Catastrophe modeling with a doubly stochastic process

Bayesian inference and applications

Masterarbeit

vorgelegt beim Fachbereich Informatik und Mathematik
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Christian Kubitza

Betreuerin:
Prof. Dr. Gaby Schneider

31. August 2015

Acknowledgement

Foremost, I would like to express my deep gratitude to my supervisor Prof. Gaby Schneider for her continuous support, motivation, enthusiasm, open-mindedness, insightful comments, and immense knowledge. Her guidance helped me in all stages of the Master's thesis and she always encouraged me to explore new aspects of the topic.

Also, I am very thankful to Stefan Albert for generously providing immense computational power for simulations, Dr. Kevin Leckey for his helpful contribution of knowledge, Lorenz Ebermann for spending his time searching for mistakes in the work and Prof. Gavin Gibson for an inspiring and in-depth introduction to Bayesian inference and computational methods. Moreover, I specially thank the international catastrophe database EM-DAT for providing me with granular data about the occurrence of tropical cyclones.

Most importantly, I am grateful for the patience and encouragement of both my family and girlfriend - thanks for always supporting me and sticking by my side!

Contents

1. Introduction **1**
- 1.1. Motivation . . . 1
  - 1.1.1. Examples for collective risk models and point processes . . . 2
  - 1.1.2. Overview of the thesis . . . 4
- 1.2. Fit of the Poisson process . . . 5
- 1.3. Observations in empirical data . . . 7

2. Preliminaries **11**
- 2.1. The GLO model . . . 11
- 2.2. Bayesian inference and conjugate priors . . . 22
- 2.3. MCMC Methods . . . 26
  - 2.3.1. Gibbs sampling . . . 26
  - 2.3.2. The Metropolis-Hastings algorithm . . . 27
  - 2.3.3. MCMC convergence . . . 28
  - 2.3.4. MCMC diagnostics . . . 31

3. Bayesian inference in the GLO process **37**
- 3.1. Bayesian framework . . . 37
- 3.2. Estimation procedure . . . 46
  - 3.2.1. Choice of initial values . . . 50
  - 3.2.2. What N is *best*? . . . 52
- 3.3. Choice of priors . . . 55
  - 3.3.1. Non-informative and informative priors . . . 55
  - 3.3.2. Priors for the labels . . . 58
- 3.4. Estimation characteristics and precision . . . 58
  - 3.4.1. Influence of prior distributions . . . 66
  - 3.4.2. Influence of N . . . 71
  - 3.4.3. Influence of initial values . . . 73
  - 3.4.4. Estimation precision . . . 73

4. Data analysis, forecasting and further questions **80**
- 4.1. Data analysis for tropical cyclones . . . 80
  - 4.1.1. Africa . . . 81
  - 4.1.2. Americas . . . 82
  - 4.1.3. Asia . . . 84
  - 4.1.4. Oceania . . . 85

4.2. Forecasting . . . 86
4.2.1. The a posteriori expected number of events . . . 88
4.2.2. The a posteriori probability of default . . . 89
4.3. A collective risk model for hail claims . . . 92
4.3.1. Forecasting the probability of default . . . 94
4.4. Possible model extensions . . . 97
4.4.1. Negative Binomial distribution . . . 97
4.4.2. Trends . . . 98
4.4.3. Non-Gaussian distributions . . . 98
4.4.4. Non-bursty mode . . . 98
4.4.5. Modeling catastrophes and claims . . . 99
4.4.6. Multivariate GLOs . . . 99

5. Conclusion **101**

Appendices **104**

A. The sample data set **105**

B. List of Abbreviations **106**

C. List of Figures **107**

D. List of Tables **109**

E. Bibliography **110**

1. Introduction

1.1. Motivation

Natural catastrophes pose one of the largest risks to non-life insurance companies. However, if the severity or frequency of claims is underestimated, an insurer may not withhold sufficient capital to pay incurring losses and might even default. Therefore, it is of utter importance to obtain good estimates of future claims and the insurer's risk. For this purpose a variety of claim models exists. In non-life insurance many of these models build on the so called *collective risk model* (cf. Bowers et al. (1997, p.367)), which consists of two components: Firstly, a claim arrival process Φ, which models the *frequency* of claims, and, secondly, a claim size distribution F_Y, which models the *severity* of claims. Note that in this work we will not distinguish between claims and losses although in reality the total loss is sometimes only known several years after a claim was made (e.g. in case of liability or health insurance).

The claim arrival process Φ is often given by a point process as an increasing sequence of events $0 \leq S_1 \leq S_2 \leq ...$ (cf. Embrechts et al. (1997, p.222)). In the following work we will regularly denote the number of events of a point process Φ in $[t_1, t_2)$ as $\Phi(t_1, t_2) = |\{S_i \in [t_1, t_2) : i \in \mathbb{N}\}|$ for $t_1 < t_2$ and $\Phi(t) := \Phi(0, t)$ for $0 < t$.

In addition to the claim occurrences, the claim sizes $(Y_i)_{i\in\mathbb{N}}$ are assumed to be positive, independently and identically distributed (iid) random variables having a common non-lattice distribution function F_Y. Then, the *aggregate claims* up to time t are given by (cf. Embrechts et al. (1997, p.23) or Bowers et al. (1997, p.367))

$$L(t) = \sum_{i=1}^{\Phi(t)} Y_i. \tag{1.1}$$

A crucial quantity in insurance risk modeling is the probability of default, which is not only important from a theoretical point of view but also a practically one. For instance, the European Insurance and Occupational Pensions Authority (EIOPA) requires all European insurers to assess and constrain their one-year-default probability to 0.5%. This requirement is vital to the European regulatory insurance regime Solvency II, which will be in force from January 2016 on (cf. European Parliament (2015)). In general, *the ruin probability* Ψ *at time* t is the probability of an insurer's surplus process $U(t, u)$ falling below zero at time $t < \infty$, i.e.

$$\Psi(t, u) = \mathbb{P}(U(t, u) < 0), \tag{1.2}$$

where in the classical risk model the surplus $U(t,u)$ is given as (cf. Bowers et al. (1997, p.399))

$$U(t,u) = u + ct - L(t), \quad t \geq 0, \tag{1.3}$$

where $u \geq 0$ denotes initial capital and $c > 0$ the premium income rate.

In this work, we will exclusively focus on modeling the frequency component Φ of the aggregate claims process. Once a model for Φ is fixed, suitable loss distributions may be fitted by employing Maximum Likelihood methods, the method of moments or other methods to the empirical loss data. Frequently used loss distributions are the Log-Normal, Exponential, Gamma or Pareto distribution. One might as well consider the Fréchet, Gumbel, Weibull, Burr, Generalized Pareto or log-α-Stable distribution to be able to fit heavy tails, which are particularly present in catastrophe losses. Applications for some of these distributions may be found in Chernobai et al. (2005), McNeil (1997) or Beirlant et al. (2001).

1.1.1. Examples for collective risk models and point processes

A very prominent example for a collective risk model is the *Crámer-Lundberg* model (cf. Embrechts et al. (1997, p.22)). In this model all claim sizes are assumed to be independent of Φ, whereas Φ follows a homogeneous Poisson process with rate $\lambda > 0$:

Definition 1.1 (Poisson process)**.** *(Durrett (2012, p.103)) Let $\lambda(x)$ be a non-negative real valued measurable function. A point process Φ is called inhomogeneous Poisson process, if*

(i) $\Phi(0) = 0$,

(ii) $\Phi(t)$ *has independent increments, and*

(iii) $\Phi(t) - \Phi(s)$ *is Poisson with mean* $\int_s^t \lambda(x)dx$

for any $s < t$. If $\lambda(x)$ is constant, we call Φ a homogeneous Poisson process.

There are two main advantages of modeling claim arrivals with a homogeneous Poisson process: Firstly, the theoretical justification of the Poisson distribution due to the Poisson approximation (the frequency of very rare events is approximately Poisson distributed, cf. Klenke (2006, p.79)) and, secondly, its simplicity. For example, parameter estimation is straightforward e.g. by means of Maximum-Likelihood methods, the distribution of $L(t)$ is the well known *Compound-Poisson* distribution (cf. Heilmann (1988, p.95)) and quantities of interest as the expected value, $\mathbb{E}[L(t)]$, or variance, $\text{var}(L(t))$, are known in closed form. Moreover, several approximations and closed-form solutions for special cases of ruin probabilities exist (cf. Embrechts et al. (1997, p.28) and Heilmann (1988, p.245-268)).

Nonetheless, the *Crámer-Lundberg* model exhibits major drawbacks: Most crucially, the assumption of a constant intensity $\lambda > 0$ of the claim arrival process makes it impossible to model any kind of seasonal behavior (e.g. for snow storms, which only occur in winter time), trends (e.g. the frequency of cyber attacks has increased heavily over the last few

decades), or *overdispersion*. The latter describes the property of a probability distribution exhibiting a larger variance than expected value. This is not possible for claim numbers in the *Crámer-Lundberg* model since the expected value of the Poisson distribution equals its variance, i.e. $\mathbb{E}[\Phi(t)] = \text{var}(\Phi(t)) = \lambda t$.

In response to these shortcomings, a variety of extensions of the *Crámer-Lundberg* model were developed. For example, Boudreault et al. (2014) and Baumgartner et al. (2015) analyze collective risk models that incorporate dependent severity and frequency components. Focussing only on the frequency component, Chernobai et al. (2005) add an oscillatory component by considering Φ to be an inhomogeneous Poisson process with a deterministic sinusoidal intensity function $\lambda(t) = a + b2\pi \cdot \sin(2\pi(t-c))$, for which the constants $a, b, c > 0$ are fitted to a given dataset by least squares estimation. Clearly, their model is able to include a trend, as well.

To allow for overdispersion, the Negative Binomial distribution represents a natural extension of the Poisson distribution: It incorporates a dispersion parameter $r \in \mathbb{N}$ such that the probability mass function (pmf) for $X \sim NB(r, p)$ is given by (cf. Evans et al. (2010))

$$\mathbb{P}(X = k) = \binom{k + r - 1}{k} p^k (1 - p)^r. \tag{1.4}$$

Here, $k \in \mathbb{N}_0$ represents the number of successes which occur in a sequence of independent Bernoulli trails with probability $p \in (0, 1)$ before a target number of failures $r \in \mathbb{N}$ is reached. One may reparametrize the distribution in terms of its mean $m = r\frac{p}{1-p}$. In this case we have $\mathbb{E}[X] = m$ and $\text{var}(X) = m\left(1 + \frac{m}{r}\right)$. Thus, r may be interpreted as a dispersion parameter accounting for a larger variance $\text{var}(X) > \mathbb{E}[X]$. Also, by substituting $\binom{k+r-1}{k} = \frac{\Gamma(k+r)}{\Gamma(k+1)\Gamma(r)}$ with the Gamma function $\Gamma(\cdot)$ one may extend the distribution for a real parameter $r > 0$. Moreover, the Negative Binomial distribution converges to a Poisson distribution with intensity m as $r \to \infty$ (cf. Hilbe (2007)). Nevertheless, as Eastoe and Tawn (2010) point out, a model that assumes the number of events in a certain time window (e.g. one year) to be Negative Binomially distributed does not only lack a theoretical justification, but is also highly inflexible, as it can not easily be extended to a point process (although Gregoire (1983) shows the existence of Negative Binomial processes).

Another class of point processes is given by *Poisson-cluster processes*, which model clusters of events at every point of a Poisson processes (cf. e.g. Onof et al. (2000)). These processes are a special case of *Cox processes*. The latter were introduced by Cox (1955) as doubly stochastic Poisson processes with random intensity function $\lambda(t)$. One main assumption of *Poisson-cluster processes* is the irregular (i.e. not cyclical) occurrence of bursts of events. However, as we will show in sections 1.3 and 4.3, natural catastrophes like tropical cyclones or hail occur in rather regular bursts, i.e. the distance between bursts is comparatively stable. Therefore, it is reasonable to allow for other background rhythms, which is done by Møller and Torrisi (2005) in the form of *Generalized Shot Noise Cox processes*.

1.1.2. Overview of the thesis

In this work we will analyze two datasets: Mainly, we will analyze data from the international catastrophe database EM-DAT (Guha-Sapir et al. (2014)) that report the occurrence of tropical cyclones in Africa, Americas, Asia and Oceania. In section 1.2 we will fit a homogeneous Poisson process to the empirical data, which will underline the shortcomings of this model. Therefore, in section 1.3 the empirical data will be reviewed more intensively. This will motivate the application of the GLO model from Bingmer et al. (2011), which will be analyzed in section 2.1.

Additionally, in section 4.3 we will analyze the occurrence and claim sizes of hail damage of a large German insurance company. This application will illustrate the particular use of the GLO model for risk management in insurance companies.

The remainder of the work will be concerned with developing a Bayesian estimation framework for the GLO model. In chapter 2 we will provide the necessary mathematical background for the estimation procedure: We will introduce basic notations and give a short introduction to Bayesian inference and computational methods. The Bayesian estimation framework for the GLO process will be developed in chapter 3. Since the estimation procedure will rely on a Markov chain Monte Carlo (MCMC) algorithm with several input parameters, we will present methods how to choose these parameters and assess the influence of the input parameters on the estimation procedure. Moreover, for nine different parameter combinations we will analyze the estimation precision of the proposed estimators.

At the end, in chapter 4 the inference framework will be applied to the data and open questions and possible extensions of the inference framework and the model itself will be discussed. Also, we show how to forecast the future occurrence of events by using samples from the MCMC algorithm. Chapter 5 concludes. Additional details about the data set are reported in the appendix.

All data evaluations and simulations were performed with the statistical software RStudio (version 0.98.1028), which bases on R version 3.0.2. We mainly employed a Mac Mini with 2.6 GHz Intel Core i7, 16 GB RAM and Mac OS X Yosemite 10.10.4. To compute the estimation errors in section 3.4.4 we performed simulations on a Mac Pro with 2.7 GHz 12 Core Intel, 32 GB RAM and Mac OS X Yosemite 10.10.2.

1.2. Fit of the Poisson process

The sample data set consists of the occurrence times $S_1^{(\text{continent})}, ..., S_{n_{\text{continent}}}^{(\text{continent})}$ (in days) of tropical cyclones for the continents Africa, Americas, Asia and Oceania. For simplicity we will drop the continent-index in the following, since the data is analyzed for each continent separately. The data was generously made available by EM-DAT (Guha-Sapir et al. (2014); for more information about EM-DAT and the data format see chapter A, p. 105, in the appendices). In this section we will review the fit to a homogeneous Poisson process.

For a homogeneous Poisson process with intensity $\lambda > 0$ the Inter-Event-Interval (IEI) distribution is given by an Exponential distribution with rate λ, $Exp(\lambda)$ (cf. Durrett (2012, p.97)). Therefore, we employ the Maximum-Likelihood estimator (MLE) of the Exponential distribution to estimate the parameter λ. In particular, the probability density function (pdf) of $Exp(\lambda)$ is given as

$$f_{Exp(\lambda)}(x) = \lambda e^{-\lambda x}. \tag{1.5}$$

Therefore, the log-Likelihood function is given as

$$\ell(\lambda; IEI) = \sum_{i=1}^{n-1} \left(\log \lambda - \lambda\, IEI_i\right), \tag{1.6}$$

where $IEI = (S_2 - S_1, S_3 - S_2, ..., S_n - S_{n-1})$. Since $\frac{\partial \ell}{\partial \lambda^2}(\lambda) = -\frac{n-1}{\lambda^2} < 0$, the MLE is the solution of

$$\frac{\partial \ell}{\partial \lambda}(\lambda) = \frac{n-1}{\lambda} - \sum_{i=1}^{n-1} IEI_i = 0, \tag{1.7}$$

which is

$$\hat{\lambda}_{ML} = \left(\frac{1}{n-1}\sum_{i=1}^{n-1} IEI_i\right)^{-1} = \left(\overline{IEI}\right)^{-1}. \tag{1.8}$$

The resulting estimates $\hat{\lambda}_{ML}$ are reported in Table 1.1. Additionally, we compare the sample variance

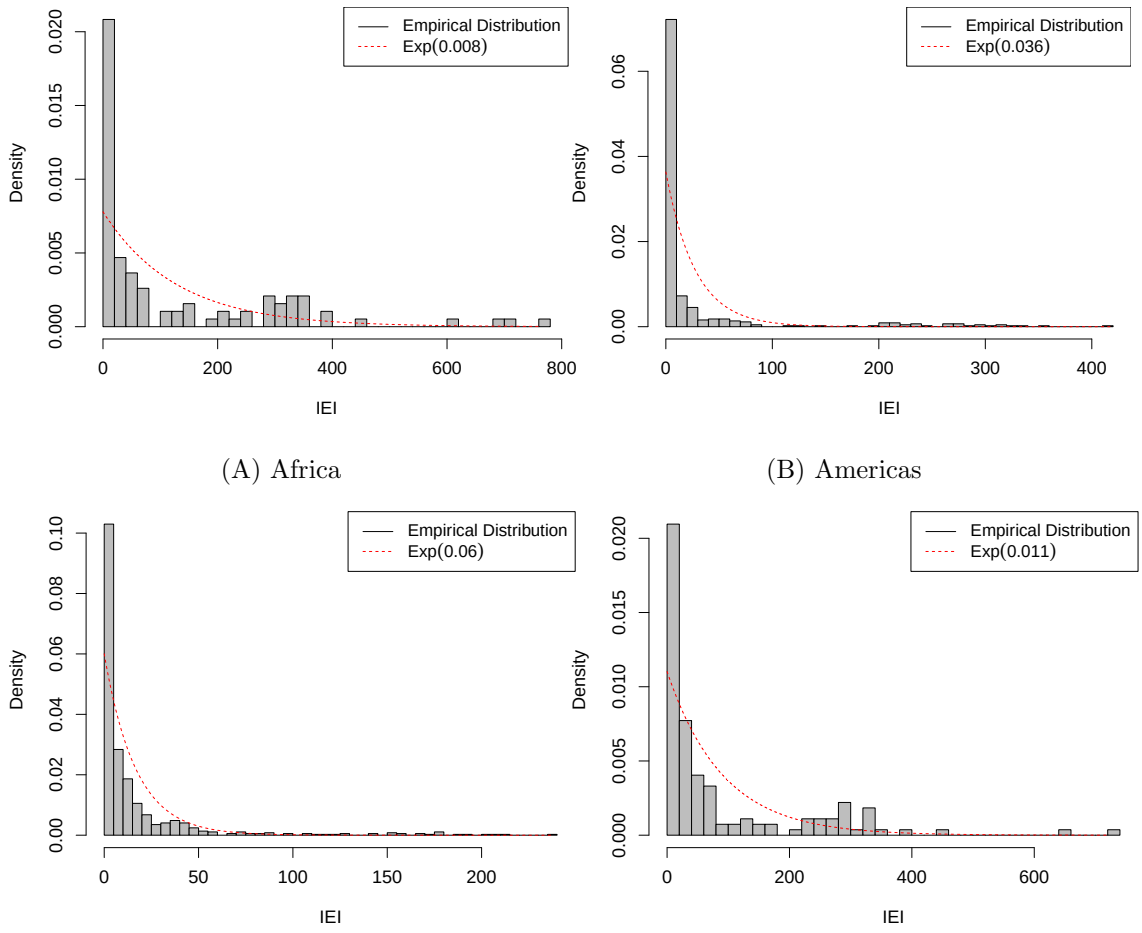
$$s^2(IEI) := \frac{1}{n-2}\sum_{i=1}^{n-1} \left(IEI_i - \overline{IEI}\right)^2 \tag{1.9}$$

with the theoretical variance $\hat{\lambda}_{ML}^{-2}$ of the fitted distribution: For all continents the sample variance is substantially larger.

Moreover, the fitted Exponential distribution does not provide a good fit to the empirical IEI distributions as Figure 1.1 shows. In conclusion, the model lacks important characteristics of the data like a large variance and two-peaked density and, therefore, is inappropriate.

Continent	MLE, $\hat{\lambda}_{ML}$	Theoretical variance, $\hat{\lambda}_{ML}^{-2}$	Sample variance, $s^2(IEI)$
Africa	0.008	16485.5	30730.5
Americas	0.036	757	4826.3
Asia	0.060	277.3	1077.8
Oceania	0.011	8228.9	17431.2

Table 1.1.: Tropical cyclones: ML estimates, theoretical and sample variance of the Poisson model



(A) Africa

(B) Americas

(C) Asia

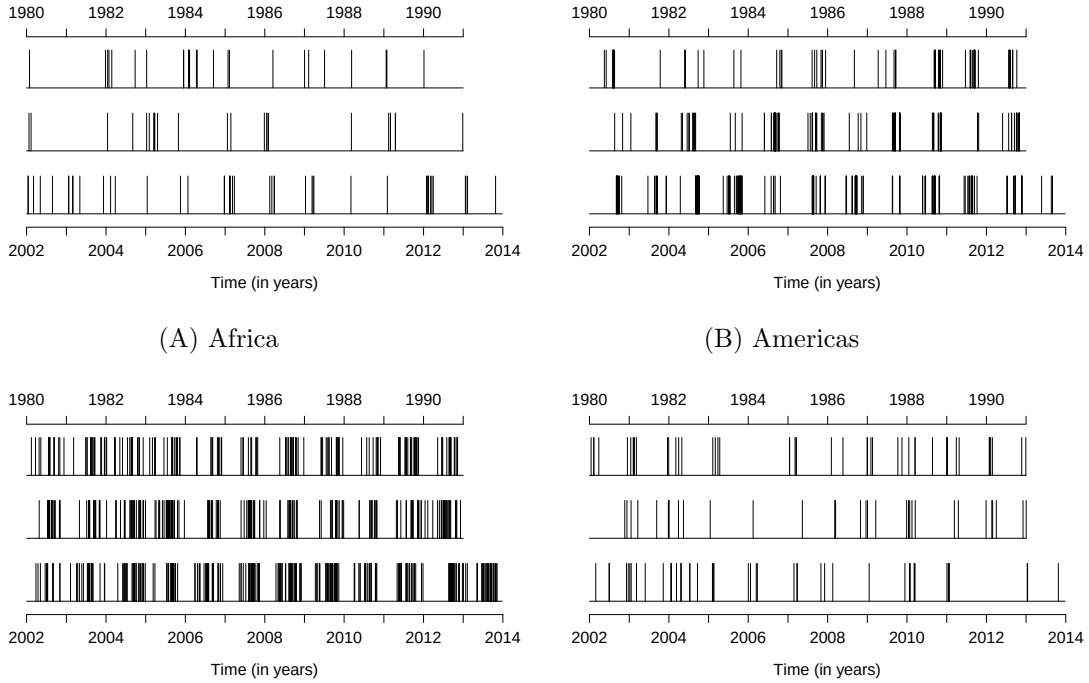
(D) Oceania

Figure 1.1.: IEI distributions of tropical cyclones (empirical density (grey/black) and fitted exponential density (red, dotted)) for (A) Africa, (B) Americas, (C) Asia and (D) Oceania.

1.3. Observations in empirical data

In this section the data is analyzed more intensively using rasterplots, inter-event-interval distributions and autocorrelation histograms. For this purpose we mainly follow the methodology of Bingmer (2012).

Rasterplots: In these plots the raw data is shown: For each event a bar is plotted at its time of occurrence. Figure 1.2 shows the rasterplots of the occurrence times of tropical cyclones in Africa, Americas, Asia and Oceania. Clearly, tropical cyclones tend to occur in bursts, i.e. moments of increased cyclone activity are followed by moments without any cyclones. These bursts are rather large in Americas and Asia and small in Africa and Oceania. Furthermore, in all four continents the number of events per burst is not constant and usually larger than one and the bursts seem to follow a very rhythmic behavior.



(A) Africa

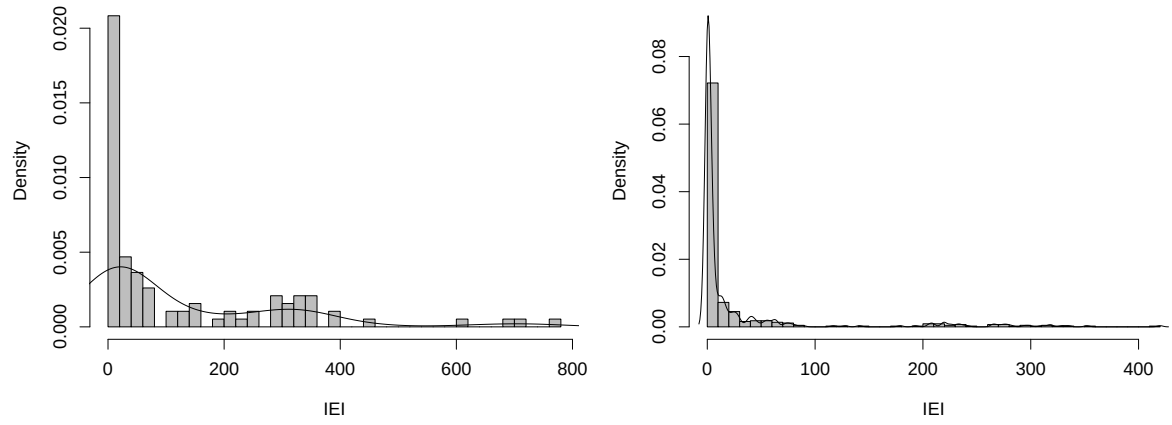
(B) Americas

(C) Asia

(D) Oceania

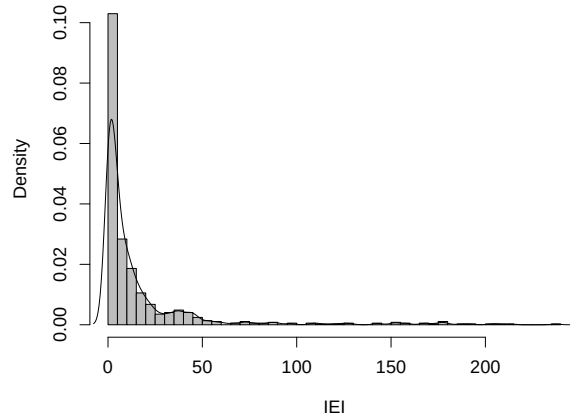
Figure 1.2.: Rasterplots of tropical cyclones in (A) Africa, (B) Americas, (C) Asia and (D) Oceania.

Inter-Event-Interval distribution: In Figure 1.3 the histograms with a density kernel estimate of the empirical IEI distribution of tropical cyclones is shown. Clearly, the IEI distributions exhibit two peaks. Intuitively, the first peak corresponds to small IEIs within bursts whereas the second peak corresponds to large IEIs between different neighboring bursts.

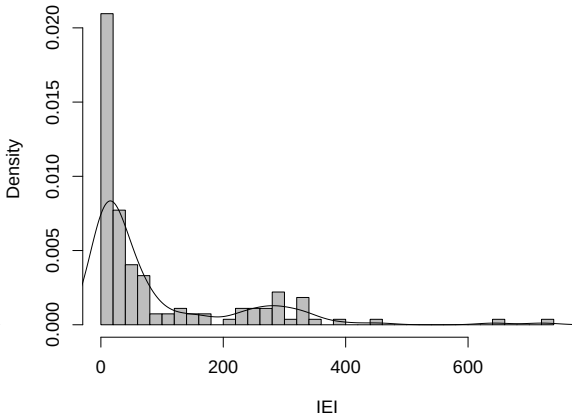


(A) Africa

(B) Americas



(C) Asia

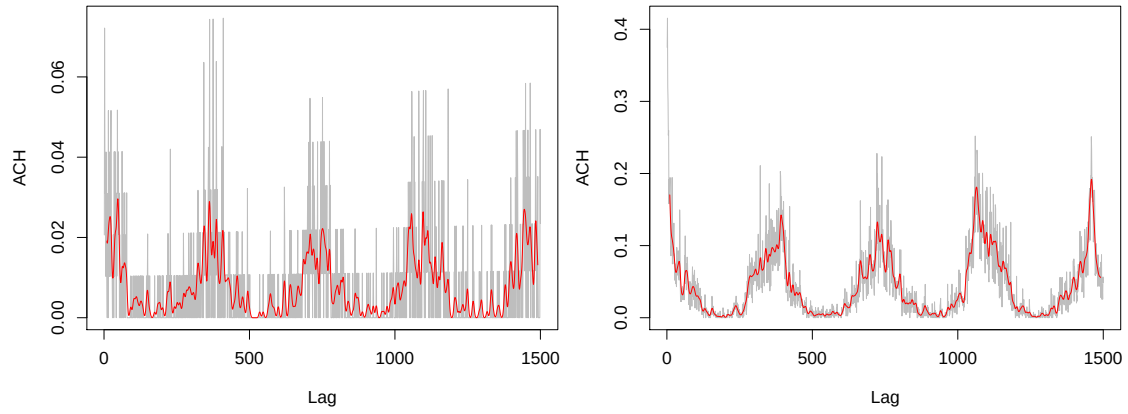


(D) Oceania

Figure 1.3.: IEI distributions of tropical cyclones in (A) Africa, (B) Americas, (C) Asia and (D) Oceania.

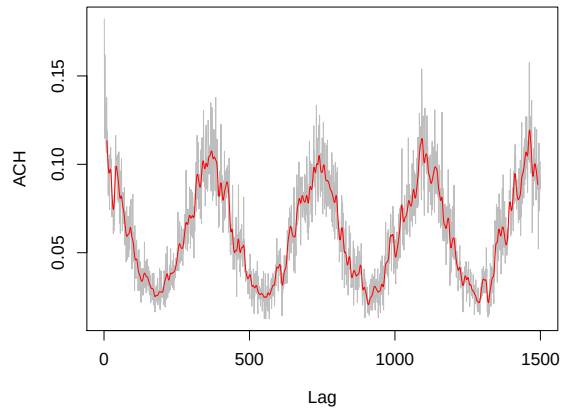
Autocorrelation histogram: An autocorrelation histogram (ACH) is very useful to analyze the oscillatory behavior of point processes. It is essentially a histogram of the distances $S_j - S_i$ of all pairs of event times (S_i, S_j) and, thus, an estimate for the ACF of a point process (cf. Remark 2.6). For a more detailed introduction to ACHs and their properties see e.g. Moore et al. (1966) or Perkel et al. (1967). Also, Bingmer (2012, p.65-73) and Bingmer et al. (2011) show how to interpret ACHs in the context of the GLO model.

To give an example for an ACH, consider a very rhythmic process that deterministically produces an event in time steps of size $\mu > 0$. For a realization of this process the ACH intuitively exhibits peaks at $k\mu$, $k \in \mathbb{Z}$, and equals zero otherwise. Consequently, regular processes exhibit multiple, repetitive peaks. Furthermore, for bursty processes there exists a peak near zero, since intervals between events belonging to the same burst are usually quite small. The visual inspection of the ACHs of tropical cyclones in Figure 1.4 clearly indicates a regular (oscillatory) and bursty occurrence behavior.

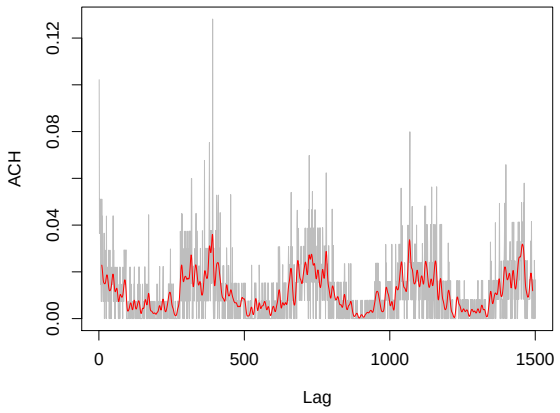


(A) Africa

(B) Americas



(C) Asia



(D) Oceania

Figure 1.4.: ACHs of tropical cyclones in (A) Africa, (B) Americas, (C) Asia and (D) Oceania.

In conclusion, all figures indicate a bursty and regular occurrence behavior of tropical cyclones. An explanation for these properties is given by the creation process of tropical cyclones: According to Gray (1968), Gray (1975) and Gray (1979) several conditions have to be in place to allow a tropical cyclone to evolve. Among these are warm ocean waters (a maximum sea surface temperature of at least 26.5°C) whereas the atmosphere needs to cool down rapidly with increasing height to allow releasing the heat of condensation which powers the tropical cyclone. Moreover, high humidity and a pre-existing system of disturbed weather is necessary. For an overview about the genesis of tropical cyclones we also refer to Henderson-Sellers et al. (1998). Clearly, it is not common to have all necessary conditions fulfilled at the very same time. In fact, there are only distinctive months in distinctive regions in which tropical cyclones can develop (mainly summer time in equatorial areas). This explains why bursts of tropical cyclones occur regularly. Since the conditions normally last for some time, during that time many cyclones occur, which in turn explains the burstiness.

To conclude, the data exhibits properties that can not be found in Poisson processes, namely regularly occurring bursts. However, the GLO model proposed by Bingmer et al. (2011) is able to model this particular occurrence behavior as it is shown in the next chapter.

2. Preliminaries

In the following chapter, firstly, we will review the definition and some basic properties of the GLO-process as it is introduced in Bingmer et al. (2011). In the second and third part of the chapter Bayesian inference and computational methods for Bayesian inference will be presented.

2.1. The GLO model

In the following we will review the GLO model as it is introduced in Bingmer et al. (2011). For the construction of a Bayesian estimation framework it will be sufficient to focus on the basic definitions and general characteristics of the process. For a more in-depth introduction and discussion of the GLO process and its properties we refer to Bingmer (2012). An introduction to the theory of point processes can be found for example in Daley and Vere-Jones (1988), König and Schmidt (1992) or Cox and Isham (1980).

The GLO process builds on an oscillatory background rhythm that follows a random walk and places a Poisson distributed number of events around every beat of this rhythm according to a Gaussian distribution. Due to the dominant role of the Gaussian distribution in this model the process is called "Gaussian Locking to a free Oscillator" – GLO. It is very similar to the ELO ("Exponential Locking to a free Oscillator"), which contains the same background rhythm but distributes events at beats with exponentially decreasing firing intensity (cf. Schneider (2008)). The GLO process is constructed in 3 steps:

1. **Background rhythm:** The background rhythm consists of random background beats $\mathcal{B} = (B_k)_{k\in\mathbb{Z}}$ that follow a stationary random walk with normally distributed increments $B_{k+1} - B_k \sim \mathcal{N}(\mu, \sigma_1^2)$.

2. **Number of events:** The number of events P_k at beat B_k is assumed to be Poisson distributed with firing rate $\gamma > 0$, $P_k \sim Pois(\gamma)$.

3. **Event variation:** At every beat B_k all events $S_{k,1}, ..., S_{k,P_k}$ are normally distributed around the beat with independent increments, i.e.

$$S_{k,i} = B_k + Z_{k,i} \quad \text{with } Z_{k,i} \overset{iid}{\sim} \mathcal{N}(0, \sigma_2^2). \tag{2.1}$$

All random variables Z_{i_1,i_2}, P_{i_3} are independent for all $i_1, i_3 \in \mathbb{Z}$ and $i_2 \in \mathbb{N}$ and independent of the random walk $(B_k)_{k\in\mathbb{Z}}$. If Φ is a GLO process, we denote $\Phi \sim \text{GLO}(\mu, \sigma_1, \sigma_2, \gamma)$.

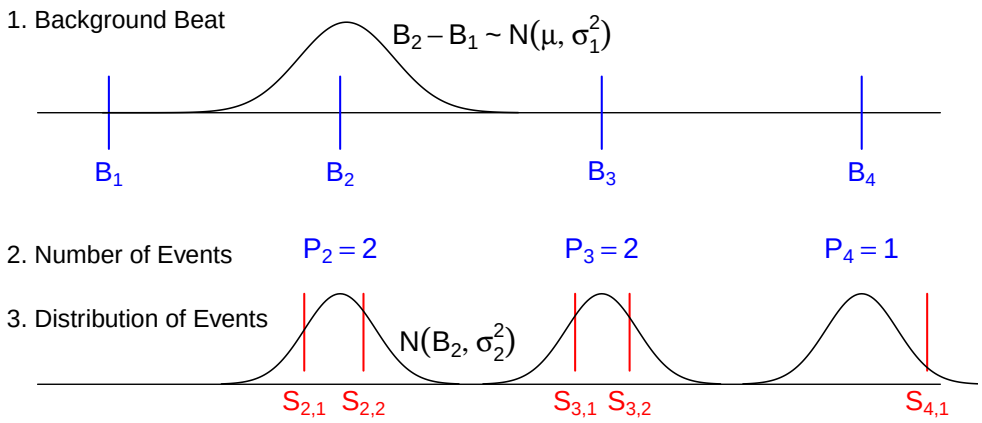


Figure 2.1.: Construction of the GLO process: The background rhythm (blue) has normally distributed increments $B_{k+1} - B_k \sim \mathcal{N}(\mu, \sigma_1^2)$. At each beat B_k the number of events P_k (blue) is Poisson distributed, $P_k \sim Pois(\gamma)$, and every event $S_{k,i}$ (red) is placed around its birth beat B_k according to $\mathcal{N}(B_k, \sigma_2^2)$.

The stationary random walk: The construction of a random walk $(B_k)_{k\in\mathbb{Z}}$ in step 1 is as follows: At first, we construct a random walk with $B'_1 = 0$ and independent and normally distributed increments $B'_{k+1} - B'_k$,

$$\left(B'_{k+1} - B'_k\right) \overset{iid}{\sim} \mathcal{N}(\mu, \sigma_1^2), \quad \text{for } k \in \mathbb{Z}. \tag{2.2}$$

Stationarity is achieved by setting the origin of the time axis at random. This is done by taking a large boundary a and setting B_1 to the first time that $(B'_k)_{k\in\mathbb{Z}}$ reaches a. Then, the steps of the stationary random walk are given as (cf. Bingmer (2012, p.17))

$$B_k := \lim_{a\to\infty} B'_{k+\tau_a-1} - a, \quad \text{for } k \in \mathbb{Z}, \tag{2.3}$$

where τ_a is the time at which the random walk first reaches the height a, or ∞,

$$\tau_a := \inf\{k \in \mathbb{Z} : B'_k > a\}. \tag{2.4}$$

Bingmer (2012, p.31) shows that with this construction the GLO process is a stationary process. However, note that conditional on knowing one or more beats the intensity of the process is not constant any more.

Remark 2.1 (Non-bursty mode and stationarity)**.** *Bingmer et al. (2011) also allow for a non-bursty mode, in which the random number of events at one beat is Bernoulli-distributed, $P_k \sim Bern(\gamma)$, $\gamma \in [0, 1]$. However, we do not consider this mode in this work since the given dataset exhibits clusters of events with more than one event, i.e. the process is clearly bursty (cf. section 1.3). Nonetheless, in section 4.4.4 we will describe how the non-bursty mode may be embedded into the estimation framework.*

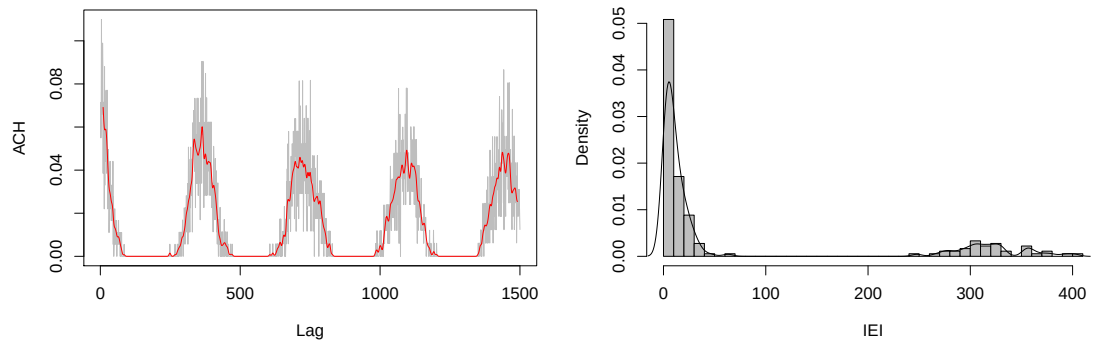
The GLO model was first introduced by Bingmer et al. (2011) particularly for the purpose of modeling the behavior of neuronal spike trains. Therefore, it is possible to distinguish between regular and irregular as well as between bursty and non-bursty processes. A regular (irregular) spike train exhibits (lacks of) dominant rhythmic activity and oscillation, whereas in bursty spike trains events tend to occur in clusters, i.e. in times of high intensity, which are followed by periods without any events. With respect to this definition of regularity the Poisson process is an irregular, non-bursty process (cf. Bingmer (2012, p.52)). The parameters of a GLO process directly indicate its level of regularity and burstiness: μ and σ_1 describe the mean increment and variability of the background rhythm. γ is the firing rate (average number of events per burst) whereas σ_2 represents the burst width. In the following we will denote the parameter vector as $\eta = (\mu, \sigma_1, \sigma_2, \gamma)$.

Remark 2.2 (GLO simulation)**.** *Bingmer (2012, p.41 and p.137) develops an algorithm (Algorithm 1) to simulate events of a GLO process in the time interval* $[0, T]$*. This algorithm also allows us to also save the simulated (original) beat locations and labels (i.e. markers that indicate to which beat each event belongs; for a thorough definition see chapter 3).*

In order to ensure stationarity of the simulated background rhythm the algorithm starts simulating beats at $-K_0\mu$*. We set* $K_0 = 30000$ *to allow the process to converge to stationarity. Then, only beats in the interval* $[-K, T + K]$ *with* $K > 0$ *are considered to place events in* $[0, T]$*. However, if* K *is very small and* σ_2 *large, there may exist beats outside* $[-K, T + K]$ *that place events in* $[0, T]$ *but are not considered.*

Therefore, Bingmer (2012, p.39-42) introduces a precision parameter $\varepsilon > 0$ *that controls the size of the interval* $[-K(\varepsilon), T+K(\varepsilon)]$ *in which beats are considered. By choosing* $K(\varepsilon) = F^{-1}_{\mathcal{N}(0,\sigma_2^2)}(1-\varepsilon)$ *the probability that beats larger than* $T+K(\varepsilon)$ *place events in* $[0, T]$ *is bounded by* ε *and analogously the probability that beats smaller than* $-K(\varepsilon)$ *place events in* $[0, T]$ *is bounded by* ε*. We choose* $\varepsilon = 10^{-9}$ *to ensure a very high precision of the simulation procedure. Both values,* $K_0 = 30000$ *and* $\varepsilon = 10^{-9}$*, remain fixed throughout the remaining chapters whenever the simulation algorithm is employed.*

To illustrate the behavior of a typical GLO process we employ the aforementioned algorithm to simulate events in $[0, 13000]$ with parameters $\eta = (364, 20, 20, 5)$. In Figure 2.2 we show the resulting autocorrelation histogram (ACH) and inter-event-interval (IEI) distribution. As it is typical for bursty processes (cf. Bingmer (2012, chapter 3)) the ACH exhibits a narrow peak near zero due to small intervals within bursts. Moreover, this peak is followed by clear, repetitive peaks, which indicate the oscillatory behavior. Analogously, the IEI distribution exhibits two peaks: one (small) peak representing intervals within bursts and one (large) peak representing intervals between bursts. Note that these peaks tend to move together for more irregular processes (cf. Bingmer (2012, chapter 3)).



(A) ACH (B) IEI distribution

Figure 2.2.: (A) ACH and (B) IEI distribution of a simulated GLO process with parameters $\mu = 364$, $\sigma_1 = 20$, $\sigma_2 = 20$ and $\gamma = 5$.

Since the model is constructed to describe regular oscillations with distinguishable bursts, σ_1/μ and σ_2/μ should be small. Particularly if $\sigma_1/\mu < 1/3$, the process very likely maintains the property of monotone increasing beat locations (in this case $\mathbb{P}(B_{k+1} - B_k < 0) < 0.0014$, cf. Bingmer (2012, p.23)).

In the following we will present several representations and properties of the GLO process, which are helpful to assess its characteristics in the context of collective risk models. We start with the representation of the GLO as a Cox process:

Proposition 2.1. *(Bingmer (2012, p.34))* $\Phi \sim GLO(\mu, \sigma_1, \sigma_2, \gamma)$ *is a Cox process with random intensity*

$$\lambda(t) = \gamma \sum_{k \in \mathbb{Z}} f_{\mathcal{N}(B_k, \sigma_2^2)}(t), \tag{2.5}$$

i.e. Φ *is a Poisson process* $\varphi(\cdot \mid \Lambda)$ *with intensity measure*

$$\Lambda = \int_A \lambda(t)\, dt, \quad A \in \mathbb{B}(\mathbb{R}). \tag{2.6}$$

From Proposition 2.1 it can easily be inferred that for $\sigma_2 \to \infty$ the intensity of the process converges to a constant since

$$\lim_{\sigma_2 \to \infty} \frac{\partial \lambda(t)}{\partial t} = \lim_{\sigma_2 \to \infty} \frac{\gamma}{\sigma_2 \sqrt{2\pi}} \sum_{k \in \mathbb{Z}} -\frac{(t - B_k)}{\sigma_2^2} \exp\left(-\frac{(t - B_k)^2}{2\sigma_2^2}\right) = 0. \tag{2.7}$$

This observation gives an intuition about the following convergence behavior:

Proposition 2.2 (Convergence to Poisson process)**.** *(Bingmer (2012, p.35-36))*
Let $\Phi \sim GLO(\mu, \sigma_1, \sigma_2, \gamma)$ *with arbitrary* $\mu \in \mathbb{R}$, $\sigma_1, \gamma > 0$. *Then, if* σ_2 *tends to infinity, the GLO* Φ *converges weakly to a Poisson process with intensity* $\lambda = \gamma/\mu$.

When considering a collective risk model

$$L(t_1, t_2) = \sum_{i=1}^{\Phi(t_1,t_2)} Y_i \tag{2.8}$$

it is important to assess the number of events in $(t_1, t_2]$, $t_2 > t_1$, which is for the GLO process given as

$$\Phi(t_1, t_2) = \sum_{k\in\mathbb{Z}} \sum_{i=1}^{P_k} \mathbb{1}_{\{S_{k,i}\in(t_1,t_2]\}}. \tag{2.9}$$

Then, if we assume the claim severities Y_i to be independent from their frequency Φ, the expected aggregate claims in $(t_1, t_2]$ are given as

$$\mathbb{E}[L(t_1, t_2)] = \mathbb{E}\left[\sum_{i=1}^{\Phi(t_1,t_2)} Y_i\right] = \mathbb{E}[\Phi(t_1, t_2)]\mathbb{E}[Y_1] \tag{2.10}$$

by employing Wald's theorem (cf. Klenke (2006, p.99)). Moreover, the variance is by the theorem of Blackwell-Girshick (cf. Klenke (2006, p.103)) given as

$$\mathrm{var}(L(t_1, t_2)) = \mathrm{var}\left(\sum_{i=1}^{\Phi(t_1,t_2)} Y_i\right) = \mathbb{E}[Y_1]^2\mathrm{var}(\Phi(t_1, t_2)) + \mathbb{E}[\Phi(t_1, t_2)]\mathrm{var}(Y_1). \tag{2.11}$$

Therefore, it may be useful to compute the expected value and variance of $\Phi(t_1, t_2)$.

Remark 2.3. *For a homogeneous Poisson process φ with intensity λ the number of events in $(t_1, t_2]$ is $\varphi(t_1, t_2) \sim Pois((t_2-t_1)\lambda)$. According to Proposition 2.2 for the GLO we obtain $\mathcal{L}\big(\Phi(t_1, t_2)\big) \to Pois((t_2 - t_1)\gamma/\mu)$ if $\sigma_2 \to \infty$.*

Additionally, we obtain a similar convergence behavior for $\sigma_2 \to 0$: In this case the event times converge weakly to the beat locations, i.e. for all $k \in \mathbb{Z}$ and $i = 1, ..., P_k$ (we apply the dominated convergence theorem, cf. Klenke (2006, p.135))

$$\lim_{\sigma_2\downarrow 0} \mathbb{P}(S_{k,i} \in (t_1, t_2]) = \mathbb{E}\left[\lim_{\sigma_2\downarrow 0} \mathbb{P}\left(B_k + Z_{k,i} \in (t_1, t_2] \mid B_k\right)\right] = \mathbb{P}(B_k \in (t_1, t_2]). \tag{2.12}$$

Due to the conditional independence of the events it follows that the characteristic function of $\Phi(t_1, t_2)$ converges: For all $u \in \mathbb{R}$ we have (in this case i is the imaginary unit):

$$\lim_{\sigma_2\downarrow 0} \mathbb{E}\left[\, e^{iu\Phi(t_1,t_2)} \mid \mathcal{B}\right] = \lim_{\sigma_2\downarrow 0} \mathbb{E}\left[\prod_{k\in\mathbb{Z}} \mathbb{E}\left[\, e^{iu\sum_{j=1}^{P_k} \mathbb{1}_{\{S_{k,j}\in(t_1,t_2]\}}} \mid P_k, \mathcal{B}\right] \mid \mathcal{B}\right] \tag{2.13}$$

$$= \mathbb{E}\left[\prod_{k\in\mathbb{Z}} \prod_{j=1}^{P_k} \mathbb{E}\left[\, e^{iu\mathbb{1}_{\{B_k\in(t_1,t_2]\}}} \mid \mathcal{B}\right] \mid \mathcal{B}\right] = \mathbb{E}\left[\, e^{iu\sum_{k\in\mathbb{Z}}\sum_{j=1}^{P_k} \mathbb{1}_{\{B_k\in(t_1,t_2]\}}} \mid \mathcal{B}\right]. \tag{2.14}$$

Since $\sum_{k\in\mathbb{Z}}\sum_{i=1}^{P_k}\mathbb{1}_{\{B_k\in(t_1,t_2]\}}$ *is conditional on the background rhythm* $\mathcal{B}$ *the superposition of thinned Poisson distributions (cf. Proposition 2.1 and below), we have*

$$\lim_{\sigma_2\downarrow 0}\mathcal{L}\left(\Phi(t_1,t_2)\mid\mathcal{B}\right)=Pois\left(\gamma\sum_{k\in\mathbb{Z}}\mathbb{1}_{\{B_k\in(t_1,t_2]\}}\right). \tag{2.15}$$

Therefore, if σ_2 *is sufficiently small and the number of beats in* $(t_1,t_2]$ *is known, we can expect* $\Phi(t_1,t_2)$ *to be approximately Poisson distributed. If the beats are unknown, the unconditional limiting probability for observing* $v\in\mathbb{N}_0$ *events in* $(t_1,t_2]$ *is*

$$\lim_{\sigma_2\to 0}\mathbb{P}(\Phi(t_1,t_2)=v)=\int_{\mathbb{R}^{\mathbb{Z}}}\frac{\left(\gamma\sum_{k\in\mathbb{Z}}\mathbb{1}_{\{b_k\in(t_1,t_2]\}}\right)^v}{v!}e^{-\gamma\sum_{k\in\mathbb{Z}}\mathbb{1}_{\{b_k\in(t_1,t_2]\}}}d\mathbb{P}_{\mathcal{B}}(b), \tag{2.16}$$

which may be computed e.g. by means of (Markov Chain) Monte Carlo methods.

Proposition 2.2 and Remark 2.3 give an intuition about the distribution of the number of GLO events for very large and very small burst width σ_2. However, σ_2 may in general not be large or small enough to obtain good approximations. Nonetheless, we can still calculate the conditional distribution and expectation of $\Phi(t_1,t_2)$:

Bingmer (2012, p.32) shows that a priori the expected intensity of the process is given as

$$\mathbb{E}[\Phi(0,1)]=\frac{\gamma}{\mu}. \tag{2.17}$$

However, in this work we assume to have already observed a number of events. Therefore, we are interested in distributional properties of $\Phi(t_1,t_2)$ conditional on background beats. For this purpose, we firstly focus on one beat B_k with Poisson distributed number of events $P_k\sim Pois(\gamma)$. The number of events that B_k places in $(t_1,t_2]$ is given by

$$\Phi_k(t_1,t_2):=\sum_{i=1}^{P_k}\mathbb{1}_{\{S_{k,i}\in(t_1,t_2]\}}. \tag{2.18}$$

Clearly, $\mathcal{L}\left(\mathbb{1}_{\{S_{k,i}\in(t_1,t_2]\}}\mid B_k\right)=Bernoulli\left(\nu_k(t_1,t_2)\right)$ with

$$\nu_k(t_1,t_2):=F_{\mathcal{N}(B_k,\sigma_2^2)}(t_2)-F_{\mathcal{N}(B_k,\sigma_2^2)}(t_1). \tag{2.19}$$

Moreover, we may interpret P_k as the number of events in $(0,1]$ originating from a homogeneous Poisson process with intensity γ.

Assume that B_k is known. Then, due to the thinning property of the Poisson distribution (see Durrett (2012, p. 106)) $\Phi_k(t_1,t_2)$ exhibits the same distribution as the number of events in $(0,1]$ of a homogeneous Poisson process with intensity $\gamma\,\nu_k(t_1,t_2)$. Since conditional on $\mathcal{B}$ the processes $\Phi_k(t_1,t_2)$ and $\Phi_l(t_1,t_2)$, $l\neq k\in\mathbb{Z}$, are independent, we may interpret $\Phi(t_1,t_2)=\sum_{k\in\mathbb{Z}}\Phi_k(t_1,t_2)$ as the superposition of independent Poisson processes with intensities $\gamma\,\nu_k(t_1,t_2)$, $k\in\mathbb{Z}$. Thus, $\Phi(t_1,t_2)$ is conditional on $\mathcal{B}$ Poisson distributed.

Proposition 2.3 (Number of events)**.** *Let $\Phi \sim GLO(\mu, \sigma_1, \sigma_2, \gamma)$. Then, the number of events in $(t_1, t_2]$ is conditional on the background rhythm Poisson distributed,*

$$\mathcal{L}\left(\Phi(t_1,t_2) \mid \mathcal{B}\right) = Pois\left(\gamma \sum_{k\in\mathbb{Z}} \left(F_{\mathcal{N}(B_k,\sigma_2^2)}(t_2) - F_{\mathcal{N}(B_k,\sigma_2^2)}(t_1)\right)\right). \tag{2.20}$$

Proof. This follows directly from the construction of $\Phi(t_1, t_2)$ given above and the superposition property of homogeneous Poisson processes (see Durrett (2012, p.108)).

As Bingmer (2012, p.34) points out, this property also follows from the representation of the GLO process as a Cox process by conditioning the random intensity measure Λ on the beats $\mathcal{B}$ (cf. Proposition 2.1). □

According to Proposition 2.3 the expected value and variance of $\Phi(t_1, t_2)$ conditional on the beats is given by $\gamma \sum_{k\in\mathbb{Z}} \nu_k(t_1, t_2)$. However, the summands $\Phi_k(t_1, t_2)$ are unconditionally not independent any more and, thus, $\Phi(t_1, t_2)$ is unconditionally not Poisson distributed and may exhibit a different expectation and variance, i.e. $\mathbb{E}[\Phi(t_1, t_2)] \neq \text{var}(\Phi(t_1, t_2))$.

To give an intuition about the unconditional moments, we will in the following calculate the expected number of events conditional on N beats, $\mathbb{E}[\Phi(t_1, t_2) \mid B_1, ..., B_N]$.

Lemma 2.1 (Expected number of events)**.** *Let $\Phi \sim GLO(\mu, \sigma_1, \sigma_2, \gamma)$. Then, the expected number of events in $(t_1, t_2]$ conditional on $B_1, ..., B_N$ is given by*

$$\begin{aligned}
\mathbb{E}[\Phi(t_1,t_2) \mid B_1,...,B_N] &= \gamma \sum_{k=1}^{N} \left(F_{\mathcal{N}(B_k,\sigma_2^2)}(t_2) - F_{\mathcal{N}(B_k,\sigma_2^2)}(t_1)\right) \\
&+ \gamma \sum_{k\in\{-1,-2,...\}} \left(F_{\mathcal{N}(B_1+k\mu,\sigma_2^2+|k|\sigma_1^2)}(t_2) - F_{\mathcal{N}(B_1+k\mu,\sigma_2^2+|k|\sigma_1^2)}(t_1)\right) \\
&+ \gamma \sum_{k\in\{1,2,...\}} \left(F_{\mathcal{N}(B_N+k\mu,\sigma_2^2+|k|\sigma_1^2)}(t_2) - F_{\mathcal{N}(B_N+k\mu,\sigma_2^2+|k|\sigma_1^2)}(t_1)\right).
\end{aligned} \tag{2.21}$$

Proof. We employ the theorem of monotone convergence (cf. Klenke (2006, p.91)), the independence between P_k and $S_{k,1}, ..., S_{k,P_k}$ (conditional on B_k) and Wald's theorem (cf. Klenke (2006, p.99)), which yields

$$\begin{aligned}
&\mathbb{E}[\Phi(t_1,t_2) \mid B_1,...,B_N] \\
&= \mathbb{E}\left[\sum_{k\in\mathbb{Z}} \sum_{i=1}^{P_k} \mathbb{1}_{\{S_{k,i}\in(t_1,t_2]\}} \mid B_1,...,B_N\right] && (2.22) \\
&= \sum_{k\in\mathbb{Z}} \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{P_k} \mathbb{1}_{\{S_{k,i}\in(t_1,t_2]\}} \mid B_1,...,B_N,B_k\right] \mid B_1,...,B_N\right] && (2.23) \\
&= \sum_{k\in\mathbb{Z}} \mathbb{E}\left[P_k\right] \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{\{S_{k,1}\in(t_1,t_2]\}} \mid B_1,...,B_N,B_k\right] \mid B_1,...,B_N\right]. && (2.24)
\end{aligned}$$

Therefore,

$$\mathbb{E}[\Phi(t_1,t_2) \mid B_1,...,B_N] = \gamma \sum_{k=1}^{N} \left(F_{\mathcal{N}(B_k,\sigma_2^2)}(t_2) - F_{\mathcal{N}(B_k,\sigma_2^2)}(t_1) \right) \tag{2.25}$$
$$+\gamma \sum_{k\in\mathbb{Z}\setminus\{1,...,N\}} \mathbb{E}\left[\int_{t_1}^{t_2} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2\sigma_2^2}(B_k-x)^2} dx \mid B_1,...,B_N \right].$$

Now, we will use the random walk structure of the background rhythm, which gives

$$\mathcal{L}(B_k \mid B_1,...,B_N) = \mathcal{N}\left(B_1 + (k-1)\mu, |k-1|\sigma_1^2\right), \text{ for } k<1, \tag{2.26}$$
$$\text{and } \mathcal{L}(B_k \mid B_1,...,B_N) = \mathcal{N}\left(B_N + (k-N)\mu, |k-N|\sigma_1^2\right), \text{ for } k>N, \tag{2.27}$$

since the random walk is Markovian and its increments are iid $\mathcal{N}(\mu,\sigma_1^2)$, i.e.

$$B_k = B_1 - \underbrace{(B_1 - B_0)}_{\sim\mathcal{N}(\mu,\sigma_1^2)} - \underbrace{(B_0 - B_{-1})}_{\sim\mathcal{N}(\mu,\sigma_1^2)} - .... - \underbrace{(B_{k+1} - B_k)}_{\sim\mathcal{N}(\mu,\sigma_1^2)}, \text{ for } k<1, \tag{2.28}$$
$$\text{and } B_k = B_N + \underbrace{B_{N+1} - B_N}_{\sim\mathcal{N}(\mu,\sigma_1^2)} + \underbrace{B_{N+2} - B_{N+1}}_{\sim\mathcal{N}(\mu,\sigma_1^2)} ... + \underbrace{B_k - B_{k-1}}_{\sim\mathcal{N}(\mu,\sigma_1^2)}, \text{ for } k>N. \tag{2.29}$$

Let $k<1$. As a next step we can identify the distribution of $\left(\frac{B_k-x}{\sqrt{2}\sigma_2}\right)^2$ conditional on B_1: It is known (cf. Tiku (2004)) that for $Y \sim \mathcal{N}(\sqrt{\delta},1)$ we have $Y^2 \sim \chi_1^2(\delta)$, which is a non-central chi-squared distribution with one degree of freedom and non-centrality parameter δ. Thus,

$$Y_k^2 := \left(\frac{\frac{B_k-x}{\sqrt{2}\sigma_2}}{\sqrt{\text{var}\left(\frac{B_k-x}{\sqrt{2}\sigma_2}\right)}} \right)^2 = \left(\frac{\frac{B_k-x}{\sqrt{2}\sigma_2}}{\sqrt{\frac{|k-1|\sigma_1^2}{2\sigma_2^2}}} \right)^2 = \left(\frac{B_k-x}{\sqrt{|k-1|\sigma_1^2}} \right)^2 \sim \chi_1^2(\delta_k) \tag{2.30}$$

with non-centrality parameter

$$\delta_k = (\mathbb{E}[Y_k])^2 = \frac{(B_1 + (k-1)\mu - x)^2}{|k-1|\sigma_1^2}. \tag{2.31}$$

Moreover, we may apply the fact that for a random variable $Y^2 \sim \chi_1^2(\delta)$ the moment generating function is given by (cf. Tiku (2004))

$$\mathbb{E}\left[e^{uY^2}\right] = \frac{1}{\sqrt{1-2u}} e^{\frac{\delta u}{1-2u}}, \quad \text{for } u<1/2. \tag{2.32}$$

Thus,

$$\mathbb{E}\left[e^{-\left(\frac{B_k-x}{\sqrt{2}\sigma_2}\right)^2} \mid B_1\right] = \mathbb{E}\left[e^{-\frac{|k-1|\sigma_1^2}{2\sigma_2^2}Y_k^2} \mid B_1\right] = \frac{1}{\sqrt{1+2\frac{|k-1|\sigma_1^2}{2\sigma_2^2}}} \exp\left[-\frac{\delta_k \frac{|k-1|\sigma_1^2}{2\sigma_2^2}}{1+2\frac{|k-1|\sigma_1^2}{2\sigma_2^2}}\right] \tag{2.33}$$

$$= \frac{\sigma_2}{\sqrt{\sigma_2^2+|k-1|\sigma_1^2}} \exp\left[-\frac{(B_1+(k-1)\mu-x)^2}{2(\sigma_2^2+|k-1|\sigma_1^2)}\right] \tag{2.34}$$

and we obtain with the dominated convergence theorem (Klenke (2006, p.170))

$$\mathbb{E}\left[\int_{t_1}^{t_2} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2\sigma_2^2}(B_k-x)^2} dx \mid B_1\right] = \int_{t_1}^{t_2} \frac{1}{\sqrt{2\pi}\sigma_2} \mathbb{E}\left[e^{-\left(\frac{B_k-x}{\sqrt{2}\sigma_2}\right)^2} \mid B_1\right] dx \tag{2.35}$$

$$= \frac{\sigma_2}{\sqrt{\sigma_2^2+|k-1|\sigma_1^2}} \int_{t_1}^{t_2} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(B_1+(k-1)\mu-x)^2}{2(\sigma_2^2+|k-1|\sigma_1^2)}} dx \tag{2.36}$$

$$= \int_{t_1}^{t_2} \frac{1}{\sqrt{2\pi(\sigma_2^2+|k-1|\sigma_1^2)}} e^{-\frac{(B_1+(k-1)\mu-x)^2}{2(\sigma_2^2+|k-1|\sigma_1^2)}} dx \tag{2.37}$$

$$= F_{\mathcal{N}(B_1+(k-1)\mu,\sigma_2^2+|k-1|\sigma_1^2)}(t_2) - F_{\mathcal{N}(B_1+(k-1)\mu,\sigma_2^2+|k-1|\sigma_1^2)}(t_1). \tag{2.38}$$

The calculation is analog for $k > N$. By shifting the summation index one yields the form of (2.21).

□

Remark 2.4. *The proof of Lemma 2.1 shows that conditional on one beat B_1 the probability that an event of beat B_k is placed in $(t_1, t_2]$ is given by*

$$\mathbb{P}(S_{1,i} \in (t_1, t_2] \mid B_1) = F_{\mathcal{N}(B_1+(k-1)\mu,\sigma_2^2+|k-1|\sigma_1^2)}(t_2) - F_{\mathcal{N}(B_1+(k-1)\mu,\sigma_2^2+|k-1|\sigma_1^2)}(t_1), \tag{2.39}$$

thus, $\mathcal{L}(S_{1,i} \mid B_1) = \mathcal{N}(B_1+(k-1)\mu, \sigma_2^2+|k-1|\sigma_1^2)$.

This allows for an intuitive interpretation of $\mathbb{E}[\Phi(t_1,t_2) \mid B_1, ..., B_N]$: For every beat the mean number of events at one beat is weighted by the probability that the beat places an event in $(t_1, t_2]$.

Remark 2.5 (Variance). *The second moment of $\Phi(t_1,t_2)$ conditional on $B_1,...,B_N$ is given as*

$$\mathbb{E}\left[\Phi(t_1,t_2)^2 \mid B_1,...,B_N\right] \tag{2.40}$$

$$= \mathbb{E}\left[\left(\sum_{k\in\mathbb{Z}}\sum_{i=1}^{P_k}\mathbb{1}_{\{S_{k,i}\in(t_1,t_2]\}}\right)^2 \mid B_1,...,B_N\right] \tag{2.41}$$

$$= \mathbb{E}\left[\sum_{k\in\mathbb{Z}}\left(\sum_{i=1}^{P_k}\mathbb{1}_{\{S_{k,i}\in(t_1,t_2]\}}\right)^2 \mid B_1,...,B_N\right]$$

$$+ \mathbb{E}\left[\sum_{k\in\mathbb{Z}}\sum_{l\in\mathbb{Z}\setminus\{k\}}\sum_{i=1}^{P_k}\sum_{j=1}^{P_l}\mathbb{1}_{\{S_{k,i}\in(t_1,t_2]\}}\mathbb{1}_{\{S_{l,j}\in(t_1,t_2]\}} \mid B_1,...,B_N\right] \tag{2.42}$$

$$= \sum_{k\in\mathbb{Z}}\Bigg(\underbrace{\mathbb{E}\left[\left(\sum_{i=1}^{P_k}\mathbb{1}_{\{S_{k,i}\in(t_1,t_2]\}}\right)^2 \mid B_1,...,B_N\right]}_{(A)}$$

$$+ \sum_{l\in\mathbb{Z}\setminus\{k\}}\underbrace{\mathbb{E}\left[\sum_{i=1}^{P_k}\sum_{j=1}^{P_l}\mathbb{1}_{\{S_{k,i},S_{l,j}\in(t_1,t_2]\}} \mid B_1,...,B_N\right]}_{(B)}\Bigg). \tag{2.43}$$

It is straightforward to adopt the proof of Wald's theorem given by Klenke (2006, p.99-100) for (A) and (B) since P_k, P_l and $\mathbb{1}_{\{S_{k,i},S_{l,j}\in(t_1,t_2]\}}$ are independent for $l\neq k$ and $\mathbb{1}_{\{S_{k,i_1},S_{l,j_1}\in(t_1,t_2]\}}$ and $\mathbb{1}_{\{S_{k,i_2},S_{l,j_2}\in(t_1,t_2]\}}$ are conditional on B_k and B_l independent and identically distributed for all $i_1,i_2\in\{1,...,P_k\}$, $j_1,j_2\in\{1,...,P_l\}$. Thus, for the first expectation (A) in (2.43) it follows that

$$\mathbb{E}\left[\left(\sum_{i=1}^{P_k}\mathbb{1}_{\{S_{k,i}\in(t_1,t_2]\}}\right)^2 \mid B_1,...,B_N\right] \tag{2.44}$$

$$= \gamma\mathbb{E}\left[\nu_k(t_1,t_2) \mid B_1,...,B_N\right]$$

$$+ \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{P_k}\sum_{\substack{j=1,...,P_k\\ j\neq i}}\mathbb{1}_{\{S_{k,i}\in(t_1,t_2]\}}\mathbb{1}_{\{S_{k,j}\in(t_1,t_2]\}} \mid B_1,...,B_N,B_k\right] \mid B_1,...,B_N\right] \tag{2.45}$$

$$= \gamma\mathbb{E}\left[\nu_k(t_1,t_2) \mid B_1,...,B_N\right] + \mathbb{E}\left[P_k(P_k-1)\right]\mathbb{E}\left[\nu_k(t_1,t_2)^2 \mid B_1,...,B_N\right] \tag{2.46}$$

$$= \gamma\mathbb{E}\left[\nu_k(t_1,t_2) \mid B_1,...,B_N\right] + \gamma^2\mathbb{E}\left[\nu_k(t_1,t_2)^2 \mid B_1,...,B_N\right], \tag{2.47}$$

where $\nu_k(t_1,t_2) = \mathbb{P}(S_{k,1}\in(t_1,t_2] \mid B_k) = \int_{t_1}^{t_2}\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{1}{2\sigma_2^2}(B_k-x)^2}dx$.

Furthermore, for the second expectation (B) in (2.43) it similarly follows that

$$\mathbb{E}\left[\sum_{i=1}^{P_k}\sum_{j=1}^{P_l}\mathbb{1}_{\{S_{k,i},S_{l,j}\in(t_1,t_2]\}} \mid B_1,...,B_N\right] \tag{2.48}$$

$$= \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{P_k}\sum_{j=1}^{P_l}\mathbb{1}_{\{S_{k,i},S_{l,j}\in(t_1,t_2]\}} \mid B_1,...,B_N,B_k,B_l\right] \mid B_1,...,B_N\right] \tag{2.49}$$

$$= \mathbb{E}\left[P_kP_l\right]\mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{\{S_{k,1},S_{l,1}\in(t_1,t_2]\}} \mid B_1,...,B_N,B_k,B_l\right] \mid B_1,...,B_N\right] \tag{2.50}$$

$$= \gamma^2\mathbb{E}\left[\int_{t_1}^{t_2}\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{1}{2\sigma_2^2}(B_k-x)^2}dx\int_{t_1}^{t_2}\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{1}{2\sigma_2^2}(B_l-x)^2}dx \mid B_1,...,B_N\right]. \tag{2.51}$$

However, since B_k and B_l are not independent, the integrals in Equation (2.51) are conditional on $B_1,...,B_N$ not independent for $k,l \notin \{1,...,N\}$, which makes it difficult to find an analytical solution for the second moment.

Remark 2.6 (Parameter estimation)**.** *Given $n \in \mathbb{N}$ observed events $S_1,...,S_n$ of a GLO process we would like to estimate the four parameters $\eta = (\mu,\sigma_1,\sigma_2,\gamma)$. However, we did neither observe the background rhythm $\mathcal{B}$ nor do we know which events belong to the same burst and are, thus, not able to write down the likelihood function for η. Therefore, it is not possible to compute a Maximum-Likelihood estimate for η.*

Bingmer et al. (2011) suggest to fit the theoretical to the empirical intensities. For this purpose, they show that the autocorrelation function (ACF) for the GLO is given as

$$f(l) = \gamma\sum_{k\in\mathbb{Z}} f_{\mathcal{N}(k\mu,|k|\sigma_1^2+2\sigma_2^2)}(l). \tag{2.52}$$

Note, that in contrast to the ACF for a time series the ACF for point processes corresponds to the intensity of events at time l conditioned on an event at time 0 (cf. Gerstein and Kiang (1960)), whereas the ACF for a time series is the correlation between two events with lag l (cf. Brockwell and Davis (2010) and Definition 2.5, p. 33). Since the autocorrelation histogram (ACH) (after standardizing, cf. Bingmer (2012, p.71)) is an estimate for the ACF of a point process, Bingmer et al. (2011) propose to fit the empirical ACH to the ACF by minimizing the weighted residual sum of squares.

However, this procedure exhibits several shortcomings: Firstly, and most importantly, the obtained estimates do not meet a thorough theoretical justification as, in contrast, Maximum-Likelihood estimates are consistent and asymptotically unbiased and Bayesian estimates result as subjective beliefs from Bayes' theorem. Secondly, the procedure leads to large estimation errors for large σ_2/μ or low firing rates γ. Thirdly, the uncertainty about true parameter values can only be assessed in two ways: As a first approach, the fitting routine of the R-function nls() provides estimates for the standard errors. However, these are underestimating the true variability most of the time (cf. Bingmer (2012, p.95)). Thus,

Bingmer et al. (2011) propose the use of bootstrap methods to obtain confidence intervals. Nonetheless, this is rather time consuming and exposed to additional estimation errors itself.

Furthermore, the fitting method is not able to cover extensions of the GLO model that e.g. allow for overdispersed number of events at bursts: Since the ACF of the GLO is defined for $l > 0$ as (cf. Bingmer (2012, p.67))

$$f(l) = \lim_{\delta_1,\delta_2 \to 0+} \frac{\mathbb{E}[\Phi(l, l+\delta_1) \mid \Phi(-\delta_2, 0] > 0]}{\delta_1} \tag{2.53}$$

$$= \lim_{\delta_1,\delta_2 \to 0+} \frac{\mathbb{E}[\sum_{k\in\mathbb{Z}} \sum_{i=1}^{P_k} \delta_{B_k+Z_{k,i}}(l, l+\delta_1) \mid \Phi(-\delta_2, 0] > 0]}{\delta_1}, \tag{2.54}$$

regardless of var(P_k) only the average burst size $\mathbb{E}[P_k]$ is considered in the ACF. Finally, forecasting future events, which is an important tool for risk management in general, is not possible with ACH estimates since historical beats for observed events are not estimated.

To conclude, in addition to the successful modeling of neuronal activity the GLO model exhibits many attractive features for modeling claim arrivals as well: It constructs a full point process, does generally not restrict the expected value or variance of the number of events in a given time window and is particularly suitable for seasonal events. Moreover, with increasing burst width it converges to a homogeneous Poisson process and it is straightforward to extend the model with e.g. Negative Binomially distributed number of events per burst or trends (we will suggest several extensions in section 4.4).

However, there does not exist a coherent and flexible theoretical framework for parameter estimation. Addressing this issue, in the next chapter we will review methods of Bayesian inference in order to develop a Bayesian estimation framework for the parameters of the GLO model.

2.2. Bayesian inference and conjugate priors

The main idea of Bayesian inference is to assess someone's subjective belief about unknown quantities of a model. Thus, probabilities are seen as subjective and interpreted conditional on available information. This contrasts the classical (frequentist) approach, in which probabilities are seen as limiting frequencies under infinite hypothetical replications of the situation under consideration, i.e. model parameters are considered as fixed, albeit unknown, constants (cf. Wakefield (2013, p.22-23) or Chu and Zhao (2011, p.244)).

To perform a Bayesian analysis one has to assign a *prior distribution* $\pi(\vartheta)$ to the parameter (vector) ϑ, which represents a person's belief (or knowledge) about the value of ϑ before having observed any data. When new information is obtained, the prior knowledge is revised. By combining the prior $\pi(\vartheta)$ and the likelihood $L(\vartheta; \mathcal{S}) = p(\mathcal{S} \mid \vartheta)$ of the observed data $\mathcal{S}$ the *posterior distribution* $p(\vartheta \mid \mathcal{S})$ can be calculated by employing Bayes' theorem

(cf. e.g. Hoff (2009, p.2) or Koch (2007, p.31)):

$$p(\vartheta \mid \mathcal{S}) = \frac{L(\vartheta; \mathcal{S})\, \pi(\vartheta)}{\int_\Theta L(v; \mathcal{S})\, \pi(v)\, dv}, \tag{2.55}$$

where Θ is a parameter-space for ϑ. The posterior distribution represents the updated belief about ϑ after having observed the data $\mathcal{S}$. Since the normalizing constant $\int_\Theta L(v; \mathcal{S})\, \pi(v)\, dv$ does not contain any information about ϑ, it is frequently omitted, in which case Bayes' theorem is represented by

$$p(\vartheta \mid \mathcal{S}) \propto L(\vartheta; \mathcal{S})\, \pi(\vartheta), \tag{2.56}$$

where $\propto$ denotes proportionality. A justification for the use of Bayes' theorem can be found in the results of Cox (1946), Cox (1961) or Savage (1972), who show that it provides an optimal method for updating a person's belief about ϑ given the new information $\mathcal{S}$.

To obtain analytically tractable posterior distributions, the concept of *conjugancy* is immensely helpful:

Definition 2.1 (Conjugate)**.** *(Hoff (2009, p.38)) A class $\mathcal{P}$ of prior distributions for ϑ is called conjugate for a sampling model $p(\mathcal{S} \mid \vartheta)$ if*

$$\pi(\vartheta) \in \mathcal{P} \;\Rightarrow\; p(\vartheta \mid \mathcal{S}) \in \mathcal{P}. \tag{2.57}$$

Thus, if one chooses a conjugate prior, the posterior distribution will originate from the same distributional family. Example 1 shows that a Gamma prior for the precision ψ (which is the inverse of the variance σ^2) and a Normal prior for the mean μ of a Normal distribution are its conjugate priors. In this work we parametrize the Gamma distribution in terms of its shape $\alpha > 0$ and rate $\beta > 0$, such that the expected value is α/β and the pdf is given by

$$f_{Ga(\alpha,\beta)}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad \text{for } x > 0, \tag{2.58}$$

where $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ is the Gamma function.

Note that in calculations as well as algorithms we mostly parametrize the Normal distribution in terms of its precision $\psi = \sigma^{-2}$, particularly to exploit conjugacy. However, we use σ^2 and ψ^{-1} interchangeably and report priors, posterior distributions and estimates in terms of the variance or standard deviation, since these are generally more intuitive to interpret and often desired to estimate. Because precision and variance are deterministically linked, i.e. $\psi = 1/\sigma^2$, samples can be transformed easily. Moreover, the inverse-Gamma distribution provides a straightforward link between the priors for ψ and σ^2, i.e. if $\psi \sim Ga(\alpha, \beta)$, then σ^2 is inverse-Gamma distributed, $\sigma^2 = \psi^{-1} \sim IGa(\alpha, \beta)$, with pdf

$$f_{IGa(\alpha,\beta)}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left[-\frac{\beta}{\alpha} \right], \quad \text{for } x > 0. \tag{2.59}$$

Example 1 (Joint inference for the normal distribution)**.** *(cf. Hoff (2009, p.73-75))*
Let $\mathcal{S} = (S_1, ..., S_n)$ be $n \in \mathbb{N}$ independent, identically and normally distributed observations, i.e. $S_1, ..., S_n \overset{iid}{\sim} \mathcal{N}(\mu, \psi^{-1})$, and suppose that a priori ψ and μ are independent and the prior distributions are $\mu \sim \mathcal{N}\left(\mu_\mu, \sigma_\mu^2\right)$ and $\psi \sim Ga(\alpha, \beta)$. Then, by Bayes' theorem the joint posterior distribution is proportional to

$$p(\mu, \psi \mid \mathcal{S}) \tag{2.60}$$

$$\propto f_{\mathcal{N}(\mu_\mu, \sigma_\mu^2)}(\mu) \times f_{Ga(\alpha,\beta)}(\psi) \times \prod_{i=1}^{n} f_{\mathcal{N}(\mu, \psi^{-1})}(S_i) \tag{2.61}$$

$$\propto \psi^{n/2+\alpha-1} \exp\left[-\frac{1}{2\sigma_\mu^2}(\mu - \mu_\mu)^2 - \beta\psi - \frac{\psi}{2}\sum_{i=1}^{n}(S_i - \mu)^2\right]. \tag{2.62}$$

One can easily identify the full conditional posterior densities: For μ one yields

$$p(\mu \mid \psi, \mathcal{S}) \propto p(\mu, \psi \mid \mathcal{S}) \propto \exp\left[-\frac{1}{2\sigma_\mu^2}(\mu - \mu_\mu)^2 - \frac{\psi}{2}\sum_{i=1}^{n}(S_i - \mu)^2\right] \tag{2.63}$$

$$\propto \exp\left[-\frac{\psi}{2}\left(\sum_{i=1}^{n}(S_i - \overline{\mathcal{S}})^2 + n(\overline{\mathcal{S}} - \mu)^2\right) - \frac{1}{2\sigma_\mu^2}(\mu - \mu_\mu)^2\right] \tag{2.64}$$

$$\propto \exp\left[-\frac{1}{2}\left(\mu^2\left(n\psi + \frac{1}{\sigma_\mu^2}\right) - 2\mu\left(n\overline{\mathcal{S}}\psi + \frac{\mu_\mu}{\sigma_\mu^2}\right)\right)\right] \tag{2.65}$$

$$\propto \exp\left[-\frac{1}{2}\left(\mu^2 a - 2\mu b\right)\right] \tag{2.66}$$

$$\propto \exp\left[-\frac{a}{2}\left(\mu^2 - 2\mu\frac{b}{a}\right)\right] \propto \exp\left[-\frac{a}{2}\left(\mu - \frac{b}{a}\right)^2\right], \tag{2.67}$$

which is proportional to the density of a normal distribution with variance $1/a$,

$$\frac{1}{a} := \frac{1}{n\psi + \frac{1}{\sigma_\mu^2}} = \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma_\mu^2}} = \frac{\sigma^2\sigma_\mu^2}{n\sigma_\mu^2 + \sigma^2}, \tag{2.68}$$

and expected value b/a,

$$\frac{b}{a} := \frac{n\overline{\mathcal{S}}\psi + \frac{\mu_\mu}{\sigma_\mu^2}}{a} = \left(\frac{n\overline{\mathcal{S}}}{\sigma^2} + \frac{\mu_\mu}{\sigma_\mu^2}\right)\frac{\sigma^2\sigma_\mu^2}{n\sigma_\mu^2 + \sigma^2} = \frac{n\sigma_\mu^2}{n\sigma_\mu^2 + \sigma^2}\overline{\mathcal{S}} + \frac{\sigma^2}{n\sigma_\mu^2 + \sigma^2}\mu_\mu. \tag{2.69}$$

Thus, the conditional posterior expectation of the parameter μ is a weighted average of the sample mean $\overline{\mathcal{S}}$ and the prior expectation μ_μ, whereas $\overline{\mathcal{S}}$ gains more weight with increasing number of observations.

The full conditional posterior distribution of ψ is proportional to

$$\begin{aligned} p(\psi \mid \mu, \mathcal{S}) &\propto p(\psi, \mu \mid \mathcal{S}) \\ &\propto \psi^{\alpha+n/2-1} \exp\left[-\frac{1}{2\sigma_\mu^2}(\mu-\mu_\mu)^2 - \beta\psi - \frac{\psi}{2}\sum_{i=1}^{n}(S_i-\mu)^2\right] \\ &\propto \psi^{\alpha+n/2-1} \exp\left[-\psi\left(\beta + \frac{1}{2}\sum_{i=1}^{n}(S_i-\mu)^2\right)\right], \end{aligned} \tag{2.70}$$

which is proportional to the density of a $Ga\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum_{i=1}^{n}(S_i-\mu)^2\right)$ *distribution. The conditional posterior expectation is a weighted average analogously to that of* μ*:*

$$\mathbb{E}[\psi \mid \mu, \mathcal{S}] = \frac{\alpha + \frac{n}{2}}{\beta + \frac{1}{2}\sum_{i=1}^{n}(S_i-\mu)^2} \tag{2.71}$$

$$= \frac{\beta}{\beta + \frac{1}{2}\sum_{i=1}^{n}(S_i-\mu)^2}\,\frac{\alpha}{\beta} + \frac{\frac{1}{2}\sum_{i=1}^{n}(S_i-\mu)^2}{\beta + \frac{1}{2}\sum_{i=1}^{n}(S_i-\mu)^2}\,\frac{n}{\sum_{i=1}^{n}(S_i-\mu)^2}. \tag{2.72}$$

However, the full conditional posterior distribution of μ *depends on* ψ *and vice versa, thus, one has to compute the marginal posterior distributions to obtain posterior knowledge about the single parameters itself. These are given by*

$$p(\mu \mid \mathcal{S}) = \int_0^\infty p(\mu, \psi \mid \mathcal{S})\, d\psi \tag{2.73}$$

$$\text{and } p(\psi \mid \mathcal{S}) = \int_{-\infty}^\infty p(\mu, \psi \mid \mathcal{S})\, d\mu. \tag{2.74}$$

An alternative way to gain posterior information about single parameters without solving the integrals in (2.73) and (2.74) analytically but drawing samples from $p(\mu \mid \psi, \mathcal{S})$ *and* $p(\psi \mid \mu, \mathcal{S})$ *is Gibbs sampling, which will be described in section 2.3.*

Remark 2.7 (Consistency)**.** *Example 1 shows that for Bayesian inference for a Normal distribution the conditional posterior distribution of* μ *shrinks towards* $\overline{\mathcal{S}}$ *with increasing sample size: The conditional posterior mean converges to* $\overline{\mathcal{S}}$ *and the variance converges to zero. This phenomenon is consistent with the Bernstein-von Mises theorem which states that in the limit of infinite trials the posterior distribution converges independently of the initial prior to a Gaussian distribution around the true parameter.*

This theorem is sometimes referred to as consistency of Bayesian estimates (cf. Wakefield (2013, p.90)). However, the first proof of the theorem was only given in finite probability spaces (cf. Doob (1949)) and, moreover, the theorem is proven to be almost surely wrong in infinite probability spaces (cf. Freedman (1965)). Thus, prior distributions do generally matter, even in the case of many observations.

2.3. MCMC Methods

Often the joint posterior distribution $p(\vartheta \mid \mathcal{S})$ is not a standard distribution and, thus, difficult to handle. If at least $p(\vartheta \mid \mathcal{S})$ were a distribution one could take samples from, Monte Carlo approximations for posterior quantities such as the posterior mean, variance or credible intervals could be computed. However, many, particularly multidimensional, posterior distributions are not easy to sample from.

In these situations, it may be easier to employ Markov chain Monte Carlo (MCMC) methods. Essentially, a MCMC algorithm constructs a Markov chain with equilibrium distribution $p(\vartheta \mid \mathcal{S})$. If the chain converges, it is convenient to discard the first BI iterations of the chain (the so called *burn-in phase*) until convergence is reached and to consider the remaining samples as samples from $p(\vartheta \mid \mathcal{S})$ (cf. Koch (2007, p.218) and Roberts and Rosenthal (2004, p.22)). In this work the length of the burn-in phase BI will be identified by visual inspection, examples are given in section 2.3.4. In the following we will shortly review two frequently used MCMC-methods, which will also be employed in the Bayesian inference framework for the GLO described in chapter 3.

2.3.1. Gibbs sampling

The Gibbs sampler was introduced by Geman and Geman (1984) in the context of Bayesian image analysis. Its main idea is to decompose the parameter space by sampling from the full conditional distributions of the posterior distribution:

Definition 2.2 (Full Conditional Posterior Distribution)**.** *For the k-th component of a parameter vector $\vartheta = (\vartheta_1, ..., \vartheta_d)$, $k \in \{1, ..., d\}$, the full conditional posterior distribution (FCPD) is the posterior distribution (density) of the k-th component given everything else and can be computed with Bayes' theorem as follows:*

$$p(\vartheta_k \mid \vartheta_{-k}, \mathcal{S}) = \frac{p(\vartheta_1, ..., \vartheta_d \mid \mathcal{S})}{p(\vartheta_1, ..., \vartheta_{k-1}, \vartheta_{k+1}, ..., \vartheta_d \mid \mathcal{S})} \tag{2.75}$$

$$\propto p(\vartheta_1, ..., \vartheta_d \mid \mathcal{S}), \tag{2.76}$$

where $\vartheta_{-k} := (\vartheta_1, \vartheta_2, ..., \vartheta_{k-1}, \vartheta_{k+1}, ..., \vartheta_d)$.

Since the denominator on the right-hand side of (2.75) corresponds to a normalizing constant, the FCPD for ϑ_k is found such that in the joint posterior density $p(\vartheta_1, ..., \vartheta_d \mid \mathcal{S})$ only the component ϑ_k is considered being variable (cf. Koch (2007, p.217-218)). In Example 1 we already obtained the FCPDs for μ and ψ in a similar way.

Algorithm 2.1 (Gibbs sampler)**.** *(Koch (2007, p.218)) The Gibbs sampler begins with arbitrary starting values* $\vartheta_1^{(0)}, ..., \vartheta_d^{(0)}$*. Then random values are sequentially drawn from the FCPDs of* $\vartheta_1, ..., \vartheta_d$ *to complete one iteration. Hence, in the* $(r+1)$*-th iteration one draws*

$$\begin{aligned}
&\vartheta_1^{(r+1)} \quad \textit{from} \quad p(\vartheta_1 \mid \vartheta_2^{(r)}, ..., \vartheta_d^{(r)}, \mathcal{S}),\\
&\vartheta_2^{(r+1)} \quad \textit{from} \quad p(\vartheta_2 \mid \vartheta_1^{(r+1)}, \vartheta_3^{(r)}, ..., \vartheta_d^{(r)}, \mathcal{S}),\\
&...\\
&\vartheta_d^{(r+1)} \quad \textit{from} \quad p(\vartheta_d \mid \vartheta_1^{(r+1)}, \vartheta_2^{(r+1)}, ..., \vartheta_{d-1}^{(r+1)}, \mathcal{S}).
\end{aligned}$$

In many situations it is possible to draw samples from FCPDs, particularly if conjugate priors are used. However, once that is not possible for one or more components of ϑ, one can not employ the Gibbs sampling algorithm. In these cases, the Metropolis-Hastings (MH) algorithm represents an alternative way to sample from the posterior distribution.

2.3.2. The Metropolis-Hastings algorithm

Suppose that we have already obtained sample values $\left\{\vartheta^{(1)}, ..., \vartheta^{(r)}\right\}$ from the posterior distribution $p(\vartheta \mid \mathcal{S})$ and would like to add a new value $\vartheta^{(r+1)} \in \left\{\vartheta^{(r)}, \vartheta'\right\}$ for some $\vartheta' \in \Theta$. Should we add ϑ' or another instance of $\vartheta^{(r)}$? Intuitively, if $p(\vartheta' \mid \mathcal{S}) > p(\vartheta^{(r)} \mid \mathcal{S})$, we would want more instances of ϑ' than of $\vartheta^{(r)}$ in the sample set, and vice versa. Thus, it is reasonable to require the ratio of the number of instances of ϑ' relative to the number of instances of $\vartheta^{(r)}$ to match the ratio of the posterior distributions, which by Bayes' theorem is equal to

$$z = \frac{p(\vartheta' \mid \mathcal{S})}{p(\vartheta^{(r)} \mid \mathcal{S})} = \frac{p(\mathcal{S} \mid \vartheta')\pi(\vartheta')}{p(\mathcal{S})} \frac{p(\mathcal{S})}{p(\mathcal{S} \mid \vartheta^{(r)})\pi(\vartheta^{(r)})} = \frac{p(\mathcal{S} \mid \vartheta')\pi(\vartheta')}{p(\mathcal{S} \mid \vartheta^{(r)})\pi(\vartheta^{(r)})}. \tag{2.77}$$

Note that the normalizing constants do not appear on the right-hand side of equation (2.77). Thus, it is sufficient to calculate posterior densities only proportional to constants.

Additionally, the Metropolis-Hastings (MH) algorithm, which was first introduced by Metropolis et al. (1953) and subsequently generalized by Hastings (1970), also takes into account the proposal distribution $q\left(\vartheta \mid \vartheta^{(r)}\right)$ for ϑ' given $\vartheta^{(r)}$:

Algorithm 2.2 (Metropolis-Hastings)**.** *(Wakefield (2013, p.123))*
In the $(r+1)$*-th iteration*

1) propose a new value $\vartheta' \sim q\left(\vartheta \mid \vartheta^{(r)}\right)$*,*

2) accept this value, i.e. set $\vartheta^{(r+1)} = \vartheta'$*, with probability*

$$p_{acc} = \min\left\{1, z \times \frac{q(\vartheta^{(r)} \mid \vartheta')}{q(\vartheta' \mid \vartheta^{(r)})}\right\} = \min\left\{1, \frac{p(\vartheta' \mid \mathcal{S})}{p(\vartheta^{(r)} \mid \mathcal{S})} \times \frac{q(\vartheta^{(r)} \mid \vartheta')}{q(\vartheta' \mid \vartheta^{(r)})}\right\}, \tag{2.78}$$

and set $\vartheta^{(r+1)} = \vartheta^{(r)}$*, otherwise.*

An important special case of proposal distributions are *symmetric proposal distributions*: If the proposal distribution satisfies $q(x \mid y) = q(y \mid x)$ for all $x, y \in \Theta$, the acceptance probability (2.78) reduces to $p_{acc} = \min\{1, z\}$. In this case the algorithm is simply called *Metropolis algorithm* (cf. Wakefield (2013, p.123)).

The MH algorithm represents a rather simple method to sample from a posterior distribution. However, it also exhibits one major drawback in comparison to the Gibbs sampler since the proposal distribution has to be chosen: Clearly, if the proposal distribution is far off the posterior distribution, the simulated Markov chain tends to stay in the same realizations for long times and, thus, maintains very slow convergence.

By applying MCMC algorithms one expects that the distribution of the random vector $\vartheta^{(r)}$ converges to the target distribution $p(\vartheta \mid \mathcal{S})$ if $r \to \infty$, which is indeed true under certain conditions (cf. Geman and Geman (1984) and Roberts and Smith (1994), more references can be found in Hoff (2009, p.104)). However, assessing the convergence speed and accuracy in general is difficult and, thus, one often relies on methods of MCMC diagnostics instead. Several convergence conditions and diagnostic methods are described in the next two sections.

2.3.3. MCMC convergence

Clearly, there may be two major drawbacks when employing Markov chain Monte Carlo methods: slow convergence and no convergence at all. The latter issue may be addressed by the theory of Markov chains and their stationary distributions. Therefore, in this section we will firstly summarize some basic results about (finite) Markov chains in discrete time and state space and show that in this case Gibbs and MH algorithm both produce Markov chains with $p(\vartheta \mid \mathcal{S})$ as their stationary distribution. A more detailed introduction to Markov chains may be found in Durrett (2012) or Häggström (2002). Secondly, we will present the convergence conditions given by Roberts and Smith (1994) which hold for arbitrary state spaces. However, for the sake of simplicity we will avoid any measure-theoretical details arising with the non-discrete case and refer to Roberts and Smith (1994) and Roberts and Rosenthal (2004) for further details.

Definition 2.3 (Markov chain)**.** *(Durrett (2012, p.2))* *$(\vartheta^{(0)}, \vartheta^{(1)}, ...)$ is a discrete time Markov chain with state space Θ and transition matrix P if for any $y, x, x_{r-1}, ..., x_0 \in \Theta$*

$$\mathbb{P}(\vartheta^{(r+1)} = y \mid \vartheta^{(r)} = x, \vartheta^{(r-1)} = x_{r-1}, ..., \vartheta^{(0)} = x_0) = P(x, y). \tag{2.79}$$

It is straightforward to show that the two MCMC methods presented in section 2.3 construct a Markov chain since $\vartheta^{(r+1)}$ depends only on $\vartheta^{(r)}$: For the MH algorithm a move from state $x \in \Theta$ to $y \in \Theta$ is proposed with probability $q(y \mid x)$ and is accepted with probability $p_{acc}(y \mid x) = \min\left\{1, \frac{\tilde{p}(y)}{\tilde{p}(x)} \times \frac{q(x|y)}{q(y|x)}\right\}$, where $\tilde{p}(\vartheta) := p(\vartheta \mid \mathcal{S})$. Thus, the transition probability is given by

$$P(x, y) = q(y \mid x) p_{acc}(y \mid x). \tag{2.80}$$

In the case of the Gibbs sampler updating dimension k we have $y = (y_1, ..., y_d)$ with $y_i = x_i$ for all $i \in \{1, ..., d\} \backslash \{k\}$. We can think of y_k as being proposed according to the FCPD $q(y_k \mid x) = \tilde{p}\,(y_k \mid x_{-k})$ and accepted with the corresponding MH acceptance probability $p_{acc}(y \mid x)$, which is

$$\frac{\tilde{p}(y)}{\tilde{p}(x)} \frac{q(x \mid y)}{q(y \mid x)} = \frac{\tilde{p}(y_k \mid y_{-k})\tilde{p}(y_{-k})}{\tilde{p}(x_k \mid x_{-k})\tilde{p}(x_{-k})} \frac{\tilde{p}(x_k \mid y_{-k})}{\tilde{p}(y_k \mid x_{-k})} = \frac{\tilde{p}(y_k \mid x_{-k})\tilde{p}(x_{-k})}{\tilde{p}(x_k \mid x_{-k})\tilde{p}(x_{-k})} \frac{\tilde{p}(x_k \mid x_{-k})}{\tilde{p}(y_k \mid x_{-k})} = 1. \tag{2.81}$$

Therefore, the Gibbs sampler is essentially a MH algorithm with special proposal distribution and it is sufficient to develop convergence criteria for the MH algorithm. Moreover, this justifies to use both, Gibbs and MH updating steps, in one MCMC algorithm, e.g. a MH step to update dimension l and a Gibbs step to update dimension k.

Definition 2.4 (Irreducibility and aperiodicity)**.** *(Durrett (2012, p.17, 29)) A Markov chain $(\vartheta^{(0)}, \vartheta^{(1)}, ...)$ with state space Θ and transition matrix P is*

a) ***irreducible*** *if for all $x, y \in \Theta$ there exist $n \in \mathbb{N}$ such that $P^n(x, y) > 0$,*

b) ***aperiodic*** *if all states $x \in \Theta$ have period 1, i.e. the greatest common divisor is 1.*

Lemma 2.2. *(Durrett (2012, p.29)) $\left(\vartheta^{(0)}, \vartheta^{(1)}, ...\right)$ is aperiodic, if $P(x, x) > 0$ for all $x \in \Theta$.*

Theorem 2.1 (Convergence Theorem)**.** *(Durrett (2012, p.31)) A distribution $\tilde{p}$ is defined to be* ***stationary*** *for a Markov chain with transition matrix P, if*

$$\sum_{x \in \Theta} \tilde{p}(x) P(x, y) = \tilde{p}(y). \tag{2.82}$$

Let $(\vartheta^{(0)}, \vartheta^{(1)}, ...)$ be an irreducible, aperiodic Markov chain with transition matrix P and stationary distribution $\tilde{p}(\vartheta)$. Then as $n \to \infty$, $P^n(\vartheta^{(0)}, y) \to \tilde{p}(y)$.

If $p(\vartheta \mid \mathcal{S})$ is discrete, it is straightforward to show that the MH algorithm constructs a Markov chain with stationary distribution $p(\vartheta \mid \mathcal{S})$:

Lemma 2.3 (Stationary distribution of the MH algorithm)**.** *Let $(\vartheta^{(0)}, \vartheta^{(1)}, ...,)$ be a Markov chain that is constructed by the MH algorithm (Algorithm 2.2). Then $p(\vartheta \mid \mathcal{S})$ is stationary for the chain.*

Proof. We show that the MH algorithm constructs a Markov chain that is reversible with respect to $\tilde{p}(x) = p(\vartheta \mid \mathcal{S})$, i.e. that $\tilde{p}(x)P(x, y) = \tilde{p}(y)P(y, x)$:

Suppose, that $\tilde{p}(y)q(x \mid y) \leq \tilde{p}(x)q(y \mid x)$. Then,

$$\tilde{p}(x)P(x, y) = \tilde{p}(x)\, q(y \mid x)\, p_{acc}(y \mid x) = \tilde{p}(x)\, q(y \mid x) \times \frac{\tilde{p}(y)}{\tilde{p}(x)} \times \frac{q(x \mid y)}{q(y \mid x)} \tag{2.83}$$

$$= \tilde{p}(y)\, q(x \mid y) \tag{2.84}$$

$$= \tilde{p}(y)\, q(x \mid y)\, p_{acc}(x \mid y) = \tilde{p}(y)P(y, x). \tag{2.85}$$

Due to symmetry, the calculation is analog for $\tilde{p}(y)q(x \mid y) > \tilde{p}(x)q(y \mid x)$. Therefore,

$$\sum_{x \in \Theta} \tilde{p}(x)P(x,y) = \sum_{x \in \Theta} \tilde{p}(y)P(y,x) = \tilde{p}(y) \sum_{x \in \Theta} P(y,x) = \tilde{p}(y). \tag{2.86}$$

□

According to Lemma 2.3 it is sufficient to show that a particular MH algorithm is aperiodic and $\tilde{p}$-irreducible in order to obtain convergence. This will be done by the next two theorems for both the MH algorithm and Gibbs sampler, that also apply to the general (non-discrete) case.

Theorem 2.2 (Convergence of the MH algorithm)**.** *(Roberts and Smith (1994, p.215))*

i) If q is aperiodic or $\mathbb{P}\left(\vartheta^{(r)} = \vartheta^{(r-1)}\right) > 0$ for some $r \geq 1$, then the MH algorithm is aperiodic.

ii) If q is $\tilde{p}$-irreducible and $q(y \mid x) = 0$ if, and only if, $q(x \mid y) = 0$, then the MH algorithm is $\tilde{p}$-irreducible.

If the MH algorithm is aperiodic and $\tilde{p}$-irreducible with stationary distribution $\tilde{p}$, then the Markov chain's distribution converges to $\tilde{p}$.

Theorem 2.3 (Convergence of the Gibbs sampler)**.** *(Roberts and Smith (1994, p.213)) Let $\tilde{p}(\vartheta)$ be a mgf or pdf with respect to the d-dimensional Lebesgue-measure which is l.s.c at 0, i.e. for all $\vartheta_0 \in \Theta$ with $\tilde{p}(\vartheta_0) > 0$ there exists an open neighborhood $N_{\vartheta_0} \ni \vartheta_0$ and $\varepsilon > 0$ such that for all $\vartheta \in N_{\vartheta_0}$ we have $\tilde{p}(\vartheta) \geq \varepsilon > 0$.*

Furthermore, let $\int \tilde{p}(\vartheta)\, d\vartheta_i$ be locally bounded for $i = 1, ..., d$ and Θ be connected. Then, the Markov chain's distribution converges to $\tilde{p}$.

Remark 2.8. *The conditions from Theorem 2.3 are comparably weak: With $\tilde{p}$ being l.s.c at 0 the transition kernel of the Gibbs sampler (the FCPD) is well-defined, since it implies that $\int \tilde{p}(\vartheta)\, d\vartheta_i > 0$ (cf. Definition 2.2). Moreover, together with locally bounded $\int \tilde{p}(x)\, dx_i$ aperiodicity is established (cf. Roberts and Smith (1994, p.212)). The chain is irreducible if, additionally, Θ is connected (cf. Roberts and Smith (1994, p.213)).*

Neither Theorem 2.2 nor Theorem 2.3 include any assumptions about the initial state $\vartheta^{(0)}$ in order to guarantee convergence. Nonetheless, it may take a lot of time for the chain to converge if $\vartheta^{(0)}$ lies in a region $A \subset \Theta$ with small probability mass $\tilde{p}(A)$. Put differently, until now we have only inferred conditions under which the distribution of Gibbs or MH samples converges to the posterior distribution, however, the respective convergence rate may be intolerably small.

Addressing this issue, Gelman et al. (1996) obtain convergence rates and efficient proposal distributions for the Metropolis algorithm for simulating from a normal distribution. Their results suggest to calibrate the proposal distribution q such that the average acceptance rate of the algorithm is roughly 1/4. However, in case of much more complicated target densities there exist, to the best of my knowledge, no general rules how to choose q and, thus, it is easier to rely on visual diagnostics. For this purpose we will outline some possible ways how to identify and avoid slow convergence in the following section.

2.3.4. MCMC diagnostics

While introducing the MH algorithm in section 2.3.2 we already noticed that slow convergence may result from employing unsuitable proposal distributions, in particular if proposed values lie in regions with small probability mass of $p(\vartheta \mid \mathcal{S})$. In this case, the MH algorithm tends to get "stuck". To give an example we will show how both Gibbs and MH algorithm may be implemented in the setting of Example 1. The results will highlight the difference between good and bad *mixing behavior* and the importance of choosing appropriate proposal distributions for the MH algorithm:

Example 2 (Joint inference for the parameters of a normal distribution (con't))**.** *In the setting of Example 1, p. 23, the* ***Gibbs sampler*** *takes the following form:*

0) Set $r = 0$. Choose arbitrary starting values $\mu^{(0)}$ and $\psi^{(0)}$.

1) Draw $\mu^{(r+1)} \sim \mathcal{N}(c, 1/a)$, where

$$a = n\psi^{(r)} + \frac{1}{\sigma_\mu^2} \tag{2.87}$$

$$\text{and } c = \frac{n\sigma_\mu^2}{n\sigma_\mu^2 + (\psi^{(r)})^{-1}}\overline{\mathcal{S}} + \frac{(\psi^{(r)})^{-1}}{n\sigma_\mu^2 + (\psi^{(r)})^{-1}}\mu_\mu. \tag{2.88}$$

2) Draw $\psi^{(r+1)} \sim Ga\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum_{i=1}^n \left(S_i - \mu^{(r+1)}\right)^2\right)$.
Set $r = r + 1$ and go to 1).

Alternatively, we may apply the ***Metropolis-Hastings algorithm****. As proposal distributions we choose $\mathcal{N}\left(\mu^{(r)}, \sigma_{prop}^2\right)$ and $Exp(\lambda)$ for the $(r+1)$-th update of μ and ψ, respectively. Note that $\mathcal{N}\left(\mu^{(r)}, \sigma_{prop}^2\right)$ is a symmetric proposal distribution, whereas $Exp(\lambda)$ is not. The MH algorithm is given by*

0) Set $r = 0$. Choose arbitrary starting values $\mu^{(0)}$ and $\psi^{(0)}$.

1) Propose a new value for the mean, $\mu' \sim \mathcal{N}\left(\mu^{(r)}, \sigma_{prop}^2\right)$.
Let $U_1^{(r)} \sim unif(0,1)$ and accept μ', i.e. set $\mu^{(r+1)} = \mu'$, if

$$U_1^{(r)} \leq \min\left\{1, \frac{f_{\mathcal{N}(c,1/a)}(\mu')}{f_{\mathcal{N}(c,1/a)}\left(\mu^{(r)}\right)} \times \frac{f_{\mathcal{N}\left(\mu', \sigma_{prop}^2\right)}\left(\mu^{(r)}\right)}{f_{\mathcal{N}\left(\mu^{(r)}, \sigma_{prop}^2\right)}(\mu')}\right\} \tag{2.89}$$

$$= \min\left\{1, \exp\left[-\frac{a}{2}\left(\left(\mu'\right)^2 - \left(\mu^{(r)}\right)^2 - 2c\left(\mu' - \mu^{(r)}\right)\right)\right]\right\}, \tag{2.90}$$

otherwise set $\mu^{(r+1)} = \mu^{(r)}$.

2) Propose a new value for the precision, $\psi' \sim Exp(\lambda)$.
Let $U_2^{(r)} \sim unif(0,1)$ *and accept* ψ'*, i.e. set* $\psi^{(r+1)} = \psi'$*, if*

$$U_2^{(r)} \leq \min\left\{ 1, \frac{f_{Ga(\tilde{\alpha},\tilde{\beta})}(\psi')}{f_{Ga(\tilde{\alpha},\tilde{\beta})}\left(\psi^{(r)}\right)} \times \frac{f_{Exp(\lambda)}\left(\psi^{(r)}\right)}{f_{Exp(\lambda)}\left(\psi'\right)} \right\} \tag{2.91}$$

$$= \min\left\{ 1, \left(\frac{\psi'}{\psi^{(r)}}\right)^{\tilde{\alpha}-1} e^{(\tilde{\beta}-\lambda)\left(\psi^{(r)}-\psi'\right)} \right\}, \tag{2.92}$$

where $\tilde{\alpha} = \alpha + \frac{n}{2}$ *and* $\tilde{\beta} = \beta + \frac{1}{2}\sum_{i=1}^{n}\left(S_i - \mu^{(r+1)}\right)^2$*, otherwise set* $\psi^{(r+1)} = \psi^{(r)}$.
Set $r = r + 1$ *and go to 1).*

We applied both algorithms to a dataset of 100 samples that were independently drawn from $\mathcal{N}(\mu_0, \sigma_0^2)$. The parameter calibration is reported in Table 2.1.

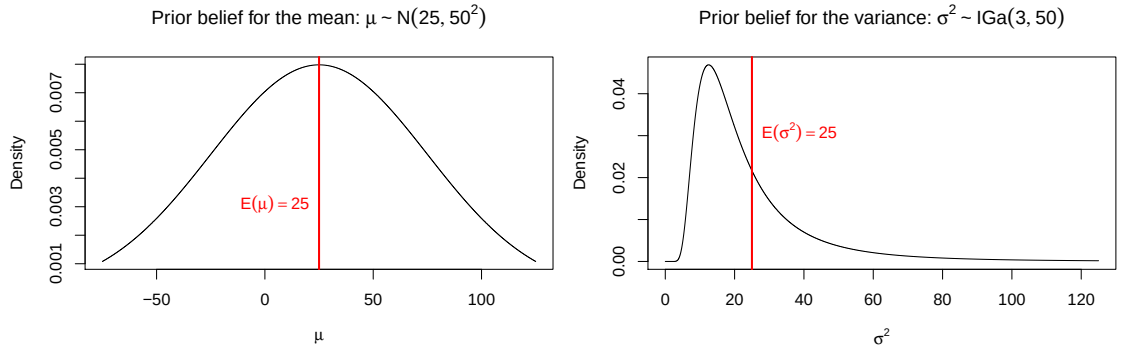
True sample parameters	
μ_0	50.5
σ_0	30.5
Prior parameters for μ	
μ_μ	25
σ_μ	50
Prior parameters for ψ	
α	3
β	50
Parameters for the MCMC algorithms	
R	100000
$\mu^{(0)}$	$30 \times \mu_0$
$\psi^{(0)}$	$\sigma_0^{-2}/5$
σ_{prop}	5
λ	10

Table 2.1.: Example 1: True parameters for data simulation, prior distributions and MCMC algorithms.

In Figure 2.3 the prior distributions are shown. Clearly, the prior expectations for μ and σ^2 are much smaller than the true parameters, i.e. $\mathbb{E}[\mu] = 25 < 50.5 = \mu_0$ and $\mathbb{E}[\sigma^2] = 25 < \sigma_0^2 = 930.25$. Nonetheless, we might expect the posterior distribution to approach the true parameters μ_0 and σ_0^2 since the number of observations is quite large (cf. the conditional distributions in Example 1, p. 23, and Remark 2.7, p. 25). To be able to analyze the convergence behavior of the MCMC algorithms, we initialize both algorithms rather far away from the true parameter values, i.e. $\mu^{(0)} > \mu_0$ and $\psi^{(0)} < \psi_0 = \sigma_0^{-2}$.

Note that for the MH algorithm we choose a proposal distribution for $\mu^{(r+1)}$ that is symmetric around the latest sample $\mu^{(r)}$ and exhibits a small variance.

For the precision $\psi^{(r+1)}$ we choose an $Exp(10)$-distribution, thus, independent of $\psi^{(r)}$ the average proposed value for $\psi^{(r+1)}$ is $1/10 = 0.1$. This is far away from the true precision, which is $\psi_0 = \sigma_0^{-2} \approx 0.0011$. Therefore, we may expect the MH algorithm to exhibit small acceptance probabilities, thus, to move very slowly (particularly in the precision's component) and to exhibit rather high autocorrelations.



(A) Prior for μ (B) Prior for σ^2

Figure 2.3.: Example 2: Prior beliefs for (A) μ and (B) σ^2.

We applied both algorithms to sample $R = 100000$ values for both parameters. Figures 2.4 and 2.5 analyze the evolution of the simulated Markov chains. Firstly, we show *trace plots* of the first few samples (to identify the burn-in period BI) and a trace plot of following samples (to check for stationarity and trends). Secondly, we plot the *sample autocorrelation function* (ACF) based on all samples after the burn-in period. The ACF for a time series $(\vartheta^{(r)})_{r \in H}$ is defined as follows:

Definition 2.5 ((Sample) Autocorrelation function)**.** *(Cowpertwait and Metcalfe (2009, p.33-34)) For a stochastic process $\left(\vartheta^{(r)}\right)_{r \in H}$, with $H = \mathbb{R}$ or $H = \mathbb{Z}$, the autocorrelation function (ACF) at time $r \in H$ and lag $l > 0$ is given by*

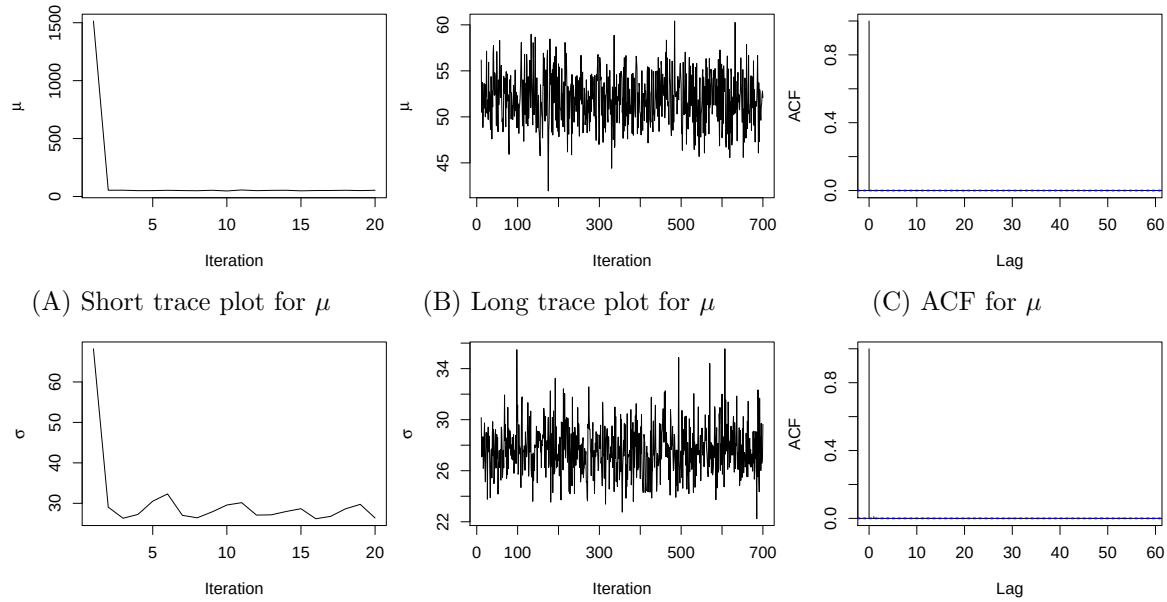
$$f(l, r) = \frac{cov\left(\vartheta^{(r+l)}, \vartheta^{(r)}\right)}{\sqrt{var\left(\vartheta^{(r+l)}\right) \; var\left(\vartheta^{(r)}\right)}}. \tag{2.93}$$

If the process $\left(\vartheta^{(r)}\right)$ is weak-sense stationary, i.e. $\mathbb{E}\left[\vartheta^{(r)}\right] = \mathbb{E}\left[\vartheta^{(0)}\right]$ and $cov\left(\vartheta^{(r+l)}, \vartheta^{(r)}\right) = cov\left(\vartheta^{(l)}, \vartheta^{(0)}\right)$ for all $r \in H, l > 0$, the ACF is time-independent, i.e. $f(l, r) \equiv f(l)$.

For samples $\vartheta = \left(\vartheta^{(1)}, ..., \vartheta^{(R)}\right)$ of a weak-sense stationary time series the ACF can be approximated by the sample ACF, which is given as

$$\hat{f}(l) = \frac{\frac{1}{R}\sum_{r=1}^{R-l}\left(\vartheta^{(r)} - \overline{\vartheta}\right)\left(\vartheta^{(r+l)} - \overline{\vartheta}\right)}{\frac{1}{R}\sum_{r=1}^{R}\left(\vartheta^{(r)} - \overline{\vartheta}\right)\left(\vartheta^{(r)} - \overline{\vartheta}\right)}. \tag{2.94}$$

As shown in Hoff (2009, p.102) the variance of a MCMC estimate for the posterior mean is equal to the (standard) Monte Carlo variance plus a (generally positive) term that depends on the correlation of samples within the Markov chain. Thus, the higher the autocorrelation of the chain, the worse the approximation is. This highlights the importance of a fast decreasing ACF for obtaining a good posterior approximation.



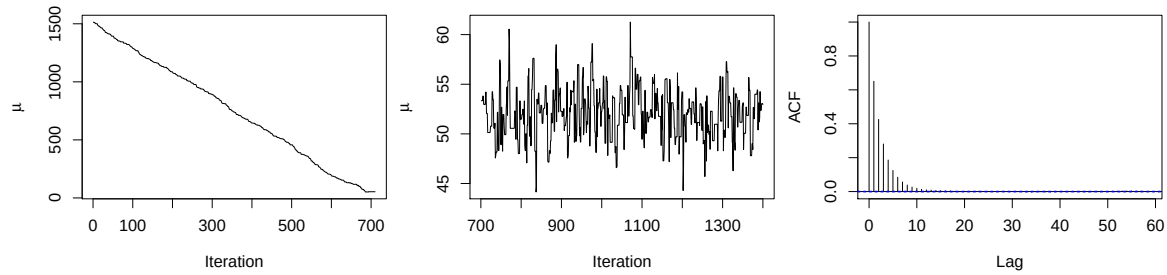
(A) Short trace plot for μ (B) Long trace plot for μ (C) ACF for μ

(D) Short trace plot for σ (E) Long trace plot for σ (F) ACF for σ

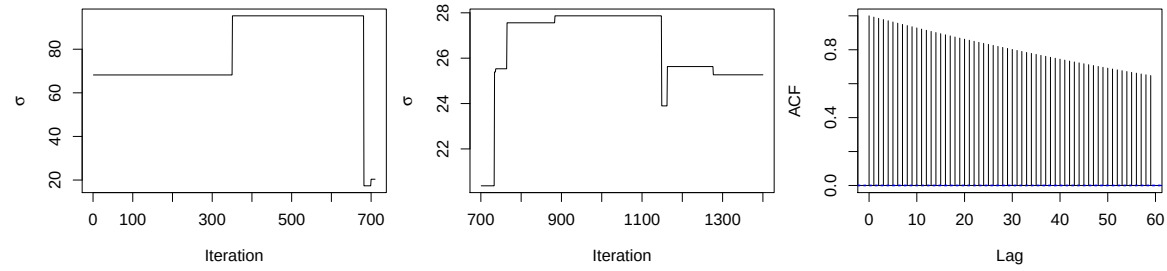
Figure 2.4.: Example 2: Gibbs sampler: Trace plots and ACF for (A)-(C) μ and (D)-(F) σ.

Clearly, the Gibbs sampler exhibits desirable properties: Due to a very small burn-in period (we set $BI = 10$, however, one might even suspect it to equal 1) we may assume that the chain has *achieved stationarity* at least after 10 iterations. Moreover, samples are uncorrelated if taken after the burn-in period and the chain does not get stuck in distinctive regions of the parameter space and exhibits no long-term trend. In summary, the chain exhibits a *very good mixing behavior*.

Contrarily, for the MH algorithm we obtain a Markov chain that is *poorly mixing*: The burn-in period is very large (we set $BI = 700$), samples for μ are significantly correlated for lags smaller than 10 and samples for σ are extremely correlated for all lags up to $l = 50$ and larger. Nonetheless, the histograms of samples for both parameters (constructed with every 10th observation after the burn-in period for μ and every observation after the burn-in period for $\sigma = \psi^{-1/2}$) exhibit very similar properties to these obtained with the Gibbs sampler, in particular the posterior means are equal. The samples' histograms, including density approximations (with Gaussian kernel), the posterior means and true parameters are shown in Figure 2.6.

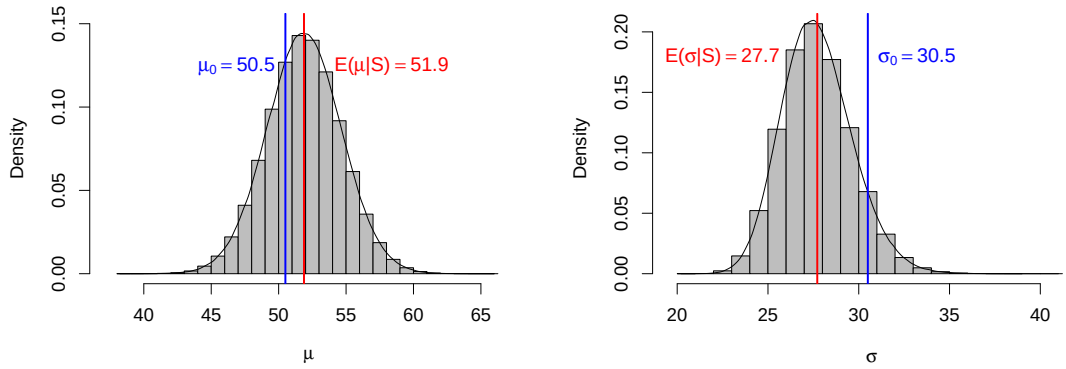


(A) Short trace plot for μ (B) Long trace plot for μ (C) ACF for μ

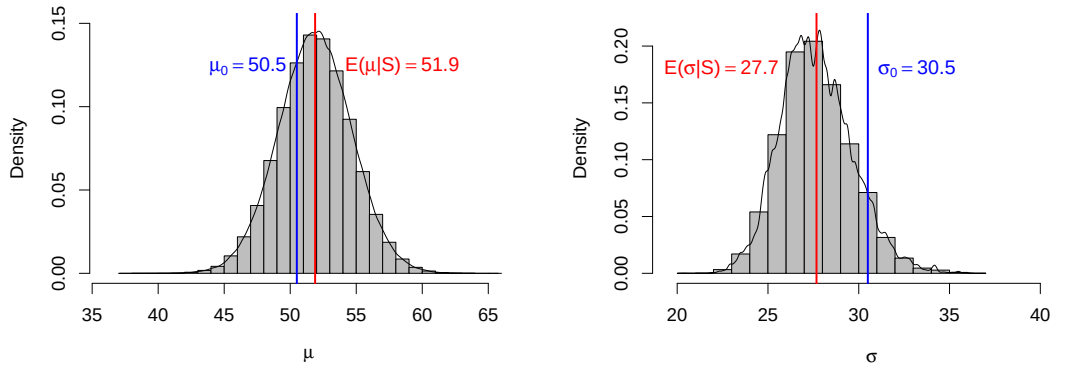


(D) Short trace plot for σ (E) Long trace plot for σ (F) ACF for σ

Figure 2.5.: Example 2: MH algorithm: Trace plots and ACF for (A)-(C) μ and (D)-(F) σ.



(A) Posterior Distribution: μ (B) Posterior Distribution: σ



(C) Posterior Distribution: μ (D) Posterior Distribution: σ

Figure 2.6.: Example 2: Posterior distributions: (A)-(B) Gibbs and (C)-(D) MH algorithm.

In conclusion, we may assume that both algorithms achieved stationarity. However, Gibbs sampling obtained a very well mixing Markov chain with a small burn-in period, whereas samples from the MH algorithm are poorly mixing but may still be used to obtain an approximation for the posterior distribution.

3. Bayesian inference in the GLO process

In this chapter we assume the following underlying setting: In the observation window $[0, T]$ we observed the occurrence (or arrival/event) times $\mathcal{S} := (S_1, ..., S_n),\ S_i \leq S_{i+1}$, of $n \in \mathbb{N}$ events that originated from $N \in \mathbb{N}$ beats from a GLO process $\Phi \sim \text{GLO}(\mu, \sigma_1, \sigma_2, \gamma)$. Unfortunately, we did neither observe these beats $\mathcal{B} := (B_1, ..., B_N)$ nor do we know how many or which events belong to which beat. We will occasionally call the events originating from one beat the beat's *event family*.

In the following we will construct a Bayesian framework that enables us to obtain posterior information about the beats, their event-families and the parameters of the GLO process. For this purpose we introduce *labels* $\mathcal{J} = (J_1, ..., J_n)$ that link events to beats: For every event S_i, $i = 1, ..., n$, and beat B_k, $k = 1, ..., N$, label J_i equals k if event S_i originates from beat B_k. Moreover, n_k counts the *number of events originating from* B_k:

$$n_k := |\{i \in \{1, ..., n\} : J_i = k\}|, \quad \text{for } k = 1, ..., N. \tag{3.1}$$

3.1. Bayesian framework

We consider a parameter vector that is $(4+n+N)$-dimensional and consists of the parameters of the GLO model, beat locations and labels:

$$\vartheta := (\mu, \psi_1, \psi_2, \gamma, B_1, ..., B_N, J_1, ..., J_n)\,. \tag{3.2}$$

Note that similar to Example 1, p. 23, we will work with the precisions ψ_1 and ψ_2 instead of σ_1 and σ_2 to be able to employ conjugate priors. Nonetheless, in the end we will be interested in estimates for σ_1 and σ_2 since these are more straightforward to interpret. Moreover, from MCMC posterior samples $\psi_i^{(1)}, ..., \psi_i^{(R)}$ we can easily obtain samples $\sigma_i^{(r)} = \left(\psi_i^{(r)}\right)^{-1/2}$, $r = 1, ..., R,\ i = 1, 2$.

Note that, although the (unobserved) beats and labels are not parameters in the GLO model, we still treat them as if they were parameters whose posterior distribution we wanted to compute. This method is called *data augmentation* and is widely used in models that deal with unobserved data. For example, Scott (1999) applies this method for a hidden Markov model and Loza-Reyes et al. (2014) for mixture models augmented with allocation variables. For more examples and a more general introduction to data augmentation we refer to Tanner and Wong (1987).

Since the event times, S_i, and numbers of events per burst, P_k, are independent, the likelihood can be split into two parts. Recall from section 2.1 that, firstly, for every beat B_k the number of events, $P_k \sim Pois(\gamma)$, is drawn and, secondly, the events are distributed around the beat with independent normally distributed increments with variance σ_2^2. If the GLO process exhibits a large variance σ_2^2 or beats near the limits of the observation window $[0, T]$, we may not for all beats B_k, $k \in \{1, ..., N\}$, observe all P_k events, but in fact only $n_k \leq P_k$.

Nevertheless, due to the thinning property of the Poisson distribution (see Durrett (2012, p. 106) and section 2.1) n_k is still Poisson distributed with parameter

$$\gamma_k := \gamma \, \mathbb{P}(S_{k,1} \in [0, T]) = \gamma \underbrace{\left(F_{\mathcal{N}(B_k, \sigma_2^2)}(T) - F_{\mathcal{N}(B_k, \sigma_2^2)}(0) \right)}_{=:\nu_k}, \tag{3.3}$$

which is also used in Proposition 2.3, p. 17. Since all observed events are restricted to $[0, T]$, we have $S_i \sim \mathcal{N}(B_{J_i}, \sigma_2^2 \mid [0, T])$ with pdf (cf. Griffiths (2002))

$$f_{\mathcal{N}(B_{J_i}, \sigma_2^2 | [0,T])}(x) = \nu_{J_i}^{-1} f_{\mathcal{N}(B_{J_i}, \sigma_2^2)}(x) = \frac{\nu_{J_i}^{-1}}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2\sigma_2^2}\left(x - B_{J_i}\right)^2}, \tag{3.4}$$

for $0 \leq x \leq T$ and 0 otherwise. Note that in the definition of the GLO process the i-th event at beat B_k is denoted as $S_{k,i}$. However, here we count the observed events in increasing (chronological) order and denote S_i for the i-th observed event.

In conclusion, the likelihood is given by

$$\begin{aligned} L(\vartheta; \mathcal{S}) &= \prod_{i=1}^{n} f_{\mathcal{N}(B_{J_i}, \sigma_2^2 | [0,T])}(S_i) \times \prod_{k=1}^{N} f_{Pois(\gamma_k)}(n_k) \\ &= \prod_{k=1}^{N} \prod_{i: J_i = k} \nu_k^{-1} f_{\mathcal{N}(B_k, \psi_2^{-1})}(S_i) \times \prod_{k=1}^{N} \frac{\gamma_k^{n_k}}{n_k!} e^{-\gamma_k} \\ &= \left(\frac{\psi_2}{2\pi} \right)^{n/2} \prod_{k=1}^{N} (\nu_k)^{-n_k} \exp\left[-\frac{\psi_2}{2} \sum_{i: J_i = k} (S_i - B_k)^2 \right] \times \gamma^n \prod_{k=1}^{N} \frac{\nu_k^{n_k}}{n_k!} e^{-\nu_k \gamma}. \end{aligned} \tag{3.5}$$

In order to simplify the calculations, we will from now on assume the following conjugate prior distributions:

1. $\mu \sim \mathcal{N}\left(\mu_\mu, \sigma_\mu^2\right)$
2. $\psi_1 \sim Ga\left(\alpha_{\psi_1}, \beta_{\psi_1}\right)$
3. $\psi_2 \sim Ga\left(\alpha_{\psi_2}, \beta_{\psi_2}\right)$
4. $\gamma \sim Ga\left(\alpha_\gamma, \beta_\gamma\right)$

Since in the GLO model the background rhythm is based on a stationary random walk, we have $\mathcal{L}\left(B_{k+1} - B_k \mid \mu, \psi_1\right) = \mathcal{N}\left(\mu, \psi_1^{-1}\right)$. Moreover, the "true" distribution of B_1 is endogenously given by the asymptotic residual waiting time (see Bingmer (2012, p.25-27)). However, a priori we assume B_1 to be independently distributed according to a flat normal distribution, $B_1 \sim \mathcal{N}(\mu_{B_1}, \sigma^2_{B_1})$ because, firstly, given any other beat B_1 is normally distributed, secondly, B_1 will mostly be determined by the first few observations since we can choose the prior to be rather uninformative, and, thirdly, with this assumption we can substantially simplify calculations.

For J_i we do not know of one particular conjugate prior distribution, thus, we will a priori either assume $\mathcal{L}(J_i) = unif(\{1, ..., N\})$ or $\mathcal{L}(J_i) = Binomial(N, i/n)$.

All priors are a priori assumed to be independent. All prior assumptions as well as other possible priors will be discussed more extensively in section 3.3 and and the influence of the prior choice in section 3.4.1. With the specified prior distributions the joint posterior distribution is proportional to

$$\begin{aligned} p(\vartheta \mid \mathcal{S}) \propto L(\vartheta; \mathcal{S}) & f_{\mathcal{N}\left(\mu_\mu, \sigma^2_\mu\right)}(\mu) f_{Ga\left(\alpha_{\psi_1}, \beta_{\psi_1}\right)}(\psi_1) f_{Ga\left(\alpha_{\psi_2}, \beta_{\psi_2}\right)}(\psi_2) f_{Ga(\alpha_\gamma, \beta_\gamma)}(\gamma) \\ & \times f_{\mathcal{N}\left(\mu_{B_1}, \sigma^2_{B_1}\right)}(B_1) \prod_{k=2}^{N} f_{\mathcal{N}\left(\mu, \psi_1^{-1}\right)}(B_k - B_{k-1}) \prod_{i=1}^{n} \pi(J_i). \end{aligned} \tag{3.6}$$

Since this distribution is not straightforward to work with, we will calculate the parameter's FCPDs in the following and present ways how to perform Gibbs or MH updates. To simplify notations we will analogously to ϑ_{-k} denote with $\vartheta_{-\mu}$ the vector of all components of the parameter vector ϑ except for μ, with ϑ_{-B_k} the vector of all components except for B_k, and so forth.

Lemma 3.1 (FCPD of γ)**.** *Let a priori $\gamma \sim Ga(\alpha_\gamma, \beta_\gamma)$. Then, the full conditional posterior distribution of γ is given by*

$$\mathcal{L}\left(\gamma \mid \vartheta_{-\gamma}, \mathcal{S}\right) = Ga\left(\alpha_\gamma + n,\ \beta_\gamma + \sum_{k=1}^{N} \nu_k\right), \tag{3.7}$$

where

$$\nu_k := F_{\mathcal{N}(B_k, \psi_2^{-1})}(T) - F_{\mathcal{N}(B_k, \psi_2^{-1})}(0). \tag{3.8}$$

Proof. According to Bayes' theorem the FCPD is proportional to

$$p(\gamma \mid \vartheta_{-\gamma}, \mathcal{S}) \propto L(\vartheta; \mathcal{S})\ p(\mu, \psi_1, \psi_2, \gamma, \mathcal{B}, \mathcal{J}) \tag{3.9}$$

$$\propto \gamma^{\alpha_\gamma + n - 1} e^{-\beta_\gamma \gamma} \prod_{k=1}^{N} \frac{\nu_k^{n_k}}{n_k!} e^{-\nu_k \gamma} \propto \gamma^{\alpha_\gamma + n - 1} \exp\left[-\gamma\left(\beta_\gamma + \sum_{k=1}^{N} \nu_k\right)\right], \tag{3.10}$$

which is proportional to the pdf of $Ga\left(\alpha_\gamma + n,\ \beta_\gamma + \sum_{k=1}^{N} \nu_k\right)$. □

Lemma 3.2 (FCPD of μ)**.** *Let a priori $\mu \sim \mathcal{N}\left(\mu_\mu, \sigma_\mu^2\right)$ and $\pi(B_1)$ independent of μ. Then, the full conditional posterior distribution of μ is given by*

$$\mathcal{L}\left(\mu \mid \vartheta_{-\mu}, \mathcal{S}\right) = \mathcal{N}\left(\frac{\psi_1 \sum_{k=2}^{N}(B_k - B_{k-1}) + \sigma_\mu^{-2}\mu_\mu}{\psi_1(N-1) + \sigma_\mu^{-2}},\ \frac{1}{\psi_1(N-1) + \sigma_\mu^{-2}}\right). \tag{3.11}$$

Proof. According to Bayes' theorem the FCPD is proportional to

$$p(\mu \mid \vartheta_{-\mu}, \mathcal{S}) \propto L(\vartheta; \mathcal{S})\ p(\mu, \psi_1, \psi_2, \gamma, \mathcal{B}, \mathcal{J}) \tag{3.12}$$

$$\propto p(\mathcal{B} \mid \mu, \psi_1)\, \pi(\mu). \tag{3.13}$$

By rewriting $p(\mathcal{B} \mid \mu, \psi_1)$ in terms of the increments of the random walk we obtain

$$p(\mu \mid \vartheta_{-\mu}, \mathcal{S}) \propto \pi(\mu)\ \pi(B_1) \prod_{k=2}^{N} f_{\mathcal{N}\left(\mu, \psi_1^2\right)}(B_k - B_{k-1}) \tag{3.14}$$

$$\propto \exp\left[-\frac{1}{2}\left(\psi_1 \sum_{k=2}^{N}(B_k - B_{k-1} - \mu)^2 + \frac{1}{\sigma_\mu^2}(\mu - \mu_\mu)^2\right)\right]. \tag{3.15}$$

Similar to Example 1, p. 23, with $a := \psi_1(N-1) + \sigma_\mu^{-2}$ and $b := \psi_1 \sum_{k=2}^{N}(B_k - B_{k-1}) + \sigma_\mu^{-2}\mu_\mu$ and $c \in \mathbb{R}$ independent of μ equation (3.15) is proportional to

$$\exp\left[-\frac{1}{2}\left(a\mu^2 - 2b\mu + c\right)\right] \propto \exp\left[-\frac{a}{2}\left(\mu - \frac{b}{a}\right)^2\right], \tag{3.16}$$

which is proportional to the pdf of a normal distribution with variance

$$\frac{1}{a} = \frac{1}{\psi_1(N-1) + \sigma_\mu^{-2}} \tag{3.17}$$

and mean

$$\frac{b}{a} = \frac{1}{\psi_1(N-1) + \sigma_\mu^{-2}}\left(\psi_1 \sum_{k=2}^{N}(B_k - B_{k-1}) + \sigma_\mu^{-2}\mu_\mu\right) \tag{3.18}$$

$$= \frac{\sigma_\mu^2(N-1)}{\psi_1^{-1} + \sigma_\mu^2(N-1)} \frac{1}{N-1} \sum_{k=2}^{N}(B_k - B_{k-1}) + \frac{\psi_1^{-1}}{\psi_1^{-1} + \sigma_\mu^2(N-1)}\mu_\mu. \tag{3.19}$$

□

Lemma 3.3 (FCPD of ψ_1)**.** *Let a priori $\psi_1 \sim Ga(\alpha_{\psi_1}, \beta_{\psi_1})$ and $\pi(B_1)$ independent of ψ_1. Then, the full conditional posterior distribution of ψ_1 is given by*

$$\mathcal{L}\left(\psi_1 \mid \vartheta_{-\psi_1}, \mathcal{S}\right) = Ga\left(\alpha_{\psi_1} + \frac{N-1}{2},\ \beta_{\psi_1} + \frac{1}{2}\sum_{k=2}^{N}(B_k - B_{k-1} - \mu)^2\right). \tag{3.20}$$

Proof. According to Bayes' theorem the FCPD is (similar to Lemma 3.2) proportional to

$$p(\psi_1 \mid \vartheta_{-\psi_1}, \mathcal{S}) \propto L(\vartheta; \mathcal{S})\, p(\mu, \psi_1, \psi_2, \gamma, \mathcal{B}, \mathcal{J}) \tag{3.21}$$

$$\propto p(\mathcal{B} \mid \mu, \psi_1)\, \pi(\psi_1) \tag{3.22}$$

$$\propto \left(\frac{\psi_1}{\sqrt{2\pi}}\right)^{(N-1)/2} \exp\left[-\frac{1}{2}\left(\psi_1 \sum_{k=2}^{N}(B_k - B_{k-1} - \mu)^2\right) - \beta_{\psi_1}\psi_1\right] \psi_1^{\alpha_{\psi_1}-1} \tag{3.23}$$

$$\propto \psi_1^{\alpha_{\psi_1}+(N-1)/2-1} \exp\left[-\psi_1\left(\beta_{\psi_1} + \frac{1}{2}\sum_{k=2}^{N}(B_k - B_{k-1} - \mu)^2\right)\right], \tag{3.24}$$

which is proportional to the pdf of $Ga\left(\alpha_{\psi_1} + \frac{N-1}{2},\, \beta_{\psi_1} + \frac{1}{2}\sum_{k=2}^{N}(B_k - B_{k-1} - \mu)^2\right)$. □

The beats require Bayesian inference for the truncated normal distribution. A Gibbs sampling approach for this problem is described by Griffiths (2002). The author introduces latent variables $z_1, ..., z_n$ that are distributed according to a non-truncated normal distribution and deterministically linked to the truncated observations $S_1, .., S_n$. In the following we will adopt this idea and adjust the estimation framework to fit to the GLO process.

Lemma 3.4 (FCPD of ψ_2)**.** *Let a priori $\psi_2 \sim Ga(\alpha_{\psi_2}, \beta_{\psi_2})$. Then, a Gibbs update for $\psi_2^{(r)}$ is given by*

(1) Compute $z_i, i = 1, ..., n$, by

$$z_i = B_{J_i} + \left(\psi_2^{(r)}\right)^{-1/2} F^{-1}_{\mathcal{N}(0,1)}\left(\frac{F_{\mathcal{N}(0,1)}\left(\frac{S_i - B_{J_i}}{\left(\psi_2^{(r)}\right)^{-1/2}}\right) - F_{\mathcal{N}(0,1)}\left(\frac{0 - B_{J_i}}{\left(\psi_2^{(r)}\right)^{-1/2}}\right)}{F_{\mathcal{N}(0,1)}\left(\frac{T - B_{J_i}}{\left(\psi_2^{(r)}\right)^{-1/2}}\right) - F_{\mathcal{N}(0,1)}\left(\frac{0 - B_{J_i}}{\left(\psi_2^{(r)}\right)^{-1/2}}\right)}\right). \tag{3.25}$$

(2) Draw $\psi_2^{(r+1)}$ from $Ga\left(\alpha_{\psi_2} + \frac{n}{2},\, \beta_{\psi_2} + \frac{1}{2}\sum_{i=1}^{n}(z_i - B_{J_i})^2\right)$.

Proof. In the following we will denote by $F := F_{\mathcal{N}(0,1)}$ the pdf of the standard Normal distribution. Let $z_i \overset{iid}{\sim} \mathcal{N}\left(B_{J_i}, \psi_2^{-1}\right)$, $i = 1, .., n$. One may draw samples z_i from $\mathcal{N}(B_{J_i}, \psi_2^{-1})$ and S_i from $\mathcal{N}\left(B_{J_i}, \psi_2^{-1} \mid [0, T]\right)$ by independently drawing uniformly distributed random variables $U_1, ..., U_n \sim unif(0, 1)$ and calculating

$$z_i = B_{J_i} + \psi_2^{-1/2} F^{-1}(U_i) \tag{3.26}$$

and

$$S_i = B_{J_i} + \psi_2^{-1/2} F^{-1}\left[F\left(\frac{0 - B_{J_i}}{\psi_2^{-1/2}}\right) + U_i\left(F\left(\frac{T - B_{J_i}}{\psi_2^{-1/2}}\right) - F\left(\frac{0 - B_{J_i}}{\psi_2^{-1/2}}\right)\right)\right]. \tag{3.27}$$

The latter result can be found for example in Albert and Chib (1996). Hence, given values for B_{J_i}, ψ_2 and S_i, a value for U_i can be calculated by equation (3.27). Then, the corresponding value for z_i can be computed from equation (3.26) and is given by equation (3.25). Thus, to sample from the FCPD of ψ_2 one can use Bayes' theorem in the following way (we denote $\mathcal{Z} := (z_1, ..., z_n)$):

$$p(\psi_2 \mid \vartheta_{-\psi_2}, \mathcal{S}) \propto p(\psi_2, \mathcal{Z} \mid \mathcal{S}, \mathcal{B}) \tag{3.28}$$

$$\propto p(\mathcal{S} \mid \psi_2, \mathcal{B}, \mathcal{Z})\; p(\psi_2, \mathcal{Z} \mid \mathcal{B}) \tag{3.29}$$

$$\propto p(\mathcal{S} \mid \psi_2, \mathcal{B}, \mathcal{Z})\; p(\mathcal{Z} \mid \psi_2, \mathcal{B})\; \pi(\psi_2). \tag{3.30}$$

Now, $\pi(\mathcal{S} \mid \psi_2, \mathcal{B}, \mathcal{Z}) = 1$ if equation (3.25) holds for all $i = 1, .., n$ and in this case we get

$$p(\psi_2 \mid \vartheta_{-\psi_2}, \mathcal{S}) \propto p(\mathcal{Z} \mid \psi_2, \mathcal{B})\; \pi(\psi_2) \tag{3.31}$$

$$\propto \psi_2^{n/2} \exp\left[-\frac{\psi_2}{2}\sum_{i=1}^{n}(z_i - B_{J_i})^2\right] \times \psi_2^{\alpha_{\psi_2}-1} \exp\left[-\psi_2\beta_{\psi_2}\right] \tag{3.32}$$

$$\propto \psi_2^{\alpha_{\psi_2}+n/2-1} \exp\left[-\psi_2\left(\beta_{\psi_2} + \frac{1}{2}\sum_{i=1}^{n}(z_i - B_{J_i})^2\right)\right], \tag{3.33}$$

which is proportional to the pdf of $Ga\left(\alpha_{\psi_2} + \frac{n}{2}, \beta_{\psi_2} + \frac{1}{2}\ \sum_{i=1}^{n}(z_i - B_{J_i})^2\right)$. □

Lemma 3.5 (FCPD of $\mathcal{B}$)**.** *Let a priori $B_1 \sim \mathcal{N}\left(\mu_{B_1}, \sigma_{B_1}^2\right)$. Then, a Gibbs update for beat $B_k^{(r)}$ is given by*

(1) Compute $z_j^{(k)}, j = 1, ..., n_k$, by evaluating

$$z_j^{(k)} = B_k^{(r)} + \psi_2^{-1/2} F_{\mathcal{N}(0,1)}^{-1}\left(\frac{F_{\mathcal{N}(0,1)}\left(\frac{S_{i_j} - B_k^{(r)}}{\psi_2^{-1/2}}\right) - F_{\mathcal{N}(0,1)}\left(\frac{0 - B_k^{(r)}}{\psi_2^{-1/2}}\right)}{F_{\mathcal{N}(0,1)}\left(\frac{T - B_k^{(r)}}{\psi_2^{-1/2}}\right) - F_{\mathcal{N}(0,1)}\left(\frac{0 - B_k^{(r)}}{\psi_2^{-1/2}}\right)}\right), \tag{3.34}$$

where $\{i_j : j = 1, ..., n_k\} = \{i \in \{1, ..., n\} : J_i = k\}$.

(2) If $k = 1$ draw $B_1^{(r+1)}$ from

$$\mathcal{N}\left(\frac{\sigma_{B_1}^{-2}\mu_{B_1} + \psi_1(B_2 - \mu) + \psi_2\sum_{j=1}^{n_1} z_j^{(1)}}{\psi_1 + n_1\psi_2 + \sigma_{B_1}^{-2}},\ \frac{1}{\psi_1 + n_1\psi_2 + \sigma_{B_1}^{-2}}\right). \tag{3.35}$$

If $k \in \{2, ..., N-1\}$ draw $B_k^{(r+1)}$ from

$$\mathcal{N}\left(\frac{1}{2\psi_1 + n_k\psi_2}\left(\psi_1(B_{k+1} + B_{k-1}) + \psi_2\sum_{j=1}^{n_k} z_j^{(k)}\right),\ \frac{1}{2\psi_1 + n_k\psi_2}\right). \tag{3.36}$$

If $k = N$ *draw* $B_N^{(r+1)}$ *from*

$$\mathcal{N}\left(\frac{1}{\psi_1 + n_N\psi_2}\left(\psi_1(B_{N-1}+\mu) + \psi_2\sum_{j=1}^{n_N} z_j^{(N)}\right),\ \frac{1}{\psi_1 + n_N\psi_2}\right). \tag{3.37}$$

Proof. Let $k \in \{2, ..., N-1\}$. The FCPD for B_k is given by

$$\begin{aligned} p\left(B_k \mid \vartheta_{-B_k}, \mathcal{S}\right) &\propto L(\vartheta; \mathcal{S})\ p(\mu, \psi_1, \psi_2, \gamma, \mathcal{B}, \mathcal{J}) && (3.38)\\ &\propto L(\vartheta; \mathcal{S})\ p(B_{k+1} \mid \mu, \psi_1, B_k)\ p(B_k \mid \mu, \psi_1, B_{k-1}), && (3.39)\end{aligned}$$

since only the likelihood and the beats around B_k depend on B_k, which is due to the independence of the increments of a random walk. Then, we have

$$\begin{aligned} &p(B_k \mid \vartheta_{-B_k}, \mathcal{S})\\ &\propto \prod_{i:J_i=k} \nu_k^{-1} f_{\mathcal{N}\left(B_k, \psi_2^{-1}\right)}(S_i)\ f_{\mathcal{N}\left(\mu, \psi_1^{-1}\right)}(B_{k+1} - B_k)\ f_{\mathcal{N}\left(\mu, \psi_1^{-1}\right)}(B_k - B_{k-1}) && (3.40)\\ &\propto \prod_{i:J_i=k} \nu_k^{-1} f_{\mathcal{N}\left(B_k, \psi_2^{-1}\right)}(S_i)\ f_{\mathcal{N}\left(B_k, \psi_1^{-1}\right)}(B_{k+1} - \mu) f_{\mathcal{N}\left(B_k, \psi_1^{-1}\right)}(B_{k-1} + \mu), && (3.41)\end{aligned}$$

where

$$\nu_k := F_{\mathcal{N}\left(B_k, \psi_2^{-1}\right)}(T) - F_{\mathcal{N}\left(B_k, \psi_2^{-1}\right)}(0). \tag{3.42}$$

Now, B_k only appears in the mean of the single densities, hence, this setting is similar to inference for the posterior mean of truncated normal distributions with flat prior $\pi(B_k) \propto 1$ and known variance. However, in this case not all distributions are truncated at the same points. In fact, the joint distribution (3.41) includes n_k densities that are truncated such that $0 \leq S_i \leq T$, while the two densities for the beats are not truncated. Therefore, we adopt the idea from Lemma 3.4 and change it slightly to fit to the considered problem:

We introduce latent random variables $z_1^{(k)}, ..., z_{n_k}^{(k)}$ with $z_1^{(k)}, ..., z_{n_k}^{(k)} \overset{iid}{\sim} \mathcal{N}\left(B_k, \psi_2^{-1}\right)$ and define $\mathcal{Z}^{(k)} := \left(z_1^{(k)}, .., z_{n_k}^{(k)}\right)$. Note that we do not need latent variables for the non-truncated observations. Furthermore, let $\{i_1, .., i_{n_k}\} = \{i : J_i = k\}$ be the labels of the beat's event-family $S^{(k)} := (S_{i_1}, ..., S_{i_{n_k}})$. Analogously to Lemma 3.4 sampling from z_j and S_{i_j} via inversion of the cdf is straightforward and yields the relationship that is given by equation (3.34). Moreover, with Bayes' theorem we obtain

$$\begin{aligned} &p(B_k \mid \vartheta_{-B_k}, \mathcal{S})\\ &\propto p\left(S^{(k)} \mid \mathcal{Z}^{(k)}, \mu, \psi_1, \psi_2, B_{k-1}, B_k, B_{k+1}\right)\ p\left(\mathcal{Z}^{(k)}, \mu, \psi_1, \psi_2, \gamma, B_{k-1}, B_k, B_{k+1}\right) && (3.43)\\ &\propto p\left(S^{(k)} \mid \mathcal{Z}^{(k)}, B_k\right)\ p(B_{k+1} \mid B_k, \mu, \psi_1)\, p\left(\mu, \psi_1, B_{k-1}, B_k, \mathcal{Z}^{(k)}\right) && (3.44)\\ &\propto p\left(S^{(k)} \mid \mathcal{Z}^{(k)}, B_k\right)\ p(B_{k+1} \mid B_k, \mu, \psi_1)\, p(B_k \mid B_{k-1}, \mu, \psi_1)\, p\left(\mathcal{Z}^{(k)} \mid B_k, \mu, \psi_1\right). && (3.45)\end{aligned}$$

Clearly, $p\left(S^{(k)} \mid \mathcal{Z}^{(k)}, B_k\right) = 1$ if equation (3.34) holds for all $j = 1, .., n_k$ and in this case we yield

$$
\begin{aligned}
&p(B_k \mid \vartheta_{-B_k}, \mathcal{S}) \\
&\propto p(B_{k+1} \mid B_k, \mu, \psi_1)\, p(B_k \mid B_{k-1}, \mu, \psi_1)\, p\left(\mathcal{Z}^{(k)} \mid B_k, \mu, \psi_1\right) && (3.46)\\
&\propto f_{\mathcal{N}(\mu, \psi_1^{-1})}(B_{k+1} - B_k)\, f_{\mathcal{N}(\mu, \psi_1^{-1})}(B_k - B_{k-1}) \prod_{j=1}^{n_k} f_{\mathcal{N}(B_k, \psi_2^{-1})}\left(z_j^{(k)}\right) && (3.47)\\
&\propto \exp\left[-\frac{1}{2}\ \left(\psi_1(B_{k+1} - B_k - \mu)^2 + \psi_1(B_k - B_{k-1} - \mu)^2 + \psi_2 \sum_{j=1}^{n_k}\left(z_j^{(k)} - B_k\right)^2\right)\right] && (3.48)\\
&\propto \exp\left[-\frac{1}{2}\ \left(aB_k^2 - 2bB_k + c\right)\right] \propto \exp\left[-\frac{a}{2}\ \left(B_k - \frac{b}{a}\right)^2\right], && (3.49)
\end{aligned}
$$

where $a = 2\psi_1 + n_k\psi_2$ and $b = \psi_1(B_{k+1} + B_{k-1}) + \psi_2 \sum_{j=1}^{n_k} z_j^{(k)}$ and $c \in \mathbb{R}$. Thus, B_k is a posteriori conditionally on $\mathcal{Z}, \mu, \psi_1$ and ψ_2 normally distributed with variance

$$\frac{1}{a} = \frac{1}{2\psi_1 + n_k\psi_2} \tag{3.50}$$

and mean

$$\frac{b}{a} = \frac{1}{2\psi_1 + n_k\psi_2}\left(\psi_1(B_{k+1} + B_{k-1}) + \psi_2 \sum_{j=1}^{n_k} z_j^{(k)}\right). \tag{3.51}$$

For $k = 1$ we obtain (if equation (3.34) holds for all $j = 1, .., n_1$)

$$
\begin{aligned}
p(B_1 \mid \vartheta_{-B_1}, \mathcal{S}) &\propto f_{\mathcal{N}(\mu, \sigma_1^2)}(B_2 - B_1)\, f_{\mathcal{N}\left(\mu_{B_1}, \sigma_{B_1}^2\right)}(B_1) \prod_{j=1}^{n_1} f_{\mathcal{N}(B_1, \sigma_2^2)}\left(z_j^{(1)}\right) && (3.52)\\
&\propto \exp\left[-\frac{1}{2}\ \left(\psi_1(B_2 - B_1 - \mu)^2 + \psi_2 \sum_{j=1}^{n_1}\left(z_j^{(1)} - B_1\right)^2 + \sigma_{B_1}^{-2}(B_1 - \mu_{B_1})^2\right)\right] && (3.53)\\
&\propto \exp\left[-\frac{a}{2}\ \left(B_1 - \frac{b}{a}\right)^2\right], && (3.54)
\end{aligned}
$$

with $a = \psi_1 + n_1\psi_2 + \sigma_{B_1}^{-2}$ and $b = \sigma_{B_1}^{-2}\mu_{B_1} + \psi_1(B_2 - \mu) + \psi_2 \sum_{j=1}^{n_1} z_j^{(1)}$, which yields (3.35).

Analogously, if $k = N$, the FCPD is (if equation (3.34) holds for all $j = 1, .., n_N$)

$$p(B_N \mid \vartheta_{-B_N}, \mathcal{S}) \propto f_{\mathcal{N}(\mu,\psi_1^{-1})}(B_N - B_{N-1}) \prod_{j=1}^{n_N} f_{\mathcal{N}(B_N,\sigma_2^2)}\left(z_j^{(N)}\right) \tag{3.55}$$

$$\propto \exp\left[-\frac{1}{2}\left(\psi_1(B_N - B_{N-1} - \mu)^2 + \psi_2 \sum_{j=1}^{n_N}\left(z_j^{(N)} - B_N\right)^2\right)\right] \tag{3.56}$$

$$\propto \exp\left[-\frac{a}{2}\left(B_N - \frac{b}{a}\right)^2\right], \tag{3.57}$$

where $a = \psi_1 + n_N\psi_2$ and $b = \psi_1(B_{N-1} + \mu) + \psi_2 \sum_{j=1}^{n_N} z_j^{(N)}$, which yields (3.37). □

Lemma 3.6 (FCPD of $\mathcal{J}$)**.** *Let $\pi_1(\cdot), ..., \pi_n(\cdot)$ be the independent prior distributions for labels $J_1, ..., J_n$. Then, the FCPD of J_i, $i = 1, ..., n$, is proportional to*

$$p\left(J_i \mid \vartheta_{-J_i}, \mathcal{S}\right) \propto \exp\left[-\frac{\psi_2}{2}(S_i - B_{J_i})^2 - \nu_{J_i}\gamma\right] \times \frac{\nu_{J_i}^{n_{J_i}-1}}{n_{J_i}!} \times \pi_i(J_i). \tag{3.58}$$

To sample from this distribution, we will make use of the Metropolis-Hastings algorithm, which yields the following updating procedure:

(1) For $J_i^{(r)} = k$ propose a switch to beat k', where k' is uniformly distributed among surrounding beats in distance $\Delta^{(switch)}$, i.e. $k' \sim unif(V_k)$ with

$$V_k := \left\{\max\{k - \Delta^{(switch)}, 1\}, ..., \min\{k + \Delta^{(switch)}, N\}\right\}. \tag{3.59}$$

(2) If $k' \neq k$, accept the new beat, i.e. set $J_i^{(r+1)} = k'$, with probability

$$p_{acc}^{(switch)}(k' \mid k) = \min\left\{1, \tilde{p}\right\}, \tag{3.60}$$

where

$$\tilde{p} = \exp\left[-\frac{\psi_2}{2}\left(B_{k'}^2 - B_k^2 + 2S_i(B_k - B_{k'})\right)\right] \times \frac{n_k}{n_{k'} + 1} \times \frac{\pi(k')}{\pi(k)} \times \frac{|V_k|}{|V_{k'}|}.$$

Otherwise, set $J_i^{(r+1)} = k$.

Proof. The FCPD is proportional to

$$p(J_i \mid \vartheta_{-J_i}, \mathcal{S}) \propto L(\vartheta; S)\, p(\mu, \psi_1, \psi_2, \gamma, \mathcal{B}, \mathcal{J}) \tag{3.61}$$

$$\propto L(\vartheta; S)\, p(\mathcal{B} \mid \mu, \psi_1)\, \pi(\mu, \psi_1, \psi_2, \gamma, \mathcal{J}) \tag{3.62}$$

$$\propto \nu_{J_i}^{-1} f_{\mathcal{N}\left(B_{J_i},\psi_2^{-1}\right)}(S_i) \times f_{Pois(\gamma\,\nu_{J_i})}(n_{J_i}) \times \pi(J_i) \tag{3.63}$$

$$\propto \nu_{J_i}^{-1} \exp\left[-\frac{\psi_2}{2}(S_i - B_{J_i})^2\right] \times \frac{\nu_{J_i}^{n_{J_i}}}{n_{J_i}!} \exp\left[-\nu_{J_i}\gamma\right] \times \pi(J_i). \tag{3.64}$$

If $k = k'$, we have $p_{acc}^{(switch)} = 1$ (cf. Algorithm 2.2). In contrast, a switch from $J_i^{(r)} = k$ to $J_i^{(r+1)} = k'$ with $k' \neq k$ affects both n_k and $n_{k'}$, thus, the acceptance probability is given by $p_{acc}^{(switch)}(k' \mid k) = \min\{1, \tilde{p}\}$ with

$$\begin{aligned}
\tilde{p} &= \frac{\nu_{k'}^{-1}}{\nu_k^{-1}} \times \frac{\exp\left[-\frac{\psi_2}{2}(S_i - B_{k'})^2\right] \times \frac{\nu_k^{n_k - 1}}{(n_k-1)!} \times \frac{\nu_{k'}^{n_{k'}+1}}{(n_{k'}+1)!} \exp\left[-(\nu_{k'} + \nu_k)\gamma\right] \times \pi(k')}{\exp\left[-\frac{\psi_2}{2}(S_i - B_k)^2\right] \times \frac{\nu_k^{n_k}}{n_k!} \times \frac{\nu_{k'}^{n_{k'}}}{n_{k'}!} \exp\left[-(\nu_{k'} + \nu_k)\gamma\right] \times \pi(k)} \times \frac{|V_{k'}|^{-1}}{|V_k|^{-1}} \\
&= \frac{\nu_k}{\nu_{k'}} \times \exp\left[-\frac{\psi_2}{2}\left((S_i - B_{k'})^2 - (S_i - B_k)^2\right)\right] \times \frac{\nu_k^{n_k - 1}\, \nu_{k'}^{n_{k'}+1}\, n_k!\, n_{k'}!}{\nu_k^{n_k}\, \nu_{k'}^{n_{k'}}\, (n_k - 1)!\, (n_{k'} + 1)!} \times \frac{\pi(k')\,|V_k|}{\pi(k)\,|V_{k'}|} \\
&= \frac{\nu_k}{\nu_{k'}} \times \exp\left[-\frac{\psi_2}{2}\left(B_{k'}^2 - B_k^2 + 2S_i(B_k - B_{k'})\right)\right] \times \frac{\nu_{k'}\, n_k}{\nu_k\, (n_{k'} + 1)} \times \frac{\pi(k')\,|V_k|}{\pi(k)\,|V_{k'}|} \\
&= \exp\left[-\frac{\psi_2}{2}\left(B_{k'}^2 - B_k^2 + 2S_i(B_k - B_{k'})\right)\right] \times \frac{n_k}{n_{k'} + 1} \times \frac{\pi(k')}{\pi(k)} \times \frac{|V_k|}{|V_{k'}|}. \qquad (3.65)
\end{aligned}$$

□

3.2. Estimation procedure

With the FCPDs and updating steps derived in section 3.1 the complete MCMC algorithm for sampling from $p(\vartheta \mid \mathcal{S})$ is given as follows:

Algorithm 3.1 (Complete MCMC procedure)**.** *Firstly, values for the following quantities have to be chosen:*

- *Number of MCMC-iterations:* $R \in \mathbb{N}$*, Number of beats:* $N \in \mathbb{N}\backslash\{1\}$
- *Prior parameters for* γ : $\alpha_\gamma > 0, \beta_\gamma > 0$
- *Prior parameters for* μ : $\mu_\mu \in \mathbb{R}, \sigma_\mu > 0$
- *Prior parameters for* ψ_1 : $\alpha_{\psi_1} > 0, \alpha_{\psi_1} > 0$
- *Prior parameters for* ψ_2 : $\alpha_{\psi_2} > 0, \alpha_{\psi_2} > 0$
- *Prior parameters for* B_1 : $\mu_{B_1} \in \mathbb{R}, \sigma_{B_1} > 0$
- *Update parameters for* $\mathcal{J}$*:*
  - $\pi_i(\cdot)$ *(prior distribution for* J_i*)*
  - $n^{(switch)} = \max\{1, \lfloor np^{(switch)} \rceil\} \in \{1, ..., n\}$ *(number of labels to be updated)*
  - $\Delta^{(switch)} \in \{1, ..., \lceil (n-1)/2 \rceil\}$ *(MH proposal step-size)*

Then, the MCMC procedure for sampling from $p(\vartheta \mid \mathcal{S})$ *is as follows:*

0) *Set initial values for* $\mu^{(0)}, \psi_1^{(0)}, \psi_2^{(0)}, \gamma^{(0)}, B_1^{(0)}, ..., B_N^{(0)}, J_1^{(0)}, ..., J_n^{(0)}$ *and* $r = 0$.

1) *Update γ: Draw* $\gamma^{(r+1)} \sim Ga\left(\alpha_\gamma + n,\ \beta_\gamma + \sum_{k=1}^{N} \nu_k^{(r)}\right)$

2) *Update μ: Draw*

$$\mu^{(r+1)} \sim \mathcal{N}\left(\frac{\psi_1^{(r)} \sum_{k=2}^{N}\left(B_k^{(r)} - B_{k-1}^{(r)}\right) + \sigma_\mu^{-2}\mu_\mu}{\psi_1^{(r)}(N-1) + \sigma_\mu^{-2}},\ \frac{1}{\psi_1^{(r)}(N-1) + \sigma_\mu^{-2}}\right)$$

3) *Update ψ_1: Draw*

$$\psi_1^{(r+1)} \sim Ga\left(\alpha_{\psi_1} + \frac{N-1}{2},\ \beta_{\psi_1} + \frac{1}{2}\sum_{k=2}^{N}\left(B_k^{(r)} - B_{k-1}^{(r)} - \mu^{(r+1)}\right)^2\right)$$

4) *Update ψ_2:*

4.1) *Compute z_i, $i = 1, ..., n$, by*

$$z_i = B_{J_i}^{(r)} + \left(\psi_2^{(r)}\right)^{-1/2} F_{\mathcal{N}(0,1)}^{-1}\left(\frac{F_{\mathcal{N}(0,1)}\left(\frac{S_i - B_{J_i}^{(r)}}{\left(\psi_2^{(r)}\right)^{-1/2}}\right) - F_{\mathcal{N}(0,1)}\left(\frac{0 - B_{J_i}^{(r)}}{\left(\psi_2^{(r)}\right)^{-1/2}}\right)}{F_{\mathcal{N}(0,1)}\left(\frac{T - B_{J_i}^{(r)}}{\left(\psi_2^{(r)}\right)^{-1/2}}\right) - F_{\mathcal{N}(0,1)}\left(\frac{0 - B_{J_i}^{(r)}}{\left(\psi_2^{(r)}\right)^{-1/2}}\right)}\right)$$

4.2) *Draw*

$$\psi_2^{(r+1)} \sim Ga\left(\alpha_{\psi_2} + \frac{n}{2},\ \beta_{\psi_2} + \frac{1}{2}\sum_{i=1}^{n}\left(z_i - B_{J_i}^{(r)}\right)^2\right)$$

5) *Update B_1:*

5.1) *Compute z_j, $j = 1, ..., n_1^{(r)}$, by evaluating*

$$z_j = B_1^{(r)} + \left(\psi_2^{(r+1)}\right)^{-1/2} F_{\mathcal{N}(0,1)}^{-1}\left(\frac{F_{\mathcal{N}(0,1)}\left(\frac{S_{i_j} - B_1^{(r)}}{\left(\psi_2^{(r+1)}\right)^{-1/2}}\right) - F_{\mathcal{N}(0,1)}\left(\frac{0 - B_1^{(r)}}{\left(\psi_2^{(r+1)}\right)^{-1/2}}\right)}{F_{\mathcal{N}(0,1)}\left(\frac{T - B_1^{(r)}}{\left(\psi_2^{(r+1)}\right)^{-1/2}}\right) - F_{\mathcal{N}(0,1)}\left(\frac{0 - B_1^{(r)}}{\left(\psi_2^{(r+1)}\right)^{-1/2}}\right)}\right),$$

where $\left\{i_j : j = 1, ..., n_1^{(r)}\right\} = \left\{i \in \{1, ..., n\} : J_i^{(r)} = 1\right\}$.

5.2) *Draw*

$$B_1^{(r+1)} \sim \mathcal{N}\left(\frac{\sigma_{B_1}^{-2}\mu_{B_1} + \psi_1^{(r+1)}\left(B_2^{(r)} - \mu^{(r+1)}\right) + \psi_2^{(r+1)}\sum_{j=1}^{n_1^{(r)}} z_j}{\psi_1^{(r+1)} + n_1^{(r)}\psi_2^{(r+1)} + \sigma_{B_1}^{-2}},\ \frac{1}{\psi_1^{(r+1)} + n_1^{(r)}\psi_2^{(r+1)} + \sigma_{B_1}^{-2}}\right)$$

6) *For* $k = 2, ..., N-1$ *update* B_k*:*

6.1) *Compute* z_j, $j = 1, ..., n_k^{(r)}$, *by evaluating*

$$z_j = B_k^{(r)} + \left(\psi_2^{(r+1)}\right)^{-1/2} F_{\mathcal{N}(0,1)}^{-1} \left(\frac{F_{\mathcal{N}(0,1)}\left(\frac{S_{i_j} - B_k^{(r)}}{\left(\psi_2^{(r+1)}\right)^{-1/2}} \right) - F_{\mathcal{N}(0,1)}\left(\frac{0 - B_k^{(r)}}{\left(\psi_2^{(r+1)}\right)^{-1/2}} \right)}{F_{\mathcal{N}(0,1)}\left(\frac{T - B_k^{(r)}}{\left(\psi_2^{(r+1)}\right)^{-1/2}} \right) - F_{\mathcal{N}(0,1)}\left(\frac{0 - B_k^{(r)}}{\left(\psi_2^{(r+1)}\right)^{-1/2}} \right)} \right)$$

where $\left\{ i_j : j = 1, ..., n_k^{(r)} \right\} = \left\{ i \in \{1, ..., n\} : J_i^{(r)} = k \right\}$.

6.2) *Draw*

$$B_k^{(r+1)} \sim \mathcal{N}\left(\frac{\psi_1^{(r+1)} \left(B_{k+1}^{(r)} + B_{k-1}^{(r+1)} \right) + \psi_2^{(r+1)} \sum_{j=1}^{n_k^{(r)}} z_j}{2\psi_1^{(r+1)} + n_k^{(r)} \psi_2^{(r+1)}}, \frac{1}{2\psi_1^{(r+1)} + n_k^{(r)} \psi_2^{(r+1)}} \right)$$

7) *Update* B_N*:*

7.1) *Compute* z_j, $j = 1, ..., n_N^{(r)}$, *by evaluating*

$$z_j = B_N^{(r)} + \left(\psi_2^{(r+1)}\right)^{-1/2} F_{\mathcal{N}(0,1)}^{-1} \left(\frac{F_{\mathcal{N}(0,1)}\left(\frac{S_{i_j} - B_N^{(r)}}{\left(\psi_2^{(r+1)}\right)^{-1/2}} \right) - F_{\mathcal{N}(0,1)}\left(\frac{0 - B_N^{(r)}}{\left(\psi_2^{(r+1)}\right)^{-1/2}} \right)}{F_{\mathcal{N}(0,1)}\left(\frac{T - B_N^{(r)}}{\left(\psi_2^{(r+1)}\right)^{-1/2}} \right) - F_{\mathcal{N}(0,1)}\left(\frac{0 - B_N^{(r)}}{\left(\psi_2^{(r+1)}\right)^{-1/2}} \right)} \right)$$

where $\left\{ i_j : j = 1, ..., n_N^{(r)} \right\} = \left\{ i \in \{1, ..., n\} : J_i^{(r)} = N \right\}$.

7.2) *Draw*

$$B_N^{(r+1)} \sim \mathcal{N}\left(\frac{\psi_1^{(r+1)} \left(B_{N-1}^{(r+1)} + \mu^{(r+1)} \right) + \psi_2^{(r+1)} \sum_{j=1}^{n_N^{(r)}} z_j}{\psi_1^{(r+1)} + n_N^{(r)} \psi_2^{(r+1)}}, \frac{1}{\psi_1^{(r+1)} + n_N^{(r)} \psi_2^{(r+1)}} \right)$$

8) *Draw* $H \subseteq \{1, ..., n\}$, $|H| = n^{(switch)}$, *randomly without replacing. For each* $i \in H$ *update* J_i*:*

8.1) *For* $J_i^{(r)} = k$ *draw* $k' \sim unif(V_k)$ *with*

$$V_k = \left\{ \max\{k - \Delta^{(switch)}, 1\}, ..., \min\{k + \Delta^{(switch)}, N\} \right\}$$

8.2) *If* $k' \neq k$, *set* $J_i^{(r+1)} = k'$ *with probability* $p_{acc}^{(switch)}(k' \mid k) = \min\{1, \tilde{p}\}$, *where*

$$\tilde{p} = \exp\left[-\frac{\psi_2^{(r+1)}}{2} \left(\left(B_{k'}^{(r+1)} \right)^2 - \left(B_k^{(r+1)} \right)^2 + 2S_i \left(B_k^{(r+1)} - B_{k'}^{(r+1)} \right) \right) \right] \frac{n_k^*}{n_{k'}^* + 1} \frac{\pi(k')}{\pi(k)} \frac{|V_k|}{|V_{k'}|}.$$

Otherwise, set $J_i^{(r+1)} = k$.

9) *Set* $r = r + 1$. *If* $r < R$*: Go to 1), else: stop.*

Note, that during the execution of step 8) the size of beat B_k*'s family,* n_k, *may change. However, always the latest (updated) value of* n_k *should be used, which we denote by* n_k^*.

Proposition 3.1 (Convergence)**.** *The distribution of the Markov chain constructed by Algorithm 3.1 converges to the target distribution $p(\vartheta \mid \mathcal{S})$.*

Proof. Firstly, we will focus on updating step 8) for the labels: If $n^{(switch)} < n$, $J_i^{(r)}$ is not updated with positive probability, thus $\mathbb{P}(J_i^{(r+1)} = k \mid J_i^{(r)} = k) > 0$. If $n^{(switch)} = n$, J_i is updated in each iteration. In this case, since $k \in V_k$, we have

$$\mathbb{P}\left(J_i^{(r+1)} = k \mid J_i^{(r)} = k\right) = |V_k|^{-1} \, p_{acc}^{(switch)}(k \mid k) = |V_k|^{-1} > 0. \tag{3.66}$$

Thus, this step is aperiodic according to Lemma 2.2, p. 29, for all $i = 1, ..., n$. Moreover, the proposal distribution is irreducible because it reaches any state $k^* \in \{1, ..., N\}$ from any state $k \in \{1, ..., N\}$ after $x \geq \lceil |k^* - k| / \Delta^{(switch)} \rceil$ iterations with positive probability, and we have that $k' \notin V_k$ if, and only if, $k \notin V_{k'}$. Thus, the updates of all label components J_i satisfy Theorem 2.2, p. 30.

Secondly, we analyze the remaining Gibbs updates. Therefore, we focus on the components $\vartheta_{-\mathcal{J}} := (\vartheta_1, \vartheta_2, ..., \vartheta_{4+N})$ and treat the labels $\mathcal{J} = (\vartheta_{4+N+1}, ..., \vartheta_{4+N+n})$ as given. Therefore, joint prior density is given by

$$\begin{aligned} \pi(\vartheta_{-\mathcal{J}}) := f_{\mathcal{N}(\mu_\mu, \sigma_\mu^2)}(\mu) f_{Ga\left(\alpha_{\psi_1}, \beta_{\psi_1}\right)}(\psi_1) f_{Ga\left(\alpha_{\psi_2}, \beta_{\psi_2}\right)}(\psi_2) f_{Ga(\alpha_\gamma, \beta_\gamma)}(\gamma) \\ \times f_{\mathcal{N}\left(\mu_{B_1}, \sigma_{B_1}^2\right)}(B_1) \prod_{k=2}^{N} f_{\mathcal{N}\left(\mu, \psi_1^{-1}\right)}(B_k - B_{k-1}) \end{aligned} \tag{3.67}$$

and the target distribution is given by the joint posterior distribution (cf. equation (3.6)),

$$\tilde{p}(\vartheta_{-\mathcal{J}}) := p(\vartheta_{-\mathcal{J}} \mid \mathcal{S}, \mathcal{J}) = \frac{p(\mathcal{S} \mid \vartheta_{-\mathcal{J}}, \mathcal{J}) \, \pi(\vartheta_{-\mathcal{J}})}{\int_{\Theta_{-\mathcal{J}}} p(\mathcal{S} \mid \vartheta_{-\mathcal{J}}, \mathcal{J}) \, \pi(\vartheta_{-\mathcal{J}}) \, d\vartheta_{-\mathcal{J}}}. \tag{3.68}$$

According to Theorem 2.3, p. 30, it is sufficient to proof that *a)* the target distribution $\tilde{p}$ is l.s.c. at 0, *b)* $\int \tilde{p}(\vartheta_{-\mathcal{J}}) d\vartheta_i$ is locally bounded and *c)* that $\Theta_{-\mathcal{J}} := \Theta_{1,...,4+N}$ is connected. The last condition is fulfilled since

$$\Theta_{-\mathcal{J}} = \underbrace{(-\infty, \infty)}_{(\mu)} \times \underbrace{(0, \infty)}_{(\psi_1)} \times \underbrace{(0, \infty)}_{(\psi_2)} \times \underbrace{(0, \infty)}_{(\gamma)} \times \underbrace{(-\infty, \infty)^N}_{(B_1, ..., B_N)}. \tag{3.69}$$

Now, recall that $p(\mathcal{S} \mid \vartheta_{-\mathcal{J}}, \mathcal{J}) \, \pi(\vartheta_{-\mathcal{J}})$ is the product of continuous Lebesgue densities, thus, continuous and positive on $\Theta_{-\mathcal{J}}$. It follows $\int_{\Theta_{-\mathcal{J}}} p(\mathcal{S} \mid \vartheta_{-\mathcal{J}}, \mathcal{J}) \, \pi(\vartheta_{-\mathcal{J}}) \, d\vartheta_{-\mathcal{J}} > 0$, hence, $\tilde{p}(\vartheta_{-\mathcal{J}})$ is well-defined on $\Theta_{-\mathcal{J}}$ and l.s.c. at 0, which is condition *a)*.

To show that condition *b)* holds let $i \in \{1, ..., 4 + N\}$, $\vartheta_0 \in \Theta_{-\mathcal{J},-i}$ and $F_i(\vartheta_{-\mathcal{J},-i}) := \int \tilde{p}(\vartheta_{-\mathcal{J}}) d\vartheta_i$. We need to show that there exists $M > 0$ and an open neighborhood $N_{\vartheta_0} \subset \Theta_{-\mathcal{J},-i}$ containing ϑ_0 such that $|F_i(\vartheta)| \leq M$ for all $\vartheta \in N_{\vartheta_0}$. This property follows directly since all FCPDs are well-defined (cf. Lemmas 3.2 to 3.5). For example, take $i = 1$ and arbitrary $\vartheta_0 \in (0, \infty)^3 \times (-\infty, \infty)^N$. Then, ϑ_1 (which is μ) only appears in N parts of the

joint prior and we have for all $\vartheta \in N_{\vartheta_0}$ (cf. Lemma 3.2)

$$F_1(\vartheta) = \int_{-\infty}^{\infty} \frac{p(\mathcal{S} \mid \vartheta_{-\mathcal{J}}, \mathcal{J})\, \pi(\vartheta_{-\mathcal{J}})}{\int_{\Theta_{-\mathcal{J}}} p(\mathcal{S} \mid \vartheta_{-\mathcal{J}}, \mathcal{J})\, \pi(\vartheta_{-\mathcal{J}})\, d\vartheta_{-\mathcal{J}}} d\vartheta_1 \tag{3.70}$$

$$\leq C \int_{-\infty}^{\infty} \underbrace{f_{\mathcal{N}\left(\mu_\mu, \sigma_\mu^2\right)}(\mu) \prod_{k=2}^{N} f_{\mathcal{N}\left(\mu, \psi_1^{-1}\right)}(B_k - B_{k-1})}_{=g(\mu)}\, d\mu, \tag{3.71}$$

where $C > 0$ is a positive constant. Since $g(\mu)$ can be normalized with $\tilde{C} > 0$ such that $\tilde{C} g(\mu)$ is the density of a normal distribution (namely the FCPD of μ, cf. Lemma 3.2), we get $|F_1(\vartheta)| \leq C\tilde{C}$ for all $\vartheta \in N_{\vartheta_0}$. The calculation is analog for all other components. □

Remark 3.1 (Numerical stability)**.** *Some steps of the MCMC algorithm (Algorithm 3.1) are prone to numerical instability issues. This affects particularly the updates for $B_1, ..., B_N$ and ψ_2, since in these steps the quantile and cumulative probability function of the normal density is employed. To give an example, if $B_{J_i} > 0$ is very large in comparison to S_i, than both $F_{\mathcal{N}(0,1)}\left(\frac{S_i - B_{J_i}}{\psi_2^{-1/2}}\right)$ and $F_{\mathcal{N}(0,1)}\left(\frac{0 - B_{J_i}}{\psi_2^{-1/2}}\right)$ may approximately be zero (cf. equation (3.25)). However, if the algorithm sets both to zero, for the latent variable z_i in step 4.1) we yield*

$$|z_i| \approx \left| B_{J_i} + \psi_2^{-1/2} \lim_{x \to 0} F^{-1}_{\mathcal{N}(0,1)} \left(\frac{x}{F_{\mathcal{N}(0,1)}\left(\frac{T - B_{J_i}}{\psi_2^{-1/2}}\right) - F_{\mathcal{N}(0,1)}\left(\frac{0 - B_{J_i}}{\psi_2^{-1/2}}\right)} \right) \right| = \infty. \tag{3.72}$$

Therefore, the parameter values should be restricted in the numerical procedure.

3.2.1. Choice of initial values

To complete the estimation procedure we need to set need initial values for all components of $\vartheta^{(0)}$. Initial values in areas with large probability mass of the target distribution $\tilde{p}$ are vital to shorten burn-in periods, as Example 2, p. 31, demonstrates. Since one may expect good (*frequentists'*) point estimates for the parameters to exhibit a large probability/density (as indicated by Example 1, p. 23, and Remark 2.7, p. 25), we will in the following propose several procedures to obtain first estimates for the parameters.

As a first approach, we require to determine beat locations and number of beats by visual inspection. Based on these, events are matched to their nearest beat and initial values for the other parameters are given by their Maximum-Likelihood estimates or inverse of the corrected sample variance (since $\psi_1 = \sigma_1^{-2}$ and $\psi_2 = \sigma_2^{-2}$).

Algorithm 3.2 (Simple guess of initial values)**.**

1) Choose $B_1^{(0)}, ..., B_N^{(0)}$ and N by visual inspection.

2) Set $\mu^{(0)} = \frac{1}{N-1}\sum_{k=2}^{N}\left(B_k^{(0)} - B_{k-1}^{(0)}\right)$
and $\psi_1^{(0)} = \left(\frac{1}{N-2}\sum_{k=2}^{N}\left(B_k^{(0)} - B_{k-1}^{(0)} - \mu^{(0)}\right)^2\right)^{-1}$.

3) Set

$$J_i^{(0)} := \arg\min_{k\in\{1,...,N\}} \left|B_k^{(0)} - S_i\right| \quad \text{for } i = 1, ..., n. \tag{3.73}$$

4) Set $\psi_2^{(0)} = \left(\frac{1}{n}\sum_{i=1}^{n}\left(S_i - B_{J_i}^{(0)}\right)^2\right)^{-1}$ *and* $\gamma^{(0)} = \frac{1}{N}\sum_{k=1}^{N}\left|\left\{i \in \{1, ..., n\} : J_i^{(0)} = k\right\}\right|$.

Algorithm 3.2 is straightforward to implement and may provide very good estimates if the initial values for beats and number of beats fit. Nonetheless, particularly visual estimation of the beat locations may be rather time consuming and prone to unsystematic (subjective) failures. Alternatively, one may, firstly, estimate N visually and, secondly, partition $[0, T]$ into N subintervals and place one beat in the middle of each of these intervals. Then, step *1)* of Algorithm 3.2 is replaced by the following two steps:

1.1) Choose N *by visual inspection.*

1.2) Set $B_k^{(0)} = \frac{T}{2N} + (k-1)\frac{T}{N}$ *for* $k = 1, ..., N$.

Nevertheless, this procedure will provide very poor parameter estimates if N does not fit or the intervals' mid points do not fit to the beat locations. Hence, we may consider an algorithm that estimates the beat locations based on estimates for μ, σ_1 and σ_2. The algorithm builds on a kernel density estimator (KDE) which estimates the pdf of $\mathcal{S}$ under the assumption that $\mathcal{S}$ consists of iid samples from one distribution (an introduction to KDEs may be found in Parzen (1962) or Silverman (1986)). Thus, for events $\mathcal{S}$ originating from a GLO process we expect the KDE to exhibit peaks at every burst and pits between bursts. Furthermore, since the events are normally distributed around the beats, the choice of a Gaussian kernel is reasonable.

Algorithm 3.3 (Estimating beat locations)**.** *(Mohapatra (2013))*

1) Compute the kernel density estimate (KDE) of $\mathcal{S}$ *with a Gaussian kernel and bandwidth* σ.

2) Identify the turnpoints (maxima and minima) of the KDE with the R function turnpoints().

3) For every turnpoint x *(starting with the global maximum of the KDE) consider the interval* $I(a, b) = [x + \mu - a\sigma_1 - b\sigma_2, x + \mu + a\sigma_1 + b\sigma_2]$*:*
Let $a = 1$ *and* $b = 0$.
If there exist turnpoints in $I(a, b)$*: Identify the respective turnpoint as beat that fits best to the background rhythm.*
Otherwise: Increase the interval size with default values for a *and* b*. If for all default values there does not exist a turnpoint in* $I(a, b)$*, locate an additional beat at* $x + \mu$.

Since in the GLO model events are distributed with variance σ_2^2, it is convenient to choose $\sigma = \sigma_2$ as bandwidth. However, we may increase (decrease) σ in order to decrease (increase) the resulting estimated number of beats, N.

With Algorithm 3.3 it is possible to obtain initial values for N and $\mathcal{B}^{(0)}$. Moreover, $\mathcal{J}^{(0)}$ may be set analogously to step *3)* in Algorithm 3.2. However, Algorithm 3.3 requires estimates for μ, σ_1 and σ_2 as input parameters. These may be obtained by applying the ACH estimation procedure developed in Bingmer et al. (2011) by fitting the empirical ACH to the theoretical ACF of the GLO model (cf. Remark 2.6). In conclusion, we obtain the following procedure:

Algorithm 3.4 (Complex guess of initial values)**.**

1) Obtain $\hat{\eta}_{ACH} = \left(\mu^{(0)}, \sigma_1^{(0)}, \sigma_2^{(0)}, \gamma^{(0)}\right)$ *as proposed by Bingmer et al. (2011).*

2) Set $\psi_1^{(0)} = \left(\sigma_1^{(0)}\right)^{-2}$ *and* $\psi_2^{(0)} = \left(\sigma_2^{(0)}\right)^{-2}$.

3) Apply Algorithm 3.3 with $\hat{\eta}_{ACH}$ *to obtain values for* N *and* $B_1^{(0)}, ..., B_N^{(0)}$.

4) Set

$$J_i^{(0)} := \arg \min_{k \in \{1,...,N\}} \left| B_k^{(0)} - S_i \right|. \tag{3.74}$$

3.2.2. What N is best?

Clearly, the number of beats, N, is a very crucial parameter to the whole estimation procedure. However, it is not straightforward to extend the Bayesian estimation framework in order to estimate N, since for different $N_1 \neq N_2$ the parameter vector exhibits a different dimension, i.e. $4+n+N_1 \neq 4+n+N_2$. Hence, we essentially deal with different models.

To distinguish between one GLO model with different fixed numbers of beats, we define by $\text{GLO}_N(\mu, \sigma_1, \sigma_2, \gamma)$ one GLO process $\text{GLO}(\mu, \sigma_1, \sigma_2, \gamma)$ with the property that every event $S_i \in [0, T]$ originates from one of N beats. Furthermore, we interpret GLO_N as the *hypothesis* that the (observed) events $\mathcal{S} = (S_1, ..., S_n)$ originate from N beats and are distributed according to a GLO process. There exist several possibilities how to evaluate different hypotheses $\text{GLO}_{N_1}, ..., \text{GLO}_{N_h}$, which we will shortly review in the following:

A) **Bayes Factors:**
The standard Bayesian solution to test and select one hypothesis GLO_{N_1} against its complementary hypothesis GLO_{N_0} is to compute the Bayes factor (cf. Lewis and Raftery (1997))

$$B = \left(\frac{p(\text{GLO}_{N_1} \mid \mathcal{S})}{p(\text{GLO}_{N_0} \mid \mathcal{S})}\right) \Big/ \left(\frac{p(\text{GLO}_{N_1})}{p(\text{GLO}_{N_0})}\right) = \frac{p(\mathcal{S} \mid \text{GLO}_{N_1})}{p(\mathcal{S} \mid \text{GLO}_{N_0})}, \tag{3.75}$$

which can be seen as the likelihood ratio of the two hypotheses. Essentially, it describes if the observations $\mathcal{S}$ have increased or decreased the odds on hypothesis GLO_{N_1} relative to hypothesis GLO_{N_0} (cf. Chu and Zhao (2011, p.245)).

However, for the GLO model it is rather difficult to evaluate (3.75) since the integral (which is often referred to as *evidence* for GLO_N given $\mathcal{S}$)

$$Z = p(\mathcal{S} \mid \text{GLO}_N) = \int_{\Theta} L(\vartheta; \mathcal{S} \mid \text{GLO}_N)\, \pi(\vartheta \mid \text{GLO}_N)\, d\vartheta, \tag{3.76}$$

where we denote by $L(\vartheta; \mathcal{S} \mid \text{GLO}_N) = p(\mathcal{S} \mid \vartheta, \text{GLO}_N)$ the likelihood function under hypothesis GLO_N, can not be solved analytically. However, note that $Z = \mathbb{E}_{\pi(\vartheta|\text{GLO}_N)}\left[L(\vartheta; \mathcal{S} \mid \text{GLO}_N)\right]$. Thus, a simple Monte Carlo estimator may be given by

$$\hat{Z} = \frac{1}{R} \sum_{r=1}^{R} L\left(\vartheta^{(r)}; \mathcal{S} \mid \text{GLO}_N\right), \tag{3.77}$$

where $\vartheta^{(1)}, ..., \vartheta^{(R)}$ are samples from the joint prior distribution $\pi(\vartheta \mid \text{GLO}_N)$. However, this estimator is prone to extreme estimation errors and, thus, requires huge computational effort (see Lewis and Raftery (1997) and references therein).

Moreover, it is not straightforward to use MCMC methods to estimate (3.76) since the MCMC algorithm aims to draw samples of ϑ that are distributed as $p(\vartheta \mid \mathcal{S}, \text{GLO}_N)$. Nonetheless, Newton and Raftery (1994) point out that

$$L(\vartheta; \mathcal{S} \mid \text{GLO}_N) \times \pi(\vartheta \mid \text{GLO}_N) = Z \times p(\vartheta \mid \mathcal{S}, \text{GLO}_N). \tag{3.78}$$

By rearranging (3.78) they yield the harmonic mean of the samples $\vartheta^{(1)}, ..., \vartheta^{(R)}$ that are drawn from $p(\vartheta \mid \mathcal{S}, \text{GLO}_N)$ as an estimator for Z, e.g.

$$\hat{Z}^* = \left[\frac{1}{R} \sum_{r=1}^{R} \frac{1}{L\left(\vartheta^{(r)}; \mathcal{S} \mid \text{GLO}_N\right)} \right]^{-1}. \tag{3.79}$$

However, $\hat{Z}^*$ is prone to domination by a few outlying terms with abnormally small values of $L\left(\vartheta^{(r)}; \mathcal{S} \mid \text{GLO}_N\right)$. Exemplary simulations indicate that due to this reason the harmonic mean estimator is not of particular use for the GLO model. For a proper discussion about the harmonic mean estimator and other more elaborate estimators for Z that employ MCMC methods we refer to Weinberg (2012) and references therein. A different approach to estimate Z is given by RJMCMC methods:

B) **Reversible jump Markov chain Monte Carlo (RJMCMC) algorithms:**
Green (1995) first introduced the RJMCMC algorithm as an integrative approach to deal with the hypothesis selection problem. The main idea of the algorithm is the following: Additionally to a simple MCMC algorithm the RJMCMC *jumps* between

different parameter spaces. To contain the dimension-mismatch for different models $\mathrm{GLO}_{N_1}, ..., \mathrm{GLO}_{N_h}$ additional pseudo random variables are introduced. During the simulation the algorithm proposes visits to other hypotheses and accepts these visits analogously to the MH algorithm (cf. section 2.3.2). Then, $p(\mathrm{GLO}_N \mid \mathcal{S})$ can be approximated by the time the simulated Markov chain stays in hypothesis GLO_N.

For example, Chu and Zhao (2011) apply the RJMCMC algorithm for a change-point model. In their model events occur according to a homogeneous Poisson process with intensity λ. However, λ is able to change at k change-points, thus, under each hypothesis H_k, $k = 1, ..., h$, there exist k change-points. The locations of the change-points as well as the Bayes factors are estimated via a RJMCMC algorithm. For a more detailed introduction to RJMCMC methods see e.g. Hastie and Green (2012).

In the context of the GLO model one may e.g. propose an additional beat near the largest burst (to increase N) or propose to delete the beat with the smallest family (to decrease N). Nonetheless, implementing the RJMCMC algorithm does not only require to carefully choose additional quantities which affect the mixing properties of the algorithm, as for example a proposal distribution. It is also related to supplementary computational effort.

C) **Means for model selection:**
A rather straightforward method is given by, firstly, applying the Bayesian estimation procedure (Algorithm 3.1) for several fixed $N_1, ..., N_h$ and, secondly, computing measures that assess the *Goodness of Fit* of the models $\mathrm{GLO}_{N_1}, ..., \mathrm{GLO}_{N_h}$. Then, the respective number of beats should be chosen that maximizes the *Goodness of Fit*.

The natural measure for the *Goodness of Fit* in Bayesian models is the aforementioned Bayes Factor. However, we may also consider means for model selection from the *frequentist* world, in which ϑ is seen as a fixed parameter with estimate $\hat{\vartheta}$. In this world the central quantity for model selection is the (maximized) likelihood function: For example, Akaike's information criterion (AIC) is given as

$$AIC\left(\mathrm{GLO}_N\right) = -2\left(\log L\left(\hat{\vartheta}; \mathcal{S} \mid \mathrm{GLO}_N\right) - \log L\left(\hat{\vartheta}_s; \mathcal{S} \mid \mathrm{GLO}_s\right)\right) + 2p, \tag{3.80}$$

where p is the number of estimated parameters in the model and GLO_s is the full (saturated) model (cf. deLeeuw (1992)). The first part of the AIC is called *deviance*, i.e.

$$D = -2\left(\log L\left(\hat{\vartheta}; \mathcal{S} \mid \mathrm{GLO}_N\right) - \log L\left(\hat{\vartheta}_s; \mathcal{S} \mid \mathrm{GLO}_s\right)\right), \tag{3.81}$$

and assesses the fit of the model to the data, whereas the second part ($2p$) accounts for model complexity. From the perspective of this measure the smaller the AIC is, the better is the *Goodness of Fit*.

Spiegelhalter et al. (2002) extend the idea of the AIC to Bayesian inference and propose to use the *deviance information criterion* (DIC) for Bayesian model selection, i.e.

$$DIC\left(\mathrm{GLO}_N\right) = \mathbb{E}_{\vartheta|\mathcal{S}}\left[D\right] + p_D, \tag{3.82}$$

where in this context $D = -2\left(\log L\left(\vartheta; \mathcal{S} \mid \mathrm{GLO}_N\right) - \log L\left(\hat{\vartheta}_s; \mathcal{S} \mid \mathrm{GLO}_s\right)\right)$ is a random variable and p_D a measure for the complexity of the model (the analog to $2p$ in the AIC). Hence, the DIC can be interpreted as the a posteriori expected deviance modified by an additional complexity parameter.

Since $DIC\left(\mathrm{GLO}_N\right)$ is straightforward to estimate with MCMC samples, we will rely on this measure to choose a value for N. However, the additive term $\log L\left(\hat{\vartheta}_s; \mathcal{S} \mid \mathrm{GLO}_s\right)$ is irrelevant for model selection, thus, we will drop this term.

Moreover, we assume that only small changes in the value of N will be considered (because a first estimate may easily be found by visual inspection or Algorithm 3.3). Since these changes will be small relative to the total number of parameters (which is $4 + N + n$), the changes of the model complexity, p_D, would be small. Therefore, and for the sake of simplicity, we will drop p_D. In conclusion, the choice of N will be based on the maximization of the expected log-likelihood, i.e.

$$\mathbb{E}_{\vartheta|\mathcal{S}}\left[\log L(\vartheta; \mathcal{S} \mid \mathrm{GLO}_N)\right], \tag{3.83}$$

which will be approximate with samples $\vartheta^{(1)}, ..., \vartheta^{(R)}$ from the posterior distribution by

$$\overline{\ell}(\mathrm{GLO}_N) := \frac{1}{R} \sum_{r=1}^{R} \log L\left(\vartheta^{(r)}; \mathcal{S} \mid \mathrm{GLO}_N\right). \tag{3.84}$$

3.3. Choice of priors

3.3.1. Non-informative and informative priors

Generally, one distinguishes between proper and improper prior distributions: While the first are proper distributions on their own, the latter may not integrate to one, like e.g. $\pi = \mathbb{1}_{(0,\infty)}$. However, improper priors may still lead to proper posterior distributions. As an example, the posterior distribution for the mean μ of a normal distribution with known variance σ^2 and prior $\pi(\mu) \propto 1$ is given as

$$\begin{aligned} p(\mu \mid \mathcal{S}) &= \frac{L(\mu; \mathcal{S})\pi(\mu)}{\int_{-\infty}^{\infty} L(v; \mathcal{S})\pi(v)dv} \\ &= \frac{1}{\int_{-\infty}^{\infty} \prod_{i=1}^{n} f_{\mathcal{N}(v,\sigma^2)}(S_i)dv} \prod_{i=1}^{n} f_{\mathcal{N}(\mu,\sigma^2)}(S_i). \end{aligned} \tag{3.85}$$

Improper priors are mostly motivated by the idea of *non-informative priors*. A very intuitive non-informative prior is given by the *flat prior* $\pi = \mathbb{1}_{\Theta}$. However, flat priors are not

necessarily the most non-informative priors. For a discussion about the non-informativeness of (flat) priors we refer to Zhu and Lu (2004). It may also be desirable for a non-informative prior to be invariant under reparametrization. For this reason, Jeffreys (1946) showed that the use of

$$\pi(\vartheta) \propto |I(\vartheta)|^{1/2}, \tag{3.86}$$

where $I(\vartheta)$ is Fisher's expected information, leads to a prior that is in fact invariant under reparametrization of ϑ. This prior is called *Jeffreys prior*. Other families of prior distributions include e.g. the Maximum Entropy Prior, that maximizes uncertainty (cf. Koch (2007, p.57-58)). For an overview of non-informative priors we refer to Yang and Berger (1998).

Still, the informativeness of informative priors can also be controlled: In case of a Normal prior the prior is more informative (i.e. it gains more weight in the posterior distribution) the smaller the prior variance is (cf. Example 1, p. 23), whereas for Gamma priors the prior is more informative the larger the parameters are. An example for the latter is given by the FCPD of γ in the GLO model, i.e. (cf. Lemma 3.1)

$$p(\gamma \mid \vartheta_{-\gamma}, \mathcal{S}) = Ga\left(\alpha_\gamma + n,\ \beta_\gamma + \sum_{k=1}^{N} \nu_k\right). \tag{3.87}$$

Analogously to the Normal distribution, the less data there are (the smaller n and $\sum_{k=1}^{N} \nu_k$) and the larger the prior parameters α_γ and β_γ are, the more influence the data has on the posterior distribution and e.g. on the conditional posterior mean, which is

$$\mathbb{E}[\gamma \mid \vartheta_{-\gamma}, \mathcal{S}] = \frac{\alpha_\gamma + n}{\beta_\gamma + \sum_{k=1}^{N} \nu_k} = \frac{\beta_\gamma}{\beta_\gamma + \sum_{k=1}^{N} \nu_k} \underbrace{\frac{\alpha_\gamma}{\beta_\gamma}}_{=\mathbb{E}[\gamma]} + \frac{\sum_{k=1}^{N} \nu_k}{\beta_\gamma + \sum_{k=1}^{N} \nu_k} \frac{n}{\sum_{k=1}^{N} \nu_k}. \tag{3.88}$$

Hence, the posterior mean converges to the prior mean, $\mathbb{E}[\gamma \mid \vartheta_{-\gamma}, \mathcal{S}] \to \mathbb{E}[\gamma]$, for $\beta_\gamma \to \infty$ and the prior influence vanishes, $\mathbb{E}[\gamma \mid \vartheta_{-\gamma}, \mathcal{S}] \to \frac{n}{\sum_{k=1}^{N} \nu_k}$, for $\beta_\gamma \to 0$.

In Table 3.1 the conjugate and Jeffreys priors and the resulting posterior distributions for μ, ψ_1, ψ_2, γ and B_1 are reported (as references we refer to Yang and Berger (1998) and Koch (2007, p.57)). Note that for ψ_2 and B_1 the posterior distribution denotes the sampling distribution for the second step of the Gibbs sampling procedure described in Lemma 3.4 and Lemma 3.5, respectively. Although the Jeffreys prior for ψ_1, ψ_2 and γ exhibits a similar form as the Gamma distribution, it does not integrate to one. Nonetheless, it is possible to use the reported parameters as input parameters in Algorithm 3.1 since the resulting posterior distributions are proper distributions. The fully specified parameters for the conjugate prior and respective posterior distributions can be found in Lemmas 3.2 to 3.5. Note that due to the invariance property of the Jeffreys prior in case of $\pi(\psi_1) \propto \psi_1^{-1}$ and $\pi(\psi_2) \propto \psi_2^{-1}$ we have $\pi(\sigma_1) \propto \sigma_1^{-1}$ and $\pi(\sigma_2) \propto \sigma_2^{-1}$, respectively.

Parameter	**Conjugate**		**Jeffreys**	
	Prior	**FCPD**	**Prior**	**FCPD**
μ	$\mathcal{N}(\mu_\mu, \sigma_\mu^2)$	$\mathcal{N}(a_\mu, b_\mu)$	$\pi(\mu) \propto 1$	$\mathcal{N}\left(\frac{\sum(B_k - B_{k-1})}{N-1}, \frac{1}{\psi_1(N-1)}\right)$
ψ_1	$Ga(\alpha_{\psi_1}, \beta_{\psi_1})$	$Ga(a_{\psi_1}, b_{\psi_1})$	$\pi(\psi_1) \propto \psi_1^{-1} \approx Ga(0,0)$	$Ga\left(\frac{N-1}{2}, \frac{1}{2}\sum_{k=2}^{N}(B_k - B_{k-1} - \mu)^2\right)$
ψ_2	$Ga(\alpha_{\psi_2}, \beta_{\psi_2})$	$Ga(a_{\psi_2}, b_{\psi_2})$	$\pi(\psi_2) \propto \psi_2^{-1} \approx Ga(0,0)$	$Ga\left(\frac{n}{2}, \frac{1}{2}\sum_{i=1}^{n}(z_i - B_{J_i})^2\right)$
γ	$Ga(\alpha_\gamma, \beta_\gamma)$	$Ga(a_\gamma, b_\gamma)$	$\pi(\gamma) \propto \gamma^{-1/2} \approx Ga\left(\frac{1}{2}, 0\right)$	$Ga\left(1/2 + n, \sum_{k=1}^{N}\nu_k\right)$
B_1	$\mathcal{N}(\mu_{B_1}, \sigma_{B_1}^2)$	$\mathcal{N}(a_{B_1}, b_{B_1})$	$\pi(B_1) \propto 1$	$\mathcal{N}\left(\frac{\psi_1(B_2-\mu)+\psi_2\sum_{j=1}^{n_1} z_j}{\psi_1+n_1\psi_2}, \frac{1}{\psi_1+n_1\psi_2}\right)$

Table 3.1.: Conjugate and Jeffreys prior and posterior distributions for μ, ψ_1, ψ_2, γ and B_1.

3.3.2. Priors for the labels

To complete the Bayesian estimation framework we also need to assign a prior to the labels $\mathcal{J} = (J_1, ..., J_n)$. For this purpose, we assume that the labels are a priori independent from each other and from all other parameters, i.e.

$$\pi(\mu, \psi_1, \psi_2, \gamma, B_1, J_1, ..., J_n) = \pi(\mu)\,\pi(\psi_1)\,\pi(\psi_2)\,\pi(\gamma)\,\pi(B_1)\,\pi(J_1) \times ... \times \pi(J_n). \tag{3.89}$$

This assumption simplifies the calculations substantially, although, it may not completely be fulfilled in the GLO model since other parameter values might provide information about the labels. For example, if $\sigma_2/\mu = \psi_2^{-1/2}/\mu$ and $\sigma_1/\mu = \psi_1^{-1/2}/\mu$ are very small, the labels $J_1, ..., J_n$ will very likely be monotone increasing, which may not be true otherwise. Nevertheless, simulation studies suggest that the independence assumption still leads to reasonable estimators (cf. section 3.4).

A proper prior for label J_i should take values in $\{1, ..., N\}$. Moreover, it seems natural that a priori $\mathbb{E}[J_i]$ should be increasing with increasing i, i.e. $\mathbb{E}[J_i] \leq \mathbb{E}[J_{i+1}]$, since we assume the observations to be sorted, i.e. $S_i \leq S_{i+1}$. Therefore, a binomial prior with parameters $Bin\left(N, \frac{i}{n}\right)$ might represent an appropriate choice, i.e.

$$\pi(J_i) = \binom{N}{J_i} \left(\frac{i}{n}\right)^{J_i} \left(\frac{n-i}{n}\right)^{N-J_i}, \tag{3.90}$$

which yields $\mathbb{E}[J_i] = \frac{iN}{n}$. However, this prior restricts the value of J_n to N a.s.. On the one hand, this may be an advantage since it anchors the last observation S_n to the last beat B_N. On the other hand, if N is too large, B_N can not be pushed out of the observation window $[0, T]$.

An alternative to the binomial prior is given by a uniform prior, $J_i \sim unif(\{1, ..., N\})$, which might also be interpreted as *non-informative* prior. In general, it is also possible to simply assign arbitrary probability weights $w_{i,1}, ..., w_{i,N}$ with $\sum_{l=1}^{N} w_{i,l} = 1$ for every label J_i, $i \in \{1, ..., n\}$.

3.4. Estimation characteristics and precision

The estimation procedure depends on several input parameters, namely (the parameters of) the prior distributions, initial values for the parameters, $\vartheta^{(0)}$, and the number of beats, N. In this section we will analyze the influence of these quantities on the parameter estimates and the mixing behavior of the MCMC algorithm. Moreover, we will analyze the general estimation precision of the Bayesian estimates.

To enhance comparability between the estimation errors, we will focus on the *mean absolute relative error* (MRE):

For an estimator $\hat{\eta} = (\hat{\eta}_1, ..., \hat{\eta}_d)$ with realizations $\hat{\eta}^{(1)}, ..., \hat{\eta}^{(M)}$ and true parameter value $\eta_0 = (\eta_{1,0}, ..., \eta_{d,0})$ the MRE of component $p \in \{1, ..., d\}$ is given as

$$\text{MRE}(\hat{\eta}_p) = \frac{1}{M} \sum_{m=1}^{M} \frac{\left| \hat{\eta}_p^{(m)} - \eta_{p,0} \right|}{\eta_{p,0}}. \tag{3.91}$$

Moreover, the *overall relative error* (ORE) of $\hat{\eta}$ is given as the sum of its components' MREs, i.e.

$$\text{ORE}(\hat{\eta}) = \sum_{p=1}^{d} \text{MRE}(\hat{\eta}_p). \tag{3.92}$$

For the remaining work we will particularly focus on two very intuitive estimators arising from the posterior distribution, namely the a posteriori expectation and median value of the parameters. These are given as

$$\hat{\eta}_{\text{Mean}} = (\hat{\mu}_{\text{Mean}}, \hat{\sigma}_{1,\text{Mean}}, \hat{\sigma}_{2,\text{Mean}}, \hat{\gamma}_{\text{Mean}}) \tag{3.93}$$

$$= \left(\mathbb{E}[\mu \mid \mathcal{S}],\ \mathbb{E}[\sigma_1 \mid \mathcal{S}],\ \mathbb{E}[\sigma_2 \mid \mathcal{S}],\ \mathbb{E}[\gamma \mid \mathcal{S}] \right) \tag{3.94}$$

$$\text{and } \hat{\eta}_{\text{Median}} = (\hat{\mu}_{\text{Median}}, \hat{\sigma}_{1,\text{Median}}, \hat{\sigma}_{2,\text{Median}}, \hat{\gamma}_{\text{Median}}) \tag{3.95}$$

$$= \left(p^{-1}_{\mu|\mathcal{S}}(0.5),\ p^{-1}_{\sigma_1|\mathcal{S}}(0.5),\ p^{-1}_{\sigma_2|\mathcal{S}}(0.5),\ p^{-1}_{\gamma|\mathcal{S}}(0.5) \right). \tag{3.96}$$

To simulate GLO processes and, thus, realizations of the aforementioned estimators we will employ Algorithm 1 from Bingmer (2012, p.41 and p.137), which was already reviewed in section 2.1. Since we can save the original (simulated) beat locations $\mathcal{B}_0$ and labels $\mathcal{J}_0$, we are able to use these to calculate (sample) moment estimates (ME) for the parameters, which are given as

$$\hat{\mu}_{\text{ME}} = \frac{1}{N-1} \sum_{k=2}^{N} \left(B_{k,0} - B_{k-1,0} \right), \tag{3.97}$$

$$\hat{\sigma}_{1,\text{ME}} = \sqrt{\frac{1}{N-2} \sum_{k=2}^{N} \left(B_{k,0} - B_{k-1,0} - \hat{\mu}_{\text{ME}} \right)^2}, \tag{3.98}$$

$$\hat{\sigma}_{2,\text{ME}} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(S_i - B_{J_{i,0},0} \right)^2}, \tag{3.99}$$

$$\hat{\gamma}_{\text{ME}} = \frac{1}{N} \sum_{k=1}^{N} \left| \{ i \in \{1, ..., n\} : J_{i,0} = k \} \right|. \tag{3.100}$$

Since the moment estimates base on the true beats and labels, they represent benchmark parameter estimates and allow us to evaluate other parameter estimates.

Furthermore, in this section we will compare Bayesian posterior means $\hat{\eta}_{\text{Mean}}$ and medians $\hat{\eta}_{\text{Median}}$ with the ACH estimates from Bingmer et al. (2011). For this purpose we will calculate ACH estimates $\hat{\eta}_{\text{ACH}}$ with Algorithm A.10 from Bingmer (2012, p.141) with a bin size of $\delta = 10$ and analysis window $(0, 25, ..., 500)$ to incorporate the first peak of the ACH (we refer to Bingmer (2012, p.80ff.) for more details about the ACH-fitting procedure). In the following these input parameters will remain fixed.

Moreover, we will set $\Delta^{(switch)} = 2$ and $p^{(switch)} = 0.9$ for the MH updates of the labels in the MCMC algorithm. Thus, in every iteration we update $n^{(switch)} = \max\{1, \lfloor np^{(switch)} \rceil\}$ labels and during every update we propose a switch from beat B_k to $B_{k'}$ with $k' \sim unif(V_k)$ and $V_k = \{\max\{k-2, 1\}, ..., \min\{k+2, N\}\}$. These values for $\Delta^{(switch)}$ and $p^{(switch)}$ are calibrated to yield an average acceptance probability of roughly $1/4$ for typical processes analyzed in this work as proposed by Gelman et al. (1996) for MH algorithms for Normal models (cf. section 2.3.3). However, the acceptance probability varies a lot since for regular processes the labels are likely to remain constant during the simulation, whereas for irregular processes there is far more uncertainty in the labels. The proposed values for $\Delta^{(switch)}$ and $p^{(switch)}$ will remain unchanged for all next examples and the data analysis in chapter 4.

In the following, we will, firstly, analyze three exemplary GLO processes that exhibit a different level of regularity. Focussing on the ORE of the posterior means and medians and the mixing properties of the MCMC samples, we will assess the impact of the prior distribution, the number of beats and the impact of initial values. Secondly, we will examine the general estimation error of posterior means and medians for nine different parameter combinations based on a simulation study. We start by giving an example of a perfectly working estimation procedure:

Example 3. *In this example we analyze events from a GLO process with parameters* $\eta_0 = (\mu_0, \sigma_{1,0}, \sigma_{2,0}, \gamma_0) = (364, 20, 10, 10)$ *in the time window* $[0, 13000]$. *This particular GLO instance, which we call Test-Data 1, exhibits* $N_0 = 36$ *beats and* $n = 351$ *events. A rasterplot is shown in Figure 3.1. One can clearly distinguish between different bursts due to a very regular background beat and a small burst width.*

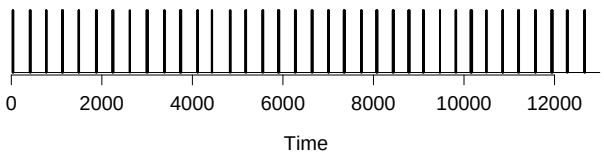


Figure 3.1.: Test-Data 1: Rasterplot.

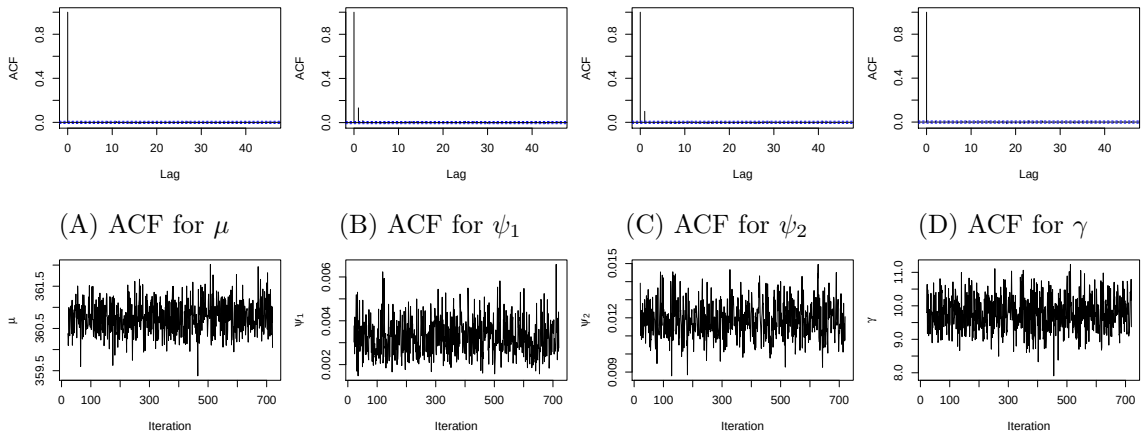
For the simulated events we employ the MCMC algorithm (Algorithm 3.1) with Jeffreys priors for all parameters μ, ψ_1, ψ_2, γ, B_1 *and Binomial priors for the labels* $\mathcal{J}$ *(cf. section 3.3). The true parameter values and original beat locations and labels are taken as initial values to draw* $R = 40000$ *posterior samples for* ϑ. *An overview of the input parameters is given in Table 3.2. Moreover, we report the burn-in period,* BI, *and computation time (in*

minutes) for the MCMC procedure on a Mac Mini with 2.6 GHz Intel Core i7, 16 GB RAM and Mac OS X Yosemite 10.10.4.

Prior for $\mu, \psi_1, \psi_2, \gamma, B_1$	Jeffreys
Prior for $\mathcal{J}$	Binomial
$p^{(switch)}$	0.9
$\Delta^{(switch)}$	2
No. of iterations, R	40000
burn-in period, BI	20
Computation time	31 min

Table 3.2.: Example 3: Input parameters, burn-in period and computation time.

As shown in Figure 3.2 the posterior samples exhibit very small levels of autocorrelation and a visually very well mixing chain. The initial labels remain constant during the full simulation, which is reasonable since we already initialized the algorithm with the true (original) labels and parameter values.



(A) ACF for μ (B) ACF for ψ_1 (C) ACF for ψ_2 (D) ACF for γ

(E) Trace plot for μ (F) Trace plot for ψ_1 (G) Trace plot for ψ_2 (H) Trace plot for γ

Figure 3.2.: Example 3: Test-Data 1: (A)-(D) Sample ACFs and (E)-(H) trace plots for $\mu, \psi_1, \psi_2, \gamma$.

Method/Estimate	$\hat{\mu}$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\gamma}$	ORE
True values, η_0	364	20	10	10	-
Moment estimates, $\hat{\eta}_{\text{ME}}$	360.81	19.64	9.30	9.75	0.12
Posterior means, $\hat{\eta}_{\text{Mean}}$	360.76	18.07	9.20	9.76	0.21
Posterior medians, $\hat{\eta}_{\text{Median}}$	360.76	17.87	9.19	9.75	0.22
ACH estimates, $\hat{\eta}_{\text{ACH}}$	360.33	10.31	12.12	7.33	0.97

Table 3.3.: Test-Data 1: Parameter estimates in Example 3.

In Table 3.3 the posterior parameter means $\hat{\eta}_{Mean}$ and medians $\hat{\eta}_{Median}$ (computed with all samples after the burn-in period $BI = 20$), moment estimates (ME) and the ACH estimates are reported. Clearly, both, posterior means and medians, provide almost identical, very good estimates with a smaller overall estimation error than the ACH estimates.

Example 4. *In this example we analyze events from a GLO process with parameters $\eta_0 = (\mu_0, \sigma_{1,0}, \sigma_{2,0}, \gamma_0) = (364, 20, 50, 10)$ in the time window $[0, 13000]$. This particular GLO instance, which we call Test-Data 2, exhibits $N_0 = 37$ beats and $n = 349$ events. The rasterplot in Figure 3.3 shows that the main difference to Test-Data 1 is a larger burst width.*

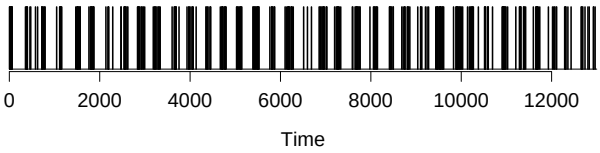


Figure 3.3.: Test-Data 2: Rasterplot.

For the Bayesian estimation we follow the same procedure with the same input parameters as in Example 3. Particularly, we assume Jeffreys priors for all parameters and Binomial priors for the labels.

Prior for $\mu, \psi_1, \psi_2, \gamma, B_1$	Jeffreys
Prior for $\mathcal{J}$	Binomial
$p^{(switch)}$	0.9
$\Delta^{(switch)}$	2
No. of iterations, R	40000
Burn-in period, BI	20
Computation time	32.3 min

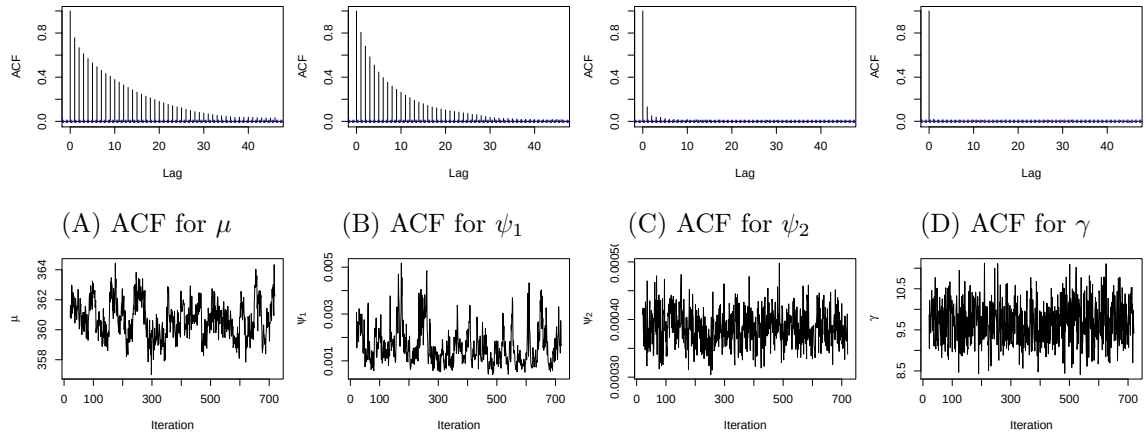
Table 3.4.: Example 4: Input parameters, burn-in period and computation time.

Similarly as in Example 3 the sample ACFs of the posterior samples for ψ_2 and γ exhibit a rapidly declining structure. However, the sample ACFs for μ, ψ_1 and all beat locations $\mathcal{B}$ indicate a rather slow mixing chain (cf. Figures 3.4 and 3.5). (Note that "autocorrelation of the samples" always corresponds to the autocorrelation of the MCMC (posterior) samples $\vartheta_j^{(BI+1)}, ..., \vartheta_j^{(R)}$ of one parameter ϑ_j after a burn-in period of length $BI \in \mathbb{N}, BI < R$.)

The reason for this increased autocorrelation is given by the updating procedure for the beats: As described in Algorithm 3.1, the (r+1)-th sample for beat B_k is distributed as

$$B_k^{(r+1)} \sim \mathcal{N}\left(\frac{\psi_1^{(r+1)} \left(B_{k+1}^{(r)} + B_{k-1}^{(r+1)} \right) + \psi_2^{(r+1)} \sum_{j=1}^{n_k^{(r)}} z_j}{2\psi_1^{(r+1)} + n_k^{(r)} \psi_2^{(r+1)}}, \frac{1}{2\psi_1^{(r+1)} + n_k^{(r)} \psi_2^{(r+1)}} \right),$$

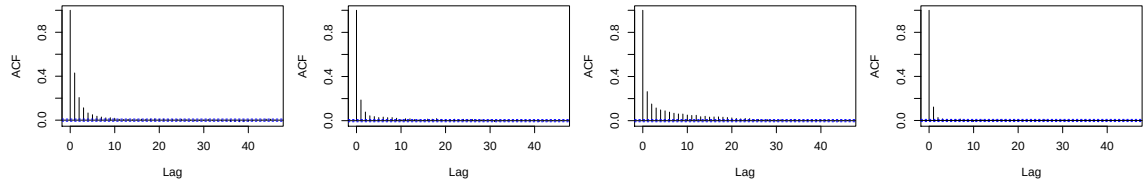
where z_j are the rescaled event locations (cf. Lemma 3.5, p. 3.5). Clearly, if σ_2 increases, $\psi_2 = \sigma_2^{-2}$ decreases and the expected value $\mathbb{E}\left[B_k^{(r+1)} \mid \vartheta_{-B_k}, \mathcal{S}\right]$ depends more on the surrounding beats $B_{k+1}^{(r)}$ and $B_{k-1}^{(r+1)}$. However, the latter are updated in an analogous way. Therefore, the expected value of $B^{(r+1)}$ depends more on $B^{(r)}$, thus, the autocorrelation of the beat locations increases. Since the FCPDs of μ and ψ_1 only depend on the beat locations and prior parameters, their autocorrelations increase as well.



(A) ACF for μ (B) ACF for ψ_1 (C) ACF for ψ_2 (D) ACF for γ

(E) Trace plot for μ (F) Trace plot for ψ_1 (G) Trace plot for ψ_2 (H) Trace plot for γ

Figure 3.4.: Example 4: Test-Data 2: (A)-(D) Sample ACFs and (E)-(H) trace plots for $\mu, \psi_1, \psi_2, \gamma$.



(A) ACF for B_1 (B) ACF for B_{11} (C) ACF for B_{21} (D) ACF for B_{31}

Figure 3.5.: Example 4: Test-Data 2: Sample ACF for (A) B_1, (B) B_{11}, (C) B_{21}, (D) B_{31}.

In Table 3.5 the posterior parameter means and medians (computed with all samples after a burn-in period of $BI = 20$ samples), moment estimates and ACH estimates are reported. As in Example 3, both the posterior means and medians for σ_2 and γ are very good estimates. Moreover, despite the high autocorrelations, this is also true for the posterior mean and median of μ. Nevertheless, the estimate for σ_1 exhibits an increased estimation error. This observation is typical for high values of σ_1 and may be due to the high autocorrelation resulting as described above or due to the prior parameters (cf. section 3.4.1).

Method/Estimator	$\hat{\mu}$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\gamma}$	ORE
True values, η_0	364	20	50	10	-
Moment estimates, $\hat{\eta}_{\text{ME}}$	361.81	18.87	51.67	9.43	0.15
Posterior means, $\hat{\eta}_{\text{Mean}}$	361.19	24.37	51.33	9.71	0.28
Posterior medians, $\hat{\eta}_{\text{Median}}$	361.32	23.76	51.27	9.70	0.25
ACH estimates, $\hat{\eta}_{\text{ACH}}$	367.39	0.002	59.35	8.95	1.30

Table 3.5.: Test-Data 2: Parameter estimates in Example 4.

Example 5. *In this example we analyze events from a GLO process with parameters* $\eta_0 = (\mu_0, \sigma_{1,0}, \sigma_{2,0}, \gamma_0) = (364, 5, 200, 5)$ *in the time window* $[0, 13000]$. *This particular GLO instance, which we call Test-Data 3, exhibits* $N_0 = 39$ *beats and* $n = 185$ *events. A rasterplot is shown in Figure 3.6. In contrast to Test-Data 2 the background rhythm is even more regular while the events are distributed much more irregularly.*

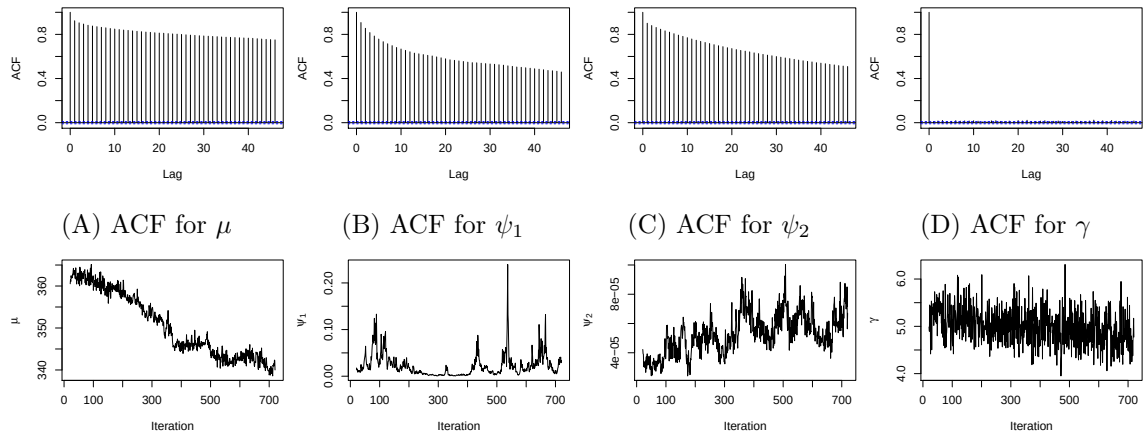
Figure 3.6.: Test-Data 3: Rasterplot.

For the Bayesian estimation we follow the same procedure with the same input parameters as in the last two Examples 3 and 4. Particularly, we assume Jeffreys priors for all parameters and Binomial priors for the labels.

Prior for $\mu, \psi_1, \psi_2, \gamma, B_1$	Jeffreys
Prior for $\mathcal{J}$	Binomial
$p^{(switch)}$	0.9
$\Delta^{(switch)}$	2
No. of iterations, R	40000
Burn-in period, BI	1250
Computation time	17.6 min

Table 3.6.: Example 5: Input parameters, burn-in period and computation time.

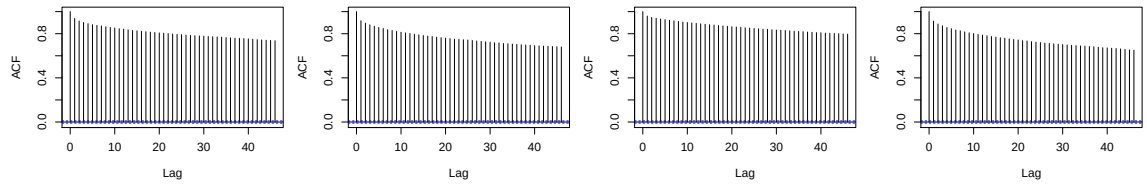
However, in this example only the γ*-component exhibits good mixing properties. All other parameters* μ, ψ_1, ψ_2 *and all beat locations* $\mathcal{B}$ *exhibit large autocorrelations (cf. Figures 3.7 and 3.8). Moreover, we can observe a very large burn-in period* $BI = 1250$ *which results from the beat locations moving together at the beginning of the simulation. Apparently, the posterior probability is larger for beats being more close to each other. In Table 3.7 the posterior parameter means and medians (computed with all samples after the burn-in period), moment estimates (ME) and ACH estimates are reported.*



(A) ACF for μ (B) ACF for ψ_1 (C) ACF for ψ_2 (D) ACF for γ

(E) Trace plot for μ (F) Trace plot for ψ_1 (G) Trace plot for ψ_2 (H) Trace plot for γ

Figure 3.7.: Example 5: Test-Data 3: (A)-(D) Sample ACFs and (E)-(H) trace plots for $\mu, \psi_1, \psi_2, \gamma$.



(A) ACF for B_1 (B) ACF for B_{11} (C) ACF for B_{21} (D) ACF for B_{31}

Figure 3.8.: Example 5: Test-Data 3: Sample ACF for (A) B_1, (B) B_{11}, (C) B_{21}, (D) B_{31}.

Analogously to the mixing properties only γ is estimated well. Regarding the other parameters, σ_1 is overestimated and, thus, its estimates capture some variability that is missing in in the estimates for σ_2, which is underestimated. The only difference between posterior means and medians is the posterior median of σ_1, which is slightly smaller. Thus, the ORE for $\hat{\eta}_{Median}$ is slightly smaller. In comparison to the ACH estimates we see no improvement in terms of the overall estimation error.

Method/Estimate	$\hat{\mu}$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\gamma}$	ORE
True values, η_0	364	5	200	5	-
Moment estimates, $\hat{\eta}_{\text{ME}}$	363.32	6.28	191.98	4.74	0.35
$\hat{\eta}_{\text{Mean}}$	332.86	28.81	129.82	4.79	5.24
$\hat{\eta}_{\text{Median}}$	332.27	19.53	127.08	4.78	3.40
ACH	312.66	0.68	93.71	4.23	1.69

Table 3.7.: Test-Data 3: Parameter estimates in Example 5.

Nonetheless, the reported point estimates exclude vital information about the posterior probabilities: Clearly, a posteriori there are two areas with a high density for σ_1. These are approximately $[0, 20]$ and $[30, 60]$, as Figure 3.9 shows. In other words, a posteriori it is unsure if σ_1 is small or large. This phenomenon might occur due to a discrepancy between the prior distribution and the data. To address this possibility we will change the prior distribution in Example 6.

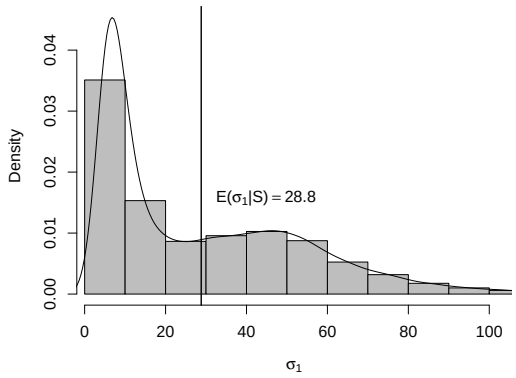


Figure 3.9.: Example 5: Test-Data 3: Posterior distribution of σ_1.

3.4.1. Influence of prior distributions

As it is shown in section 3.3 the impact of the prior distributions for $\mu, \psi_1, \psi_2, \gamma, B_1$ on the posterior distribution depends on the parameters of the prior distribution: For the Normal distribution the impact is larger the smaller the prior variance is, whereas for the Gamma distribution the impact is larger the larger the prior parameters are. To restrict the prior influence one may choose the (*non-informative*) Jeffreys prior. However, if we assume a Jeffreys prior for both ψ_1 and ψ_2, we a priori think of σ_1 and σ_2 as being identically distributed. As a comparison between Examples 5 and 6 shows, this may result in large autocorrelations and estimation errors, which are smaller for different prior parameters:

Example 6. *In this example we analyze Test-Data 3 and follow the same procedure as in Example 5 with exactly one difference: For the Gamma prior of ψ_1 we set $\alpha_{\psi_1} = 10$ instead of $\alpha_{\psi_1} = 0$. We can a priori informally think of ψ_1 to be $Ga(\alpha_{\psi_1}, \beta_{\psi_1})$ distributed and ψ_2 to be $Ga(\alpha_{\psi_2}, \beta_{\psi_2})$ distributed. Since for $Y \sim Ga(\alpha, \beta)$ the expected value is given by $\mathbb{E}[Y] = \alpha/\beta$, with $\alpha_{\psi_1} > \alpha_{\psi_2}$ and $\beta_{\psi_1} = \beta_{\psi_2}$ we a priori expect ψ_1 to exceed ψ_2, which in turn implicates σ_2 to exceed σ_1.*

However, since $\beta_{\psi_1} = \beta_{\psi_2} = 0$ the prior expectation does not exist. Thus, we may recall that the conditional expected value of ψ_1 is given as (cf. Lemma 3.3, p. 40)

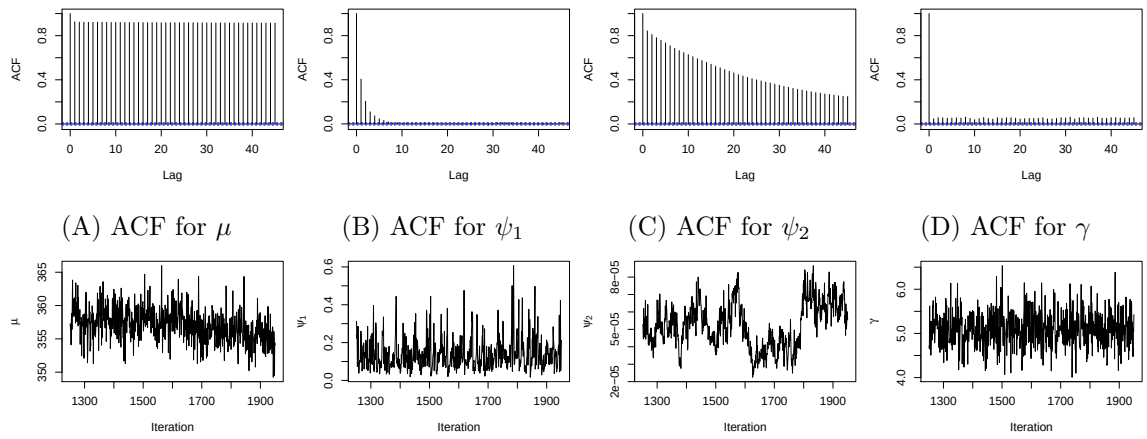
$$\mathbb{E}\left[\psi_1 \mid \vartheta_{-\psi_1}, \mathcal{S}\right] = \frac{2\alpha_{\psi_1} + (N-1)}{2\beta_{\psi_1} + \sum_{k=2}^{N} \left(B_k - B_{k-1} - \mu\right)^2}. \tag{3.101}$$

Thus, with increasing α_{ψ_1} the conditional expectation of ψ_1 increases, whereas the expectation of σ_1 decreases.

Prior for μ, $\pi(\mu)$	$\propto 1$ (Jeffreys)
Prior for ψ_1, $\pi(\psi_1)$	$Ga(10, 0)$
Prior for ψ_2, $\pi(\psi_2)$	$Ga(0, 0)$ (Jeffreys)
Prior for γ, $\pi(\gamma)$	$Ga(1/2, 0)$ (Jeffreys)
Prior for B_1, $\pi(B_1)$	$\propto 1$ (Jeffreys)
Prior for J_1, $\pi(J_i)$	$Bin(N, i/n)$ (Binomial)
$p^{(switch)}$	0.9
$\Delta^{(switch)}$	2
No. of iterations, R	40000
Burn-in period, BI	1250
Computation time	17.3 min

Table 3.8.: Example 6: Input parameters, burn-in period and computation time.

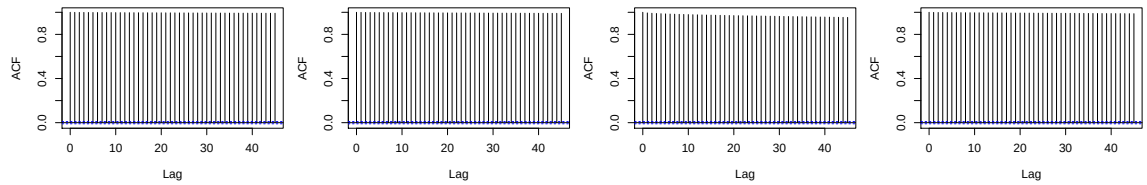
In comparison to Example 5, with $\alpha_{\psi_1} = 10$ we are able to substantially improve the mixing behavior of ψ_1 and also yield slightly smaller autocorrelations for ψ_2. As intended the posterior distribution of σ_1 does no longer exhibit two peaks but only one. This fact supports the hypothesis that the poor mixing behavior of σ_1 and the two-peaked density in Example 5 results from a discrepancy between prior distribution and data. However, here the burn-in period is still very large and all beats and μ exhibit extremely large autocorrelations.



(A) ACF for μ (B) ACF for ψ_1 (C) ACF for ψ_2 (D) ACF for γ

(E) Trace plot for μ (F) Trace plot for ψ_1 (G) Trace plot for ψ_2 (H) Trace plot for γ

Figure 3.10.: Example 6: Test-Data 3: (A)-(D) Sample ACFs and (E)-(H) trace plots for $\mu, \psi_1, \psi_2, \gamma$.



(A) ACF for B_1 (B) ACF for B_{11} (C) ACF for B_{21} (D) ACF for B_{31}

Figure 3.11.: Example 6: Test-Data 3: Sample ACF for (A) B_1, (B) B_{11}, (C) B_{21}, (D) B_{31}.

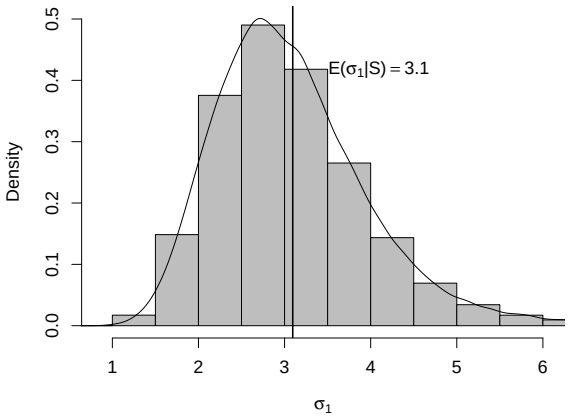


Figure 3.12.: Example 6: Test-Data 3: Posterior distribution of σ_1.

In Table 3.9 the posterior parameter means and medians (computed with all samples after the burn-in period), moment estimates (ME) and ACH estimates are reported. Also, we include the estimates from Example 5. Clearly, the major improvement in comparison to Example 5 is a much better estimate for σ_1, which is also reflected by better mixing properties of ψ_1. Consequently, the overall relative error decreases enormously.

Method/Estimate	$\hat{\mu}$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\gamma}$	ORE	$\overline{\ell}(\mathrm{GLO}_N)$
True values, η_0	364	5	200	5	-	-
Moment estimates, $\hat{\eta}_{\mathrm{ME}}$	363.32	6.28	191.98	4.74	0.35	-
Posterior means, $\hat{\eta}_{\mathrm{Mean}}$	336.29	3.09	142.52	4.84	0.78	- 1254.44
Posterior medians, $\hat{\eta}_{\mathrm{Median}}$	333.89	2.97	138.92	4.83	0.83	- 1254.44
Posterior means, $\hat{\eta}_{\mathrm{Mean}}$ (Ex. 5)	332.86	28.81	129.82	4.79	5.24	- 1235.87
Posterior medians, $\hat{\eta}_{\mathrm{Median}}$ (Ex. 5)	332.27	19.53	127.08	4.78	3.40	- 1235.87
ACH estimates, $\hat{\eta}_{\mathrm{ACH}}$	312.66	0.68	93.71	4.23	1.69	-

Table 3.9.: Test-Data 3: Parameter estimates in Example 6.

Moreover, we report the expected log-Likelihood $\overline{\ell}(GLO_N) = \mathbb{E}_{\vartheta|\mathcal{S}}\left[\log L(\vartheta; \mathcal{S} \mid GLO_N)\right]$ (cf. section 3.2.2). Although the estimation precision is better, $\overline{\ell}(GLO_N)$ is smaller for $\alpha_{\psi_1} = 10$ than for $\alpha_{\psi_1} = 0$ (Example 5).

As Example 6 shows, always choosing the Jeffreys prior does not necessarily lead to the smallest estimation error. Particularly for very different σ_1 and σ_2 the posterior distribution tends to increase both parameters if both exhibit the same prior distribution. In this case, with different prior parameters the large difference between σ_1 and σ_2 may be accounted for.

Moreover, a smaller estimation error due to a change in the prior may not necessarily be reflected in the expected log-likelihood. Contrarily, in Example 6 the expected log-likelihood even decreases. However, in Remark 3.2.2, p. 52, we introduced the expected log-likelihood $\overline{\ell}(\mathrm{GLO}_N)$ as a means for model selection whereas a change in the prior distribution does not change the underlying model but only the a priori belief about the parameters. Thus, we can not expect $\overline{\ell}(\mathrm{GLO}_N)$ to be maximized by the respective choice of prior distributions that lead to ORE-minimizing posterior parameter means.

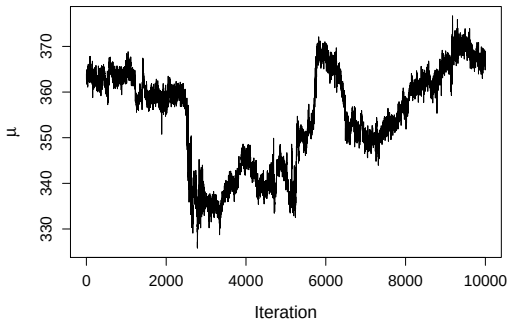
Concisely, Example 6 shows that prior distributions may serve as a kind of anchor for the posterior distribution. This is also true for the Binomial prior for the labels $\mathcal{J}$: As pointed out in section 3.3, we mainly distinguish between the Binomial and Uniform prior. Whereas with the first the last event is fixed to the last event, the latter allows the last event a priori to originate from the first beat with the same probability as from the last beat. For large σ_2 this may cause serious problems for the MCMC procedure: While for Test-Data 1 and 2 the posterior distributions with a uniform label prior correspond to these with a Binomial prior from Examples 3 and 4, the MCMC procedure for Test-Data 3 essentially fails:

Example 7. *In this example we analyze Test-Data 3 and follow the same procedure as in Example 5 with exactly one difference: For all labels J_i, $i = 1, ..., n$, we assume that a priori $J_i \sim unif(\{1, ..., N\})$. In other words, for every event S_i we a priori believe that S_i is placed by any beat B_k, $k = 1, ..., N$, with the same probability.*

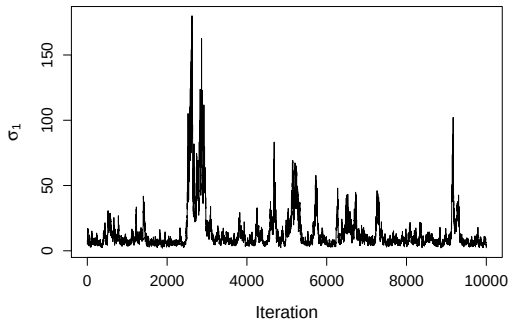
True parameter values	$\eta_0 = (\mu_0, \sigma_{1,0}, \sigma_{2,0}, \gamma_0) = (364, 5, 200, 5)$
Prior for $\mu, \psi_1, \psi_2, \gamma, B_1$	Jeffreys
Prior for $\mathcal{J}$	Uniform
$p^{(switch)}$	0.9
$\Delta^{(switch)}$	2
No. of iterations, R	40000
Computation time	19.43 min

Table 3.10.: Example 7: Input parameters and computation time.

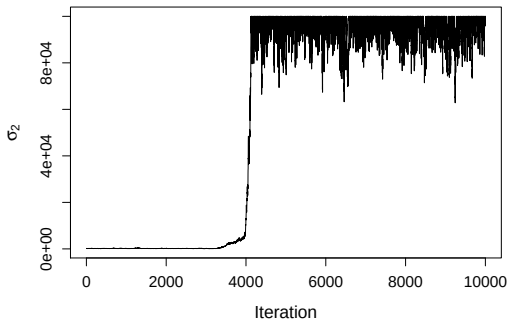
Figure 3.13 shows the first 10000 MCMC samples of μ, σ_1, σ_2 and γ. Apparently, at some point σ_2 explodes (as motivated in Remark 3.1, p. 50, in the implementation of the MCMC algorithm we restrict $\sigma_2 \leq 10^5$ to avoid problems due to numerical instability).



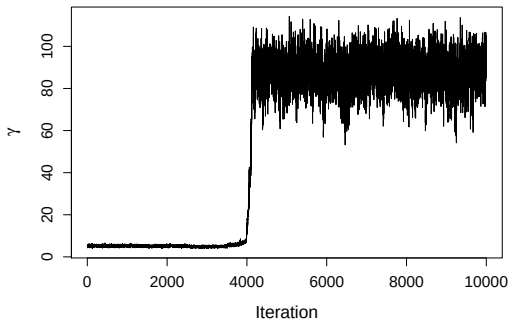
(A) First 10000 posterior samples of μ



(B) First 10000 posterior samples of σ_1



(C) First 10000 posterior samples of σ_2



(D) First 10000 posterior samples of γ

Figure 3.13.: Example 7: Test-Data 3: Evolution of (A) μ, (B) σ_1, (C) σ_2 and (D) γ.

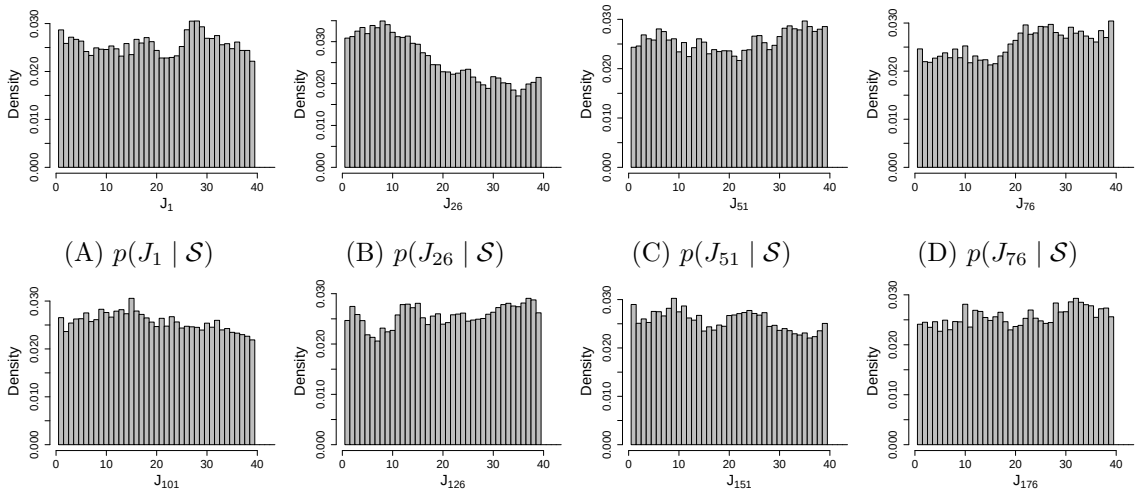
To analyze the reason for this behavior, recall from Lemma 3.6, p. 45, that the acceptance probability for a change of label J_i from $J_i = k$ to k' is given by $p_{acc}^{(switch)}(k' \mid k) = \min\{1, \tilde{p}\}$, *where*

$$\tilde{p} = \exp\left[-\frac{\psi_2}{2}\left(B_{k'}^2 - B_k^2 + 2S_i(B_k - B_{k'})\right)\right] \times \frac{n_k}{n_{k'}+1} \times \frac{\pi(k')}{\pi(k)} \times \frac{|V_k|}{|V_{k'}|}.$$

In this example we have $\psi_{2,0} = \sigma_{2,0}^{-2} = 200^{-2} = 2.5 \times 10^{-5}$ *and* $\pi(k) = \pi(k') = 1/N$. *Therefore,*

$$\tilde{p} \approx \frac{n_k}{n_{k'}+1} \times \frac{|V_k|}{|V_{k'}|}. \tag{3.102}$$

Moreover, for all $k \in \{1, ..., N : 1 + \Delta^{(switch)} < k < N - \Delta^{(switch)}\}$ *we have* $|V_k| = |V_{k'}|$ *(cf. Lemma 3.6, p. 45). Therefore, the acceptance probability does hardly depend on the location of the respective event S_i, current beat B_k or proposed beat $B_{k'}$, but only on the current number of events at beats B_k and $B_{k'}$, i.e. n_k and $n_{k'}$, respectively. Particularly, the more events there are at the current beat B_k in comparison to the proposed beat $B_{k'}$ the larger is the acceptance probability for a switch. As a result, labels are (after the burn-in period of* $BI = 4500$*) a posteriori nearly uniformly distributed, as Figure 3.14 shows.*



(A) $p(J_1 \mid \mathcal{S})$ (B) $p(J_{26} \mid \mathcal{S})$ (C) $p(J_{51} \mid \mathcal{S})$ (D) $p(J_{76} \mid \mathcal{S})$

(E) $p(J_{101} \mid \mathcal{S})$ (F) $p(J_{126} \mid \mathcal{S})$ (G) $p(J_{151} \mid \mathcal{S})$ (H) $p(J_{176} \mid \mathcal{S})$

Figure 3.14.: Example 7: Test-Data 3: Posterior pmf of (A) J_1, (B) J_{26},..., (H) J_{176}.

In other words, it is for every event almost equally likely to originate from any beat $B_1, ..., B_N$. *Therefore, the distances between beats and their families increase enormously, i.e.* $\sum_{i=1}^{n}(S_i - B_{J_i})^2$ *is large. Now, recall from Lemma 3.4, p. 41, that* ψ_2 *is updated as*

$$\psi_2 \sim Ga\left(\alpha_{\psi_2} + \frac{n}{2},\, \beta_{\psi_2} + \frac{1}{2}\sum_{i=1}^{n}(z_i - B_{J_i})^2\right), \tag{3.103}$$

where z_i *are the rescaled event locations. Since* $\sum_{i=1}^{n}(S_i - B_{J_i})^2$ *increases substantially,* $\sum_{i=1}^{n}(z_i - B_{J_i})^2$ *also increases (in most cases the rescaling will rather increase than substantially decrease the distances to the beats), thus* $\mathbb{E}[\psi_2 \mid \vartheta_{-\psi_2}, \mathcal{S}]$ *further decreases.*

Analogously, $\nu_k = F_{\mathcal{N}(B_k,\psi_2^{-1})}(T) - F_{\mathcal{N}(B_k,\psi_2^{-1})}(0)$ *decreases for decreasing* ψ_2. *Since* γ *is updated by (cf. Lemma 3.1, p. 39)*

$$\gamma \sim Ga\left(\alpha_\gamma + n,\, \beta_\gamma + \sum_{k=1}^{N}\nu_k\right), \tag{3.104}$$

$\mathbb{E}[\gamma \mid \vartheta_{-\gamma}, \mathcal{S}]$ *increases with decreasing* ψ_2.

3.4.2. Influence of N

Clearly, if N is too small, the beats have to incorporate more events than they originally placed. Thus, $\hat{\gamma}$ increases but also the beat locations will not always coincide with the mid of an observed burst. Thus, $\hat{\sigma}_1$ and $\hat{\sigma}_2$ increase, as Example 8 illustrates. Contrarily, for too large N the parameters are underestimated. More examples that illustrate the influence of N can be found in section 4.1.

Example 8. *In this example we analyze Test-Data 1. Recall from Example 3 that Test-Data 1 exhibits* $N_0 = 36$ *beats and* $n = 351$ *events and is very regular. Here, we assume the same Jeffreys and Binomial priors as in Example 3 and also use the original parameter values* η_0 *as initial values. However, we set different initial values* $\mathcal{B}^{(0)}$ *and* $\mathcal{J}^{(0)}$*: Instead of the original values we, firstly, choose* $N = 30$*, secondly, set the beats to the mids of equidistant intervals,*

$$B_k^{(0)} = \frac{T}{2N} + (k-1)\frac{T}{N} \quad \text{for } k = 1, ..., N, \tag{3.105}$$

and, thirdly, choose similar to Algorithm 3.2 (Simple guess of initial values) the nearest beat for every event, i.e.

$$J_i^{(0)} := \arg\min_{k \in \{1,...,N\}} \left| B_k^{(0)} - S_i \right| \quad \text{for } i = 1, ..., n. \tag{3.106}$$

In Table 3.11 the posterior means and medians (computed with all samples after a burn-in period of $BI = 100$ *samples), ME and ACH estimates as well as the estimates from Example 3 with correct number of beats,* $N_0 = 36$*, are reported. In comparison to Example 3 all posterior means and medians substantially increase.*

Method/Estimate	$\hat{\mu}$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\gamma}$	ORE	$\overline{\ell}(\text{GLO}_N)$
True values, η_0	364	20	10	10	-	-
Moment estimates, $\hat{\eta}_{\text{ME}}$	360.81	19.64	9.30	9.75	0.12	-
Posterior means, $\hat{\eta}_{\text{Mean}}$	439.23	115.86	95.17	12.01	13.72	- 2168.8
Posterior medians, $\hat{\eta}_{\text{Median}}$	439.22	114.46	94.48	12.00	13.58	- 2168.8
Posterior means, $\hat{\eta}_{\text{Mean}}$ (Ex. 3)	360.76	18.07	9.20	9.76	0.21	- 1363.8
Posterior medians, $\hat{\eta}_{\text{Median}}$ (Ex. 3)	360.76	17.87	9.19	9.75	0.22	- 1363.8
ACH estimates, $\hat{\eta}_{\text{ACH}}$	360.33	10.31	12.12	7.33	0.97	-

Table 3.11.: Test-Data 1: Parameter estimates in Example 8.

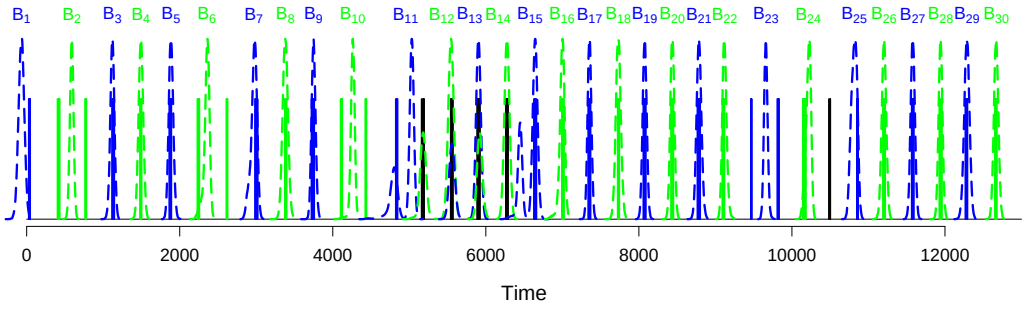


Figure 3.15.: Example 8: Test-Data 1: Observed events (solid), posterior beat densities (dashed) and labels (colors).

Moreover, we report the expected log-Likelihood $\bar{\ell}(GLO_N) = \mathbb{E}_{\vartheta|\mathcal{S}}\left[\log L(\vartheta; \mathcal{S} \mid GLO_N)\right]$ *(cf. section 3.2.2), which is substantially smaller for* $N = 30$ *than the expected log-likelihood for the correct* $N = N_0 = 36$ *in Example 3. Thus, in this example* $\bar{\ell}(GLO_N)$ *is suitable to evaluate what value for* N *provides the better model.*

Figure 3.15 shows the posterior distributions of all beats (dashed) together with the observed event locations (solid). The beats are deterministically colored in blue and green (starting with blue). The events' color corresponds to the posterior labels: If the posterior probability for event S_i *originating from beat* B_k *is larger or equal to a threshold value of* 80%*, i.e.* $\mathbb{P}(J_i = k \mid \mathcal{S}) \geq 0.8$*, the bar of event* S_i *is colored in the same color as beat* B_k *(green or blue). If there does not exist any* $k \in \{1, ..., N\}$ *with* $\mathbb{P}(J_i = k \mid \mathcal{S}) \geq 0.8$*,* S_i *is colored black.*

As expected there exist several beats that cover two bursts in their event-families (e.g. beats B_2*,* B_6*,* B_{10}*,* B_{23}*) or exhibit two peaks in their posterior distribution (e.g. beats* B_{11}*,...,*B_{15}*). Moreover, for five bursts of events (four bursts between 5000 and 7000 and one burst roughly at 10500) it is not sure to which beats they belong (with respect to the threshold of* 80%*).*

3.4.3. Influence of initial values

Similar to the other input parameters of the algorithm the initial values $\vartheta^{(0)}$ do also influence the estimation precision. To assess their impact we apply Algorithm 3.2 (*Simple guess of initial values*) and choose the correct $N = N_0$, i.e. we divide $[0, T]$ into N_0 equidistant intervals, set $\mathcal{B}^{(0)}$ to the mids of these and set $\mathcal{J}^{(0)}$ according to the nearest beat for every event, which is also done in Example 8. For the remaining parameters we take the sample estimates as proposed in Algorithm 3.2.

However, for all *Test-Data 1, 2* and *3* we still obtain exactly the same mixing behavior and posterior distributions as with the original values as initial values. This indicates that even by choosing the initial values with this simple and straightforward method the MCMC procedure works properly. Moreover, the burn-in periods do not change, although, we can identify a large jump from the initial value to the first simulated value if the initial value has low probability mass.

3.4.4. Estimation precision

In the following we will analyze the general estimation precision of the posterior parameter means and medians for a given parameter combination $\eta_0 = (\mu_0, \sigma_{1,0}, \sigma_{2,0}, \gamma_0)$. For this purpose we consider nine different parameter constellations which are reported in Table 3.12. Note that combination f corresponds to the true parameter values of *Test-Data 3* in Examples 5, 6 and 7. Since μ is a scaling parameter for the whole process, we fix $\mu_0 = 364$, which is a typical value for the mean distance between bursts of tropical cyclones. Our baseline calibration is combination b which represents a regular GLO process with beats that very

likely exhibit observable bursts of events since for $\gamma = 5$ we have $\mathbb{P}(P_k = 0) = e^{-5} \approx 0.007$. Starting with this parameter combination we increase the irregularity of the process and vary the firing rate γ_0. We are particularly interested in large values for $\sigma_{2,0}$ and the influence of low firing rates γ_0 since these lead to high estimation errors of the ACH estimates (cf. Bingmer (2012, p.96)).

Combination	μ_0	$\sigma_{1,0}$	$\sigma_{2,0}$	γ_0
a	364	5	50	1
b	364	5	50	5
c	364	5	50	10
d	364	5	50	20
e	364	5	100	5
f	364	5	200	5
g	364	15	50	5
h	364	50	50	5
i	364	100	50	5

Table 3.12.: Exemplary parameter combinations used for simulation study.

Moreover, to assess the impact of different record times T we analyze all *combinations* in three different *sets* with different values for T, which can be found in Table 3.13. We will call one distinct parameter combination together with one value for T a *setting*. Thus, we have $9 \times 3 = 27$ settings.

Set	I	II	III
Record time, T	8000	13000	23000
No. of simulations, M	1000	1000	1000
No. of MCMC iterations, R	30000	30000	30000
Burn-in period, BI	20	20	20

Table 3.13.: Exemplary record time, number of simulations and MCMC iterations and burn-in period used for simulation study.

For every setting we repeat the following steps for $m = 1, ..., M$:

1) Simulate events $\mathcal{S}^{(m)}$ of a process $\Phi \sim \text{GLO}(\mu_0, \sigma_{1,0}, \sigma_{2,0}, \gamma_0)$ in $[0, T]$.

2) Draw samples from the posterior distribution by applying the MCMC algorithm (Algorithm 3.1) with input parameters from Table 3.14

3) Calculate the estimates $\hat{\eta}^{(m)}_{\text{Mean}}$ and $\hat{\eta}^{(m)}_{\text{Median}}$ based on the last $R - BI$ MCMC samples.

With this procedure we obtain M realizations of the estimators $\hat{\eta}_{\text{Mean}}$ and $\hat{\eta}_{\text{Median}}$ and are able to calculate the MREs and ORE as described at the beginning of the section. Moreover, we compute the MREs and OREs of the corresponding ACH estimates $\hat{\eta}_{\text{ACH}}$ by following the same procedure.

Initial values, $\vartheta^{(0)}$	True parameter values η_0 and simulated (original) beats and labels
Prior for $\mu, \psi_1, \psi_2, \gamma, B_1$	Jeffreys
Prior for $\mathcal{J}$	Binomial
$p^{(switch)}$	0.9
$\Delta^{(switch)}$	2

Table 3.14.: Input parameters for MCMC algorithm in simulation study.

Firstly, we compare the ORE of $\hat{\eta}_{\text{Mean}}$ and $\hat{\eta}_{\text{Median}}$ in Figure 3.16. Clearly, only for rather regular processes (combinations a to d) there is a noticeable difference between $\text{ORE}(\hat{\eta}_{\text{Mean}})$ and $\text{ORE}(\hat{\eta}_{\text{Median}})$. Since the average overall error of $\hat{\eta}_{\text{Median}}$ of all settings, $\overline{\text{ORE}}(\hat{\eta}_{\text{Median}}) = 4.2$, is slightly smaller than $\overline{\text{ORE}}(\hat{\eta}_{\text{Mean}}) = 4.5$, we will focus on $\hat{\eta}_{\text{Median}}$ in the following.

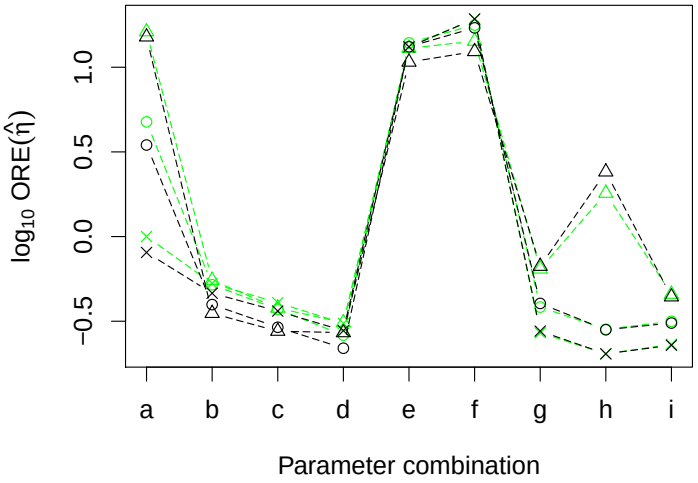
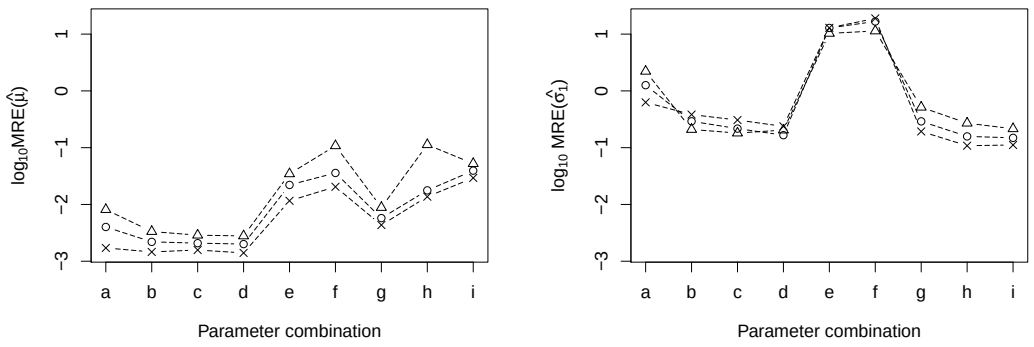
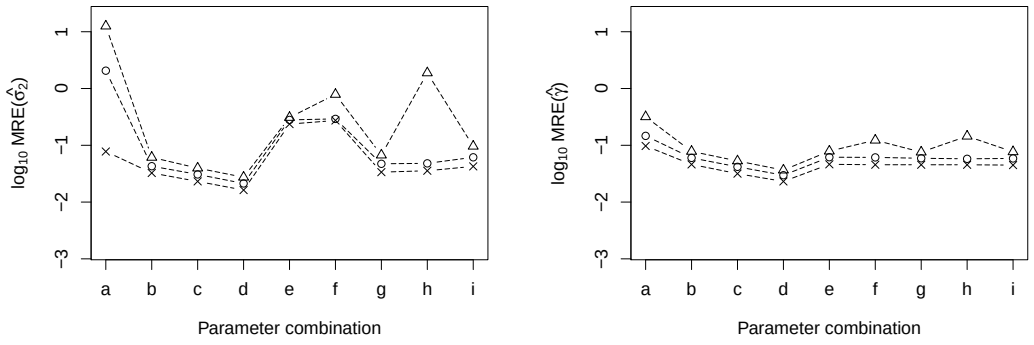


Figure 3.16.: $\log_{10}$ ORE for $\hat{\eta}_{\text{Mean}}$ (green) and $\hat{\eta}_{\text{Median}}$ (black) in simulation study for $T = 8000$ ($\triangle$), $T = 13000$ ($\bigcirc$) and $T = 23000$ ($\times$).

As Figure 3.17 shows, the MREs for σ_1 are comparably large, whereas μ exhibits the smallest estimation errors. Moreover, for most combinations the impact of the record time, T, on the MREs is rather small. Only in case of a low firing rate γ_0 (combination a), very large burst width $\sigma_{2,0}$ (combination f) or irregular background rhythm and large burst width (combination h) there is a substantial improvement for large T. For the largest considered record time $T = 23000$ the estimation error of $\hat{\gamma}_{\text{Median}}$ does not change by much for different combinations, whereas for the other parameters the MRE is particularly large in case of a large burst width (combinations e and f). Interestingly, in case of $T = 23000$ only the estimation error for $\hat{\mu}_{\text{Median}}$ increases noticeably for an irregular background rhythm (combinations h and i).



(A) $\log_{10}\mathrm{MRE}\,(\hat{\mu}_{\mathrm{Median}})$ (B) $\log_{10}\mathrm{MRE}\,(\hat{\sigma}_{1,\mathrm{Median}})$

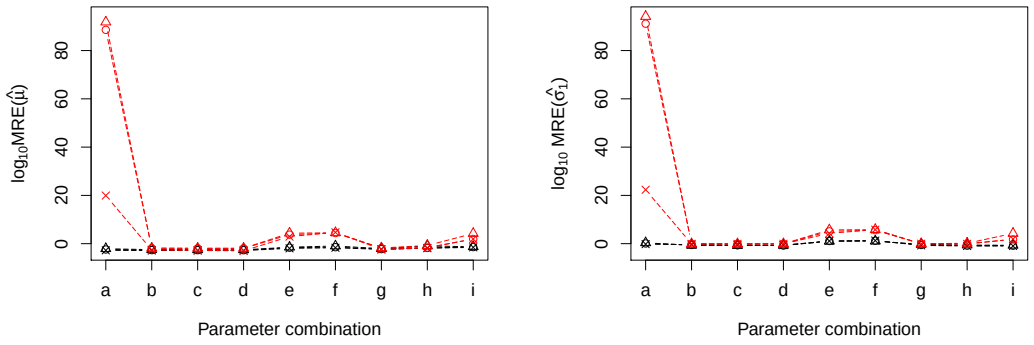


(C) $\log_{10}\mathrm{MRE}\,(\hat{\sigma}_{2,\mathrm{Median}})$ (D) $\log_{10}\mathrm{MRE}\,(\hat{\gamma}_{\mathrm{Median}})$

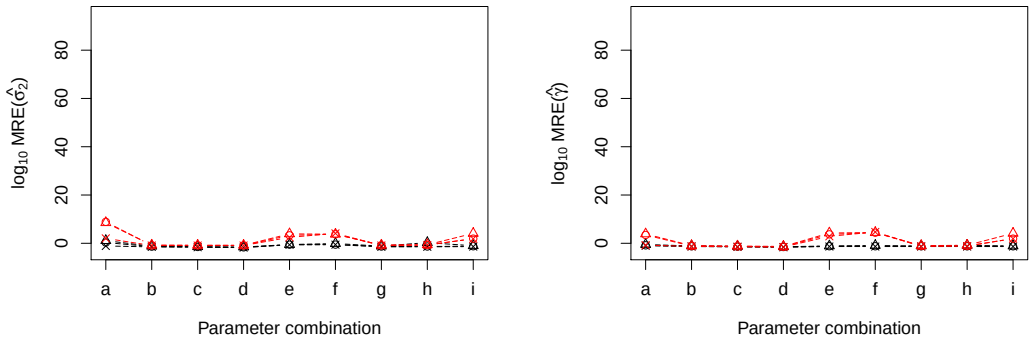
Figure 3.17.: $\log_{10}\mathrm{MRE}$ for (A) $\hat{\mu}_{\mathrm{Median}}$, (B) $\hat{\sigma}_{1,\mathrm{Median}}$, (C) $\hat{\sigma}_{2,\mathrm{Median}}$ and (D) $\hat{\gamma}_{\mathrm{Median}}$ in simulation study for $T = 8000$ ($\triangle$), $T = 13000$ ($\bigcirc$) and $T = 23000$ ($\times$).

As shown in Figure 3.18, the estimation precision of the ACH estimates can be improved substantially, in particular for μ and σ_1, in case of a small firing rate γ_0 (combination a). Moreover, there are considerable improvements for all parameters in case of large burst widths $\sigma_{2,0}$ (combinations e and f) and an irregular background rhythm (combination i). The OREs, shown in Figure 3.19, emphasize these findings. Generally, there is no setting in which the ACH estimates exhibit a considerably smaller estimation error than the posterior medians.

Additionally in Table 3.15 we report for every setting the average number of observed (simulated) events, $\overline{n}$, the average number of observed (simulated) beats, $\overline{N}$ and the average computation time (in minutes) per GLO simulation (on a Mac Pro with 2.7 GHz 12 Core Intel, 32 GB RAM and Mac OS X Yosemite 10.10.2). Furthermore, we report the number of failures relative to the number of simulations, M. Since we restrict $\sigma_2 \leq 10^5$ in the implementation of the MCMC algorithm (cf. Remark 3.1, p. 50, and Example 7, p. 69), we interpret one estimation procedure as being failed if either $\max_{i=1,\ldots,4} \hat{\eta}_i \geq 10^5$ or one estimate is missing (reported as *NA*; in R this is possible if values get to small or large for employed functions like e.g. *rgamma*). Note that we neither include the failures of the Bayesian estimation nor failures of the ACH estimation (aborts of the fitting algorithm) in Figures 3.16 to 3.19.



(A) $\log_{10}\mathrm{MRE}\,(\hat{\mu})$ (B) $\log_{10}\mathrm{MRE}\,(\hat{\sigma}_1)$



(C) $\log_{10}\mathrm{MRE}\,(\hat{\sigma}_2)$ (D) $\log_{10}\mathrm{MRE}\,(\hat{\gamma})$

Figure 3.18.: $\log_{10}\mathrm{MRE}$ for $\hat{\eta}_{\mathrm{Median}}$ (black) and $\hat{\eta}_{\mathrm{ACH}}$ (red) in simulation study for $T = 8000$ ($\triangle$), $T = 13000$ ($\bigcirc$) and $T = 23000$ ($\times$) for (A) μ, (B) σ_1, (C) σ_2 and (D) γ.

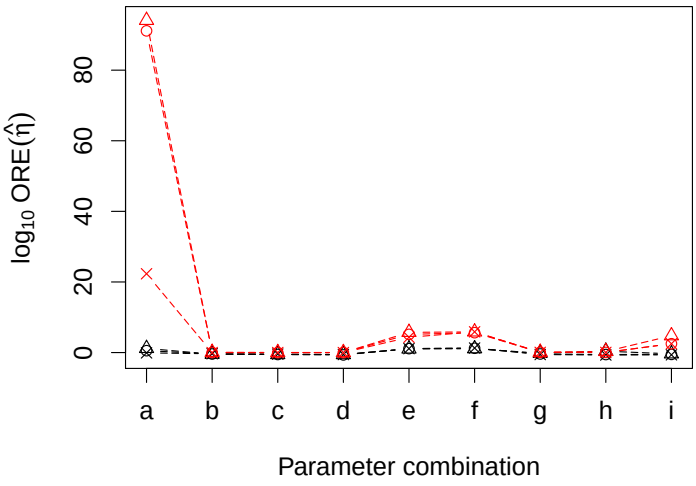


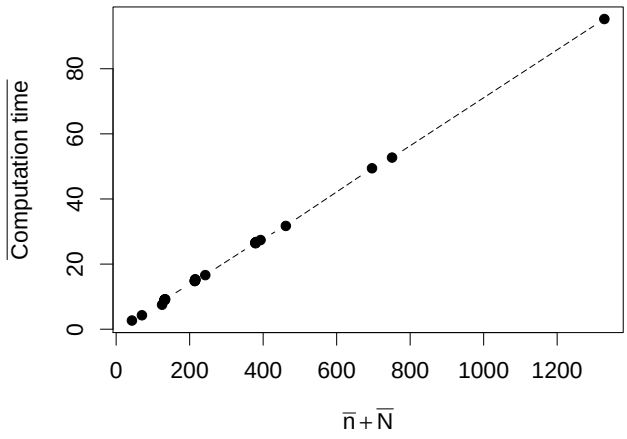
Figure 3.19.: $\log_{10}\mathrm{ORE}$ for $\hat{\eta}_{\mathrm{Median}}$ (black) and $\hat{\eta}_{\mathrm{ACH}}$ (red) in simulation study for $T = 8000$ ($\triangle$), $T = 13000$ ($\bigcirc$) and $T = 23000$ ($\times$).

T	Combination	$\overline{n}$	$\overline{N}$	Computation time	No. of failures/M
8000	a	109.70	22.28	9.09	0.001
8000	b	109.60	22.24	9.03	0.001
8000	c	110.19	22.56	9.16	0.009
8000	d	220.13	22.39	16.62	0
8000	e	439.22	22.49	31.73	0
8000	f	109.62	22.28	9.01	0.007
8000	g	109.43	22.18	8.99	0.019
8000	h	110.20	23.19	9.20	0.020
8000	i	22.19	20.75	2.67	0.061
13000	a	179.09	35.99	15.35	0
13000	b	177.94	36.01	14.80	0.001
13000	c	178.85	36.29	15.05	0.008
13000	d	356.82	36.12	27.40	0
13000	e	714.52	36.23	52.72	0
13000	f	178.25	35.97	14.95	0.001
13000	g	178.11	35.95	14.83	0.010
13000	h	178.90	36.90	15.17	0.006
13000	i	35.59	34.59	4.29	0.009
23000	a	316.91	63.48	26.75	0
23000	b	315.83	63.47	26.56	0
23000	c	315.73	63.77	26.67	0.003
23000	d	632.62	63.59	49.43	0
23000	e	1264.69	63.71	95.27	0
23000	f	315.02	63.44	26.49	0
23000	g	316.01	63.51	26.42	0.003
23000	h	316.29	64.41	26.77	0.003
23000	i	62.83	62.15	7.52	0.002

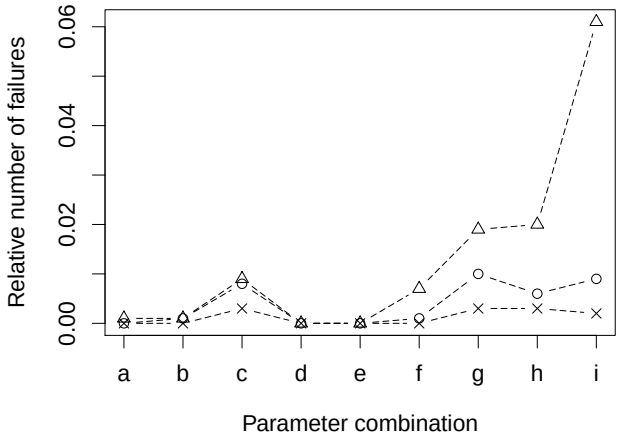
Table 3.15.: Average number of events, beats, computation time (in minutes) and average relative number of failures in simulation study.

Figure 3.20A shows that the average computation time is nearly linearly increasing with $\overline{n} + \overline{N}$. This is very intuitive since the number of calculation steps does not change with varying size of observations, n, or beats, N, except for N update steps for the beats, $\max\{1, \lfloor np^{(switch)} \rceil\}$ update steps for the labels, and n calculations for latent variables for the update of ψ_2.

Moreover, the number of failures is very sensitive for the record time, as Figure 3.20B shows. Nonetheless, for the typical record time $T = 13000$ (or larger) there are no combinations with more than 1% failures.



(A) Average computation time $\sim (\overline{n} + \overline{N})$



(B) Relative relative number of failures for $T = 8000$ ($\triangle$), $T = 13000$ ($\bigcirc$) and $T = 23000$ ($\times$)

Figure 3.20.: (A) Average computation time (in minutes) and (B) relative number of failures in simulation study.

4. Data analysis, forecasting and further questions

4.1. Data analysis for tropical cyclones

In this section we will apply the Bayesian estimation procedure to the data of tropical cyclones in Africa, Americas, Asia and Oceania that were analyzed in sections 1.2 and 1.3. For this purpose, we will employ the MCMC procedure (Algorithm 3.1) with input parameters as reported in Table 4.1. Particularly, we assume Jeffreys prior for all parameters since we do not have any prior information about the occurrence behavior of tropical cyclones.

Prior for $\mu, \psi_1, \psi_2, \gamma, B_1$	Jeffreys
Prior for $\mathcal{J}$	Binomial
Initial values	ACH estimates $\hat{\eta}_{\text{ACH}}$, *(Complex/Simple guess of initial values)*
$p^{(switch)}$	0.9
$\Delta^{(switch)}$	2
No. of iterations, R	35000

Table 4.1.: Input parameters for estimation procedure for tropical cyclones.

To obtain initial values we follow Algorithm 3.4 *(Complex guess of initial values)*, i.e. we use the ACH estimates $\hat{\eta}_{\text{ACH}}$ as initial values for $\mu^{(0)}$, $\psi_1^{(0)}$, $\psi_2^{(0)}$ and $\gamma^{(0)}$, employ Algorithm 3.3 *(Estimating beat locations)* to obtain $\mathcal{B}^{(0)}$, N, and set $\mathcal{J}^{(0)}$ such that all events are linked to their nearest beat. We will determine the ACH estimates as described in section 3.4. Regarding the KDE bandwidth σ in Algorithm 3.3 we will, firstly, set $\sigma = \hat{\sigma}_{2,\text{ACH}}$ and, secondly, analyze if a different bandwidth leads to a better model with respect to the expected log-likelihood $\overline{\ell}(\text{GLO}_N)$. If we are not able to find a bandwidth that yields a particular value for N, we set

$$B_k^{(0)} = \frac{T}{2N} + (k-1)\frac{T}{N} \tag{4.1}$$

as described in section 3.2 *(Simple guess)*. The goodness of fit of the fitted GLO model will be judged by visual inspection of the posterior beat distributions and GLO simulations.

4.1.1. Africa

The dataset incorporates $n = 97$ reports of tropical cyclones in Africa, for which we obtain the ACH estimates $\hat{\eta}_{\text{ACH}} = (366.07, 7.43, 46.64, 2.7)$.
However, by choosing $\sigma = \hat{\sigma}_{2,\text{ACH}} = 46.64$ as bandwidth, we obtain $N = 30$ beats and abnormally large autocorrelations and parameter estimates. This indicates, that N is too small. Thus, we vary the bandwidth σ. The resulting values for N are plotted in Figure 4.1. Apparently, $N = 35$ is the largest value that Algorithm 3.3 is able to find given the ACH estimates for the other parameters.

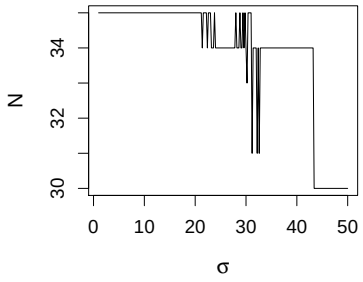


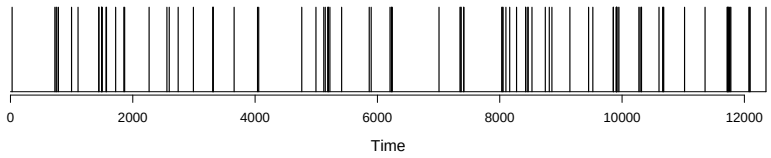
Figure 4.1.: Tropical cyclones in Africa: Different bandwidths σ with resulting N.

σ/Method	N	$\overline{\ell}(\text{GLO}_N)$	BI	$\hat{\mu}_{\text{Median}}$	$\hat{\sigma}_{1,\text{Median}}$	$\hat{\sigma}_{2,\text{Median}}$	$\hat{\gamma}_{\text{Median}}$
(Simple guess)	36	- 915.83	200	415.72	572.05	126.15	2.85
10	**35**	**- 595.19**	**100**	**366.67**	**6.06**	**50.56**	**2.89**
23	34	- 893.19	100	448.98	546.08	129.38	3.02
$\hat{\sigma}_{2,\text{ACH}} = 46.64$	30	$-\infty$	100	152.9×10^4	373.8×10^4	10^5	1296.58

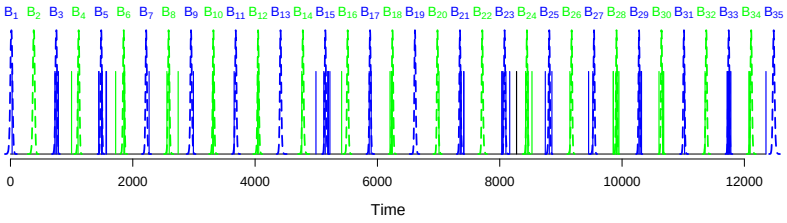
Table 4.2.: Tropical cyclones in Africa: Bandwidth, number of beats, expected log-likelihood, burn-in period and posterior parameter medians.

We employ the fitting procedure with bandwidths $\sigma = 10$ and $\sigma = 23$ and compare the resulting expected log-likelihoods in Table 4.2. Apparently, $\overline{\ell}(\text{GLO}_N)$ is larger for $N = 35$. To check if N is still too small, we also apply the estimation procedure for $N = 36$, which leads to a smaller expected log-likelihood. Therefore, $N = 35$ provides the best fit among the three considered values for N. We do not check for the existence of other local minima of $\overline{\ell}(\text{GLO}_N)(N)$ since the resulting estimates for $N = 35$ are reasonable:

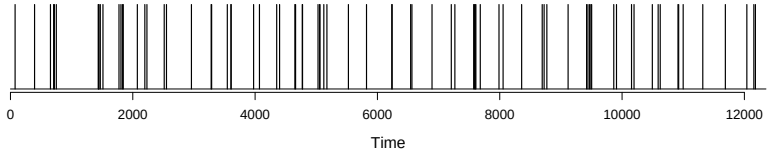
Figure 4.2B indicates that $N = 35$ is correct by showing the posterior distribution of all beats together with the observed event locations. As in Example 8, the events' color corresponds to the posterior labels: If $\mathbb{P}(J_i = k \mid \mathcal{S}) \geq 0.8$, event S_i is colored in the same color as beat B_k (green or blue). If there does not exist any $k \in \{1, ..., N\}$ with $\mathbb{P}(J_i = k \mid \mathcal{S}) \geq 0.8$, S_i is colored black. Additionally, in Figure 4.2C we show a GLO simulation with the posterior medians. Visually the estimates are able to reproduce the observed pattern.



(A) Observed events



(B) Observed events (solid), posterior beat densities (dashed) and labels (colors) ($N = 35$)



(C) GLO simulation with $\hat{\eta}_{\text{Median}}$ ($N = 35$)

Figure 4.2.: Tropical cyclones in Africa: (A) Data, (B) posterior beat densities and (C) simulation.

4.1.2. Americas

The dataset incorporates $n = 443$ reports of tropical cyclones in Americas, for which we obtain the ACH estimates $\hat{\eta}_{\text{ACH}} = (364.27, 0.21, 42.23, 14.75)$. Figure 4.3 depicts several bandwidths with resulting values for N. According to the expected log-likelihood $N = 34$ provides the best model as reported in Table 4.3. Figures 4.4B and 4.4C indicate that the beat locations are estimated reasonable and the resulting posterior medians are able to reproduce the observed occurrence pattern.

σ/Method	N	$\overline{\ell}(\text{GLO}_N)$	BI	$\hat{\mu}_{\text{Median}}$	$\hat{\sigma}_{1,\text{Median}}$	$\hat{\sigma}_{2,\text{Median}}$	$\hat{\gamma}_{\text{Median}}$
(Simple guess)	35	- 2459.88	200	357.02	42.93	39.77	12.99
5	**34**	**- 2457.65**	**200**	**367.41**	**20.07**	**40.7**	**13.35**
10	33	- 2490.38	200	380.69	55.16	43.31	13.83
$\hat{\sigma}_{2,\text{ACH}} = 42.23$	31	- 2583.58	100	417.2	103.76	53.66	14.78

Table 4.3.: Tropical cyclones in Americas: Bandwidth, number of beats, expected log-likelihood, burn-in period and posterior parameter medians.

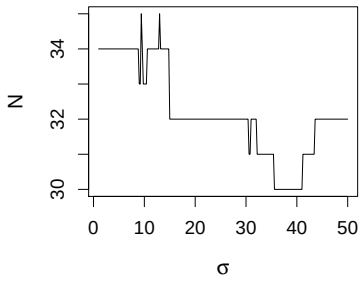
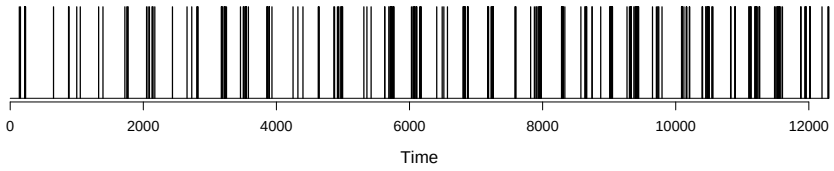
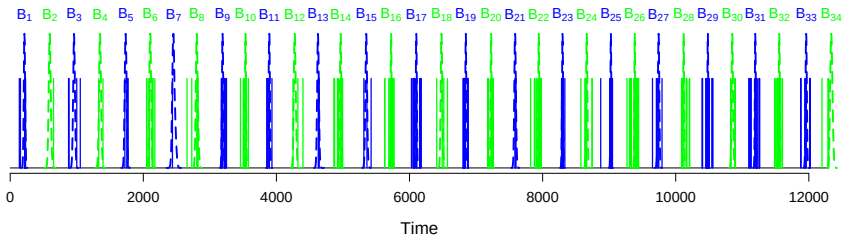


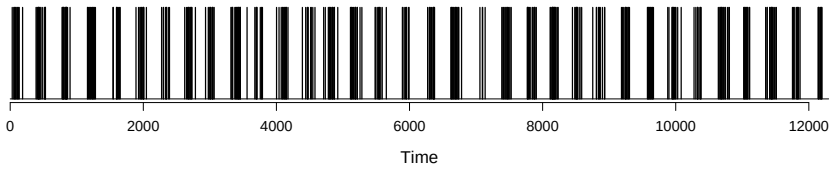
Figure 4.3.: Tropical cyclones in Americas: Different bandwidths σ with resulting N.



(A) Observed events



(B) Observed events (solid), posterior beat densities (dashed) and labels (colors) ($N = 34$)



(C) GLO simulation with $\hat{\eta}_{\text{Median}}$ ($N = 34$)

Figure 4.4.: Tropical cyclones in Americas: (A) Data, (B) posterior beat densities and (C) simulation.

However, the simulated bursts appear to be slightly larger than the observed. Moreover, the observed numbers of events per burst are much more variable than the simulated. In particular, if we take $\hat{n}_k := \left| \{i \in \{1, ..., n\} : p^{-1}_{J_i|\mathcal{S}}(0.5) = k\} \right|$ as estimate for the number of events in burst $k \in \{1, ..., N\}$, the sample mean is $\overline{\hat{n}} = 13.03$, whereas the sample variance is $s^2(\hat{n}) = 99.79$. This indicates that, firstly, large bursts have a larger influence on the

posterior estimate and, secondly, in contrast to a Poisson distribution the variance of the number of events in bursts may be larger than the expected value. This motivates a different distribution for P_k like e.g. the Negative Binomial distribution.

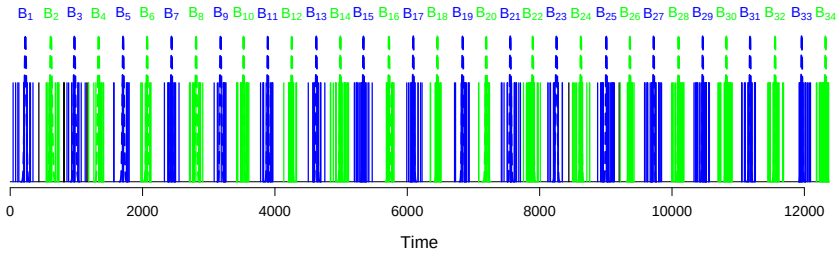
4.1.3. Asia

The dataset incorporates $n = 741$ reports of tropical cyclones in Asia, for which we obtain the following ACH estimates: $\hat{\eta}_{\text{ACH}} = (370.98, 1.07, 74.04, 22.73)$. For every bandwidth $\sigma \leq 100$ we obtain $N = 34$ (which may be due to the small ACH estimate for σ_1). However, the expected log-likelihood is larger for $N = 35$, as reported in Table 4.4.

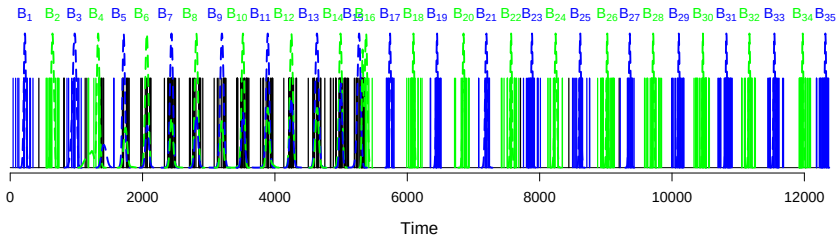
σ/Method	N	$\overline{\ell}(\text{GLO}_N)$	BI	$\hat{\mu}_{\text{Median}}$	$\hat{\sigma}_{1,\text{Median}}$	$\hat{\sigma}_{2,\text{Median}}$	$\hat{\gamma}_{\text{Median}}$
(Simple guess)	36	- 4243.38	100	345.66	74.70	63.91	20.73
(Simple guess)	**35**	**- 4242.62**	**100**	**355.80**	**53.42**	**63.87**	**21.31**
$\hat{\sigma}_{2,ACH} = 74.04$	34	- 4245.39	100	366.38	18.25	64.54	21.95

Table 4.4.: Tropical cyclones in Asia: Bandwidth, number of beats, expected log-likelihood, burn-in period and posterior parameter medians.

Since there exist several very large bursts with unusually many beats, these additional events are incorporated by additional beats with $N = 35$ (cf. Figure 4.5). Nonetheless, the GLO simulation with $\hat{\eta}_{\text{Median}}$ from $N = 35$ (cf. Figure 4.6B) is able to reproduce the observed occurrence pattern and, thus, indicates that the parameter estimates are reasonable.



(A) Observed events (solid), posterior beat densities (dashed) and labels (colors) ($N = 34$)



(B) Observed events (solid), posterior beat densities (dashed) and labels (colors) ($N = 35$)

Figure 4.5.: Tropical cyclones in Asia: Posterior beats for (A) $N = 34$ and (B) $N = 35$.

(A) Observed events

(B) GLO simulation with $\hat{\eta}_{\text{Median}}$ ($N = 35$)

Figure 4.6.: Tropical cyclones in Asia: (A) Data and (B) simulation.

4.1.4. Oceania

The dataset incorporates $n = 137$ reports of tropical cyclones in Oceania, for which we obtain the following ACH estimates: $\hat{\eta}_{\text{ACH}} = (364.37, 1.89, 60.29, 3.98)$. According to the expected log-likelihood $N = 35$ provides the best model. The estimates are reported in Table 4.5. Figures 4.8B and 4.8C indicate that the beat locations are estimated reasonably and the fitted model is able to reproduce the observed occurrence pattern.

σ/Method	N	$\overline{\ell}(\text{GLO}_N)$	BI	$\hat{\mu}_{\text{Median}}$	$\hat{\sigma}_{1,\text{Median}}$	$\hat{\sigma}_{2,\text{Median}}$	$\hat{\gamma}_{\text{Median}}$
(Simple guess)	36	- 968.19	100	377.82	200.58	53.32	3.96
$\hat{\sigma}_{2,\mathbf{ACH}} = 60.29$	**35**	**- 819.42**	**100**	**365.33**	**6.58**	**56.12**	**4.06**
$\hat{\sigma}_{2,\text{ACH}}/2 = 30.15$	34	- 864.56	100	389.26	126.62	56.55	4.22

Table 4.5.: Tropical cyclones in Oceania: Bandwidth, number of beats, expected log-likelihood, burn-in period and posterior parameter medians.

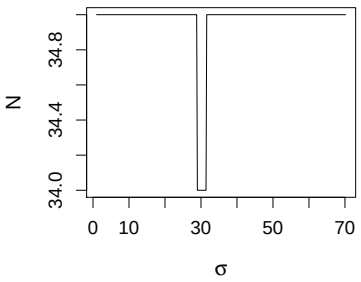
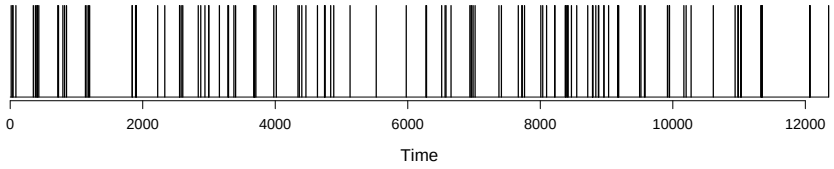
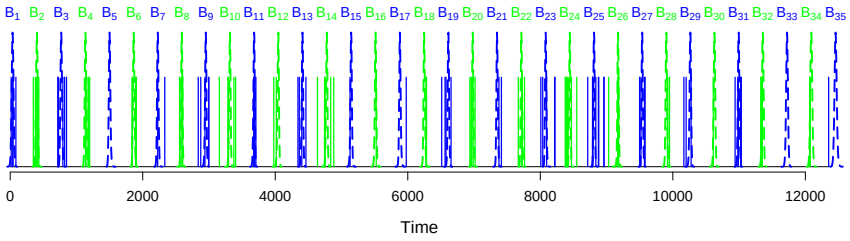


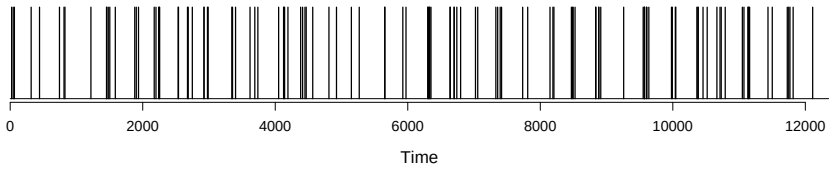
Figure 4.7.: Tropical cyclones in Oceania: Different bandwidths σ with resulting N.



(A) Observed events



(B) Observed events (solid), posterior beat densities (dashed) and labels (colors) ($N = 35$)



(C) GLO simulation with $\hat{\eta}_{\text{Median}}$ ($N = 35$)

Figure 4.8.: Tropical cyclones in Oceania: (A) Data, (B) posterior beat densities and (C) simulation.

4.2. Forecasting

For insurance companies it is of great interest to forecast future events after having observed events $\mathcal{S} \in [0, T]^n$. Thus, particularly to illustrate the use of the GLO model in combination with the Bayesian estimation framework for risk management purposes, in this section we will give approximations for two possible quantities of interest:

4.2.1 the expected number of events in $(t_1, t_2]$, $T \leq t_1 < t_2$, i.e. $\mathbb{E}[\Phi(t_1, t_2) \mid \mathcal{S}]$,

4.2.2 the probability of a default at time $t > T$.

In general, a posteriori and conditional on ϑ we know the following:

1. $n \in \mathbb{N}$ events $S_1, ..., S_n$ are located in $[0, T]$ and were placed by the beats $B_1, ..., B_N$.
2. Beat B_k placed n_k events in $[0, T]$ and $\Delta_k := P_k^* - n_k$ events in $(-\infty, 0) \cup (T, \infty)$, where P_k^* is lower-truncated Poisson distributed, $P_k^* \sim Pois(\gamma \mid \{n_k, n_k + 1, ...\})$.
3. All beats B_k with $k \notin \{1, ..., N\}$ place events in $(-\infty, 0) \cup (T, \infty)$ only.

Therefore, not only events $S_{k,i}$ with $k \notin \{1, ..., N\}$ but also beats B_k with $k \notin \{1, ..., N\}$ exhibit a posteriori a truncated distribution. However, for the sake of simplicity we will omit the truncation at several points in the following. Thus, we will only obtain approximations for the quantities of interest. Nonetheless, these approximations are very accurate if the GLO process is regular (i.e. if σ_2/μ and σ_1/μ are small), since then it is very probable that beats B_k with $k \notin \{1, ..., N\}$ and corresponding events are not in $[0, T]$ as indicated by Remark 4.1 and Remark 2.3, p. 15, respectively.

Remark 4.1 (Truncation of the background rhythm)**.** *For the distribution of beat B_{N+1} we have to distinguish between two cases:*

1) *If B_{N+1} places no events, i.e. $P_{N+1} = 0$, then B_{N+1} is able to be located in $[0, T]$, thus, $B_{N+1} \sim \mathcal{N}(B_N + \mu, \sigma_1^2)$.*

2) *If B_{N+1} does place events, i.e. $P_{N+1} > 0$, then B_{N+1} can not be in*

$$[-\max Z_{N+1},\ T - \min Z_{N+1}], \tag{4.2}$$

where $\min Z_{N+1} := \min_{i=1,...,P_k} Z_{N+1,i}$ and $\max Z_{N+1} := \max_{i=1,...,P_k} Z_{N+1,i}$ and $Z_{N+1,1}, ..., Z_{N+1,P_k} \overset{iid}{\sim} \mathcal{N}(0, \sigma_2^2)$, since otherwise there would exist an additional event $S_{N+1,i} = B_{N+1} + Z_{N+1,i} \in [0, T]$. Thus, $B_{N+1} \sim \mathcal{N}(B_N + \mu, \sigma_1^2 \mid I_{N+1})$, where

$$I_{N+1} := \mathbb{R} \setminus [-\max Z_{N+1},\ T - \min Z_{N+1}]. \tag{4.3}$$

Therefore,

$$\begin{aligned}
&\mathbb{P}(B_{N+1} \in [a, b] \mid \mathcal{S}) \\
&= \mathbb{P}(B_{N+1} \in [a, b], P_{N+1} = 0 \mid \mathcal{S}) + \mathbb{P}(B_{N+1} \in [a, b], P_{N+1} > 0 \mid \mathcal{S}) && (4.4)\\
&= e^{-\gamma} \int_a^b f_{\mathcal{N}(B_N+\mu,\sigma_1^2)}(u)\, du + \left(1 - e^{-\gamma}\right) \int_a^b f_{\mathcal{N}(B_N+\mu,\sigma_1^2 \mid I_{N+1})}(u)\, du, && (4.5)
\end{aligned}$$

where the density of the truncated normal distribution is given as (cf. Griffiths (2002))

$$f_{\mathcal{N}(B_N+\mu,\sigma_1^2 \mid I_{N+1})}(x) = \frac{f_{\mathcal{N}(B_N+\mu,\sigma_1^2)}(x)}{\int_{I_{N+1}} f_{\mathcal{N}(B_N+\mu,\sigma_1^2)}(u)\, du}, \tag{4.6}$$

for $x \in I_{N+1}$ and 0 otherwise. This is similar for all beats B_k with $k \notin \{1, ..., N\}$. Moreover, if $B_N + \mu > T - \min Z_{N+1}$, we have

$$\int_{I_{N+1}} f_{\mathcal{N}(B_N+\mu,\sigma_1^2)}(u)\, du \to 0, \text{ for } \sigma_1 \to 0. \tag{4.7}$$

Thus, for small σ_1 the truncation does not noticeably change the distribution, i.e. $f_{\mathcal{N}(B_N+\mu,\sigma_1^2 \mid I_{N+1})} \approx f_{\mathcal{N}(B_N+\mu,\sigma_1^2)}$. Similarly, $\min Z_{N+1}$ weakly converges to zero if σ_2 approaches zero. Hence, in case of small σ_1 and σ_2, and $B_N + \mu \geq T + \varepsilon$ with sufficiently large $\varepsilon > 0$, we may approximate the forecast by carrying on the background rhythm without accounting for the truncation, i.e. by taking $B_{N+1} \sim N(B_N + \mu, \sigma_1^2)$.

4.2.1. The a posteriori expected number of events

Firstly, we will focus on the expected number of events in $(t_1, t_2]$ with $T \leq t_1 < t_2$. For this purpose, we obtain from Lemma 2.1, p. 17, that conditional on knowing $B_1, ..., B_N$ and the parameter values η, the expected number of events in $(t_1, t_2]$ is given by

$$\begin{aligned}\mathbb{E}[\Phi(t_1,t_2) \mid B_1,...,B_N,\eta] &= \gamma \sum_{k=1}^{N} \left(F_{\mathcal{N}(B_k,\sigma_2^2)}(t_2) - F_{\mathcal{N}(B_k,\sigma_2^2)}(t_1)\right) \\ &+ \gamma \sum_{k\in\{-1,-2,...\}} \left(F_{\mathcal{N}(B_1+k\mu,\sigma_2^2+|k|\sigma_1^2)}(t_2) - F_{\mathcal{N}(B_1+k\mu,\sigma_2^2+|k|\sigma_1^2)}(t_1)\right) \\ &+ \gamma \sum_{k\in\{1,2,...\}} \left(F_{\mathcal{N}(B_N+k\mu,\sigma_2^2+|k|\sigma_1^2)}(t_2) - F_{\mathcal{N}(B_N+k\mu,\sigma_2^2+|k|\sigma_1^2)}(t_1)\right).\end{aligned} \tag{4.8}$$

If we do not account for the truncation of beats B_k, $k \notin \{1, ..., N\}$, and corresponding events, the a posteriori expected number of events in $(t_1, t_2]$ can similarly be approximated by

$$\mathbb{E}[\Phi(t_1,t_2) \mid \mathcal{S}] = \mathbb{E}\left[\mathbb{E}\left[\Phi(t_1,t_2) \mid \vartheta, \mathcal{S}\right] \mid \mathcal{S}\right] = \mathbb{E}\left[\mathbb{E}\left[\sum_{k\in\mathbb{Z}}\sum_{i=1}^{P_k} \mathbb{1}_{\{S_{k,i}\in(t_1,t_2]\}} \mid \vartheta, \mathcal{S}\right] \mid \mathcal{S}\right] \tag{4.9}$$

$$\begin{aligned}&\approx \sum_{k=1}^{N} \mathbb{E}[P_k^* - n_k \mid \mathcal{S}]\, \mathbb{E}\left[F_{\mathcal{N}(B_k,\sigma_2^2\mid(-\infty,0)\cup(T,\infty))}(t_2) - F_{\mathcal{N}(B_k,\sigma_2^2\mid(-\infty,0)\cup(T,\infty))}(t_1) \mid \mathcal{S}\right] \\ &+ \mathbb{E}\Bigg[\gamma \sum_{k\in\{1,2,...\}} \left(F_{\mathcal{N}(B_1-k\mu,\sigma_2^2+|k|\sigma_1^2)}(t_2) - F_{\mathcal{N}(B_1-k\mu,\sigma_2^2+|k|\sigma_1^2)}(t_1)\right) \\ &+ \gamma \sum_{k\in\{1,2,...\}} \left(F_{\mathcal{N}(B_N+k\mu,\sigma_2^2+|k|\sigma_1^2)}(t_2) - F_{\mathcal{N}(B_N+k\mu,\sigma_2^2+|k|\sigma_1^2)}(t_1)\right) \mid \mathcal{S}\Bigg].\end{aligned} \tag{4.10}$$

The pmf of the lower-truncated Poisson distribution $P_k^* \sim Pois(\gamma \mid \{n_k, n_k + 1, ...\})$ is given as

$$\mathbb{P}(P_k^* = u) = \frac{1}{\mathbb{P}(P_k \geq n_k)} \frac{\gamma^k}{k!} e^{-\gamma}, \quad \text{for } k \geq n_k, \tag{4.11}$$

where $P_k \sim Pois(\gamma)$. Thus, for the expected value we yield

$$\mathbb{E}[P_k^* \mid \gamma, n_k] = \frac{1}{1 - F_{Pois(\gamma)}(n_k - 1)} \sum_{k=n_k}^{\infty} k \frac{\gamma^k}{k!} e^{-\gamma} \tag{4.12}$$

$$= \frac{1}{1 - F_{Pois(\gamma)}(n_k - 1)} \left(\sum_{k=0}^{\infty} k \frac{\gamma^k}{k!} e^{-\gamma} - \sum_{k=0}^{n_k-1} k \frac{\gamma^k}{k!} e^{-\gamma}\right) \tag{4.13}$$

$$= \gamma \frac{1 - F_{Pois(\gamma)}(n_k - 2)}{1 - F_{Pois(\gamma)}(n_k - 1)}. \tag{4.14}$$

Therefore, to approximate $\mathbb{E}[\Phi(t_1,t_2) \mid \mathcal{S}]$ we can use samples $\vartheta^{(1)},...,\vartheta^{(R)}$ from the posterior distribution and evaluate

$$
\begin{aligned}
&\mathbb{E}[\Phi(t_1,t_2) \mid \mathcal{S}] \\
&\approx \frac{1}{R}\sum_{l=1}^{R}\left[\sum_{k=1}^{N}\left(\gamma^{(r)}\frac{1-F_{Pois\left(\gamma^{(r)}\right)}\left(n_k^{(r)}-2\right)}{1-F_{Pois\left(\gamma^{(r)}\right)}\left(n_k^{(r)}-1\right)}-n_k^{(r)}\right)\right. \qquad (4.15)\\
&\qquad\times\left(F_{\mathcal{N}\left(B_k^{(r)},\left(\sigma_2^{(r)}\right)^2\mid(-\infty,0)\cup(T,\infty)\right)}(t_2)-F_{\mathcal{N}\left(B_k^{(r)},\left(\sigma_2^{(r)}\right)^2\mid(-\infty,0)\cup(T,\infty)\right)}(t_1)\right)\\
&+\gamma^{(r)}\sum_{k\in\{1,2,...\}}\left(F_{\mathcal{N}\left(B_1^{(r)}-k\mu^{(r)},\left(\sigma_2^{(r)}\right)^2+k\left(\sigma_1^{(r)}\right)^2\right)}(t_2)-F_{\mathcal{N}\left(B_1^{(r)}-k\mu^{(r)},\left(\sigma_2^{(r)}\right)^2+k\left(\sigma_1^{(r)}\right)^2\right)}(t_1)\right)\\
&\left.+\gamma^{(r)}\sum_{k\in\{1,2,...\}}\left(F_{\mathcal{N}\left(B_N^{(r)}+k\mu^{(r)},\left(\sigma_2^{(r)}\right)^2+k\left(\sigma_1^{(r)}\right)^2\right)}(t_2)-F_{\mathcal{N}\left(B_N^{(r)}+k\mu^{(r)},\left(\sigma_2^{(r)}\right)^2+k\left(\sigma_1^{(r)}\right)^2\right)}(t_1)\right)\right].
\end{aligned}
$$

Note that we do not account for the truncation of new beats $B_k,\ k\notin\{1,...,N\}$, and corresponding events. Thus, the resulting approximation is accurate only for small σ_1 and σ_2, and if $\mathbb{P}(B_N+\mu\geq T+\varepsilon \mid \mathcal{S})$ is large with sufficiently large $\varepsilon>0$, as stated above.

4.2.2. The a posteriori probability of default

Secondly, we will focus on the a posteriori probability of default at time $t>T$, i.e.

$$\Psi_T(t,u)=\mathbb{P}\left(u+c(t-T)-L(T,t)<0 \mid \mathcal{S}\right), \qquad (4.16)$$

where $u\geq 0$ denotes the initial capital at time T, $c>0$ the premium income rate and $L(T,t)=\sum_{i=1}^{\Phi(T,t)}Y_i$ the aggregate claims occurring in $(T,t]$. Moreover, we assume that $Y_1,Y_2,...$ are independent and identically distributed according to F_Y and independent from Φ. It follows that

$$
\begin{aligned}
\Psi_T(t,u)&=\mathbb{E}\left[1_{\{u+c(t-T)-L(T,t)<0\}} \mid \mathcal{S}\right] \qquad (4.17)\\
&=\mathbb{E}\left[\mathbb{E}\left[1_{\left\{u+c(t-T)<\sum_{i=1}^{\Phi(T,t)}Y_i\right\}} \mid \vartheta,\mathcal{S}\right] \mid \mathcal{S}\right], \qquad (4.18)
\end{aligned}
$$

which can be approximated by means of Monte-Carlo methods as follows: For every sample $\vartheta^{(r)}$ from the posterior distribution simulate $M\in\mathbb{N}$ second parts of the observed GLO process and claim sizes Y_i. In other words, for all $r=1,...,R$ and $m=1,...,M$ draw samples for $\Phi(T,t)$ and $Y_1,....,Y_{\Phi(T,t)}$ conditionally on $\mathcal{S}$ and $\vartheta^{(r)}$.

Recall from Proposition 2.3, p. 17, that

$$\mathcal{L}\left(\Phi(T,t) \mid \mathcal{B}\right)=Pois\left(\gamma\sum_{k\in\mathbb{Z}}\int_T^t f_{\mathcal{N}(B_k,\sigma_2^2)}(u)\,du\right). \qquad (4.19)$$

However, for beats B_k with $k \in \{1, ..., N\}$ a posteriori we already know how many events were placed in $[0, T]$. The remaining number of events is given by Δ_k as defined above. Since conditional on the background rhythm the event locations are independent, we can sample the number of events placed by $B_k,\ k \in \{1, ..., N\}$, independently from the events originating from $B_k,\ k \notin \{1, ..., N\}$. We propose the following procedure for every posterior sample $r = 1, ..., R$ and iteration $m = 1, ..., M$:

1) For all beats $B_k^{(r)}, k \in \{1, ..., N\}$, draw

$$P_k^{(r,m)} \sim Pois\left(\gamma^{(r)} \mid \left\{n_k^{(r)}, n_{k+1}^{(r)}, ...\right\}\right) \tag{4.20}$$

and place $\Delta_k = P_k^{(r,m)} - n_k^{(r)}$ iid events outside of $[0, T]$, i.e.

$$S_{k,n_k^{(r)}+1}^{(r,m)}, ..., S_{k,P_k^{(r,m)}}^{(r,m)} \sim \mathcal{N}\left(B_k^{(r)}, \left(\sigma_2^{(r)}\right)^2 \mid (-\infty, 0) \cup (T, \infty)\right). \tag{4.21}$$

2) Carry on the background rhythm in both directions, i.e. model two random walks: The first starts in $B_N^{(r)}$ with mean increment $\mu^{(r)}$ and ends when it reaches the height $t + K$. The second starts in $B_1^{(r)}$ with mean increment $-\mu^{(r)}$ and ends when it is below $T - K$.

3) Draw the number of events in $(T, t]$ originating from beats $\left\{B_k^{(r)} : k \notin \{1, ..., N\}\right\}$, i.e.

$$Q^{(r,m)} \sim Pois\left(\gamma^{(r)} \sum_{k \notin \{1,...,N\}} \int_T^t f_{\mathcal{N}\left(B_k^{(r)}, \left(\sigma_2^{(r)}\right)^2 \mid (-\infty,0) \cup (T,\infty)\right)}(u)\, du\right) \tag{4.22}$$

4) Set $\Phi^{(r,m)}(T, t) = Q^{(r,m)} + \left|\left\{\ S_{k,j}^{(r,m)} \in (T, t] : k \in \{1, ..., N\}, j \in \left\{n_k^{(r)} + 1, ..., P_k^{(r,m)}\right\}\right\}\right|$.

5) Draw claim sizes $Y_1, ..., Y_{\Phi^{(r,m)}(T,t)} \overset{iid}{\sim} F_Y$.

6) Evaluate $h^{(r,m)} = \mathbb{1}_{\left\{u + c(t-T) < \sum_{i=1}^{\Phi^{(r,m)}(T,t)} Y_i\right\}}$.

Then, the probability of default can be approximated by

$$\Psi_T(t, u) \approx \frac{1}{MR} \sum_{r=1}^{R} \sum_{m=1}^{M} h^{(r,m)}. \tag{4.23}$$

Note that during this procedure we only account for the truncation of events but not for the newly simulated background rhythm. Thus, the resulting approximation is accurate only if σ_1 and σ_2 are small and $\mathbb{P}(B_N + \mu \geq T + \varepsilon \mid \mathcal{S})$ is large with sufficiently large $\varepsilon > 0$, as stated above.

Remark 4.2 (Simulation parameters)**.** *K should be chosen such that both the probability that beats outside $[T-K, t+K]$ place events in $(T,t]$ and the probability that the random walk jumps back from beats outside $[T-K, t+K]$ into $(T,t]$ are small. Moreover, when applying the simulation procedure it will be useful to determine constants k_1 and k_N that correspond to the number of beats to cross $T-K$ and $t+K$, respectively. Here we present guidelines for choosing these parameters, which are similar to the guidelines for the GLO simulation algorithm from Bingmer (2012, p.41 and p.137). For this purpose, we will neither consider the truncation of events nor beats.*

Firstly, we focus on the probability that beats place events in $(T,t]$: Suppose that there is a beat at $t+K'$. The location of a random event originating at this beat is $S \sim \mathcal{N}(t+K', \sigma_2^2)$, thus, we want $\mathbb{P}(S \leq t) \leq \varepsilon$ for a small $\varepsilon \in (0,1)$. It follows that

$$K' \geq -q_\varepsilon \sigma_2, \tag{4.24}$$

where $q_\varepsilon = F^{-1}_{\mathcal{N}(0,1)}(\varepsilon)$ is the ε-quantile of the standard Normal distribution.

Secondly, we analyze the random walk. As Bingmer (2012, p.39-40) shows by applying results for ruin probabilities of random walks, the probability that a random walk that is started at 0 with mean increment $-\mu < 0$ and variance σ_1^2 crosses $a > 0$ is bounded by $\varepsilon \in (0,1)$ if

$$a \geq \frac{\sigma_1^2}{2\mu} \log \varepsilon =: K''. \tag{4.25}$$

This is the same situation as having a random walk that is started at $t+K''$ with $\mu > 0$ and determining K'' such the random walk jumps below t only with probability ε. Therefore, with $K = \max\{K', K''\}$ in our setting we may consider all beats in the interval

$$[T-K, t+K], \tag{4.26}$$

Finally, we obtain the number of beats to cross $T-K$ and $t+K$: First, focus on $t+K$ and the upper random walk, that starts at B_N. Thus, $B_{N+k} - B_N \sim \mathcal{N}(k\mu, k\sigma_1^2)$. $k_N \in \mathbb{N}$ should be the smallest index value, such that the probability for the random walk to be smaller than $t+K$ is smaller or equal than $\varepsilon \in (0,1)$. As Bingmer (2012, p.42) shows, this value is given by

$$k_N \geq \frac{2(t+K-B_N)\mu + q_\varepsilon^2 \sigma_1^2}{2\mu^2} + \frac{1}{2\mu}\sqrt{q_\varepsilon^2\sigma_1^2(4(t+K-B_N) + q_\varepsilon^2\sigma_1^2}. \tag{4.27}$$

Similarly, the value for k_1 can be obtained, which is given as

$$k_1 \geq \frac{2(B_1-(T-K))\mu + q_\varepsilon^2 \sigma_1^2}{2\mu^2} + \frac{1}{2\mu}\sqrt{q_\varepsilon^2\sigma_1^2(4(B_1-(T-K)) + q_\varepsilon^2\sigma_1^2}. \tag{4.28}$$

Therefore, in step 2) we may simulate $B_{N+1}, ..., B_{N+k_N}$ and $B_0, ..., B_{-k_1+1}$.

4.3. A collective risk model for hail claims

In this section we will review a dataset of claims that were caused by hail damage of one of the six largest (according to premium income in 2014) insurance companies in Germany. Every datapoint corresponds to one particular day and the total claim size occurring at that day. To exclude trends due to (premium) growth of the company, we adjust the original (daily) claim sizes $Y_i^{(y)}$ in year y by the latest total premiums in 2014, c_{2014}, such that $Y_i'^{(y)} = Y_i^{(y)} c_{2014}/c_y$ is the premium-adjusted claim size. Due to reasons of information privacy the dates are reported in days starting at an unknown past date and instead of the premium-adjusted claim size Y_i' we only consider the normalized log-premium-adjusted claim size (in the following simply called *claim size*)

$$\widetilde{Y}_i = \frac{\log Y_i' - \overline{\log \mathcal{Y}'}}{\sqrt{s^2(\log \mathcal{Y}')}}. \tag{4.29}$$

Moreover, claims were made at almost every day in the dataset. However, in this example we will only analyze events with a premium-adjusted claim size Y_i' that exceeds a certain threshold value $\alpha > 0$, the *Peaks over Threshold* (POT) process. This type of process is often considered in *Extreme-Value theory* (cf. Embrechts et al. (1997, p.352)). There are a few reasons for using the POT process: Firstly, for the business risk of an insurance company only large claims are relevant. Secondly, one may also interpret the POT as the occurrence times of claims made in context of an *Excess of loss*-reinsurance agreement. In this case the insurance company claims a loss to a reinsurer if the loss exceeds a certain threshold. Thirdly, and most practically, the POT exhibits the regular structure of a bursty GLO process as the rasterplot in Figure 4.10A and the ACH in Figure 4.9 indicate for $\alpha = 100000$.

Although the firing rate apparently increases, we will ignore this trend for the sake of simplicity (in section 4.4.2 we will discuss how trends may be embedded in the GLO model). We choose $\alpha = 100000$, which roughly corresponds to the 90.6%-quantile of the data, thus, we only analyze the roughly 10% largest premium-adjusted claims. Since the claims are aggregated over one day, we may assume that the POT's claims are direct and fast responses to events with a large impact and, thus, indicate the true occurrence of large hail events.

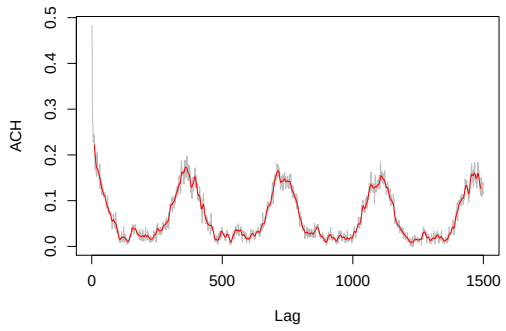
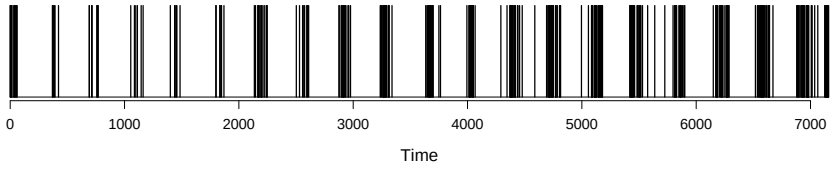


Figure 4.9.: Occurrence of Hail in Germany: ACH of the POT process for $\alpha = 100000$.

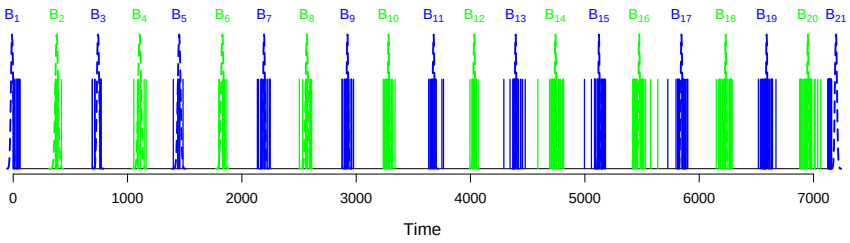
In a first step, we obtain $R = 35000$ MCMC posterior samples by employing the same fitting procedure as in section 4.1, particularly with Jeffreys and Binomial priors. The ACH estimates are $\hat{\eta}_{\text{ACH}} = (360.84, 0.0001, 47.79, 23.32)$. By comparing the expected log-likelihood for different values for N we yield $N = 21$ as optimal value and the corresponding parameter estimates reported in Table 4.6. The fit to the data is appropriate as Figure 4.10 indicates.

σ	N	$\bar{\ell}(\text{GLO}_N)$	BI	$\hat{\mu}_{\text{Median}}$	$\hat{\sigma}_{1,\text{Median}}$	$\hat{\sigma}_{2,\text{Median}}$	$\hat{\gamma}_{\text{Median}}$
(Simple guess)	22	- 2347.15	100	342.82	79.65	36.88	21.6
(Simple guess)	**21**	**- 2346.29**	**100**	**360.31**	**32.11**	**36.91**	**22.76**
$\hat{\sigma}_{2,\text{ACH}} = 47.79$	20	- 2522.06	100	373.19	25.35	55.51	23.22

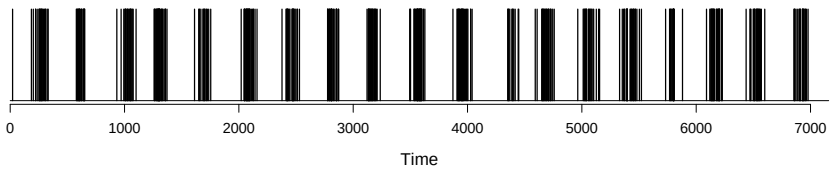
Table 4.6.: Occurrence of hail in Germany: Bandwidth, number of beats, expected log-likelihood, burn-in period and posterior parameter medians.



(A) Observed events (with $\alpha = 100000$)



(B) Observed events (solid), posterior beat densities (dashed) and labels (colors) ($N = 21$)



(C) GLO simulation with $\hat{\eta}_{\text{Median}}$ ($N = 21$)

Figure 4.10.: Occurrence of hail in Germany: (A) Data, (B) posterior beat densities and (C) simulation.

In a second step, we fit the corresponding claim sizes of the POT process to a loss distribution. For this purpose we consider the Exponential, Log-Normal and shifted Gamma distribution. The latter is given by the distribution of $\tilde{Y} + \kappa$ where $\tilde{Y} \sim Ga(\alpha, \beta)$ and $\kappa \in \mathbb{R}$. The parameters of all three distributions are fitted by the method of moments and the empirical and fitted densities are shown in Figure 4.11.

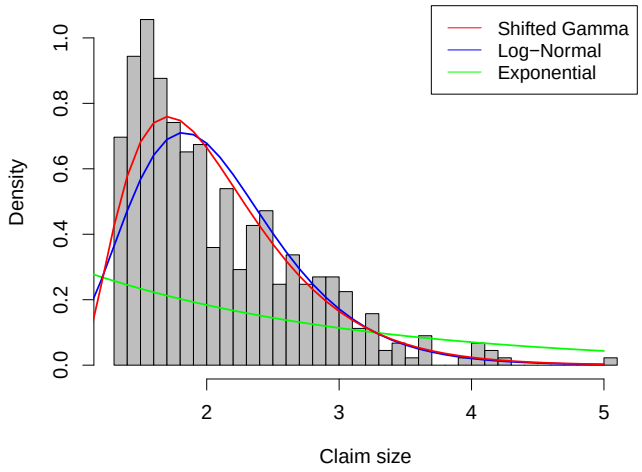


Figure 4.11.: Claim sizes due to hail in Germany: Empirical (black/grey) and fitted shifted Gamma (red), Log-Normal (blue) and Exponential (green) distribution.

Since in this work we focus on the claim arrival process, we do not need an optimal fit of the claim size distribution and evaluate the three distributions only visually by comparing the (fitted) theoretical and empirical densities. Despite of the additional third parameter, the shifted Gamma distribution exhibits only a slightly better fit than the Log-Normal distribution. Since both indicate a better fit than the Exponential distribution, we choose the Log-Normal distribution with fitted parameters $m = 0.69$ and $s = 0.29$, i.e. we assume that $\log \tilde{Y}_1, ..., \log \tilde{Y}_n \overset{iid}{\sim} \mathcal{N}(m, s^2)$. However, note that due to the truncation of the process left-truncated loss distributions may provide a better fit. Several of these are analyzed in Chernobai et al. (2005).

4.3.1. Forecasting the probability of default

In order to illustrate the use of the GLO model for risk management purposes and to compare it to the often used homogeneous Poisson process we apply the forecasting method for the probability of default which was proposed in section 4.2.2. For this purpose we assume that an insurance company in fact observed the claims shown in Figure 4.12 that were simulated according to a GLO process in the time window $[0, T]$, where the parameter values are the estimates $\hat{\eta}_{\text{Median}}$ (with $N = 21$) for the occurrence of hail in section 4.3. However, the company is uncertain about the true parameter values (and past beat locations and labels) of the GLO process.

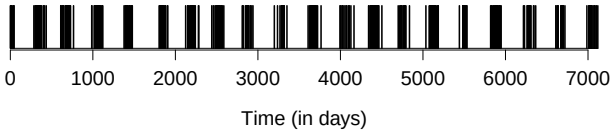


Figure 4.12.: Rasterplot of simulated data.

Furthermore, we suppose that the insurance company knows, that the claim sizes are independent, identically and log-normally distributed with parameters $m = 0.69$ and $s = 0.29$ as the fitted claim sizes due to hail from section 4.3 (i.e., we do not consider parameter uncertainty for the claim sizes). At time T we assume arbitrary values for the company's initial capital $u \geq 0$ and the continuous premium payments $c \geq 0$ as reported in Table 4.7.

Record times, $\mathbb{T}$	**Initial capital,** u	**Premium rate,** c	**Sample size,** M
6721, 6850, 6950, 7159	12	40/364	150

Table 4.7.: Exemplary values for the surplus process.

For the analysis we will consider the same observed events but, however, truncate these at four different values $T \in \mathbb{T} := \{6721, 6850, 6950, 7159\}$. For every $T \in \mathbb{T}$ we apply the following procedure:

Firstly, we apply the MCMC algorithm (Algorithm 3.1) analogously to section 4.3 (with Jeffreys and Binomial priors and the true parameter values as initial values) to obtain $R = 20000$ samples (after a burn-in period of $BI = 100$) from the posterior distribution.

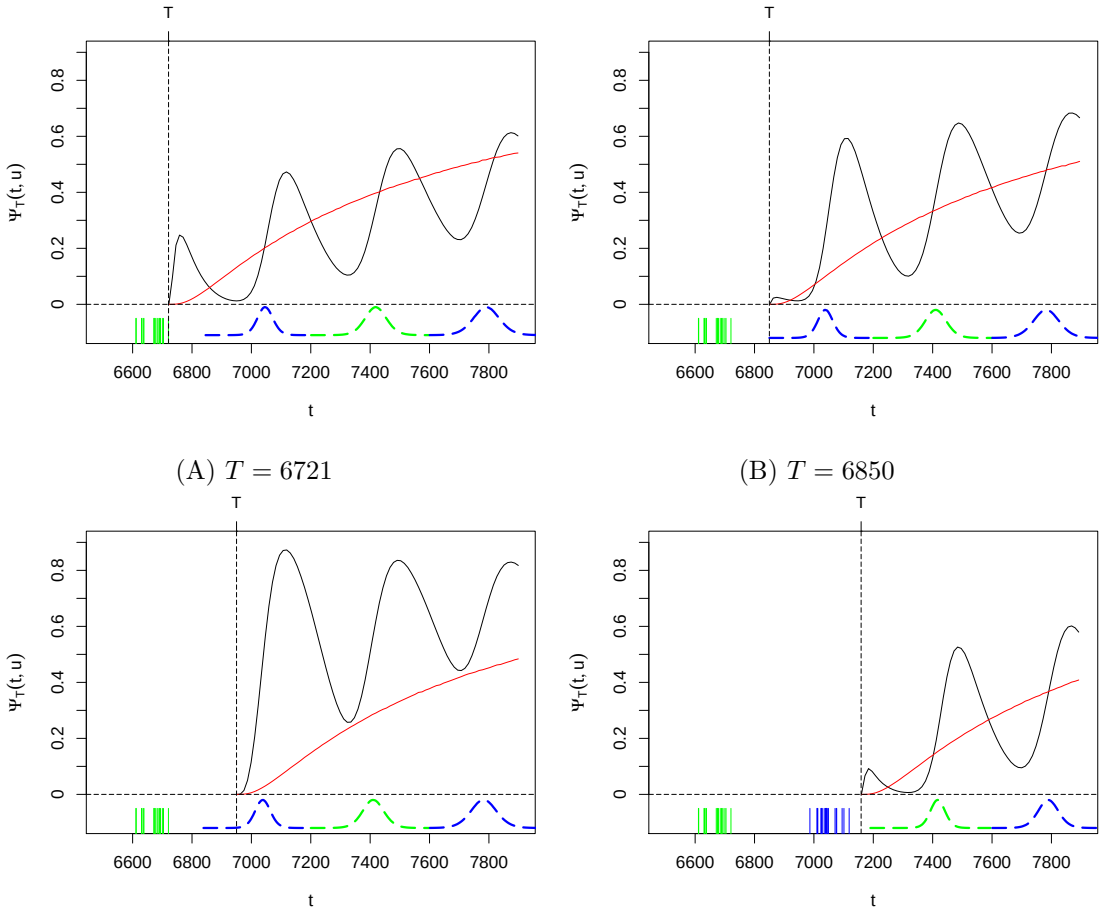
Secondly, we apply the Monte Carlo procedure described in section 4.2.2 to approximate the a posteriori probability of default (PD) at time $t > T$, $\Psi_T(t, u)$. Particularly, we do not account for the truncation of the background rhythm since the true values of σ_1 and σ_2 are rather small and $\mathbb{P}(B_N + \mu - T > \varepsilon)$ is sufficiently large with large $\varepsilon > 0$ for all $T \in \mathbb{T}$ (cf. Remark 4.1).

Additionally, we fit the events to a homogeneous Poisson process φ. To obtain comparability to $\Psi_T(t, u)$, we also account for parameter uncertainty. Since the inter-event-intervals $IEI_1, ..., IEI_{n-1}$ are independent and $Exp(\lambda)$-distributed under the Poisson process-hypothesis, we perform a Bayesian estimation with a Jeffreys prior for λ. Fisher's expected information is for the $Exp(\lambda)$-distribution given as $I(\lambda) = \mathbb{E}\left[-\frac{\partial^2 \log L(\lambda; IEI_1)}{\partial \lambda^2}\right] = \lambda^{-2}$, thus, the Jeffreys prior is given as $\pi(\lambda) \propto \lambda^{-1}$. Therefore, λ is a posteriori distributed as

$$p(\lambda \mid IEI) \propto \frac{1}{\lambda} \prod_{i=1}^{n-1} \lambda e^{-\lambda IEI_i} = \lambda^{n-2} e^{-\lambda \sum_{i=1}^{n-1} IEI_i}, \tag{4.30}$$

which is proportional to a $Ga(n-1, \sum_{i=1}^{n-1} IEI_i)$ distribution.

Moreover, due to the memorylessness of the Exponential distribution, the number of events in $(T,t]$ is conditional on λ given as $\varphi(T,t) \sim Pois(\lambda(t-T))$. Hence, to approximate the a posteriori probability of default, $\Psi_T(t,u)^{(\text{Pois})}$, we draw $R = 30000$ posterior samples for $\lambda \sim Ga(n-1, \sum_{i=1}^{n-1} IEI_i)$. Then, for each of these samples we draw $M = 120$ samples of $\varphi(T,t)$, claim sizes $Y_1, ..., Y_{\varphi(T,t)}$ and evaluate $h^{(r,m)}$ similar to the procedure in section 4.2.2.



(A) $T = 6721$ (B) $T = 6850$

(C) $T = 6950$ (D) $T = 7159$

Figure 4.13.: *Upper part*: A posteriori probability of default for the GLO (solid, black) and homogeneous Poisson process (solid, red) beginning (A)/(D) after bursts, (B) between bursts and (C) before a burst. *Lower part*: Observed events (solid), forecasted beat densities (dashed) and labels (colors).

In Figure 4.13 the resulting PDs are shown. As one might expect, $\Psi_T(t,u)$ increases substantially at the time a beat might occur, and decreases when the occurrence of events becomes unlikely and premiums increase the insurer's capital. If the observation window ends directly after bursts (Figures 4.13A and 4.13D), $\Psi_T(t,u)$ starts with a steep peak since events might be placed near T by beats from inside the observation window. This peak is rather small if T is between two bursts (Figures 4.13B). If, however, T is located directly

before a burst (Figure 4.13C), $\Psi_T(t,u)$ immediately increases with a very large peak since the insurer was not able to save premiums before the first events occur.

In contrast, $\Psi_T(t,u)^{(\text{Pois})}$ is monotone increasing with time, since the arrival process exhibits a constant intensity. Therefore, $\Psi_T(t,u)^{(\text{Pois})}$ is smaller than $\Psi_T(t,u)$ in bursty times, and larger in times without any (GLO) events if the forecast begins directly after or between bursts (Figures 4.13A, 4.13B and 4.13D). If, however, T is immediately followed by a burst (Figure 4.13C), $\Psi_T(t,u)$ is larger than $\Psi_T(t,u)^{(\text{Pois})}$ for the first few years since the insurer's losses are immediately very large compared to the insurer's capital (initial surplus and premiums).

To conclude, an insurance company that is exposed to claims that arrive according to a GLO process is particularly exposed to bursty periods with many claims. Moreover, the probability of default considerably depends on the location of the latest burst in relation to the beginning of the forecast. If, nonetheless, the claim arrivals are modeled with a homogeneous Poisson process, the oscillatory behavior of claim arrivals is not accounted for and, thus, the PD is largely underestimated in bursty times.

4.4. Possible model extensions

In this section we will discuss several possible model extensions for the GLO and analyze how they might by incorporated in the Bayesian estimation framework.

4.4.1. Negative Binomial distribution

For some processes in section 4.1 we observed a larger variance than mean of the number of events in bursts. In other words, the size of the beats' families were overdispersed. Therefore, a Negative Binomial distribution (cf. section 1.1.1) might be more preferable than the Poisson distribution, i.e. $P_k \sim NB(r,p)$ with pmf

$$\mathbb{P}(P_k = l) = \binom{l+r-1}{l} p^l (1-p)^r. \tag{4.31}$$

However, incorporating the Negative Binomial distribution in the Bayesian framework is not straightforward since there is no standard distribution serving as conjugate prior. One may consider two possible ways how to perform Bayesian inference in a Negative Binomial model:

First, one may calculate the posterior distribution for a certain prior distribution as it is done in Ganji et al. (2013). However, their resulting posterior distribution is a non-standard distribution, for which it is not clear how to sample from. Thus, their results may be used by employing the Metropolis-Hastings algorithm.

Second, we may employ the following Lemma:

Lemma 4.1. *(Liu and Dey (2007)) If* $\mathcal{L}(P \mid \gamma) = Pois(\gamma)$ *and* $\gamma \sim Ga(r, \frac{1-p}{p})$*, then* $\mathcal{L}(P) = NB(r, p)$*.*

Thus, in the MCMC algorithm one may estimate r and p conditionally on γ. This method is called *hierarchical Bayes method*. For example, Liu and Dey (2007) review Bayesian analysis of several hierarchical non-standard Poisson regression models. They incorporate the Negative Binomial model as a special case. However, for being able to employ the hierarchical Bayes method with Gibbs updating steps, one needs to find conjugate priors for the Gamma distribution which can be sampled from. This is straightforward for the first but harder for the second parameter (cf. Yang and Berger (1998)). Addressing this issue, Moala et al. (2013) and Son and Oh (2006) present ways how to employ MCMC methods in case of non-informative priors for the parameters of the Gamma distribution, whereas Pradhan and Kundu (2011) develop a Gibbs sampling procedure for informative priors.

4.4.2. Trends

It is straightforward to include a trend e.g. for P_k (as one can e.g. observe in section 4.3 for the occurrence of hail in Germany). For example, one may propose a model similar to *General Linear Models* (GLMs) with

$$\mathcal{L}\left(P_k \mid \gamma_k\right) = Pois(\gamma_k) \quad \text{and} \quad \log(\gamma_k) = \gamma_0 + \beta B_k. \tag{4.32}$$

An introduction to Bayesian inference for GLMs may be found in Wakefield (2013, p.273ff.). However, in such a model it is difficult to find conjugate prior distributions that one can sample from and, thus, it may be necessary to employ the Metropolis-Hastings algorithm.

4.4.3. Non-Gaussian distributions

For both the background rhythm as well as the event distribution one may consider other distributions. By changing the distribution of beat increments one may e.g. allow for heavy tails (for example with a t (student) distribution). Alternatively, one may employ distributions that take only positive values, like the Exponential, Log-Normal or Weibull distribution, such that beat locations are monotone increasing and the background rhythm follows a renewal process. Then, the updating procedure in the MCMC procedure (Algorithm 3.1, p. 46) can easily be adapted by employing conjugate priors or MH updates.

4.4.4. Non-bursty mode

Recall from Remark 2.1, p. 12, that Bingmer et al. (2011) allow for a *non-bursty* mode, in which the random number of events in one burst is Bernoulli-distributed, $P_k \sim Bernoulli(\gamma)$ with $\gamma \in [0, 1]$. By assuming a uniform, Jeffreys or Beta prior for γ we yield the Beta distribution as posterior distribution (cf. Yang and Berger (1998) and Hoff (2009, p.37-38)). Therefore, we can apply the Bayesian estimation framework for such a model by replacing the Gamma distribution in step 1) in the MCMC procedure (Algorithm 3.1, p. 46) with the respective Beta distribution and replacing the Poisson distribution in the remaining steps with the Bernoulli distribution.

4.4.5. Modeling catastrophes and claims

In this work we applied the GLO model for two different types of events: the occurrence of catastrophes (namely tropical cyclones, cf. section 4.1) and the occurrence of claims (due to hail damage, cf. section 4.3). However, it is also possible to model both, bursts of catastrophes and claims caused by the catastrophes, in one model. A straightforward extension of the GLO model might be given as follows:

1. **Background rhythm:** The beats $B_k, k \in \mathbb{Z}$, follow a stationary random walk.
2. **Number of events:** The number of events P_k at beat B_k is Poisson distributed, $P_k \sim Pois(\gamma_0)$.
3. **Event variation:** At every beat B_k all events are distributed around the beat with independent increments, i.e. $S_{k,i} = B_k + Z_{k,i}$ with $Z_{k,i} \overset{iid}{\sim} \mathcal{N}(0, \sigma_2^2)$.
4. **Number of claims:** For every event $S_{k,i}$ there occur $G_{k,i} \sim Pois(\gamma_1)$ claims.
5. **Claim variation:** At every event $S_{k,i}$ all claims are distributed with positive, independent increments, i.e.

$$C_{k,i,j} = S_{k,i} + W_{k,i,j} \text{ with } W_{k,i,j} \overset{iid}{\sim} F_C. \tag{4.33}$$

All random variables $Z_{i_1,i_2}, P_{i_3}, G_{i_4,i_5}, W_{i_6,i_7,i_8}$ are independent for all $i_1, i_3, i_4, i_6 \in \mathbb{Z}$ and $i_2, i_5, i_7, i_8 \in \mathbb{N}$. The Bayesian framework may easily be extended for this model since it essentially consists of an inner and outer GLO process.

4.4.6. Multivariate GLOs

The data suggests that tropical cyclones do not occur independently for different continents. In other words, the occurrence of an event (or burst) at some place increases (or decreases) the probability of an event (or burst) at a different place. For example, bursts in Americas seem to occur in times of low firing rates in Africa, as Figure 4.14 shows. Therefore, it is of great interest to model the joint behavior of several GLO processes.

The following methods may provide first leads about how to construct multivariate GLO processes: Bäuerle and Grübel (2005) discuss several methods for the construction of multivariate point processes with Poisson marginals and propose a class of models where events are produced by thinning and shifts from a homogeneous Poisson process. Such an external mechanism is avoided by Bäuerle and Grübel (2008) who assume that event intensities of the components are interacting.

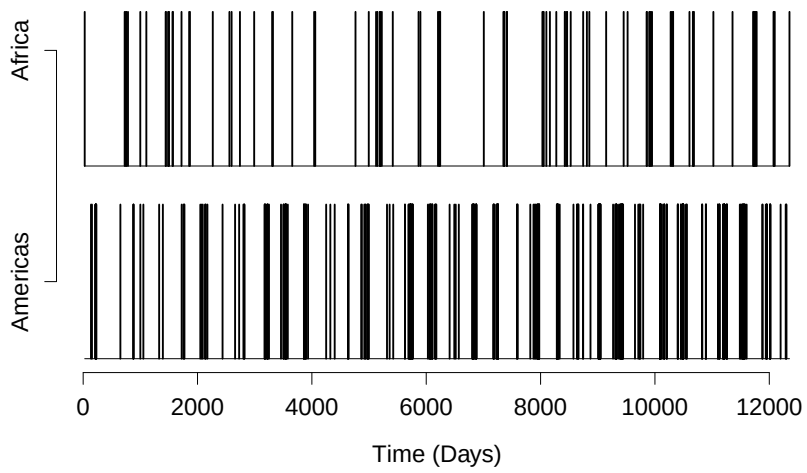


Figure 4.14.: Tropical cyclones in Africa and Americas: Rasterplot.

Furthermore, Pfeifer and Nešlehová (2004) review the construction of dependent Poisson variables by employing copulas, which are able to model various kinds of dependence structures (particularly, for any joint distribution function F with marginals $F_1, ..., F_n$ there exists a copula C such that $F(x_1, ..., x_n) = C(F_1(x_1), ..., F_n(x_n))$, cf. Embrechts et al. (2002)). They propose to construct dependent Poisson processes by either introducing multi-dimensional, dependent event-time points or dependent counting variables.

5. Conclusion

This work proposed to employ the (bursty) GLO model from Bingmer et al. (2011) to model the occurrence of tropical cyclones in Africa, Americas, Asia and Oceania. The GLO process is a doubly stochastic point process that is particularly suited to describe regularity and burstiness of a process. It builds on a stationary background process which is modeled as a stationary random walk. At every beat there is a Poisson distributed number of events. These events are distributed with independent, normally distributed increments around the beats. The GLO model is very easy to interpret in regard to the regularity of a process since it only uses four parameters: μ and σ_1 describe, respectively, the mean increment and variability of the background rhythm. γ is the average burst size (or firing rate), whereas σ_2 represents the burst width.

However, the background rhythm of a GLO process is not observable, which complicates the estimation procedure. Since the likelihood-function is not available in closed-form, Bingmer et al. (2011) propose to fit the four parameters by fitting the theoretical autocorrelation function (ACF) of the model to the empirical autocorrelation histogram (ACH) with a nonlinear least squares algorithm. Nonetheless, the resulting estimates exhibit large errors for small firing rates γ or irregular background rhythms (large σ_1/μ). Furthermore, it is not clear how to apply the fitting procedure for several extensions of the GLO model, and it is not straightforward to develop a forecasting method since beat locations need to be estimated with additional algorithms. Therefore, in this work we developed a different, Bayesian estimation framework for the GLO model.

In contrast to Frequentist inference, Bayesian inference assesses someone's subjective belief about an unknown parameter value. Therefore, parameters are interpreted as random variables. Before observations are made (*a priori*) someone's belief about the parameter value is expressed by the *prior* distribution $\pi(\vartheta)$. After observations $\mathcal{S}$ were made (*a posteriori*) the posterior distribution $p(\vartheta \mid \mathcal{S})$ incorporates the new knowledge about the parameter and can be calculated by employing Bayes' theorem.

In the context of the GLO model not only the four parameters μ, σ_1, σ_2 and γ but also the beat locations $\mathcal{B}$ and the labels $\mathcal{J}$ are unknown and, thus, need to be estimated. The labels indicate to which beat the events belong, i.e. $J_i = k$ if event S_i originates from beat B_k. Moreover, we do not know from how many beats the events originated. Therefore, we suggested to fix a certain number of beats, $N \in \mathbb{N}$, and presented methods how to evaluate which value for N leads to the best model. Particularly, we proposed to maximize the expected log-likelihood-function for this purpose.

Since it is not possible to calculate the marginal posterior distributions in closed-form, we developed a Markov chain Monte Carlo (MCMC) algorithm that builds on Gibbs (cf. Geman and Geman (1984)) and Metropolis-Hastings updates (cf. Metropolis et al. (1953) and Hastings (1970)) to sample values from the posterior distribution. We showed that the distribution of the samples converges to the target posterior distribution and presented methods how to choose initial values for the algorithm to achieve convergence very fast. Furthermore, for typical applications the algorithm is very robust with respect to the initial values.

Addressing the choice of a prior distribution, we reviewed several possible prior distributions and, particularly, distinguished between conjugate and Jeffreys priors. Whereas the first are proper distributions like e.g. the Normal distribution, the Jeffreys prior is proportional to the inverse of Fisher's expected distribution. This prior may not always integrate to one but, however, is generally invariant under reparametrization. Thus, it may be seen as *non-informative*. Moreover, its application still yields proper posterior distribution for the GLO model.

Therefore, the Jeffreys prior may be used whenever there is no prior information (or belief) about the observed phenomenon. However, we showed that in certain cases, in which σ_1 is very different from σ_2, assuming the Jeffreys prior for both of these parameters leads to high estimation errors: If a priori σ_1 and σ_2 are identically distributed, there exists a discrepancy between data and the prior assumption, thus, the posterior distribution of σ_1 or σ_2 may exhibit two peaks and the MCMC samples may be highly correlated. In such cases we suggested to assume different priors.

For label J_i we proposed to use a Binomial prior, i.e. $J_i \sim Bin(N, i/n)$. With this assumption, we have $J_n = N$ a.s., thus, the last label serves as anchor between random walk and events. If this link does not exist, e.g. with a uniform prior, $J_i \sim unif(\{1, ..., N\})$, we yield similar posterior distributions if the process is regular, but, however, the estimation procedure fails if the process is very irregular. Therefore, the Binomial prior represents a superior choice.

The posterior mean or median may serve as a Bayesian point estimate. For both, we analyzed the general estimation precision with a simulation study that covered nine different parameter combinations and three different record times. Generally, the estimation errors for σ_1 are larger than for the other parameters and the estimation precision of the posterior median is slightly better than for the posterior mean. In comparison to the ACH estimates from Bingmer et al. (2011) the posterior median yields a substantial improvement of the estimation precision for small firing rates γ, highly irregular background rhythms (large σ_1/μ) and large burst widths σ_2.

To assess the future development of the process after observing events we developed a forecasting framework. Particularly, we derived (Markov chain) Monte Carlo methods that approximate the a posteriori expected number of events in $(t_1, t_2]$ and the a posteriori probability of default $\Psi_T(t, u)$ of an insurance company. The latter builds on the commonly used *collective risk model* (cf. Bowers et al. (1997, p.367)), in which iid claims $Y_1, Y_2, ... \overset{iid}{\sim} F_Y$ occur according to a GLO process. The aggregate claims in $(T, t]$ are given as

$$L(t) = \sum_{i=1}^{\Phi(T,t)} Y_i. \tag{5.1}$$

Then, $\Psi_T(t, u)$ is the posteriori probability that $L(t)$ is larger than the insurer's capital. By applying the model and forecasting framework to a GLO process which is calibrated on large claims due to hail damage of a major German insurance company, we illustrated that $\Psi_T(t, u)$ exhibits a oscillatory behavior if Φ follows a GLO process. Moreover, if one bases the forecasting on a homogeneous Poisson process, which is the standard type of arrival process in risk theory (cf. Embrechts et al. (1997, p.22)), the probability of default is largely underestimated in bursty times.

Additionally, we suggested several extensions of the GLO model that are possible to incorporate in the Bayesian estimation framework, like e.g. Negative Binomially distributed numbers of events in bursts. However, for some of these some effort is still needed to derive the appropriate theoretical background and/or MCMC updates. Furthermore, determining the expected-log-likelihood-maximizing number of beats, N, often takes a lot of computation time. Therefore, the MCMC updating procedure may be embedded into a reversible jump Markov chain Monte Carlo (RJMCMC) algorithm (cf. Green (1995)) to estimate the parameters and N simultaneously.

In conclusion, we developed a Bayesian framework to estimate the parameters and (past) beats of a GLO process. By applying a Markov chain Monte Carlo algorithm we were able to sample from the posterior distribution and, thereby, yielded a very good estimation precision by using the posterior medians as parameter estimates. For insurance companies, the developed forecasting framework is able to substantially improve actuarial risk management if events occur in oscillatory bursts.

Appendices

A. The sample data set

In sections 1.2, 1.3 and 4.1 we analyzed the occurrence times of tropical cyclones in different continents. The corresponding data originates from the international disaster database *EM-DAT* (cf. Guha-Sapir et al. (2014)), which is maintained by the Centre for Research on the Epidemiology of Disasters (CRED) at the School of Public Health of the Université catholique de Louvain located in Brussels, Belgium. The data is compiled from various sources, including UN agencies, non-governmental organizations, insurance companies, research instituted and press agencies, and can be accessed via `www.emdat.be` in an aggregated form.

The database not only includes the dates of occurrences but also information about the country, number of deaths, missing, injured, homeless and affected persons and the estimated total damage (i.e., damage to property, crops and livestock). For a disaster to be entered into the database at least one of the following criteria must be fulfilled:

a) 10 or more people reported killed.

b) 100 or more people reported affected.

c) Declaration of a state of emergency.

d) Call for international assistance.

Since EM-DAT is a country-level database, the same disaster may be entered several times if it affects several countries. However, we still consider all these events. This is justified if taking the perspective of an insurance company that insured people in both countries. To obtain a simple point processes we applied the following procedure to the dates of occurrence for one disaster type (tropical cyclones) regarding one continent (Africa, Americas, Asia or Oceania):

1) Restrict the dataset to events occurring in 1980 or later since data for events dating back too far is unreliable.

2) Map all existing dates of occurrences to the number of days elapsed since 31.12.1979; for example 01.01.1980 $\mapsto$ 1, 10.01.1980 $\mapsto$ 10, 01.02.1980 $\mapsto$ 32 and so forth. We call d a mapped date.

3) Interpolate *missing data*: Assume that $d_1, ..., d_M$ are all (existing and missing) mapped dates, which are ordered by occurrence date. Let d_p be missing and d_l and d_u the next neighboring dates that exist, $l < p < u$. Then, we set $\hat{d}_p = d_l + \frac{d_u - d_l}{u-l}(p-l)$. If there do not exist such two neighbors, we set $\hat{d}_p$ equal to the next existing neighbor.

B. List of Abbreviations

ACF	autocorrelation function
ACH	autocorrelation histogram
AIC	Akaike's information criterion
a.s.	almost surely
cdf	cumulative density function
DIC	deviance information criterion
FCPD	full conditional posterior distribution (density)
IEI	Inter-Event-Interval
iid	independent and identically distributed
KDE	kernel density estimate
MCMC	Markov chain Monte Carlo
ME	moment estimate
MH	Metropolis Hastings
MLE	Maximum-Likelihood estimate
PD	probability of default
pdf	probability density function
pmf	probability mass function
RJMCMC	reversible jump Markov chain Monte Carlo
rv	random variable

C. List of Figures

1.1. IEI distributions of tropical cyclones (empirical density (grey/black) and fitted exponential density (red, dotted)) for (A) Africa, (B) Americas, (C) Asia and (D) Oceania. . . . 6
1.2. Rasterplots of tropical cyclones in (A) Africa, (B) Americas, (C) Asia and (D) Oceania. . . . 7
1.3. IEI distributions of tropical cyclones in (A) Africa, (B) Americas, (C) Asia and (D) Oceania. . . . 8
1.4. ACHs of tropical cyclones in (A) Africa, (B) Americas, (C) Asia and (D) Oceania. . . . 9

2.1. Construction of the GLO process: The background rhythm (blue) has normally distributed increments $B_{k+1} - B_k \sim \mathcal{N}(\mu, \sigma_1^2)$. At each beat B_k the number of events P_k (blue) is Poisson distributed, $P_k \sim Pois(\gamma)$, and every event $S_{k,i}$ (red) is placed around its birth beat B_k according to $\mathcal{N}(B_k, \sigma_2^2)$. . 12
2.2. (A) ACH and (B) IEI distribution of a simulated GLO process with parameters $\mu = 364$, $\sigma_1 = 20$, $\sigma_2 = 20$ and $\gamma = 5$. . . . 14
2.3. Example 2: Prior beliefs for (A) μ and (B) σ^2. . . . 33
2.4. Example 2: Gibbs sampler: Trace plots and ACF for (A)-(C) μ and (D)-(F) σ. 34
2.5. Example 2: MH algorithm: Trace plots and ACF for (A)-(C) μ and (D)-(F) σ. 35
2.6. Example 2: Posterior distributions: (A)-(B) Gibbs and (C)-(D) MH algorithm. 35

3.1. Test-Data 1: Rasterplot. . . . 60
3.2. Example 3: Test-Data 1: (A)-(D) Sample ACFs and (E)-(H) trace plots for $\mu, \psi_1, \psi_2, \gamma$. . . . 61
3.3. Test-Data 2: Rasterplot. . . . 62
3.4. Example 4: Test-Data 2: (A)-(D) Sample ACFs and (E)-(H) trace plots for $\mu, \psi_1, \psi_2, \gamma$. . . . 63
3.5. Example 4: Test-Data 2: Sample ACF for (A) B_1, (B) B_{11}, (C) B_{21}, (D) B_{31}. 63
3.6. Test-Data 3: Rasterplot. . . . 64
3.7. Example 5: Test-Data 3: (A)-(D) Sample ACFs and (E)-(H) trace plots for $\mu, \psi_1, \psi_2, \gamma$. . . . 65
3.8. Example 5: Test-Data 3: Sample ACF for (A) B_1, (B) B_{11}, (C) B_{21}, (D) B_{31}. 65
3.9. Example 5: Test-Data 3: Posterior distribution of σ_1. . . . 66
3.10. Example 6: Test-Data 3: (A)-(D) Sample ACFs and (E)-(H) trace plots for $\mu, \psi_1, \psi_2, \gamma$. . . . 67
3.11. Example 6: Test-Data 3: Sample ACF for (A) B_1, (B) B_{11}, (C) B_{21}, (D) B_{31}. 68
3.12. Example 6: Test-Data 3: Posterior distribution of σ_1. . . . 68

3.13. Example 7: Test-Data 3: Evolution of (A) μ, (B) σ_1, (C) σ_2 and (D) γ. . . 70
3.14. Example 7: Test-Data 3: Posterior pmf of (A) J_1, (B) J_{26},..., (H) J_{176}. . . . 71
3.15. Example 8: Test-Data 1: Observed events (solid), posterior beat densities (dashed) and labels (colors). 72
3.16. $\log_{10}$ ORE for $\hat{\eta}_{\text{Mean}}$ (green) and $\hat{\eta}_{\text{Median}}$ (black) in simulation study for $T = 8000$ ($\triangle$), $T = 13000$ ($\bigcirc$) and $T = 23000$ ($\times$). 75
3.17. $\log_{10}$ MRE for (A) $\hat{\mu}_{\text{Median}}$, (B) $\hat{\sigma}_{1,\text{Median}}$, (C) $\hat{\sigma}_{2,\text{Median}}$ and (D) $\hat{\gamma}_{\text{Median}}$ in simulation study for $T = 8000$ ($\triangle$), $T = 13000$ ($\bigcirc$) and $T = 23000$ ($\times$). . . 76
3.18. $\log_{10}$ MRE for $\hat{\eta}_{\text{Median}}$ (black) and $\hat{\eta}_{\text{ACH}}$ (red) in simulation study for $T = 8000$ ($\triangle$), $T = 13000$ ($\bigcirc$) and $T = 23000$ ($\times$) for (A) μ, (B) σ_1, (C) σ_2 and (D) γ. 77
3.19. $\log_{10}$ ORE for $\hat{\eta}_{\text{Median}}$ (black) and $\hat{\eta}_{\text{ACH}}$ (red) in simulation study for $T = 8000$ ($\triangle$), $T = 13000$ ($\bigcirc$) and $T = 23000$ ($\times$). 77
3.20. (A) Average computation time (in minutes) and (B) relative number of failures in simulation study. 79

4.1. Tropical cyclones in Africa: Different bandwidths σ with resulting N. . . . 81
4.2. Tropical cyclones in Africa: (A) Data, (B) posterior beat densities and (C) simulation. 82
4.3. Tropical cyclones in Americas: Different bandwidths σ with resulting N. . . 83
4.4. Tropical cyclones in Americas: (A) Data, (B) posterior beat densities and (C) simulation. 83
4.5. Tropical cyclones in Asia: Posterior beats for (A) $N = 34$ and (B) $N = 35$. 84
4.6. Tropical cyclones in Asia: (A) Data and (B) simulation. 85
4.7. Tropical cyclones in Oceania: Different bandwidths σ with resulting N. . . 85
4.8. Tropical cyclones in Oceania: (A) Data, (B) posterior beat densities and (C) simulation. 86
4.9. Occurrence of Hail in Germany: ACH of the POT process for $\alpha = 100000$. . 92
4.10. Occurrence of hail in Germany: (A) Data, (B) posterior beat densities and (C) simulation. 93
4.11. Claim sizes due to hail in Germany: Empirical (black/grey) and fitted shifted Gamma (red), Log-Normal (blue) and Exponential (green) distribution. . . 94
4.12. Rasterplot of simulated data. 95
4.13. *Upper part*: A posteriori probability of default for the GLO (solid, black) and homogeneous Poisson process (solid, red) beginning (A)/(D) after bursts, (B) between bursts and (C) before a burst. *Lower part*: Observed events (solid), forecasted beat densities (dashed) and labels (colors). 96
4.14. Tropical cyclones in Africa and Americas: Rasterplot. 100

D. List of Tables

1.1. Tropical cyclones: ML estimates, theoretical and sample variance of the Poisson model 6

2.1. Example 1: True parameters for data simulation, prior distributions and MCMC algorithms. 32

3.1. Conjugate and Jeffreys prior and posterior distributions for μ, ψ_1, ψ_2, γ and B_1. 57
3.2. Example 3: Input parameters, burn-in period and computation time. 61
3.3. Test-Data 1: Parameter estimates in Example 3. 61
3.4. Example 4: Input parameters, burn-in period and computation time. 62
3.5. Test-Data 2: Parameter estimates in Example 4. 64
3.6. Example 5: Input parameters, burn-in period and computation time. 64
3.7. Test-Data 3: Parameter estimates in Example 5. 65
3.8. Example 6: Input parameters, burn-in period and computation time. 67
3.9. Test-Data 3: Parameter estimates in Example 6. 68
3.10. Example 7: Input parameters and computation time. 69
3.11. Test-Data 1: Parameter estimates in Example 8. 72
3.12. Exemplary parameter combinations used for simulation study. 74
3.13. Exemplary record time, number of simulations and MCMC iterations and burn-in period used for simulation study. 74
3.14. Input parameters for MCMC algorithm in simulation study. 75
3.15. Average number of events, beats, computation time (in minutes) and average relative number of failures in simulation study. 78

4.1. Input parameters for estimation procedure for tropical cyclones. 80
4.2. Tropical cyclones in Africa: Bandwidth, number of beats, expected log-likelihood, burn-in period and posterior parameter medians. 81
4.3. Tropical cyclones in Americas: Bandwidth, number of beats, expected log-likelihood, burn-in period and posterior parameter medians. 82
4.4. Tropical cyclones in Asia: Bandwidth, number of beats, expected log-likelihood, burn-in period and posterior parameter medians. 84
4.5. Tropical cyclones in Oceania: Bandwidth, number of beats, expected log-likelihood, burn-in period and posterior parameter medians. 85
4.6. Occurrence of hail in Germany: Bandwidth, number of beats, expected log-likelihood, burn-in period and posterior parameter medians. 93
4.7. Exemplary values for the surplus process. 95

E. Bibliography

Albert, J. and Chib, S. (1996). Computation in bayesian econometrics: An introduction to markov chain monte carlo. In Hill, R. C., editor, *Advances in Econometrics Volume 11A: Computational Methods and Applications*, pages 3–24.

Bäuerle, N. and Grübel, R. (2005). Multivariate counting processes: copulas and beyond. *ASTIN Bulletin*, 35(2):379–408.

Bäuerle, N. and Grübel, R. (2008). Multivariate risk processes with interacting intensities. (available at `http://www.math.kit.edu/stoch/~baeuerle/media/baeugrueii.pdf`, last checked: August 24, 2015). *Advances in Applied Probability*, 40(2):578–601.

Baumgartner, C., Gruber, L. F., and Czado, C. (2015). Bayesian total loss estimation using shared random effects. *Insurance: Mathematics and Economics*, 62:194–201.

Beirlant, J., Matthys, G., and Dierckx, G. (2001). Heavy-tailed distributions and rating. *ASTIN Bulletin*, 31(1):37–58.

Bingmer, M. (2012). *A stochastic model for the joint evaluation of burstiness and regularity in oscillatory spike trains*. PhD thesis, Goethe-University Frankfurt am Main.

Bingmer, M., Schiemann, J., Roeper, J., and Schneider, G. (2011). Measuring burstiness and regularity in oscillatory spike trains. *Journal of Neuroscience Methods*, 201:426–437.

Boudreault, M., Cossette, H., and Marceau, É. (2014). Risk models with dependence between claim occurrences and severities for atlantic hurricanes. *Insurance: Mathematics and Economics*, 54:123–132.

Bowers, N. L., Gerber, H. U., Hickmand, J. C., Jones, D. A., and Nesbitt, C. J. (1997). *Actuarial Mathematics*. The Society of Actuaries.

Brockwell, P. and Davis, R. (2010). *Introduction to Time Series and Forecasting*. Springer, New York.

Chernobai, A., Burnecki, K., Rachev, S., Trück, S., and Weron, R. (2005). Modelling catastrophe claims with left-truncated severity distributions (extended version). Technical Report HSC/05/1, Hugo Steinhaus Center for Stochastic Methods, Wroclaw University of Technology.

Chu, P.-S. and Zhao, X. (2011). Bayesian analysis for extreme climatic events: A review. *Atmospheric Research*, 102:243–262.

Cowpertwait, P. S. P. and Metcalfe, A. V. (2009). *Introductory Time Series with R*. Springer.

Cox, D. and Isham, V. (1980). Point processes. In *CRC Monographs on Statistics & Applied Probability*. Chapman & Hall/CRC.

Cox, D. R. (1955). Some statistical methods connected with series of events. *Journal of the Royal Statistical Society. Series B (Methodological)*, 17(2):129–164.

Cox, R. T. (1946). Probability, frequency and reasonable expectation. *American Journal of Physics*, 14(1):1–13.

Cox, R. T. (1961). *The Algebra of Probable Inference*. The Johns Hopkins Press, Baltimore.

Daley, D. J. and Vere-Jones, D. (1988). *An Introduction to the theory of point processes*. Springer Berlin.

deLeeuw, J. (1992). Introduction to akaike (1973) information theory and an extension of the maximum likelihood principle. In *Breakthroughs in Statistics I*, pages 599–609. S. Kotz and N. L. Johnson.

Doob, J. L. (1949). Application of the theory of martingales. *Colloques Internationaux du C.N.R.S.. Paris*, 13:22–28.

Durrett, R. (2012). *Essentials of Stochastic Processes*. Springer, 2nd edition.

Eastoe, E. F. and Tawn, J. A. (2010). Statistical models for overdispersion in the frequency of peaks over threshold data for a flow series. *Water Resources Research*, 46.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer.

Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependency in risk management: properties and pitfalls. (available at `https://people.math.ethz.ch/~embrecht/ftp/pitfalls.pdf`, last checked: August 24, 2015). *Risk Management: Value at Risk and Beyond*, pages 176–223.

European Parliament (2015). Commission delegated regulation (eu) 2015/35 of 10 october 2014 supplementing directive 2009/138/ec of the european parliament and of the council on the taking-up and pursuit of the business of insurance and reinsurance (solvency ii). (available at `http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32015R0035&from=DE`, last checked: August 24, 2015).

Evans, S. N., Hubbard, A. E., and Shilane, D. (2010). Confidence intervals for negative binomial random variables of high dispersion. *The International Journal of Biostatistics*, 6(1).

Freedman, D. A. (1965). On the asymptotic behaviour of bayes estimates in the discrete case ii. *The Annals of Mathematical Statistics*, 36(2):454–456.

Ganji, M., Eghbali, N., and Azimian, M. (2013). Bayes and empirical bayes estimation of parameter k in negative binomial distribution. *Journal of Hyperstructures*, 2(2):185–200.

Gelman, A., Roberts, G., and Gilks, W. (1996). Efficient metropolis jumping rules. *Bayesian Statistics*, 5:599–607.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Gerstein, G. L. and Kiang, N. Y. (1960). An approach to the quantitative analysis of electrophysiological data from single neurons. *Biophysical Journal*, 1:15–28.

Gray, W. M. (1968). A global view of the origin of tropical disturbances and storms. *Monthly Weather Review*, 96(10):669–700.

Gray, W. M. (1975). Tropical cyclone genesis. Atmospheric science paper no. 234, Department of Atmospheric Science, Colorado State University.

Gray, W. M. (1979). Hurricanes: Their formation, structure, and likely role in the tropical circulation. In Shawn, D. B., editor, *Meteorology over Tropical Oceans*, pages 155–218. Royal Meteorological Society, Bracknell.

Green, P. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82:711–732.

Gregoire, G. (1983). Negative binomial distributions for point processes. *Stochastic Processes and their Applications*, 16:179–188.

Griffiths, W. (2002). A gibbs' sampler for the parameters of a truncated multivariate normal distribution a gibbs' sampler for the parameters of a truncated multivariate normal distribution. (available at `http://fbe.unimelb.edu.au/__data/assets/pdf_file/0006/805866/856.pdf`, last checked: August 24, 2015). Technical Report 856, University of Melbourne - Department of Economics.

Guha-Sapir, D., Below, R., and Hoyois, P. (2014). *EM-DAT: International Disaster Database. (available at `http://www.emdat.be`, last checked: August 24, 2015).*

Häggström, O. (2002). *Finite Markov Chains and Algorithmic Applications*. Cambridge University Press.

Hastie, D. and Green, P. J. (2012). Model choice using reversible jump markov chain monte carlo. (available at `http://www.maths.bris.ac.uk/~peter/papers/HastieGreenR1.pdf`, last checked: August 24, 2015). *Statistica Neerlandica*, 66:309–338.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.

Heilmann, W.-R. (1988). *Fundamentals of Risk Theory*. Verlag Versicherungswirtschaft e.V.

Henderson-Sellers, A., Zhang, H., Berz, G., Emanuel, K., Gray, W., Landsea, C., Holland, G., Lighthill, J., Shieh, S.-L., Webster, P., and McGuffie, K. (1998). Tropical cyclones and global climate change: A post-ipcc assessment. *Bulletin of the American Meteorological Society*, 79(1):19–38.

Hilbe, J. M. (2007). *Negative Binomial Regression*. Cambridge University Press.

Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*. Springer.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 186(1007):453–461.

Klenke, A. (2006). *Wahrscheinlichkeitstheorie*. Springer.

Koch, K.-R. (2007). *Introduction to Bayesian Statistics*. Springer, 2nd edition.

König, D. and Schmidt, V. (1992). *Zufällige Punktprozesse*. Teubner, Stuttgart.

Lewis, S. M. and Raftery, A. E. (1997). Estimating bayes factors via posterior simulation with the laplace-metropolis estimator. *Journal of the American Statistical Association*, 92(438):648–655.

Liu, J. and Dey, D. K. (2007). Hierarchical overdispersed poisson model with macrolevel autocorrelation. *Statistical Methodology*, 4:354–370.

Loza-Reyes, E., Hurn, M., and Robinson, A. (2014). Classification of molecular sequence data using bayesian phylogenetic mixture models. *Computational Statistics & Data Analysis*, 75:81–95.

McNeil, A. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin*, 27(1):117–137.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., and Teller, A. H. (1953). Equations of state calculations by fast computing machine. *Journal of Chemical Physics*, 21:1087–1092.

Moala, F. A., Ramos, P. L., and Achcar, J. A. (2013). Bayesian inference for two-parameter gamma distribution assuming different noninformative priors. *Revista Colombiana de Estadística*, 36(2):321–338.

Mohapatra, N. (2013). A model-based burst detection algorithm. Master's thesis, Goethe-University Frankfurt am Main.

Møller, J. and Torrisi, G. L. (2005). Generalised shot noise cox processes. *Advances in Applied Probability*, 37(1):48–74.

Moore, G. P., Perkel, D. H., and Segundo, J. P. (1966). Statistical analysis and functional interpretation of neuronal spike data. *Annual Review of Physiology*, 28:493–522.

Newton, M. A. and Raftery, A. E. (1994). Approximate bayesian inference by the weighted likelihood bootstrap. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(1):3–48.

Onof, C., Chandler, R., Kakou, A., Northtrop, P., Wheater, H., and Isham, V. (2000). Rainfall modelling using poisson-cluster processes: a review of developments. *Stochastic Environmental Research and Risk Assessment*, 14:384–411.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076.

Perkel, D. H., Gerstein, G. L., and Moore, G. P. (1967). Neuronal spike trains and stochastic point processes i: The single spike train. *Biophysical Journal*, 7(4):391–418.

Pfeifer, D. and Nešlehová, J. (2004). Modeling and generating dependent risk processes for irm and dfa. *ASTIN Bulletin*, 34(2):333–360.

Pradhan, B. and Kundu, D. (2011). Bayes estimation and prediction of the two-parameter gamma distribution. *Journal of Statistical Computation and Simulation*, 81(9):1187–1198.

Roberts, G. and Smith, A. (1994). Gibbs sampler and metropolis-hastings algorithms. *Stochastic Processes and their Applications*, 49:207–216.

Roberts, G. O. and Rosenthal, J. S. (2004). General state space markov chains and mcmc algorithms. *Probability Surveys*, 1:20–71.

Savage, L. J. (1972). *Foundations of Statistics*. Dover Publications Inc., New York.

Schneider, G. (2008). Messages of oscillatory correlograms - a spike train model. *Neural Computation*, 20(5):1211–1238.

Scott, S. L. (1999). Bayesian analysis of a two-state markov modulated poisson process. *Journal of Computation and Graphical Statistics*, 8(3):662–670.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

Son, Y. S. and Oh, M. (2006). Bayesian estimation of the two-parameter gamma distribution. *Communications in Statistics - Simulation and Computation*, 35(2):285–293.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Linkde, A. V. D. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B (Methodological)*, 64(4):583–639.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398):528–540.

Tiku, M. (2004). Noncentral chi-square distribution. *Ecyclopedia of Statistical Sciences*.

Wakefield, J. (2013). *Bayesian and Frequentist Regression Methods*. Springer.

Weinberg, M. D. (2012). Computing the bayes factor from a markov chain monte carlo simulation of the posterior distribution. (available at `http://projecteuclid.org/euclid.ba/1346158782`. last checked: August 24, 2015. *Bayesian Analysis*, 7(3):737–770.

Yang, R. and Berger, J. O. (1998). *A Catalog of Noninformative Priors. (available at `http://www.stats.org.uk/priors/noninformative/YangBerger1998.pdf`, last checked: June 20, 2015).*

Zhu, M. and Lu, A. Y. (2004). The counter-intuivitive non-informative prior for the bernoulli family. (available at `http://www.amstat.org/publications/jse/v12n2/zhu.pdf`, last checked: August 24, 2015). *Journal of Statistics Education*, 12(2).