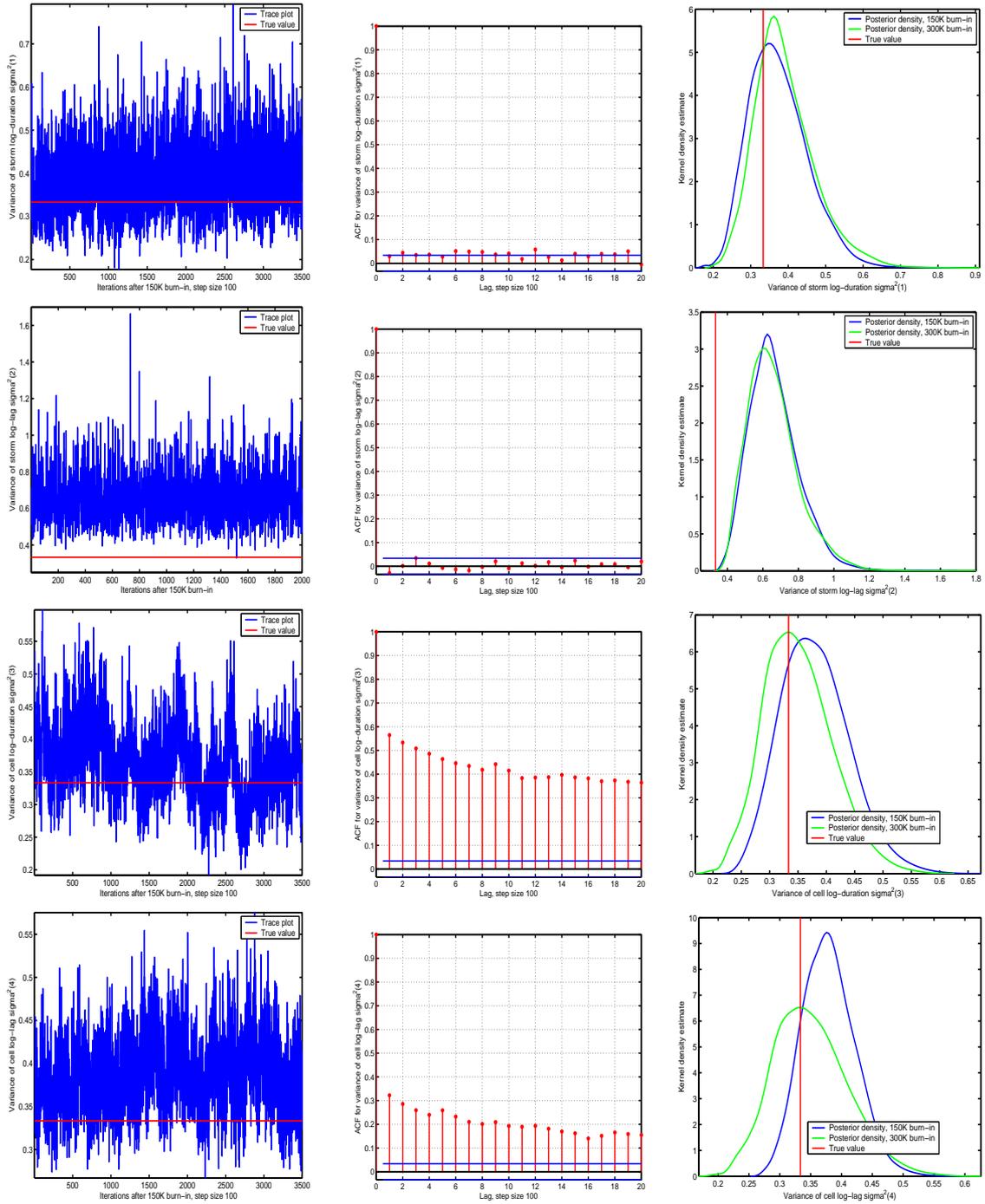
Figure A.7: Correlations $\rho_{k,l}$ for trial C1d (Model III)

Short cells, fixed ζ, mean matching, restricted Λ, 10 min intervals
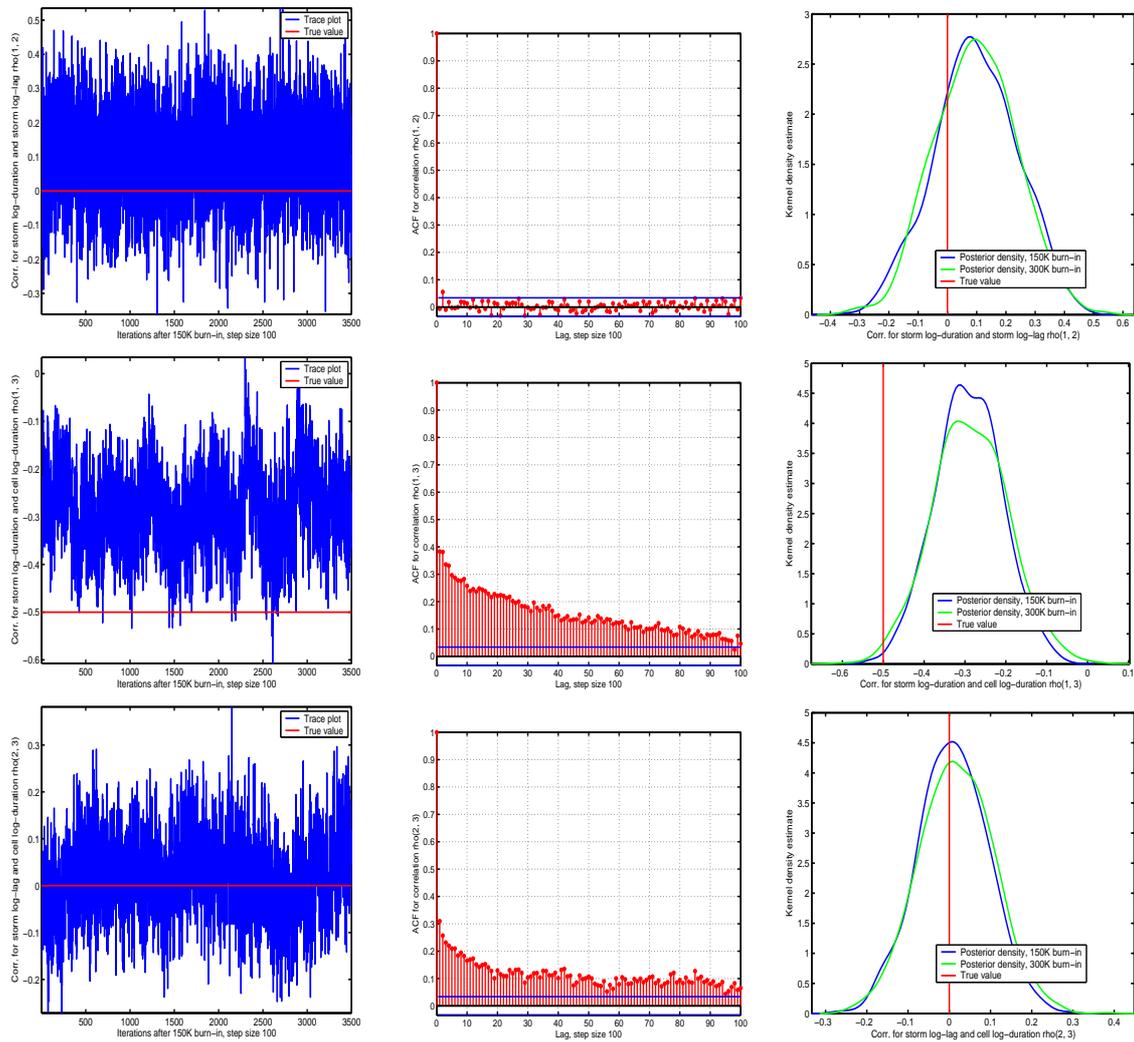
Figure A.8: Comparing the estimators for trials C1d and C1d alt (Model III)

Short cells, fixed ζ, mean matching, restricted Λ, 10 min intervals
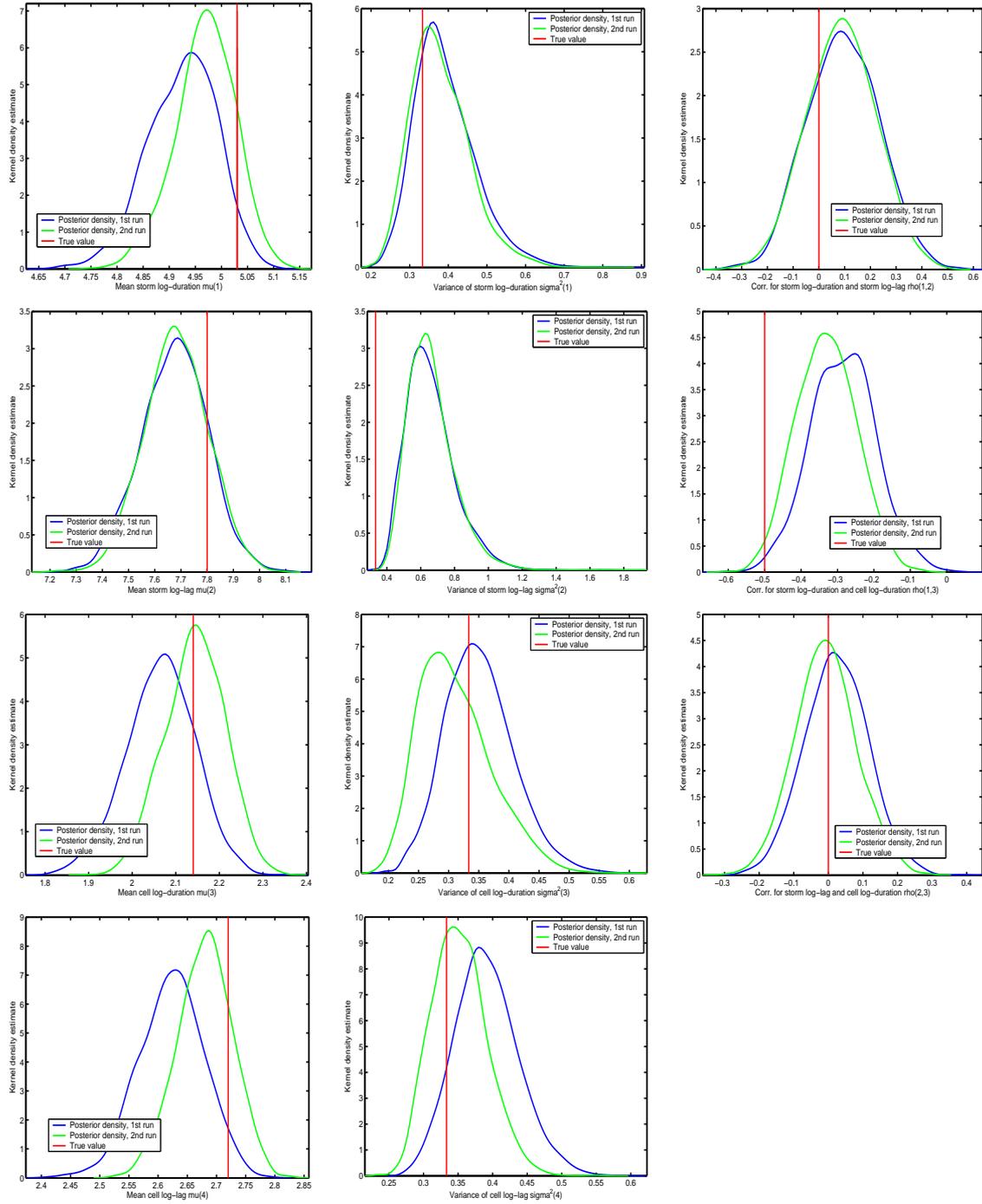
Figure A.9: Rainfall characteristics for trials C1d and C1d alt (Model III) — Part 1

Short cells, fixed ζ, mean matching, restricted Λ, 10 min intervals
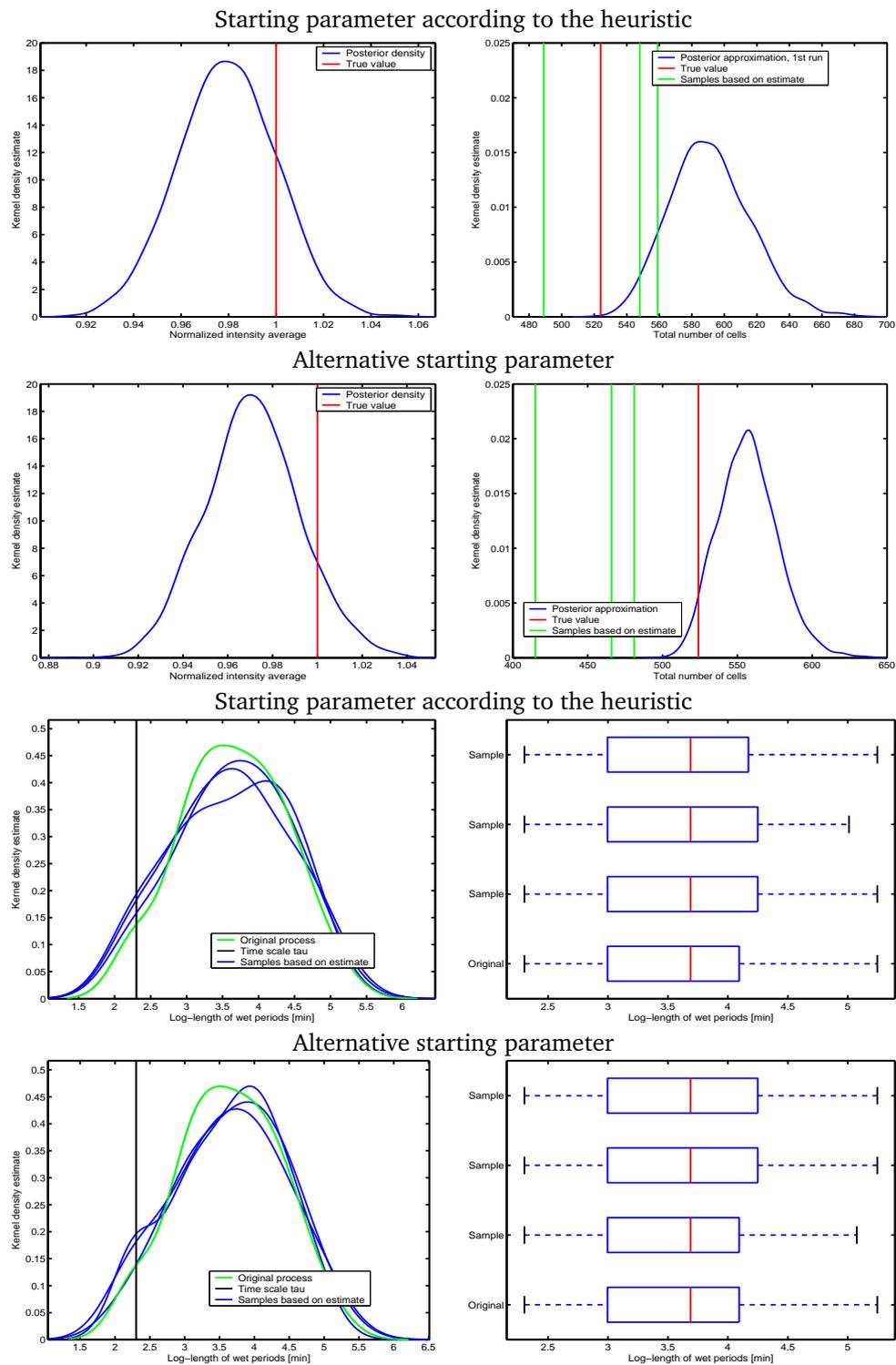
Figure A.10: Rainfall characteristics for trials C1d and C1d alt (Model III) — Part 2

Short cells, fixed ζ, mean matching, restricted Λ, 10 min intervals
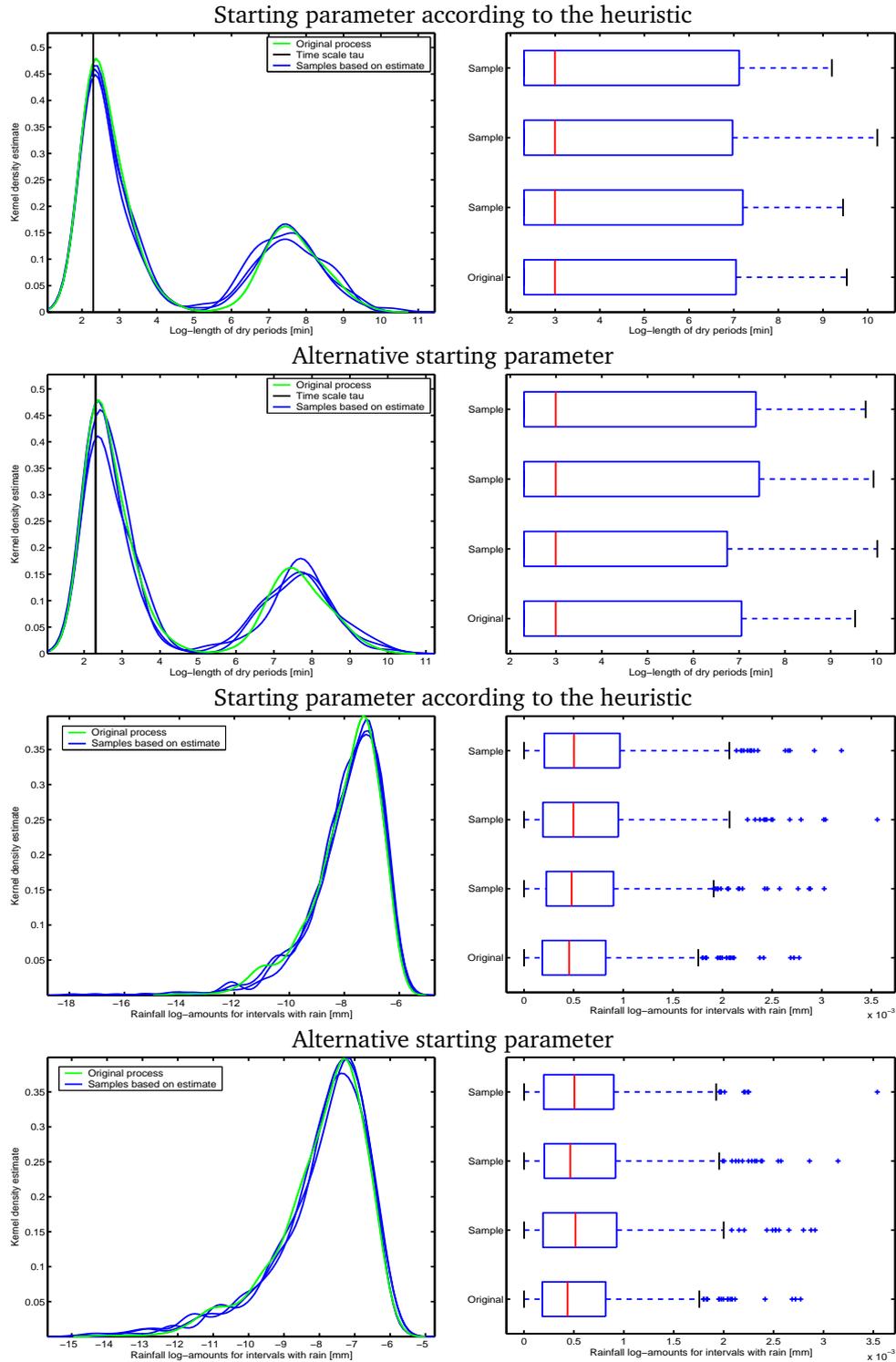
Figure A.11: Mean parameters μ_k for trial C2d (Model III)

Short cells, fixed ζ, mean matching, restricted Λ, 1 h intervals
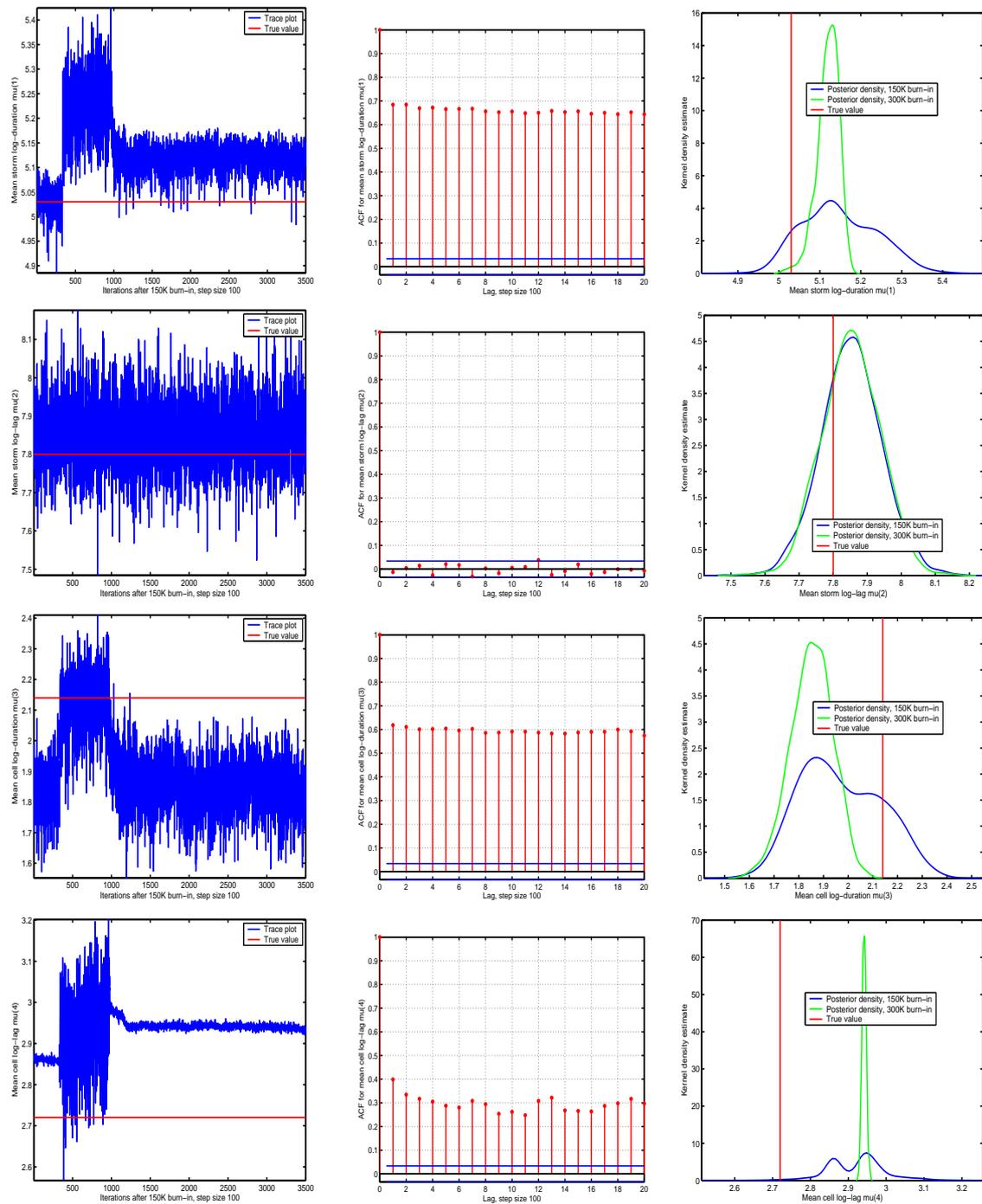
Figure A.12: Variances σ_k^2 for trial C2d (Model III)

Short cells, fixed ζ, mean matching, restricted Λ, 1 h intervals
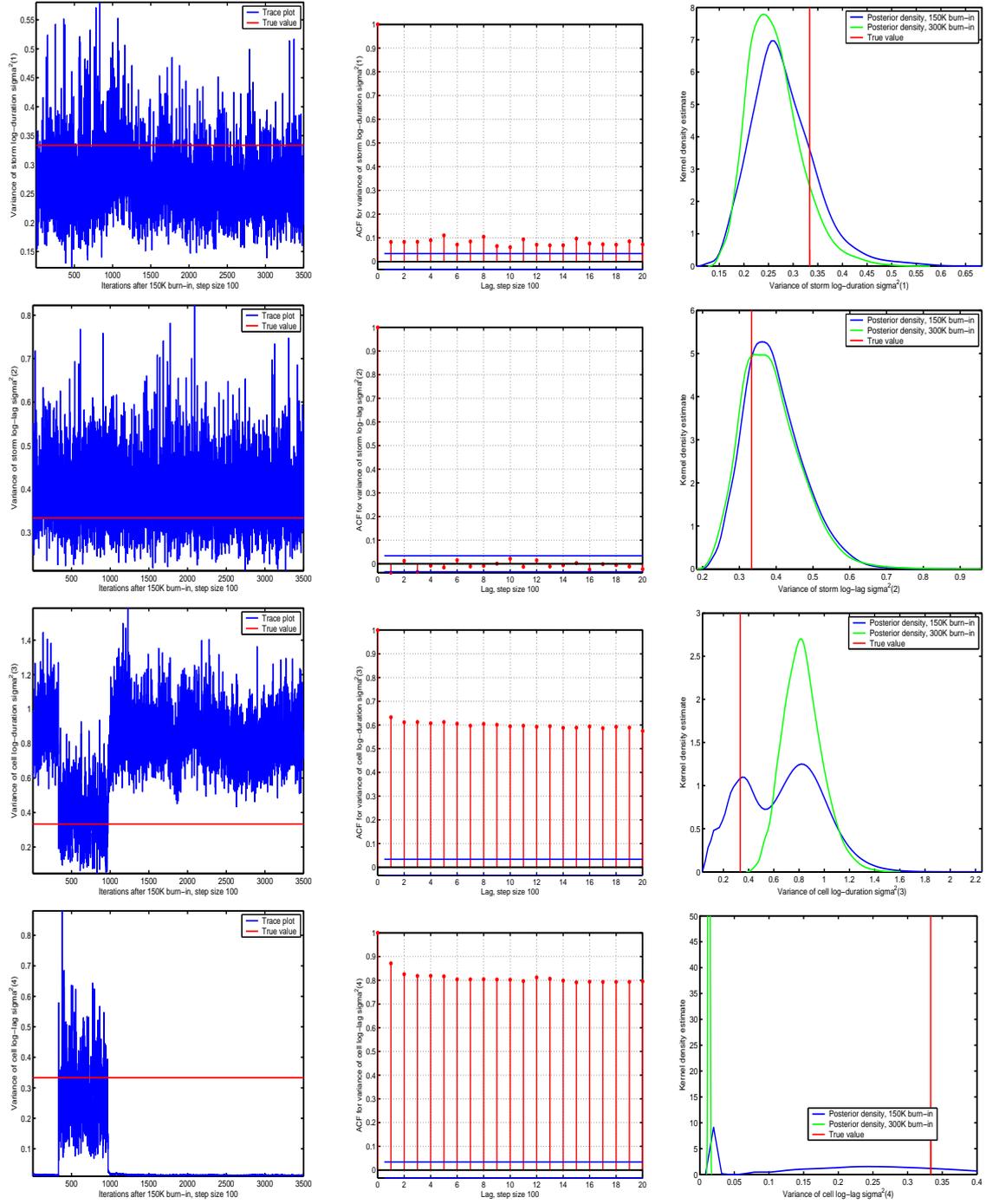
Figure A.13: Correlations $\rho_{k,l}$ for trial C2d (Model III)

Short cells, fixed ζ, mean matching, restricted Λ, 1 h intervals
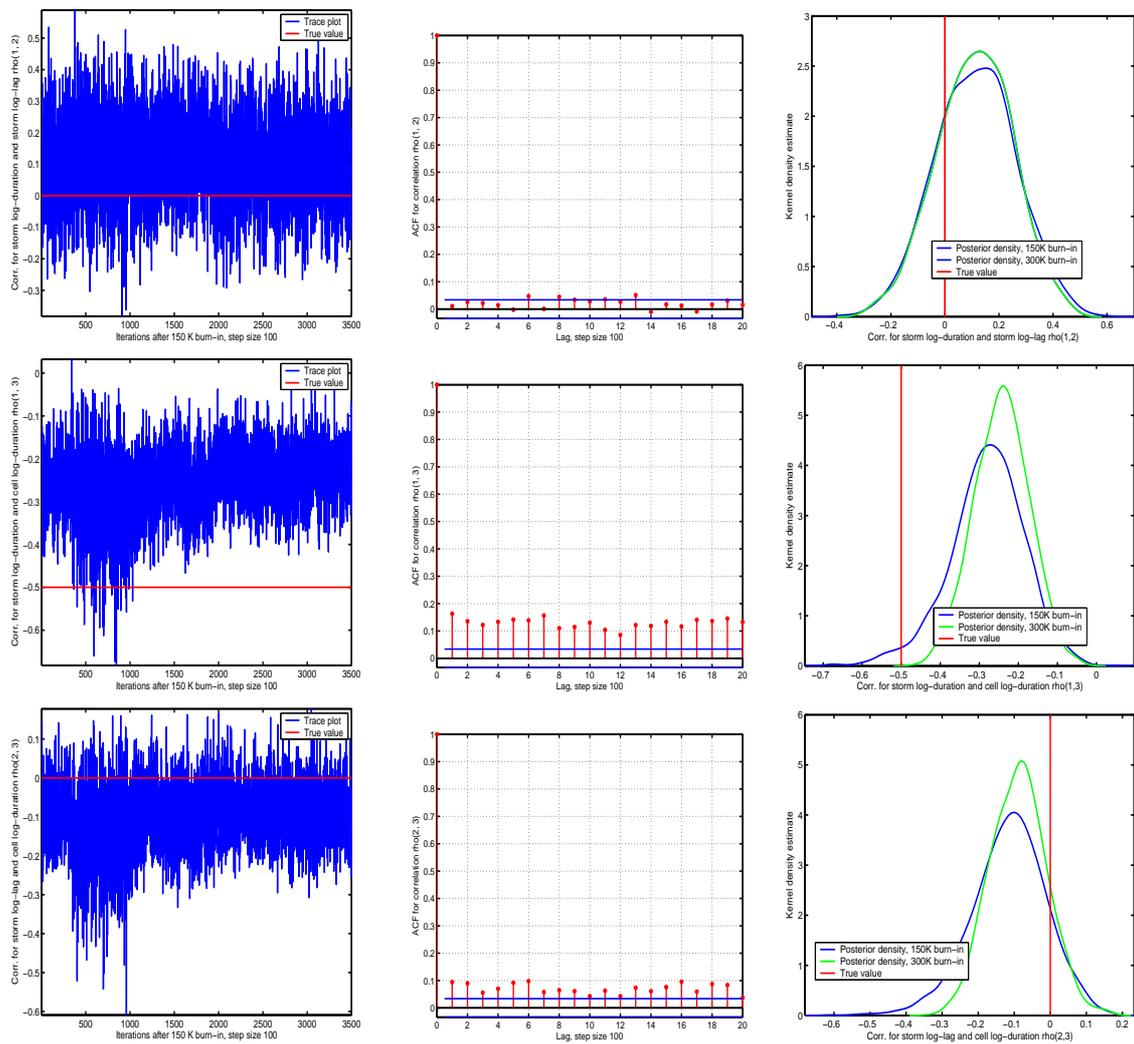
Figure A.14: Rainfall characteristics for trial C2d (Model III)

Short cells, fixed ζ, mean matching, restricted Λ, 1 h intervals
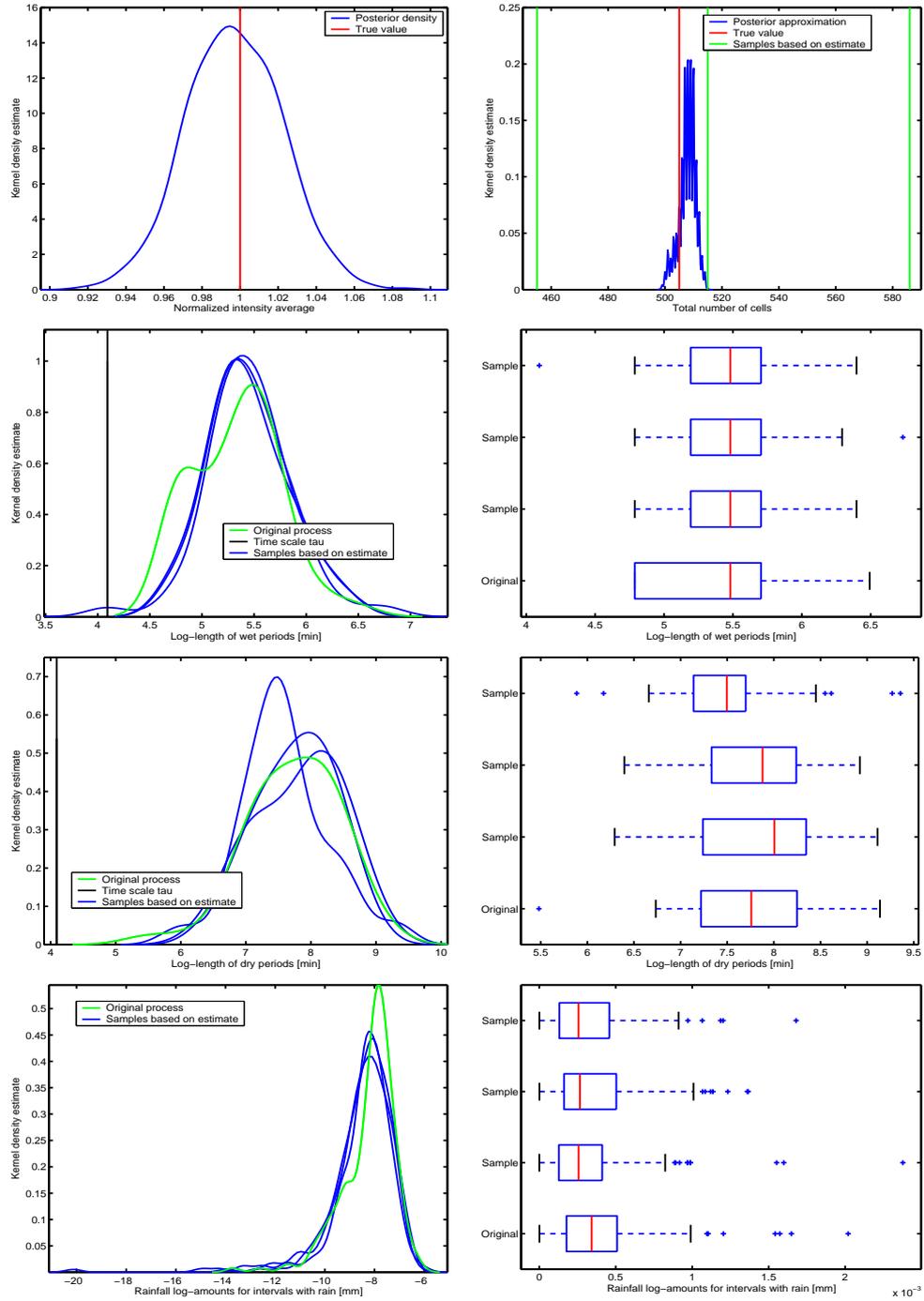
Figure A.15: Mean parameters μ_k for trial D1d (Model III)

Long cells, fixed ζ, mean matching, restricted Λ, 10 min intervals
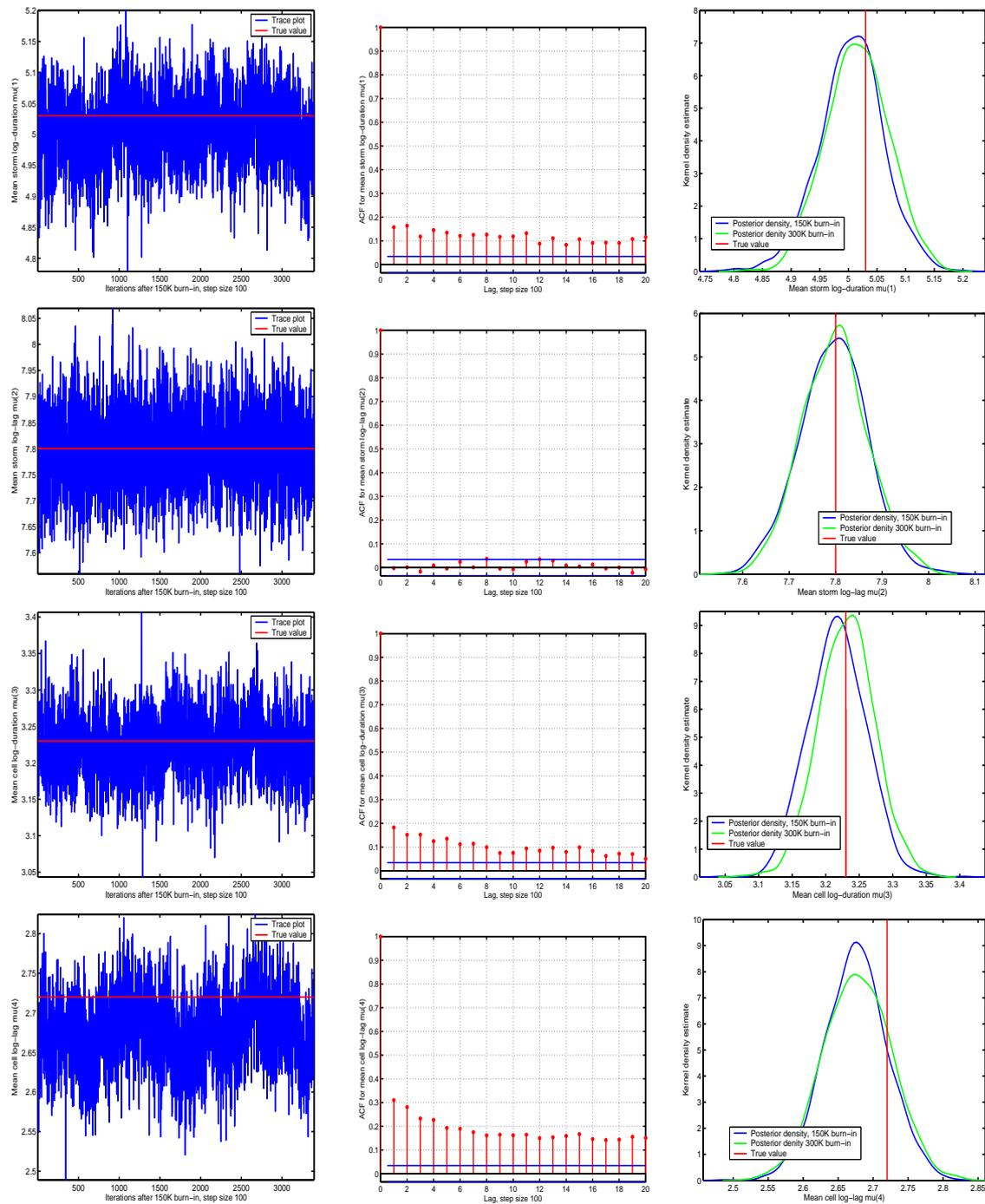
Figure A.16: Variances σ_k^2 for trial D1d (Model III)

Long cells, fixed ζ, mean matching, restricted Λ, 10 min intervals
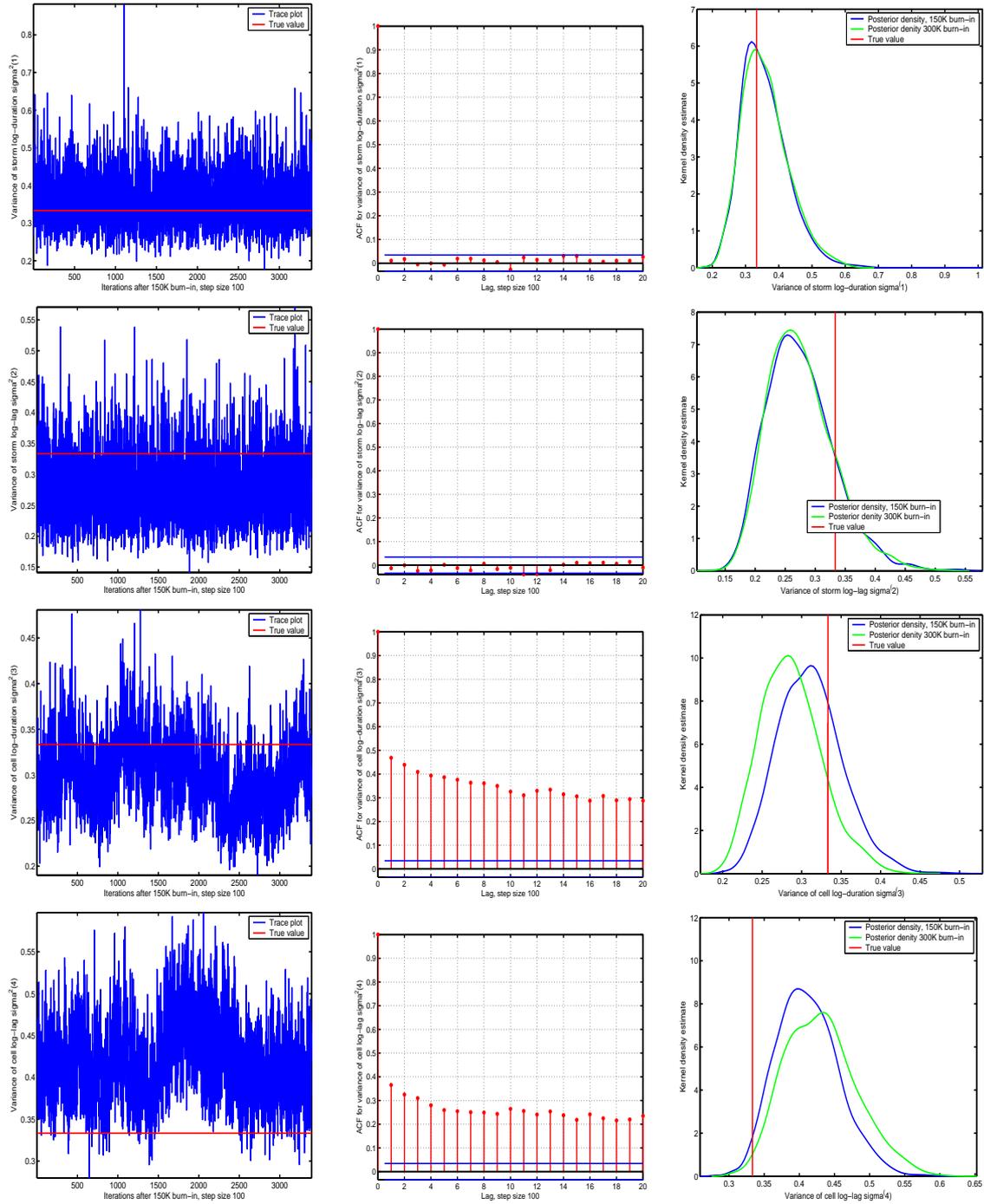
Figure A.17: Correlations $\rho_{k,l}$ for trial D1d (Model III)

Long cells, fixed ζ, mean matching, restricted Λ, 10 min intervals
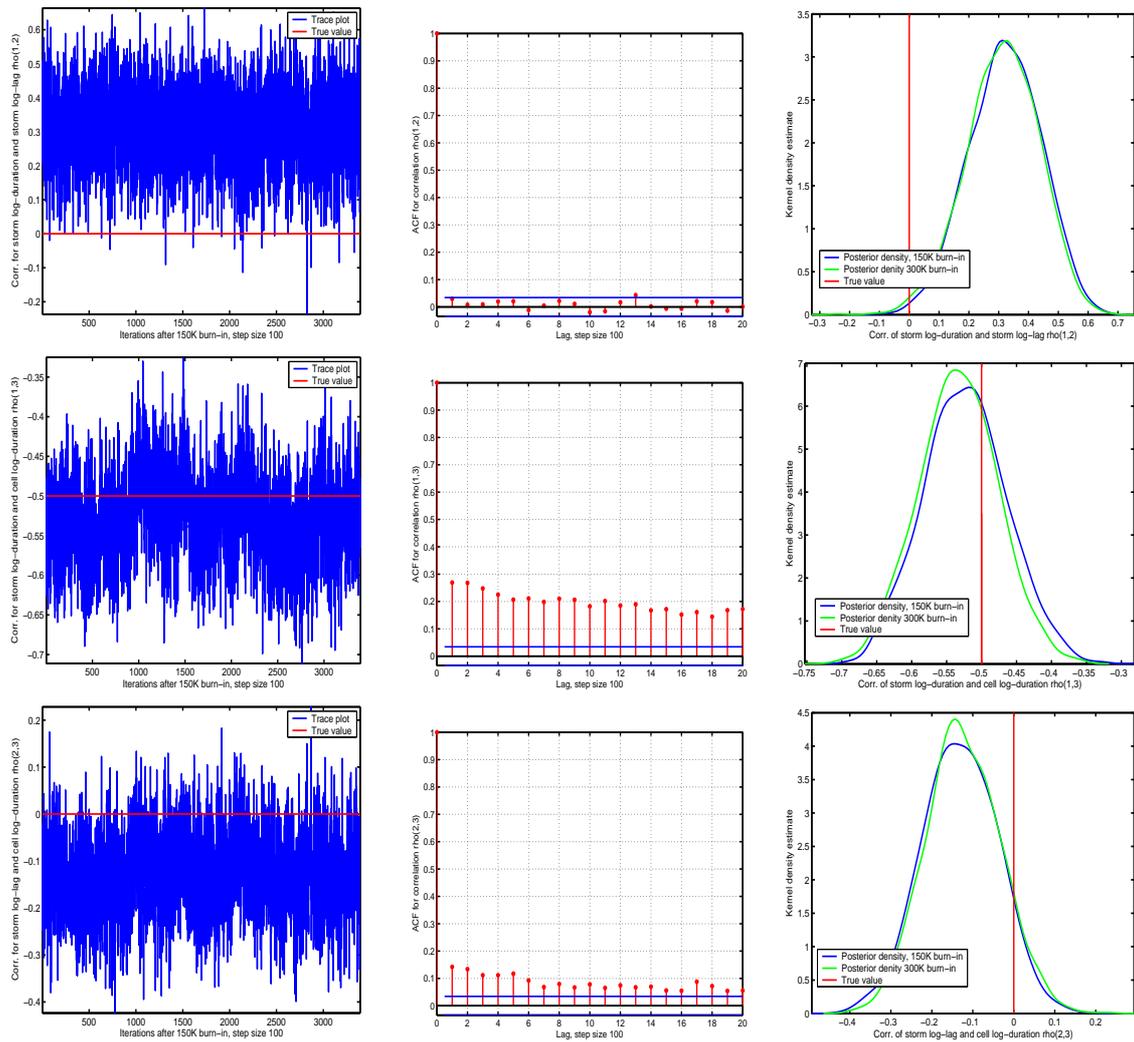
Figure A.18: Rainfall characteristics for trial D1d (Model III)

Long cells, fixed ζ, mean matching, restricted Λ, 10 min intervals
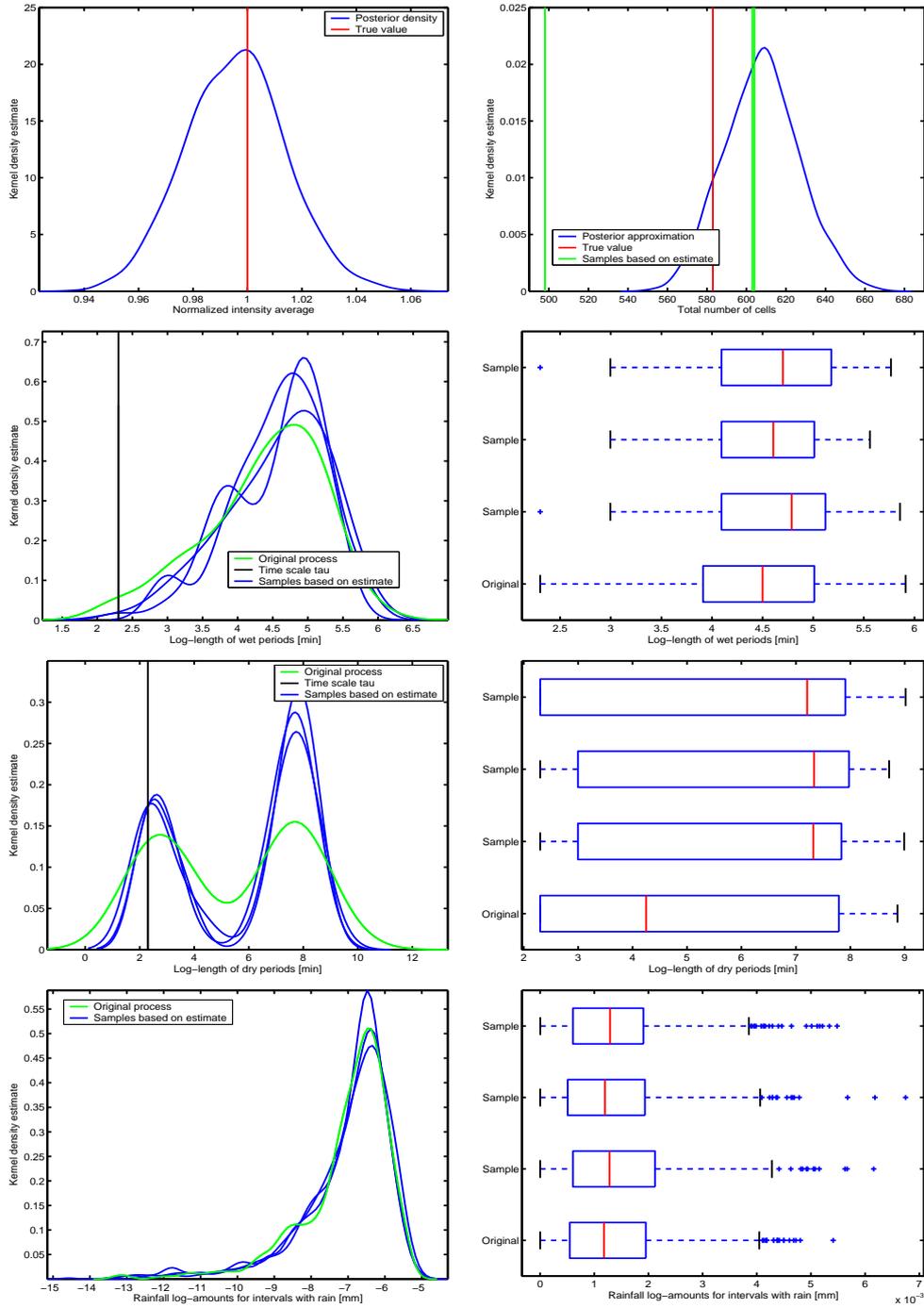
Figure A.19: Mean parameters μ_k for trial F2b (Model V)

Long cells, fixed ζ, restricted Λ, 1 h intervals
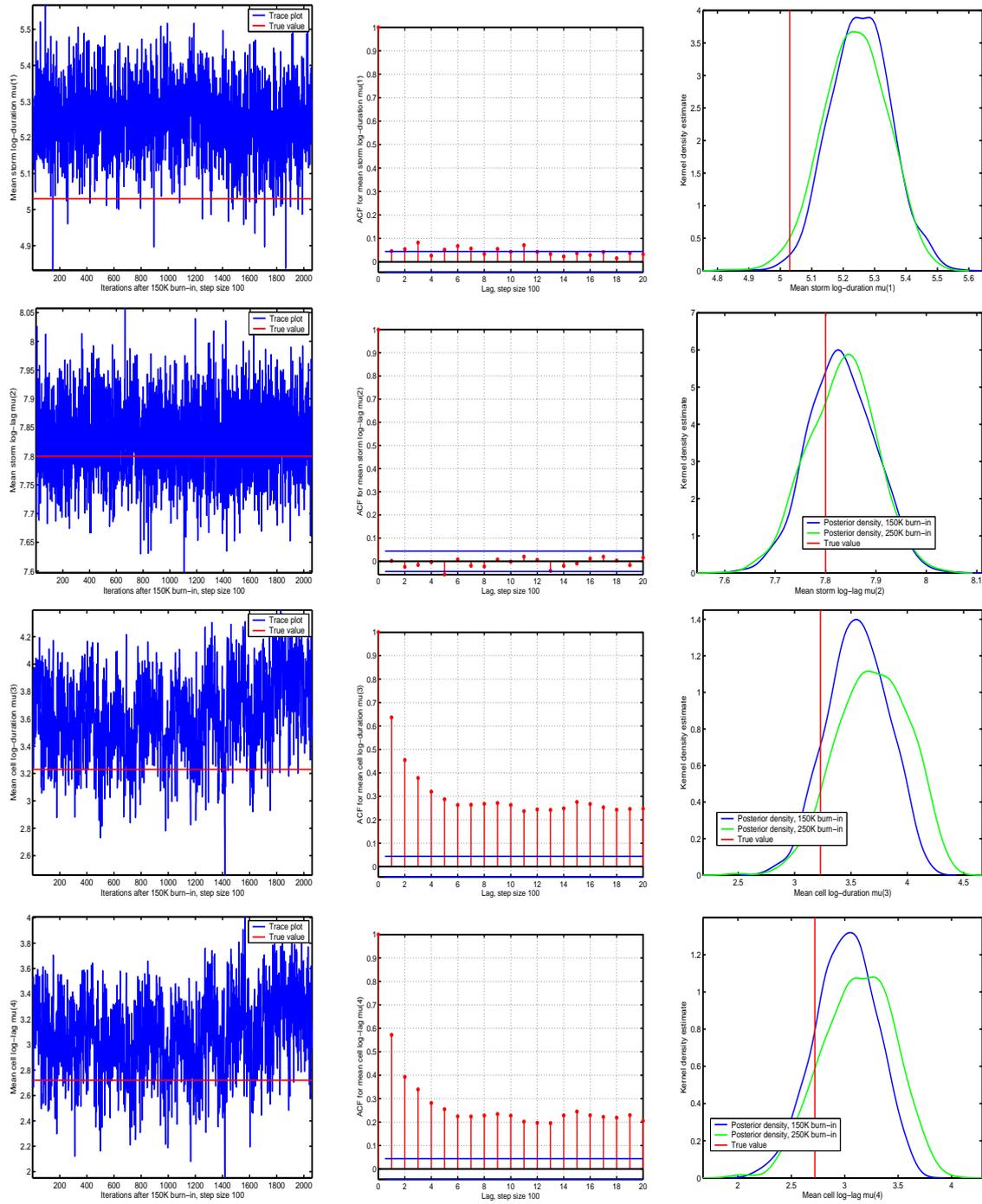
Figure A.20: Variances σ_k^2 for trial F2b (Model V)

Long cells, fixed ζ, restricted Λ, 1 h intervals
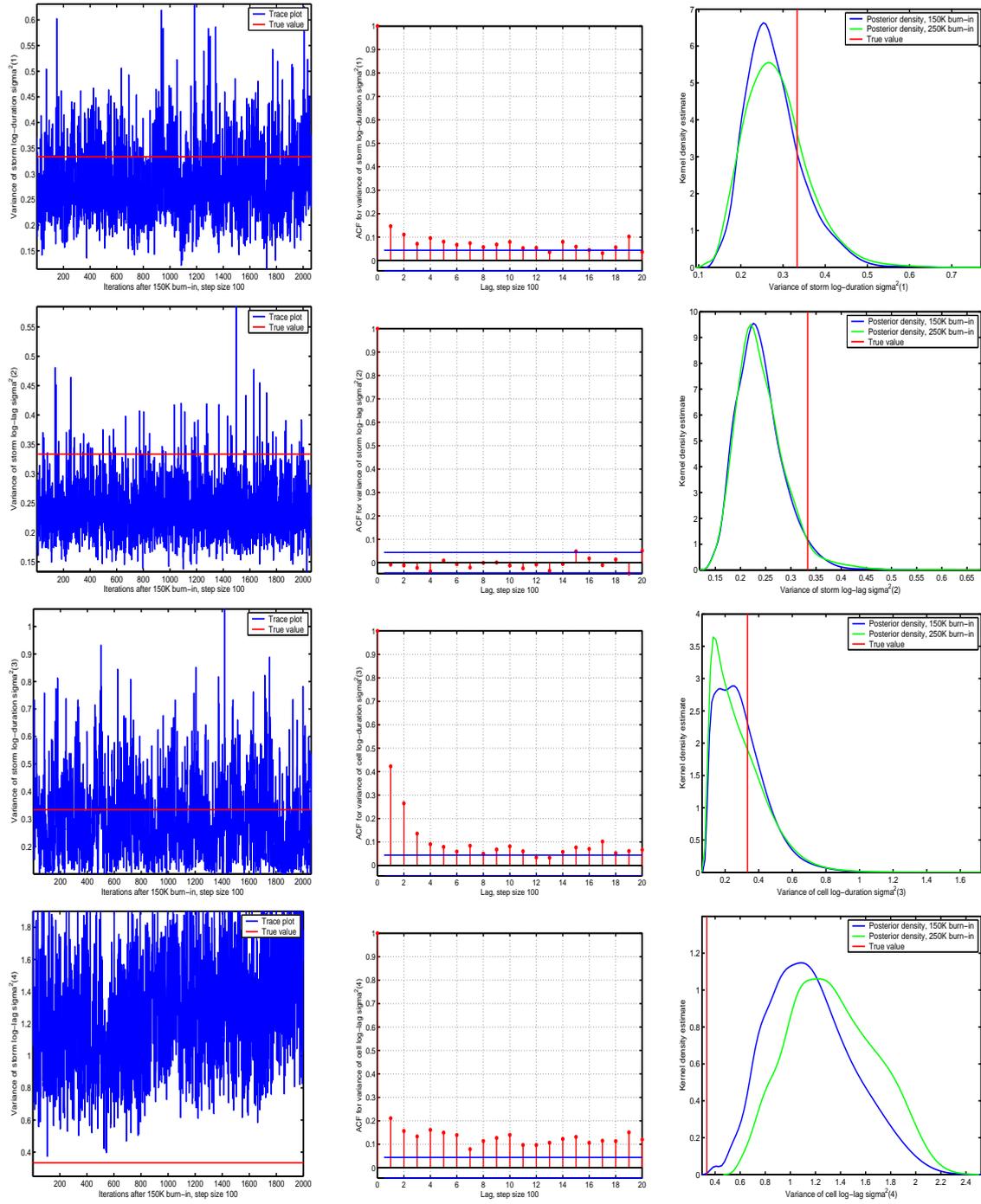
Figure A.21: Correlations $\rho_{k,l}$ for trial F2b (Model V)

Long cells, fixed ζ, restricted Λ, 1 h intervals
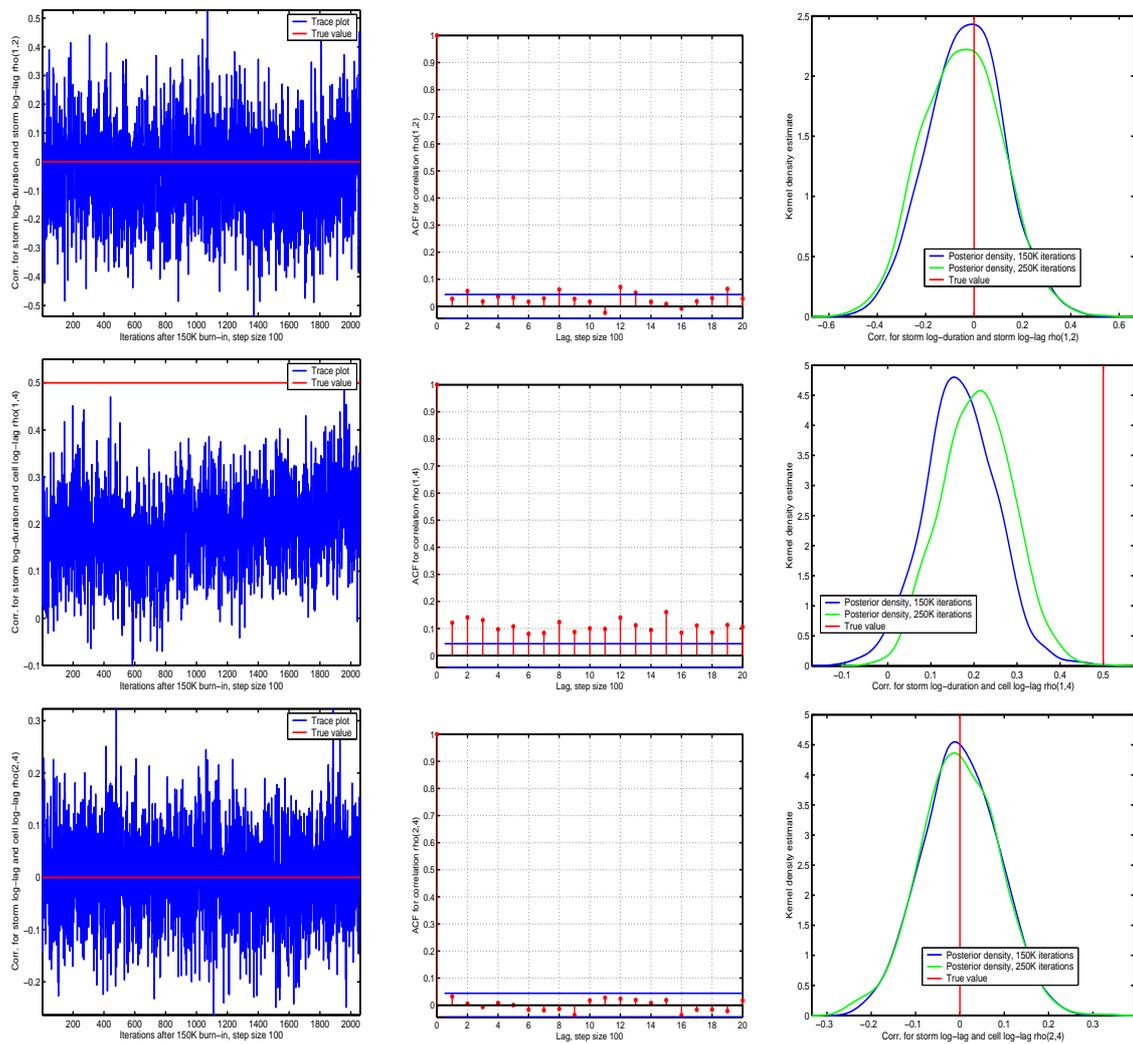
Figure A.22: Comparing the estimators for trials F2b and F2b alt (Model V)

Long cells, fixed ζ, restricted Λ, 1 h intervals
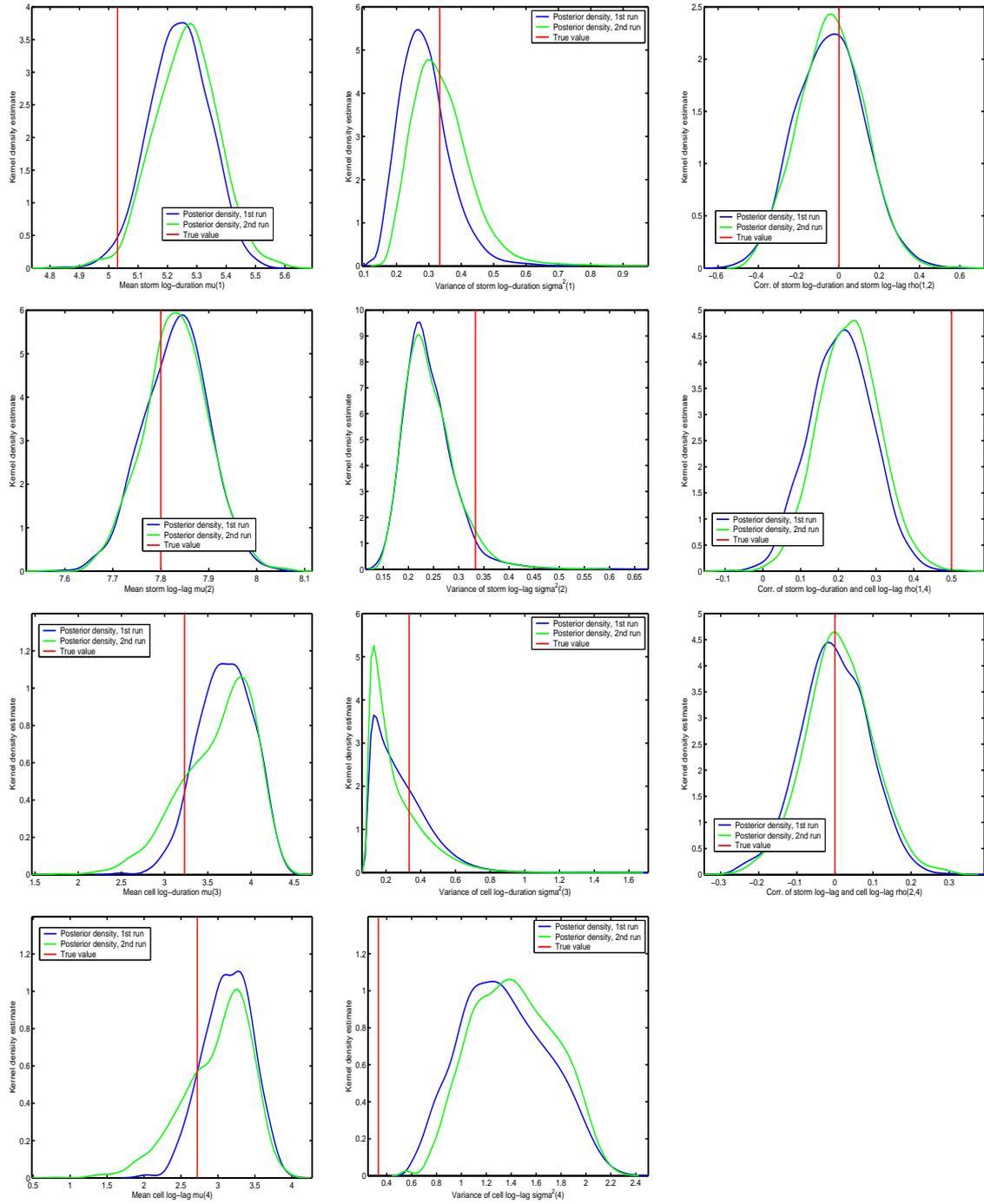
Figure A.23: Rainfall characteristics for trials F2b and F2b alt (Model V) — Part 1

Long cells, fixed ζ, restricted Λ, 1 h intervals

Starting parameter according to the heuristic
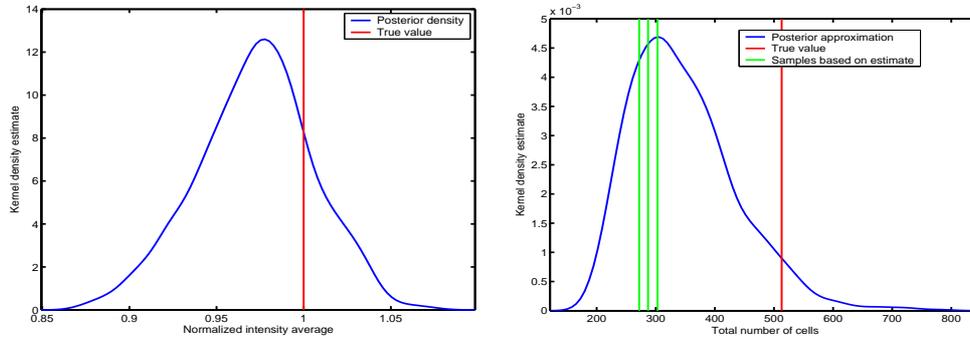
Alternative starting parameter
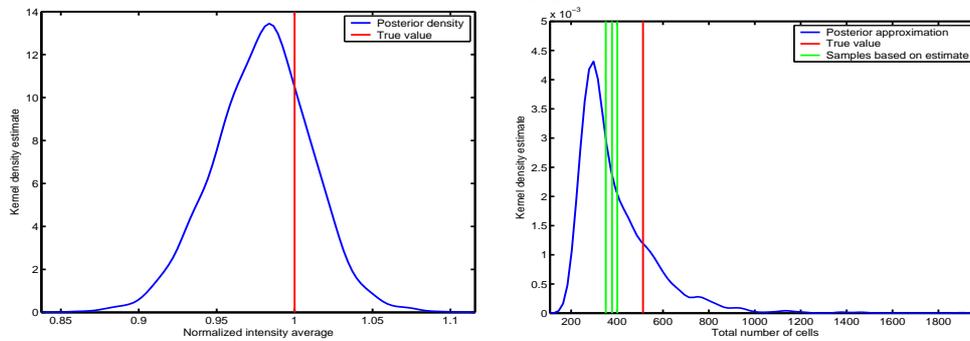
Starting parameter according to the heuristic
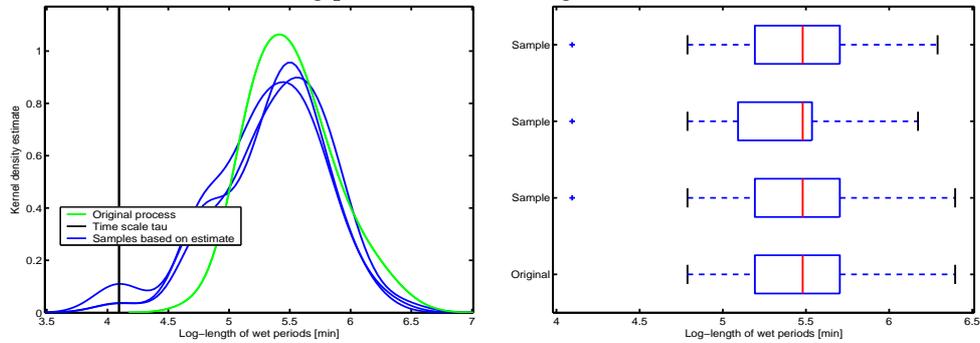
Alternative starting parameter
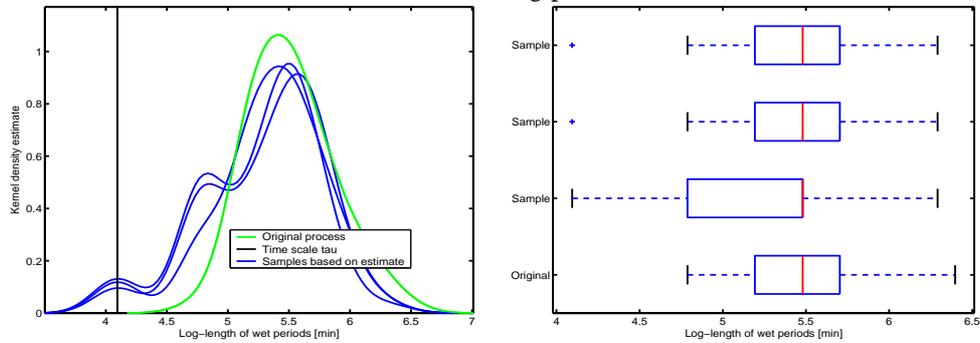
Figure A.24: Rainfall characteristics for trials F2b and F2b alt (Model V) — Part 2

Long cells, fixed ζ, restricted Λ, 1 h intervals
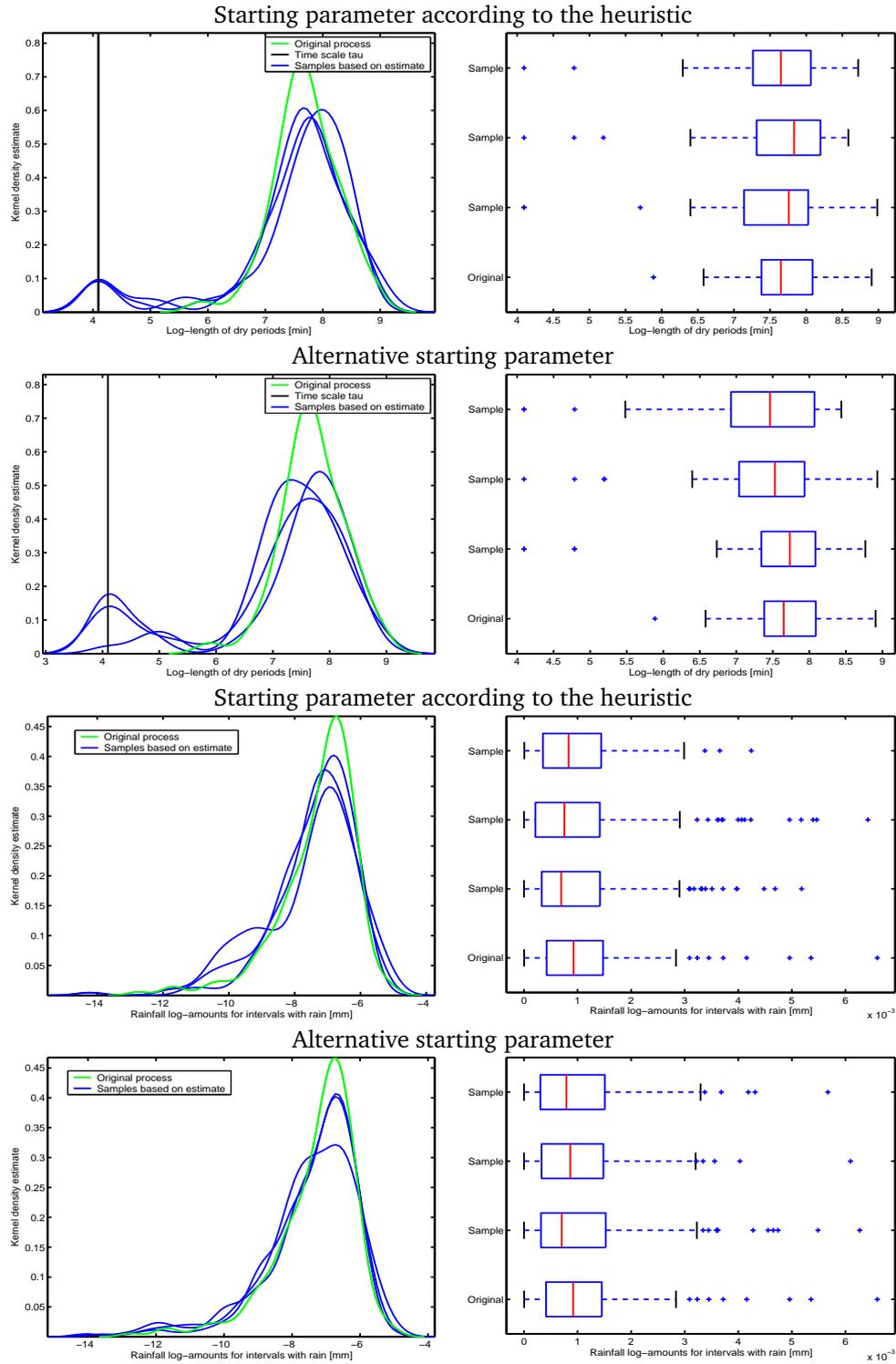
Figure A.25: Rainfall characteristics for real data and model output
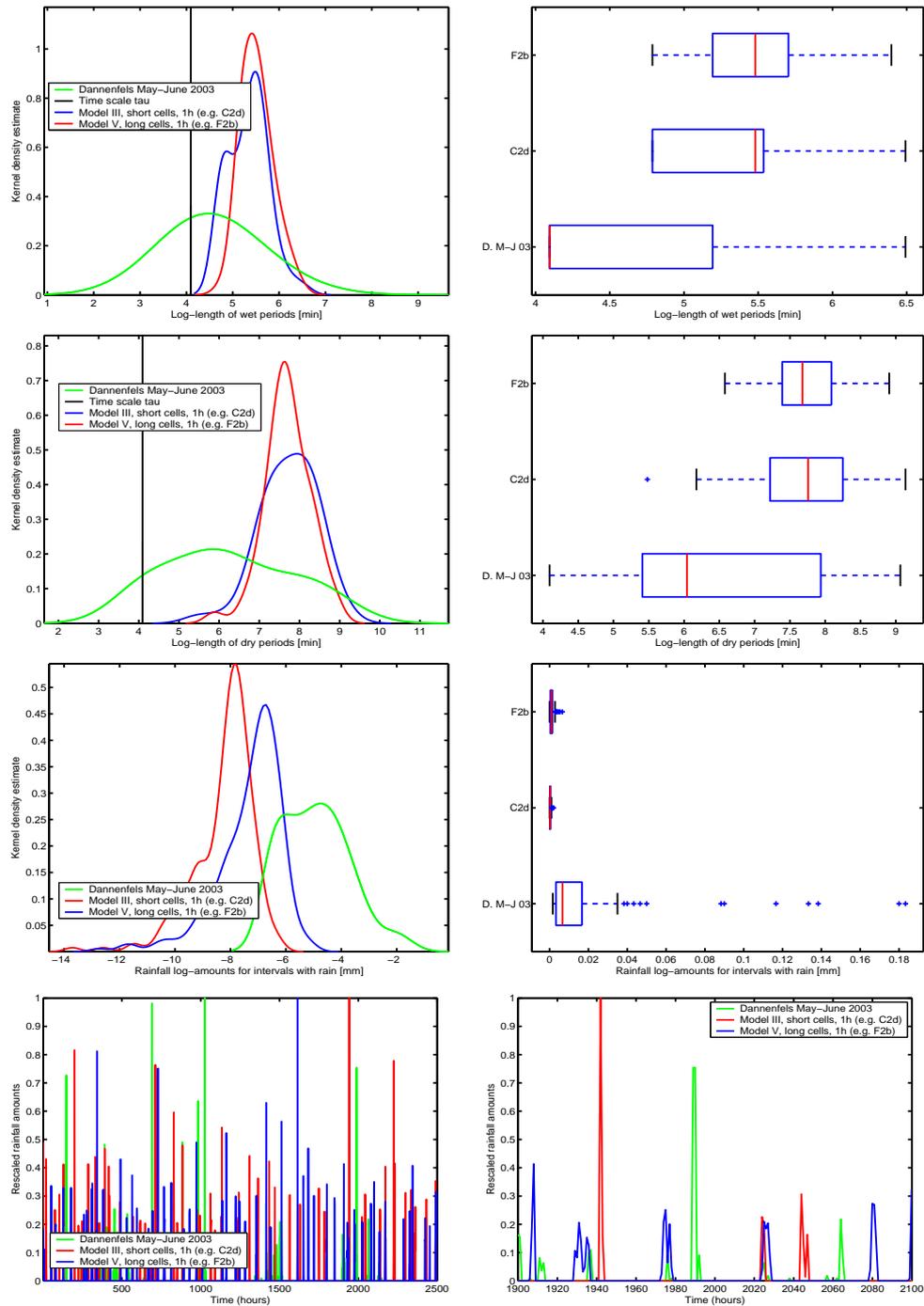
Appendix B

Numerical results — Tables

Table B.1: Artificial data — Instances and run lengths

Model	Inst.	Iterations	Model	Inst.	Iterations	Model	Inst.	Iterations
I	C1a	370,700	IV	A1c	500,000	V	E1b	248,700
	C1b	284,100		alt	500,000		alt	199,700
II	A1a	498,400		A1d	500,000		E2b	313,200
	A1b	496,900		alt	500,000		alt	313,100
	A2a	432,000		A2c	500,000		F1b	207,300
	A2b	388,100		A2d	500,000		alt	195,300
III	C1c	500,000		B1c	456,900		F2b	356,100
	alt	500,000		B1d	465,700		alt	375,600
	C1d	500,000		B2c	282,800	VI	E1e	269,100
	alt	500,000		B2d	500,000		E2e	307,000
	C2c	423,800					F1e	201,400
	C2d	500,000					F2e	321,100
	D1c	484,200						
	D1d	489,400						
	D2c	500,000						
	D2d	500,000						

The number of burn-in iterations depends on the total run length:

400K–500K iterations:	300K burn-in
300K–400K iterations:	250K burn-in
250K–300K iterations:	200K burn-in
$<$250K iterations:	150K burn-in

Table B.2: Data set C1a (Model I)

Short cells, variable ζ, full Λ, 10 min intervals								
Thinning with step-size 100, 250K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000	142	228	388	232	251	270
μ_1	5.04	5.03	4.99	5.11	5.25	5.11	5.12	5.13
μ_2	7.78	7.80	7.60	7.74	7.89	7.73	7.74	7.75
μ_3	2.40	2.14	1.33	1.72	2.01	1.58	1.69	1.80
μ_4	2.40	2.72	1.84	2.29	2.60	2.13	2.25	2.37
σ_1^2	0.33	0.33	0.37	0.47	0.61	0.47	0.48	0.49
σ_2^2	0.46	0.33	0.45	0.57	0.75	0.58	0.59	0.60
σ_3^2	1.00	0.33	0.28	0.40	0.59	0.38	0.42	0.46
σ_4^2	1.00	0.33	0.42	0.62	1.04	0.59	0.68	0.78
$\rho_{1,2}$	0.00	0.00	-0.05	0.14	0.33	0.13	0.14	0.15
$\rho_{1,3}$	0.00	-0.50	-0.17	0.11	0.30	0.02	0.08	0.15
$\rho_{1,4}$	0.00	0.00	0.17	0.38	0.50	0.31	0.35	0.40
$\rho_{2,3}$	0.00	0.00	-0.11	0.02	0.15	0.01	0.02	0.04
$\rho_{2,4}$	0.00	0.00	-0.16	-0.03	0.10	-0.05	-0.03	-0.01
$\rho_{3,4}$	0.00	0.00	0.31	0.46	0.59	0.42	0.45	0.48
$\overline{M}$	999	524	544	663	893	656	695	734
$\hat{D}$	0.81	1.00	0.96	0.99	1.02	0.99	0.99	0.99
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.3: Data set C1b (Model I)

Short cells, fixed ζ, full Λ, 10 min intervals								
Thinning with step-size 100, 200K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000						
μ_1	5.04	5.03	5.02	5.16	5.29	5.13	5.15	5.17
μ_2	7.78	7.80	7.53	7.75	7.91	7.62	7.71	7.79
μ_3	2.40	2.14	-0.19	0.65	1.17	0.20	0.56	0.91
μ_4	2.40	2.72	-0.04	1.05	1.67	0.51	0.92	1.32
σ_1^2	0.33	0.33	0.28	0.39	0.56	0.39	0.42	0.44
σ_2^2	0.46	0.33	0.43	0.59	0.94	0.50	0.77	1.04
σ_3^2	1.00	0.33	0.65	0.93	1.45	0.84	1.02	1.20
σ_4^2	1.00	0.33	1.12	1.72	2.94	1.48	1.92	2.35
$\rho_{1,2}$	0.00	0.00	-0.12	0.10	0.32	0.07	0.11	0.14
$\rho_{1,3}$	0.00	-0.50	0.15	0.24	0.34	0.23	0.24	0.25
$\rho_{1,4}$	0.00	0.00	0.30	0.40	0.49	0.38	0.39	0.41
$\rho_{2,3}$	0.00	0.00	-0.06	0.05	0.16	0.04	0.05	0.07
$\rho_{2,4}$	0.00	0.00	-0.07	0.05	0.16	0.03	0.05	0.06
$\rho_{3,4}$	0.00	0.00	0.27	0.42	0.57	0.37	0.42	0.48
$\overline{M}$	991	524	1031	1483	2736	1426	1725	2025
$\hat{D}$	0.81	1.00	0.98	1.00	1.01	1.00	1.00	1.00
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.4: Data set A1a (Model II)

	Value		Posterior quantiles			Confidence interval*		
Parameter	Start	True	10%	50%	90%	2.5%	Mean	97.5%
ζ	1,000	1,000	155	280	662	259	381	502
μ_1	5.08	5.03	5.03	5.15	5.28	5.15	5.16	5.16
μ_2	7.92	7.80	7.81	7.93	8.06	7.92	7.93	7.93
μ_3	2.43	2.14	1.82	1.95	2.07	1.93	1.95	1.96
μ_4	2.43	2.72	2.47	2.60	2.71	2.58	2.59	2.61
σ_1^2	0.31	0.33	0.30	0.39	0.52	0.40	0.40	0.41
σ_2^2	0.43	0.33	0.32	0.42	0.55	0.42	0.43	0.43
σ_3^2	1.00	0.33	0.28	0.36	0.45	0.35	0.36	0.37
σ_4^2	1.00	0.33	0.31	0.37	0.45	0.37	0.37	0.38
$\overline{M}$	978	507	528	570	630	572	577	583
$\hat{D}$	0.84	1.00	0.98	1.01	1.03	1.00	1.01	1.01

Short cells, variable ζ, diagonal Λ, 10 min intervals
Thinning with step-size 100, 300K burn-in iterations

* The bounds for the components of θ form a joint 95% confidence interval.

Table B.5: Data set A1b (Model II)

	Value		Posterior quantiles			Confidence interval*		
Parameter	Start	True	10%	50%	90%	2.5%	Mean	97.5%
ζ	1,000	1,000						
μ_1	5.08	5.03	5.03	5.16	5.28	5.16	5.16	5.17
μ_2	7.92	7.80	7.80	7.93	8.06	7.92	7.93	7.94
μ_3	2.43	2.14	1.71	1.82	1.93	1.79	1.82	1.84
μ_4	2.43	2.72	2.36	2.47	2.56	2.44	2.46	2.48
σ_1^2	0.31	0.33	0.30	0.41	0.55	0.41	0.42	0.43
σ_2^2	0.43	0.33	0.32	0.42	0.56	0.42	0.43	0.44
σ_3^2	1.00	0.33	0.31	0.40	0.52	0.39	0.41	0.43
σ_4^2	1.00	0.33	0.35	0.42	0.49	0.41	0.42	0.43
$\overline{M}$	978	507	592	634	682	630	636	642
$\hat{D}$	0.84	1.00	0.98	1.00	1.03	1.00	1.01	1.01

Short cells, fixed ζ, diagonal Λ, 10 min intervals
Thinning with step-size 100, 300K burn-in iterations

* The bounds for the components of θ form a joint 95% confidence interval.

Table B.6: Data set A2a (Model II)

Short cells, variable ζ, diagonal Λ, 1 h intervals								
Thinning with step-size 100, 300K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000	605	1635	5556	2223	2569	2914
μ_1	5.35	5.03	5.11	5.20	5.30	5.20	5.20	5.21
μ_2	7.81	7.80	7.57	7.70	7.83	7.69	7.70	7.71
μ_3	2.74	2.14	2.83	2.94	3.12	2.94	2.96	2.98
μ_4	2.74	2.72	3.68	3.71	3.83	3.71	3.74	3.76
σ_1^2	0.40	0.33	0.20	0.27	0.35	0.27	0.27	0.28
σ_2^2	0.36	0.33	0.39	0.50	0.66	0.51	0.51	0.52
σ_3^2	1.00	0.33	0.16	0.28	0.41	0.26	0.28	0.30
σ_4^2	1.00	0.33	0.01	0.02	0.05	0.02	0.03	0.04
$\overline{M}$	478	621	257	291	300	278	284	289
$\hat{D}$	0.91	1.00	0.96	1.00	1.05	1.00	1.00	1.01
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.7: Data set A2b (Model II)

Short cells, fixed ζ, diagonal Λ, 1 h intervals								
Thinning with step-size 100, 250K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000						
μ_1	5.35	5.03	5.16	5.27	5.37	5.26	5.27	5.27
μ_2	7.81	7.80	7.58	7.70	7.82	7.69	7.70	7.71
μ_3	2.74	2.14	2.91	3.00	3.12	2.99	3.01	3.02
μ_4	2.74	2.72	3.63	3.72	3.79	3.71	3.72	3.73
σ_1^2	0.40	0.33	0.21	0.28	0.37	0.28	0.29	0.29
σ_2^2	0.36	0.33	0.38	0.49	0.65	0.50	0.51	0.52
σ_3^2	1.00	0.33	0.08	0.17	0.26	0.16	0.17	0.19
σ_4^2	1.00	0.33	0.02	0.04	0.10	0.04	0.05	0.06
$\overline{M}$	429	621	266	287	306	284	287	289
$\hat{D}$	0.91	1.00	0.95	0.99	1.03	0.99	0.99	0.99
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.8: Data set C1c (Model III)

Short cells, variable ζ, mean matching, restricted Λ, 10 min intervals

Thinning with step-size 100, 300K burn-in iterations

	Value		Posterior quantiles			Confidence interval*		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000	184	416	2370	480	874	1267
μ_1	5.04	5.03	4.92	5.00	5.07	4.99	5.00	5.01
μ_2	7.78	7.80	7.55	7.70	7.85	7.69	7.70	7.71
μ_3	2.40	2.14	2.13	2.22	2.30	2.20	2.22	2.23
μ_4	2.40	2.72	2.67	2.74	2.80	2.73	2.74	2.74
σ_1^2	0.33	0.33	0.31	0.39	0.51	0.39	0.40	0.40
σ_2^2	0.46	0.33	0.49	0.63	0.83	0.64	0.65	0.66
σ_3^2	1.00	0.33	0.18	0.24	0.30	0.23	0.24	0.25
σ_4^2	1.00	0.33	0.27	0.32	0.37	0.31	0.32	0.32
$\rho_{1,2}$	0.00	0.00	-0.13	0.06	0.25	0.05	0.06	0.07
$\rho_{1,3}$	0.00	-0.50	-0.50	-0.39	-0.27	-0.40	-0.39	-0.37
$\rho_{2,3}$	0.00	0.00	-0.13	-0.01	0.12	-0.02	-0.01	0.00
$\overline{M}$	979	524	514	541	568	538	541	544
$\hat{D}$	0.83	1.00	0.94	0.96	0.99	0.96	0.96	0.96

* The bounds for the components of θ form a joint 95% confidence interval.

Table B.9: Data set C1c alt (Model III)

Short cells, variable ζ, mean matching, restricted Λ, 10 min intervals								
Thinning with step-size 100, 300K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	900	1,000	149	254	501	243	314	385
μ_1	4.53	5.03	4.97	5.04	5.11	5.03	5.04	5.04
μ_2	6.70	7.80	7.53	7.69	7.84	7.68	7.69	7.70
μ_3	2.16	2.14	2.15	2.24	2.32	2.23	2.24	2.25
μ_4	2.16	2.72	2.70	2.76	2.82	2.75	2.76	2.77
σ_1^2	0.37	0.33	0.28	0.35	0.45	0.35	0.36	0.36
σ_2^2	0.51	0.33	0.51	0.64	0.85	0.66	0.66	0.67
σ_3^2	1.11	0.33	0.17	0.23	0.30	0.22	0.23	0.24
σ_4^2	1.11	0.33	0.27	0.31	0.36	0.31	0.32	0.32
$\rho_{1,2}$	0.00	0.00	-0.09	0.10	0.28	0.09	0.10	0.11
$\rho_{1,3}$	0.00	-0.50	-0.52	-0.40	-0.27	-0.41	-0.40	-0.39
$\rho_{2,3}$	0.00	0.00	-0.15	-0.02	0.11	-0.03	-0.02	-0.01
$\overline{M}$	1,072	524	505	528	553	526	529	531
$\hat{D}$	0.86	1.00	0.94	0.96	0.99	0.96	0.96	0.97
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.10: Comparison of C1c and C1c alt (Model III)

Short cells, variable ζ, mean matching, restricted Λ, 10 min intervals									
Thinning with step-size 100, 300K burn-in iterations									
		C1c			**C1c alt**					
	True	**Posterior quantiles**			**Posterior quantiles**			**Scale reduction***		
Parameter	**Value**	**10%**	**50%**	**90%**	**10%**	**50%**	**90%**	**150K**	**300K**	**450K**
ζ	1,000	184	416	2370	149	254	501	1.22	1.10	1.08
μ_1	5.03	4.92	5.00	5.07	4.97	5.04	5.11	1.03	1.03	1.05
μ_2	7.80	7.55	7.70	7.85	7.53	7.69	7.84	1.00	1.01	1.00
μ_3	2.14	2.13	2.22	2.30	2.15	2.24	2.32	1.05	1.05	1.04
μ_4	2.72	2.67	2.74	2.80	2.70	2.76	2.82	1.02	1.02	1.02
σ_1^2	0.33	0.31	0.39	0.51	0.28	0.35	0.45	1.00	1.01	1.02
σ_2^2	0.33	0.49	0.63	0.83	0.51	0.64	0.85	1.00	1.00	1.00
σ_3^2	0.33	0.18	0.24	0.30	0.17	0.23	0.30	1.03	1.00	1.00
σ_4^2	0.33	0.27	0.32	0.37	0.27	0.31	0.36	1.00	1.00	1.00
$\rho_{1,2}$	0.00	-0.13	0.06	0.25	-0.09	0.10	0.28	1.01	1.03	1.03
$\rho_{1,3}$	-0.50	-0.50	-0.39	-0.27	-0.52	-0.40	-0.27	1.00	1.01	1.00
$\rho_{2,3}$	0.00	-0.13	-0.01	0.12	-0.15	-0.02	0.11	1.00	1.00	1.00
$\overline{M}$	524	514	541	568	505	528	553	1.17	1.12	1.11
$\hat{D}$	1.00	0.94	0.96	0.99	0.94	0.96	0.99	1.00	1.00	1.00
* Values less than 1.2 can be taken as an indicator of convergence.										

Table B.11: Data set C1d (Model III)

Short cells, fixed ζ, mean matching, restricted Λ, 10 min intervals								
Thinning with step-size 100, 300K burn-in iterations								
	Value		Posterior quantiles			Confidence interval*		
Parameter	Start	True	10%	50%	90%	2.5%	Mean	97.5%
ζ	1,000	1,000						
μ_1	5.04	5.03	4.84	4.93	5.00	4.92	4.92	4.93
μ_2	7.78	7.80	7.51	7.68	7.83	7.67	7.68	7.68
μ_3	2.40	2.14	1.96	2.07	2.16	2.05	2.06	2.08
μ_4	2.40	2.72	2.55	2.62	2.69	2.61	2.62	2.64
σ_1^2	0.33	0.33	0.30	0.38	0.49	0.39	0.39	0.40
σ_2^2	0.46	0.33	0.49	0.64	0.85	0.65	0.66	0.67
σ_3^2	1.00	0.33	0.28	0.35	0.42	0.34	0.35	0.36
σ_4^2	1.00	0.33	0.33	0.39	0.45	0.38	0.39	0.40
$\rho_{1,2}$	0.00	0.00	-0.08	0.09	0.27	0.09	0.09	0.10
$\rho_{1,3}$	0.00	-0.50	-0.40	-0.28	-0.17	-0.30	-0.28	-0.27
$\rho_{2,3}$	0.00	0.00	-0.09	0.02	0.14	0.01	0.02	0.04
$\overline{M}$	980	524	560	590	624	587	591	596
$\hat{D}$	0.79	1.00	0.95	0.98	1.01	0.98	0.98	0.98
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.12: Data set C1d alt (Model III)

Short cells, fixed ζ, mean matching, restricted Λ, 10 min intervals								
Thinning with step-size 100, 300K burn-in iterations								
	Value		Posterior quantiles			Confidence interval*		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	900	1,000						
μ_1	4.53	5.03	4.89	4.97	5.04	4.96	4.97	4.98
μ_2	6.70	7.80	7.52	7.68	7.84	7.67	7.68	7.69
μ_3	2.16	2.14	2.05	2.14	2.23	2.13	2.14	2.16
μ_4	2.16	2.72	2.62	2.68	2.74	2.67	2.68	2.69
σ_1^2	0.37	0.33	0.29	0.37	0.47	0.37	0.38	0.38
σ_2^2	0.51	0.33	0.51	0.64	0.84	0.65	0.66	0.67
σ_3^2	1.11	0.33	0.24	0.30	0.40	0.29	0.31	0.33
σ_4^2	1.11	0.33	0.30	0.35	0.40	0.35	0.35	0.36
$\rho_{1,2}$	0.00	0.00	-0.09	0.09	0.26	0.08	0.09	0.10
$\rho_{1,3}$	0.00	-0.50	-0.44	-0.33	-0.23	-0.34	-0.33	-0.32
$\rho_{2,3}$	0.00	0.00	-0.12	-0.01	0.11	-0.02	-0.01	0.00
$\overline{M}$	1,033	524	531	556	582	553	557	561
$\hat{D}$	0.87	1.00	0.94	0.97	1.00	0.97	0.97	0.97
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.13: Comparison of C1d and C1d alt (Model III)

Parameter	True Value	C1d Posterior quantiles 10%	50%	90%	C1d alt Posterior quantiles 10%	50%	90%	Scale reduction* 150K	300K	450K
Short cells, fixed ζ, mean matching, restricted Λ, 10 min intervals										
Thinning with step-size 100, 300K burn-in iterations										
μ_1	5.03	4.84	4.93	5.00	4.89	4.97	5.04	1.02	1.05	1.06
μ_2	7.80	7.51	7.68	7.83	7.52	7.68	7.84	1.00	1.00	1.00
μ_3	2.14	1.96	2.07	2.16	2.05	2.14	2.23	1.18	1.24	1.25
μ_4	2.72	2.55	2.62	2.69	2.62	2.68	2.74	1.10	1.17	1.19
σ_1^2	0.33	0.30	0.38	0.49	0.29	0.37	0.47	1.06	1.04	1.01
σ_2^2	0.33	0.49	0.64	0.85	0.51	0.64	0.84	1.00	1.00	1.00
σ_3^2	0.33	0.28	0.35	0.42	0.24	0.30	0.40	1.01	1.03	1.05
σ_4^2	0.33	0.33	0.39	0.45	0.30	0.35	0.40	1.04	1.07	1.10
$\rho_{1,2}$	0.00	-0.08	0.09	0.27	-0.09	0.09	0.26	1.00	1.01	1.00
$\rho_{1,3}$	-0.50	-0.40	-0.28	-0.17	-0.44	-0.33	-0.23	1.28	1.20	1.15
$\rho_{2,3}$	0.00	-0.09	0.02	0.14	-0.12	-0.01	0.11	1.07	1.03	1.00
$\overline{M}$	524	560	590	624	531	556	582	1.20	1.28	1.35
$\hat{D}$	1.00	0.95	0.98	1.01	0.94	0.97	1.00	1.00	1.05	1.05

* Values less than 1.2 can be taken as an indicator of convergence.

Table B.14: Data set C2c (Model III)

Short cells, variable ζ, mean matching, restricted Λ, 1 h intervals								
Thinning with step-size 100, 300K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000	346	1057	4601	1459	1893	2328
μ_1	5.19	5.03	5.06	5.09	5.12	5.09	5.09	5.09
μ_2	7.89	7.80	7.72	7.84	7.96	7.83	7.84	7.85
μ_3	2.53	2.14	1.55	1.69	1.82	1.67	1.68	1.70
μ_4	2.53	2.72	2.88	2.89	2.90	2.89	2.89	2.89
σ_1^2	0.28	0.33	0.16	0.21	0.28	0.21	0.22	0.22
σ_2^2	0.32	0.33	0.30	0.39	0.53	0.39	0.41	0.43
σ_3^2	1.00	0.33	0.94	1.18	1.46	1.16	1.19	1.21
σ_4^2	1.00	0.33	0.01	0.01	0.01	0.01	0.01	0.01
$\rho_{1,2}$	0.00	0.00	-0.14	0.05	0.24	0.04	0.05	0.07
$\rho_{1,3}$	0.00	-0.50	-0.24	-0.16	-0.07	-0.17	-0.16	-0.15
$\rho_{2,3}$	0.00	0.00	-0.13	-0.04	0.06	-0.05	-0.04	-0.03
$\overline{M}$	375	505	517	522	540	516	526	536
$\hat{D}$	0.92	1.00	0.97	1.00	1.03	1.00	1.00	1.00
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.15: Data set C2d (Model III)

Short cells, fixed ζ, mean matching, restricted Λ, 1 h intervals								
Thinning with step-size 100, 300K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000						
μ_1	5.19	5.03	5.08	5.12	5.15	5.12	5.12	5.12
μ_2	7.89	7.80	7.74	7.85	7.96	7.85	7.85	7.86
μ_3	2.53	2.14	1.74	1.85	1.96	1.84	1.85	1.85
μ_4	2.53	2.72	2.93	2.94	2.95	2.94	2.94	2.94
σ_1^2	0.28	0.33	0.20	0.25	0.33	0.26	0.26	0.26
σ_2^2	0.32	0.33	0.29	0.38	0.49	0.38	0.39	0.39
σ_3^2	1.00	0.33	0.65	0.82	1.03	0.82	0.83	0.84
σ_4^2	1.00	0.33	0.01	0.01	0.01	0.01	0.01	0.01
$\rho_{1,2}$	0.00	0.00	-0.08	0.11	0.29	0.10	0.11	0.12
$\rho_{1,3}$	0.00	-0.50	-0.33	-0.24	-0.15	-0.24	-0.24	-0.23
$\rho_{2,3}$	0.00	0.00	-0.19	-0.09	0.01	-0.09	-0.09	-0.08
$\overline{M}$	364	505	504	508	511	506	508	509
$\hat{D}$	0.93	1.00	0.97	1.00	1.03	1.00	1.00	1.00
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.16: Data set D1c (Model III)

Long cells, variable ζ, mean matching, restricted Λ, 10 min intervals								
Thinning with step-size 100, 300K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000	522	693	903	663	705	747
μ_1	5.24	5.03	4.97	5.04	5.11	5.04	5.04	5.05
μ_2	7.83	7.80	7.72	7.81	7.90	7.81	7.81	7.82
μ_3	2.57	3.23	3.20	3.26	3.31	3.25	3.26	3.26
μ_4	2.57	2.72	2.66	2.72	2.78	2.72	2.72	2.73
σ_1^2	0.27	0.33	0.29	0.38	0.49	0.38	0.39	0.39
σ_2^2	0.28	0.33	0.21	0.27	0.35	0.27	0.28	0.28
σ_3^2	1.00	0.33	0.26	0.30	0.35	0.30	0.30	0.31
σ_4^2	1.00	0.33	0.36	0.41	0.47	0.41	0.41	0.42
$\rho_{1,2}$	0.00	0.00	0.17	0.34	0.48	0.32	0.33	0.34
$\rho_{1,3}$	0.00	-0.50	-0.61	-0.53	-0.46	-0.54	-0.53	-0.53
$\rho_{2,3}$	0.00	0.00	-0.25	-0.14	-0.01	-0.14	-0.14	-0.13
$\overline{M}$	1,673	583	547	572	598	570	573	575
$\hat{D}$	0.90	1.00	0.98	1.00	1.02	1.00	1.00	1.00
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.17: Data set D1d (Model III)

Long cells, fixed ζ, mean matching, restricted Λ, 10 min intervals								
Thinning with step-size 100, 300K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000						
μ_1	5.24	5.03	4.94	5.01	5.08	5.00	5.01	5.02
μ_2	7.83	7.80	7.71	7.80	7.89	7.79	7.80	7.80
μ_3	2.57	3.23	3.18	3.23	3.28	3.22	3.23	3.23
μ_4	2.57	2.72	2.62	2.68	2.74	2.67	2.68	2.68
σ_1^2	0.27	0.33	0.28	0.35	0.45	0.35	0.36	0.36
σ_2^2	0.28	0.33	0.21	0.27	0.35	0.27	0.28	0.28
σ_3^2	1.00	0.33	0.24	0.29	0.34	0.28	0.29	0.30
σ_4^2	1.00	0.33	0.36	0.42	0.49	0.42	0.42	0.43
$\rho_{1,2}$	0.00	0.00	0.15	0.32	0.47	0.31	0.31	0.32
$\rho_{1,3}$	0.00	-0.50	-0.61	-0.53	-0.46	-0.54	-0.53	-0.53
$\rho_{2,3}$	0.00	0.00	-0.24	-0.13	-0.01	-0.13	-0.13	-0.12
$\overline{M}$	1,673	583	583	608	633	605	608	610
$\hat{D}$	0.90	1.00	0.97	1.00	1.02	0.99	1.00	1.00
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.18: Data set D2c (Model III)

Long cells, variable ζ, mean matching, restricted Λ, 1 h intervals								
Thinning with step-size 100, 300K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000	250	377	664	409	436	463
μ_1	5.37	5.03	5.08	5.18	5.28	5.16	5.18	5.20
μ_2	7.68	7.80	7.54	7.64	7.73	7.63	7.64	7.64
μ_3	2.69	3.23	2.93	3.10	3.37	3.09	3.13	3.17
μ_4	2.69	2.72	2.89	2.97	3.07	2.96	2.97	2.99
σ_1^2	0.24	0.33	0.19	0.24	0.33	0.24	0.25	0.26
σ_2^2	0.27	0.33	0.26	0.32	0.41	0.32	0.33	0.33
σ_3^2	1.00	0.33	0.30	0.62	0.84	0.56	0.60	0.64
σ_4^2	1.00	0.33	0.01	0.02	0.26	0.04	0.06	0.08
$\rho_{1,2}$	0.00	0.00	-0.05	0.12	0.28	0.11	0.12	0.13
$\rho_{1,3}$	0.00	-0.50	-0.32	-0.20	-0.12	-0.23	-0.22	-0.20
$\rho_{2,3}$	0.00	0.00	-0.13	-0.03	0.07	-0.03	-0.03	-0.02
$\overline{M}$	766	663	556	622	698	619	627	636
$\hat{D}$	1.04	1.00	0.97	1.00	1.03	1.00	1.00	1.00
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.19: Data set D2d (Model III)

Long cells, fixed ζ, mean matching, restricted Λ, 1 h intervals								
Thinning with step-size 100, 300K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000						
μ_1	5.37	5.03	5.19	5.28	5.36	5.27	5.28	5.28
μ_2	7.68	7.80	7.54	7.63	7.73	7.63	7.63	7.64
μ_3	2.69	3.23	3.34	3.44	3.54	3.43	3.44	3.45
μ_4	2.69	2.72	2.86	2.97	3.08	2.96	2.97	2.98
σ_1^2	0.24	0.33	0.26	0.35	0.48	0.36	0.36	0.37
σ_2^2	0.27	0.33	0.26	0.32	0.41	0.33	0.33	0.33
σ_3^2	1.00	0.33	0.10	0.19	0.30	0.18	0.20	0.21
σ_4^2	1.00	0.33	0.26	0.36	0.49	0.36	0.37	0.38
$\rho_{1,2}$	0.00	0.00	-0.22	-0.05	0.14	-0.05	-0.04	-0.03
$\rho_{1,3}$	0.00	-0.50	-0.56	-0.41	-0.26	-0.42	-0.41	-0.40
$\rho_{2,3}$	0.00	0.00	-0.07	0.09	0.26	0.08	0.09	0.10
$\overline{M}$	724	663	508	555	603	553	555	557
$\hat{D}$	1.02	1.00	0.96	0.99	1.02	0.99	0.99	0.99
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.20: Data set A1c (Model IV)

Short cells, variable ζ, mean matching, diagonal Λ, 10 min intervals								
Thinning with step-size 100, 300K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000	125	196	358	191	230	269
μ_1	5.08	5.03	4.97	5.04	5.11	5.04	5.04	5.05
μ_2	7.92	7.80	7.81	7.92	8.05	7.92	7.93	7.93
μ_3	2.43	2.14	1.99	2.08	2.16	2.07	2.08	2.09
μ_4	2.43	2.72	2.71	2.78	2.85	2.77	2.78	2.78
σ_1^2	0.31	0.33	0.26	0.34	0.44	0.35	0.35	0.35
σ_2^2	0.43	0.33	0.32	0.42	0.55	0.42	0.43	0.43
σ_3^2	1.00	0.33	0.26	0.34	0.45	0.34	0.35	0.36
σ_4^2	1.00	0.33	0.23	0.27	0.32	0.27	0.27	0.28
$\overline{M}$	993	507	479	504	533	504	505	507
$\hat{D}$	0.84	1.00	0.98	1.01	1.04	1.01	1.01	1.01
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.21: Data set A1c alt (Model IV)

Short cells, variable ζ, mean matching, diagonal Λ, 10 min intervals								
Thinning with step-size 100, 300K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	900	1,000	135	229	399	227	253	278
μ_1	4.53	5.03	4.96	5.04	5.11	5.03	5.04	5.05
μ_2	6.70	7.80	7.80	7.93	8.05	7.92	7.93	7.93
μ_3	2.16	2.14	2.03	2.12	2.21	2.11	2.12	2.13
μ_4	2.16	2.72	2.69	2.76	2.84	2.75	2.76	2.77
σ_1^2	0.35	0.33	0.28	0.36	0.46	0.36	0.37	0.37
σ_2^2	0.47	0.33	0.32	0.42	0.56	0.42	0.43	0.44
σ_3^2	1.11	0.33	0.25	0.32	0.41	0.32	0.33	0.34
σ_4^2	1.11	0.33	0.27	0.31	0.37	0.31	0.31	0.32
$\overline{M}$	1,079	507	473	500	530	499	501	504
$\hat{D}$	0.85	1.00	0.97	1.00	1.02	1.00	1.00	1.00
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.22: Comparison of A1c and A1c alt (Model IV)

Short cells, variable ζ, mean matching, diagonal Λ, 10 min intervals									
Thinning with step-size 100, 300K burn-in iterations									
		A1c			**A1c alt**					
	True	**Posterior quantiles**			**Posterior quantiles**			**Scale reduction***		
Parameter	**Value**	**10%**	**50%**	**90%**	**10%**	**50%**	**90%**	**150K**	**300K**	**450K**
ζ	1,000	125	196	358	135	229	399	1.00	1.02	1.01
μ_1	5.03	4.97	5.04	5.11	4.96	5.04	5.11	1.03	1.01	1.01
μ_2	7.80	7.81	7.92	8.05	7.80	7.93	8.05	1.00	1.00	1.00
μ_3	2.14	1.99	2.08	2.16	2.03	2.12	2.21	1.04	1.04	1.05
μ_4	2.72	2.71	2.78	2.85	2.69	2.76	2.84	1.05	1.03	1.02
σ_1^2	0.33	0.26	0.34	0.44	0.28	0.36	0.46	1.00	1.01	1.01
σ_2^2	0.33	0.32	0.42	0.55	0.32	0.42	0.56	1.00	1.00	1.00
σ_3^2	0.33	0.26	0.34	0.45	0.25	0.32	0.41	1.01	1.01	1.00
σ_4^2	0.33	0.23	0.27	0.32	0.27	0.31	0.37	1.07	1.08	1.11
$\overline{M}$	507	479	504	533	473	500	530	1.00	1.00	1.00
$\hat{D}$	1.00	0.98	1.01	1.04	0.97	1.00	1.02	1.10	1.07	1.07
* Values less than 1.2 can be taken as an indicator of convergence.									

Table B.23: Data set A1d (Model IV)

Short cells, fixed ζ, mean matching, diagonal Λ, 10 min intervals								
Thinning with step-size 100, 300K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000						
μ_1	5.08	5.03	4.91	4.99	5.06	4.98	4.99	4.99
μ_2	7.92	7.80	7.80	7.93	8.05	7.92	7.93	7.93
μ_3	2.43	2.14	1.95	2.02	2.10	2.01	2.02	2.04
μ_4	2.43	2.72	2.65	2.71	2.77	2.71	2.71	2.72
σ_1^2	0.31	0.33	0.28	0.36	0.48	0.37	0.38	0.38
σ_2^2	0.43	0.33	0.32	0.42	0.56	0.42	0.43	0.44
σ_3^2	1.00	0.33	0.32	0.38	0.46	0.37	0.38	0.40
σ_4^2	1.00	0.33	0.27	0.31	0.36	0.31	0.31	0.32
$\overline{M}$	993	507	506	529	553	526	529	532
$\hat{D}$	0.84	1.00	0.98	1.01	1.03	1.01	1.01	1.01
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.24: Data set A1d alt (Model IV)

	Value		Posterior quantiles			Confidence interval*		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	900	1,000						
μ_1	4.53	5.03	4.94	5.01	5.09	5.01	5.01	5.02
μ_2	6.70	7.80	7.80	7.93	8.05	7.92	7.93	7.93
μ_3	2.16	2.14	1.99	2.06	2.13	2.05	2.06	2.07
μ_4	2.16	2.72	2.66	2.72	2.78	2.71	2.72	2.73
σ_1^2	0.35	0.33	0.29	0.37	0.48	0.37	0.37	0.38
σ_2^2	0.47	0.33	0.32	0.42	0.55	0.42	0.43	0.44
σ_3^2	1.11	0.33	0.26	0.32	0.39	0.32	0.33	0.34
σ_4^2	1.11	0.33	0.30	0.35	0.40	0.35	0.35	0.36
$\overline{M}$	980	507	503	524	547	522	525	527
$\hat{D}$	0.79	1.00	0.97	1.00	1.02	1.00	1.00	1.00

Short cells, fixed ζ, mean matching, diagonal Λ, 10 min intervals
Thinning with step-size 100, 300K burn-in iterations

* The bounds for the components of θ form a joint 95% confidence interval.

Table B.25: Comparison of A1d and A1d alt (Model IV)

Short cells, fixed ζ, mean matching, diagonal Λ, 10 min intervals									
Thinning with step-size 100, 300K burn-in iterations									
		A1d			**A1d alt**					
	True	**Posterior quantiles**			**Posterior quantiles**			**Scale reduction***		
Parameter	**Value**	**10%**	**50%**	**90%**	**10%**	**50%**	**90%**	**150K**	**300K**	**450K**
μ_1	5.03	4.91	4.99	5.06	4.94	5.01	5.09	1.05	1.03	1.00
μ_2	7.80	7.80	7.93	8.05	7.80	7.93	8.05	1.00	1.00	1.00
μ_3	2.14	1.95	2.02	2.10	1.99	2.06	2.13	1.03	1.00	1.01
μ_4	2.72	2.65	2.71	2.77	2.66	2.72	2.78	1.20	1.16	1.07
σ_1^2	0.33	0.28	0.36	0.48	0.29	0.37	0.48	1.00	1.00	1.00
σ_2^2	0.33	0.32	0.42	0.56	0.32	0.42	0.55	1.00	1.00	1.00
σ_3^2	0.33	0.32	0.38	0.46	0.26	0.32	0.39	1.23	1.14	1.14
σ_4^2	0.33	0.27	0.31	0.36	0.30	0.35	0.40	1.21	1.29	1.24
$\overline{M}$	507	506	529	553	503	524	547	1.12	1.09	1.03
$\hat{D}$	1.00	0.98	1.01	1.03	0.97	1.00	1.02	1.16	1.07	1.07
* Values less than 1.2 can be taken as an indicator of convergence.									

Table B.26: Data set A2c (Model IV)

Short cells, variable ζ, mean matching, diagonal Λ, 1 h intervals								
Thinning with step-size 100, 300K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000	449	1198	4242	1723	1975	2226
μ_1	5.35	5.03	5.31	5.41	5.51	5.40	5.41	5.42
μ_2	7.81	7.80	7.57	7.70	7.83	7.69	7.70	7.71
μ_3	2.74	2.14	2.36	2.52	2.67	2.50	2.52	2.53
μ_4	2.74	2.72	3.00	3.14	3.28	3.13	3.14	3.16
σ_1^2	0.40	0.33	0.28	0.37	0.49	0.38	0.38	0.39
σ_2^2	0.36	0.33	0.39	0.49	0.64	0.50	0.51	0.51
σ_3^2	1.00	0.33	0.06	0.24	0.43	0.22	0.24	0.27
σ_4^2	1.00	0.33	0.20	0.30	0.44	0.30	0.31	0.32
$\overline{M}$	466	621	405	449	495	447	449	452
$\hat{D}$	0.90	1.00	0.96	1.00	1.03	1.00	1.00	1.00
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.27: Data set A2d (Model IV)

	Value		Posterior quantiles			Confidence interval*		
Parameter	Start	True	10%	50%	90%	2.5%	Mean	97.5%
ζ	1,000	1,000						
μ_1	5.35	5.03	5.28	5.36	5.44	5.36	5.36	5.37
μ_2	7.81	7.80	7.58	7.71	7.83	7.70	7.71	7.71
μ_3	2.74	2.14	2.32	2.45	2.58	2.44	2.45	2.46
μ_4	2.74	2.72	2.96	3.08	3.20	3.07	3.08	3.09
σ_1^2	0.40	0.33	0.25	0.32	0.42	0.32	0.33	0.33
σ_2^2	0.36	0.33	0.38	0.49	0.64	0.50	0.50	0.51
σ_3^2	1.00	0.33	0.17	0.29	0.44	0.29	0.30	0.31
σ_4^2	1.00	0.33	0.18	0.28	0.40	0.28	0.29	0.30
$\overline{M}$	466	621	430	472	513	470	472	474
$\hat{D}$	0.90	1.00	0.96	0.99	1.03	0.99	0.99	0.99

Short cells, fixed ζ, mean matching, diagonal Λ, 1 h intervals
Thinning with step-size 100, 300K burn-in iterations

* The bounds for the components of θ form a joint 95% confidence interval.

Table B.28: Data set B1c (Model IV)

Long cells, variable ζ, mean matching, diagonal Λ, 10 min intervals								
Thinning with step-size 100, 300K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000	581	707	871	693	717	741
μ_1	5.20	5.03	4.96	5.04	5.12	5.04	5.04	5.05
μ_2	7.89	7.80	7.76	7.89	8.02	7.88	7.89	7.90
μ_3	2.53	3.23	3.22	3.28	3.35	3.28	3.28	3.29
μ_4	2.53	2.72	2.71	2.77	2.83	2.76	2.77	2.77
σ_1^2	0.27	0.33	0.30	0.39	0.51	0.39	0.40	0.40
σ_2^2	0.42	0.33	0.32	0.41	0.55	0.42	0.43	0.43
σ_3^2	1.00	0.33	0.27	0.32	0.38	0.32	0.33	0.34
σ_4^2	1.00	0.33	0.29	0.34	0.40	0.34	0.34	0.35
$\overline{M}$	1,503	455	444	465	488	464	465	467
$\hat{D}$	0.88	1.00	0.91	0.94	0.96	0.94	0.94	0.94
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.29: Data set B1d (Model IV)

Long cells, fixed ζ, mean matching, diagonal Λ, 10 min intervals								
Thinning with step-size 100, 300K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000						
μ_1	5.20	5.03	4.94	5.02	5.09	5.01	5.02	5.02
μ_2	7.89	7.80	7.77	7.89	8.01	7.88	7.89	7.90
μ_3	2.53	3.23	3.16	3.22	3.28	3.21	3.22	3.23
μ_4	2.53	2.72	2.69	2.75	2.80	2.74	2.75	2.75
σ_1^2	0.27	0.33	0.30	0.38	0.51	0.39	0.39	0.40
σ_2^2	0.42	0.33	0.32	0.41	0.56	0.42	0.43	0.44
σ_3^2	1.00	0.33	0.32	0.39	0.47	0.38	0.39	0.41
σ_4^2	1.00	0.33	0.28	0.33	0.39	0.32	0.33	0.34
$\overline{M}$	1,503	455	459	478	497	477	478	480
$\hat{D}$	0.88	1.00	0.91	0.94	0.96	0.94	0.94	0.94
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.30: Data set B2c (Model IV)

Long cells, variable ζ, mean matching, diagonal Λ, 1h intervals								
Thinning with step-size 100, 200K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000	186	265	379	268	276	284
μ_1	5.35	5.03	4.97	5.06	5.14	5.05	5.06	5.07
μ_2	7.80	7.80	7.73	7.80	7.88	7.80	7.80	7.81
μ_3	2.65	3.23	2.44	2.70	2.94	2.67	2.70	2.72
μ_4	2.65	2.72	2.44	2.57	2.71	2.56	2.57	2.59
σ_1^2	0.20	0.33	0.09	0.14	0.20	0.13	0.14	0.15
σ_2^2	0.20	0.33	0.18	0.23	0.30	0.23	0.24	0.24
σ_3^2	1.00	0.33	1.06	1.48	1.99	1.45	1.50	1.56
σ_4^2	1.00	0.33	0.34	0.49	0.70	0.50	0.51	0.53
$\overline{M}$	678	564	521	559	597	556	559	561
$\hat{D}$	1.02	1.00	0.98	1.01	1.04	1.00	1.01	1.01
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.31: Data set B2d (Model IV)

Long cells, fixed ζ, mean matching, diagonal Λ, 1h intervals								
Thinning with step-size 100, 300K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000						
μ_1	5.35	5.03	5.17	5.26	5.35	5.25	5.26	5.27
μ_2	7.80	7.80	7.72	7.81	7.89	7.80	7.81	7.81
μ_3	2.65	3.23	3.35	3.56	3.74	3.53	3.55	3.57
μ_4	2.65	2.72	2.82	2.98	3.13	2.97	2.98	2.99
σ_1^2	0.20	0.33	0.32	0.45	0.63	0.46	0.47	0.47
σ_2^2	0.20	0.33	0.17	0.22	0.29	0.23	0.23	0.23
σ_3^2	1.00	0.33	0.06	0.19	0.42	0.20	0.22	0.25
σ_4^2	1.00	0.33	0.29	0.41	0.57	0.41	0.42	0.43
$\overline{M}$	650	564	409	463	520	462	465	467
$\hat{D}$	1.06	1.00	0.94	0.98	1.01	0.98	0.98	0.98
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.32: Data set E1b (Model V)

Short cells, fixed ζ, restricted Λ, 10 min intervals								
Thinning with step-size 100, 150K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000						
μ_1	5.08	5.03	5.11	5.25	5.41	5.24	5.26	5.27
μ_2	7.75	7.80	7.67	7.77	7.87	7.76	7.77	7.78
μ_3	2.43	2.14	1.09	1.38	1.57	1.26	1.35	1.44
μ_4	2.43	2.72	1.46	1.82	2.06	1.70	1.78	1.87
σ_1^2	0.31	0.33	0.35	0.48	0.68	0.48	0.50	0.53
σ_2^2	0.33	0.33	0.23	0.31	0.40	0.31	0.31	0.32
σ_3^2	1.00	0.33	0.43	0.54	0.69	0.51	0.55	0.59
σ_4^2	1.00	0.33	1.04	1.33	1.76	1.24	1.37	1.50
$\rho_{1,2}$	0.00	0.00	-0.15	0.06	0.29	0.04	0.06	0.09
$\rho_{1,4}$	0.75	0.50	0.19	0.27	0.35	0.26	0.27	0.28
$\rho_{2,4}$	0.00	0.00	-0.09	-0.01	0.08	-0.01	-0.01	0.00
$\overline{M}$	1,155	553	919	1064	1317	1044	1100	1155
$\hat{D}$	0.79	1.00	0.98	1.00	1.02	1.00	1.00	1.00
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.33: Data set E1b alt (Model V)

Short cells, fixed ζ, restricted Λ, 10 min intervals								
Thinning with step-size 100, 150K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	900	1,000						
μ_1	4.57	5.03	5.30	5.40	5.50	5.38	5.40	5.41
μ_2	6.97	7.80	7.67	7.77	7.87	7.76	7.77	7.79
μ_3	2.19	2.14	0.53	0.77	0.99	0.69	0.76	0.83
μ_4	2.19	2.72	0.87	1.10	1.33	1.04	1.11	1.17
σ_1^2	0.34	0.33	0.19	0.25	0.33	0.24	0.25	0.26
σ_2^2	0.37	0.33	0.23	0.30	0.39	0.29	0.30	0.31
σ_3^2	1.11	0.33	0.68	0.82	1.00	0.80	0.83	0.87
σ_4^2	1.11	0.33	1.77	1.98	2.00	1.91	1.93	1.95
$\rho_{1,2}$	0.00	0.00	0.03	0.23	0.40	0.19	0.22	0.24
$\rho_{1,4}$	0.75	0.50	0.09	0.16	0.23	0.15	0.16	0.17
$\rho_{2,4}$	0.00	0.00	-0.06	-0.01	0.06	-0.01	-0.00	0.00
$\overline{M}$	1,157	553	1444	1680	1970	1641	1696	1752
$\hat{D}$	0.85	1.00	0.98	1.00	1.02	1.00	1.00	1.00
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.34: Comparison of E1b and E1b alt (Model V)

Short cells, fixed ζ, restricted Λ, 10 min intervals								
Thinning with step-size 100, 150K burn-in iterations								
		E1b			**E1b alt**			
	True	**Posterior quantiles**			**Posterior quantiles**			**PSR***
Parameter	**Value**	**10%**	**50%**	**90%**	**10%**	**50%**	**90%**	**150K**
μ_1	5.03	5.11	5.25	5.41	5.30	5.40	5.50	1.89
μ_2	7.80	7.67	7.77	7.87	7.67	7.77	7.87	1.00
μ_3	2.14	1.09	1.38	1.57	0.53	0.77	0.99	2.76
μ_4	2.72	1.46	1.82	2.06	0.87	1.10	1.33	3.10
σ_1^2	0.33	0.35	0.48	0.68	0.19	0.25	0.33	2.00
σ_2^2	0.33	0.23	0.31	0.40	0.23	0.30	0.39	1.01
σ_3^2	0.33	0.43	0.54	0.69	0.68	0.82	1.00	2.06
σ_4^2	0.33	1.04	1.33	1.76	1.77	1.98	2.00	3.35
$\rho_{1,2}$	0.00	-0.15	0.06	0.29	0.03	0.23	0.40	1.37
$\rho_{1,4}$	0.50	0.19	0.27	0.35	0.09	0.16	0.23	1.80
$\rho_{2,4}$	0.00	-0.09	-0.01	0.08	-0.06	-0.01	0.06	1.00
$\overline{M}$	553	919	1064	1317	1444	1680	1970	2.81
$\hat{D}$	1.00	0.98	1.00	1.02	0.98	1.00	1.02	1.01
* Values less than 1.2 can be taken as an indicator of convergence.								

Table B.35: Data set E2b (Model V)

Short cells, fixed ζ, restricted Λ, 1 h intervals								
Thinning with step-size 100, 250K burn-in iterations								
	Value		Posterior quantiles			Confidence interval*		
Parameter	Start	True	10%	50%	90%	2.5%	Mean	97.5%
ζ	1,000	1,000						
μ_1	5.27	5.03	5.33	5.47	5.61	5.46	5.47	5.49
μ_2	7.88	7.80	7.81	7.92	8.03	7.91	7.92	7.93
μ_3	2.63	2.14	1.71	1.97	2.19	1.92	1.96	2.00
μ_4	2.63	2.72	1.88	2.25	2.58	2.18	2.24	2.30
σ_1^2	0.32	0.33	0.30	0.41	0.57	0.41	0.42	0.44
σ_2^2	0.34	0.33	0.24	0.30	0.41	0.31	0.32	0.32
σ_3^2	1.00	0.33	0.11	0.13	0.22	0.15	0.15	0.16
σ_4^2	1.00	0.33	1.53	1.86	1.99	1.78	1.80	1.83
$\rho_{1,2}$	0.00	0.00	0.15	0.34	0.51	0.32	0.34	0.36
$\rho_{1,4}$	0.75	0.50	0.12	0.22	0.31	0.21	0.22	0.23
$\rho_{2,4}$	0.00	0.00	-0.03	0.05	0.14	0.05	0.06	0.06
$\overline{M}$	430	487	489	610	764	602	620	638
$\hat{D}$	0.88	1.00	0.97	1.00	1.04	1.00	1.01	1.01
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.36: Data set E2b alt (Model V)

Short cells, fixed ζ, restricted Λ, 1 h intervals								
Thinning with step-size 100, 250K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	900	1,000						
μ_1	4.75	5.03	5.32	5.45	5.59	5.44	5.45	5.46
μ_2	7.09	7.80	7.81	7.91	8.02	7.91	7.91	7.92
μ_3	2.37	2.14	1.75	2.03	2.25	1.97	2.01	2.06
μ_4	2.37	2.72	1.92	2.31	2.68	2.25	2.31	2.38
σ_1^2	0.35	0.33	0.28	0.39	0.54	0.39	0.40	0.41
σ_2^2	0.37	0.33	0.24	0.31	0.42	0.31	0.32	0.33
σ_3^2	1.11	0.33	0.11	0.13	0.22	0.14	0.15	0.16
σ_4^2	1.11	0.33	1.44	1.83	1.99	1.72	1.76	1.80
$\rho_{1,2}$	0.00	0.00	0.03	0.25	0.44	0.22	0.24	0.27
$\rho_{1,4}$	0.75	0.50	0.14	0.21	0.31	0.21	0.22	0.23
$\rho_{2,4}$	0.00	0.00	-0.05	0.03	0.12	0.03	0.04	0.04
$\overline{M}$	467	487	457	577	744	570	590	610
$\hat{D}$	0.93	1.00	0.97	1.01	1.04	1.00	1.01	1.01
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.37: Comparison of E2b and E2b alt (Model V)

Parameter	True Value	E2b Posterior quantiles			E2b alt Posterior quantiles			PSR*	
		10%	50%	90%	10%	50%	90%	150K	300K
μ_1	5.03	5.33	5.47	5.61	5.32	5.45	5.59	1.02	1.00
μ_2	7.80	7.81	7.92	8.03	7.81	7.91	8.02	1.00	1.00
μ_3	2.14	1.71	1.97	2.19	1.75	2.03	2.25	1.01	1.00
μ_4	2.72	1.88	2.25	2.58	1.92	2.31	2.68	1.02	1.01
σ_1^2	0.33	0.30	0.41	0.57	0.28	0.39	0.54	1.03	1.00
σ_2^2	0.33	0.24	0.30	0.41	0.24	0.31	0.42	1.00	1.00
σ_3^2	0.33	0.11	0.13	0.22	0.11	0.13	0.22	1.00	1.00
σ_4^2	0.33	1.53	1.86	1.99	1.44	1.83	1.99	1.03	1.01
$\rho_{1,2}$	0.00	0.15	0.34	0.51	0.03	0.25	0.44	1.03	1.00
$\rho_{1,4}$	0.50	0.12	0.22	0.31	0.14	0.21	0.31	1.00	1.00
$\rho_{2,4}$	0.00	-0.03	0.05	0.14	-0.05	0.03	0.12	1.00	1.00
$\overline{M}$	487	489	610	764	457	577	744	1.02	1.00
$\hat{D}$	1.00	0.97	1.00	1.04	0.97	1.01	1.04	1.00	1.00

Short cells, fixed ζ, restricted Λ, 1 h intervals

Thinning with step-size 100, 250K burn-in iterations

* The bounds for the components of θ form a joint 95% confidence interval.

Table B.38: Data set F1b (Model V)

Long cells, fixed ζ, restricted Λ, 10 min intervals								
Thinning with step-size 100, 150K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000						
μ_1	5.18	5.03	5.02	5.13	5.25	5.12	5.13	5.14
μ_2	7.72	7.80	7.58	7.69	7.80	7.68	7.69	7.70
μ_3	2.50	3.23	3.06	3.13	3.19	3.12	3.13	3.14
μ_4	2.50	2.72	2.48	2.57	2.66	2.56	2.57	2.58
σ_1^2	0.25	0.33	0.33	0.42	0.55	0.43	0.44	0.45
σ_2^2	0.35	0.33	0.33	0.41	0.52	0.41	0.42	0.43
σ_3^2	1.00	0.33	0.33	0.38	0.45	0.38	0.39	0.40
σ_4^2	1.00	0.33	0.61	0.75	0.93	0.73	0.76	0.80
$\rho_{1,2}$	0.00	0.00	-0.02	0.15	0.31	0.13	0.15	0.16
$\rho_{1,4}$	0.75	0.50	0.18	0.26	0.34	0.25	0.26	0.27
$\rho_{2,4}$	0.00	0.00	0.03	0.11	0.18	0.10	0.10	0.11
$\overline{M}$	1,790	553	583	609	640	607	610	614
$\hat{D}$	0.85	1.00	0.91	0.93	0.95	0.93	0.93	0.93
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.39: Data set F1b alt (Model V)

Long cells, fixed ζ, restricted Λ, 10 min intervals								
Thinning with step-size 100, 150K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	900	1,000						
μ_1	4.66	5.03	5.03	5.14	5.26	5.13	5.14	5.15
μ_2	6.94	7.80	7.56	7.67	7.80	7.66	7.67	7.68
μ_3	2.25	3.23	3.06	3.14	3.20	3.12	3.13	3.15
μ_4	2.25	2.72	2.48	2.57	2.65	2.56	2.57	2.58
σ_1^2	0.28	0.33	0.30	0.38	0.50	0.38	0.39	0.41
σ_2^2	0.39	0.33	0.34	0.42	0.55	0.42	0.43	0.45
σ_3^2	1.11	0.33	0.30	0.37	0.44	0.35	0.37	0.39
σ_4^2	1.11	0.33	0.58	0.71	0.88	0.68	0.72	0.75
$\rho_{1,2}$	0.00	0.00	-0.06	0.13	0.29	0.11	0.13	0.15
$\rho_{1,4}$	0.75	0.50	0.14	0.22	0.29	0.20	0.22	0.23
$\rho_{2,4}$	0.75	0.00	0.01	0.09	0.17	0.08	0.09	0.10
$\overline{M}$	1,879	553	587	614	644	611	616	620
$\hat{D}$	0.92	1.00	0.90	0.93	0.95	0.93	0.93	0.93
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.40: Comparison of F1b and F1b alt (Model V)

Long cells, fixed ζ, restricted Λ, 10 min intervals								
Thinning with step-size 100, 150K burn-in iterations								
		F1b			**F1b alt**			
	True	**Posterior quantiles**			**Posterior quantiles**			**PSR***
Parameter	**Value**	**10%**	**50%**	**90%**	**10%**	**50%**	**90%**	**150K**
μ_1	5.03	5.02	5.13	5.25	5.03	5.14	5.26	1.00
μ_2	7.80	7.58	7.69	7.80	7.56	7.67	7.80	1.00
μ_3	3.23	3.06	3.13	3.19	3.06	3.14	3.20	1.03
μ_4	2.72	2.48	2.57	2.66	2.48	2.57	2.65	1.02
σ_1^2	0.33	0.33	0.42	0.55	0.30	0.38	0.50	1.05
σ_2^2	0.33	0.33	0.41	0.52	0.34	0.42	0.55	1.00
σ_3^2	0.33	0.33	0.38	0.45	0.30	0.37	0.44	1.06
σ_4^2	0.33	0.61	0.75	0.93	0.58	0.71	0.88	1.02
$\rho_{1,2}$	0.00	-0.02	0.15	0.31	-0.06	0.13	0.29	1.01
$\rho_{1,4}$	0.50	0.18	0.26	0.34	0.14	0.22	0.29	1.06
$\rho_{2,4}$	0.00	0.03	0.11	0.18	0.01	0.09	0.17	1.04
$\overline{M}$	553	583	609	640	587	614	644	1.02
$\hat{D}$	1.00	0.91	0.93	0.95	0.90	0.93	0.95	1.00
* Values less than 1.2 can be taken as an indicator of convergence.								

Table B.41: Data set F2b (Model V)

	Value		Posterior quantiles			Confidence interval*		
Long cells, fixed ζ, restricted Λ, 1 h intervals								
Thinning with step-size 100, 250K burn-in iterations								
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000						
μ_1	5.38	5.03	5.11	5.24	5.37	5.23	5.24	5.25
μ_2	7.83	7.80	7.74	7.83	7.91	7.83	7.83	7.84
μ_3	2.65	3.23	3.29	3.71	4.10	3.63	3.70	3.77
μ_4	2.65	2.72	2.70	3.14	3.53	3.05	3.13	3.21
σ_1^2	0.14	0.33	0.20	0.28	0.38	0.28	0.29	0.30
σ_2^2	0.21	0.33	0.19	0.23	0.30	0.24	0.24	0.24
σ_3^2	1.00	0.33	0.13	0.26	0.49	0.26	0.29	0.31
σ_4^2	1.00	0.33	0.92	1.33	1.82	1.31	1.34	1.38
$\rho_{1,2}$	0.00	0.00	-0.26	-0.04	0.17	-0.06	-0.04	-0.02
$\rho_{1,4}$	0.75	0.50	0.10	0.21	0.31	0.20	0.21	0.22
$\rho_{2,4}$	0.00	0.00	-0.11	-0.00	0.11	-0.01	-0.00	0.01
$\overline{M}$	676	513	245	332	469	332	347	362
$\hat{D}$	1.02	1.00	0.93	0.97	1.02	0.97	0.97	0.98

* The bounds for the components of θ form a joint 95% confidence interval.

Table B.42: Data set F2b alt (Model V)

Long cells, fixed ζ, restricted Λ, 1h intervals								
Thinning with step-size 100, 250K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	900	1,000						
μ_1	4.84	5.03	5.13	5.27	5.40	5.25	5.26	5.28
μ_2	7.05	7.80	7.74	7.83	7.92	7.83	7.83	7.84
μ_3	2.38	3.23	3.03	3.70	4.08	3.46	3.62	3.77
μ_4	2.38	2.72	2.36	3.08	3.50	2.84	2.99	3.14
σ_1^2	0.16	0.33	0.24	0.33	0.45	0.33	0.34	0.35
σ_2^2	0.24	0.33	0.19	0.23	0.31	0.24	0.24	0.25
σ_3^2	1.11	0.33	0.12	0.20	0.47	0.23	0.25	0.28
σ_4^2	1.11	0.33	1.00	1.41	1.88	1.39	1.42	1.46
$\rho_{1,2}$	0.00	0.00	-0.24	-0.04	0.16	-0.05	-0.04	-0.02
$\rho_{1,4}$	0.75	0.50	0.13	0.23	0.33	0.22	0.23	0.24
$\rho_{2,4}$	0.00	0.00	-0.10	0.01	0.12	0.00	0.01	0.02
$\overline{M}$	671	513	248	342	605	359	396	434
$\hat{D}$	1.06	1.00	0.94	0.98	1.02	0.98	0.98	0.98
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.43: Comparison of F2b and F2b alt (Model V)

Long cells, fixed ζ, restricted Λ, 1 h intervals								
Thinning with step-size 100, 250K burn-in iterations								
		F2b			**F2b alt**			
	True	**Posterior quantiles**			**Posterior quantiles**			**PSR***
Parameter	**Value**	**10%**	**50%**	**90%**	**10%**	**50%**	**90%**	**150K**	**300K**
μ_1	5.03	5.11	5.24	5.37	5.13	5.27	5.40	1.30	1.12
μ_2	7.80	7.74	7.83	7.91	7.74	7.83	7.92	1.01	1.00
μ_3	3.23	3.29	3.71	4.10	3.03	3.70	4.08	1.76	1.27
μ_4	2.72	2.70	3.14	3.53	2.36	3.08	3.50	1.85	1.32
σ_1^2	0.33	0.20	0.28	0.38	0.24	0.33	0.45	1.24	1.01
σ_2^2	0.33	0.19	0.23	0.30	0.19	0.23	0.31	1.01	1.01
σ_3^2	0.33	0.13	0.26	0.49	0.12	0.20	0.47	1.26	1.09
σ_4^2	0.33	0.92	1.33	1.82	1.00	1.41	1.88	1.47	1.27
$\rho_{1,2}$	0.00	-0.26	-0.04	0.17	-0.24	-0.04	0.16	1.05	1.01
$\rho_{1,4}$	0.50	0.10	0.21	0.31	0.12	0.20	0.47	1.00	1.02
$\rho_{2,4}$	0.00	-0.11	-0.00	0.11	-0.10	0.01	0.12	1.00	1.00
$\overline{M}$	513	245	332	469	248	342	605	1.30	1.16
$\hat{D}$	1.00	0.93	0.97	1.02	0.94	0.98	1.02	1.22	1.06
* Values less than 1.2 can be taken as an indicator of convergence.								

Table B.44: Data set E1e (Model VI)

Short cells, fixed ζ, $\rho_{1,4} \neq 0$, 10 min intervals								
Thinning with step-size 100, 200K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000						
μ_1	5.08	5.03	5.06	5.17	5.28	5.16	5.17	5.18
μ_2	7.75	7.80	7.66	7.75	7.84	7.74	7.75	7.76
μ_3	2.43	3.23	1.47	1.59	1.72	1.56	1.59	1.63
μ_4	2.43	2.72	2.01	2.17	2.32	2.13	2.16	2.19
σ_1^2	0.31	0.33	0.31	0.40	0.52	0.40	0.41	0.42
σ_2^2	0.33	0.33	0.26	0.32	0.42	0.33	0.33	0.34
σ_3^2	1.00	0.33	0.40	0.48	0.58	0.47	0.49	0.51
σ_4^2	1.00	0.33	0.71	0.86	1.02	0.84	0.86	0.88
$\rho_{1,4}$	0.75	0.50	0.29	0.36	0.43	0.35	0.36	0.37
$\overline{M}$	1,136	553	785	863	949	852	866	880
$\hat{D}$	0.81	1.00	0.98	1.00	1.03	1.00	1.00	1.00
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.45: Data set E2e (Model VI)

Short cells, fixed ζ, $\rho_{1,4} \neq 0$, 1h intervals								
Thinning with step-size 100, 250K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000						
μ_1	5.27	5.03	5.28	5.40	5.53	5.40	5.40	5.41
μ_2	7.88	7.80	7.80	7.91	8.01	7.90	7.91	7.92
μ_3	2.63	2.14	1.39	1.76	2.06	1.67	1.74	1.80
μ_4	2.63	2.72	1.72	2.19	2.62	2.09	2.18	2.26
σ_1^2	0.32	0.33	0.28	0.38	0.53	0.38	0.40	0.41
σ_2^2	0.34	0.33	0.25	0.33	0.44	0.33	0.34	0.35
σ_3^2	1.00	0.33	0.13	0.23	0.44	0.24	0.26	0.28
σ_4^2	1.00	0.33	0.89	1.23	1.63	1.19	1.24	1.29
$\rho_{1,4}$	0.75	0.50	0.20	0.28	0.38	0.27	0.28	0.29
$\overline{M}$	390	487	534	708	977	704	736	767
$\hat{D}$	0.94	1.00	0.97	1.00	1.03	1.00	1.00	1.00
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.46: Data set F1e (Model VI)

Long cells, fixed ζ, $\rho_{1,4} \neq 0$, 10 min intervals								
Thinning with step-size 100, 150K burn-in iterations								
	Value		**Posterior quantiles**			**Confidence interval***		
Parameter	**Start**	**True**	**10%**	**50%**	**90%**	**2.5%**	**Mean**	**97.5%**
ζ	1,000	1,000						
μ_1	5.18	5.03	5.03	5.15	5.27	5.14	5.15	5.16
μ_2	7.72	7.80	7.57	7.68	7.80	7.67	7.68	7.69
μ_3	2.50	3.23	3.01	3.09	3.15	3.07	3.09	3.11
μ_4	2.50	2.72	2.48	2.57	2.65	2.55	2.56	2.57
σ_1^2	0.25	0.33	0.37	0.46	0.60	0.46	0.48	0.49
σ_2^2	0.35	0.33	0.34	0.43	0.56	0.44	0.45	0.46
σ_3^2	1.00	0.33	0.33	0.40	0.48	0.38	0.40	0.42
σ_4^2	1.00	0.33	0.45	0.52	0.60	0.51	0.52	0.53
$\rho_{1,4}$	0.75	0.50	0.25	0.32	0.40	0.31	0.32	0.33
$\overline{M}$	1,779	553	604	631	667	629	634	638
$\hat{D}$	0.87	1.00	0.91	0.93	0.95	0.93	0.93	0.93
* The bounds for the components of θ form a joint 95% confidence interval.								

Table B.47: Data set F2e (Model VI)

Long cells, fixed ζ, $\rho_{1,4} \neq 0$, 1 h intervals								
Thinning with step-size 100, 250K burn-in iterations								
	Value		Posterior quantiles			Confidence interval*		
Parameter	Start	True	10%	50%	90%	2.5%	Mean	97.5%
ζ	1,000	1,000						
μ_1	5.38	5.03	5.18	5.27	5.36	5.27	5.27	5.28
μ_2	7.83	7.80	7.75	7.84	7.92	7.83	7.84	7.84
μ_3	2.65	3.23	2.50	3.06	3.62	2.98	3.05	3.12
μ_4	2.65	2.72	2.24	2.75	3.31	2.69	2.76	2.83
σ_1^2	0.14	0.33	0.12	0.16	0.23	0.17	0.17	0.17
σ_2^2	0.21	0.33	0.19	0.24	0.31	0.24	0.25	0.25
σ_3^2	1.00	0.33	0.26	0.58	1.05	0.57	0.63	0.69
σ_4^2	1.00	0.33	0.34	0.67	1.04	0.64	0.69	0.73
$\rho_{1,4}$	0.75	0.50	0.14	0.24	0.34	0.23	0.24	0.24
$\overline{M}$	626	513	344	531	770	530	550	571
$\hat{D}$	1.06	1.00	0.95	0.99	1.02	0.98	0.99	0.99
* The bounds for the components of θ form a joint 95% confidence interval.								

Bibliography

[1] Krishna B. Athreya, Hani Doss, and Jayaram Sethuraman. On the convergence of the Markov chain simulation method. *The Annals of Statistic*, 24(1):69–100, 1996.

[2] Baldassare Bacchi, Gianfranco Becciu, and Nath. T. Kottegoda. Bivariate exponential model applied to intensities and durations of extreme rainfall. *Journal of Hydrology*, 155:225–236, 1994.

[3] Richard E. Barlow and Frank Proschan. *Statistical Theory of Reliability and Life Testing*. International Series in Decision Processes. Holt, Rinehart & Winston, 1975.

[4] O. E. Barndorff-Nielsen, V. K. Gupta, V. Pérez-Abreu, and E. Waymire, editors. *Stochastic Methods in Hydrology*, volume 7 of *Advanced Series on Statistical Science & Applied Probability*. Singapore: World Scientific, 1996.

[5] J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors. *Bayesian Statistics 4*. Oxford University Press, 1992.

[6] J. M. Bernardo and A. F. M. Smith. *Bayesian Statistics*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 1994.

[7] Ming-Hui Chen, Qi-Man Shao, and Joseph G. Ibrahim. *Monte Carlo methods in Bayesian computation*. Springer series in statistics. Springer, 2000.

[8] D. R. Cox and Valerie Isham. Stochastic spatial-temporal models for rain. In Barndorff-Nielsen et al. [4], chapter 1, pages 1—24.

[9] Robert Davies. NEWRAN random number generator. http://www.robertnz.net/.

[10] N. R. Draper and H. Smith. *Applied Regression Analysis*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, second edition, 1981.

[11] Jürgen Franke, Wolfgang Härdle, and Christian Hafner. *Statistics of Financial Markets*. Springer, 2004.

[12] Sylvia Frühwirth-Schnatter. *Finite Mixture and Markov Switching Models*. Springer, 2006.

[13] Andrew Gelman. Inference and monitoring convergence. In Gilks et al. [18], chapter 8, pages 131–143.

[14] Andrew Gelman and Donald B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457–511, 1992.

[15] Andrew Gelman and Donald B. Rubin. A single series from a Gibbs sampler provides a false sense of security. In Bernardo et al. [5], pages 625–631.

[16] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[17] Charles J. Geyer. Practical Markov chain Monte Carlo. *Statistical Science*, 7(4):473–511, 1992.

[18] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, 1996.

[19] Walter R. Gilks, Sylvia Richardson, and David J. Spiegelhalter. Introducing Markov chain Monte Carlo. In Gilks et al. [18], chapter 1, pages 1–19.

[20] Walter R. Gilks and Gareth O. Roberts. Strategies for improving MCMC. In Gilks et al. [18], chapter 6, pages 89–114.

[21] V. Granville and R. L. Smith. Disaggregation of rainfall time series via Gibbs sampling. citeseer.ifi.unizh.ch/110226.html.

[22] Peter J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

[23] Bernhard Gründer. *Stochastische Modelle für die Lutschadstoffverteilung in Waldgebieten*. PhD thesis, Mathematics Department of the University of Kaiserslautern, 1993.

[24] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[25] Benjami Kedem and Long. S. Chiu. On the lognormality of rain rate. *Proc. Natl. Acad. Sci. USA*, 84:901–905, 1987.

[26] C. Kipnis and S. R. S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Commun. Math. Phys.*, 104:1–19, 1986.

[27] S. L. Lauritzen and D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *J. R. Statist. Soc. B*, 50(2):157–224, 1988.

[28] Jun S. Liu and Chiara Sabatti. Generalised Gibbs sampler and multigrid Monte Carlo for Bayesian computation. *Biometrika*, 87(2):353–369, 2000.

[29] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1092, 1953.

[30] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer Verlag, 1993. available online at probability.ca/MT.

[31] Roger B. Nelsen. *An Introduction to Copulas*. Lecture Notes in Statistics. Springer Verlag, 1999.

[32] E. Nummelin. *General irreducible Markov chains and non-negative operators*. Cambridge University Press, 1984.

[33] Harry Pavlopoulos and Benjamin Kedem. Stochastic modeling of rain rate processes: a diffusion model. *Commun. Statist.–Stochastic Models*, 8(3):397–420, 1992.

[34] M. B. Priestley. *Spectral analysis and time series*, volume 1. Academic Press, 1981.

[35] D. Revuz. *Markov Chains*. North-Holland, 1975.

[36] Brian D. Ripley. *Spatial Statistics*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 1981.

[37] I. Rodriguez-Iturbe, D. R. Cox, and Valerie Isham. Some models for rainfall based on stochastic point processes. *Proc. R. Soc. Lond. A*, 410:269–288, 1987.

[38] I. Rodriguez-Iturbe, D. R. Cox, and Valerie Isham. A point process model for rainfall: further developments. *Proc. R. Soc. Lond. A*, 417:283–298, 1988.

[39] Jonathan R. Stroud, Peter Müller, and Bruno Sansó. Dynamic models for spatiotemporal data. *Journal of the Royal Statistical Society, Series B*, 63(4):673–689, 2001.

[40] Luke Tierney. Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics*, 22(4):1701–1762, 1994.

[41] Luke Tierney. Introduction to general state space Markov Chain theory. In Gilks et al. [18], chapter 4, pages 59–74.

Lebenslauf

1996	Abitur am Albert Schweitzer Gymnasium in Kaiserslautern
1997–2002	Studium der Wirtschaftsmathematik an der Universität Kaiserslautern
2002	Diplom in Wirtschaftsmathematik mit Abschlussarbeit *Locating Train Stations: Set Covering with Almost Consecutive Ones Property*
2002–2008	Wissenschaftlicher Mitarbeiter am Fachbereich Mathematik der Universität Kaiserslautern
seit 2008	Mitarbeiter am Fraunhofer Institut für Techno- und Wirtschaftsmathematik (ITWM) in Kaiserslautern

Curriculum Vitae

1996	Abitur at Albert Schweitzer Gymnasium in Kaiserslautern
1997–2002	Studies in financial mathematics at the University of Kaiserslautern
2002	Diplom in financial mathematics with the thesis *Locating Train Stations: Set Covering with Almost Consecutive Ones Property*
2002–2008	Research assistant at the Mathematics Department of the University of Kaiserslautern
since 2008	Researcher at the Fraunhofer Institute for Industrial and Financial Mathematics (ITWM) in Kaiserslautern