Statistical Models for Spatio-Temporal Extrema and Dependencies

a thesis presented for the degree of

Doctor of Philosophy of Imperial College London

and the

Diploma of Imperial College

by

Ragnhild Cassel Noven

Department of Mathematics

Imperial College London

180 Queen's Gate, London SW7 2AZ

2016

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Ragnhild Cassel Noven

COPYRIGHT

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work

Thesis advisors:
Professor Axel Gandy,
Dr. Almut E.D. Veraart.

Author: Ragnhild Cassel Noven

Statistical Models for Spatio-Temporal Extrema and Dependencies

Abstract

This thesis develops statistical models for capturing the dependence characteristics and distributional structures of spatio-temporal data.

In the first part we propose a parsimonious stochastic process model for characterising rainfall time series. It is shown that the model is sufficiently flexible to capture important features of the rainfall process at different locations and time scales. We also study an application to the pricing of rainfall derivatives that introduces the market price of risk via the Esscher transform, and calculate futures prices based on empirical rainfall data.

In the second part we consider statistical models based on the class of trawl processes, which are stationary, infinitely divisible stochastic processes. We discuss properties of this class and introduce a novel representation of trawl processes in terms of independent random variables.

The trawl process class is used to construct a new model for characterising temporal dependence in exceedances above a threshold, which has the proper limiting distribution from extreme value theory. We illustrate the model by applying it to time series of rainfall and pollution measurements, and demonstrate its ability to capture the dependence structure of extreme value data.

The third part of the thesis introduces a framework for hierarchical models based on trawl processes, which includes the extreme value model as a special case. We develop a targeted MCMC algorithm for this class of models, which exploits the trawl process representation from the second part. We also introduce two specific models in this framework, and fit the second model to a time series of wind production.

In the final part we consider a wider class of processes for spatio-temporal modelling, which contains the specific models considered previously. We discuss the construction of models with specific covariance functions, and introduce a discrete approximation for these processes.

ACKNOWLEDGMENTS

First and foremost, it gives me great pleasure to thank my supervisors, Dr. Almut E. D. Veraart and Professor Axel Gandy, for their continual support, guidance, generosity and encouragement throughout the past years.

I am grateful for receiving financial support from the Grantham institute for Climate Change and the Environment and from Climate-KIC, which enabled me to carry out the research presented in this thesis.

Thanks are also due to the Department of Mathematics at Imperial, and the Statistics section in particular, for additional funding and helpful guidance from members of staff.

Special thanks are due to all my friends and family, for their love and support throughout my time at Imperial.

CONTENTS

List of figures **ix**

List of tables **xi**

1 Introduction **1**
- 1.1 Motivation . . . 1
- 1.2 Overview . . . 2
- 1.3 Publications . . . 4

2 Integrated CARMA rainfall model **5**
- 2.1 Introduction . . . 5
  - 2.1.1 Rainfall modelling . . . 5
  - 2.1.2 Chapter summary . . . 6
- 2.2 Characteristics of observed rainfall . . . 8
- 2.3 Rainfall Model . . . 10
  - 2.3.1 Primary model structure . . . 10
  - 2.3.2 Extension to CARMA process . . . 13
- 2.4 Fitting procedure . . . 16
  - 2.4.1 Seasonality function . . . 16
  - 2.4.2 Autocovariance structure . . . 17
  - 2.4.3 Driving Lévy process . . . 19
- 2.5 Assessing model performance . . . 22
  - 2.5.1 Marginal distribution . . . 22
  - 2.5.2 Autocorrelation and path properties . . . 24
  - 2.5.3 Zero proportion . . . 25
  - 2.5.4 Monthly fit . . . 26
- 2.6 Conclusion . . . 29
- 2.7 Appendix . . . 29

3 Rainfall derivative pricing **30**
- 3.1 Introduction . . . 30
  - 3.1.1 Pricing methodology . . . 31
  - 3.1.2 Esscher transform for integrated moving average processes 33
  - 3.1.3 Pricing illustration . . . 35
- 3.2 Conclusion . . . 37

3.3 Appendix 38

4 Trawl process background **43**
4.1 Ambit processes 44
4.2 Lévy basis 44
4.2.1 Lévy basis specification 47
4.2.2 Poisson point process representation 49
4.3 Trawl processes 51
4.3.1 Trawl process definition 51
4.3.2 Distributional properties of trawl processes 53
4.4 Partition representation 56
4.5 Temporal dependence structure of trawl processes 60
4.5.1 Autocovariance function 61

5 Latent trawl model for exceedances **64**
5.1 Introduction 64
5.2 Background on extreme value analysis 67
5.2.1 Convergence of maxima 67
5.2.2 Generalised Pareto distribution 70
5.3 Latent trawl process model 70
5.3.1 Basic structure 71
5.3.2 Latent trawl process 74
5.3.3 Autocovariance structure 77
5.3.4 Marginal transformation model 78
5.4 Model fitting and evaluation 80
5.4.1 Pairwise likelihood 80
5.4.2 Conditional tail dependence coefficient 82
5.4.3 Heathrow data 86
5.4.4 Pollution data 87
5.5 Conclusion 89
5.6 Appendix 90

6 Hierarchical trawl process framework **93**
6.1 General model construction 95
6.1.1 Defining the trawl process 95
6.1.2 Defining the observational distribution 96
6.2 Likelihood-based estimation approaches 98
6.2.1 Likelihood function 98
6.2.2 Pairwise likelihood 99
6.3 MCMC estimation methods 101
6.3.1 Monte Carlo within Metropolis 101
6.3.2 Data augmentation 102
6.4 Gibbs sampling algorithm 102
6.4.1 Truncated partition representation 103
6.4.2 Sampling slices 104

6.4.3 Updating scheme . . . 105
6.5 Strategies for improving mixing . . . 107
6.6 Transformation moves . . . 109
6.6.1 Theoretical setup . . . 109
6.6.2 Quantile moves . . . 113
6.7 Extended MCMC algorithm . . . 114
6.7.1 Diagnosing convergence . . . 115
6.8 Gaussian model . . . 117
6.8.1 Model specification . . . 118
6.8.2 Parametrisation issues . . . 119
6.8.3 Sampling algorithm . . . 119
6.8.4 Simulation study . . . 120
6.9 Mean-variance mixture model . . . 123
6.9.1 Data description . . . 123
6.9.2 Model specification . . . 123
6.9.3 Sampling scheme . . . 125
6.10 Model fit . . . 126
6.11 Conclusion . . . 128

7 Towards spatio-temporal models 129
7.1 Kernels versus trawl sets in temporal processes . . . 130
7.2 Spatio-temporal model setup . . . 131
7.2.1 Kernel-Covariance duality . . . 132
7.3 Lévy basis approximation . . . 133
7.4 Ambit process examples . . . 135
7.5 Conclusion . . . 136

8 Conclusion and future work 137
8.1 Conclusion . . . 137

Appendix A Background on Lévy-driven processes 140
A.1 Lévy processes . . . 140
A.1.1 Moving average processes . . . 142
A.2 Copyright statement . . . 145

9 Bibliography 147

List of symbols 157

LIST OF FIGURES

2.1 Empirical rainfall time series 8
2.2 Rainfall frequency plots 9
2.3 Empirical rainfall ACF 10
2.4 Frequency plots for hourly model simulations 23
2.5 Frequency plots for daily model simulations 23
2.6 Simulated rainfall QQ-plots 24
2.7 Fitted and empirical rainfall ACF 25
2.8 Hourly path plots 25
2.9 Daily path plots 26
2.10 Empirical and fitted monthly means 27
2.11 QQ-plots for monthly rainfall distribution 28

4.1 Gaussian Lévy basis and trawl process 54
4.2 Gamma Lévy basis and trawl process 54
4.3 Poisson Lévy basis and trawl process 54
4.4 Illustration of slice partition. 60

5.1 Illustration of latent trawl model structure. 75
5.2 Autocorrelation functions of latent trawl models 79
5.3 Original and transformed model structures 80
5.4 Heathrow rainfall exceedances. 87
5.5 Estimated extremal index 87
5.6 Ozone level exceedances 88
5.7 Estimated extremal index for transformed model 89

6.1 Truncated slice grid 104
6.2 Latent trawl model simulation traceplot 122
6.3 Latent trawl model simulation histogram 122
6.4 Wind energy production time series 123

6.5 QQ-plot and ACF for mean-variance mixture model 126
6.6 Path plots for mean-variance mixture model 127
6.7 Sampled mean curve and observations 127

7.1 Illustration of discrete Lévy basis approximation 135
7.2 Ambit processes with exponential kernel 135
7.3 Trawl process with circular kernel 136

LIST OF TABLES

2.1 Zero percentage 9
2.2 Parameter estimates for Heathrow rainfall 21
2.3 Parameter estimates for Detroit rainfall 21
2.4 Proportion of implied zero values 26

3.1 Prices of monthly rainfall contracts for Detroit. 37
3.2 Estimated values of risk premium parameter 37

5.1 Parameter estimates for exceedance models 86
5.2 Parameter estimates for transformed exceedance models 89

6.1 Latent trawl model parameter values, first simulation 121
6.2 Latent trawl model parameter values, second simulation 121

1

INTRODUCTION

1.1 Motivation

Data that arise from phenomena observed over space and time will typically exhibit intrinsic dependencies, and capturing these dependencies is crucial for achieving a detailed understanding of the process under investigation. For processes coming from environmental applications, these dependencies arise from the dynamics of the underlying physical systems. Similar dependence structures are also found in many other applications, as measurement units contained in local regions of time and space tend to be influenced by the same underlying factors.

When modelling spatio-temporal phenomena as stochastic processes, a first step consists in constructing a model that can replicate the covariance structure and distributional characteristics of the observations. A more detailed understanding comes from considering the observations as representing a path over time, or a field in space, which gives an additional level of characteristics. In particular, a process may have a stationary marginal distribution, but at the level of individual realisations the process will deviate from its marginal distribution in random, highly localised regions of time and space. For example, wind turbulence is characterised by a local increase in the variance of the air velocity and pressure, and so-called convective rainfall corresponds to an interspersion of high and low intensity in localised areas.

Naturally, an ideal model of a spatio-temporal process will be able to describe such localised irregularities in the process of interest. This may require us to go beyond the simplest statistical models, which tend to be based on Gaussian noise plus linear dependence on the nearest neighbours in time and space. Indeed, the rainfall and wind production data discussed in Chapters 2 and 6, respectively, both illustrate how the presence of such irregularities may

result in a somewhat pathological distributions, with features such as extreme skewness or heavier tails, for which a Gaussian model is unsuitable. Similarly, the model for extreme values in Chapter 5 demonstrates how the temporal dependence of extreme values can be explained more accurately by moving beyond the assumption of a Markov dependence structure.

1.2 Overview

This thesis introduces several statistical models that aim to capture the distributional and dependence characteristics in spatio-temporal data as described above. The models are all based on continuous-time stochastic processes, which can be written in the most general form as

$$Y(x,t) = \int g(t,s;\, x,\xi)\, dL(s,\xi), \tag{1.1}$$

where t represents time, x space, and L is either a standard Lévy process or its higher-dimensional generalisation, called a Lévy basis, and the kernel function g gives a measure of the proximity of points in space and time. This construction means that we are collecting up the random values of L over the field $\{(s,\xi)\}$, but assigning more weight to values close to the point (t,x). Thus the kernel function g is responsible for creating the dependence structure in Y, and the Lévy basis L controls its probability distribution.

The construction above results in a spatio-temporal process $Y(x,t)$ with a flexible dependence structure. Furthermore, by working with the class of Lévy bases as opposed to the more standard Gaussian random fields, we obtain a richer class of marginal distributions as well.

We begin in Chapter 2 with a well-known version of the above construction, the Lévy-driven Ornstein-Uhlenbeck process (Barndorff-Nielsen and Shephard, 2001), and use it to construct a new model for capturing the characteristic features of observed rainfall. The motivation for this model is that a typical rainfall time series has characteristic features that are difficult to capture in a simple model, including a distinctly non-Gaussian distribution, a large proportion of zero values, and frequent large fluctuations. We also consider an extension to the more general class of Lévy-driven CARMA process (Brockwell, 2001; Brockwell and Lindner, 2009), which leads to an effective

procedure for parameter estimation. We illustrate the model by fitting it to rainfall data on both hourly and daily time scales, and show that the model gives a good fit to the marginal distribution and autocorrelation function of the rainfall process.

In Chapter 3 we exploit the analytical tractability of the model construction in order to derive a formula for the prices of rainfall futures, which can be considered as a form of insurance against too much or too little rainfall. In contrast to much of the pricing literature, which focuses on reproducing the observed prices as a standalone process, we derive prices that are directly related to the empirical features of the observed rainfall process, via the model described in Chapter 2,

The remainder of the thesis is mostly concerned with the class of trawl processes; these are temporal processes that constitute a special case of (1.1), where the kernel function g is the indicator function of a deterministic set. This so-called trawl set gives an intuitive representation of the dependence structure of the process as well as specific temporal features such as causality. Chapter 4 presents the theoretical background for the class of trawl processes, focusing on the properties that are relevant for statistical models, and showing how these relate to the mathematical construction of the process. Furthermore, we introduce a new representation of trawl processes in terms of a collection of independent random variables.

In Chapter 5 we construct a model for time series of exceedances above a threshold, which uses the trawl process framework introduced in Chapter 4. The model is based on a representation of the generalised Pareto distribution (GPD) as a Gamma mixture of exponential random variables, which allows us to introduce dependence via a latent process in a hierarchical model, as was done in Bortot and Gaetan (2014). We extend this construction by using a latent trawl process instead of a Markov chain, and introduce a modification of the model that allows for a negative-valued shape parameter in the GPD. We apply the model to time series measuring pollution and rainfall, and show that using a latent trawl process gives a more flexible dependence structure for the extreme values of those data, compared to the model in Bortot and Gaetan (2014).

In Chapter 6 we develop a general framework for hierarchical trawl process models, which includes the extreme value model of Chapter 5 as a special

case. We describe the construction of models in this framework, and discuss approaches to parameter estimation in the general case. Based on this discussion we develop a targeted MCMC algorithm for this class of models, which exploits the representation of the trawl process from Chapter 4. This algorithm introduces a type of MCMC moves called transformation moves, which are a special case of reversible-jump MCMC moves (Green, 1995), where the dimension of the parameter space is fixed.

We demonstrate the hierarchical trawl process framework by introducing two specific models, the first is a simple Gaussian model, and the second a mean-variance mixture model, which is fit to a data set of daily wind energy production. Simulation studies for the Gaussian model show that the algorithm works well for parameter estimation. The mean-variance mixture model demonstrates the flexibility of the latent trawl framework, and is able to capture the dependence structure and fluctuations in the wind energy data.

In Chapter 7 we consider the wider class of ambit processes, discussing how the models used in previous chapters fit into this framework. We describe the construction of spatio-temporal models based on the ambit class of processes, and introduce a discrete approximation that makes the resulting models more tractable. Finally, we show examples of spatial processes based on this discretisation.

1.3 Publications

This thesis contains work that has been published or submitted for publication.

The material in Chapters 2 and 3 have been published as a paper entitled "A Lévy-driven rainfall model with applications to future pricing" (Noven et al., 2015). Permission has been given to use the article in this thesis, as described in the copyright statement in Appendix A.2.

The material in Chapters 4 and 5 is available as a preprint at `http://arxiv.org/abs/1511.08190`, and has been submitted for publication.

2

INTEGRATED CARMA RAINFALL MODEL

2.1 Introduction

In this chapter we propose a stochastic process model for characterising the distributional and temporal properties of rainfall. The motivation for this model comes from the properties of typical rainfall time series that are difficult to capture in a simple statistical model, including a heavily skewed marginal distribution that is distinctly non-Gaussian, a large proportion of zero values, and frequent large fluctuations. Thus there is a need for specialised models for rainfall that can capture the unique characteristics of this type of process.

The model we present is based on an integrated, Lévy-driven Ornstein–Uhlenbeck process. These processes were popularised in Barndorff-Nielsen and Shephard (2001) in the context of modelling stochastic volatility. The choice of this process is motivated by the characteristics of the observed rainfall process, which include abrupt jumps and a large proportion of zero values; these can be modelled by choosing an appropriate driving process with jump-behaviour and a zero-inflated distribution.

2.1.1 Rainfall modelling

The existing literature on modelling rainfall is large and spread over fields such as hydrology, atmospheric sciences, environmental risk analysis and statistics. Onof et al. (2000) classifies the different approaches into four categories: meteorological models based on large sets of differential equations, multi-scale models concerned with the spatial evolution of rainfall, statistical models that capture spatial and temporal trends, and finally stochastic process models that make simple assumptions in order to remain parsimonious. In the following we will focus on models of the last category.

Many of the early attempts at modelling rainfall use a simple model that represents the rainfall occurrence process as a two-state Markov chain, and models the intensity of rainfall with a Gamma distribution (Katz, 1977; Chin, 1977; Woolhiser and Roldán, 1982; Coe and Stern, 1982). This model is easy to interpret and enables the direct use of likelihood methods for fitting. However, it makes several restrictive assumptions on the rainfall process, and may require a high-order Markov chain with many parameters to capture observed temporal dependence.

There is also a large literature on modelling rainfall for hydrological applications based on a form of the Poisson-cluster model, first developed by Rodriguez-Iturbe et al. (1987) and Cox and Isham (1988). This model is based on a hierarchical structure, with a primary Poisson process controlling the arrival of storms and a secondary process generating cells from each storm, which then deposit rainfall. There have been numerous extensions of the Poisson-cluster model that focus on fitting specific properties of the observed rainfall process. For example, Cowpertwait (1994) considers an extended model that allows for different rainfall cell types, where each cell has a random duration and intensity depending on its type, allowing for the different types of precipitation that are observed in practice.

In terms of fitting specific properties of rainfall that are of interest for hydrological applications, extensions to the Poisson-cluster model generally perform very well. However, as remarked by Onof et al. (2000), there is always a trade-off between the inclusion of more features and the mathematical tractability of the resulting models. Poisson-cluster based models are usually fitted by the method of moments, which involves matching analytical expressions for properties such as the mean, variance, and proportion of dry intervals to their empirical equivalents. Chandler (1997), remarks that the method of moments approach "suffers from the disadvantage that the parameter estimates can vary greatly depending on the properties used in the fitting procedure", and proposes a spectral estimation method for estimating rainfall models based on point processes.

2.1.2 Chapter summary

The contribution of this chapter is the development of a parsimonious and analytically tractable model, which captures the distributional features and

autocorrelation structure of the observed rainfall time series. We relate the model to the framework of Lévy-driven, continuous-time ARMA (CARMA) processes, and use this connection to develop a suitable fitting method, which is illustrated using empirical rainfall data. A second contribution is presented in Chapter 3, where we derive a formula specifying the so-called risk-neutral distribution of a general class of Lévy-driven stochastic processes, which includes our model as a special case. We then use this result to calculate prices for rainfall futures based on our model.

The structure of our model may be interpreted as a non-clustered Poisson model with multiple cell types (Cowpertwait, 1994), where rainfall cells of a random type arrive according to a Poisson process. In comparison to classical Poisson-cluster models, the main benefit of the model presented in this chapter is its parsimonious formulation based on a stochastic integral, making it suitable for applications where mathematical tractability and fitting methodology are of primary importance.

In contrast to the Markov chain and independent increment (Benth and Šaltytė Benth, 2013) models that have been used for pricing rainfall derivatives, our proposed model has the advantage of not making assumptions about temporal independence or Markovianity of the rainfall process increments. Furthermore, the model and the fitting method used are based on the assumption that the available data represents accumulated rainfall, i.e. the instantaneous intensity cannot be directly observed, which is typically the case for applications. By allowing for temporal dependence and considering CARMA processes of arbitrary order, we obtain a model with a flexible autocorrelation structure, which is particularly relevant for data on finer time scales. This flexibility is illustrated by fitting the model to hourly rainfall data. We also compare our model to that given in Wilks (1998), which was used in López Cabrera et al. (2013) for pricing rainfall futures based on daily data from Detroit.

This chapter is structured as follows: Section 2.2 discusses characteristic features of the rainfall process in light of data from different locations and time scales. Section 2.3 presents the rainfall model and shows how it fits into the continuous-time ARMA (CARMA) model framework. Section 2.4 gives details on the fitting method, Section 2.5 investigates the model performance using empirical data and Section 2.6 concludes. In Chapter 3, this model is

used to derive a method for pricing rainfall derivatives.

2.2 Characteristics of observed rainfall

In this section we motivate the structure of our model by illustrating some of the characterising features of rainfall time series. We base this illustration on two data sets which will be used throughout this chapter: the first consists of hourly accumulated rainfall amounts at Heathrow (UK) over the years 1980-2012, provided by the UK Meteorological Office (2012). The second data set gives daily accumulated rainfall amounts in Detroit (US) over the years 1980-2010, provided by Bloomberg Professional Service.

Figure 2.1 shows the rainfall time series for both locations over the years 2008-2010. These plots illustrate that the rainfall process is subject to sudden transitions between periods with little or no rain and periods of higher intensity, causing the large spikes in the graph (note the different y-scales of the plots, which are chosen to illustrate these path properties).

There is also a large amount of zero values in both data sets, as shown in Table 2.1. Because the data is rounded to the nearest unit of measurement (0.1mm for the Heathrow data and 3/100 inch for the Detroit data), these zero values represent all data points with a value smaller than this unit. The proportion of zero values is dependent on the measurement time scale, with fewer zero values for the daily data. This is because periods with no rainfall must persist throughout the measurement time interval in order to induce a value of zero in the data.

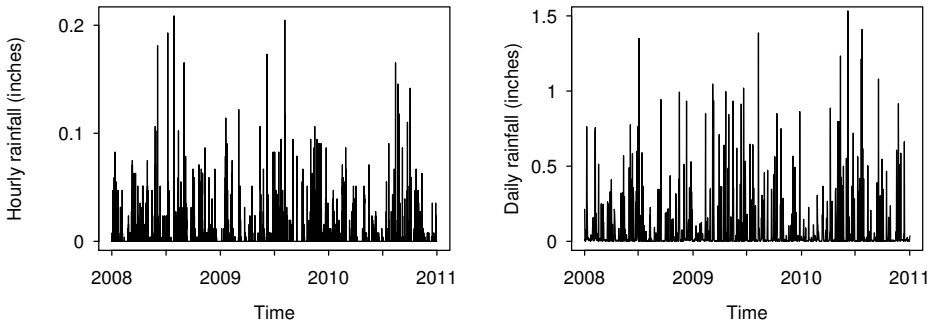

Figure 2.1: Time series of empirical rainfall over three years for Heathrow (left) and Detroit (right).

Location	Heathrow	Heathrow	Detroit
Time scale	Hourly	Daily	Daily
Zero values (%)	91.27	53.15	48.14

Table 2.1: Percentage of rainfall measurements equal to zero.

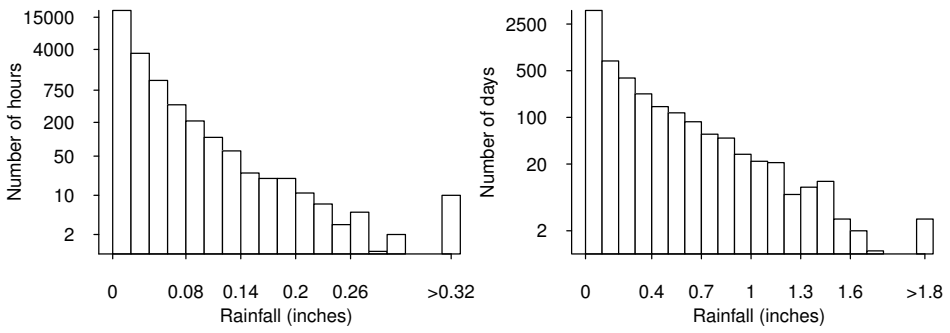

Figure 2.2: Frequency plots of non-zero rainfall for Heathrow (left) and Detroit (right), with frequencies on a log scale and the upper end of the range grouped together.

Figure 2.2 shows frequency plots of the non-zero (i.e. positive-valued) data, which illustrates the non-normality and skewness of the empirical distributions. In general, hourly data has more pronounced skewness, for Heathrow the coefficient is 15.87, compared with 4.18 when the data is aggregated to the daily time scale. This fits with the general observation that measuring accumulated rainfall on larger time scales has a smoothing effect, which makes characterising features such as large skewness and frequent zero values less evident.

The empirical autocorrelation functions of the Heathrow and Detroit rainfall time series are shown in Figure 2.3. For the hourly Heathrow data there is clearly a non-trivial autocorrelation part which decays smoothly up to lag 10. For the daily Detroit data the autocorrelation function (ACF) decays steeply after lag 1, indicating that there is less relevant time-dependence in this rainfall process, as one would expect from the daily time scale. As will be seen in Section 2.4, these differences in the autocorrelation structure lead us to fit models of different orders to the two data sets.

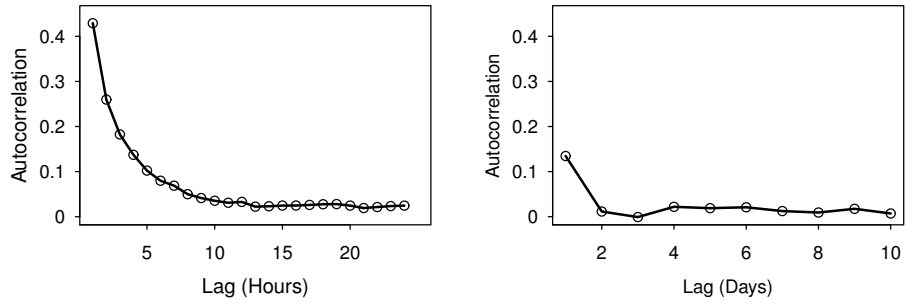

Figure 2.3: Empirical rainfall autocorrelation functions for Heathrow (left) and Detroit (right).

2.3 Rainfall Model

In this section we present the basic structure of our model and relate its properties to the observed rainfall dynamics. Based on this we construct a model extension where the rainfall intensity process belongs to the class of continuous-time ARMA (CARMA) models.

2.3.1 Primary model structure

We propose to model the accumulated rainfall R by

$$R(t_i) - R(t_{i-1}) = S(t_{i-1})(Y(t_i) - Y(t_{i-1})), \tag{2.1}$$

where $0 = t_0 < t_1 < \ldots < t_n$ are discrete measurement times such that $t_i - t_{i-1} = \delta$, and S is a deterministic seasonal component, which will be discussed in detail in Section 2.4.1. We define the continuous-time stochastic process $(Y(t))_{t\geq 0}$ as the integral of a Lévy-driven Ornstein-Uhlenbeck (OU) process $(X(t))_{t\geq 0}$ (Barndorff-Nielsen and Shephard, 2001), i.e.

$$\begin{aligned} Y(t) &= \int_0^t X(s)\, ds, \\ X(s) &= X(0)e^{-\lambda s} + \int_0^s e^{-\lambda(s-v)}\, dL(v), \end{aligned} \tag{2.2}$$

where λ is a positive parameter and $(L(t))_{t\geq 0}$ is a Lévy subordinator (see Appendix A.1 for details on these processes). We interpret $X(s)$ as the instantaneous rainfall intensity at time s, and so $Y(t)$ measures the accumulated rainfall over the time interval $[0, t]$ - up to the seasonal adjustment given by

S. We let $X(0)$ be a random variable which is independent of $(L(t))_{t\geq 0}$ and has distribution

$$X(0) \stackrel{d}{=} \int_0^\infty e^{-\lambda v}\, dL(v),$$

making the resulting OU process X strictly stationary.

Using the Fubini theorem for stochastic integrals (c.f. Protter (2005) p. 210 and Barndorff-Nielsen (1997)), we can exchange the order of integration in the definition of Y and obtain a simpler expression in terms of a single stochastic integral:

$$Y(t) = X(0)\left(\frac{1-e^{-\lambda t}}{\lambda}\right) + \int_0^t \frac{1-e^{-\lambda(t-v)}}{\lambda}\, dL(v). \tag{2.3}$$

For our particular rainfall model we let $(L(t))_{t\geq 0}$ be a compound Poisson process with Gamma-distributed jumps, called the Hougaard process (Lee and Whitmore, 1993; Grigelionis, 2011). This means that L is a pure-jump Lévy process, more specifically a subordinator. The choice of a pure-jump Lévy process is motivated by the intermittent behaviour of the observed rainfall process, in particular the abrupt switches from exact zero to large positive values, which are modelled by jumps in the driving process L.

The marginal distribution of $L(1)$ is a member of the Tweedie distribution family (Jørgensen, 1997), which was used by Dunn (2004) to model the monthly rainfall in Australia. In the following we will parametrise $L(1)$ as a Tweedie random variable, which has parameters (μ, ρ, κ) such that

$$\begin{aligned} \mathrm{E}(L(1)) &= \mu, \\ \mathrm{Var}(L(1)) &= \rho\mu^\kappa. \end{aligned} \tag{2.4}$$

The stochastic process X defined in (2.2), which represents the rainfall intensity, has an interpretation in terms of the physical dynamics of the rainfall process. In this interpretation the jumps of the driving process L represent the arrival of storm events, generating a jump in the intensity of random size. As the storm dissipates, this intensity decays smoothly towards zero at a rate determined by the parameter λ in the OU process.

By using this interpretation we see the parallel to the Poisson-cluster models discussed in the chapter introduction, which are based on the idea of storms

arriving according to a Poisson process. More specifically, the basic model presented in this section can be interpreted as a non-clustered Poisson model, i.e. the special case where each storm has only one cell. At each storm arrival, the current intensity increases by a random, Gamma-distributed amount, and then decays exponentially from the increased level.

A similar approach was considered in Samuel (1999), under the name of "Poisson Decaying pulse model", corresponding to a non-clustered Poisson model where each cell has an exponentially decreasing intensity, with the addition that each cell has a random lifetime after which its intensity is set to zero. An attempt was made to fit this model using the spectral method (Chandler, 1997), however it was found that there is no unique solution to this estimation problem.

Coming back to our model structure, it is clear from (2.1) that the discrete-time process $(\triangle Y(t_i))_{i=1,\ldots,n}$ given by

$$\triangle Y(t_i) := Y(t_i) - Y(t_{i-1})$$

should have features resembling those of deseasonalised empirical rainfall. As shown in Section 2.5, Figures 2.4 and 2.5, the empirical marginal distribution of our rainfall data is well approximated by the marginal distribution of $\triangle Y$. In fact, this was the motivation for specifying the driving process L to be the Hougaard process.

When it comes to approximating the empirical autocorrelation structure, the present model is rather limited, because the autocovariance function $C_{\triangle Y}(h)$ is restricted to take the form of an exponential decay. Specifically, we get the following form of the covariance:

$$\begin{aligned} C_{\triangle Y}(0) &= \frac{\rho\mu^\kappa}{\lambda^3}(e^{-\lambda\delta} + \lambda\delta - 1), \\ C_{\triangle Y}(h) &= \frac{\rho\mu^\kappa}{2\lambda^3}(e^{-\lambda\delta} - 1)^2 e^{-\lambda(h-1)\delta}, \quad h \geq 1. \end{aligned} \tag{2.5}$$

This expression can be derived by writing

$$\triangle Y(0)\triangle Y(h) = \int_{-\delta}^{0}\int_{(h-1)\delta}^{h\delta} X(u)X(v)\,du\,dv,$$

and using Lemmas A.1.15 and A.1.17.

As illustrated in Figure 2.3, the empirical autocovariance functions do not necessarily take such a simple form. This restrictive form of the autocovariance function motivates the following extension of the model.

2.3.2 Extension to CARMA process

In this subsection we consider an extension of our model which admits a more flexible autocovariance structure. This extension is based on generalising the Ornstein-Uhlenbeck process X in (2.2) to a continuous-time ARMA (CARMA) process. We first give a brief overview of the construction of Lévy-driven CARMA processes, and then show how the extension of X is obtained.

CARMA processes

A CARMA processes is a continuous-time analogue of the discrete-time ARMA process. Here we will consider Lévy-driven CARMA processes, following the expositions in Brockwell (2001); Brockwell and Lindner (2009).

To illustrate the correspondence to the discrete-time setting, we start by considering the ARMA(p, q) process (V_n) defined by the difference equation

$$a(B)V = b(B)L,$$

where B is the backward shift operator, L is a white noise sequence, meaning a sequence of i.i.d. random variables, and a, b are polynomials given by

$$\begin{aligned} a(x) &= x^p + a_1 x^{p-1} + \ldots + a_p, \\ b(x) &= b_0 + b_1 x + \ldots + b_q x^q. \end{aligned}$$

We can consider formally replacing B with the differential operator D to obtain a stochastic differential equation (SDE) for the CARMA(p, q) process V driven by the process L (see e.g. Applebaum (2009) for background on stochastic differential equations).

This SDE will contain expressions of the form $D^k L$, which may not be well-defined. Therefore it is customary to consider an equivalent definition of CARMA processes via the state-space representation. This representation

defines the observation and state equations

$$V(t) = \mathbf{b}^T \mathbf{Z}(t),$$
$$d\mathbf{Z}(t) - A\mathbf{Z}(t)dt = \mathbf{e}\, dL(t),$$

where $\mathbf{b}$ is the vector of coefficients of $b(x)$, A is the matrix

$$\begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -a_p & -a_{p-1} & -a_{p-2} & \dots & -a_1 \end{pmatrix},$$

where a_i is the i'th coefficient of $a(x)$, $\mathbf{e}$ is the p'th unit vector and L is the driving Lévy process, where $E\left[\max(0, \log|L(1)|)\right] < \infty$. Provided all eigenvalues of A have negative real parts, the SDE for Z can be solved to give the following expression for $(V(t))_{t\geq 0}$:

$$V(t) = \mathbf{b}^T e^{At}\mathbf{e}V(0) + \int_0^t \mathbf{b}^T e^{A(t-u)}\mathbf{e}\, dL(u), \tag{2.6}$$

where V is strictly stationary. We can now obtain a representation of the CARMA process V that extends the Lévy-driven OU process (Brockwell, 2004). Assume that A has distinct eigenvalues $\{\alpha_i\}$ (equivalently, that the polynomial $a(x)$ has distinct roots $\{\alpha_i\}$), with corresponding eigenvectors

$$(1, \alpha_i, \dots, \alpha_i^{p-1}).$$

Then we can obtain the spectral expansion

$$\mathbf{b}^T e^{A(t-u)}\mathbf{e} = \sum_{i=1}^{p} \frac{b(\alpha_i)}{a'(\alpha_i)} e^{\alpha_i(t-u)}.$$

Substituting this expansion into (2.6) gives

$$V(t) = V(0)\left(\sum_{i=1}^{p} \frac{b(\alpha_i)}{a'(\alpha_i)} e^{\alpha_i t}\right) + \int_0^t \left(\sum_{i=1}^{p} \frac{b(\alpha_i)}{a'(\alpha_i)} e^{\alpha_i(t-u)}\right) dL(u),\ t \geq 0. \tag{2.7}$$

If we now set $p = 1$ and let $a(z) = z + \lambda$, $b(z) = 1$, we recover the OU process

X given in (2.2), with $\lambda = -\alpha_1$. Thus CARMA processes generalise Ornstein-Uhlenbeck processes, and this motivates the extended model described in the following.

Extended model

We define the extended model of order p by

$$R(t_i) - R(t_{i-1}) = S(t_{i-1})(Y(t_i) - Y(t_{i-1})),$$

where

$$Y(t) = \int_0^t X(s)\, ds, \tag{2.8}$$

$$X(s) = \left(\sum_{i=1}^{p} w_i X_i(0) e^{-\lambda_i s}\right) + \int_0^s \left(\sum_{i=1}^{p} w_i e^{-\lambda_i (s-v)}\right) dL(v) \tag{2.9}$$

$$:= \sum_{i=1}^{p} w_i X_i(s), \tag{2.10}$$

with $\sum_i w_i = 1$. Here X is a CARMA process of the form given in (2.7), where the coefficients of the polynomial b can be found by solving $b(-\lambda_i)/a'(-\lambda_i)$ for w_i, with $b_q = 1$. $\{w_i, \lambda_i\}$

In the following we will assume that b has order $q = p - 1$, making X a CARMA$(p, p-1)$ process. The restriction $q = p-1$ is necessary for obtaining the implied ARMA process representation (Brockwell and Lindner, 2013) used in the fitting method described in Section 2.4.

For our applications we will only consider the cases $p = 1$ and $p = 2$, with w_i, λ_i being positive real numbers, and so in particular the resulting CARMA processes are non-negative when we take L to be a subordinator. When using the general CARMA process to represent rainfall intensity we need to restrict to the non-negative case, which requires L to be a subordinator and the kernel $g(u) = \boldsymbol{b}^T e^{Au} \boldsymbol{e}$ to be non-negative; conditions for this to hold are given in Tsai and Chan (2005) in terms of the polynomials a and b.

Similarly to the OU case, each $X_i(0)$ is chosen to be independent of $(L_t)_{t \geq 0}$,

with

$$\sum_{i=1}^{p} w_i X_i(0) e^{-\lambda_i s} \stackrel{d}{=} \int_0^\infty \left(\sum_{i=1}^{p} w_i e^{-\lambda_i (s+v)} \right) dL(v).$$

This extension of X can also be seen as a mixture of *dependent* OU processes X_i, driven by the same subordinator L.

Following the interpretation given in the previous subsection, storms arrive in a Poisson process, generating a jump in the intensity which is Gamma distributed. In the extended model, the intensity until the next arrival is given by a weighted sum of p intensity processes, which decay from the *same* initial level at different rates λ_i. This could be taken to mean that a typical storm has p components whose intensity dissipates at different rates.

As we will see in equation (2.12), the autocovariance function $C_{\triangle Y}$ with Y defined in (2.8) is a mixture of exponential decays with separate rates λ_i. Hence we can get more complex autocovariance structures by increasing the order p of the CARMA process X.

2.4 Fitting procedure

In this section we discuss a fitting approach for the extended model defined in Section 2.3.2. The fitting is done in three parts, firstly the deterministic seasonality function S is estimated, then we estimate the autocovariance parameters via the CARMA representation, and finally we find moment-based estimates of the driving Lévy process parameters.

2.4.1 Seasonality function

The multiplicative seasonality function S is estimated in an ad hoc fashion by fitting a truncated Fourier series with an annual period to the empirical mean of each month, specifically, we have

$$S(t) = \frac{a_0}{2} + \sum_{i=1}^{n} a_i \cos(2\pi i t/12) + b_i \sin(2\pi i t/12), \tag{2.11}$$

for time t on a monthly scale, where a_i, b_i are the fitted parameters. The Fourier series was fitted by considering a linear model where the responses

are the empirical monthly means and the covariates are the corresponding values of the sin and cos terms. The order of truncation n was then chosen to minimise the AIC of this linear model, which occurs at order $n = 2$, giving a total of 5 parameters for the seasonality function.

Using a multiplicative seasonality function has the advantage of enabling straightforward fitting of a single integrated CARMA process Y to all the available data. However, it does not account for seasonal variations in the autocorrelation structure or higher-order moments of $\triangle Y$. An alternative approach would be to first fit the model separately to data from each month, in order to detect any significant seasonal changes in the parameters. The monthly parameter values could then be allowed to vary between months according to a suitably chosen, smoothly varying function. Using this setup, the model can be fitted by simultaneously minimising the squared prediction errors and moment differences for the individual monthly models.

Here we will only consider the multiplicative seasonality function, and focus on fitting the integrated CARMA process to the deseasonalised data. After fitting the seasonality function S using a truncated Fourier series as described above, we rewrite (2.1) as

$$\frac{R(t_i) - R(t_{i-1})}{S(t_{i-1})} = Y(t_i) - Y(t_{i-1}),$$

which shows that we can fit $\triangle Y$ to the discrete observations $\triangle R/S$. Thus in the following we will only consider fitting the model given by $\triangle Y$.

2.4.2 Autocovariance structure

We now show how to use the CARMA representation of the process X to develop a fitting method for the parameters $\{\lambda_i, w_i\}$. This approach relies on Theorem 2 in Brockwell and Lindner (2013), which states that under certain conditions* on the polynomials a and b, we have that for a causal and invertible CARMA$(p, p-1)$ process V, the discrete process I_n^Δ given by

$$I_n^\Delta = \int_{(n-1)\Delta}^{n\Delta} V(s)ds,$$

*The conditions are as follows: a and b have no common zeroes, the roots of a have multiplicity 1, and $\mathrm{Im}(\lambda_i) \in (-\frac{\pi}{\delta}, \frac{\pi}{\delta})$.

is a weak ARMA(p, p) process. This implied ARMA process takes the form

$$\phi(B) I_n^{\Delta} = \theta(B) \epsilon_n,$$

where $\{\epsilon_i\}$ is a weak white noise sequence, i.e. the terms are uncorrelated but possibly dependent. Here the parameters of the process L driving the CARMA process V only affect the sequence $\{\epsilon_i\}$, not the polynomials ϕ and θ. Furthermore, the theorem also states that there is a one-to-one correspondence between the coefficients $\{w_i, \lambda_i\}$ of V and the coefficients (ϕ_i, θ_i) of the corresponding ARMA(p, p) process.

By using the CARMA$(p, p-1)$ representation of X as defined in (2.10), we can write

$$\triangle Y(t_i) = \int_{t_{i-1}}^{t_i} X(s)\, ds = I_{t_i/\delta}^{\delta},$$

where X has associated polynomial a with distinct roots $\alpha_i = -\lambda_i$, so it is a causal and invertible CARMA process fulfilling the conditions of the theorem. Hence the observed increments of Y can be seen as observations from the implied weak ARMA process.

We can also obtain the autocovariance of $\triangle Y$ from the integrated CARMA representation (Brockwell and Lindner, 2013, Corollary 2):

$$\begin{aligned} C_{\triangle_Y}(0) &= \sum_{\lambda_i} 2\beta(\lambda_i)\lambda_i^{-2}(e^{-\lambda_i \delta} - 1 + \lambda_i \delta), \\ C_{\triangle_Y}(h) &= \sum_{\lambda_i} \beta(\lambda_i)\lambda_i^{-2}(e^{-\lambda_i \delta} - 1)^2 e^{-\lambda_i (h-1)\delta}, \quad h \geq 1, \\ \beta(\lambda_i) &= \sigma^2 \frac{b(-\lambda_i) b(\lambda_i)}{a'(-\lambda_i) a(\lambda_i)}, \end{aligned} \tag{2.12}$$

where σ^2 is the variance of the driving process increment $L(1)$, and a, b are the polynomials in the CARMA representation of X. For the Hougaard process we have $\sigma^2 = \rho\mu^{\kappa}$.

We now follow Brockwell and Lindner (2013) in estimating $\{w_i, \lambda_i\}$ by minimising the weighted sum of the one-step prediction errors of the implied ARMA(p, p) process, which is equivalent to minimising with respect to $\{w_i, \lambda_i\}$ due to the one-to-one correspondence. Initial values for the parameters in the minimisation can be obtained by setting the values of the autocovariance function of $\triangle Y$ for the first few lags equal to the corresponding empirical values.

The estimation procedure based on minimising the prediction errors is shown to be strongly consistent by Brockwell and Lindner (2013).

We will use a CARMA$(1,0)$ model, meaning an OU process, for the intensity process X corresponding to the Detroit rainfall data, and a CARMA$(2,1)$ model for X corresponding to the Heathrow data. These orders are chosen to be as low as possible while ensuring that the ACF of the fitted model can adequately replicate the shape of the empirical ACF.

If the order p of the model is chosen too high, some of the weight parameters w_i may have estimates equal to 0, meaning that the process X_i has no influence on X. Thus the fitted model is equivalent to specifying X with a lower order $p' < p$. This was found to be the case for the Detroit data when using the model with $p = 2$, resulting in the estimates $\hat{w_1} \approx 1, \hat{w_2} \approx 0$. This motivates the use of a CARMA$(1,0)$ process for the Detroit data. In general we expect that higher order models are more suitable for high-frequency data, which has more significant dependence structure.

For the CARMA$(1,0)$ model representing the Detroit data we have the injective mapping $\lambda \to (\phi, \theta)$ given by

$$\begin{aligned} \phi(\lambda) &= e^{-\lambda}, \\ \theta(\lambda) &= -r - \sqrt{r^2 - 1}, \\ r &= \frac{1 - \lambda - e^{-2\lambda}(1+\lambda)}{1 - 2\lambda e^{-\lambda} - e^{-2\lambda}}. \end{aligned}$$

For the CARMA$(2,1)$ model, the mapping between $(w_1, w_2, \lambda_1, \lambda_2)$ and $(\phi_1, \phi_2, \theta_1, \theta_2)$ is found by numerically solving for the autocovariance function of the implied ARMA process, under the constraint $w_1 + w_2 = 1$.

2.4.3 Driving Lévy process

Having estimated $\{w_i, \lambda_i\}$ it remains to estimate the parameters (μ, ρ, κ) of the driving process L. The parameter estimation is done by the method of

moments applied to the process $\triangle Y$, which can be written as

$$\begin{aligned}\triangle Y(t_i) = &\int_0^{\infty} \sum_{k=1}^{p} w_k \left(\frac{e^{-\lambda_k t_{i-1}} - e^{-\lambda_k t_i}}{\lambda_k} \right) e^{-\lambda_k v} \, dL^*(v) \\ &+ \int_0^{t_{i-1}} \sum_{k=1}^{p} w_k \left(\frac{e^{-\lambda_k (t_{i-1}-v)} - e^{-\lambda_k (t_i - v)}}{\lambda_k} \right) dL(v) \\ &+ \int_{t_{i-1}}^{t_i} \sum_{k=1}^{p} w_k \left(\frac{1 - e^{-\lambda_k (t_i - v)}}{\lambda_k} \right) dL(v),\end{aligned} \tag{2.13}$$

where the processes L, L^* are independent and have the same characteristic triplet, given by $(0, 0, \nu(\cdot))$, corresponding to the Hougaard process.

Using the above representation we can obtain an analytic expression for the characteristic function of $\triangle Y$, as shown in Appendix 2.7. Using this expression, we can find the moments of $\triangle Y$ and thus fit the driving Lévy process. It is readily shown that

$$\mathrm{E}(\triangle Y) = \sum_{k=1}^{p} \frac{\mu \delta}{\lambda_k},$$

and furthermore the variance of $\triangle Y$ has been given in (2.12). The third moment is calculated numerically from the characteristic function.

We now replace the autocovariance parameters $\{w_i, \lambda_i\}$ in the expressions for the moments with their least-squares estimates. Comparing the theoretical moments to those of the observed increments $\triangle R/S$ gives three equations with unknowns (μ, ρ, κ), which can be solved to obtain estimates for these parameters.

To obtain confidence intervals for the estimated parameters we use the block bootstrap method (Politis and Romano, 1994; Künsch, 1989) to resample from the empirical distribution under the assumption of dependent data. In this resampling the block size has a geometric distribution with a specified mean value, which ensures stationarity of the resampled sequence in the case of a weakly dependent process. For the process $\triangle Y$, the precise dependence conditions would need to be investigated further, and we use the block bootstrap as an ad hoc method for obtaining confidence intervals.

In order to choose the mean block size we first generated a sample of model simulations with the parameters held fixed at their estimated values. Then we applied the block bootstrap with fixed mean block size to the simulated

Parameter	w_1	λ_1	λ_2
Estimated value	0.92	4.79	0.31
95% CI	$(0.88, 0.95)$	$(3.58, 7.86)$	$(0.28, 0.36)$

Parameter	μ	ρ	κ
Estimated value	2.15	143.01	1.85
95% CI	$(1.64, 3.48)$	$(130.63, 158.53)$	$(1.81, 1.92)$

Table 2.2: Estimated parameters and confidence intervals (CI) for Heathrow rainfall data.

Parameter	λ	μ	ρ	κ
Estimated value	4.54	4.55	14.85	1.62
95% CI	$(4.06, 5.25)$	$(4.07, 5.26)$	$(14.71, 14.99)$	$(1.60, 1.64)$

Table 2.3: Estimated parameters and confidence intervals (CI) for Detroit rainfall data.

data sets to obtain a bootstrap confidence interval for each simulation. We repeated this procedure for several choices of the mean block size, and chose the one that optimised the coverage rate of the confidence intervals relative to the 95% nominal rate.

After obtaining a bootstrap sample as described above, we calculated 95% confidence intervals by taking the lower and upper bounds to equal, respectively, the 2.5 and 97.5 percentiles of the bootstrap sample. Tables 2.2 and 2.3 show the estimated parameters and confidence intervals for the Heathrow and Detroit data.

The parameter estimates for the hourly Heathrow rainfall data show that the autocorrelation structure consists of one quickly decaying component with rate $\lambda_1 = 4.79$, and one slowly decaying component with rate $\lambda_2 = 0.31$. These components could be taken to represent different types of storms, perhaps corresponding to the "light" and "heavy" rainfall cell types considered in Cowpertwait (1994).

The confidence intervals indicate that for both data sets the estimates of λ_1 and μ are quite variable. However, the bootstrap estimates of w_1, λ_1 and λ_2 could be affected by seasonal variation in the autocorrelation structure, which is not accounted for by dividing out the seasonality component S, as discussed in Section 2.4.1.

2.5 Assessing model performance

To assess the fit of the model we first compare several properties of the simulations from the fitted model to the corresponding empirical properties, specifically considering the overall marginal distribution, autocorrelation function and zero proportion of the rainfall process. In the last subsection we compare properties that are important for the pricing application in Section 3 over individual months.

A simulation from the process Y can be obtained by using the compound Poisson process representation of the driving process L, which gives an expression for Y as a weighted sum of the jumps of L. Multiplying by the seasonality function S then gives a simulation from the full model for the accumulated rainfall increments $\triangle R$.

2.5.1 Marginal distribution

Figures 2.4 and 2.5 show frequency plots of the empirical rainfall time series, together with an estimate of the corresponding model-based frequencies. This estimate is obtained by averaging the frequencies over 100 simulations for the hourly Heathrow model, and 500 simulations for the daily Detroit model, which keeps the computational effort reasonable. Each simulated time series uses the parameter estimates given in Tables 2.2 and 2.3, and has the same length as the empirical time series. For the daily Detroit data we also include the frequencies averaged over 500 simulations from the model given in Wilks (1998), which was fitted to the empirical data. This model was used in López Cabrera et al. (2013) as the basis for a pricing method, and so in view of the application in Chapter 3 it is a natural choice for comparison.

We see that on both time scales the model manages to capture the characteristic shape of the rainfall distribution quite well. For the daily Detroit data the fit is somewhat better than that of the model given in Wilks (1998), especially in the lower part of the range where the majority of the data is found.

Figure 2.6 shows QQ-plots comparing empirical quantiles to simulation quantiles, where the latter are obtained by combining the data from the respective collections of 100 and 500 simulations for Heathrow and Detroit. These plots confirm the goodness-of-fit in the lower part of the range, however there

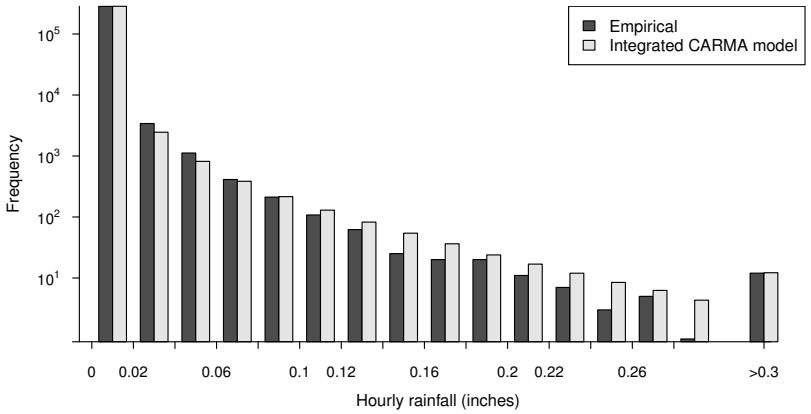

Figure 2.4: Frequency plots of empirical hourly Heathrow rainfall and simulations from the integrated CARMA model, with frequencies on a log scale and the upper end of the range grouped together.

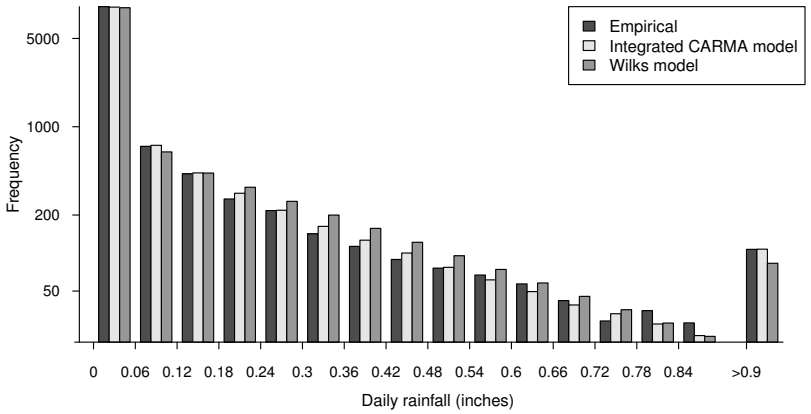

Figure 2.5: Frequency plots of empirical daily Detroit rainfall and simulations from the integrated CARMA model and the Wilks model, with frequencies on a log scale and the upper end of the range grouped together.

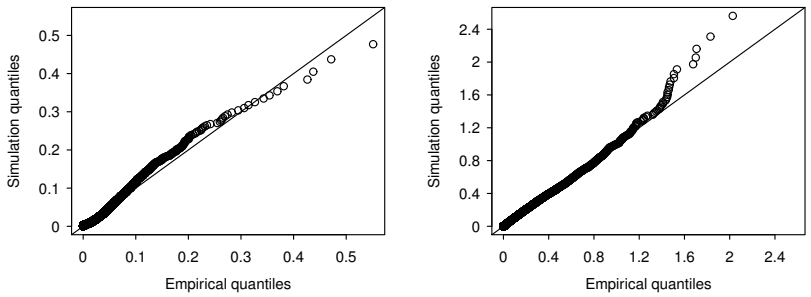

Figure 2.6: QQ-plots of empirical and simulated rainfall for Heathrow (left) and Detroit (right).

are some deviations in the extreme quantiles, especially for the daily Detroit model.

2.5.2 Autocorrelation and path properties

Figure 2.7 shows the autocorrelation function of the deseasonalised data along with the theoretical ACF of the fitted model for both time scales. For the hourly Heathrow data the gradual decay is captured very well by the fitted ACF from the CARMA$(2, 1)$ model, especially below lag 10. For higher lags it appears that the empirical ACF decays somewhat more slowly than the fitted ACF, however the difference is very small, and may be interpreted as noise, or as an effect of the deseasonalisation. If there is evidence of long-range dependence in the data, this could potentially be modelled by using a superposition of OU processes (Barndorff-Nielsen, 2001), although this was not the case for the Heathrow data.

The right panel in Figure 2.7 shows that the fitted ACF from the CARMA$(1, 0)$ model is very similar to the empirical ACF; in both cases the autocorrelation decays to zero almost immediately. Thus we see that our model manages to capture the autocorrelation structure of the rainfall process accurately for both hourly and daily time scales.

Figures 2.8 and 2.9 show path plots of the empirical data together with one particular simulation from the fitted model. For the daily Detroit data the paths look very similar; the model clearly reproduces the intermittent jump behaviour of the observed rainfall process. For the hourly Heathrow data the simulation also has intermittent jumps and periods of no activity, similar to the empirical process. However, there appears to be some clustering of time

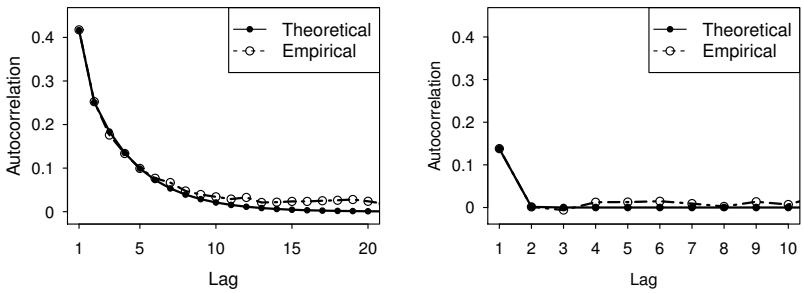

Figure 2.7: Theoretical and empirical ACF for Heathrow (left) and Detroit (right) rainfall.

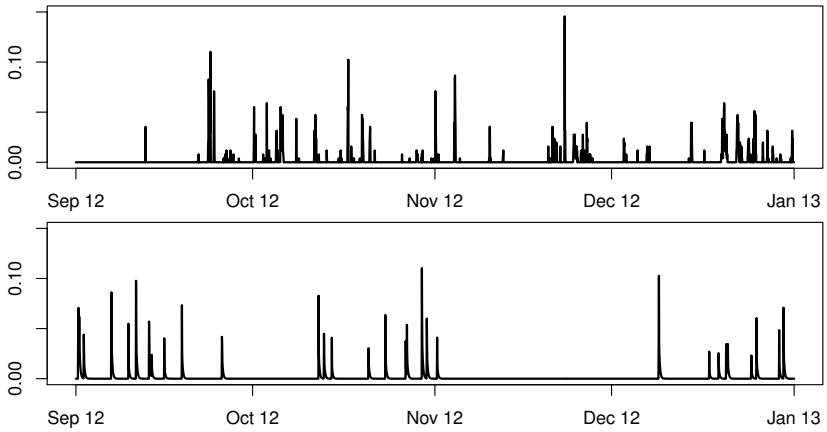

Figure 2.8: Path plots of hourly Heathrow empirical data (top) and simulated data (bottom).

intervals with higher intensity in the empirical process which is not reflected by the model.

2.5.3 Zero proportion

Table 2.4 shows the proportion of zero values in the empirical data, together with the corresponding zero proportion averaged over 20 simulations. As mentioned in Section 2.2, the zero values in the empirical data come from data points with a value below the measurement threshold, and we performed the same rounding for the simulated data to get the *implied* zero proportion shown. We see that the simulated time series for the hourly rainfall have very similar zero proportions to the empirical data, whereas the daily rainfall simulations are somewhat less accurate, they overestimate the zero proportion by about 24%.

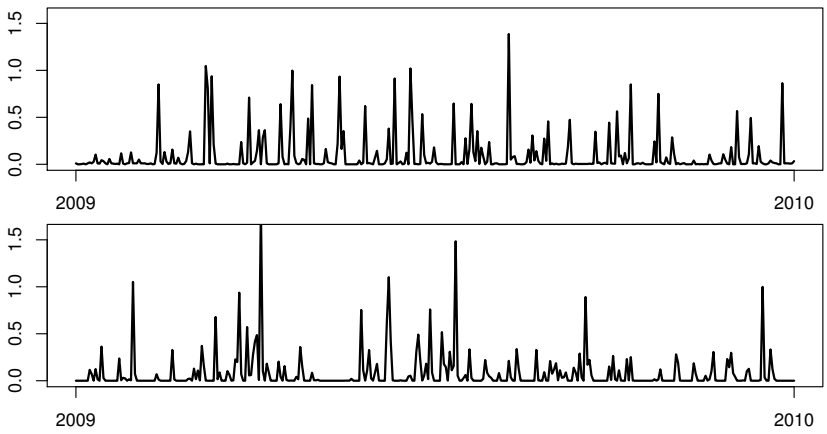

Figure 2.9: Path plots of daily Detroit empirical data (top) and simulated data (bottom).

Location	Time scale	Type	Implied zero proportion (%)
Heathrow	Hourly	Simulation average	90.39
		Data	91.27
Detroit	Daily	Simulation average	59.52
		Data	48.14

Table 2.4: Proportion of implied zero values in empirical data and corresponding simulation average.

For the purpose of pricing rainfall derivatives that depend on accumulated rainfall amounts, it is not particularly important to precisely match the zero proportions of the data. For other applications, a different approach for estimating the parameters of L may be more appropriate. One alternative would be to use a simulation-based generalised method of moments, with one of the moment conditions specifying that the proportion of implied zero values in empirical and simulated data match.

2.5.4 Monthly fit

In this subsection we consider the fit of the model on a monthly basis, which is especially relevant for the pricing application in Chapter 3. Although the current derivatives at the CME only consider total monthly accumulated rainfall, rainfall derivatives are sold over-the-counter, i.e. traded directly between two parties, so they can be tailored to specific needs. Thus one can consider derivatives that depend on daily accumulations or other relevant quantities. Hence it is important to have a flexible modelling framework that allows for adjustments to different time scales.

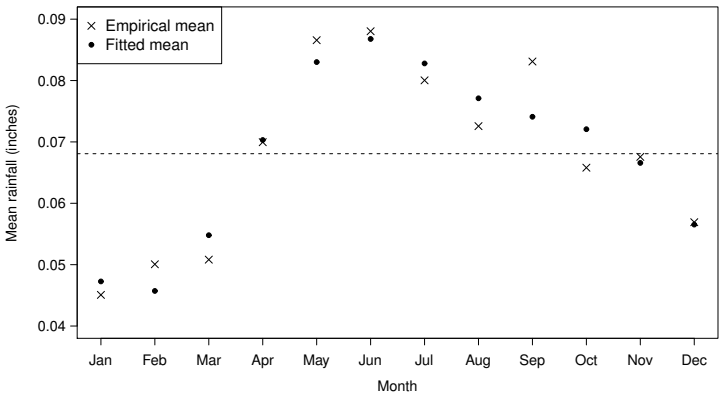

Figure 2.10: Empirical and fitted monthly means for daily Detroit rainfall, stippled line shows overall mean.

Figure 2.10 shows the fitted and empirical monthly means for the Detroit rainfall. Because the mean of $\triangle Y$ is constant, this plot gives a measure of the fit of the truncated Fourier series used for the seasonality function S. The plot also demonstrates the fit to the monthly totals that form the basis of the current CME derivatives. We see that the fitted seasonality matches the yearly trend quite well, except for the months of September and October, where the empirical means deviate from the smooth curve.

Figure 2.11 shows plots comparing the quantiles of the empirical and simulated Detroit daily rainfall for the months from March to October, which are the months considered for the CME rainfall derivatives currently on offer. As mentioned above, derivatives could be made to depend on rainfall accumulated over different time scales, thus it is reasonable to consider the fit to the finest available time scale, i.e. daily data. The simulation quantiles are based on combined data from the 500 simulations used for the frequency plots in Section 2.5.1. Similar to the overall QQ-plot in Figure 2.6, the overall fit is acceptable, especially in the lower end of the data range (the 99.9 percentile of the empirical daily rainfall is 1.46).

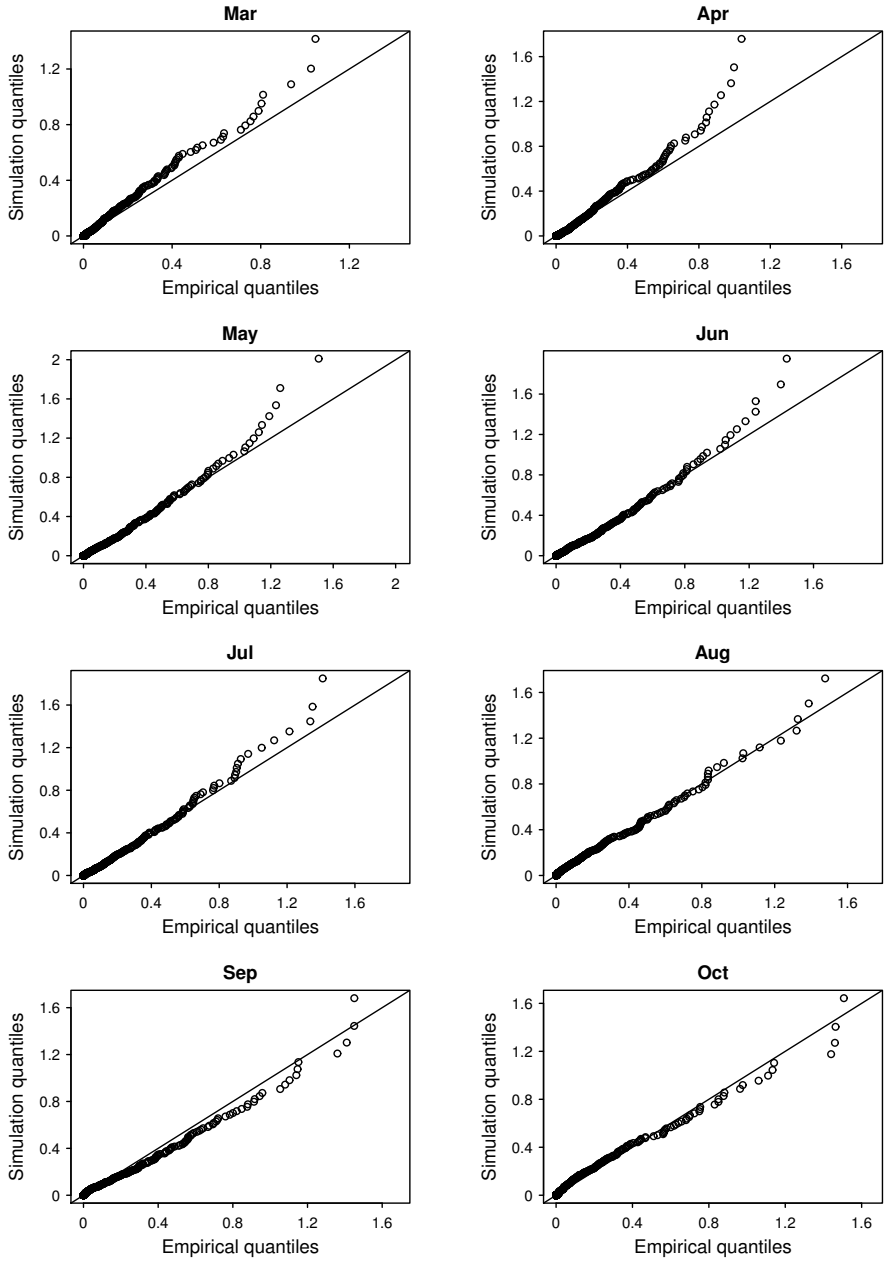

Figure 2.11: QQ-plots comparing simulated and empirical daily Detroit rainfall for months included in current CME rainfall derivatives.

2.6 Conclusion

In this chapter we have introduced a new class of continuous-time stochastic processes, driven by the Hougaard Lévy process, and shown how it can be used to construct a parsimonious and analytically tractable model for rainfall. By generalising the Ornstein-Uhlenbeck process representing rainfall intensity to a continuous-time ARMA (CARMA) process, we obtain a model with a flexible autocorrelation structure. We presented a general fitting method for this class that exploits a correspondence between integrated CARMA and ARMA processes.

We showed that the model fits the marginal distribution of the rainfall very well on both hourly and daily time scales. In particular, the marginal fit for daily rainfall is better than that of the standard model described in Wilks (1998), and we also obtain an adequate fit to the mean and marginal distribution on a month-by-month basis. By virtue of the CARMA generalisation, the extended model manages to accurately reproduce the autocorrelation structure of the observed rainfall, a characterising feature of the process which becomes increasingly significant for smaller time scales.

2.7 Appendix

The characteristic function of $\triangle Y$ is given by

$$\varphi_{\triangle Y} = \exp\left\{\int_0^\infty \psi(ug_1(s))\,ds + \int_0^1 \psi(ug_2(s))\,ds\right\},$$

where

$$g_1(s) = \sum_{k=1}^{p} w_k \frac{e^{-\lambda_k s} - e^{-\lambda_k(\delta+s)}}{\lambda_k},$$
$$g_2(s) = \sum_{k=1}^{p} w_k \frac{1 - e^{-\lambda_k(\delta-s)}}{\lambda_k},$$

and ψ is the cumulant function of the driving process L.

This follows immediately from applying Lemma A.1.15 to the expression given in (2.13), and noting that because $\triangle Y$ is stationary we can set $t_{i-1} = t_0 = 0$, causing the second integral to vanish.

3

RAINFALL DERIVATIVE PRICING

3.1 Introduction

In this chapter we calculate prices for rainfall futures contracts based on the daily rainfall model presented in Chapter 2. This is done by first deriving the characteristic function of a general Lévy-driven stochastic process under a risk-neutral measure.

Rainfall derivatives were introduced at the Chicago Merchantile Exchange (CME) in 2010, as a recent addition to the class of weather-related products. These products have a large potential market in all economic sectors that depend on favourable weather conditions, such as farming and energy development. The literature on rainfall derivatives pricing is currently rather limited. Because the underlying rainfall cannot be traded directly, the rainfall derivative market is incomplete, and thus there is no single fixed price for the derivative. Due to this incompleteness, there are several distinct methods that can be used for rainfall derivative pricing. It appears that most current approaches rely on either the utility indifference approach or risk-neutral pricing using the Esscher transform.

The utility indifference approach is used in Carmona and Diko (2005). In this paper a modification of the Poisson-cluster model is considered, which makes the rainfall intensity a Markov jump process, thus enabling maximum likelihood estimation. This modification relies on the assumption that the data used, though inevitably discrete, approximates continuous-time observations of the rainfall intensity. Leobacher and Ngare (2011) also use the utility indifference approach for pricing hypothetical rainfall derivatives in Kenya, based on a Markov-Gamma model with seasonality.

A natural choice for pricing based on Lévy process models is the Esscher transform (Esscher, 1932), as it is structure-preserving (Esche and Schweizer,

2005) and moreover gives rise to a minimal entropy martingale measure (Frittelli, 2000). López Cabrera et al. (2013) use the Esscher transform for pricing based on a version of the daily rainfall model by Wilks (1998). They fit simulated monthly rainfall totals to a normal inverse Gaussian distribution, and use the Esscher transform to obtain a risk-neutral distribution. Benth and Šaltytė Benth (2013) also use the Esscher transform for pricing, but base their underlying rainfall model on an independent increment process.

3.1.1 Pricing methodology

Classical asset pricing theory is based on the assumption of a complete market, where the risk associated with any derivative can be completely hedged against by replicating the derivative through a portfolio that includes holdings of the underlying asset. Then the derivative has a unique fair price equal to that of the replicating portfolio, and we say that the market is complete. This price can also be specified as the expected final payoff of the derivative under an equivalent measure Q, called the risk-neutral measure. Under this measure the discounted price processes of all tradeable assets are martingales.

For rainfall derivatives the underlying "asset" is an index $(I(t))_{t\geq 0}$ measuring accumulated rainfall, which cannot be directly traded, and so the hedging argument cannot be applied. Thus the market for rainfall derivatives is incomplete, meaning that there is no unique fair price of the derivative. Hence there exist many possible choices of equivalent probability measures. In this chapter we construct one such measure by using the Esscher transform on the underlying rainfall process, which we specify through our Lévy-driven rainfall model.

The Esscher transform is a generalised Girsanov transform for jump processes; it was first introduced by Esscher (1932) as a change of probability measure, and Gerber and Shiu (1994) generalised the transform to stochastic processes driven by a Lévy process. As shown in Esche and Schweizer (2005), the Esscher transform preserves the Lévy properties of the process to be transformed. This property makes it a natural choice for constructing a risk-neutral measure when the underlying is driven by a Lévy process, and contributes to achieving analytical tractability.

In the following we consider a finite time horizon $T < \infty$, and assume all

derivatives expire before that time. We also consider a Lévy subordinator $(L(t))_{t\geq 0}$ (assumed to be càdlàg), and extend L to a two-sided Lévy process $(L^*(t))_{t\in\mathbb{R}}$ by defining the process $\hat{L}$ to be an independent (càdlàg) copy of L such that L and $\hat{L}$ have the same characteristic triplet, and letting

$$L^*(t) = \begin{cases} L(t), & \text{for } t \geq 0 \\ -\hat{L}(-(t-)), & \text{for } t < 0, \end{cases} \tag{3.1}$$

which makes L^* càdlàg. In the following we will take L to mean the two-sided process L^* in order to simplify notation. We define the so-called *increment filtration* (Basse-O'Connor et al., 2014) by

$$\mathcal{F}_t = \sigma(L_u - L_s : -\infty < s < u \leq t),\ t \in \mathbb{R}, \tag{3.2}$$

so that $(L(t))_{t\in\mathbb{R}}$ is a Lévy process in this filtration.

We will use the generalised version of the Esscher transform for a Lévy process $(L(t))_{t\in\mathbb{R}}$ with filtration $(\mathcal{F}_t)_{t\in\mathbb{R}}$ as above, which is defined by giving the Radon-Nikodym derivative

$$\left.\frac{dQ}{dP}\right|_{\mathcal{F}_t} = Z(t) = \frac{\exp\left\{\int_0^t \theta(s)dL(s)\right\}}{E\left[\exp\{\int_0^t \theta(s)dL(s)\}\right]}, \tag{3.3}$$

where $\theta(s)$ is a time-dependent parameter, as opposed to the standard transform where it is constant. This parameter can be interpreted as a measure of risk-aversion, called the *market price of risk* (MPR), and is used to calibrate Q such that theoretical and observed market prices match. Specifically, the investor selling a derivative at time t will have to pay an amount given by the payoff function of the index I at the time of maturity τ. This amount is determined by the jumps of the driving process L in the future time interval $[t, \tau]$. Thus the investor is exposed to risk from these jumps, and the Esscher transform reflects the corresponding risk premium by exponentially tilting the jump measure.

Having defined Q via the Esscher transform, we find derivative prices by taking expected values of payoffs at maturity under Q, conditional on the information known at the current time, similar to the complete market case. For simplicity we assume a zero interest rate. Then for a rainfall index I

adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$, we get that the futures price process given by

$$f_\tau(t) = E_Q[I(\tau)|\mathcal{F}_t],$$

will be a Q-martingale by construction (provided it is integrable), which is required since the derivative contract is itself a tradeable asset. However, since the market is incomplete we do not require the underlying rainfall index process $(I(t))_{t\geq 0}$ to be a Q-martingale, since it cannot be directly traded.

3.1.2 Esscher transform for integrated moving average processes

In this subsection we show the result of applying the Esscher transform to the class of moving average processes, which includes our rainfall model as a special case.

Consider a two-sided Lévy subordinator $(L(t))_{t\in\mathbb{R}}$ with associated filtration $(\mathcal{F}_t)_{t\in\mathbb{R}}$, as defined in (3.1) and (3.2). We now define the stochastic process $X(t)$ by

$$X(s) = \int_{-\infty}^{s} h(s-v)\, dL(v) = \underbrace{\int_{-\infty}^{0} h(s-v)\, dL(v)}_{:=\tilde{A}(s)} + \int_{0}^{s} h(s-v)\, dL(v),$$

where $h : \mathbb{R}_+ \to \mathbb{R}_+$, and we assume that h is càglàd with $h \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. We also require that the Lévy measure ν of L is such that

$$\int_{|x|>1} |x|\, \nu(dx) < \infty.$$

The resulting process X is strictly stationary, and is called a *moving average process* (see Appendix A.1.1). A moving average process can be seen as a general form of the Ornstein-Uhlenbeck process; for the OU equation given by (2.2) we have $h(s) = \exp\{-\lambda s\}$, with $\tilde{A}(s) = e^{-\lambda s}X(0)$.

If we now integrate X over the interval $[0, t]$ and exchange the order of integration by using the stochastic Fubini theorem, we get the integrated moving

average process, which is similar to the primary rainfall model given in (2.3):

$$\begin{aligned} Y(t) &= \int_0^t X(s)\,ds = \underbrace{\int_{-\infty}^0 \tilde{g}(t,v)\,dL(v)}_{:=A(0,t)} + \int_0^t g(t,v)\,dL(v), \\ g(t,v) &= \int_v^t h(s-v)\,ds, \quad \tilde{g}(t,v) = \int_0^t h(s-v)\,ds, \end{aligned} \tag{3.4}$$

with $A(0,t) \in \mathcal{F}_0$, $\forall t$.

We now want to calculate the characteristic function of the process Y under the probability measure Q specified by the Esscher transform defined in (3.3). In order to ensure that the Radon-Nikodym derivative Z is well-defined, we assume that L satisfies the exponential moment condition, which states that there exists a constant $k > 0$ such that

$$E[\exp(kL(t))] < \infty, \tag{3.5}$$

for $t < T$, where T is our time horizon. For the particular case given by our rainfall model, L is the Hougaard Lévy process $L(\mu, \rho, \kappa)$, which has exponential moments for $k < \mu^{1-\kappa}/(\rho(\kappa - 1))$.

As discussed in Section 3.1.1, derivative prices are calculated in terms of the expected payoff at maturity under the measure Q, conditional on the current information $(\mathcal{F}_t)$. We want to find prices for a general payoff function $f(\text{Ind}(\tau_1, \tau_2))$, where $\text{Ind}(\tau_1, \tau_2) = Y(\tau_2) - Y(\tau_1)$ is the index measuring accumulated rainfall in the interval $[\tau_1, \tau_2]$. We follow Benth and Šaltytė Benth (2013) in using Fourier methods for these calculations, where we define the Fourier transform and its inverse by

$$\begin{aligned} \hat{f}(y) &= \int_{\mathbb{R}} f(x) e^{-ixy}\,dx, \\ f(x) &= \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(y) e^{ixy}\,dy, \end{aligned} \tag{3.6}$$

assuming $f, \hat{f} \in L^1(\mathbb{R})$. To account for the case when f is not integrable, we consider the function $f_\delta(x) = e^{-\delta x} f(x)$, and assume there exists a $\delta \geq 0$ such that f_δ is integrable. We then have the following result specifying the risk-neutral expectation of $f(Y(\tau_2) - Y(\tau_1))$.

Proposition 3.1.1: Let f be a payoff function such that $\hat{f}_\delta \in L^1$, where $\hat{f}$ is the Fourier transform of f as defined in (3.6). Assume that

$$\sup_{v\in[t,\tau_2]} (\delta|g(\tau_2,v) - g(\tau_1,v)| + |\theta(v)|) < k,$$

for k given by the exponential moment condition. Then, when Y is an integrated moving average process as specified in (3.4), and Q corresponds to the Esscher transform, we have that

$$\begin{aligned}
&E_Q\left(f\left(Y(\tau_2) - Y(\tau_1)\right)|\mathcal{F}_t\right) \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} \hat{f}_\delta(\xi)\exp\left\{(\delta+i\xi)\left(A(0,\tau_2) - A(0,\tau_1) + \int_0^t [g(\tau_2,v) - g(\tau_1,v)]\,dL(v)\right)\right. \\
&\left. + \int_t^{\tau_1} \psi_\theta(v,(\delta+i\xi)[g(\tau_2,v) - g(\tau_1,v)])\,dv + \int_{\tau_1}^{\tau_2} \psi_\theta(v,(\delta+i\xi)g(\tau_2,v))\,dv\right\} d\xi,
\end{aligned} \tag{3.7}$$

for $t < \tau_1 < \tau_2 < T$, where we define

$$\psi_\theta(s,\gamma c(s)) := \int_{\mathbb{R}_+} e^{\theta(s)y}(e^{(\gamma c(s)y} - 1)\nu(dy),$$

for a complex variable γ and real-valued function c, where $\nu(\cdot)$ is the Lévy measure of L.

The proof of Proposition 3.1.1 is given in Appendix 3.3.

We stress that the resulting derivative price *does* depend on the time points t, τ_1, τ_2, and as such our model constitutes an important generalisation of the independent increment model considered in Benth and Šaltytė Benth (2013). In their setup, the resulting futures prices did not exhibit any dynamics in t, which is a serious restriction in practical applications.

3.1.3 Pricing illustration

We now illustrate how to calculate the prices for futures written on the Detroit rainfall by using the model presented in Section 2.3, with the estimated parameters given in Table 2.3. For rainfall futures the payoff equals the index $Y(\tau_2) - Y(\tau_1)$, hence such contracts are also called swap contracts.

Recall that for the Detroit rainfall we used an integrated CARMA model of order $p = 1$, which reduces to the integrated OU model. This model can be written in the general form given in (3.4) above, with

$$g(t,s) = \frac{1 - e^{-\lambda(t-s)}}{\lambda}.$$

Letting $f(x) = e^{iux}$, we obtain the characteristic function $E_Q(e^{iu(Y(\tau_2)-Y(\tau_1))}|\mathcal{F}_t)$ as the exponential term in (3.7) evaluated at $\delta = 0$ and $\xi = u$. Taking derivatives with respect to u and evaluating at $u = 0$ gives

$$E_Q\left[Y(\tau_2) - Y(\tau_1) \,|\, \mathcal{F}_t\right] = A(0,\tau_2) - A(0,\tau_1) + \left(\frac{e^{-\lambda\tau_1} - e^{-\lambda\tau_2}}{\lambda}\right)\int_0^t e^{\lambda v}\, dL(v)$$
$$+ \int_t^{\tau_1} \psi'_\theta(v,0)\left(\frac{e^{-\lambda\tau_1} - e^{-\lambda\tau_2}}{\lambda}\right)e^{\lambda v}\, dv + \int_{\tau_1}^{\tau_2} \psi'_\theta(v,0)\left(\frac{1 - e^{-\lambda(\tau_2 - v)}}{\lambda}\right) dv, \tag{3.8}$$

where ψ'_θ denotes the derivative with respect to the second argument. The explicit form of ψ_θ for this model can be found analytically in terms of the parameters of the Hougaard process, and is specified in Appendix 3.3. When L is a Hougaard process the exponential moment condition (3.5) restricts the values of $\theta(v)$ to be below $\mu^{1-\kappa}/\rho(\kappa - 1)$, which equals 0.042 when using the estimated parameters. Note that this restriction does not affect the range of the prices, as the price explodes when $\theta(v)$ approaches the upper limit.

For our rainfall data we need to evaluate this expression based on the discrete observations $(Y(t_i))$, meaning that we do not observe L or $A(0,\tau_i)$ directly. To obtain an explicit value for the price, we approximate these terms by their expected value. Using this approximation, the sum of the first three terms in (3.8) reduce to

$$\frac{\mu}{\lambda^2}(e^{-\lambda(\tau_1 - t)} - e^{-\lambda(\tau_2 - t)}),$$

hence these terms decay to zero as $\tau_1 - t$ and $\tau_2 - t$ become large. This shows that when the time interval $[\tau_1, \tau_2]$ of interest is sufficiently far removed from the time t, then the mean approximation does not significantly affect the value of the price; this was also confirmed by the numerical calculations for the monthly rainfall contracts considered in the following.

We calculated prices for monthly rainfall contracts in 2011 for a time t corre-

	θ	Mar 11	Apr 11	May 11	Jun 11	Jul 11	Aug 11	Sep 11	Oct 11
CME price		4.2	4.4	3.2	5.0	4.5	4.3	4.2	4.6
Model price	-0.01	1.20	1.50	1.83	1.85	1.83	1.70	1.58	1.59
	0.00	1.69	2.10	2.57	2.60	2.57	2.39	2.22	2.24
	0.01	2.61	3.24	3.96	4.01	3.96	3.69	3.43	3.45
	0.02	4.72	5.86	7.15	7.25	7.16	6.67	6.19	6.23
	0.03	12.12	15.04	18.38	18.63	18.39	17.12	15.91	16.01
	0.04	153.69	190.80	233.11	236.27	233.21	217.19	201.83	203.13

Table 3.1: Prices of monthly rainfall contracts for Detroit.

Month	Mar 11	Apr 11	May 11	Jun 11	Jul 11	Aug 11	Sep 11	Oct 11
θ	0.0183	0.0156	0.0054	0.0142	0.0125	0.0130	0.01314	0.0142

Table 3.2: Estimated values of parameter θ based on CME prices.

sponding to the 31st of December 2010. The final price equals

$$S_{m_i}(t)E_Q[Y(\tau_2) - Y(\tau_1) \mid \mathcal{F}_t],$$

where $S_{m_i}(t)$ is the monthly average of the seasonality function $S(t)$. Table 3.1 shows a range of the resulting prices corresponding to different values of the parameter θ, representing the risk premium.

The first row of the table shows market prices reported by the CME, and we see that by adjusting θ we can calibrate the prices obtained from the model to match the market price. The values of θ resulting from this calibration are shown in Table 3.2. We remark that rainfall derivatives are a very recent addition to the CME portfolio, and their trading volume is currently close to zero, thus the reported CME prices for 2011 do not accurately reflect the market value of these products at the current time. Hence the corresponding estimates of θ for the specific 2011 prices may also differ from their true value. These values do however demonstrate how the rainfall model and associated pricing methodology provides a unified and flexible framework for studying the market view of the risk associated with rainfall.

3.2 Conclusion

In this chapter we derived a result specifying the risk-neutral expectation of a function of the rainfall process, which can be used for pricing general derivatives written on a precipitation index. To construct a risk-neutral measure we

used the Esscher transform, with a time-dependent parameter representing the risk premium. We stated the result for a general moving average process, a class which includes our model as a special case. The pricing methodology constitutes an important generalisation of the independent increment model considered in Benth and Šaltytė Benth (2013), which does not allow for price dynamics of derivatives.

We illustrated the pricing method by calculating futures prices based on empirical daily rainfall data from Detroit, and showed how they can be calibrated to observed prices. Given a sufficient amount of price data, one can study the properties of the market price of risk (MPR) that is implied by this calibration. Härdle and Cabrera (2012) discuss the market price of weather risk, focusing on temperature derivatives, and provides an example of modelling the MPR of temperature derivatives as a deterministic function of the seasonal temperature variation. As the market for rainfall derivatives matures, one could conduct a similar study for the MPR of rainfall derivatives.

3.3 Appendix

Proof of Proposition 3.1.1. By construction of f_δ, we have that

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}_\delta(\xi) e^{(\delta + i\xi)x} \, d\xi,$$

and hence by the Fubini theorem, it follows that

$$E_Q[\, f(Y(\tau_2) - Y(\tau_1)) |\, \mathcal{F}_t] = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}_\delta(\xi) E_Q[\, e^{(\delta + i\xi)(Y(\tau_2) - Y(\tau_1))} \big|\, \mathcal{F}_t] \, d\xi,$$

similar to the proof of Proposition 8.4 in Benth and Šaltytė Benth (2013).

We now calculate the expectation involving the integrated moving average process Y defined in (3.4). To this end, we first split the integrals in the

expression for Y as follows:

$$\begin{aligned}
&E_Q\left[\exp\left\{(\delta+i\xi)(Y(\tau_2)-Y(\tau_1))\right\}\middle|\,\mathcal{F}_t\right] \\
&=\exp\left\{(\delta+i\xi)\left(A(0,\tau_2)-A(0,\tau_1)+\int_0^t\left[g(\tau_2,v)-g(\tau_1,v)\right]\,dL(v)\right)\right\} \\
&\times\underbrace{E_Q\left[\exp\left\{(\delta+i\xi)\int_t^{\tau_1}\left[g(\tau_2,v)-g(\tau_1,v)\right]\,dL(v)\right\}\middle|\,\mathcal{F}_t\right]}_{(\mathbf{A})} \\
&\times\underbrace{E_Q\left[\exp\left\{(\delta+i\xi)\int_{\tau_1}^{\tau_2}g(\tau_2,v)\,dL(v)\right\}\middle|\,\mathcal{F}_t\right]}_{(\mathbf{B})},
\end{aligned} \tag{3.9}$$

using that $A(0,t)\in\mathcal{F}_0\,\forall t$.

By the abstract Bayes formula (see e.g Øksendal, 2000), for the measure Q such that $dQ/dP|_{\mathcal{F}_t}=Z(t)$, with X being $\mathcal{F}_\tau$-measurable and $t<\tau$, we have that

$$E_Q(X|\mathcal{F}_t)=E\left(X\frac{Z(\tau)}{Z(t)}\middle|\,\mathcal{F}_t\right).$$

Recall that we are working with the Esscher transform, so we have

$$Z(t)=\frac{\exp\left\{\int_0^t\theta(v)\,dL(v)\right\}}{E\left[\exp\left\{\int_0^t\theta(v)\,dL(v)\right\}\right]}.$$

Applying the Esscher transform then gives

$$\begin{aligned}
(\mathbf{A})=&E\left[\exp\left\{\int_t^{\tau_1}(\delta+i\xi)\left[g(\tau_2,v)-g(\tau_1,v)\right]\,dL(v)\right\}\frac{Z(\tau_1)}{Z(t)}\middle|\,\mathcal{F}_t\right] \\
=&E\left[\exp\left\{\int_t^{\tau_1}\left((\delta+i\xi)\left[g(\tau_2,v)-g(\tau_1,v)\right]+\theta(v)\right)\,dL(v)\right\}\right] \\
&\times\frac{E\left[\exp\left\{\int_0^t\theta(v)\,dL(v)\right\}\right]}{E\left[\exp\left\{\int_0^{\tau_1}\theta(v)\,dL(v)\right\}\right]},
\end{aligned}$$

where we get an unconditional expectation due to the independent increments of L.

Similar to the calculations in Benth and Šaltytė Benth (2013, Chapter 8), we

consider complex arguments of the cumulant function, and write

$$E\left[\exp\left\{\int_0^t (a(v)+ib(v))\,dL(v)\right\}\right] = \exp\left\{\int_0^t \psi_1(-ia(v)+b(v))\,dv\right\}, \tag{3.10}$$

where the term on the RHS is defined as

$$\exp\left\{\int_0^t \log E\left[\exp\left\{(a(v)+ib(v))L(1)\right\}\right]\,dv\right\},$$

and we have that

$$\left|E\left(e^{(a(v)+ib(v))L(1)}\right)\right| \leq E(e^{a(v)L(1)}),$$

and so if $\sup_v a(v) < k$, then the last term is bounded by the exponential moment condition given in (3.5), and so the cumulant function is well-defined.

Applying (3.10) to the terms in (**A**) gives

$$(\mathbf{A}) = \exp\left\{\int_t^{\tau_1} \psi_1\Big(\xi\left[g(\tau_2,v)-g(\tau_1,v)\right] - i\left[\theta(v)+\delta\left(g(\tau_2,v)-g(\tau_1,v)\right)\right]\Big) - \psi_1\left(-i\theta(v)\right)\,dv\right\},$$

where the requirement $\sup_{v\in[t,\tau_2]}\left(|\theta(v)|+\delta|g(\tau_2,v)-g(\tau_1,v)|\right) < k$ ensures that the terms in the above equation are well-defined.

Now we consider the Lévy-Khintchine formula for subordinators, which takes the form (see Appendix A.1)

$$E(e^{iuL(t)}) = \exp\left\{t\psi(u)\right\} = \exp\left\{t\int_{\mathbb{R}_+}(e^{iuy}-1)\nu(dy)\right\},$$

where $\nu(\cdot)$ is the Lévy measure associated with L. We can analytically continue this formula to complex arguments given the bounding assumption in the original proposition, and so we get that

$$(\mathbf{A}) = \exp\left\{\int_t^{\tau_1}\int_{\mathbb{R}_+} e^{\theta(v)y}(e^{(\delta+i\xi)[g(\tau_2,v)-g(\tau_1,v)]y}-1)\nu(dy)\,dv\right\}.$$

We note that this expression can also be written as

$$E\left[\exp\left\{\int_t^{\tau_1}(\delta+i\xi)\left[g(\tau_2,v)-g(\tau_1,v)\right]\,dL_Q(v)\right\}\right],$$

where $L_Q(v)$ is now a non-stationary stochastic process with jump measure depending on time, namely

$$\nu_\theta(dv,dy)=e^{\theta(v)y}\nu(dy)dv.$$

Thus we see that conditioning with respect to the measure Q has the effect of exponentially tilting the jump measure of L at time v according to $\theta(v)$, so the jumps of L at times v will be weighted more or less in the expectation depending on the sign of $\theta(v)$.

Now defining

$$\psi_\theta(v,\gamma c(v)):=\int_{\mathbb{R}_+}e^{\theta(v)y}(e^{\gamma c(v)y}-1)\nu(dy), \tag{3.11}$$

we get that

$$(\mathbf{A})=\exp\left\{\int_t^{\tau_1}\psi_\theta\Big(v,(\delta+i\xi)\left[g(\tau_2,v)-g(\tau_1,v)\right]\Big)\,dv\right\};$$

and by similar arguments

$$(\mathbf{B})=\exp\left\{\int_{\tau_1}^{\tau_2}\psi_\theta\Big(v,(\delta+i\xi)g(\tau_2,v)\Big)\,dv\right\}.$$

Substituting these expressions into (3.9) then gives the result.

Hougaard process. The Hougaard process has Lévy measure given by (Grigelionis, 2011)

$$\nu(dy)=\left(\rho^{\frac{1}{\kappa-1}}\Gamma\left(\frac{\kappa}{\kappa-1}\right)(\kappa-1)^{\kappa/(\kappa-1)}\right)^{-1}y^{\frac{3-2\kappa}{\kappa-1}}\exp\left\{-\frac{\mu^{1-\kappa}}{\rho(\kappa-1)}y\right\}dy,$$

in terms of the Tweedie parametrisation. We also have that when L is the

Hougaard process, the function ψ_θ defined in (3.11) takes the form

$$\psi_\theta(v, \gamma c(v)) = \frac{\mu^{2-\kappa}}{\rho(2-\kappa)\left(1 - \frac{\theta\rho(\kappa-1)}{\mu^{1-\kappa}}\right)^{\frac{2-\kappa}{\kappa-1}}} \left[\frac{1}{\left(1 - \frac{\gamma c(v)\rho(\kappa-1)}{\mu^{1-\kappa} - \theta(v)\rho(\kappa-1)}\right)^{\frac{2-\kappa}{\kappa-1}}} - 1 \right].$$

4

TRAWL PROCESS BACKGROUND

In this chapter we present the theoretical background underlying the class of trawl processes, which forms the basis of the models presented in the coming chapters. There are two main components of a trawl process: the Lévy basis, which is a random measure on a Borel space S, and a fixed subset of S called the trawl set. The Lévy basis extends the Lévy process from $\mathbb{R}_+$ to S, whereas the trawl set is used to introduce a given dependence structure.

Trawl processes are particularly useful for statistical modelling due to their intuitive and mathematically tractable structure, coupled with a flexibility in their distributional form and covariance structure. We will make use of this flexibility to construct a variety of models, with distributions and covariance functions tailored to particular applications.

We begin by describing the more general class of ambit processes and some related literature. The second section reviews the theoretical background for Lévy bases, following the expositions in Barndorff-Nielsen (2011) and Barndorff-Nielsen et al. (2015); Section 4.2.2 also describes a link between Lévy bases and a class of standard Poisson point process models. Section 4.3 introduces the trawl process and discusses its properties.

The remaining sections develop some new representations of trawl processes. Section 4.4 introduces a representation in terms of independent random variables that is particularly useful for statistical modelling purposes, it enables exact simulation and forms the basis of the MCMC algorithm used in Chapter 6. Section 4.5 considers trawl processes from a modelling perspective, showing that they can be characterised as stationary processes whose values are accumulations of random shocks, where the observed intensity of each shock changes as we move back in time relative to the current time point.

4.1 Ambit processes

The class of trawl processes belongs to the more general class of ambit processes, which were originally developed as general spatio-temporal models for studying turbulence (see Barndorff-Nielsen and Schmiegel (2007) and Barndorff-Nielsen et al. (2015) for an introduction to the ambit process modelling framework, and Barndorff-Nielsen (2011) for a discussion of trawl processes).

The ambit process framework has similarities to several stochastic modelling approaches in the literature. The monograph by Samorodnitsky and Taqqu (1994) considers stable, non-Gaussian stochastic processes for modelling heavy-tailed data, and shows how they can be represented as integrals with respect to a stable random measure. Wolpert and Taqqu (2005) give examples of trawl processes under the name "upstairs representation" and use them for studying workloads of communication networks.

Ambit processes are also related to solutions of stochastic partial differential equations (SPDEs) driven by a random noise process, as discussed in Barndorff-Nielsen et al. (2011). SPDEs form the basis of several models in the spatio-temporal modelling literature: Lindgren et al. (2011) points out that a Gaussian field with a Mátern covariance function is a solution to a particular SPDE driven by Gaussian noise, whereas Wallin and Bolin (2015) discusses non-Gaussian extensions. In contrast to SPDE methods, the ambit process framework provides a more direct and tractable way of representing the process of interest via the stochastic integral formulation.

4.2 Lévy basis

Consider a probability space $(\Omega, \mathcal{F}, P)$ and let S be a Borel set in $\mathbb{R}^n$, with the associated Borel σ-algebra $\mathcal{S} = \mathcal{B}(S)$ and Lebesgue measure λ^{Leb}. Let $\mathcal{B}_b(S)$ be the subsets of S with finite Lebesgue measure, i.e.

$$\mathcal{B}_b(S) = \{A \in \mathcal{S} : \lambda^{Leb}(A) < \infty\}.$$

A *Lévy basis* is a particular type of random measure, which we define as follows:

Definition 4.2.1: A *random measure on* $(S, \mathcal{B}(S))$ is a collection of $\mathbb{R}$-valued random variables

$$\{M(A) : A \in \mathcal{B}_b(S)\}$$

such that for any sequence $A_1, A_2, \ldots$ of disjoint elements of $\mathcal{B}_b(S)$ with bounded union, i.e. $\cup_{j=1}^{\infty} A_j \in \mathcal{B}_b(S)$, we have

$$M\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} M(A_j) \quad a.s.$$

Remark 4.2.2: The above definition only concerns sets of bounded Lebesgue measure, this is sufficient, however, as $\mathcal{B}(S)$ is generated by the sets in $\mathcal{B}_b(S)$. As will be seen, the area $\lambda^{Leb}(A)$ of a set A enters into the distribution of the random variable $M(A)$, and thus it is convenient to work with sets of bounded area.

Remark 4.2.3: This definition of a random measure differs from that in Kallenberg (1976), where random measures are defined as maps from the probability space Ω to the set $\mathcal{M}(S)$ of Radon measures on S. In contrast, the random measure defined above is analogous to a stochastic process, indexed by sets in $\mathcal{B}_b(S)$ as opposed to time points.

To illustrate this difference, consider the set function $\phi_\omega : S \to \mathbb{R}$ defined on $\mathcal{B}_b(S)$ via $\phi_\omega(A) := M(A)(\omega)$. Because this function is allowed to take negative values, it may fail to be an ordinary measure, and it may even fail to be a signed measure. The reason for this is that the additivity condition in the random measure definition only requires almost sure equality, where the associated null-set may depend on the particular sequence $\{A_i\}$.

In particular, consider the example where W is a standard Brownian motion, then we can define a random measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ in the sense of the above definition via

$$M((s,t]) = W_t - W_s, \ s < t.$$

The paths of the Brownian motion are given by the function $f_\omega(t) = W_t(\omega)$, and it is well known that this function is of infinite variation. Hence the corresponding set function ϕ_ω of M defined on $\mathcal{B}(\mathbb{R})$ also has infinite variation, so in particular it is not a signed measure.

To emphasise that the above definition differs from that of Kallenberg (1976),

some authors use the term *random noise* for the random measure M defined above. However, the distinction between random measures and random noise will not be important for the models described in the following, as it is usually the case that a finite collection of sets can be identified a priori from the measurements, and thus we will only be concerned with finite-dimensional distributions.

The following definition gives an analogue of the independent increment property of Lévy processes applied to random measures.

Definition 4.2.4: We say that a random measure M on $(S, \mathcal{S})$ is *independently scattered* if for any sequence $A_1, A_2, \ldots$ of disjoint elements of $\mathcal{B}_b(S)$, the random variables $M(A_1), M(A_2), \ldots$ are independent.

Remark 4.2.5: The Lévy basis itself may not be a good model for most physical processes, because the independent scattering property amounts to the assumption that events in disjoint regions of time and space are independent. However, as will be shown in the trawl process section, the Lévy basis can be combined with other elements to create models that incorporate dependence, while preserving the analytical tractability that is due to the independent scattering property.

Finally, we define infinite divisibility for random measures.

Definition 4.2.6: A random measure M on $(S, \mathcal{S})$ is called *infinitely divisible* if for each $n \in \mathbb{N}$ there exist n independent, identically distributed random measures $Z_1^n, \ldots, Z_n^n$ such that

$$M \stackrel{d}{=} Z_1^n + \ldots + Z_n^n.$$

In particular, infinite divisibility implies that for any finite collection $A_1, \ldots, A_n$ of elements of $\mathcal{B}_b(S)$, the random vector $(M(A_1), \ldots, M(A_n))$ is infinitely divisible in $\mathbb{R}^n$.

With the above definitions to hand, we are now ready to define a Lévy basis.

Definition 4.2.7: A *Lévy basis* L on $(S, \mathcal{S})$ is a random measure that is

independently scattered and infinitely divisible.

4.2.1 Lévy basis specification

Having defined a Lévy basis, a natural question is how we can construct such an object. This subsection introduces a characterisation that enables us to specify homogeneous Lévy bases explicitly via a generating random variable.

We start with the observation that the random variable $L(A)$ is infinitely divisible for any set $A \in \mathcal{B}_b(S)$, and hence, by the Lévy-Khintchine theorem (A.1.5), it has a characteristic function of form

$$E[\exp\{iuL(A)\}] = \exp\left\{iu\widehat{a}(A) - \frac{1}{2}\widehat{b}(A)u^2 + \int_{\mathbb{R}} (e^{iux} - 1 - iux\mathbb{I}_{[-1,1]}(x))\, n(dx, A)\right\}, \tag{4.1}$$

for $u \in \mathbb{R}$, where $\widehat{a}(A) \in \mathbb{R}$, $\widehat{b}(A) > 0$ and $n(\cdot, A)$ is a Lévy measure for fixed A, i.e.

$$n(\{0\}, A) = 0, \quad \int (|x|^2 \wedge 1)\, n(dx, A) < \infty.$$

We say that $(\widehat{a}, \widehat{b}, n(\cdot, A))$ is the *characteristic triplet* of $L(A)$.

The precise relationship between L and the characteristic triplet is described by Proposition 2.1 in Rajput and Rosinski (1989), which is restated below in our notation:

Proposition 4.2.8:

1. Let L be a Lévy basis with characteristic function given by (4.1). Then $\widehat{a}$ and $\widehat{b}$ are signed and ordinary measures on $\mathcal{B}_b(S)$, respectively, $n(\cdot, A)$ is a Lévy measure on $\mathbb{R}$ for fixed A, and $n(B, \cdot)$ is a measure on $\mathcal{B}_b(S)$ for fixed $B \in \mathcal{B}(\mathbb{R})$ such that $0 \notin \bar{B}$.
2. Given measures $\widehat{a}, \widehat{b}$ and n as described in part (1), there exists a unique Lévy basis such that (4.1) holds.

Part (2) of the above proposition implies that we can define Lévy basis by specifying measures $\widehat{a}, \widehat{b}$ and n satisfying the conditions in part (1). This Lévy basis is defined on a probability space $(\Omega, \mathcal{F}, P)$, which accommodates the

collection of random variables $L(A)$ for all $A \in \mathcal{B}_b(S)$, with the law of $L(A)$ determined by (4.1).

In the following we will work with a so-called *homogeneous* Lévy basis L, meaning that the random measure is stationary, in addition to the standard Lévy basis properties:

Definition 4.2.9: A random measure on $(S, \mathcal{S})$ is called *stationary* if for any point $s \in S$ and finite collection $A_1, A_2, \ldots, A_n \in \mathcal{B}_b(S)$ such that $A_i + \boldsymbol{s} \subset S$, we have that

$$(M(A_1 + \boldsymbol{s}), M(A_2 + \boldsymbol{s}), \ldots, M(A_n + \boldsymbol{s})) \stackrel{d}{=} (M(A_1), M(A_2), \ldots, M(A_n)).$$

When defining the Lévy basis via the measures $\widehat{a}, \widehat{b}$ and n, it is convenient to specify them via their densities with respect to the Lebesgue measure on the space $(S, \mathcal{S})$. These densities are denoted a, b and $\nu(\cdot, z)$, where

$$\widehat{a}(A) = \int_A a(z)\, dz, \quad \widehat{b}(A) = \int_A b(z)\, dz, \quad n(\cdot, A) = \int_A \nu(\cdot, z)\, dz. \tag{4.2}$$

It is clear that in order to obtain a homogeneous Lévy measure L, the measures $\widehat{a}, \widehat{b}$ and n must be stationary on $(S, \mathcal{S})$, implying that a, b are constant and $\nu(\cdot, z) = \nu(\cdot)$ is independent of z.

Using the above definitions gives a concrete procedure for specifying a homogeneous Lévy basis. Specifically, consider an infinitely divisible random variable L' with characteristic triplet (γ, σ^2, Π), and define the densities $a = \gamma$, $b = \sigma^2$ and $\nu = \Pi$. It follows that the resulting measures $\widehat{a}, \widehat{b}$ and n satisfy the Lévy basis conditions in Proposition 4.2.8. In particular, ν is a Lévy measure, so the relation

$$n(\cdot, A) = \int_A \nu(\cdot)\, dz = \nu(\cdot)\lambda^{Leb}(A),$$

implies that $n(\cdot, A)$ is a Lévy measure for fixed A in $\mathcal{B}_b(S)$. The random variable L' is called the *Lévy seed* of L.

Using the Lévy-Khintchine representation (4.1), relation (4.2) and the characteristic triplet $(a, b, \nu(\cdot))$ of L', we obtain the fundamental relation

$$E\left[\exp\{iuL(A)\}\right] = \exp\left\{\lambda^{Leb}(A) \log E\left[\exp\left\{iuL'\right\}\right]\right\}, \tag{4.3}$$

which shows that the law of $L(A)$ is fully determined by the Lévy seed L' and the Lebesgue measure of the set A.

We now give some examples of homogeneous Lévy bases, where throughout A is an arbitrary set in $\mathcal{B}_b(S)$.

Gaussian basis: Let $\nu(dx) \equiv 0$, so that the Lévy seed L' has characteristic triplet $(a, b, 0)$. Then L is a Gaussian basis, with

$$L' \sim N(a, b), \quad L(A) \sim N\left(a\lambda^{Leb}(A), b\lambda^{Leb}(A)\right).$$

In particular, if we let S be the real line, set $a = 0$, $b = \sigma^2$ and define a process by $\Lambda_t = L((0, t])$, then Λ has independent increments and $\Lambda_t \sim N(0, \sigma^2 t)$, so we recover the standard Brownian motion.

Poisson basis: Let $a = b = 0$ and $\nu(dx) = \gamma\delta_1(x)$, where δ_1 is the Dirac delta measure having a point mass at 1. Then

$$L' \sim \text{Poisson}(\gamma), \quad L(A) \sim \text{Poisson}(\gamma\lambda^{Leb}(A)),$$

so L is a Poisson Lévy basis. Working on the real line and defining $\Lambda_t = L((0, t])$ gives the familiar Poisson process with intensity γ.

Gamma basis: Let $a = b = 0$ and $\nu(dx) = \alpha x^{-1} e^{-\beta x}$. Then

$$L'(A) \sim \text{Gamma}(\alpha, \beta), \quad L(A) \sim \text{Gamma}\left(\alpha\lambda^{Leb}(A), \beta\right),$$

so L is a Gamma Lévy basis.

These examples illustrate how the area $\lambda^{Leb}(A)$ scales the cumulants linearly, as follows from the characteristic function relation (4.3).

4.2.2 Poisson point process representation

This subsection shows how the Lévy basis construction can be related to a Poisson point process, and consequently to some well-known models in the literature.

In particular, we will focus on the case when the random measure L in definition 4.2.1 can be interpreted as a map from the probability space to the set of (positive) measures on S, as opposed to a collection of real-valued random

variables. It is clear that this requires L to take non-negative values a.s. This connects our definition of a random measure to the concept of a "completely random measure", as described in Kingman (1967). In the more general case, this correspondence is a consequence of the Lévy-Itô decomposition for independently scattered random measures which is given in Pedersen (2003).

To illustrate the correspondence, let L be a non-negative Lévy basis, with characteristic triplet defined in terms of densities as in (4.2), where $a \in \mathbb{R}$ is constant, ν is a Lévy measure, and b is necessarily zero, as L has no Gaussian part. By analogy with Lévy processes having only positive increments (called subordinators, cf. Appendix A.1), the Lévy-Khintchine representation of L can be rewritten as

$$E[e^{iuL(A)}] = \exp\left\{iu\tilde{a}\lambda^{Leb}(A) + \lambda^{Leb}(A)\int_0^\infty (e^{iux} - 1)\nu(dx)\right\}, \quad u \in \mathbb{R}.$$

From the above representation it can be seen that L is a completely random measure, as defined in Kingman (1967). Such measures can be decomposed into three parts; the first consisting of fixed atoms, the second a deterministic part (corresponding to the term $\tilde{a}$ above), and the third a countable collection of random atoms. It can be seen from the above Lévy-Khintchine representation that, as a measure on S, L has no fixed atoms with probability one. Hence the non-deterministic part of L can be written as

$$\tilde{L} = \sum_{i=1}^{\infty} \gamma_i \delta_{s_i} \quad a.s., \tag{4.4}$$

where (s_i, γ_i) come from an inhomogeneous Poisson point process on $S \times \mathbb{R}^+$ with mean measure $n(dz, dx) = dz\nu(dx)$.

As described in Jordan (2010), this representation can be used to obtain random probability measures by normalisation. In particular, using Campbell's theorem we have that the sum $\sum_i \gamma_i$ is finite iff $\int_{\mathbb{R}^+}(x \wedge 1)\nu(dx)$ is also finite, which is necessarily true as ν is the Lévy measure of a subordinator. Thus we can normalise the above representation, replacing γ_i by $\gamma_i / \sum_j \gamma_j$, such that $\tilde{L}(S) = 1$ almost surely, resulting in a random probability measure. A special case is obtained by starting with the Gamma process, giving rise to the following example:

Dirichlet process: Let L be a Gamma process, i.e. a Lévy basis on S with

characteristic triplet given by $a = 0, b = 0$ and $\nu(dx) = \alpha x^{-1} e^{-\beta x} dx$. Normalising the corresponding Poisson point process results in the familiar Dirichlet process, which is the starting point for much of Bayesian nonparametric statistics, (Ferguson, 1973; Antoniak, 1974).

The Poisson point process representation also links the Lévy basis to other models in the literature. For example, the papers by Møller and Torrisi (2005) and Brix (1999) discuss generalisations of shot-noise Cox processes for use in spatial modelling, where the random intensity measure can be represented as a Poisson point process analogous to (4.4).

4.3 Trawl processes

In this section we introduce trawl processes, which are a class of stationary, infinitely divisible stochastic processes defined in terms of an underlying Lévy basis (Barndorff-Nielsen, 2011; Barndorff-Nielsen et al., 2015). Sections 4.3.1 and 4.3.2 introduce trawl processes and present basic distributional properties that have been derived in e.g. Barndorff-Nielsen et al. (2015), which are stated in our notation for completeness.

The remainder of this section introduces two novel representations of the trawl process. In Section 4.4 we derive a representation in terms of a collection of independent random variables; this is particularly important for modelling purposes because it enables exact simulation, as well as forming the basis of the sampling-based parameter estimation method described in Chapter 6. This representation makes the trawl process easier to work with in practice than the more general ambit or Lévy semistationary processes, which require computationally intensive approximations (Bennedsen et al., 2014; Hedevang and Schmiegel, 2014).

Section 4.5 focuses on the temporal structure of the trawl process, in particular the autocorrelation function, demonstrating that the trawl process can be interpreted as accumulations of random pulses from a causal system.

4.3.1 Trawl process definition

In the following we work with the space $S = \mathbb{R}^d \times \mathbb{R}$, with associated σ-algebra $\mathcal{B}(S)$, and the collection $\mathcal{B}_b(S)$ of sets with bounded Lebesgue measure $\lambda^{Leb}(\cdot)$.

Definition 4.3.1: A trawl process Λ_t is generated by taking a homogeneous Lévy basis L and evaluating it over a set A called the trawl. More specifically, let A be any set in $\mathcal{B}_b(\mathbb{R}^d \times \mathbb{R})$, and define a collection of trawl sets $\{A_t\}$ by shifting A along the $\mathbb{R}$-axis corresponding to the last coordinate, which represents time:

$$A_t = A + (\mathbf{0}, t) := \{(a_1, \ldots, a_d, a_{d+1} + t) : (a_1, \ldots, a_{d+1}) \in A\}.$$

The trawl process Λ_t is then defined as

$$\Lambda_t = L(A_t). \tag{4.5}$$

Remark 4.3.2: The trawl process definition can be written as a stochastic integral, which will become useful for calculations in the following. Specifically, we write

$$\Lambda_t = \int_{\mathbb{R}^d \times \mathbb{R}} \mathbb{I}_{A_t}(\xi, s) L(d\xi, ds) = \int_{\mathbb{R}^d \times \mathbb{R}} \mathbb{I}_A(\xi, s - t) L(d\xi, ds),$$

where points in $\mathbb{R}^{d+1}$ are denoted (ξ, s) for $\xi \in \mathbb{R}^d, s \in \mathbb{R}$, so the last component corresponds to the time axis. The general stochastic integral is defined in Rajput and Rosinski (1989) and reviewed in Barndorff-Nielsen et al. (2015); its characteristic function is given in Proposition A.1.16. This integral form also illustrates the connection to the wider class of ambit processes, discussed in Section 4.1, which take the form of a stochastic integral with respect to a Lévy basis; hence the trawl process is an ambit process where the integrand is an indicator function.

The stochastic integral definition implies that a trawl process is also a so-called mixed moving-average process; it was shown in Fuchs and Stelzer (2013) that such processes are mixing, so it follows that trawl processes are mixing and ergodic.

Figures 4.1 and 4.2 illustrate the trawl process definition for $d = 1$, showing Gaussian and Gamma Lévy bases, respectively, with a trawl set superimposed at different time points. These images show a discretisation of the actual Lévy basis, which is obtained by taking a fine grid covering a sufficiently large subset of $[0, 1] \times \mathbb{R}$, evaluating the random measure L over each of the grid squares and simulating the resulting random variables. The resulting sample paths

are also shown, which are discretisations of the full paths, with time steps equal to the grid spacing.

Figure 4.3 shows a Poisson Lévy basis, represented exactly as points in $[0,1] \times \mathbb{R}$, with the same trawl sets superimposed, and the resulting sample path. The trawl set A_t used in these figures is defined as the area between the time axis $\mathbb{R}$ and the function

$$d_{exp}(s) = \exp(-3(t-s)), \quad s \in (-\infty, t], \tag{4.6}$$

which is called an *exponential trawl*.

4.3.2 Distributional properties of trawl processes

The marginal distribution of the trawl process value $\Lambda_t = L(A_t)$ is specified via the characteristic function in equation (4.3), where L is generated by the Lévy seed L' as described in Section 4.2.1. From this equation it follows that Λ_t only depends on the trawl set A_t through its size, i.e. the Lebesgue measure $\lambda^{Leb}(A_t)$. By construction, the shape of the trawl set is constant over time, hence $\lambda^{Leb}(A_t) = \lambda^{Leb}(A)$ for all t, resulting in a stationary process.

The following Lemma shows that when using an appropriate normalisation for the cumulant of the Lévy seed L', the particular choice of trawl set does not affect the trawl process marginally.

Lemma 4.3.3: For any infinitely divisible distribution $\mathcal{D}$ and set $A \in B_b(\mathbb{R}^d \times \mathbb{R})$, there exists a trawl process having trawl set A and marginal law $\mathcal{D}$.

Proof. Consider a random variable $X \sim \mathcal{D}$, and define a Lévy seed L' via its cumulant function (cf. A.1.6)

$$C\left(u\,; L'\right) := \log E\left(\exp\{iuL'\}\right) = \frac{C(u; X)}{\lambda^{Leb}(A)}.$$

Using (4.3) then gives

$$E\left[\exp\{iu\Lambda_t\}\right] = E\left[\exp\{iuL(A_t)\}\right] = \exp\{\lambda^{Leb}(A_t) C\left(u\,; L'\right)\} = \exp\{C\left(u\,; X\right)\},$$

such that $\Lambda_t \sim \mathcal{D}$ for any $t \geq 0$. □

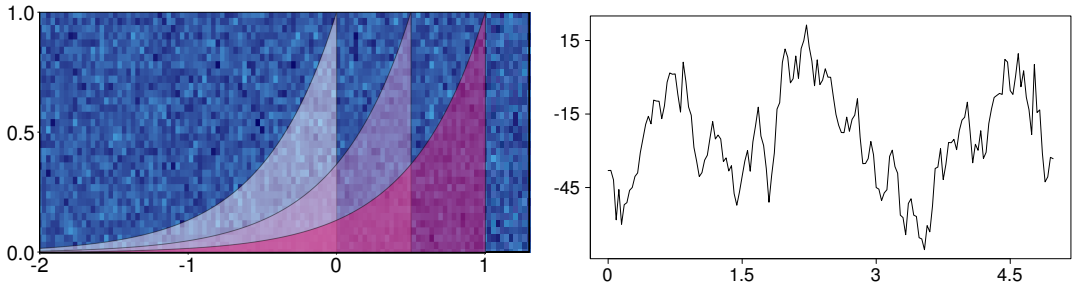

Figure 4.1: Gaussian discretised Lévy basis with superimposed exponential trawl sets (left) and resulting trawl process (right).

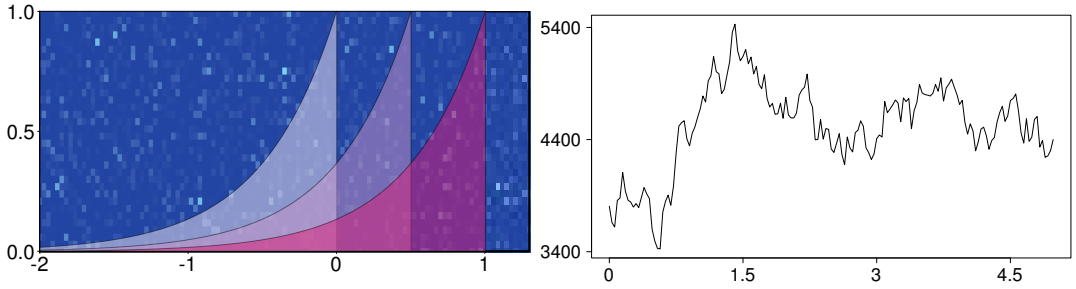

Figure 4.2: Gamma discretised Lévy basis with superimposed exponential trawl sets (left) and resulting trawl process (right).

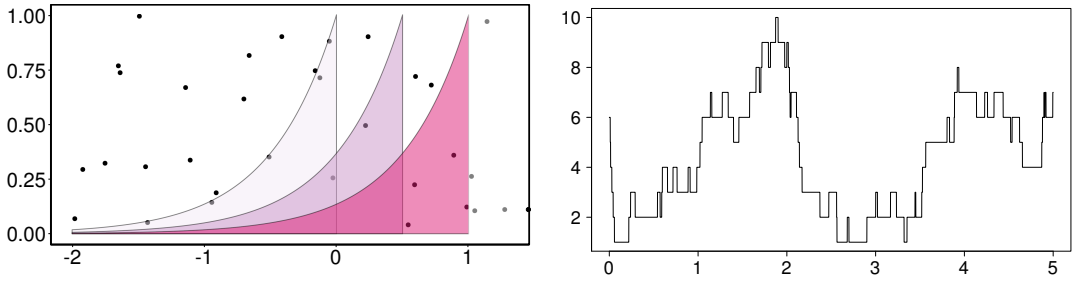

Figure 4.3: Poisson Lévy basis shown as points with superimposed exponential trawl sets (left) and resulting trawl process (right).

As will be shown in Section 4.5, the autocorrelation structure of the trawl process is fully characterised by the shape of the trawl set. Thus the temporal and marginal properties of the trawl process can be separated, which is a useful property for statistical applications.

We now turn to the joint distribution of trawl process observations, which is specified by the following Lemma:

Lemma 4.3.4: Let $\Lambda_t = L(A_t)$ be a trawl process, then the joint characteristic function of $\Lambda_{t_1}, \ldots, \Lambda_{t_k}$ is given by

$$E\left[\exp\left\{i\sum_{m=1}^{k} u_m \Lambda_{t_m}\right\}\right] = \exp\left\{\int_{\mathbb{R}^d\times\mathbb{R}} C\left(\sum_{m=1}^{k} u_m \mathbb{I}_{A_{t_m}}(y)\ ;\ L'\right)\, dy\right\},$$

Proof. Proposition 5 in Barndorff-Nielsen et al. (2015) gives the cumulant function of the quantity $\mu(\Lambda) = \int_{\mathbb{R}} \Lambda_t\, \mu(dt)$ for any measure μ, assuming this integral exists a.s. and is finite. In particular, they show that

$$C\left(\theta\ ;\mu(\Lambda)\right) = \int_{\mathbb{R}^d\times\mathbb{R}} C\left(\theta\left(\int_{\mathbb{R}} \mathbb{I}_A(\xi, s-t)\,\mu(dt)\right)\ ;L'\right)\, d\xi ds.$$

The result follows from this expression when using the measure $\mu(dt) = \sum_{m=1}^{k} u_m \delta_{t_m}(dt)$, with δ being the Dirac delta measure. □

The form of the above expression appears to differ from that of the marginal characteristic function given in (4.3). The connection becomes clearer if we rewrite the marginal expression by using an integral for the area of the trawl set, giving

$$\begin{aligned} E\left[\exp\left\{iu\Lambda_t\right\}\right] &= E\left[\exp\left\{iuL(A_t)\right\}\right] = \exp\left\{\lambda^{Leb}(A)\log E[\exp iuL']\right\} \\ &= \exp\left\{\left(\int_{\mathbb{R}} \mathbb{I}_{A_t}(x)\,dx\right) C\left(u\ ;L'\right)\right\} = \exp\left\{\int_{\mathbb{R}} C\left(u\mathbb{I}_{A_t}(x)\ ;L'\right)\, dx\right\}, \end{aligned}$$

as $C\left(0\ ;L'\right) = 0$, hence we recover the form of the above Lemma with $k = 1$.

From the above Lemma, it is clear that the transition from the marginal to the joint distribution adds complexity to the expression for the characteristic function, as we can no longer reduce the integral to the area of a trawl set. In particular, the presence of several discontinuities in the argument of the cumulant function means that numerical approximations of the integral will

usually give poor results.

The expression for the characteristic function can be simplified, however, by noting that the argument of the integrand $C(\cdot\,; L')$ is the sum of contributions u_m from all trawl sets containing the point y; hence $C(\cdot\,; L')$ can only take finitely many different values as y ranges over $\mathbb{R}^d \times \mathbb{R}$, assuming the total number of observations k is fixed. This observation forms the basis of the following section, where we use the set of possible values of the cumulant function to define a corresponding set of random variables. This leads to a useful representation of the trawl process, which also gives a simplified expression for the joint characteristic function.

4.4 Partition representation

In this section we develop a novel representation of the trawl process in terms of independent random variables. This representation provides a method for simulating exactly from a general trawl process. It also forms the basis of the parameter estimation algorithm for trawl-based models that is presented in Chapter 6. We will also consider the particular case of a monotonic trawl, where the representation can be simplified further.

Consider the case where we have k observations $\Lambda_{t_1}, \ldots, \Lambda_{t_k}$ of a trawl process, as defined in (4.5), with k corresponding trawl sets $A_{t_1}, \ldots, A_{t_k}$. As described in the following definition, these trawl sets are generated by a particular collection of subsets.

Definition 4.4.1: We consider the collection of trawl sets $\{A_{t_i}\}_{i\in I}, I = \{1, \ldots, k\}$, and define the *slice partition* as the finest partition of $\bigcup_{i\in I} A_{t_i}$ such that each set in the partition takes the form

$$\left(\bigcap_{j\in J} A_{t_j}\right) \setminus \left(\bigcap_{l\in L} A_{t_l}\right), \quad J, L \subset I, J \neq \emptyset.$$

The sets in the slice partition are called *slice sets*, and such a partition is illustrated in Figure 4.4. In particular, all trawl sets and intersections of trawl sets can be written in the form above, and thus they can also be written as unions of slice sets.

For example, the slice partition for $k = 2$ consists of three slice sets: $\{A_{t_1} \cap A_{t_2},\ A_{t_2} \setminus A_{t_1},\ A_{t_1} \setminus A_{t_2}\}$. In general there are $2^k - 1$ slice sets in the slice partition, which we will denote by B_{π_j}, with n_B being the number of slice sets in the partition.

We now make two key observations:

1. Each observation Λ_{t_i} is a sum of independent random variables $L(B_{\pi_j})$. We can see this by writing

$$\Lambda_{t_i} = L(A_{t_i}) = L\left(\bigcup_{j:B_{\pi_j}\in\mathcal{T}(t_i)} B_{\pi_j}\right),$$

   where $\mathcal{T}(t_i)$ gives the collection of slice sets whose union equals A_{t_i}. We can then use the additivity and independent scattering of the Lévy basis to write

$$L\left(\bigcup_{j:B_{\pi_j}\in\mathcal{T}(t_i)} B_{\pi_j}\right) = \sum_j L(B_{\pi_j}) \quad a.s.,$$

   where the terms $L(B_{\pi_j})$ are independent.

2. Any two observations $\Lambda_{t_i}, \Lambda_{t_j}$ are conditionally independent given $L(B_{i,j})$, where $B_{i,j}$ is the union of slice sets giving the intersection of A_{t_i}, A_{t_j}. This follows from the independent scattering property of L. Furthermore, this property can be extended to a set of observations $\Lambda_{t_1}, \ldots, \Lambda_{t_n}$ by conditioning on $\{L(B_k)\}$ for the collection of sets $\{B_k\}$ giving all the intersections of $A_{t_1}, \ldots A_{t_n}$.

Note that the distribution of $L(B_{\pi_j})$ is defined via its cumulant function $|B_{\pi_j}|\, C\,(u\,; L')$; for many infinitely divisible distributions this corresponds to a distribution in the same class as L'.

From the above discussion we see that the observations $\Lambda_{t_1}, \ldots, \Lambda_{t_k}$ can be written in terms of sums of independent random variables $L(B_{\pi_j})$, which have known distributions. Thus the random variables $L(B_{\pi_j})$ constitute a tractable representation of the trawl process.

Definition 4.4.2: Let Λ_t be a trawl process with Lévy basis L, observed at times $t_1, \ldots, t_k$, with corresponding slice partition $\{B_{\pi_j}\}$. The random variables $\{L(B_{\pi_j})\}$ are then called *slices*, and B_{π_j} are the slice sets.

The above discussion is summarised in the following proposition:

Proposition 4.4.3: A trawl process can be simulated exactly by using the slice partition, conditional on the following assumptions:

1. The area $\lambda^{Leb}(B_{\pi_j})$ of each slice is known.
2. We can simulate from the distribution corresponding to the cumulant function $C(u, \cdot) = \alpha C(u\,; L')$ for any $\alpha > 0$.

The slice representation can be used to obtain a simplified version of the joint characteristic function for the trawl process.

Lemma 4.4.4: The joint characteristic function of $(\Lambda_{t_1}, \ldots, \Lambda_{t_k})$ is given by

$$E\left[\exp\left\{i\sum_{m=1}^{k} u_m \Lambda_{t_m}\right\}\right] = \prod_{j=1}^{n_B} \exp\left\{\lambda^{Leb}(B_{\pi_j})\, C\left(\sum_{m:A_m \supset B_{\pi_j}} u_m\,; L'\right)\right\}.$$

Proof. By construction of the slices, we have

$$\begin{aligned} E\left[\exp\left\{i\sum_{m=1}^{k} u_m \Lambda_{t_m}\right\}\right] &= E\left[\exp\left\{i\sum_{m=1}^{k} u_m \left(\sum_{j:B_{\pi_j}\subset A_m} L(B_{\pi_j})\right)\right\}\right] \\ &= E\left[\exp\left\{i\sum_{j=1}^{n_B}\left(\sum_{m:A_m\supset B_{\pi_j}} u_m\right) L(B_{\pi_j})\right\}\right], \end{aligned}$$

and the result then follows from the independence of $L(B_{\pi_j})$ and relation (4.3). □

Although the slice partition is very useful in theory, the fact that it has size of order 2^k for k observations means that it quickly becomes computationally intractable for simulation and estimation purposes. This issue motivates the next subsection, which shows that the slice partition simplifies considerably for a monotonic trawl.

Monotonic trawl

We now consider the special case where the trawl set $A = A_0$ lies in $\mathbb{R}^2$ and is given by the area between the time axis and a monotonically increasing

function d on $(-\infty, 0]$, so we have

$$A_t = \{(\xi, s) : s \leq t,\ 0 \leq \xi \leq d(s-t)\}, \tag{4.7}$$

which is called a monotonic trawl. As will be discussed in Section 4.5, this form of trawl set is quite natural when considering causal systems.

When using a monotonic trawl, the expression for the intersection of multiple trawl sets can be simplified, as shown in the following Lemma:

Lemma 4.4.5: If the trawl is monotonic, as specified in (4.7), the full collection of slice sets for time points $t_1, \ldots, t_k$ is given by

$$\left\{ A^c_{t_{i-1}} \cap (A_{t_i} \cap A_{t_j}) \cap A^c_{t_{j+1}},\ 1 \leq i \leq j \leq k, \right\} := \{S(i,j)\}, \tag{4.8}$$

with the convention $A_{t_0} = A_{t_{k+1}} = \emptyset$, giving a partition of size $k(k+1)/2$.

Proof. We start by considering a slice given by the intersection of trawl sets; note that this is just the set bounded on the right by the first time point t_i, and bounded above by the smallest value of $d(s-t_j)$, which occurs at the last time point t_j, so the intermediate time points are irrelevant. Specifically, for $t_i < t_j$ we have $d(s-t_j) \leq \ldots \leq d(s-t_i)$, which gives

$$A_{t_i} \cap A_{t_{i+1}} \cap \ldots \cap A_{t_j} = \{(s,\xi) :\ s \leq t_i\ 0 \leq \xi \leq d(s-t_j)\} = A_{t_i} \cap A_{t_j}.$$

In particular, this implies that there are no sets in the partition of form $A_{t_i} \cap A^c_{t_k} \cap A_{t_j}, \quad t_i \leq t_k \leq t_j$.

Another simplification occurs when considering intersections involving trawl set complements, by using $d(s-t_j) > \ldots > d(s-t_k)$ to give

$$A_{t_i} \cap A^c_{t_{i+1}} \cap \ldots \cap A^c_{t_k} = \{(s,\xi) :\ s \leq t_i,\ \xi \leq d(s-t_i),\ \xi > d(s-t_{i+1})\} = A_{t_i} \cap A^c_{t_{i+1}}.$$

Similarly we obtain

$$A^c_{t_i} \cap A^c_{t_{i+1}} \cap \ldots \cap A_{t_j} = A^c_{t_{j-1}} \cap A_{t_j},$$

and the result then follows. □

We can obtain explicit expressions for the slice sets in terms of the trawl

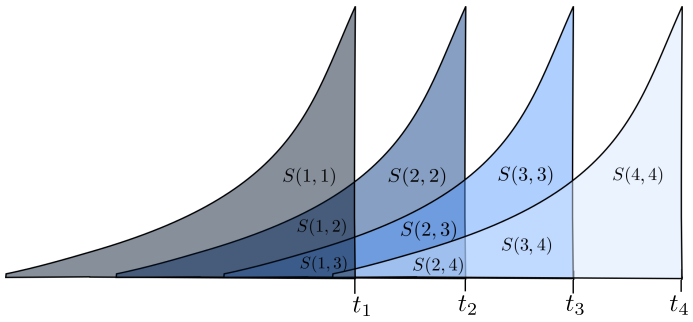

Figure 4.4: Illustration of slice partition.

function d that bounds the trawl set above, using the definition in (4.8) we have

$$\begin{aligned} S(i,j) &= \{\xi, s : d(s - t_{j+1}) \leq \xi \leq d(s - t_j),\ t_{i-1} \leq s \leq t_i\}, \quad i,j \neq 1, \\ S(1,j) &= \{\xi, s : d(s - t_{j+1}) \leq \xi \leq d(s - t_j),\ -\infty \leq s \leq 0\}, \end{aligned} \tag{4.9}$$

as illustrated in Figure 4.4.

The area of these sets is given by

$$|S(i,j)| = D(t_i - t_j) - D(t_i - t_{j+1}) - D(t_{i-1} - t_j) + D(t_{i-1} - t_{j+1}), \tag{4.10}$$

for $i, j \neq 1$, where

$$D(t) = \int_{-\infty}^{t} d(s)\, ds.$$

4.5 Temporal dependence structure of trawl processes

We now discuss in detail how the trawl set determines the autocorrelation structure of the trawl process, which provides a basis for choosing which trawl set to use in a given model.

In the trawl process definition 4.3.1, we only required the trawl set A to be in $\mathcal{B}_b(\mathbb{R}^d \times \mathbb{R})$, with points in A denoted by (ξ, s). It will be useful in the following to consider the d-dimensional cross-sections of A, defined as

$$\mathcal{K}_s = \{\xi \in \mathbb{R}^d \,\big|\, (\xi, s) \in A\}, \tag{4.11}$$

where we assume the trawl is bounded in the subspace $\mathbb{R}^d$, such that each $\mathcal{K}_s$ has finite Lebesgue measure. These cross-sections can be used to rewrite the

cumulant function of the trawl process Λ_t, which is given in (4.3) in terms of the Lévy seed L'. In particular,

$$\begin{aligned} C\left(u\,;L(A_t)\right) = \lambda^{Leb}(A_t)C\left(u\,;L'\right) &= \left(\int_{s:(\xi,s)\in A_t} \lambda^d(\mathcal{K}_s)\,ds\right) C\left(u\,;L'\right) \\ &= \int_{s:(\xi,s)\in A_t} \lambda^d(\mathcal{K}_s)C\left(u\,;L'\right)\,ds, \end{aligned} \tag{4.12}$$

where λ^d denotes the d-dimensional Lebesgue-measure. This representation shows that, in terms of its cumulants, $L(A)$ can be interpreted as a superposition of variables L_s, which have cumulant function

$$C\left(u\,;L_s\right) = \lambda^d(\mathcal{K}_s)C\left(u\,;L'\right). \tag{4.13}$$

In this interpretation, the variable L_s is seen as the contribution from some abstract event at time s to the value of the trawl process $\Lambda_t = L(A_t)$, and Λ_t can in turn be seen as an accumulation of random pulses generated in the past. This shows why it is common to consider trawl sets A_t where the last coordinate is bounded on the right by t, as this means that the value of Λ_t is only influenced by events at times $s < t$, corresponding to a causal system.

For most causal systems it is reasonable to expect that the contribution of recent events will be greater than that of events far in the past, which can be stated more precisely as

$$E(L_s) \leq E(L_v), \quad s < v \leq t. \tag{4.14}$$

Using equation (4.13) we get that $E(L_s) = \lambda^d(\mathcal{K}_s)E(L')$, so the above assumption implies that the cross-section area $\lambda^d(\mathcal{K}_s)$ increases monotonically with s. In particular, this motivates the assumption of a monotonic trawl, as discussed in Section 4.4.

4.5.1 Autocovariance function

By characterising the trawl set in terms of cross-sections we obtain a simple expression for the autocovariance function of the trawl process Λ_t. Using

Lemma 4.3.4, we can differentiate with respect to u_j to show that

$$E[\Lambda_i] = \lambda^{Leb}(A_{t_i})E(L'),$$
$$E[\Lambda_i\Lambda_j] = \lambda^{Leb}(A_{t_i} \cap A_{t_j})E(L'^2).$$

Hence the autocovariance is given by

$$\varphi(h) = \lambda^{Leb}(A_0 \cap A_h)\operatorname{Var}(L'). \tag{4.15}$$

The area of the trawl set intersection can be written in terms of the cross-sections, specifically

$$\lambda^{Leb}(A_0 \cap A_h) = \int_{s:(\xi,s)\in A_0\cap A_h} \lambda^{Leb}(\mathcal{K}_s \cap \mathcal{K}_{s-h})\, ds$$

This expression can be simplified if we refine the assumptions about the cross-sections mentioned above. First, we assume that the trawl set A_0 is bounded on the right, so that it only contains points (ξ, s) for $s \in (-\infty, 0]$, corresponding to a causal system. Second, we assume that the cross-sections are nested, such that $\mathcal{K}_s \cap \mathcal{K}_{s-h} = \mathcal{K}_{s-h}$. This is a slightly stronger assumption than (4.14), which implies that the cross-sections have increasing areas, as we are now concerned with geometric properties. However, it is clear from (4.12) that the distribution of the trawl process Λ_t is fully specified in terms of the areas of the cross-sections, and so there is no loss of generality for the trawl process in making the nesting assumption.

Having made these assumptions, we can now write

$$\begin{aligned}\varphi(h) &= \operatorname{Var}(L')\int_{-\infty}^{0} \lambda^{Leb}(\mathcal{K}_{s-h})\, ds = \operatorname{Var}(L')\int_{h}^{\infty} \lambda^{Leb}(\mathcal{K}_{-v})\, dv, \\ &= \operatorname{Var}(L')\left(\lambda^{Leb}(A) - \int_0^h \lambda^{Leb}(\mathcal{K}_{-v})\, dv\right).\end{aligned}$$

This expression can be used to solve for the area $\lambda^{Leb}(\mathcal{K}_v)$ in terms of the autocovariance function by differentiating the above expression to get

$$\lambda^{Leb}(\mathcal{K}_{-h}) = -\varphi'(h)/\operatorname{Var}(L'), \quad h \geq 0.$$

Since the area $\lambda^{Leb}(\mathcal{K}_{-h})$ must be non-negative, and by assumption is mono-

tonically decreasing with increasing v, it follows that $\varphi'(\cdot)$ must be non-positive and have a magnitude that is monotonically decreasing.

The above relation shows that given a monotonically decreasing autocovariance function $\varphi(\cdot)$ with an absolutely decreasing derivative, we can construct $\mathcal{K}_s$ and the corresponding trawl set A such that the process $\Lambda_t = L(A_t)$ has autocovariance function φ. This is shown in Barndorff-Nielsen et al. (2014) for integer-valued trawl processes with trawl sets $A \in \mathbb{R} \times \mathbb{R}$.

5

LATENT TRAWL MODEL FOR EXCEEDANCES

5.1 Introduction

This chapter presents a new model for characterising temporal dependence in exceedances above a threshold, which makes use of the trawl process class introduced in Chapter 4. The model uses the same hierarchical structure as Bortot and Gaetan (2014), thus ensuring that the marginal distribution of the exceedances is generalised Pareto, which is the proper limiting distribution for exceedances from classical extreme value theory.

Modelling dependencies in extreme value data is a topic of growing importance, with applications in a number of fields such as hydrology (de Haan and de Ronde, 1998), oceanography (Coles and Tawn, 1991), financial risk management (Ledford and Tawn, 2003; Embrechts et al., 1997) and environmental science (Davison et al., 2012; Heffernan and Tawn, 2004).

For extreme value data it is particularly valuable to have a model that captures the dependence structure, for two reasons: first, it allows us to borrow strength across multiple observations, which is especially relevant as extreme data are scarce by definition. Second, the impact of extreme events may depend on the occurrence of other events in the past, present or future; for example, the risk of flooding depends heavily on whether rainfall or storm surge events tend to occur close together in time.

Historically, the focus of extreme value analysis has been on modelling the tails of univariate i.i.d. observations. A series of papers, summarised in Leadbetter et al. (1983), expanded the existing theory to hold for stationary sequences under some mild assumptions. There has also been much work on models for multivariate extremes, which have been applied to both time series and observations from multiple sources (Coles and Tawn, 1991; Bortot and Tawn, 1998).

A recurring problem when modelling dependent extremes is the trade-off between the tractability of a given model and its ability to capture the full dependence structure. The class of max-stable processes is promising due to its flexibility and grounding in extremal theory, and has been used for spatio-temporal modelling (see Huser and Davison, 2014, and references therein). However, the computational burden associated with estimation appears to be a general issue for this class of models. Furthermore, the models fail to represent the typical processes found in environmental applications as they do not allow for realistic temporal dynamics – they produce similar dependence structures in time and space (this is discussed in Embrechts et al., 2015, who propose an alternative class of max-stable models).

In order to obtain tractable models, strong assumptions are often made about the dependence structure of extremes. It is often the case that a given time series suggests a natural block division (e.g. with blocks corresponding to months or years), in which case the modelling is done on block maxima, which are assumed to be independent. Although this assumption is justified asymptotically, this procedure makes inefficient use of the data, which in the extreme value case are scarce to begin with; furthermore, it does not describe the dependence structure within blocks, which may be of independent interest.

A popular alternative to working with block maxima is to consider exceedances above a threshold, based on the "peaks over threshold" (POT) method used in hydrology (Davison and Smith, 1990), which fits the exceedances to a generalised Pareto distribution. This method is often used together with a declustering method, which tries to identify clusters in the data and use these to extract approximately independent observations of the quantity of interest. Declustering methods are based on a fundamental result in Hsing et al. (1988), which shows that the normalised point process of exceedances converges to a compound Poisson process as the threshold increases, and that the cluster maxima are independent in the limit. Although they are theoretically justified, declustering methods will inevitably introduce additional uncertainty associated with the choice of declustering method and associated parameters (though see Ferro and Segers, 2003, for an automated selection procedure), and as with the block maxima approach they do not use all of the available data.

Among models that try to model the dependence explicitly, a common as-

sumption is that a given time series of extremes has the Markov property. This assumption motivates the approach used by Smith et al. (1997), who consider a Markov chain with transition probabilities derived from bivariate extremal dependence models. Ledford and Tawn (1997) consider a more general dependence structure based on an asymptotic expansion of the joint survivor function, which leads to a wider class of bivariate dependence models.

As shown by the above discussion, the main current approaches to modelling time series of extremes do not allow for a complete description of the temporal dependence structure for processes that are not essentially independent or Markov. One could in theory work with higher order Markov chains and use multivariate extremes models of higher dimensions, but this quickly becomes intractable in practice due to the computational burden and the lack of a finite parametrisation for the joint distributions of such models. A more fundamental issue is the application of multivariate extremes models to time series without any consideration for the time-specific aspects such as causality, which could potentially be used to reduce the dimensionality of the problem.

The main contribution of this chapter is a new extreme value model that is motivated by the above concerns. By using a trawl process in the latent part of the hierarchical structure, we allow for a flexible dependence structure, only imposing some assumptions that have clear physical interpretations (as described in Section 4.5), in contrast with the Markov chain setup used in Bortot and Gaetan (2014). In Section 5.4.2 we derive a measure of extremal dependence that is adapted to our model. We also consider a modified version of the model that enables it to be used for generalised Pareto distributions with a negative shape parameter; this modification also has the effect of separating the marginal and dependence structures.

This chapter is structured as follows: The following section reviews the theory underlying extreme value statistics, focusing on methods for capturing dependence in time series of extremes. The model for dependent extremes is introduced in Section 5.3, and finally Section 5.4 discusses parameter estimation and applies the model to different environmental time series.

5.2 Background on extreme value analysis

This section discusses some fundamentals of extreme value theory, which form the basis of the models discussed in this chapter. We pay particular attention to theory for observations with temporal dependence.

5.2.1 Convergence of maxima

We start with a fundamental relation which links the distribution of the maxima of an i.i.d sequence to the tail of its distribution function, which can be found in Leadbetter et al. (1983, p. 13):

Proposition 5.2.1: Let $M_n = \max(X_1, \ldots, X_n)$ for (X_i) an i.i.d sequence with distribution function F, and let $u_n \in \mathbb{R}$ be a sequence of constants, then the following are equivalent:

$$\begin{aligned} n(1 - F(u_n)) &\to \tau, \quad n \to \infty, \\ P(M_n \leq u_n) &\to e^{-\tau} \quad n \to \infty. \end{aligned} \tag{5.1}$$

This proposition has a simple justification – let N be the number of indices i such that $X_i > u_n$, then N is a Binomial random variable with n trials and success probability $1 - F(u_n)$. Then as n increases, $1 - F(u_n)$ goes to zero, and we get the standard convergence to a Poisson distribution with rate τ, where $P(M_n \leq u_n)$ is just the Poisson probability of a zero value.

From this relation we see the crucial role of the constants u_n in analysing the limiting properties of the maximum. The basic idea is that as n increases, the maximum of $X_1, \ldots, X_n$ will also increase, and depending on the properties of the distribution function F, we can choose u_n to increase in a coordinated way such that $P(M_n \leq u_n)$ converges.

The fundamental theorem in extreme values, the Fisher-Tippett or Extremal types theorem, states that if the sequence of normalised maxima of i.i.d random variables does converge, then this distribution must in fact belong to one of three classes. More precisely, we have (see e.g. Embrechts et al., 1997):

Theorem 5.2.2 (Extremal types theorem): Consider a sequence of i.i.d. random variables $X_1, \ldots, X_n$, and assume that there exists a non-degenerate

distribution function H such that

$$P\left(\frac{\max(X_1,\ldots,X_n)-d_n}{c_n}\leq x\right)\to H(x),\ x\in\mathbb{R},$$

for constants $c_n>0, d_n\in\mathbb{R}$, then H corresponds to one of three distributions:

$$\text{Fréchet: } \Phi_\alpha(x)=\begin{cases}0, & x\leq 0\\ \exp\{-x^{-\alpha}\}, & x>0\end{cases}\quad \alpha>0.$$

$$\text{Weibull: } \Psi_\alpha(x)=\begin{cases}\exp\{-(-x)^{\alpha}\}, & x\leq 0\\ 1, & x>0\end{cases}\quad \alpha>0.$$

$$\text{Gumbel: } \Lambda(x)=\exp\{-e^{-x}\},\quad x\in\mathbb{R}.$$

The normalising constants c_n, d_n above correspond to the relation in Proposition 5.2.1 via $u_n=c_nx+d_n$, with $P(M_n\leq u_n)\to H(x)$.

The three distributions in the Extremal Types theorem can be written as a single class, the *Generalized Extreme Value* distribution (GEV), which has distribution function

$$F(x)=\exp\left\{-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right\},$$

for $1+\xi\frac{x-\mu}{\sigma}>0,\ \mu\in\mathbb{R},\ \sigma>0,\ \xi\in\mathbb{R}$. The case $\xi>0$ corresponds to the Fréchet distribution, $\xi<0$ the Weibull distribution and $\xi=0$ the Gumbel distribution.

The theorem above provides the justification for modelling maxima of observed random variables by the GEV distribution. Fitting this distribution to data allows us to characterise the extremal behaviour of empirical processes and make statements about the probability of future extreme events. However, the fundamental assumption of independent observations is usually not realistic in practice. Important extensions to stationary sequences are developed in Leadbetter et al. (1983), where it is shown that the basic theory holds for stationary sequences under the so-called Leadbetter's D condition:

Definition 5.2.3: Let u_n be a given sequence of constants, and F the distribution function of a stationary sequence (ξ_n). We say that condition $D(u_n)$

holds if for any integers

$$1 \leq i_1 < \ldots < i_p < j_1 < \ldots < j_q \leq n,$$

where $j_1 - i_p \geq l_n$, we have

$$|F_{i_1,\ldots,i_p,j_1,\ldots,j_q}(u_n) - F_{i_1,\ldots,i_p}(u_n)F_{j_1,\ldots,j_q}(u_n)| \leq \alpha_{n,l_n},$$

where $\alpha_{n,l_n} \to 0$ as $n \to \infty$ for some sequence $l_n = o(n)$. Note that here we take $F_{i_1,\ldots,i_p}(u_n)$ to mean $F_{i_1,\ldots,i_p}(u_n,\ldots,u_n)$, i.e. the joint distribution evaluated at $(u_n,\ldots,u_n)$.

The following theorem is given in Leadbetter et al. (1983, p.66), and provides a modified version of relation (5.1) for stationary sequences:

Theorem 5.2.4: Assume that for some $\tau > 0$, and constants $u_n(\tau)$, we have that $n(1 - F(u_n(\tau))) \to \tau$ and that $D(u_n(\tau))$ holds for each such τ. Then there exists a constant $0 \leq \theta \leq 1$ such that if $P(M_n \leq u_n(\tau))$ converges for some $\tau > 0$, then the limit equals $e^{-\theta\tau}$. If this is the case we say that θ is the *extremal index*.

Here θ quantifies the degree of dependence in the sequence, where $\theta = 0$ is the degenerate case of complete dependence, and $\theta = 1$ gives complete independence, thus recovering the i.i.d result.

Finally, we present the result that shows how the same fundamental results hold in the stationary case under the D condition: (Leadbetter et al., 1983, p.69)

Theorem 5.2.5: Let $X_1,\ldots,X_n$ be a stationary sequence with extremal index $\theta > 0$, and let $\tilde{X}_i$ be a sequence of independent random variables with the same marginal distribution function as that of X_i. Then (X_i) has a nondegenerate limiting distribution if and only if $(\tilde{X}_i)$ does, the limiting distribution are of the same type, and the same normalising constants can be used.

5.2.2 Generalised Pareto distribution

The results above are concerned with the maximum M_n of a sequence $X_1, \ldots, X_n$ of random variables. In statistical applications it is often more useful to work with the distribution of the exceedances, i.e. the values above some threshold u. The distribution of the exceedances can be derived from a conditional version of the extremal types theorem. In particular, it can be shown that if X_i are i.i.d. random variables with distribution function F and $\bar{F} = 1 - F$, such that

$$\bar{F}(u + xa(u))/\bar{F}(u) \to \bar{G}(x),$$

for $u \to x_F$, with x_F being the upper endpoint of F and $a(u) > 0$, then the limiting distribution function G must be generalised Pareto (Leadbetter, 1991; Davison and Smith, 1990). The density of the generalised Pareto distribution (GPD) takes the form

$$f(x|\xi, \sigma) = \frac{1}{\sigma}\left(1 + \xi\frac{x}{\sigma}\right)_+^{-(1/\xi+1)},$$

where $(y)_+ = \max(0, y)$, $\xi \in \mathbb{R}$, $\sigma > 0$.

This result is used to motivate the common practice of modelling observations above a threshold by the generalised Pareto distribution, where the threshold u is fixed at a sufficiently high value to obtain a good empirical fit to the GPD.

5.3 Latent trawl process model

In this section we use the framework of trawl processes to construct a new model for extreme values observed over time, which extends the hierarchical setup of Bortot and Gaetan (2014).

The starting point for the model is the observation that the marginal distribution of exceedances converges to the generalised Pareto distribution (GPD), see Davison and Smith (1990). In Bortot and Gaetan (2014), a decomposition of the GPD is used to construct a hierarchical model for exceedances that preserves this distribution marginally. We adopt this hierarchical structure, and then proceed to introduce a new model by using the properties of trawl processes described in Section 4.

The original approach of Bortot and Gaetan (2014) involved using a Markov chain to generate dependence in the exceedances; by using a trawl process instead of a Markov chain we obtain a more flexible dependence structure. Moreover, the trawl process framework provides a unified procedure for generating processes with a given infinitely divisible distribution and autocovariance function, whereas Bortot and Gaetan (2014) consider two different specifications for the latent Markov chain with the same marginal distribution and autocovariance function.

We also consider a modification of the new model that enables it to be used for any generalised Pareto distribution, thus removing the restriction $\xi > 0$ on the shape parameter that was inherited from the original model of Bortot and Gaetan (2014). This means that the modified model can be used for processes with less heavy-tails, which are often found in environmental applications, as illustrated by the application to pollution data in Section 5.4.4. The modification also improves the interpretability of the parameters, and appears to make the estimation procedure more efficient.

5.3.1 Basic structure

This subsection introduces the hierarchical structure used for the extreme value model, which is taken from Bortot and Gaetan (2014). Consider a time-series $\{Y_j\}_{j=1,\dots,k}$, assumed to be strictly stationary. Our aim is to model the observations where the value of Y_j is considered to be extreme, meaning that $Y_j > u$ for a fixed threshold u.

In order to focus on the extreme values, we will only consider the values and occurrence times of exceedances. To this end, define the exceedances X_j by

$$X_j := \max(Y_j - u, 0), \quad j = 1, \dots, k. \tag{5.2}$$

As discussed in Section 5.2.2, assuming $\{Y_j\}$ are in the domain of attraction of some extreme value distribution, the conditional exceedances $\{X_j | X_j > 0\}$ converge to a generalised Pareto distribution (GPD) for an appropriate sequence of thresholds $u_n \to \infty$. Based on this result we will assume that the conditional distribution of X_j given $X_j > 0$ can be approximated by a GPD,

for a sufficiently large threshold u. The density of the GPD is written as

$$f_{\text{GPD}}(x|\alpha,\beta) = \frac{\alpha}{\beta}\left(1+\frac{x}{\beta}\right)_+^{-(\alpha+1)} \qquad x \geq 0,\ \alpha,\beta > 0,$$

where $(y)_+ = \max(0, y)$, which is a reparametrisation of the standard density with shape parameter $\xi = 1/\alpha$ and scale parameter $\sigma = \beta/\alpha$.

Following Bortot and Gaetan (2014) we now observe that the GPD can be represented as a mixture of exponential random variables with Gamma distributed parameters, motivating a hierarchical specification for the exceedance process $\{X_j\}$. In particular, the distribution of an observation X_j depends on the value of a latent process Λ at time t_j, denoted Λ_j. This latent process determines both the probability of observing an exceedance, corresponding to $X_j > 0$, and the distribution of the exceedances.

Specifically, we assume that the density of X_j is

$$f(x_j\,|\lambda_j) = \begin{cases} 1 - e^{-\kappa\lambda_j}, & x_j = 0, \\ e^{-\kappa\lambda_j}\lambda_j e^{-\lambda_j x_j}, & x_j > 0. \end{cases} \tag{5.3}$$

where the density f is defined with respect to a measure μ such that $\mu(dx) = \delta_0(dx) + dx$. This construction shows that conditionally on the value of Λ_j, the exceedance X_j is generated by a two-stage process: at the first stage X_j is set to zero with probability $1 - e^{-\kappa\lambda_j}$; at the second stage the distribution of X_j given $\{X_j > 0,\, \Lambda_j = \lambda_j\}$ is exponential with parameter λ_j. The latent process Λ may be interpreted as an inverse intensity, as higher values of Λ give a lower probability of exceeding the threshold u and a smaller expected value of exceedances.

We also require that the observations X_j are independent for distinct j, conditional on the corresponding values of Λ. Hence the conditional joint distribution of $(X_1, \ldots, X_k)$ can be written as a product:

$$f(x_1, \ldots, x_k\,|\lambda_1, \ldots, \lambda_k) = \prod_{j=1}^{k} f(x_j|\lambda_j). \tag{5.4}$$

This specification implies that any dependence between observations $X_1, \ldots, X_k$ comes from the dependence between corresponding elements of the latent process Λ.

To complete the GPD mixture construction, the latent process is required to have a Gamma marginal law, i.e.

$$\begin{aligned} &\Lambda_j \sim \text{Gamma}(\alpha, \beta), \quad \alpha, \beta > 0, \\ &C(u, \Lambda_j) = -\alpha \log(1 - iu/\beta), \end{aligned} \tag{5.5}$$

where C is the cumulant function as defined in Proposition 4.3.3. This specification introduces the restriction $\alpha > 0$, which means that the model can only capture data belonging to the Fréchet distribution class. Section 5.3.4 presents a modified version of the model that removes this restriction.

We now show that when using the above specification, the exceedances $\{X_j : X_j > 0\}$ have a GPD marginal law. This is done by integrating out the value of Λ to get the marginal distribution of X, where we drop subscripts in the following calculations to ease notation. Specifically, consider a set $B \in \mathcal{B}(\mathbb{R})$, then we have

$$P(\{X \in B, X > 0\}) = \int_B \int_0^\infty f(x, \lambda)\, d\lambda\, dx = \int_B \int_0^\infty \lambda e^{-\lambda(\kappa + x)} f(\lambda)\, d\lambda\, dx.$$

This can be written as an expectation involving Λ to give

$$\begin{aligned} &P(\{X \in B\} \cap \{X > 0\}) = \int_B E_\Lambda[\Lambda \exp\{-\Lambda(x + \kappa)\}]\, dx \\ &= \int_B -\frac{\partial}{\partial u} E_\Lambda[\exp\{-\Lambda u\}]\Big|_{u = x+\kappa} dx = \int_B \frac{\alpha}{\beta} \left(1 + \frac{x + \kappa}{\beta}\right)^{-(\alpha+1)} dx, \end{aligned}$$

where the cumulant function of Λ was used in the last step. Similarly, the unconditional probability of observing an exceedance is given by

$$P(X > 0) = E_\Lambda[e^{-\kappa\Lambda}] = \left(1 + \frac{\kappa}{\beta}\right)^{-\alpha} dx, \tag{5.6}$$

and hence the distribution of the exceedances is given by

$$P(X \in B | X > 0) = \int_B \frac{\alpha}{\beta + \kappa} \left(1 + \frac{x}{\beta + \kappa}\right)^{-(\alpha+1)} dx,$$

showing that the conditional distribution of $\{X\,|X > 0\}$ is given by GPD$(\alpha, \beta + \kappa)$.

5.3.2 Latent trawl process

The previous subsection describes a general hierarchical model setup, using the same structure as in Bortot and Gaetan (2014). So far, the latent process Λ has only been specified as having a Gamma marginal law. Now we depart from the approach used in Bortot and Gaetan (2014), where the latent process is assumed to be a Markov chain; rather, we consider a new model where Λ is a trawl process, as introduced in Chapter 4.

The conditional independence assumption of the hierarchical model means that any dependence between observations comes from the latent process, and hence this process should have a flexible dependence structure. This motivates the use of a trawl process, which can capture a wide range of dependence structures, as discussed in Section 4.5. Using a trawl process also means that the observations X_j can be seen as coming from a continuous-time process (X_t), which is useful for statistical applications where there may be missing or irregularly spaced data.

Bortot and Gaetan (2014) consider two particular classes for the latent Markov chain, the Gaver and Lewis process (G-LP) and the Warren process (WP), and proceed to show that these two classes result in different asymptotic properties of the extremes, even though they have the same autocorrelation function. In contrast, the latent trawl process in our model is specified by its trawl set, which corresponds to a particular autocorrelation function. As will be shown in Section 5.4.2, the resulting process is asymptotically independent, where the form of the dependence structure is influenced by the trawl set.

Marginal distribution

Section 4 introduced the definition of a Lévy basis and discussed how to construct a particular Lévy basis L from a random variable L', called the Lévy seed. We now use this approach to obtain a trawl process Λ with a Gamma marginal law, following Lemma 4.3.3. Specifically, we fix a trawl set A in $\mathbb{R} \times \mathbb{R}$, and let the Lévy seed have a normalised Gamma distribution, i.e.

$$L' \sim \text{Gamma}\left(\frac{\alpha}{\lambda^{Leb}(A)}, \beta\right), \tag{5.7}$$

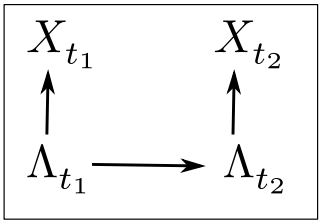

Figure 5.1: Illustration of latent trawl model structure.

then the trawl process defined by $\Lambda_t = L(A_t)$ has a Gamma(α, β) distribution.

Combining the trawl process Λ with the hierarchical model presented in the previous subsection, we obtain a stochastic process (X_t) with finite-dimensional densities given by

$$f(x_1, \ldots, x_k) = \begin{cases} \int_{\mathbb{R}_+^k} \left(\prod_{j=1}^k (1 - e^{-\kappa\lambda_j}) \right) dF(\lambda_1, \ldots, \lambda_k), & x_j = 0, \\ \int_{\mathbb{R}_+^k} \left(\prod_{j=1}^k \lambda_j e^{-\lambda_j(\kappa + x_j)} \right) dF(\lambda_1, \ldots, \lambda_k), & x_j > 0\,, \end{cases} \tag{5.8}$$

using the independence assumption in (5.4). These densities depend on the joint distribution of $(\Lambda_1, \ldots, \Lambda_k)$, which is fully specified by the trawl set A and the Lévy seed L', as shown in Lemma 4.4.4. The structure of the model is illustrated in figure 5.1.

Trawl set

To complete the definition of the trawl process Λ it now remains to specify the trawl set A. For our model we use the so-called exponential trawl set; this is the set $A_t \in [0, 1] \times \mathbb{R}$ that is bounded below by the time axis and above by the function

$$d_{exp}(s) = \exp(\rho s), \quad s \in (-\infty, t),$$

for some $\rho > 0$. The resulting process is then called the *exponential trawl process*, and hence we obtain a hierarchical model with parameters $(\rho, \alpha, \beta, \kappa)$.

As discussed in Section 4.5, the cross-section $\mathcal{K}_x$ of the trawl set can be associated with a random variable L_x, which is interpreted as the contribution to the value of Λ_0 coming from events x time units in the past. The expected value of L_x is scaled by the Lebesgue measure of the cross-section, seen as a subset of $\mathbb{R}^d$. For the exponential trawl with $d = 1$, the cross-sections are

vertical intervals, so

$$\mathcal{K}_x = \{(x, y) : y \leq e^{\rho x}\}, \quad x \leq 0,$$

with one-dimensional Lebesgue-measure $\lambda^{Leb}(\mathcal{K}_x) = \exp(\rho x)$.

The autocovariance function of the trawl process is given by $\varphi(h) = \lambda^{Leb}(A \cap A_h)\mathrm{Var}(L')$, and using the exponential trawl gives

$$\lambda^{Leb}(A) = \int_{-\infty}^{0} d(s)\, ds = \frac{1}{\rho}, \quad \lambda^{Leb}(A \cap A_h) = \int_{-\infty}^{0} d(s-h)\, ds = \frac{e^{-\rho h}}{\rho}.$$

Combining with (5.7) gives $\mathrm{Var}(L') = (\alpha\rho)/\beta^2$, resulting in the autocovariance function

$$\varphi(h) = e^{-\rho h}\frac{\alpha}{\beta^2} = e^{-\rho h}\mathrm{Var}(\Lambda_t).$$

Thus the autocorrelation function of the exponential trawl process has the same shape as the trawl set, which is a particular property of the exponential trawl.

We can also consider a general exponential trawl set, which is constructed from linear combinations of basic exponential trawls. Specifically, define the general exponential trawl set of order p to be bounded above by the function

$$d_p(x) = \sum_{i=1}^{p} w_i e^{\rho_i x},$$

with $\sum_i w_i = 1$. The latter restriction is necessary to make the parameters identifiable, as any scaling factor in the weights w_i will scale the area of the trawl, and thus be cancelled by the normalisation in (5.7). When using the general exponential trawl set, the resulting trawl process has an autocorrelation function $r(h)$ given by the corresponding linear combination of $e^{-\rho_i h}$. This construction can be seen as a deterministic version of a superposition-type trawl where the decay parameter ρ is randomised, as in Barndorff-Nielsen et al. (2014, Section 4); see also Barndorff-Nielsen (2001) for a similar approach applied to Ornstein-Uhlenbeck-type processes.

Summarising the above discussion, we have constructed a trawl process Λ with a marginal Gamma distribution and an autocorrelation function based on exponential decays. Using this as the latent process in the hierarchical structure results in a continuous-time process (X_t), where the exceedances

$\{X_t > 0\}$ have a generalised Pareto distribution. Furthermore, the model allows for dependence between observations $\{X_t\}$, which is derived directly from the dependence in the latent trawl process.

5.3.3 Autocovariance structure

We now consider the autocovariance structure of the exceedance process (X_t) from the latent model. As X is a stationary process, it has autocovariance function

$$\varphi(h) = E[X_0 X_h] - E^2[X].$$

The expected value of X is obtained by conditioning on the events $\{X = 0\}$ and $\{X > 0\}$, with probabilities given by (5.6), and using that $\{X|X > 0\}$ has a GPD$(\alpha, \beta + \kappa)$ distribution, which gives

$$E[X] = (1 + \kappa/\beta)^{-\alpha}(\beta + \kappa)/(\alpha - 1), \quad \alpha > 1.$$

The expectation of the product is more complicated, as it involves the joint density of the corresponding values of the latent trawl process:

$$\begin{aligned} E[X_0 X_h] &= 0 + \int_{\mathbb{R}_+^2} x_0 x_h \left(\int_{\mathbb{R}_+^2} f(x_0, x_h|\lambda_0, \lambda_h) f(\lambda_0, \lambda_h)\, d\lambda_0 d\lambda_h \right) dx_0\, dx_h \\ &= \int_{\mathbb{R}_+^2} x_0 x_h \left(\int_{\mathbb{R}_+^2} \lambda_0 e^{-\lambda_0(\kappa + x_0)} \lambda_h e^{-\lambda_h(\kappa + x_h)} f(\lambda_0, \lambda_h)\, d\lambda_0\, d\lambda_h \right) dx_0\, dx_h. \end{aligned}$$

Using Fubini's theorem to repeatedly exchange the order of integration gives

$$\begin{aligned} E[X_0 X_h] &= \int_{\mathbb{R}_+^2} \lambda_0 \lambda_h e^{-(\lambda_0\kappa + \lambda_h\kappa)} f(\lambda_0, \lambda_h) \left(\int_{\mathbb{R}_+^2} x_0 x_h e^{-(x_0\lambda_0 + x_h\lambda_h)}\, dx_0\, dx_h \right) d\lambda_0\, d\lambda_h \\ &= \int_{\mathbb{R}_+^2} \lambda_0 \lambda_h e^{-(\lambda_0\kappa + \lambda_h\kappa)} f(\lambda_0, \lambda_h) \left(\frac{1}{\lambda_0^2 \lambda_h^2} \right) d\lambda_0\, d\lambda_h \\ &= E_{\Lambda_0, \Lambda_h} \left[\frac{e^{-(\Lambda_0\kappa + \Lambda_h\kappa)}}{\Lambda_0 \Lambda_h} \right] = \int_\kappa^\infty \int_\kappa^\infty E_{\Lambda_0, \Lambda_h} \left[e^{-(\Lambda_0 u_0 + \Lambda_h u_h)} \right] du_0\, du_h. \end{aligned}$$

The joint expectation can now be found using the partitioning method described in Section 4.4. Specifically, partition the trawl sets A_0, A_h into disjoint sets

$$B_0 = A_h \setminus A_h,\ B_{0,h} = A_0 \cap A_h,\ B_h = A_h \setminus A_0,$$

and define the slice variables $\{S_i = L(B_i)\}$, where L is the Lévy basis derived from (5.7), so in particular these variables are independent. Furthermore, we have $\Lambda_0 \stackrel{d}{=} S_0 + S_{0,h}$ and $\Lambda_h \stackrel{d}{=} S_{0,h} + S_h$.

Let b_i be the shape parameter of S_i, which is given by $\alpha\lambda^{Leb}(B_i)/\lambda^{Leb}(A)$, then we can rewrite the above expression as

$$\begin{aligned} E[X_0 X_h] &= \int_\kappa^\infty \int_\kappa^\infty E_{S_0,S_{0,h},S_h}\left[\exp\{-\left[u_0 S_0 + (u_0+u_h)S_{0,h} + u_h S_h\right]\}\right] du_0\, du_h \\ &= \int_\kappa^\infty \int_\kappa^\infty \left(1+\frac{u_0}{\beta}\right)^{-b_0} \left(1+\frac{u_0+u_h}{\beta}\right)^{-b_{0,h}} \left(1+\frac{u_h}{\beta}\right)^{-b_h} du_0\, du_h. \end{aligned}$$

The above integral can be done numerically to obtain the autocovariance of X for given parameters $(\alpha, \beta, \rho, \kappa)$, where the areas b_i are functions of ρ and h.

The autocovariance function given above for the exceedance process X is discontinuous at zero, which is a consequence of the mixture construction; the latent trawl process itself is continuous in probability.

As remarked in Section 4.3.2, the trawl process separates the parameters controlling the marginal and dependence properties of the model. However, this is not the case when considering the full hierarchical model, as the parameters α, β and κ in the marginal distribution also influence the autocovariance structure of the process. This is illustrated in figure 5.2, which shows two different autocorrelation functions obtained by varying the parameters α and β. This conflation of marginal and dependence parameters motivates the model in the following subsection.

5.3.4 Marginal transformation model

This subsection considers a modification of the latent trawl model that has the effect of separating the marginal and dependence properties. This modification allows the model to have generalised Pareto distributions with negative shape parameter. The original restriction to positive values of the shape parameter was highlighted as a potential problem in the conclusion of Bortot and Gaetan (2014), where a similar modification was suggested but not explored further.

The model resulting from this modification is also easier to interpret, as the

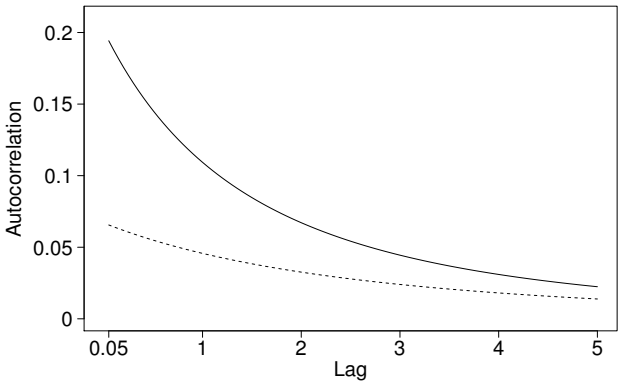

Figure 5.2: Autocorrelation functions of latent trawl models, the solid line corresponds to a model with $\alpha = 4, \beta = 4$, the dashed line to $\alpha = 9, \beta = 1$. Both models set $\rho = 0.2$ and κ such that the probability of an exceedance equals 0.05.

role of each parameter is uniquely defined in terms of controlling either the marginal distribution, probability of exceedance or dependence properties. This also contributes to identifiability of the parameters, in particular we found that the estimation procedure appears to be more efficient for the modified model.

The modified model is derived from the original model in two steps: the first is to fix the parameters α, β of the latent Gamma distribution such that only the parameters associated with the trawl set will influence the trawl process, and thus also the dependence of the exceedances. In the following we will work with $\alpha = \beta = 1$, such that the exceedances have marginal law given by $\mathrm{GPD}(1, 1 + \kappa)$.

The second step is to add an extra layer to the modified model, which uses a standard probability integral transform to give the marginals a $\mathrm{GPD}(\xi, \sigma)$ distribution, specifically

$$Z_j = F^{-1}_{\mathrm{GPD}(\xi,\sigma)}(F_{\mathrm{GPD}(1,1+\kappa)}(X_j)) := g(X_j) \quad \xi \in \mathbb{R}, \sigma > 0.$$

This modified version will be called the marginal transformation (MT) model, it has parameters $(\rho, \kappa, \xi, \sigma)$, where $\xi = 1/\alpha,\ \sigma = \beta/\alpha$, and finite-dimensional

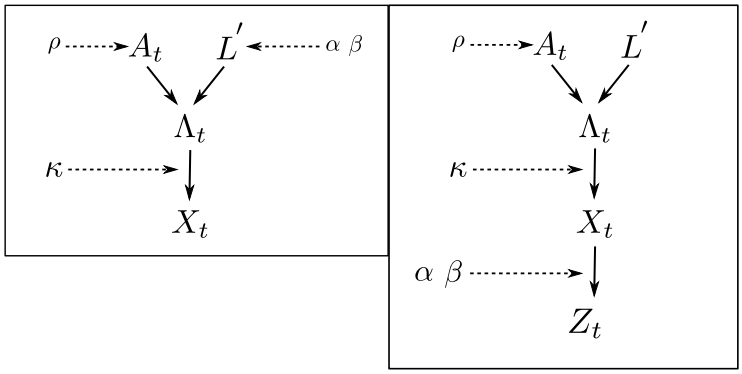

Figure 5.3: Illustration of model structures, original model on the left, transformation model on the right.

densities given by

$$
\begin{aligned}
&f(z_1, \ldots, z_k) = \\
&\begin{cases}
\int_{\mathbb{R}_+^k} \left(\prod_{j=1}^k \left(1 - \exp\{-\kappa \lambda_j\}\right) \right) dF(\lambda_1, \ldots, \lambda_k), & z_j = 0 \\
\int_{\mathbb{R}_+^k} \left(\prod_{j=1}^k J(z_j)\, \lambda_j \exp\{-\lambda_j(\kappa + g^{-1}(z_j))\} \right) dF(\lambda_1, \ldots, \lambda_k), & z_j > 0,
\end{cases} \\
&J(z_j) = \frac{f_{\mathrm{GPD}(\xi,\sigma)}(z_j)}{f_{\mathrm{GPD}(1,1+\kappa)}(g^{-1}(z_j))},
\end{aligned}
$$

with respect to the measure $\mu(dx) = \delta_0(dx) + dx$. This follows by noting that the transformation g maps the event $\{X = 0\}$ to $\{Z = 0\}$, thus leaving the atom at zero unchanged, whereas the transformation of the continuous part on $\{X > 0\}$ introduces a standard Jacobian term.

Figure 5.3 illustrates the structures of the original hierarchical model and the marginal transformation model.

5.4 Model fitting and evaluation

5.4.1 Pairwise likelihood

We now consider parameter estimation for the latent trawl model described above. Given observations $\{Y_j\}$ from a stationary time series, we transform them as in section 5.3.1 to get exceedances $X_j := \max(Y_j - u, 0)$, $j = 1, \ldots, k$. We also assume there are l positive observations $X_{p_1}, \ldots, X_{p_l}$, and m observations $X_{q_1}, \ldots, X_{q_m}$ taking the value zero, which we will call exceedances and

non-exceedances, respectively.

The likelihood of the observations $\{X_j\}$ under the original latent trawl model now follows from (5.8), and can be written as

$$f(x_1,\ldots,x_k) = \int_{\mathbb{R}_+^k} \prod_{r=1}^{l} \lambda_{p_r} \exp\{-(x_{p_r}+\kappa)\lambda_{p_r}\} \prod_{s=1}^{m} (1-\exp\{-\kappa\lambda_{q_s}\})\; dF(\lambda_1,\ldots,\lambda_k),$$

where F is the joint density function of the trawl process observations $\Lambda_1,\ldots,\Lambda_k$. The integrand above can be expanded to obtain a sum involving 2^m number of terms, by first defining $\mathcal{S}_t$ as the collection of subsets of $\{q_1,\ldots,q_m\}$ of size t, and letting $u_r = x_{p_r}+\kappa$, to give

$$\sum_{t=0}^{m}(-1)^t \sum_{\pi_t\in\mathcal{S}_t} \prod_{r=1}^{l} \lambda_{p_r} \exp\left\{-u_r\lambda_{p_r}\right\} \prod_{s_j\in\pi_t} \exp\left\{-\kappa\lambda_{s_j}\right\},$$

with the convention $\prod_{s_j\in\pi_0}(\ldots)=1$. This can be rewritten as

$$\sum_{t=0}^{m}\sum_{\pi_t\in\mathcal{S}_t}(-1)^{t+l}\frac{\partial}{\partial u_1}\cdots\frac{\partial}{\partial u_l}\exp\left\{-\sum_{r=1}^{l}u_r\lambda_{p_r}-\sum_{s_j\in\pi_t}\kappa\lambda_{s_j}\right\}.$$

Using this expression and exchanging integrals and partial derivatives we see that the full likelihood reduces to

$$\begin{aligned}
&\sum_{t=0}^{m}\sum_{\pi_t\in\mathcal{S}_t}(-1)^{t+l}\frac{\partial}{\partial u_1}\cdots\frac{\partial}{\partial u_l}\int_{\mathbb{R}_+^k}\exp\left\{-\sum_{r=1}^{l}u_r\lambda_{p_r}-\sum_{s_j\in\pi_t}\kappa\lambda_{s_j}\right\}dF(\lambda_1,\ldots,\lambda_k)\\
=&\sum_{t=0}^{m}\sum_{\pi_t\in\mathcal{S}_t}(-1)^{t+l}\frac{\partial}{\partial u_1}\cdots\frac{\partial}{\partial u_l}\,E\left[\exp\left\{-\sum_{r=1}^{l}u_r\lambda_{p_r}-\sum_{s_j\in\pi_t}\kappa\lambda_{s_j}\right\}\right],
\end{aligned}$$

where the expectation is with respect to the corresponding variables $\{\Lambda_{p_r}\}$ and $\{\Lambda_{s_j}\}$ determined by π_t.

The expected values in these terms are just joint characteristic functions, as given in Lemma 4.4.4, which involve the parameters of the Lévy seed L and trawl intersection areas; hence the complete likelihood reduces to a sum of partial derivatives of characteristic functions.

The likelihood as given above is not easy to compute in practice, for two reasons. First, it would require multiple numerical partial derivatives to be

performed. Second, the number of non-exceedances m is usually close to the number of observations k; this is because the latent model is defined to have a GP distribution, and to justify this assumption we need to consider a sufficiently high threshold for exceedances, often given by a large percentile of the observations. Now the two first sums in the likelihood above have 2^m terms in total, and hence the likelihood becomes computationally intractable for any reasonable sample size k.

Because of the computational issues associated with the sample size, we consider using a pairwise likelihood approach for model fitting, which is a particular kind of composite likelihood (Varin, 2008; Cox and Reid, 2004; Varin et al., 2011). As stated in Varin (2008), composite likelihood estimators have good properties when the data can be seen as consisting of roughly independent blocks, i.e. the autocorrelation function decays sufficiently fast. Thus using the pairwise likelihood should provide reasonable performance for the latent trawl model with an exponential trawl set.

Given observations $x_1, \ldots, x_k$, the pairwise likelihood f_{PL} takes the form

$$f_{PL}^{\triangle}(x_1, \ldots, x_k) = \prod_{i=1}^{k-1} \prod_{j=i+1}^{\min(i+\triangle, k)} f(x_i, x_j),$$

where $f(\cdot)$ is the original likelihood function and $\triangle$ denotes the maximum separation between observations. For the latent trawl model, each pairwise likelihood term $f(x_i, x_j)$ involves at most 4 terms in the sum above, and so these terms can be evaluated explicitly in terms of the parameters of the trawl process. There are four different cases, as each of x_i and x_j can be an exceedance or not; the explicit forms of $f(x_i, x_j)$ are given in Appendix 5.6.

5.4.2 Conditional tail dependence coefficient

We now consider the extremal dependence structure of our model. A common measure of dependence at high levels is the extremal index defined in Theorem 5.2.4. The extremal index can also be characterised as

$$\theta = \lim_{n \to \infty} P(X_i \leq u_n, 2 \leq i \leq l_n | X_1 > u_n),$$

where u_n is an increasing sequence of thresholds and $l_n = o(n)$ (Ancona-Navarrete and Tawn (2000), following O'Brien (1987)).

The extremal index θ essentially describes the dependence across blocks of observations whose length tends to infinity; it can also be defined as the reciprocal mean cluster length, where a cluster is a collection of exceedances in a block. Thus to estimate the extremal index one has to consider very high-dimensional joint distributions, which makes it analytically intractable in many cases, in particular for our latent trawl model. For the applications in the following subsections we will instead consider simulation-based estimates of the extremal index.

There are also measures of extremal dependence that work on a shorter range of observations than the extremal index. Coles et al. (1999) quantify the dependence between the extreme values of two random variables X_1, X_2 in the upper tail-dependence coefficient, given by *

$$\chi = \lim_{u \to 1} P(F(X_2) > u | F(X_1) > u),$$

where F is the common marginal distribution of X_1, X_2. In other words, χ gives the limiting probability of X_2 exceeding the threshold u given an exceedance X_1, with both variables on a uniform scale. When X_1, X_2 come from a stationary time series, χ can be seen as the probability of observing consecutive exceedances given a single exceedance. To get a broader characterisation of the extremal dependence one can also consider the complete function

$$\chi(u_1, u_2) = P(F(X_2) > u_2 | F(X_1) > u_1),$$

defined on $[0, 1]^2$, where $\chi(u_1, 0) = 1$ and $\chi(0, u_2) = 1 - u_2$.

We would like to use the function $\chi(u_1, u_2)$ and the limiting measure χ to evaluate the dependence between two observations X_1, X_2 from the latent trawl model as a function of the lag $t_2 - t_1$. However, the above definition is based on the assumption that X_1, X_2 are continuously-valued random variables with distribution function F defined on their entire range, and this does not fit with the exceedance framework where the limiting distribution can only be assumed to hold above a threshold. Some extreme value models (see e.g. Ledford and Tawn (1996)) do not consider exceedances separately, but specify the same

*this quantity is sometimes denoted λ_U, and should not be confused with the "coefficient of tail dependence" defined in Ledford and Tawn (1996).

probability density function f for the whole range of observations, and then treat observations below the threshold u as censored, so they have probability

$$\int_0^u f(s)ds.$$

Our model differs in that it models the occurrence of exceedances explicitly, resulting in a distribution for exceedances X only with an atom at zero, and hence the standard definition of χ cannot be applied directly. However, we can construct an analogue to the tail dependence coefficient by conditioning on both exceedances being positive, but care must be taken to ensure that the resulting measure is uniform on the u_2 margin, as shown in the following definition.

Definition 5.4.1: For the latent model as defined in Section 5.3.1, we define the *conditional tail dependence coefficient* as

$$\begin{aligned} \varphi &= \lim_{u_1,u_2\to 1} \varphi(u_1,u_2) \\ &:= \lim_{u_1,u_2\to 1} P(F_{2e}(X_2) > u_2 \,|F_{2e}(X_1) > u_1, X_1 > 0, X_2 > 0), \end{aligned} \tag{5.9}$$

where

$$F_{2e}(x) = P(X_1 \leq x \,|X_1 > 0, X_2 > 0) = P(X_2 \leq x \,|X_1 > 0, X_2 > 0).$$

The conditional tail dependence function $\varphi(u_1, u_2)$ can be calculated explicitly in terms of the parameters of the latent trawl model. Specifically, by conditioning on the latent trawl process we obtain

$$\begin{aligned} &P(F_{2e}(X_1) > u_1, F_{2e}(X_2) > u_2, X_1 > 0, X_2 > 0) \\ =&P(X_1 > F_{2e}^{-1}(u_1), X_2 > F_{2e}^{-1}(u_2), X_1 > 0, X_2 > 0) \\ =&\int \exp\left\{-\lambda_1(F_{2e}^{-1}(u_1)+\kappa)\right\} \exp\left\{-\lambda_2(F_{2e}^{-1}(u_2)+\kappa)\right\} f(\lambda_1,\lambda_2)\, d\lambda_1 d\lambda_2 \\ =&E_{\Lambda_1,\Lambda_2} \exp\left\{-\Lambda_1(F_{2e}^{-1}(u_1)+\kappa) - \Lambda_2(F_{2e}^{-1}(u_2)+\kappa)\right\}. \end{aligned}$$

This expectation can be calculated by using the partition representation of the trawl process, similar to the calculation in Section 5.3.3. Specifically, we

use the partition

$$B_0 = A_1 \setminus A_2,\ B_{0,h} = A_1 \cap A_2,\ B_h = A_2 \setminus A_1,$$

with corresponding slice variables S_i, where A_i is the trawl set at the observation time t_i of the variable X_i. Letting b_i denote the shape parameter of the slice S_i, we have

$$\begin{aligned}
&P(F_{2e}(X_1) > u_1\, F_{2e}(X_2) > u_2, X_1 > 0, X_2 > 0) \\
&= E_{S_1,S_2,S_3}\left[\exp\left\{-S_1\left(F_{2e}^{-1}(u_1) + \kappa\right) - S_2\left(F_{2e}^{-1}(u_1) + F_{2e}^{-1}(u_2) + 2\kappa\right)\right.\right. \\
&\left.\left. -S_3\left(F_{2e}^{-1}(u_2) + \kappa\right)\right\}\right] \\
&= \left(1 + \frac{\kappa + F_{2e}^{-1}(u_1)}{\beta}\right)^{-b_1} \left(1 + \frac{2\kappa + F_{2e}^{-1}(u_1) + F_{2e}^{-1}(u_2)}{\beta}\right)^{-b_2} \left(1 + \frac{\kappa + F_{2e}^{-1}(u_2)}{\beta}\right)^{-b_3}.
\end{aligned}$$

Setting $u_2 = 0$ in this expression and noting that $F_{2e}^{-1}(0) = 0$ shows that

$$\begin{aligned}
&P(F_{2e}(X_1) > u_1, X_1 > 0, X_2 > 0) \\
&= \left(1 + \frac{\kappa + F_{2e}^{-1}(u_1)}{\beta}\right)^{-b_1} \left(1 + \frac{2\kappa + F_{2e}^{-1}(u_1)}{\beta}\right)^{-b_2} \left(1 + \frac{\kappa}{\beta}\right)^{-b_3},
\end{aligned} \tag{5.10}$$

giving the denominator of φ, so that we get

$$\varphi(u_1, u_2) = \left(1 + \frac{F_{2e}^{-1}(u_2)}{\beta + 2\kappa + F_{2e}^{-1}(u_1)}\right)^{-b_2} \left(1 + \frac{F_{2e}^{-1}(u_2)}{\beta + \kappa}\right)^{-b_3}.$$

We can also write

$$F_{2e}(x) = 1 - \frac{P(X_1 > x, X_1 > 0, X_2 > 0)}{P(X_1 > 0, X_2 > 0)},$$

where the probabilities can be derived from (5.10) by the substitution $u_1 \to F_{2e}(x)$, resulting in

$$F_{2e}(x) = 1 - \left(1 + \frac{x}{\beta + 2\kappa}\right)^{-b_2} \left(1 + \frac{x}{\beta + \kappa}\right)^{-b_1}.$$

In particular, we now get the same marginal scaling as the original χ, namely $\varphi(u_1, 0) = 1, \varphi(0, u_2) = 1 - u_2$.

Using this expression, we see that for the original latent trawl model, the limiting coefficient as $u_1, u_2 \to 1$ equals zero, so according to this measure the

	α	β	ρ	κ
Latent trawl	6.33	20.12	0.27	12.18
G-LP	6.43	20.64	0.70	12.25
WP	6.30	19.94	0.78	12.15

Table 5.1: Estimated parameters for Heathrow rainfall data using latent trawl model and latent Markov chain model with G-LP and WP chains.

model is asymptotically independent. Furthermore, the speed of decay to zero increases with the value of b_3; in other words, the larger the intersection of the trawl sets given by X_1, X_2, the slower the model decays to independence as the threshold increases.

5.4.3 Heathrow data

In this subsection we use the latent trawl model to analyse a data set consisting of daily accumulated rainfall amounts at Heathrow (UK) over the years 1949-2012, provided by the UK Meteorological Office (2012). This is the same data set that was used in Section 2.2, but aggregated to daily values. We set the threshold u at the 95 percentile of the original data (8.9mm), resulting in the time series of exceedance values shown in Figure 5.4.

We fitted both the latent trawl model described in Section 5.3.2 and the latent Markov chain model of Bortot and Gaetan (2014), using both the G-LP and Warren process for the Markov chain. The parameter estimation was done using pairwise likelihood as described in Section 5.4.1, with separation parameter $\triangle = 4$; the resulting estimates are shown in Table 5.1. We see that the marginal parameters are similar across the models, which is reasonable as the models all have a marginal GPD$(\alpha, \beta + \kappa)$ distribution and κ controls the marginal exceedance probability.

The parameter ρ controls the latent dependence structure of the models. For the latent trawl process it is the decay parameter of the exponential trawl function, whereas for the latent Markov chains it enters in the autocorrelation function $\phi(h) = \rho^h = \exp(h \log(\rho))$. Using this relation we see from the fitted values of ρ that all three latent processes have similar autocorrelation functions, where the G-LP and WP processes have the fastest and slowest decay, respectively.

Figure 5.5 shows estimates of the extremal index for the latent trawl and latent

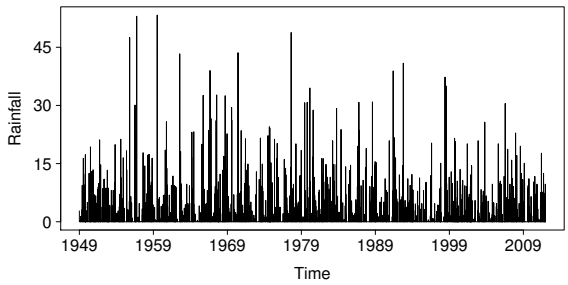

Figure 5.4: Heathrow rainfall exceedances.

Markov chain models. These estimates are based on simulating time-series of length 1,000,000 from the fitted models, and then estimating θ as the inverse cluster length using the R package `evd`, where a cluster is defined as ending when three consecutive exceedances fall below the threshold.

These estimates indicate that the latent trawl and latent Markov chain models both manage to capture the main dependence structure in the extremes. The only visible difference occurs at high thresholds, where the G-LP model appears to underestimate θ, i.e. overestimate the dependence. This discrepancy at high levels could be explained by the fact that the G-LP model is asymptotically dependent, as shown in Bortot and Gaetan (2014), whereas the empirical estimates of θ indicate that the rainfall data is asymptotically independent.

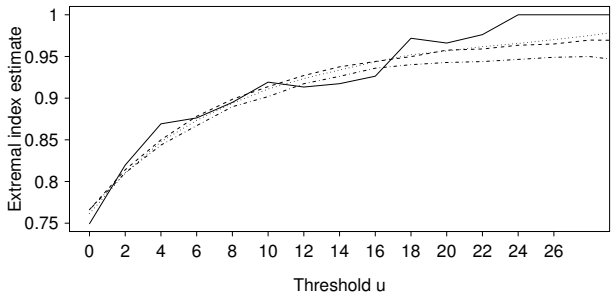

Figure 5.5: Estimated extremal index for empirical rainfall data (solid), latent trawl (dashed), G-LP (dash-dotted) and WP (dotted).

5.4.4 Pollution data

We now consider a second application that illustrates the marginal transformation model introduced in Section 5.3.4. We use this model to analyse measurements of Ozone levels in Bloomsbury, London; specifically, the data gives

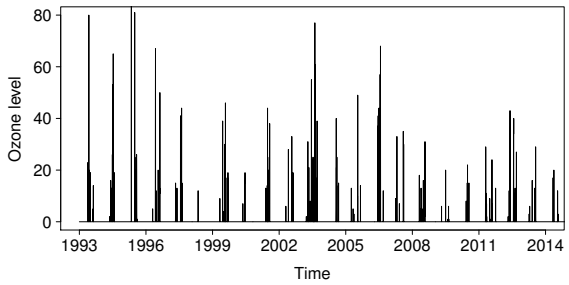

Figure 5.6: Ozone level exceedances

the daily maximum of the 8-hour running mean, measured in units of $\mu g/m^3$ (microgrammes per cubic metre). They are obtained from the Environmental Research Group, King's College London (2015).

Although there is evident seasonality in the original data, the effect diminishes significantly as the threshold increases, and so we do not adjust for seasonality in the extreme values for this application. We chose the threshold u to be the 97 percentile of the original data ($81\mu g/m^3$), resulting in the exceedances shown in Figure 5.6.

Fitting the GPD distribution to the data directly indicates a negative ξ-value, i.e. a finite upper bound, which cannot be captured by the standard model. Thus we use the transformed model described in section 5.3.4, which corresponds to taking the basic hierarchical model (either based on latent trawl or latent Markov chain as in Bortot and Gaetan (2014)) and fixing $\alpha, \beta = 1$, and then transforming the marginals to GPD(ξ, σ).

The models were fitted using pairwise likelihood with $\triangle = 4$. Table 5.2 shows the resulting estimates for the marginal transformed versions of the latent trawl and latent Markov chain models.

Figure 5.7 shows estimates of the extremal index based on simulations of length 1,000,000, obtained by the same method as for the rainfall data above. The estimates of the extremal index indicate that the transformed latent Markov chain model does not accurately capture the extremal dependence structure in this example. In particular, the Warren process model appears to underestimate the extremal dependence (i.e. θ is too high), whereas the opposite is the case for the GL-P model.

These results show that when the marginal parameters α, β are fixed in the latent layer, the latent Markov process model has less flexibility in the depen-

	ξ	σ	ρ	κ
Latent trawl	-0.11	20.73	0.17	32.69
G-LP	-0.04	21.09	0.56	32.19
WP	-0.11	20.74	0.96	32.44

Table 5.2: Estimated parameters for Ozone data using transformed versions of latent trawl model and latent Markov chain model with G-LP and WP chains.

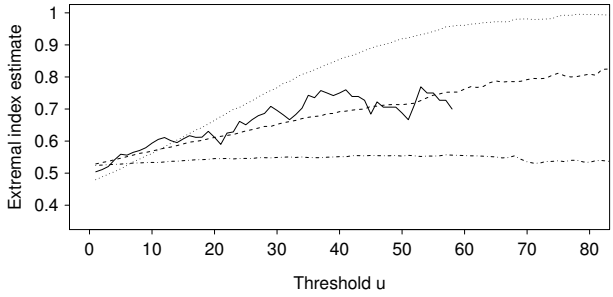

Figure 5.7: Estimated extremal index for empirical Ozone data (solid), and transformed versions of the latent trawl (dashed), G-LP (dash-dotted) and WP (dotted) models.

dence structure than the latent trawl process model. This indicates that in the original hierarchical structure, the marginal parameters also have a strong influence on the dependence structure, and this contributes to the flexibility of the model. The transformed model has the advantage that its parameters are clearly interpretable as contributing to either the dependence or the marginal distribution. In our experience, the parameter estimation procedure for the transformed model also appears to be more reliable, which is not too surprising as the modification of the original model is a form of parameter orthogonalisation.

5.5 Conclusion

In this chapter we have investigated a new model for time series of extremes based on trawl processes. Based on the results in Chapter 4, we constructed an extreme value model that uses the trawl process framework to obtain a flexible dependence structure. This was done by replacing the latent Markov chain in the setup of Bortot and Gaetan (2014) with a trawl process. In contrast with other hierarchical models, this construction has the advantage of preserving the generalised Pareto distribution for the marginals, which is consistent with extreme value theory.

We have also considered a modification to the original model structure that extends the parameter space, allowing for negative shape parameters. To evaluate the extremal dependence we have also developed an adapted version of the tail-dependence coefficient, which can be evaluated analytically for the trawl process model. According to this coefficient the latent trawl model is asymptotically independent.

The original and modified models were used to analyse two environmental time series, and compared with the latent Markov chain models of Bortot and Gaetan (2014). For the application using the original model, the results were very similar in terms of capturing the extremal dependence of the data. The advantage of using the latent trawl process was clearer when using the transformed model, where the trawl-based model captured the extremal dependence better due to the added flexibility in the latent dependence structure.

5.6 Appendix

Latent trawl model pairwise likelihood

The pairwise likelihood can be derived by using lemma 4.4.4 in the likelihood expression. Consider two observations x_1, x_2 corresponding to time points t_1, t_2. Then we can write down the explicit partition of the trawl sets as follows:

$$\{B_1 = A_{t_1} \setminus A_{t_2},\ B_2 = A_{t_1} \cap A_{t_2},\ B_3 = A_{t_1} \cap A_{t_2}\}.$$

The likelihood of x_1, x_2 is given below for each of the four possible cases, where we use the absolute value $|\cdot|$ to denote the Lebesgue measure λ^{Leb}.

1. No exceedances, so $X_1 = X_2 = 0$:

$$\begin{aligned} f_{X_1,X_2}(x_1, x_2) &= E_{\Lambda_1,\Lambda_2}\left[(1 - e^{-\kappa\Lambda_1})(1 - e^{-\kappa\Lambda_2})\right] \\ &= 1 - 2\left(1 + \frac{\kappa}{\beta}\right)^{-\alpha} + \left(1 + \frac{\kappa}{\beta}\right)^{-\alpha\rho|B_1|-\alpha\rho|B_3|}\left(1 + \frac{2\kappa}{\beta}\right)^{-\alpha\rho|B_2|} \end{aligned}$$

   where A is the area of the trawl set, and we have used the substitution $u_i \to iu_i$ in the expression for the characteristic function to obtain $E[e^{-u_i\Lambda_i}]$, which is well defined for all arguments $u_i > 0$.

2. One exceedance, we consider $X_1 > 0, X_2 = 0$, the case $X_1 = 0, X_2 > 0$ is similar. We have

$$\begin{aligned} f_{X_1,X_2}(x_1,x_2) =& E_{\Lambda_1,\Lambda_2}\left[e^{-\kappa\Lambda_1}\Lambda_1 e^{-\Lambda_1 x_1}\left(1-e^{-\kappa\Lambda_2}\right)\right] \\ =& E_{\Lambda_1,\Lambda_2}\left[\Lambda_1 e^{-(\kappa+x_1)\Lambda_1}\right] - E_{\Lambda_1,\Lambda_2}\left[\Lambda_1 e^{-\Lambda_1(\kappa+x_1)-\kappa\Lambda_2}\right] \\ =& -\frac{\partial}{\partial u_1} E_{\Lambda_1,\Lambda_2}\left[e^{-(u_1\Lambda_1+u_2\Lambda_2)}\right]\Big|_{u_1=\kappa+x_1,u_2=0} \\ &+ \frac{\partial}{\partial u_1} E_{\Lambda_1,\Lambda_2}\left[e^{-(u_1\Lambda_1+u_2\Lambda_2)}\right]\Big|_{u_1=\kappa+x_1,u_2=\kappa}. \end{aligned}$$

We have that

$$\begin{aligned} &\frac{\partial}{\partial u_1} E_{\Lambda_1,\Lambda_2}\left[e^{-(u_1\Lambda_1+u_2\Lambda_2)}\right] \\ &= -\frac{\alpha\rho|B_1|}{\beta}\left(1+\frac{u_1}{\beta}\right)^{-\alpha\rho|B_1|-1}\left(1+\frac{u_1+u_2}{\beta}\right)^{-\alpha\rho|B_2|}\left(1+\frac{u_2}{\beta}\right)^{-\alpha\rho|B_3|} \\ &- \left(1+\frac{u_1}{\beta}\right)^{-\alpha\rho|B_1|}\frac{\alpha\rho|B_2|}{\beta}\left(1+\frac{u_1+u_2}{\beta}\right)^{-\alpha\rho|B_2|-1}\left(1+\frac{u_2}{\beta}\right)^{-\alpha\rho|B_3|}. \end{aligned}$$

Evaluating, simplifying and using that $|B_1| + |B_2| = |A|$, this gives

$$\begin{aligned} f_{X_1,X_2}(x_1,x_2) &= \frac{\alpha\rho|A|}{\beta}\left(1+\frac{\kappa+x_1}{\beta}\right)^{-\alpha\rho|A|-1} \\ &- \frac{\alpha\rho}{\beta}\left(1+\frac{\kappa+x_1}{\beta}\right)^{-\alpha\rho|B_1|-1}\left(1+\frac{2\kappa+x_1}{\beta}\right)^{-\alpha\rho|B_2|-1}\left(1+\frac{\kappa}{\beta}\right)^{-\alpha\rho|B_3|} \\ &\times\left(|A|\left(1+\frac{\kappa+x_1}{\beta}\right)+\frac{|B_1|\kappa}{\beta}\right). \end{aligned}$$

3. Two exceedances, so $X_1 = X_2 = 1$:

$$
\begin{aligned}
&E_{\Lambda_1,\Lambda_2}\left[\Lambda_1 e^{-(\kappa+x_1)\Lambda_1}\Lambda_2 e^{-(\kappa+x_2)\Lambda_2}\right] \\
&= \frac{\partial}{\partial u_1}\frac{\partial}{\partial u_2}\, E_{\Lambda_1,\Lambda_2}\left[e^{-u_1\Lambda_1-u_2\Lambda_2}\right]\Big|_{u_1=\kappa+x_1,u_2=\kappa+x_2} \\
&= \frac{\alpha^2\rho^2}{\beta^2}\left(1+\frac{\kappa+x_1}{\beta}\right)^{-\alpha\rho|B_1|-1}\left(1+\frac{2\kappa+x_1+x_2}{\beta}\right)^{-\alpha\rho|B_2|-2} \\
&\times\left(1+\frac{\kappa+x_2}{\beta}\right)^{-\alpha\rho|B_3|-1}\Bigg[|B_1||B_2|\left(1+\frac{2\kappa+x_1+x_2}{\beta}\right)\left(1+\frac{\kappa+x_2}{\beta}\right) \\
&+|B_1||B_3|\left(1+\frac{2\kappa+x_1+x_2}{\beta}\right)^2 \\
&+|B_2|(|B_2|+1/(\alpha\rho))\times\left(1+\frac{\kappa+x_1}{\beta}\right)\left(1+\frac{\kappa+x_2}{\beta}\right) \\
&+|B_2|B_3|\left(1+\frac{\kappa+x_1}{\beta}\right)\left(1+\frac{2\kappa+x_1+x_2}{\beta}\right)\Bigg].
\end{aligned}
$$

6

HIERARCHICAL TRAWL PROCESS FRAMEWORK

In this chapter we use the class of trawl processes to create a general framework for modelling time series data, which includes the extreme value model in Chapter 5 as a special case.

The framework is based on a hierarchical model structure with two levels, analogous to classical state-space models. At the response level, the observations $\{Y_{t_i}\}_{i=1,\dots,n}$ are assumed to be independent random variables, conditional on a latent process (Λ_t). At the latent level, Λ_t is taken to be a trawl process, thus inducing temporal dependence between observations in a flexible way.

The models in this hierarchical framework can be summarised by three main components:

1. The set of observational distributions specifying the conditional likelihood $Y_{t_i}|\Lambda_{t_i}$.
2. The Lévy seed, which specifies the marginal distribution of each trawl process value Λ_{t_i}.
3. The trawl set A, whose shape determines the correlation structure of the process (Λ_t).

These components can be combined in different ways to obtain a wide spectrum of models with a shared structure. The formal construction of these latent trawl models is described in Section 6.1.

In the special case where the trawl process is Gaussian, and the observational distributions belong to the exponential family with mean and variance linked to the trawl process, the resulting latent trawl model can be regarded as a Generalized Linear Mixed Model (GLMM), where the trawl process values enter as correlated random effects.

In comparison to GLMM and more classical state-space models, the latent trawl framework allows for a wider class of model specifications, as it does not impose restrictions on the observational distribution, and allows the latent process to have any infinitely divisible distribution. In particular, this enables us to take advantage of known mixture decompositions in order to obtain a certain marginal distribution for the observations. For example, the latent trawl model in Chapter 5 uses a latent Gamma process with exponential observations to obtain a Generalized Pareto distribution, and in Section 6.9 we use a mean-variance mixture that results in a Variance Gamma distribution.

Due to the assumption of conditionally independent observations, the temporal dependence structure of our hierarchical model is derived from the correlation in the latent trawl process, which is in turn determined by the trawl set A. As described in Section 4.5, we can choose a trawl set that induces a given covariance function for the trawl process, under some mild conditions. Hence the latent trawl framework is very flexible in terms of both distributional and dependence properties, as it refrains from the assumptions of linearity, Gaussianity and Markovianity that are common in the standard time-series and state-space models.

As is usually the case, flexibility in a statistical model comes at a cost, and for the latent trawl model the primary drawback is the difficulty of parameter estimation. This difficulty arises from two primary sources, the first being that the trawl process itself is difficult to estimate, as we do not have a closed form expression for its joint density in terms of the trawl set and Lev́y seed parameters. In fact, this is one of the reasons why we consider models where the trawl process is not directly observed, as this approach allows us to sidestep the calculation of the trawl process likelihood and consider sampling-based estimation approaches. The second difficulty is common to all state-space models, and consists in the need for integrating over all possible values of the latent process to obtain the marginal likelihood $L(\theta|\, y_{t_1}, \ldots, y_{t_n})$, where θ are the parameters. These difficulties are discussed in detail in Sections 6.2 and 6.3, in the context of different approaches for parameter estimation,

A major contribution of this chapter is the development of an MCMC algorithm that enables inference for the general latent trawl model. Section 6.4 describes the basic version of our algorithm, which samples the trawl process by exploiting the partition representation of Section 4.4. In the following

sections, the basic algorithm is extended in order to obtain a more efficient sampling scheme. This extension introduces a type of MCMC moves called transformation moves, which are a special case of reversible-jump MCMC moves (Green, 1995), where the dimension of the parameter space is fixed.

The final part of this chapter introduces two specific models in the latent trawl framework, and demonstrates how to adapt the extended MCMC algorithm for parameter estimation in each case. The model in Section 6.8 is a basic model, which we use to illustrate the latent trawl framework and estimation method, where both the latent trawl and the observations have a Gaussian distribution. In Section 6.9 we consider a model based on a mean-variance mixture, and apply it to a data set of wind energy production.

6.1 General model construction

In this section we describe the formal construction of the general latent trawl process model and introduce the notation that is used throughout this chapter.

The latent trawl model is specified in terms of a sequence of measurements $\{Y_{t_i}\}$ at times $t_1, \ldots, t_n$. To simplify notation we will drop the explicit measurement times t_i in the following, for example taking Y_i to mean Y_{t_i}.

6.1.1 Defining the trawl process

The fundamental model in our framework is based on an underlying trawl process $(\Lambda_t)_{t\in\mathbb{R}_+}$, which can be defined on some probability space $(\Omega, \mathcal{F}, P)$ via the Lévy seed L' and trawl set A, as described in Chapter 4. In particular, we will normalise the cumulant function of the Lévy seed, as described in Lemma 4.3.3, such that the trawl set A does not influence the marginal distribution of $L(A)$. This normalisation justifies using the notation

$$\Lambda_i = L(A) \sim \mathcal{D},$$

where $\mathcal{D}$ is some infinitely divisible distribution, as a shorthand for the specification

$$C\left(u\,;L'\right) = \frac{C\left(u\,;X\right)}{\lambda^{Leb}(A)},$$

where λ^{Leb} denotes the Lebesgue measure, $C(u\,;L')$ is the cumulant function of the Lévy seed L' (c.f. A.1.6), and $C(u\,;X)$ is the cumulant function of a random variable X with distribution $\mathcal{D}$. We assume that the Lévy seed L' and trawl set A are specified in parametric form, such that the cumulant function $C(u\,;X)$ depends on parameters denoted θ_s, and the trawl set A depends on parameters θ_t.

6.1.2 Defining the observational distribution

Having defined the trawl process Λ via the trawl set A and marginal distribution $\mathcal{D}$, we model the collection of observations $\{Y_i\}_{i=1,\dots,n}$ as random variables whose distributions are conditional on the value of Λ at the corresponding time.

More specifically, assume that we have a collection of random variables $\{Y_i\}$ taking values in the Borel space $(\mathcal{Y}, \mathcal{F}_{\mathcal{Y}})$, usually $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and that the distribution of Y_i given Λ_i is specified by a conditional density $f_i(y|\,\lambda)$ with respect to a dominating measure μ_f. This gives the conditional expectation

$$E(g(Y_i)|\,\lambda_i) = \int g(y) f_i(y_i|\,\lambda_i)\,\mu_f(dy),$$

for any measurable function g on $\mathcal{Y}$.

The density $f_i(\cdot|\,\lambda_i)$ is assumed to take a known parametric form, with an associated vector of parameters θ_y; we will use the notations $f_i(y_i|\,\lambda)$ and $f_i(y_i|\,\lambda_i, \theta_y)$ interchangeably.

Remark 6.1.1: The above construction explicitly allows for the case where the conditional distribution of Y_i depends on the time index i, meaning that the distribution of Y can change over time to capture seasonality or trends.

An example of a non-trivial density is given by the model in Chapter 5, where the distribution has an atom at zero, with density given by

$$f_i(y_i\,|\,\lambda_i) = \begin{cases} 1 - e^{-\kappa\lambda_i}, & y_i = 0, \\ \lambda_i e^{-\lambda_i(\kappa + y_i)}, & y_i > 0, \end{cases}$$

with respect to $\mu_f(dy) = \delta_0(dy) + \lambda^{Leb}(dy)$; in this case we have $\theta_y = \kappa$.

Given the densities $\{f_i(\cdot|\,\lambda_i)\}$, we now define a conditional distribution on the product space $(\mathcal{Y}^n, \mathcal{F}_{\mathcal{Y}}^n)$ via conditional independence, i.e.

$$\begin{aligned}&E\left[g_1(Y_1)g_2(Y_2)\ldots g_n(Y_n)\,|\,\lambda_1,\ldots,\lambda_n\right]\\ =&\prod_{i=1}^{n} E[g(Y_i)|\,\lambda_i] = \prod_{i=1}^{n}\int g_i(y_i)f_i(y_i|\,\lambda_i)\,\mu_f(dy_i).\end{aligned}$$

We can also obtain a marginal probability measure on $(\mathcal{Y}^n, \mathcal{F}_{\mathcal{Y}}^n)$ by integrating with respect to the joint distribution of the trawl values $\{\Lambda_i\}$, specifically

$$\begin{aligned}&P_Y^n(B_1,\ldots,B_n)\\ =&\int_{\mathbb{R}^n}\left[\prod_{i=1}^{n}\int_{B_i} f_i(y_i|\,\lambda_i)\,\mu_f(dy_i)\right]P_{\mathbf{\Lambda}}(d\lambda_1,\ldots,d\lambda_n),\end{aligned}\tag{6.1}$$

where $P_{\mathbf{\Lambda}}$ is the joint distribution function of $\Lambda_1,\ldots,\Lambda_n$. This distribution depends on the parameters θ_s that specify the Lévy seed L', and the parameters θ_t, which specify the trawl set. This completes the specification of the general latent trawl model, which is summarised in the following definition.

Definition 6.1.2: The general latent trawl model has three main parts:

1. The trawl set A, with parameters θ_t.
2. The distribution $\mathcal{D}$ that specifies the normalised Lévy seed L', with parameters θ_s.
3. The densities $\{f_i(y_i|\,\lambda_i)\}$ and measure μ_f that specify the observational distributions, with parameters θ_y.

Example 6.1.3: To illustrate the construction described above, we consider the basic model that will be used in Section 6.8, which is given by

$$\begin{aligned}&Y_i|\Lambda_i \sim \mathcal{N}(\Lambda_i, \sigma_y^2),\\ &\Lambda_i = L(A) \sim \mathcal{N}(\mu, \sigma^2),\\ &A = \{(s,\xi) : -\infty < s \le 0,\ 0 \le \xi \le d_{exp}(s;\rho)\},\end{aligned}$$

where the function $d_{exp}(s;\rho) = \exp\{\rho s\}$ defines the exponential trawl with

decay parameter ρ. This gives a latent trawl model as described above, with

$$\theta_y = \sigma_y^2, \quad \theta_s = (\mu, \sigma^2), \quad \theta_t = \rho.$$

6.2 Likelihood-based estimation approaches

In this section we discuss issues associated with using maximum-likelihood estimation for the latent trawl model. The main purpose of this section is to show that calculating the likelihood and performing the numerical maximisation involves significant challenges, and thus we will not pursue this approach further in the following, but instead base our inference on the extended Gibbs sampling algorithm presented in Section 6.7.

6.2.1 Likelihood function

The likelihood function of the general trawl model can be obtained from (6.1) by using Fubini's theorem to exchange the integrals, giving

$$\begin{aligned} &P_Y^n(B_1, \ldots, B_n) \\ &= \int_{B_1} \cdots \int_{B_n} \int_{\mathbb{R}^n} \left(\prod_{i=1}^n f_i(y_i |\, \lambda_i, \theta_y) \right) P_{\mathbf{\Lambda}}(d\lambda_1, \ldots, d\lambda_n |\, \theta_s, \theta_t)\, \mu_f(dy_1) \ldots \mu_f(dy_n), \end{aligned}$$

and hence

$$L(\theta_y, \theta_s, \theta_t \,|\, y_1, \ldots, y_n) = \int_{\mathbb{R}^n} \left(\prod_{i=1}^n f_i(y_i |\, \lambda_i, \theta_y) \right) P_{\mathbf{\Lambda}}(d\lambda_1, \ldots, d\lambda_n |\, \theta_s, \theta_t), \tag{6.2}$$

where $\theta_y, \theta_s, \theta_t$ are as in Definition 6.1.2.

The above likelihood function cannot be calculated directly, for several reasons. First, the likelihood involves an n-dimensional integral, where n is the number of observations, which quickly becomes computationally intractable for any realistic application. Second, the integral in the likelihood above depends on the joint distribution of the trawl process observations $\Lambda_1, \ldots, \Lambda_n$, which is not available in analytical form for the general trawl process.

In some special cases, the likelihood integral reduces to an expression involving

the characteristic function of the trawl process, as was the case for the model in Chapter 5. As shown in Lemma 4.4.4, this characteristic function can be obtained in analytic form as a function of θ_s, θ_t, by using the partition representation. However, such reductions to the characteristic function will not occur for most models in the general framework. In the general case we would require an n-dimensional Fourier transform to obtain the joint density of the trawl process observations, which compounds the computational difficulty.

6.2.2 Pairwise likelihood

To avoid having to perform high-dimensional integrations, we could follow the example in the previous chapter and work with the pairwise likelihood, which takes the form

$$PL^{\triangle}(\theta|\, y_1, \ldots, y_n) = \prod_{i=1}^{n-1} \prod_{j=i+1}^{\min(i+\triangle, n)} L(\theta \,|\, y_i, y_j),$$

for some fixed integer $\triangle$, where $L(\theta \,|\, y_i, y_j)$ is the likelihood of the two observations y_i, y_j only. From the discussion above, we know that the likelihood of y_i, y_j takes the form of a 2-dimensional integral with respect to the unknown joint density of Λ_i, Λ_j, and if we use Fourier transformation to obtain the joint density from the characteristic function, this still leaves us with a 4-dimensional integral for each term $L(\theta \,|\, y_i, y_j)$ above.

An alternative to the Fourier transform approach is to calculate the integral in Fourier space, via an application of Plancherel's theorem. This approach is often used in the context of option pricing, see for example Lewis (2001). Under certain regularity conditions, we can write the likelihood integral as

$$\begin{aligned} L(\theta \,|\, y_i, y_j) &= \int_{\mathbb{R}^2} f_i(y_i|\, \lambda_i) f_j(y_j|\, \lambda_j) P_{\mathbf{\Lambda}}(d\lambda_i, d\lambda_j) \\ &= \int_{\mathbb{R}^2} \hat{f}_i(\cdot|\, \lambda_i)(u_i) \hat{f}_j(\cdot|\, \lambda_j)(u_j) \varphi_{\Lambda_i, \Lambda_j}(u_i, u_j)\, du_i du_j, \end{aligned}$$

where $\varphi_{\Lambda_i, \Lambda_j}$ is the pairwise characteristic function of Λ_i, Λ_j, and $\hat{f}$ is the Fourier transform of f.

This Fourier-space approach solves the problem of the unknown joint densities, as it reduces the trawl process involvement to the pairwise characteristic

function, which has a simple form involving the Lévy seed and the intersection of two trawl sets. There may be special cases where the above integral is analytically available, but in the general case we have to resort to some form of numerical approximation, resulting in an approximate pairwise term $\tilde{L}(\theta|\, y_i, y_j)$. In the context of calculating the full pairwise likelihood, this procedure becomes computationally intensive, as we require a separate integral for each pair of observation times. A more serious problem is that the total pairwise likelihood at some value θ will inherit errors from multiple (of order $n\triangle$) numerical integrals, such that the overall error may be quite large. This error may in turn have an adverse effect on the numerical optimisation procedure that is used to find the maximum likelihood estimator.

In order to test the feasibility of the approach described above, we performed a simulation study using the simple model described in Section 6.8, where we calculated the pairwise likelihood integrals by brute force numerical integration, and used the resulting estimates to calculate the pairwise likelihood function $PL^{\triangle}(\theta|\, y_1, \ldots, y_n)$, which was then used in a numerical optimisation procedure. Using this approach, we found that in many cases the likelihood optimisation procedure did not converge properly. We tried various remedies, such as increasing the required accuracy of the numerical optimisation routine, and employing a stochastic optimisation procedure (Robbins and Monro, 1951), but these did not significantly improve the convergence properties. As a comparison, we performed the same optimisation using the analytical form of the pairwise likelihood, which happens to be available for this simple model; in this case the convergence properties improved significantly. Based on these results, it appears that the lack of convergence in the likelihood optimisation are mainly due to the errors in the numerical integration.

The above discussion suggests that for the general latent trawl model, the combination of pairwise likelihood and brute-force numerical integration in the Fourier space leads to serious issues with the likelihood optimisation procedure. One could try using more advanced methods for estimating the pairwise likelihood integrals, for example using spline basis functions, where the approximation of each likelihood $L(\theta|\, y_i, y_j)$ would be fixed across different values of θ, as opposed to having to re-evaluate a numerical integral and introducing new sources of error. However, this approach would require a large amount of additional effort in order to obtain an approximate pairwise likelihood function, and may introduce additional uncertainty due to tuning

parameters, in addition to the loss of efficiency from using the pairwise instead of the full likelihood.

6.3 MCMC estimation methods

The discussion above suggests that it is prohibitively difficult to calculate the likelihood with the precision required to perform maximum likelihood estimation, especially when seeking a method that can be used for the general class of latent trawl models. These difficulties, together with the latent structure of the model, suggest the use of sampling-based estimation methods, more specifically MCMC-type approaches.

6.3.1 Monte Carlo within Metropolis

A natural approach to approximating the likelihood given by (6.2) is to replace the integral in the likelihood by an MCMC approximation, using i.i.d samples from the entire trawl process for given values of the parameters. Then we are left with an MCMC approximation of the likelihood function, which can be optimised numerically to obtain parameter estimates, or embedded in a Bayesian estimation procedure, where a standard Metropolis-Hastings(MH) move is used to sample from the posterior. The latter procedure is known as *Monte Carlo within Metropolis* (MCWM) in Andrieu and Roberts (2009), and there are several variations of this basic method that can improve the efficiency of the sampler.

We attempted to use the basic MCWM algorithm to approximate the likelihood, but found that the likelihood estimate obtained by sampling independent trawl processes had a prohibitively large variance. This in turn means that the overall MH algorithm does not sample accurately from the posterior distribution of the parameters. The large variance issue is not very surprising, given the large size of the space of trawl processes, as seen from the slice representation.

The variance of the MCMC estimator could potentially be reduced by using a proposal distribution that is closer to the posterior distribution of the trawl process. This could be done by considering one of the large variety of methods for sampling from latent processes in state-space models, see e.g. Andrieu

et al. (2010). However, these methods are not easily applicable to the latent trawl model, as they usually assume the latent process is Markov, which is not the case for the trawl process; using these approaches would require the construction of specialised proposal densities for the one-step transition probabilities of the trawl process.

6.3.2 Data augmentation

Another common approach for sampling from latent models is to use data augmentation, which alternates between sampling the latent data, in our case the trawl process $\{\Lambda_i\}$, and sampling the parameters from the augmented posterior distribution with density

$$\pi\left(\theta|\{\Lambda_i\},\{Y_i\}\right),$$

where θ represents the model parameters (Tanner and Wong, 1987). In terms of the computational effort required for a single update, data augmentation is much more efficient than MCWM, as it does not require multiple new samples of the trawl process for each parameter update. On the other hand, it is well known that the resulting samplers often exhibit poor mixing, and our latent trawl models are no exception to this problem. However, the basic algorithm based on data augmentation can be extended in order to produce a more efficient sampler, which is the topic of Section 6.6. To set the stage for this improved sampler, the next section describes the data augmentation approach in detail as it applies to latent trawl process models.

6.4 Gibbs sampling algorithm

In this section we describe a Gibbs sampling algorithm for sampling from the joint posterior distribution of the parameters and trawl process values. In particular, this means that we are working in the Bayesian framework, where we treat the parameters as random and assign appropriate prior distributions to them.

A more classical alternative would be to use the data augmentation approach to construct an EM-type algorithm, which alternates between sampling the trawl process, as described in the following, and maximising the resulting

likelihood approximation. Due to the variability in the trawl sampling step, however, this approach is likely to suffer from problems with the optimisation step, similar to the likelihood approximation methods described in the previous section. In contrast, the Bayesian approach does not require numerical optimisation, but allows for a sampling-based procedure that accounts for the joint uncertainty in the parameters and the latent trawl process given the data. Having samples from the full posterior distribution is especially useful for the models in the latent trawl framework, as we can study the dependencies between parameters at different layers of the hierarchy, as well as their relation to the latent trawl process.

6.4.1 Truncated partition representation

We now describe the specific algorithm for sampling from the augmented posterior distribution of the parameters. In order to sample from the trawl process $\Lambda_1, \ldots, \Lambda_n$, we will use the partition representation that was described in Section 4.4, which decomposes the trawl process into a set of independent random variables S_i, called slices. Because the slices uniquely determine the values of the trawl process, we may replace the trawl process by the slices in the augmented data. This reduces the problem of sampling the entire path of the trawl process to sampling the individual slices. As explained in Section 4.4, the number of slices is $2^n - 1$ for a general trawl set, and $n(n+1)/2$ for a monotonic trawl set, which implies that merely keeping the slices in memory becomes a large computational burden as n grows.

To avoid the computational issues arising from the full slice representation, we introduce an approximation based on only using the slices that are deemed most important. This approximation is motivated by the discussion in Section 4.5, where the intersection of trawl sets $A_{t_{i-\tau}}$ and A_{t_i}, say, represents the contribution of an event at time $t_{i-\tau}$ to the value of the process at time t_i. Using a monotonic trawl then corresponds to the assumption that this contribution decreases with time $\tau > 0$.

To obtain our approximation, we will restrict the trawl set to be monotonic, and make the additional assumption that for some reasonably small τ, the contribution from time $t_{i-\tau}$ is essentially zero, i.e. that the dependence between trawl process values decays quickly as a function of the lag $t_i - t_{i-\tau}$. For most natural processes this is a reasonable assumption, with the exception

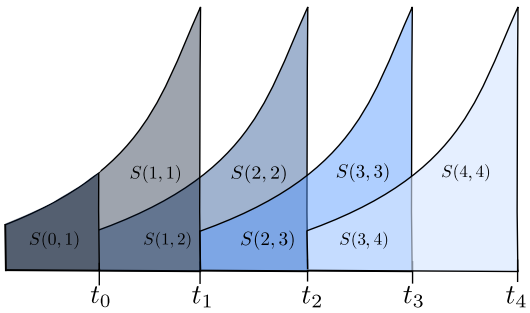

Figure 6.1: Truncated slice grid of order $\tau = 2$ for $n = 4$.

of those exhibiting long memory or having high frequency observations.

The assumption of rapid decay suggests that any slices generated by intersections of trawl sets at least τ lags apart can be omitted from the slice representation in (4.8), resulting in a truncated grid of order τn, as illustrated in Figure 6.1 for $\tau = 2$.

In some cases we may even assume a priori that observations are independent when separated by more than lag τ. This corresponds to a trawl set of bounded width, and hence we can get a sparse collection of slices without imposing a truncated grid approximation.

Remark 6.4.1: It can be shown that the truncated slice grid gives rise to a state-space representation of the trawl process of dimension 2τ, where the vector components are given by slices. Based on this representation, one could consider using particle MCMC-type methods, see e.g. Andrieu et al. (2010). These methods are known to be highly efficient for suitable problems, but are computationally intensive and may require non-trivial adaptations to particular models, and we do not pursue this approach further here.

6.4.2 Sampling slices

Having defined the truncated grid, we now consider how to sample the slices. As described in Section 4.4, we can define the distribution of each slice via its cumulant function, which takes the form

$$C(u, S_j) = \lambda^{Leb}(B_j)C(u, L'),$$

where L' is the Lévy seed that represents an infinitely divisible distribution, and $\lambda^{Leb}(B_j)$ is the Lebesgue measure of the slice intersection set. By an abuse of notation we will let S_i and the term "slice" denote both the intersection set and the corresponding random variable, which allows us to write $C(u, S_j) = \lambda^{Leb}(S_j)C(u, L')$.

As described in Section 6.1.1, we will work with a normalised trawl process, which implies that if the marginal law of the trawl process is defined by a cumulant function $C(u\,; X)$, then we have

$$C(u, S_j) = \lambda^{Leb}(S_j)\frac{C(u\,; X)}{\lambda^{Leb}(A)}. \tag{6.3}$$

In the following we assume that we can sample from the distribution with cumulant function $\alpha C(u, X)$ for any $\alpha \in \mathbb{R}_+$. Hence we can use equation (6.3) to simulate from any slice for which the area $\lambda^{Leb}(S_j)$ is known.

6.4.3 Updating scheme

We now return to the overall problem of sampling from the augmented posterior distribution

$$\pi(\{S_j\}, \theta \,|\, \{Y_i\}).$$

To simplify notation, we let $\mathcal{O}$ denote the observations $Y_1, \ldots, Y_n$, $\mathcal{S}$ denote the collection of slices, and define the set $\mathcal{T}_i$ to be the slices that make up the trawl at time i, i.e.

$$j \in \mathcal{T}_i \Leftrightarrow S_j \subset A_i,$$

where A_i is the trawl set for observation i. Furthermore, we use the decomposition of the parameter vector into $\theta_y, \theta_s, \theta_t$, as introduced in Definition 6.1.2.

With this notation to hand, the augmented posterior density can be decomposed as follows:

$$\begin{aligned}\pi\left(\mathcal{S}, \theta | \,\mathcal{O}\right) &\propto p\left(\mathcal{O}, \mathcal{S}, \theta\right) = p(\mathcal{O} \,|\, \mathcal{S}, \theta_y)\, p(\mathcal{S} \,|\, \theta_s, \theta_t)\, \pi(\theta_y, \theta_s, \theta_t) \\ &= \left[\prod_{i=1}^{n} f_i\left(Y_i \,\middle|\, \sum_{j \in \mathcal{T}_i} S_j \,, \theta_y\right)\right] \left[\prod_{j=1}^{N_s} p_j(S_j \,|\, \theta_s, \theta_t)\right] \pi(\theta_y)\pi(\theta_s)\pi(\theta_t),\end{aligned}$$

where N_s is the number of slices and $\pi(\cdot)$ is the prior density of the parameters.

To sample from this posterior, we use a Gibbs sampling scheme that updates each of the variables

$$(S_1, \ldots, S_{N_s}, \theta_t, \theta_s, \theta_y),$$

in turn, by sampling from their conditional distribution given the other variables and the data $\mathcal{O}$.

For the observational parameters θ_y, we get the conditional density

$$p\left(\theta_y \,|\, \mathcal{O}, \mathcal{S}, \theta_s, \theta_t\right) \propto \left[\prod_{i=1}^{n} f_i\left(Y_i \,\middle|\, \sum_{j \in \mathcal{T}_i} S_j \,, \theta_y\right)\right] \pi(\theta_y).$$

This corresponds to seeing n independent observations from the conditional distributions of Y_i, given the trawl set value resulting from the slices in $\mathcal{T}_i$. A standard approach is to target this density by using a Metropolis-Hastings (MH) step with a random walk proposal. Alternatively, the distribution given by f_i may admit a conjugate prior $\pi(\theta_y)$, which allows for direct sampling of θ_y from the conditional posterior.

For the parameters θ_s and θ_t we get similar forms of the conditional densities:

$$\begin{aligned} p\left(\theta_s |\, \mathcal{O}, \mathcal{S}, \theta_y, \theta_t\right) &\propto p(\mathcal{S} |\, \theta_s, \theta_t)\pi(\theta_s) \\ p\left(\theta_t |\, \mathcal{O}, \mathcal{S}, \theta_y, \theta_s\right) &\propto p(\mathcal{S} |\, \theta_s, \theta_t)\pi(\theta_t), \end{aligned}$$

which only depend on the observations indirectly through the current value of the slices. As for θ_y, we can sample from these conditional posteriors using either a random walk MH step, or by using conjugate priors when these are available.

Finally, the conditional density of a single slice is given by the following, where the dependence on the observations is only through the trawl process values that involve the slice:

$$p\left(S_i |\, \{S_j\}_{j \neq i}, \mathcal{O}, \theta\right) \propto \left[\prod_{k : i \in \mathcal{T}_k} f_k\left(Y_k |\, \sum_{j \in \mathcal{T}_k} S_j, \theta_y\right)\right] p(S_i |\, \theta_s, \theta_t).$$

This conditional density will not usually take a known parametric form, and we will use a random-walk MH step to sample from it.

6.5 Strategies for improving mixing

It is well known that using Gibbs-style sampling for augmented data usually leads to a high autocorrelation in the posterior samples, making it inefficient for parameter estimation. In particular, Liu et al. (1994) analyse the covariance structure of Gibbs samplers for augmented data, and show that a chain which alternates between sampling $x|y$ and $y|x$ will have autocovariance bounded by the correlation between x and y.

Since the likelihood of the latent trawl model involves the density of $\mathcal{S}|\theta$, we expect the variables $\{S_i\}$ and the parameters of the trawl process to be strongly correlated in the posterior distribution, leading to poor mixing rates of the sampled Markov chain. Essentially, given starting values of the trawl process parameter θ_s, θ_t, the slice updates cannot move too far away from the expected values given θ_s, θ_t without significantly decreasing the joint slice likelihood $p(\mathcal{S}|\,\theta_s, \theta_t)$. Conversely, the parameters θ_s, θ_t have conditional distributions that only depend on the observations indirectly through the slices, and thus the parameter updates will be constrained by the current slice values.

The discussion above shows that the basic augmented Gibbs sampler is likely to be highly inefficient, due to the strong dependence between the slices and the parameters θ_s, θ_t, which is an inevitable consequence of the trawl process parametrisation. This problem could be remedied by using a sampling scheme that updates the slices and their parameters simultaneously. However, the simplest type of moves, which propose new points independently of the target density $p(\mathcal{S}, \theta|\,\mathcal{O})$, are not likely to work well, as the slice space has very large dimension.

To see what types of moves may be useful for updating the slices and their parameters, we consider a concrete example where the trawl process observations have a Gaussian marginal distribution with mean μ and variance σ^2. We assume that the current values of $\{S_j\}$ and μ, σ^2 are samples from the posterior distribution, which takes the form

$$\begin{aligned}&\pi(\mu, \sigma^2, \{S_j\}|\,\theta_y, \theta_t, \{y_i\}) \\ &\propto \left[\prod_{i=1}^{n} f_i\left(y_i|\sum_{j\in\mathcal{T}_i} S_j, \theta_y\right)\right]\left[\prod_{j=1}^{N_s}\mathcal{N}(S_j;\,\mu_j, \sigma_j^2)\right]\pi(\theta_y)\pi(\mu, \sigma^2)\pi(\theta_t),\end{aligned}$$

where the mean μ_j and variance σ_j^2 of the slice distribution are given by scaling

μ and σ^2 by the factor $c_j = \lambda^{Leb}(S_j)/\lambda^{Leb}(A)$, which is a function of the trawl set parameters θ_t.

Because the number of slices is of order at least n^2, this posterior is likely to be dominated by the second term, which is a joint product of Gaussian densities. Each of these slice densities can be visualised as a function in $\mathbb{R}^2$, which takes the form $h(S_j, \mu_j) = p_{\mathcal{N}}(S_j;\, \mu_j, \sigma_j^2)$, where $p_{\mathcal{N}}$ is the Gaussian density. This function has a ridge along the diagonal where $S_j = \mu_j$, and hence the current values of (S_j, μ_j) are likely to lie along this ridge. In order to avoid proposing new slice values at points of low density, which leads to a high rate of rejected moves, we should propose values that stay close to the ridge in each subspace given by (S_j, μ_j). One way of achieving this is to perform a so-called *scaling move*, where we multiply all slices and means by some common value α, which can be visualised as moving along diagonal lines in the plane.

The scaling move described above is a particular example of a large class of MCMC moves, hereafter called *transformation moves*, which generalise the two-step construction of the transition kernel implicit in Metropolis-Hastings moves. As will be shown in the following section, transformation moves are a special case of the reversible jump MCMC framework of Green (1995). A similar construction was considered in Liu and Sabatti (2000), where they constructed a so-called Generalized Gibbs sampler.

Returning to the example above, where we consider sampling in the space $X = (S_j, \mu_j)$, it is clear that the most efficient moves are those that travel along the diagonal, which is a curve of high posterior density in X. In contrast, standard Gibbs sampling moves always travel parallel to the coordinate axes, and thus they are most efficient when the joint density of a general point $x = (x_1, \ldots, x_d)$ factorises, meaning that the corresponding parameters are independent. By considering transformation moves that act on the entire space X, instead of individual coordinate moves, we allow for the design of much more efficient samplers in the presence of dependence. In particular, transformation moves can be used to obtain the same effects as reparametrisation methods.

In the following we will develop a more efficient MCMC sampling algorithm for the latent trawl model which incorporates transformation moves. This algorithm involves steps that are similar to standard Metropolis-Hastings moves, but with a subtly different acceptance probability. In order to understand

how these moves come about, we first present some theoretical background for these moves, which is the topic of the following section.

6.6 Transformation moves

In this section we introduce the class of transformation moves, which will be used to create the extended MCMC algorithm used in the following.

The basic idea behind transformation moves is to consider points x in some space $X \subset \mathbb{R}^d$, which can represent parameters or augmented variables, and shift these points by sampling a deterministic transformation $g_u : X \to X$ from a group of transformations, i.e.

$$\begin{aligned} u &\sim q(\cdot), \quad u \in \mathcal{U} \\ x &\to g_u(x), \end{aligned}$$

where $q(\cdot)$ is a density on the space $\mathcal{U}$ of transformation indices. By accepting these transformation moves with a particular probability, we can construct a transition kernel that leaves some specified target distribution π invariant, and hence these moves can be embedded in a wider MCMC sampling scheme. As will be seen, this formalism allows us to consider transformations that shift a large number of points $(x_1, \ldots, x_d) = x$ in a single move, which has the potential for significantly improving the efficiency of the MCMC algorithm.

The theory behind these moves is presented in Section 6.6.1, following results in Green (1995) and Kou et al. (2005). Section 6.6.2 gives examples of specific transformation moves, in particular introducing the so-called *quantile move*, which to the best of our knowledge has not been mentioned in the literature before.

6.6.1 Theoretical setup

Consider a space $X \subset \mathbb{R}^d$, and a set of transformations $\{g_u\}$ indexed by a set $\mathcal{U} \subset \mathbb{R}^k$, where each g_u maps X to itself. We make the following assumptions:

1. The transformations $\{g_u\}_{u \in \mathcal{U}}$ are reversible, meaning that for each $u \in \mathcal{U}$ there exists a unique $u^{-1} \in \mathcal{U}$ such that $g_{u^{-1}}$ is the inverse transforma-

tion, i.e.

$$g_{u^{-1}}(g_u(x)) = x, \quad \forall x \in X.$$

2. All functions $g_u(\cdot)$, $u \in \mathcal{U}$, are differentiable on X.

3. The function $\tau : \mathcal{U} \to \mathcal{U}$, $\tau(u) = u^{-1}$ is differentiable on $\mathcal{U}$.

In Theorem 6.6.2 we will show how these transformation moves can be used as proposals in an MCMC sampling scheme for targeting a given probability distribution, by accepting the transformation with a given probability. To motivate this theorem and illustrate the transformation move setup, we first consider some concrete examples.

Example 6.6.1: A simple example of such transformations is obtained by letting $\mathcal{U} = \mathbb{R}$, $u^{-1} = -u$, proposing $u \sim \mathcal{N}(0, \sigma^2)$, and using

$$g_u(x_1, \ldots, x_d) = (x_1 + u, x_2, \ldots, x_d),$$

corresponding to a standard Metropolis-within-Gibbs move for the first coordinate, with a random walk proposal.

Another example is the global scaling move, where $\mathcal{U} = \mathbb{R}_+$, $u^{-1} = 1/u$, and

$$g_u(x_1, \ldots, x_d) = (ux_1, ux_2, \ldots ux_d),$$

where u is sampled from an arbitrary distribution on $\mathbb{R}_+$. This example illustrates the separation of the random proposal $u \in \mathbb{R}_+$ from the transformation acting on $\mathbb{R}^d$, which means we can construct moves that shift all points $x_1, \ldots, x_n$, while using proposal distributions defined on a lower-dimensional space.

We now consider using the transformations $\{g_u\}_{u \in \mathcal{U}}$ to create a transition kernel $K(\cdot, \cdot)$ on $X \times X$ with some invariant distribution π. This can be done by applying the following theorem, which describes a special subclass of the reversible MCMC jumps introduced in Green (1995).

Theorem 6.6.2: Fix a probability space $(\Omega, \mathcal{F}, P)$, a sampling space $X \subset \mathbb{R}^d$, a set of indices $\mathcal{U} \subset \mathbb{R}^k$ and a transformation set $\{g_u\}_{u \in \mathcal{U}}$, such that the conditions $(1) - (3)$ above hold. Let Q be a probability distribution on $\mathcal{U}$ with

density q (with respect to the Lebesgue measure on $\mathcal{U}$), and let x be a point in X with $x \sim \pi$, where π is a density on X.

Then the transition kernel that corresponds to sampling $u \sim q$ and accepting the move $x \to g_u(x)$ with probability

$$a(x,u) = \min\left(1, \frac{\pi(g_u(x))q(u^{-1})\,|\partial\tau/\partial u||J_{g_u}(x)|}{\pi(x)q(u)}\right), \tag{6.4}$$

with $J_{g_u}(x) = \partial g_u(x)/\partial x$, has π as invariant distribution.

Remark 6.6.3: The acceptance probability above looks like a standard MH acceptance probability, except for the terms $|\partial\tau/\partial u||J_{g_u}(x)|$ in the numerator. Intuitively speaking, the Jacobian comes into play because we sample $u \in \mathcal{U} \subset \mathbb{R}^k$ from the density $q(\cdot)$ with respect to the Lebesgue measure on $\mathcal{U}$, instead of sampling x' from a density $f_x(\cdot)$ with respect to the Lebesgue measure on $\mathbb{R}^d$. The correction term $|\partial\tau/\partial u|$ comes from using the ratio $q(u^{-1})/q(u)$ instead of $f_{x'}(x)/f_x(x')$.

Proof. The theorem follows by considering a special case of the reversible jump moves described in Green (1995) and Green (2002), where we work with transformations on the state space X, as opposed to jumps between parameter spaces of different dimension.

In particular, the reversible jump moves are described as transformations given by a deterministic, differentiable bijection $\mathcal{M} : (x, u_1) \to (x', u_2)$, with a differentiable inverse move $\mathcal{M}' : (x', u_2) \to (x, u_1)$, (i.e. $\mathcal{M}$ is a diffeomorphism). Here x and x' represent the points in our sampling space X, and u_1, u_2 are the random innovations that determine the particular transformation. In the general case, u_i has dimension m_i and x, x' have dimension n_1, n_2, respectively, and to match dimensions one requires $n_1 + m_1 = n_2 + m_2$.

For the transformation moves described in our theorem, we have that the map $\mathcal{M}(\cdot, u_1)$ is given by $g_{u_1}(\cdot)$, where the inverse move is generated by letting $u_2 = u_1^{-1}$, so that $\mathcal{M}'(\cdot, u_2) = \mathcal{M}(\cdot, u_1^{-1}) = g_{u_1^{-1}}(\cdot)$. In particular, the dimensions are fixed throughout, so $n_1 = n_2$, $m_1 = m_2$.

In the reversible jump framework, the innovations u_1, u_2 are sampled from

densities $q_1(\cdot), q_2(\cdot)$, respectively, and accepted with probability

$$\min\left(1, \frac{\pi(x')q_2(u_2)}{\pi(x)q_1(u_1)}\left|\frac{\partial(x', u_2)}{\partial(x, u_1)}\right|\right),$$

and it is shown that the resulting transition kernel leaves the distribution π invariant.

If the assumptions of Theorem 6.6.2 hold, the acceptance probability in the theorem follows by simplifying the Jacobian term $|\frac{\partial(x', u_2)}{\partial(x, u_1)}|$ given above. The simplification comes from observing that $\partial x'/\partial u_1 = \partial u_2/\partial x{=}0$, because the inverse $u^{-1} = u_2$ is by construction independent of $x \in X$. This means that the Jacobian reduces to $|\partial x'/\partial x\, \partial u_2/\partial u_1| = |\partial g_{u_1}(x)/\partial x||\partial u_1^{-1}/\partial u_1|$, and we recover the acceptance probability (6.4) with $u = u_1$ and $q_1 = q_2 = q$, writing $\partial\tau/\partial u$ for the derivative $\partial u^{-1}/\partial u$. □

Remark 6.6.4: Theorem 6.6.2 can be seen as a particular case of the more abstract framework used in Liu and Sabatti (2000), which uses groups of transformation moves to construct the so-called Generalized Gibbs sampler. In Kou et al. (2005), the Generalised Gibbs sampling rule is presented as a Metropolis-Hastings move, with an acceptance probability given by Theorem 1 in that paper. In our notation, this theorem states that if $\{g_u, u \in \mathcal{U}\}$ is a locally compact group of transformations on X, and $\pi(x)$ is our target probability distribution, then to leave π invariant we should draw u from $f(u)H(du)$, where H is the left Haar measure (see e.g. Cohn, 1990), set $x' = g_u(x)$, and accept the move with probability

$$a(x, u) = \min\left(1, \frac{\pi(x')f(u^{-1})|J_{g_u}(x)|}{\pi(x)f(u)}\right).$$

If we assume that H is absolutely continuous with respect to the Lebesgue measure, we can introduce a density q such that $f(u)H(du) = q(u)du$. Assuming the transformation group is abelian, the left Haar measure coincides with the right Haar measure, such that $H(du) = H(du^{-1})$, and using this fact we can perform a change of measure within the transition kernel $T_x(u) = a(x, u)f(u)H(du)$, with $a(x, u)$ as given above, resulting in the acceptance probability (6.4).

6.6.2 Quantile moves

We now consider a particular type of transformation moves that act on a vector (θ, x), where θ is a parameter, and x is a collection of random variables whose distribution depend on θ. These so-called quantile moves will be used to obtain a joint transformation of the slices and their parameters in the latent trawl representation, which should improve the mixing compared to moving the components individually, as discussed in Section 6.5.

We describe the quantile moves in a general setting, where we are given some collection of random variables in $\mathbb{R}^d$, denoted x, which have a distribution with density f_θ and CDF F_θ. Furthermore, we assume that we want to sample (x, θ) conditional on some observations $y \sim f_y(\cdot|x)$, corresponding to a hierarchical model setup.

When targeting a distribution that involves the likelihood $f_\theta(x)$, it is useful to apply a transformation to the tuple (θ, x) that preserves the distributional properties of x with respect to f_θ, in order to avoid low values of $f_\theta(x)$. Of course, this could be done simply by proposing some new θ' and letting $x' \sim f_{\theta'}(\cdot)$. However, this proposal is independent of the current value of x, and so it completely disregards any information about the observations y that may be present in the current value of x. This independent proposal is in effect the opposite of a random-walk MH move, which is highly dependent on the current value of x, and proposes $x' = x + \epsilon$, without regard for the distribution f_θ.

The quantile moves attempt to propose new points that take into account the current value of x, while proposing a new value that is adapted to the distribution f_θ. This is done in three steps: first x is transformed to the uniform scale given the current value of θ, then a new θ' is sampled from a density q, and in the third step x is transformed back to the distributional scale under $f_{\theta'}$. We will write this transformation move as

$$g_u^F(\theta, x) = \left(\theta', F_{\theta'}^{-1}(F_\theta(x))\right),$$

where $\theta' = h(u, \theta)$, for a deterministic function h such that $h(u^{-1}, \theta') = \theta$.

The simplest example of this move uses $\mathcal{U} = \mathbb{R}$, $u^{-1} = -u$, with $h(u, \theta) = \theta + u$, and thus

$$g_u^F(\theta, x) = \left(\theta + u, F_{\theta+u}^{-1}(F_\theta(x))\right).$$

In the following we use the term *quantile moves* for these transformations, and denote them by $g_u^F(\theta, x)$.

To apply these moves in an MCMC algorithm, we use Theorem 6.6.2, where the transformation g_u^F is applied to $\mathbf{x} = (\theta, x) \in \mathbb{R}^{d+1}$, and the Jacobian $J_{g_u}(\theta, x)$ of the quantile move is given by

$$\left|\frac{\partial h(u,\theta)}{\partial \theta}\right| \frac{f_\theta(x)}{f_{\theta'}(F_{\theta'}^{-1}(F_\theta(x)))}.$$

For the latent trawl model, we can apply the quantile move to the slices $\{S_j\} = x$ and their parameters θ_s, and observe that the second term in the Jacobian cancels with the ratio of slice densities. such that the acceptance ratio only involves the observation densities and priors. This makes the quantile moves computationally cheap compared to proposing MH moves for the parameters, which requires the evaluation of all slice densities.

6.7 Extended MCMC algorithm

In this section we use the quantile transformation moves to obtain a more efficient MCMC algorithm for the latent trawl models. This is done by replacing the random-walk MH steps for the trawl process parameters θ_s, θ_t by quantile moves on the slices and parameters, resulting in Algorithm 1, hereafter called the extended MCMC algorithm.

In the algorithm description we let $l_y(\mathcal{O}|\,\{s_j\}, \theta_y)$ denote the log-likelihood of the observations given the slice values, and $l_s(\cdot|c_i, \theta_s)$ denote the likelihood of the slice given the cumulant scaling factor c_i, which corresponds to the term $\lambda^{Leb}(S_i)/\lambda^{Leb}(A)$ in (6.3). For example, if $L(A) \sim \mathcal{N}(\mu, \sigma^2)$, then $l_s(\cdot|c_i, \mu, \sigma^2)$ is the log-density of the distribution $\mathcal{N}(c_i\mu, c_i\sigma^2)$. Furthermore, $F(\cdot|\,c_i, \theta_s)$ is the CDF of the scaled slice distribution, $\theta[i]$ denotes the i'th parameter in the vector θ, and $\{s_{j[-i]}\}$ denotes the collection of slices with index $j \neq i$.

Remark 6.7.1: The extended algorithm is not explicitly dependent on a particular model specification, but takes a general model structure as input, which is given by the slice distribution, the scaled area of each slice, the observational distribution and the priors. This is exploited in the computational implementation of the latent trawl framework, where the model specification

can be changed easily without affecting the implementation of the MCMC algorithm.

There are three main parts of Algorithm 1, the first part updates the slices individually, the second samples the observational parameters θ_y, and the third performs transformation moves that jointly update the parameters θ_s, θ_t and the slices. By construction, each move in this algorithm has invariant distribution given by the augmented posterior:

$$\begin{aligned}\pi\left(\mathcal{S}, \theta \middle| \mathcal{O}\right) &\propto p\left(\mathcal{O}, \mathcal{S}, \theta\right) = p(\mathcal{O} \,|\, \mathcal{S}, \theta_y)\, p(\mathcal{S} \,|\, \theta_s, \theta_t)\, \pi(\theta_y, \theta_s, \theta_t) \\ &= \left[\prod_{i=1}^{n} f_i\left(Y_i \,\middle|\, \sum_{j \in \mathcal{T}_i} S_j \,, \theta_y\right)\right] \left[\prod_{j=1}^{N_s} p(S_j \,|\, \theta_s, \theta_t)\right] \pi(\theta_y)\pi(\theta_s)\pi(\theta_t).\end{aligned}$$

Furthermore, one iteration comprising these three parts gives each parameter and slice a positive probability of taking any value in their domain, and thus the resulting Markov Chain is irreducible. Asmussen and Glynn (2010) show that an irreducible Markov chain with one-step transition kernels of MH type will be positive Harris recurrent if the acceptance probability $a(x)$ of the MH move is positive for all x in the sampling space X, which is the case for our algorithm when we consider the three parts as one MH move.

Although the sampled Markov chain will eventually visit every point $x \in X$ as time goes to infinity, in practice we are interested in getting a good estimate of the posterior with reasonable computational effort, which can only be achieved if the chain has decent mixing properties. We found that introducing transformation moves improves the mixing substantially compared to the basic Gibbs sampling algorithm, in particular for the parameters θ_t of the trawl set. However, the fact still remains that we are sampling in a space X of dimension roughly equal to the number of slices N_s, which has order at least n^2, with n being the number of observations. This means that the Markov chain resulting from the extended algorithm will inevitably be slowly mixing compared to algorithms that only sample a parameter vector θ of fixed dimension, say 5, with the likelihood function known.

6.7.1 Diagnosing convergence

Several formal methods have been introduced that attempt to diagnose the convergence of MCMC chains to their stationary distribution, see e.g. Cowles

Algorithm 1: Extended MCMC sampler for latent trawl models

Data: $\mathcal{O} = \{y_i\}, \sigma_u^2, \sigma_t^2, \sigma_s^2$

`/* Update slices one at a time, using scaled slice area */`

for $i \leftarrow 1$ **to** N_s **do**

$c_i = \lambda^{Leb}(S_i(\theta_t))/\lambda^{Leb}(A(\theta_t))$

$u \sim \mathcal{N}(0, c_i\,\sigma_u^2), \quad s_i' = s_i + u$

$a(s_i', u) = \exp\{l_y(\mathcal{O}|\, s_i', \{s_{j[-i]}\}, \theta_y) - l_y(\mathcal{O}|\, s_i, \{s_{j[-i]}\}, \theta_y)$
$+ l_s(s_i'|\, c_i, \theta_s) - l_s(s_i|\, c_i, \theta_s)\}$

Accept s_i' with probability $\min(1, a(s_i', u))$.

end

for $\theta_y[i] \in \theta_y$ **do**

Sample θ_y from posterior or MH step.

end

`/* Apply quantile move to trawl set parameters and slices */`

for $\theta_t[i] \in \theta_t$ **do**

$u \sim \mathcal{N}(0, \sigma_t^2), \quad \theta_t'[i] = \theta_t[i] + u$

for $i \leftarrow 1$ **to** N_s **do**

$c_i = \lambda^{Leb}(S_i(\theta_t))/\lambda^{Leb}(A(\theta_t)), \quad c_i' = \lambda^{Leb}(S_i(\theta_t'))/\lambda^{Leb}(A(\theta_t'))$

$q_i = F(s_i|\, c_i, \theta_s), \quad s_i' = F^{-1}(q_i|\, c_i', \theta_s)$

end

$a(\theta_t', \{s_j'\}, u) = \exp\{l_y(\mathcal{O}|\, \{s_j'\}, \theta_y) - l_y(\mathcal{O}|\, \{s_j\}, \theta_y)\} + \pi(\theta_t') - \pi(\theta_t)$

Accept $\theta_t', \{s_j'\}$ with probability $a(\theta_t', \{s_j'\}, u)$

end

`/* Apply quantile move to Lévy seed parameters and slices */`

for $\theta_s[i] \in \theta_s$ **do**

$u \sim \mathcal{N}(0, \sigma_s^2), \quad \theta_s'[i] = \theta_s[i] + u$

for $i \leftarrow 1$ **to** N_s **do**

$c_i = \lambda^{Leb}(s_j(\theta_t))/\lambda^{Leb}(A(\theta_t))$

$q_i = F(s_j|\, c_i, \theta_s), \quad s_j' = F^{-1}(q_i|\, c_i, \theta_s')$

end

$a(\theta_s', \{s_j'\}, u) = \exp\{l_y(\mathcal{O}|\, \{s_j'\}, \theta_y)\} - l_y(\mathcal{O}|\, \{s_j\}, \theta_y)\} + \pi(\theta_s') - \pi(\theta_s)$

Accept $\theta_s', \{s_j'\}$ with probability $a(\theta_s', \{s_j'\}, u)$

end

and Carlin (1996). For the latent trawl model we do not employ formal convergence methods, partly because they require a large amount of additional computation effort, and more importantly, the experience from our simulation studies suggests that we can diagnose convergence of the extended MCMC algorithm by studying the output of the sampler. More specifically, we use three key indicators to diagnose convergence:

1. Individual parameter traceplots, which typically show an initial drift, followed by fluctuations within a smaller region, indicating convergence.

2. Plots of the running estimates of the log-likelihood components:

$$l_{obs}[k] = \prod_{i=1}^{n} f_i \left(Y_i \,\middle|\, \sum_{j \in \mathcal{T}_i} S_j^k \,, \theta_y \right)$$
$$l_{slice}[k] = \prod_{j=1}^{N_s} p(S_j^k \mid \theta_s^k, \theta_t^k),$$

   where k denotes the iteration index in the MCMC algorithm. As long as the likelihood components are steadily increasing, this indicates that the sampled chain has not yet converged. The conditional likelihood of the observations, l_{obs}, is particularly indicative of convergence, as it tends to have very small fluctuations in the stationary regime.

3. The final indicator is the empirical CDF (ECDF) of the trawl process values; after convergence we expect the trawl values to be sampled from $p(\Lambda | \, \mathcal{O}, \theta_s, \theta_t)$, with θ_s, θ_t sampled from the posterior distribution, and thus the ECDF of $\{\Lambda_i\}$ should be fairly stable after convergence.

6.8 Gaussian model

In this section we present a simple latent trawl model, which serves to illustrate the model framework and extended MCMC algorithm.

6.8.1 Model specification

We consider the model where both the latent trawl process and observations have a Gaussian distribution, i.e.

$$\begin{aligned} &Y_i|\,\Lambda_i \sim \mathcal{N}(\Lambda_i, \sigma_y^2), \\ &\Lambda_i = L(A) \sim \mathcal{N}(\mu, \sigma^2), \\ &A = \{(s,\xi) : -\infty < s \leq 0,\ 0 \leq \xi \leq d_{exp}(s;\rho)\}, \end{aligned}$$

where $d_{exp}(\rho)$ stands for the function $h(x) = \exp\{\rho x\}$ that defines the exponential trawl with decay parameter ρ. Thus the model is fully specified by the parameters $(\mu, \sigma^2, \sigma_y^2, \rho)$, and in terms of the notation used in Section 6.1, we have

$$\theta_y = \sigma_y^2, \quad \theta_s = (\mu, \sigma^2), \quad \theta_t = \rho.$$

This specification gives the marginal distribution

$$Y_i \sim \mathcal{N}(\mu, \sigma_y^2 + \sigma^2),$$

and by using (4.15), we get the covariance function of the trawl process as

$$\varphi_\Lambda(h) = \sigma^2 \exp(-\rho h) = \mathrm{Var}(L(A)) \exp(-\rho h),$$

giving an exponential correlation function.

The correlation of the observations Y_i becomes

$$\mathrm{Corr}(Y_i, Y_j) = \frac{\mathrm{Corr}(\Lambda_i, \Lambda_j)}{1 + \sigma_y^2/\sigma^2} = \frac{\exp(-\rho|i-j|)}{1 + \sigma_y^2/\sigma^2}.$$

The formula above shows that if the observational variance σ_y^2 is small, then the observations inherit most of the correlation from the trawl process, whereas for large σ_y^2, most of the trawl dependence structure will be lost in the observations.

6.8.2 Parametrisation issues

The specification of the normal-normal model, although conceptually simple, is not ideal for parameter estimation. In particular, the marginal variance of the observations is $\sigma_y^2 + \sigma^2$, and this creates dependence in the joint posterior distribution of σ_y^2, σ^2. Correctly identifying σ_y^2, σ^2 and ρ relies on knowing about the covariance function of the latent trawl process, and the observations $\{y_i\}$ will be much more informative about the marginal mean μ and variance $\sigma_y^2 + \sigma^2$ than about the latent covariance function. Similar identifiability issues have been discussed in the context of hierarchical models with random effects, see e.g. Gelfand et al. (1995); Papaspiliopoulos and Roberts (2003); Yan et al. (2007). For the geostatistical model presented in Yan et al. (2007), a reparametrisation was proposed with new parameters

$$\sigma_o^2 = \sigma_y^2 + \sigma^2, \quad \kappa = \sigma_y^2/\sigma_o^2.$$

We will use a similar approach for estimating the parameters of the Gaussian model, but instead of using a Gibbs sampler for the new parameters, we use the extended MCMC algorithm in the original parametrisation, and introduce additional transformation moves that are equivalent to scaling the parameters σ_o^2 and κ.

6.8.3 Sampling algorithm

For this model we use an extended MCMC sampling scheme, with conjugate priors for the parameters $\mu, \sigma^2, \sigma_y^2$, specifically

$$\mu \sim \mathcal{N}(0, 1), \quad \sigma^2, \sigma_y^2 \sim IG(1, 1),$$

where $IG(\alpha, \beta)$ stands for the Inverse Gamma distribution, with density

$$p(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} \exp(-\beta/x).$$

The Inverse Gamma priors can be considered uninformative, as they are reasonably flat close to zero, with heavy tails. The dependence parameter ρ is given a Gamma$(1, 1)$ prior.

For the sampling procedure, we introduce two transformation moves that are

used in addition to the steps in Algorithm 1. The first move uses $\mathcal{U} = \mathbb{R}_+$, $u^{-1} = 1/u$, samples u from a Gamma distribution and performs the transformation

$$\sigma_y^2 \to u\sigma_y^2, \quad \sigma^2 \to \sigma^2 + (1-u)\sigma_y^2, \quad \{s_i\} \to g_u^F(\sigma^{2\prime}, \{s_i\}),$$

which amounts to shifting the variance parameters while preserving their sum, and performing a quantile move on the slices. In the alternative parametrisation described in the previous subsection, this is equivalent to the move $\kappa \to u\kappa$, $\{s_i\} \to g_u^F(\kappa', \{s_i\})$.

The second transformation move also uses $\mathcal{U} = \mathbb{R}_+$, $u^{-1} = 1/u$, with a Gamma proposal, and scales the variance parameters together:

$$\sigma_y^2 \to u\sigma_y^2, \quad \sigma^2 \to u\sigma^2, \quad \{s_i\} \to g_u^F(\sigma^{2\prime}, \{s_i\}).$$

In the alternative parametrisation, this amounts to the move $\sigma_o^2 \to u\sigma_o^2$, $\{s_i\} \to g_u^F(\sigma_o^{2\prime}, \{s_i\})$, hence the two additional transformation moves are orthogonal in this parametrisation.

The slice updating and transformation moves described in Algorithm 1 depend on tuning parameters that specify the variance of the random walk proposals, and the two additional scaling moves depend on the variance of the Gamma proposal distribution. Because all parameters in the model are dependent in the posterior distribution, it is difficult to manually tune the algorithm to obtain an exact acceptance rate of the various moves. However, we obtained good results when aiming for an acceptance rate of about 0.2 for the simple random-walk proposals (inspired by the asymptotically optimal rate of 0.234 in Roberts et al. (1997)), and $0.2 - 0.4$ for the transformation moves.

6.8.4 Simulation study

We performed a simulation study to test the extended MCMC algorithm for the Gaussian model described above. Each run used a simulation of length $n = 3000$, with truncation $\tau = 10$, and the parameters set to random values. The algorithm was run for 30,000 iterations, with a burn-in period of 3000 iterations. In order to reduce the autocorrelation in the output, we used a subsample consisting of every 10th iteration, which forms the basis of the

	σ_y^2	μ	σ^2	ρ
Value	1.38	-0.21	0.35	0.31
Posterior mean	1.4	-0.19	0.32	0.35
Posterior CI	(1.28,1.51)	(-0.25, -0.14)	(0.23,0.44)	(0.17,0.58)

Table 6.1: Parameter values, posterior means and credible intervals for first simulation.

	σ_y^2	μ	σ^2	ρ
Value	3.64	0.55	3.11	1.2
Posterior mean	3.8	0.62	3.10	1.32
Posterior CI	(1.79,5.65)	(0.52,0.71)	(1.26,5.12)	(0.68,1.91)

Table 6.2: Parameter values, posterior means and 95% credible intervals for second simulation.

distributional summaries.

Here we present the results of two specific simulation runs, which are chosen to have significantly different parameter values. Figure 6.2 shows traceplots of the entire first run before subsampling. These plots show that the dependence parameter ρ has the slowest mixing, which is to be expected as it is only identified through the unobserved trawl process, and does not enter into the marginal distribution of the observations. Figure 6.3 shows histograms of the parameters together with 95% credible intervals calculated from posterior quantiles, showing that the posterior samples centre around a point close to the true parameter values, which are well within the credible intervals. Similar results were also found for the second simulation run.

Tables 6.1 and 6.2 give the values of the true parameters, together with the posterior means and credible intervals (CI) for both simulation runs. The coverage properties of the CIs are generally good, but for the second simulation, the intervals for the variance parameters and ρ are quite wide. This is presumably due to the parametrisation issues discussed in Section 6.8.2, which lead to strong posterior dependence between the variance parameters, combined with the large value of the observational variance σ_y^2, which means the data carry less information about the latent trawl process.

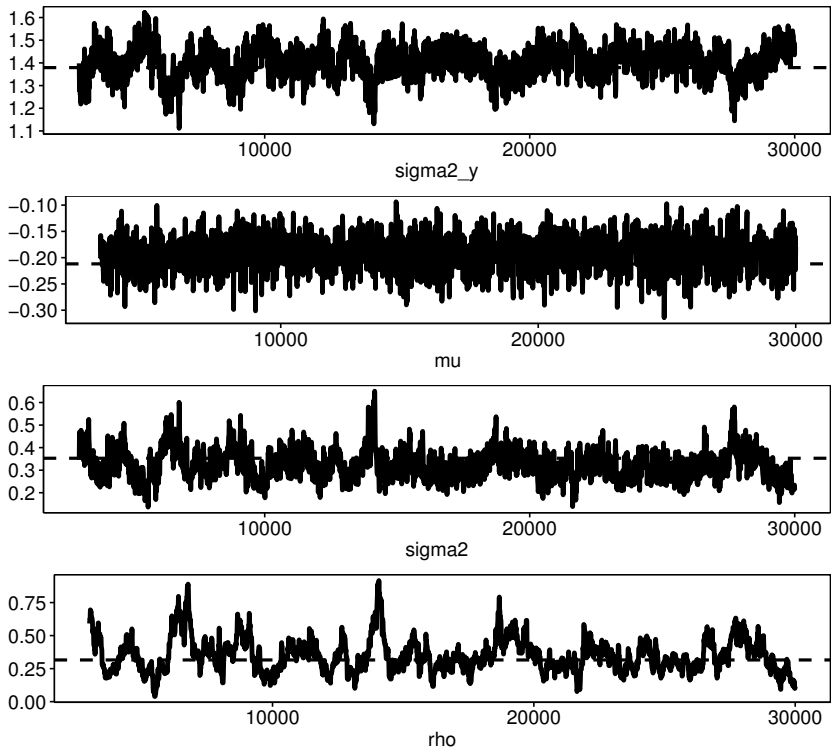

Figure 6.2: Traceplot for first simulation, dashed horizontal lines give true values.

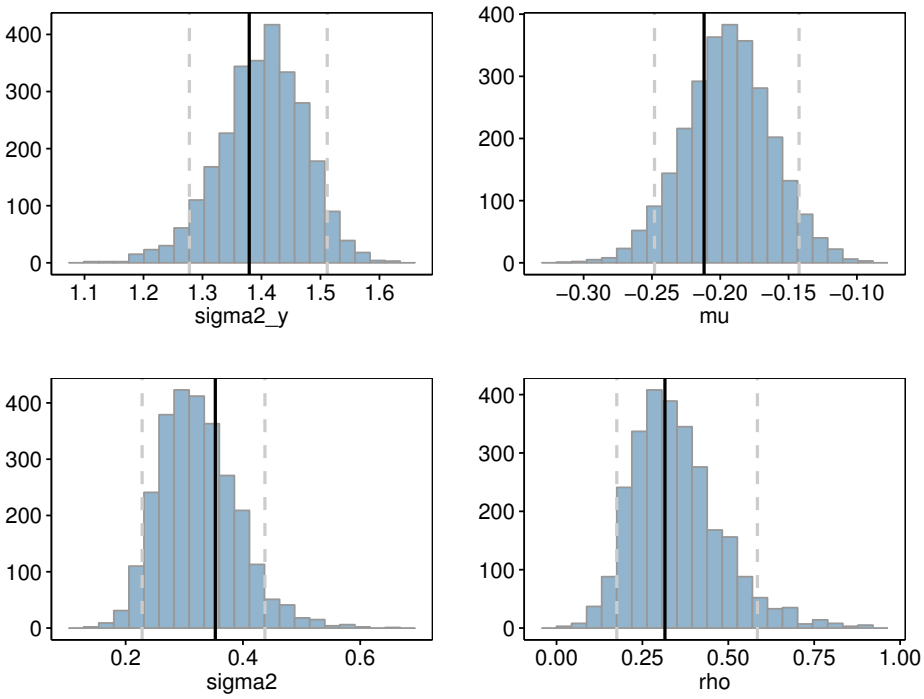

Figure 6.3: Histograms for first simulation, solid vertical lines give true values, dashed horizontal lines give boundaries of credible intervals.

6.9 Mean-variance mixture model

In this section we consider a mean-variance mixture model in the latent trawl framework, and apply it to a data set of wind production from a German transmission system operator.

6.9.1 Data description

The data used are published by Amprion GmbH (2016), and give the energy feed-in to the electricity grid in MWh, based on a selection of metered wind energy plants.

We used data from the years 2009-2015, and aggregate to daily mean values, giving a time series of length 2555. We also normalised the data to work with units of GWh, and divided out a simple linear trend, resulting in the values shown in Figure 6.4.

6.9.2 Model specification

For the wind production data we use a mean-variance mixture (MVM) model with a latent trawl process. In this model the trawl process represents the underlying amount of wind power, where the mean and variance of the observed value are assumed to be proportional to the latent process. To capture the yearly seasonality that can be seen in Figure 6.4, we use a multiplicative

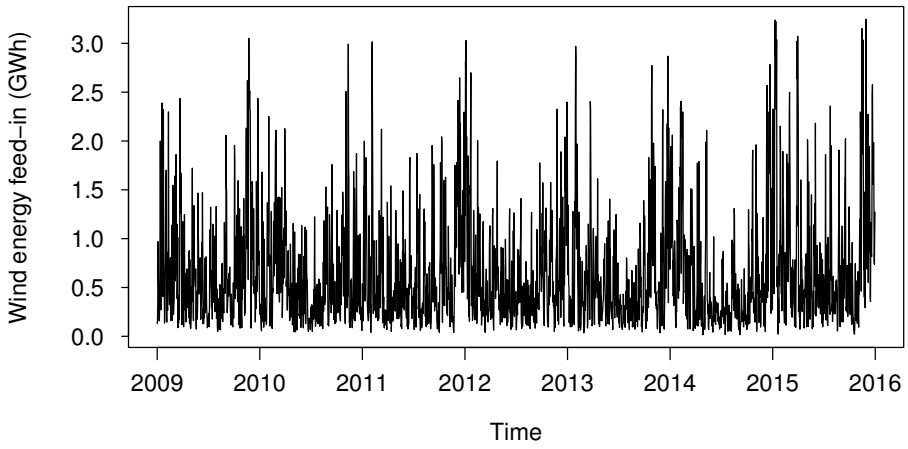

Figure 6.4: Time series of wind energy production over seven years.

seasonality function η given by the exponential of an order 2 Fourier series, specifically

$$\begin{aligned}\log(\eta(t_i)) =& a_1 \cos(2\pi t_i/365) + a_2 \sin(2\pi t_i/365) \\ &+ a_3 \cos(4\pi t_i/365) + a_4 \sin(4\pi t_i/365).\end{aligned}$$

The observations are then modelled as $\eta_i Y_i$, where Y_i comes from the latent trawl model, and the Fourier coefficients $\{a_i\}$ are considered part of the observational parameters θ_y.

Specifically, the latent trawl model uses a reparametrisation of a Variance Gamma (VG) distribution, where the observations are Gaussian, conditional on a latent Gamma random variable:

$$\begin{aligned}Y_i &\sim \mathcal{N}(\mu + \kappa\Lambda_i, \gamma^2\Lambda_i), \\ L(A) &\sim \text{Gamma}(\alpha, 1), \\ A &= d_{exp}(\rho),\end{aligned}$$

where d_{exp} is the exponential trawl, and the usual scale parameter of the Gamma distribution is represented here by its inverse value γ^2.

It can be shown that the correlation of the observations from this model is given by

$$Corr(Y_i, Y_j) = \frac{\rho_\Lambda(i,j)}{1 + \gamma^2/\kappa^2},$$

where ρ_Λ gives the correlation of the trawl process.

Clearly this is a rather simplistic model for wind production, as it does not account for any particulars of the process, for example the possibility that the production is adjusted manually in the presence of large or small amounts of available wind power. Furthermore, the model allows the observations to be negative, although the fitted distribution has only a few negative observations overall, which are truncated to zero in the output.

6.9.3 Sampling scheme

We now describe the sampling algorithm for the MVM model. In our standard notation, we have the parameters

$$\theta_y = (\mu, \kappa, \gamma^2, \{a_i\}), \quad \theta_s = \alpha, \quad \theta_t = \rho.$$

The MCMC sampling is done by using Algorithm 1, where the observational parameters μ, κ, γ^2 are sampled with an random-walk MH move.

Similar to the Gaussian model, the MVM model also suffers from identifiability issues. Specifically, the two observational parameters (κ, γ^2) are confounded with α, presumably because the trawl process value is only observed though the product with either κ or γ^2. Although the parameters should technically be identifiable though the marginal VG distribution, we were not able to obtain good estimates by using the extended MCMC algorithm with somewhat weak priors on all parameters.

To improve the sampling, we introduced two multiplicative transformation moves of the form

$$u \in \mathbb{R}_+,\, u^{-1} = 1/u, \quad \kappa \to u\kappa, \quad \gamma^2 \to u\gamma^2, \quad s_i \to 1/u s_i.$$
$$u \in \mathbb{R}_+,\, u^{-1} = 1/u, \quad \mu \to u\mu, \quad \alpha \to \alpha + \mu(1-u)/\kappa, \quad \{s_i\} \to g_u^F(\alpha', \{s_i\}).$$

Although this improved the mixing properties of the algorithm, the marginal fit was still not satisfactory. We therefore resorted to using an Empirical Bayes approach for the VG parameters in θ_y, with prior distributions set to $\mathcal{N}(0, \sigma_o^2)$ for μ, κ, and Gamma$(\alpha_0, 1)$ for γ^2, where the hyperparameters σ_o^2, α_0 are derived from a maximum-likelihood estimation of the marginal VG parameters. This approach is not ideal, and will likely lead to an underestimation of the variance in the observational parameters θ_y, but it gives a way of focusing on the estimation of the latent trawl process without compromising the marginal fit.

Finally, the priors for the trawl process parameters are given by $\pi(\alpha), \pi(\rho) \sim Gam(1, 1)$, and the coefficients $\{a_i\}$ have $\mathcal{N}(0, 0.1)$ priors.

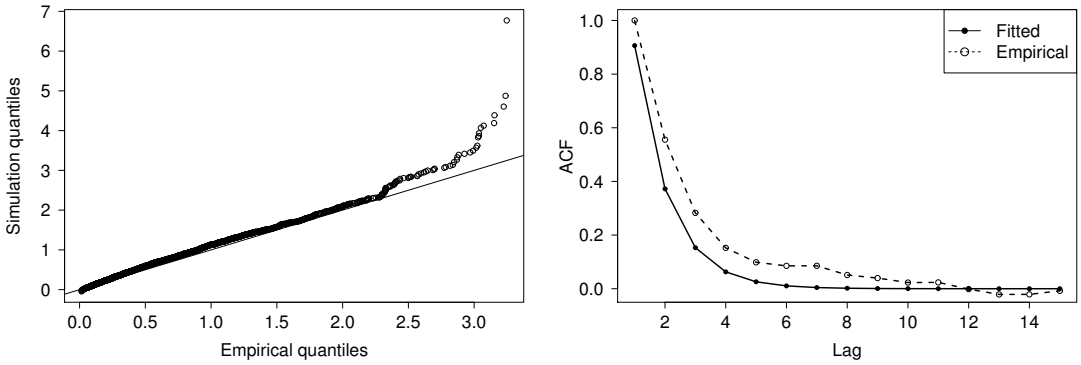

Figure 6.5: QQ-plot(left) and comparison of autocorrelation functions (right) for the mean-variance mixture model.

6.10 Model fit

Figure 6.5 shows a QQ-plot and a comparison of autocorrelation functions for the fitted MVM model. These plots are based on a simulation of length 7300, equivalent to 20 years of daily data, and the ACF plot uses the deseasonalised versions of the data and simulation, based on the fitted seasonality functions. We see that the marginal fit is acceptable, although the simulations have heavier tails than the data; this could be due to the existence of a cutoff value for the empirical wind production, which is not included in the model. The shapes of the autocorrelation functions for the empirical data and simulation are similar, with a slight underestimation of the correlation by the model. This ACF fit could possibly be improved by using a more flexible trawl set parametrisation, such as a mixture of exponential functions.

Figure 6.6 shows path plots for the empirical and simulated data. This also shows the lack of fit in the tails, however, the overall pattern of fluctuations seems to be captured by the model.

Finally, Figure 6.7 demonstrates the filtration obtained from the MCMC algorithm. Specifically, each iteration of the MCMC algorithm gives a sample of the full trawl process $\{\Lambda_i\}$, conditional on the data. This gives an expected value $\mu + \kappa\Lambda_i$ of each observation Y_i, which is plotted in the Figure for the last iteration, together with the deseasonalised observations. This indicates that the sampled trawl process adapts well to the fluctuations in the observed values.

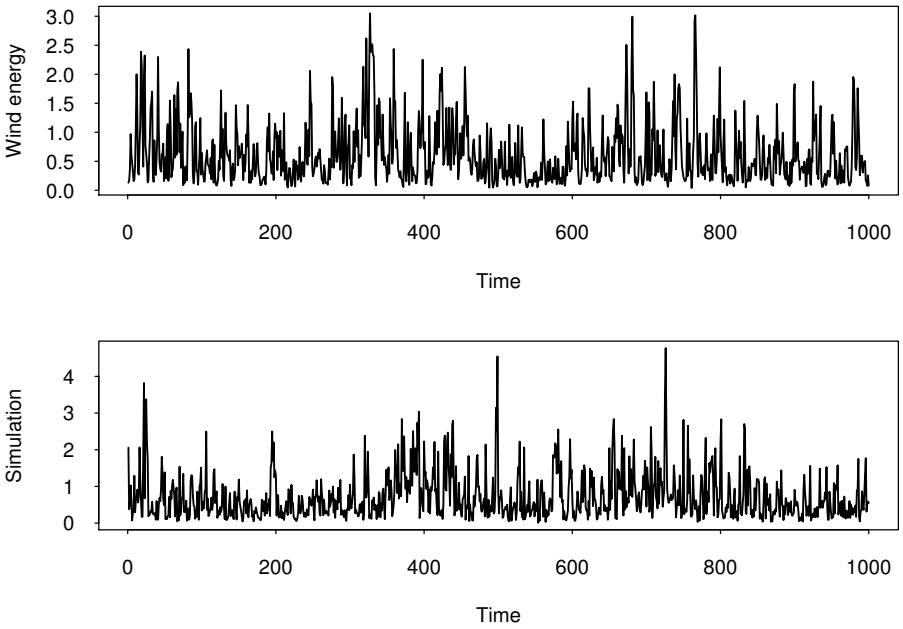

Figure 6.6: Path plots for mean-variance mixture model, top panel shows empirical values, bottom panel shows simulations.

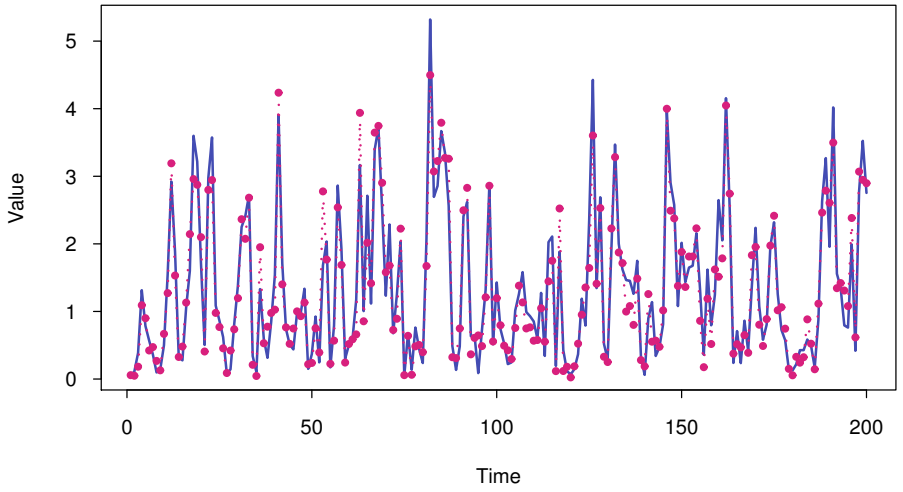

Figure 6.7: Value of sampled mean curve $\mu + \kappa\Lambda_i$ for VG model (red), and deseasonalised observations (blue).

6.11 Conclusion

In this chapter we have introduced a hierarchical model framework based on the class of trawl processes, demonstrating how to construct a general model with flexible distributional and dependence properties. Due to the hierarchical construction, the likelihood of the models in our framework is not available directly, and we discussed possible approaches to parameter estimation, which attempt to integrate out the latent trawl process that appears in the likelihood function. It was found that the calculation of the likelihood is prohibitively difficult, and thus we considered sampling-based estimation approaches, which exploit the slice representation presented in Section 4.4.

We showed how to construct a basic Gibbs sampling algorithm for latent trawl models, and introduced an approximation of the slice representation to improve computational efficiency. We also extended this algorithm by including so-called transformation moves, resulting in a more efficient sampling scheme. Finally, we introduced two specific models in the latent trawl framework, a basic Gaussian model and a mean-variance mixture model. Simulation studies for the Gaussian model demonstrate that the extended MCMC algorithm works well for the purposes of parameter estimation, although one needs to take care with the particular model parametrisation.

The mean-variance mixture model shows that the latent trawl framework can accommodate heteroscedasticity by letting the variance of the observations depend on the latent process. When fitting this model to a time series of wind production data, we found that the model has problems with identifiability, and to counteract these we used empirical priors for the observational parameters based on the initial marginal fit. The results show that the model captures the dependence structure of the observations, and that the sampled trawl process adapts to the observed fluctuations in the wind energy level.

7

TOWARDS SPATIO-TEMPORAL MODELS

In this chapter we consider spatio-temporal models based on ambit processes, which is a large class that includes trawl processes, as well as the Lévy-driven OU process used in Chapter 2. Thus it is a natural generalisation of the models we have considered so far.

Some literature related to ambit processes was discussed briefly in Section 4.1. This class is studied in detail in Barndorff-Nielsen et al. (2015), and has been applied for modelling turbulence (Hedevang and Schmiegel, 2014), tumour growth (Jónsdóttir et al., 2008) and brain imaging (Jónsdóttir et al., 2011). In Barndorff-Nielsen et al. (2015), the focus is on the theoretical properties of these processes, as well as the introduction of volatility modulation, whereas the remaining papers focus on a particular application.

In the existing literature on ambit processes, fitting is typically done in two steps: first, the parameters of the autocovariance function are fitted by least-squares, and then the Lévy basis is fitted by matching moments or cumulants. This approach is similar to that used for our rainfall model in Chapter 2, and gives a quick and simple method for parameter estimation when the data are generated directly from the ambit process. However, the resulting estimators will not have the optimality properties that maximum likelihood or Bayesian estimators enjoy. Furthermore, it may become difficult to calculate higher-order moments or cumulants when the ambit process is embedded in a larger statistical model.

In contrast to the approaches described above, this chapter takes a view toward wider statistical models based on ambit processes, such as the latent trawl framework in the previous chapter. The aim is then to develop a unified inference method that uses the full model specification. As a step towards this aim, we develop a discrete representation of the Lévy basis that makes

the resulting process more tractable for simulation and inference in the spatio-temporal case.

7.1 Kernels versus trawl sets in temporal processes

In this section we discuss the connection between the stochastic processes that have been used in our models thus far, focusing on the difference between a general ambit process and the special trawl process case.

Up until this point we have worked with various temporal processes of the form

$$Y(t) = \int_{\mathbb{R}^k} g(t,u)\, dL(u),$$

which can be classified as a temporal ambit process, where L is a Lévy basis and g is called the kernel function.

Definition 7.1.1: In the following we will take the term *kernel process* to mean an ambit process where the kernel is *not* an indicator function, in order to contrast with the trawl process case. In this terminology, the Lévy-driven OU process and the CARMA generalisation considered in Chapter 2 are kernel processes.

We now discuss the difference between the trawl and kernel subclasses, which is illustrated by the models considered in previous chapters. In particular, the Lévy-driven OU process used in Chapter 2 in defined by a stochastic integral with dimension $k = 1$, with an exponential kernel function

$$g(t,u) = e^{-\rho(t-u)},\ u \leq t.$$

On the other hand, the trawl processes introduced in Chapter 4 have integration dimension $k = 2$, and taking an exponential trawl set gives

$$\begin{aligned} g(t,u) &= \mathbb{I}_{A_t}(u_1,u_2), \\ A_t &= \{(u_1,u_2) : u_1 \leq t, u_2 \leq \exp(-\rho(t-u_1)\}. \end{aligned}$$

The two constructions described above both result in processes with an exponential autocovariance function. If we let these processes be driven by the

same Lévy-basis L, a natural question is whether they are in fact different representations of the same stochastic process. In general this is not the case, one difference being that the ambit process with $k = 1$ has the Markov property relative to an appropriate filtration $\mathcal{F}_t$, whereas the trawl process with $k = 2$ is in general not Markovian. Furthermore, the trawl process has the useful property of separating marginal and dependence structures, whereas for the kernel process, the marginal distribution involves the kernel function g, which enters as an integrand in the cumulant function.

This contrast between trawl and OU-type processes is discussed in Wolpert and Taqqu (2005), where they use the term "upstairs" representation for the former, and "downstairs" representation for the latter, referring to the location of the kernel function g, which appears either in the main integrand or as the upper limit of a second integral. They point out that if L is Gaussian, then these processes have the same joint distributions, as the Gaussian process is characterised by its covariance function.

7.2 Spatio-temporal model setup

We now consider a quite general spatio-temporal ambit process, which we will write in the form

$$Y(t,x) = \int_{\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}} g_1(t,u)g_2(x,y)\, L(du,dy), \tag{7.1}$$

where $t \in \mathbb{R}$ represents time, $x \in \mathbb{R}^2$ represents space, and g_1, g_2 are temporal and spatial kernel functions, respectively.

A major advantage of using the general ambit formulation is that in many cases it is possible to derive an explicit expression for the kernel function g such that the resulting process $Y(t,x)$ has a given covariance function. To show this, we first separate the temporal and spatial parts of the covariance function. This is done by applying a version of Lemma A.1.17 for Lévy bases, resulting in the covariance function

$$\begin{aligned}\varphi(l,h) &:= \mathrm{Cov}(Y(0,0),Y(l,h)) = \sigma^2 \int_{\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}} g_1(0,u)g_1(l,u)\, g_2(0,y)g_2(h,y)\, du\, dy \\ &:= \sigma^2\varphi_1(l)\varphi_2(h),\end{aligned}$$

where $\sigma^2 = \text{Var}(L')$ and L' is the Lévy seed of L.

This shows that by writing the overall kernel as a product of temporal and spatial kernel functions, we obtain a separable covariance structure, and in the following we will assume that all covariance functions of interest are separable. Hence we can focus on the spatial case with kernel g_2 and covariance function φ_2 without loss of generality, as the same results apply in the temporal case for one dimension fewer. Furthermore, for the temporal trawl process we have already derived a correspondence between trawl sets and covariance functions in Section 4.5.1.

7.2.1 Kernel-Covariance duality

Having separated out the temporal part, we are left with the problem of deriving a kernel function $g(\cdot, \cdot)$ on $\mathbb{R}^d \times \mathbb{R}^d$ that induces a given covariance function φ for the spatial process

$$Y(x) = \int_{\mathbb{R}^d} g(x, y) L(dy). \tag{7.2}$$

It is shown in Hedevang and Schmiegel (2014) and Jónsdóttir et al. (2011) that this can be done by linking the spectral density of φ to the kernel function g. In particular, we assume $\varphi \in L^1(\lambda^{Leb})$ and consider a symmetric kernel function k such that $g(x, y) = k(||x - y||)$. Then we can perform a Fourier transform of the correlation function φ, to give

$$\begin{aligned}
\mathcal{F}\left[\varphi(h)/\sigma^2\right](u) &= \frac{1}{\sqrt{2\pi}} \int \int e^{-ihu} k(z) k(h+z)\, dz\, dh \\
&= \frac{1}{\sqrt{2\pi}} \int k(z) \left(\int e^{-ihu} k(h+z)\, dh \right) dz \\
&= \frac{1}{\sqrt{2\pi}} \int k(z) \left(\sqrt{2\pi} e^{iuz} \mathcal{F}[k](u) \right) dz \\
&= \left(\mathcal{F}[k](u)\right)^2,
\end{aligned}$$

and hence we can obtain the kernel function k as

$$k(y) = \mathcal{F}^{-1}\left(\sqrt{\mathcal{F}\left[\varphi(h)/\sigma^2\right]} \right)(y).$$

Jónsdóttir et al. (2011) give specific examples of kernel functions that give rise to the Matérn family of covariance functions (Matérn, 1960), which is a well-known family that is typically used for geostatistical applications.

7.3 Lévy basis approximation

The result in the previous section shows that ambit processes can be adapted to a given covariance structure, and we also know that they have flexible distributional structures that are derived from the Lévy basis L. However, the joint density of ambit processes is not available in closed form, which presents difficulties for parameter estimation. Furthermore, the characteristic function for a kernel process involves an integral over the entire space $\mathbb{R}^d$, in contrast to the trawl case, where it reduces to the analytical form in Lemma 4.4.4. Finally, the ambit processes cannot in general be simulated exactly, as there is no finite-dimensional representation like the slices; typically one works with approximations that evaluate L over a fine grid, but this becomes very computationally intensive in higher dimensions.

The lack of analytical and computational tractability discussed above presents challenges when using kernel processes for statistical modelling and estimation. In order to get a more tractable process, we will consider using a discrete approximation of the Lévy basis, which makes the resulting process much easier to work with.

A simple way of obtaining a discrete Lévy basis would be to just consider the subclass of infinitely divisible distributions that can be represented as compound Poisson processes, so that by construction the Lévy basis is a marked point process.

A second option involves an approximation of a given infinitely divisible distribution by a finite number of points. This discrete Lévy basis approximation is motivated by the connection to completely random measures that was described in Section 4.2.2, where a non-negative Lévy basis L with no drift term, defined on a space S, can be written as

$$\tilde{L} = \sum_{i=1}^{\infty} \gamma_i \delta_{s_i}, \quad a.s.,$$

where (s_i, γ_i) are locations and marks that come from an inhomogeneous Poisson point process on $S \times \mathbb{R}^+$, with mean measure $n(dz, dx) = \lambda^{Leb}(dz)\nu(dx)$, where ν is the Lévy-measure of L.

There also exists a natural extension to completely random *signed* measures (CRSM), defined in Hellmund (2009), where it is shown that a Lévy basis can be written as a CRSM under a finite variation-type condition on L. From this it is clear that for Lévy bases with infinite variation, such as the Gaussian case, the realisations of $\tilde{L}$ do not have the same properties when seen as a measure on S. However, these are limiting properties that are not usually important for statistical applications, so we will not focus on these, but rather on how to use $\tilde{L}$ in an ambit process to obtain given marginal distributions and covariance structures.

We now consider an approximation of $\tilde{L}$ with finitely many terms. The basic principle is similar to the grid approximation of an ambit process, where the value of the Lévy basis in each grid square comes from a random variable whose cumulant function is scaled by the area of the square.

Definition 7.3.1: We define a Lévy basis approximation as follows: fix a space Ω with unit area, let L be a Lévy basis with Lévy seed L', and let N be some large integer. Let the locations $\{s_i\}_{i=1,\dots,N}$ be uniformly distributed over Ω, and let $\{\gamma_i\}_{i=1,\dots,N}$ be i.i.d. random variables, which are independent of the locations $\{s_i\}$ and defined via their cumulant function

$$C\left(u\,;\gamma_i\right) = \frac{1}{N} C(u, L').$$

Then we can define the approximation

$$L_N = \sum_{i=1}^{N} \gamma_i \delta_{s_i},$$

where in the limit as $N \to \infty$, the cumulant function of $L_N(A)$ becomes $\lambda^{Leb}(A)C(u, L')$, for $A \in \mathcal{B}_b(S)$.

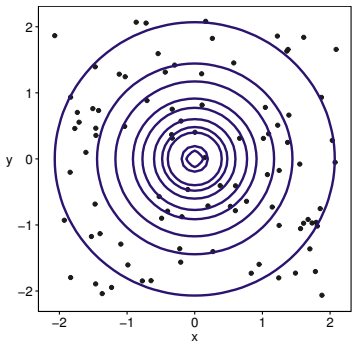

Figure 7.1: Illustration of discrete Lévy basis approximation, with contours of exponential kernel around $x = 0$ in blue.

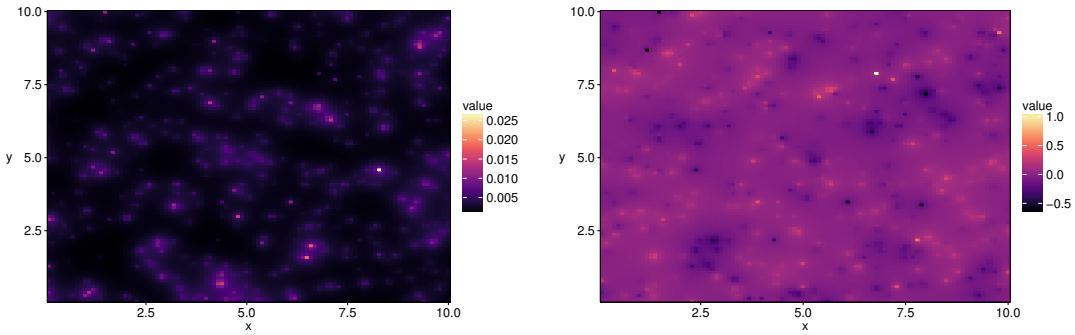

Figure 7.2: Ambit processes with exponential kernel and Lévy basis approximation, with values γ_i from an exponential (left) and Gaussian (right) distribution.

7.4 Ambit process examples

We now show some examples of spatial ambit processes of the form (7.2), where we use the approximation L_N in place of the Lévy basis L. In particular, this means that we get a process of form

$$Y_N(x) = \sum_{i=1}^{N} g(x, s_i)\gamma_i,$$

so that each value γ_i gets weighted according to the distance between its associated location s_i and x. This is illustrated in Figure 7.1, which shows simulated points from the discrete approximation, together with the contours of the kernel k corresponding to the Matérn covariance function.

Figures 7.2 and 7.3 show examples of particular model constructions with different distributions and kernel functions.

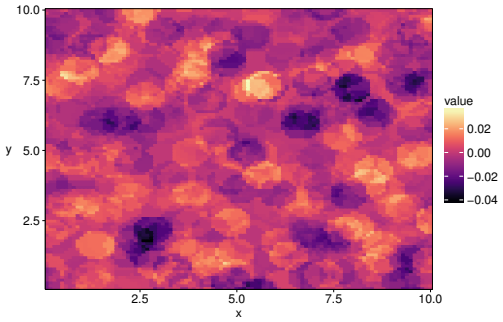

Figure 7.3: Trawl process with circular kernel, radius 0.5 and Lévy basis approximation with Gaussian distribution.

7.5 Conclusion

In this chapter we have outlined a construction of spatio-temporal models based on ambit processes. It was shown that these processes constitute a unifying class for the models considered in this thesis, and we discussed the contrast between the general ambit form and the special trawl process case. We also considered using a discrete approximation of the Lévy basis, which has the potential for simplifying the parameter estimation problem for models based on this class of processes. Finally, we showed examples of realisations from the resulting spatial processes.

8

CONCLUSION AND FUTURE WORK

8.1 Conclusion

In this thesis we have introduced several statistical models for capturing distributional and dependence structures in spatio-temporal data. The models are all based on stochastic processes that arise from the convolution of an infinitely divisible random measure with a kernel function. We have demonstrated that using these processes enables us to construct models with flexible distributions and dependence structures.

In Chapter 2 we introduced a new class of continuous-time stochastic processes, driven by the Hougaard Lévy process, and showed how it can be used to construct a parsimonious and analytically tractable model for rainfall. This process belongs to the class of continuous-time ARMA (CARMA) process, which results in a model with a flexible autocorrelation structure.

We presented a general fitting method for this class that exploits a correspondence between integrated CARMA and ARMA processes. The model fits the marginal distribution of the rainfall well on both hourly and daily time scales, where the daily fit compares favourably to the standard model described in Wilks (1998). By virtue of the CARMA generalisation, the extended model manages to accurately reproduce the autocorrelation structure of the observed rainfall.

In Chapter 3 we derived a result specifying the risk-neutral expectation of a function of the rainfall process, which can be used for pricing general derivatives written on a precipitation index. We used the Esscher transform to construct a risk-neutral measure, with a time-dependent parameter representing the risk premium. The pricing methodology generalises the independent increment model considered in Benth and Šaltytė Benth (2013), which does not

allow for price dynamics of derivatives. The pricing method was illustrated by calculating futures prices based on empirical daily rainfall data from Detroit.

In Chapter 4 we presented the theoretical background for the class of trawl processes. In particular, we introduced a novel representation in terms of independent random variables, called slices, and discussed the temporal dependence structure of trawl processes in the context of statistical modelling.

In Chapter 5 we investigated a new model for time series of exceedances, which uses the trawl process framework to obtain a flexible dependence structure. This is done by replacing the latent Markov chain in the setup of Bortot and Gaetan (2014) with a trawl process. In contrast with other hierarchical models, this construction has the advantage of preserving the generalised Pareto distribution for the marginals, which is consistent with extreme value theory. We also considered a modification to the original model structure that extends the parameter space, allowing for negative shape parameters in the Generalised Pareto distribution. The original and modified extreme value models were then used to analyse time series of rainfall and pollution measurements.

There are several aspects of the temporal exceedance model that could benefit from further investigation. First, we used a simple exponential trawl set, and it would be interesting to look at the result of using more flexible parametrisations. Another possibility is to consider a model where the threshold u for defining an exceedance is allowed to vary, and then letting the trawl set depend on u, which should result in a wider range of dependence levels across thresholds.

In Chapter 6 we introduced a hierarchical model framework based on the class of trawl processes, demonstrating how to construct a general model with flexible distributional and dependence properties. We discussed possible approaches to parameter estimation, and developed an MCMC sampling algorithm that exploits the slice representation presented in Chapter 4. The MCMC algorithm was extended by including so-called transformation moves, resulting in a more efficient sampling scheme. Finally, we introduced two specific models in the latent trawl framework, a basic Gaussian model and a mean-variance mixture model. Simulation studies for the Gaussian model demonstrate that the extended MCMC algorithm works well for parameter estimation. The mean-variance mixture model was fitted to a data set of

wind production, and it was found that the model captures the dependence structure of the observations, and that the sampled trawl process adapts to the observed fluctuations in the wind energy level.

The latent trawl framework could be expanded in several directions. In particular, the general structure of the latent trawl model suggests the exploration of a wider range of model specifications, which can be tailored to particular applications. It would also be useful to develop more efficient approaches for parameter estimation; these could be based on likelihood approximations as discussed in Chapter 6, or one could explore alternative MCMC methods such as particle filtering.

In Chapter 7 we outlined a construction of spatio-temporal models based om ambit processes. It was shown that these processes constitute a unifying class for the models considered in this thesis, and we discussed the contrast between the more general ambit form and the special trawl process case. We also considered using a discrete approximation of the Lévy basis, which has the potential for simplifying the parameter estimation problem for models based on this class of processes.

The spatio-temporal ambit models are particularly promising for future research. The first challenge would be to develop an effective fitting method, which could be based on an MCMC algorithm that resamples the points in the discrete Lévy basis approximation. Another important direction is to investigate more general ambit or trawl processes; for example, the trawl set itself could be randomised, and the Lévy basis and kernel could be made to depend on covariates in a wider statistical model.

A

BACKGROUND ON LÉVY-DRIVEN PROCESSES

This appendix presents some selected theory for Lévy processes, stochastic integrals and Lévy-driven Ornstein Uhlenbeck processes. Most of this material is taken from Applebaum (2009) Chapters 1, 2, 4 and 6, as well as Sato (1999) Chapters 1-2. We omit proofs throughout for the sake of brevity.

A.1 Lévy processes

Definition A.1.1: A stochastic process X defined on a probability space $(\Omega, \mathcal{F}, P)$ is called a *Lévy process* if

1. $X(0) = 0$ a.s.
2. X has independent and stationary increments.
3. X is continuous in probability, so for all $\epsilon > 0$,

$$\lim_{t \to s} P\left(|X(t) - X(s)| > \epsilon\right) = 0.$$

Definition A.1.2: We say that a random variable X has an *infinitely divisible distribution* if for any $n \in \mathbb{N}$ there exist i.i.d random variables $X_1, \ldots, X_n$ such that

$$X \stackrel{d}{=} X_1 + \ldots + X_n.$$

Proposition A.1.3: A Lévy process X is infinitely divisible, meaning that $X(t)$ is an infinitely divisible random variable for each $t \geq 0$.

Definition A.1.4: (Sato, 1999, pp. 33-34). If X is infinitely divisible, it can be shown that its characteristic function φ has no zeroes, and then we can

define a unique continuous function $\psi : \mathbb{R} \to \mathbb{C}$ such that $\varphi(z) = \exp(\psi(z))$. This function is called the *distinguished logarithm* of φ.

Theorem A.1.5 (Lévy-Khintchine): Let X be a random variable taking values in $\mathbb{R}$, then X has an infinitely divisible distribution iff its characteristic function is given by

$$\varphi(u) := E[e^{iuX}] = \exp\left\{iau - \frac{1}{2}bu^2 + \int (e^{iux} - 1 - iux\mathbb{I}_{[-1,1]}(x))\,\nu(dx)\right\}, \tag{A.1}$$

with $b \geq 0, a \in \mathbb{R}$ and ν a Lévy measure, i.e.

$$\nu(\{0\}) = 0, \ \int (|x|^2 \wedge 1)\,\nu(dx) < \infty.$$

For given X, the so-called *characteristic triplet* (a, b, ν) is uniquely determined by X. Conversely, for a given canonical triple there exists an infinitely divisible distribution with characteristic function given by (A.1).

Theorem A.1.6: For a Lévy process X, the distinguished logarithm of $\varphi_t(u) = E[\exp iuX(t)]$ takes the form

$$\psi_t(u) = t\psi(u),$$

where ψ is the distinguished logarithm of $X(1)$. We say that ψ is the *cumulant function* of X, and denote it by $C\,(u\,; X)$.

Definition A.1.7: A (standard) *Brownian motion* on $\mathbb{R}$ is a Lévy process $(B(t))_{t\geq 0}$ such that

1. $B(t) \sim N(0, t)$ for each $t \geq 0$.
2. B has continuous sample paths.

Definition A.1.8: A *Poisson process* with intensity λ is a Lévy process taking values in $0, 1, 2, \ldots$ such that

$$P(N(t) = n) = \frac{(\lambda t)^n}{n!}e^{-\lambda t}.$$

Definition A.1.9: Let (Z_i) be a sequence of i.i.d. real-valued random variables and let N be a Poisson process of intensity λ, independent of all Z_i, then the *compound Poisson process* is defined by

$$\begin{aligned} Y(0) &= 0, \\ Y(t) &= \sum_{i=1}^{N(t)} Z_i, \quad t > 0. \end{aligned}$$

It can be shown that the compound Poisson process is a Lévy process.

Definition A.1.10: A *subordinator* is a Lévy process that is almost surely non-decreasing.

Proposition A.1.11: If T is a subordinator, then its cumulant function takes the form

$$C(u;T) = ibu + \int_0^\infty (e^{iuy} - 1)\,\nu(dy),$$

where $b \geq 0$ and ν satisfies

$$\nu(-\infty, 0) = 0, \quad \int_0^\infty \min(y, 1)\nu(dy) < \infty.$$

A compound Poisson process is a subordinator iff the generating variables Z_i take values in $\mathbb{R}_+$. In this case $b = 0$, so the process consists only of jumps, and we call it a *pure-jump subordinator*.

A.1.1 Moving average processes

Definition A.1.12: We follow Applebaum (2009), Chapter 4, and define a general moving-average process of the form

$$Z(t) = \int_{-\infty}^\infty f(s - t)\,dL(s),$$

where L is Lévy process and $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ is a deterministic càglàd function, and the Lévy measure ν of L is assumed to be such that

$$\int_{|x|>1} |x|\,\nu(dx) < \infty.$$

It can be shown that this process is strictly stationary.

Definition A.1.13: We can now define the *Ornstein-Uhlenbeck* process as a particular moving-average process. In particular, let $f(s) = \exp(\lambda s)\mathbb{I}_{(-\infty,0]}(s)$, then we get the process

$$Z(s) = \int_{-\infty}^{t} e^{-\lambda(t-s)}\, dL(s).$$

Proposition A.1.14: The Ornstein-Uhlenbeck process is the unique solution to the stochastic differential equation

$$dZ(t) = -\lambda Z(t)dt + dL(t).$$

The following Lemma describes the characteristic function of moving average processes, and is taken from (Cont and Tankov, 2004, Lemma 15.1).

Lemma A.1.15: Let $f : [0, T] \to \mathbb{R}$ be a left-continuous function and $L(t)$ a Lévy process. Then

$$E\left[\exp\left\{\int_0^t if(s)dL(s)\right\}\right] = \exp\left\{\int_0^t \psi(f(s))ds,\right\}$$

where $\psi(t)$ is the cumulant function of L, given by

$$\exp\{\psi(u)\} = E[e^{iuL(1)}].$$

The above Lemma is a special case of a fundamental result for integrals involving Lévy bases, which is derived in Rajput and Rosinski (1989, Proposition 2.6), and stated here in our notation for a homogeneous Lévy basis:

Proposition A.1.16: Let L be a homogeneous Lévy basis and f an L-integrable function, then

$$E\left[\exp\left\{iu\int_S f(s)L(ds)\right\}\right] = \exp\left\{\int_S C\left(uf(s)\,;L'\right)\right\} ds, \quad u \in \mathbb{R}.$$

Lemma A.1.17: We have that

$$\mathrm{Cov}\left(\int_0^t f(s)\,dL(s)\int_0^t g(s)\,dL(s)\right) = \mathrm{Var}(L^2(1))\int_0^t f(s)g(s)\,ds.$$

This follows from a standard Itô-type isometry for stochastic integrals that can be found in e.g. Protter (2005).

A.2 Copyright statement

This section reproduces the copyright statement for the paper entitled "A Lévy-driven rainfall model with applications to future pricing" (Noven et al., 2015), which is based on the work presented in Chapters 2 and 3, where the section relevant for this thesis is highlighted in bold.

> The copyright to this article, including any graphic elements therein (e.g. illustrations, charts, moving images), is hereby assigned for good and valuable consideration to Springer-Verlag Berlin Heidelberg effective if and when the article is accepted for publication and to the extent assignable if assignability is restricted for by applicable law or regulations (e.g. for U.S. government or crown employees). Author warrants (i) that he/she is the sole owner or has been authorized by any additional copyright owner to assign the right, (ii) that the article does not infringe any third party rights and no license from or payments to a third party is required to publish the article and (iii) that the article has not been previously published or licensed.
>
> The copyright assignment includes without limitation the exclusive, assignable and sublicensable right, unlimited in time and territory, to reproduce, publish, distribute, transmit, make available and store the article, including abstracts thereof, in all forms and media of expression now known or developed in the future, including pre- and reprints, translations, photographic reproductions and microform. Springer may use the article in whole or in part in electronic form, such as use in databases or data networks (e.g. the Internet) for display, print or download to stationary or portable devices. This includes interactive and multimedia use as well as posting the article in full or in part or its abstract on social media accounts closely related to the Journal, and the right to alter the article to the extent necessary for such use.
>
> Authors may self-archive the Authors accepted manuscript of their articles on their own websites. Authors may also deposit this version of the article in any repository, provided it is only made publicly available 12 months after official publication or later. He/she may not use the publisher's version (the final article), which is

posted on SpringerLink and other Springer websites, for the purpose of self-archiving or deposit. Furthermore, the Author may only post his/her version provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be provided by inserting the DOI number of the article in the following sentence: "The final publication is available at Springer via http://dx.doi.org/[insert DOI]".

Prior versions of the article published on non-commercial pre-print servers like arXiv.org can remain on these servers and/or can be updated with Author's accepted version. The final published version (in pdf or html/xml format) cannot be used for this purpose. Acknowledgement needs to be given to the final publication and a link must be inserted to the published article on Springer's website, by inserting the DOI number of the article in the following sentence: "The final publication is available at Springer via http://dx.doi.org/[insert DOI]". **Author retains the right to use his/her article for his/her further scientific career by including the final published journal article in other publications such as dissertations and postdoctoral qualifications provided acknowledgement is given to the original source of publication.**

Articles disseminated via link.springer.com are indexed, abstracted and referenced by many abstracting and information services, bibliographic networks, subscription agencies, library networks, and consortia.

After submission of the agreement signed by the corresponding author, changes of authorship or in the order of the authors listed will not be accepted by Springer.

9

BIBLIOGRAPHY

Amprion GmbH (2016), 'Wind energy grid feed-in'. Available from `http://www.amprion.net/en/wind-feed-in`.

Ancona-Navarrete, M. A. and Tawn, J. A. (2000), 'A Comparison of Methods for Estimating the Extremal Index', *Extremes* **3**(1), 5–38.

Andrieu, C., Doucet, A. and Holenstein, R. (2010), 'Particle Markov chain Monte Carlo methods', *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **72**(3), 269–342.

Andrieu, C. and Roberts, G. O. (2009), 'The pseudo-marginal approach for efficient Monte Carlo computations', *The Annals of Statistics* **37**(2), 697–725.

Antoniak, C. E. (1974), 'Mixtures of Dirichlet processes with Applications to Bayesian Nonparametric Problems', *The Annals of Statistics* **2**(6), 1152–1174.

Applebaum, D. (2009), *Lévy Processes and Stochastic Calculus*, second edn, Cambridge University Press.

Asmussen, S. and Glynn, P. (2010), 'Harris recurrence and MCMC: A simplified approach', *Thiele Research Reports* (06).

Barndorff-Nielsen, O. E. (1997), 'Processes of normal inverse Gaussian type', *Finance and Stochastics* **2**(1), 41–68.

Barndorff-Nielsen, O. E. (2001), 'Superposition of Ornstein–Uhlenbeck type processes', *Theory of Probability & Its Applications* **45**(2), 175–194.

Barndorff-Nielsen, O. E. (2011), 'Stationary Infinitely Divisible Processes', *Brazilian Journal of Probability and Statistics* **25**(3), 294–322.

Barndorff-Nielsen, O. E., Benth, F. E. and Veraart, A. E. D. (2011), Ambit Processes and Stochastic Partial Differential Equations, *in* G. Di Nunno and B. Øksendal, eds, 'Advanced Mathematical Methods for Finance', Springer Berlin Heidelberg, pp. 35–74.

Barndorff-Nielsen, O. E., Benth, F. E. and Veraart, A. E. D. (2015), 'Recent advances in ambit stochastics with a view towards tempo-spatial stochastic volatility/intermittency', *Banach Center Publications* **104**, 25–60.

Barndorff-Nielsen, O. E., Lunde, A., Shephard, N. and Veraart, A. E. D. (2014), 'Integer-valued Trawl Processes: A Class of Stationary Infinitely Divisible Processes', *Scandinavian Journal of Statistics* **41**(3), 693–724.

Barndorff-Nielsen, O. E. and Schmiegel, J. (2007), Ambit processes: with applications to turbulence and cancer growth, *in* F. E. Benth, G. Di Nunno, T. Lindstrøm, B. Øksendal and T. Zhang, eds, 'Stochastic Analysis and Applications: The Abel Symposium 2005.', Springer, Heidelberg, pp. 93–124.

Barndorff-Nielsen, O. E. and Shephard, N. (2001), 'Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics', *Journal of the Royal Statistical Society - Series B: Statistical Methodology* **63**(2), 167–241.

Basse-O'Connor, A., Graversen, S.-E. and Pedersen, J. (2014), 'Stochastic integration on the real line', *Theory of Probability and Its Applications* **58**(2), 193–215.

Bennedsen, M., Lunde, A. and Pakkanen, M. S. (2014), 'Discretization of Levy semistationary processes with application to estimation', *arXiv:1407.2754* .

Benth, F. and Šaltytė Benth, J. (2013), *Modeling and pricing in financial markets for weather derivatives*, first edn, World Scientific Publishing.

Bortot, P. and Gaetan, C. (2014), 'A Latent Process Model for Temporal Extremes', *Scandinavian Journal of Statistics* **41**(3), 606–621.

Bortot, P. and Tawn, J. A. (1998), 'Models for the extremes of Markov chains', *Biometrika* **85**(4), 851–867.

Brix, A. (1999), 'Generalized gamma measures and shot-noise Cox processes', *Advances in Applied Probability* **31**(4), 929–953.

Brockwell, P. J. (2001), 'Lévy-Driven CARMA Processes', *Annals of the Institute of Statistical Mathematics.* **53**(1), 113–124.

Brockwell, P. J. (2004), 'Representations of continuous-time ARMA processes', *Journal of Applied Probability* **41**(2004), 375–382.

Brockwell, P. J. and Lindner, A. M. (2009), 'Existence and uniqueness of stationary Lévy-driven CARMA processes', *Stochastic Processes and their Applications* **119**(8), 2660–2681.

Brockwell, P. J. and Lindner, A. M. (2013), 'Integration of CARMA processes and spot volatility modelling', *Journal of Time Series Analysis* **34**(2), 156–167.

Carmona, R. and Diko, P. (2005), 'Pricing Precipitation Based Derivatives', *International Journal of Theoretical and Applied Finance* **08**(07), 959–988.

Chandler, R. E. (1997), 'A Spectral Method for Estimating Parameters in Rainfall Models', *Bernoulli* **3**(3), 301–322.

Chin, E. H. (1977), 'Modeling daily precipitation occurrence process with Markov Chain', *Water Resources Research* **13**(6), 949–956.

Coe, R. and Stern, R. (1982), 'Fitting models to daily rainfall data', *Journal of Applied Meteorology* **21**(7), 1024–1031.

Cohn, D. L. (1990), *Measure theory*, Birkhäuser Boston.

Coles, S. G. and Tawn, J. A. (1991), 'Modelling Extreme Multivariate Events', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **53**(2), 377–392.

Coles, S., Heffernan, J. and Tawn, J. A. (1999), 'Dependence Measures for Extreme Value Analyses', *Extremes* **2**(4), 339–365.

Cont, R. and Tankov, P. (2004), *Financial modelling with jump processes*, Chapman & Hall/CRC.

Cowles, M. K. and Carlin, B. P. (1996), 'Markov chain Monte Carlo convergence diagnostics: A comparative review', *J Am Stat Assoc* **91**(434), 883–904.

Cowpertwait, P. S. P. (1994), 'A Generalized Point Process Model for Rainfall', *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **447**(1929), 23–37.

Cox, D. R. and Isham, V. (1988), 'A simple spatial-temporal model of rainfall', *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **415**(1849), 317–328.

Cox, D. R. and Reid, N. (2004), 'A note on pseudolikelihood constructed from marginal densities', *Biometrika* **91**(3), 729–737.

Davison, A. C., Padoan, S. A. and Ribatet, M. (2012), 'Statistical Modeling of Spatial Extremes', *Statistical Science* **27**(2), 161–186.

Davison, A. C. and Smith, R. L. (1990), 'Models for exceedances over high thresholds', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **52**(3), 393–442.

de Haan, L. and de Ronde, J. (1998), 'Sea and Wind: Multivariate Extremes at Work', *Extremes* **1**(1), 7–45.

Dunn, P. K. (2004), 'Occurrence and quantity of precipitation can be modelled simultaneously', *International Journal of Climatology* **24**(10), 1231–1239.

Embrechts, P., Koch, E. and Robert, C. (2015), 'Space-time max-stable models with spectral separability', *arXiv:1507.07750* .

Embrechts, P., Mikosch, T. and Klüppelberg, C. (1997), *Modelling extremal events for insurance and finance*, Springer Verlag, Berlin.

Environmental Research Group, King's College London (2015), 'London Air Quality Network database'. Available from `http://www.londonair.org.uk`.

Esche, F. and Schweizer, M. (2005), 'Minimal entropy preserves the Lévy property: how and why', *Stochastic Processes and their Applications* **115**(2), 299–327.

Esscher, F. (1932), 'On the probability function in the collective theory of risk', *Scandinavian Actuarial Journal* **15**(3), 175–195.

Ferguson, T. S. (1973), 'A Bayesian Analysis of Some Nonparametric Problems', *The Annals of Statistics* **1**(2), 209–230.

Ferro, C. A. T. and Segers, J. (2003), 'Inference for clusters of extreme values', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**(2), 545–556.

Frittelli, M. (2000), 'The minimal entropy martingale measure and the valuation problem in incomplete markets', *Mathematical Finance* **10**(1), 39–52.

Fuchs, F. and Stelzer, R. (2013), 'Mixing conditions for multivariate infinitely divisible processes with an application to mixed moving averages and the supOU stochastic volatility model', *ESAIM: Probability and Statistics* **17**, 455–471.

Gelfand, A. E., Sahu, S. K. and Carlin, B. P. (1995), 'Efficient parametrizations for normal linear mixed models', *Biometrika* **82**(3), 479–488.

Gerber, H. and Shiu, E. (1994), 'Option pricing by Esscher transforms', *Transactions of the Society of Actuaries* **46**, 99–191.

Green, P. (2002), 'Trans-dimensional markov chain monte carlo'. Available from `http://www.maths.bris.ac.uk/~mapjg/slides/tdtut4.pdf`.

Green, P. J. (1995), 'Reversible jump Markov chain Monte Carlo computation and Bayesian model determination', *Biometrika* **82**(4), 711–732.

Grigelionis, B. (2011), 'On the Hougaard subordinated Gaussian Lévy processes', *Statistics & Probability Letters* **81**(8), 998–1002.

Härdle, W. K. and Cabrera, B. L. (2012), 'The Implied Market Price of Weather Risk', *Applied Mathematical Finance* **19**(1), 59–95.

Hedevang, E. and Schmiegel, J. (2014), 'A Lévy Based Approach to Random Vector Fields: With a View Towards Turbulence', *International Journal of Nonlinear Sciences and Numerical Simulation* **15**(7-8), 411–435.

Heffernan, J. E. and Tawn, J. A. (2004), 'A conditional approach for multivariate extreme values', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66**(3), 497–546.

Hellmund, G. (2009), 'Completely random signed measures', *Statistics and Probability Letters* **79**(7), 894–898.

Hsing, T., Hüsler, J. and Leadbetter, M. (1988), 'On the Exceedance Point Process for a Stationary Sequence', *Probability Theory and Related Fields* **78**(1), 97–112.

Huser, R. and Davison, A. C. (2014), 'Space-time modelling of extreme events', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**(2), 439–461.

Jónsdóttir, K. Ý., Rønn-Nielsen, A., Mouridsen, K. and Jensen, E. B. V. (2011), Lévy based modelling in brain imaging. Available from `http://pure.au.dk/portal/files/52568744/imf_csgb_2011_02.pdf`.

Jónsdóttir, K. Ý., Schmiegel, J. and Jensen, E. B. V. (2008), 'Lévy-based growth models', *Bernoulli* **14**(1), 62–90.

Jordan, M. (2010), Hierarchical models, nested models and completely random measures, *in* M.-H. Chen, P. Müller, D. Sun, K. Ye and D. Dey, eds, 'Frontiers of Statistical Decision Making and Bayesian Analysis: in Honor of James O. Berger', Springer-Verlag New York, pp. 207–216.

Jørgensen, B. (1997), *The Theory of Dispersion models*, Chapmann & Hall.

Kallenberg, O. (1976), *Random Measures*, Academic Press Inc. (London) Ltd.

Katz, R. W. (1977), 'Precipitation as a chain-dependent process', *Journal of Applied Meteorology* **16**(7), 671–676.

Kingman, J. F. C. (1967), 'Completely random measures', *Pacific Journal Of Mathematics* **21**(1), 59–78.

Kou, S. C., Xie, X. S. and Liu, J. S. (2005), 'Bayesian analysis of single-molecule experimental data', *Journal of the Royal Statistical Society. Series C: Applied Statistics* **54**(3), 469–506.

Künsch, H. R. (1989), 'The Jackknife and the Bootstrap for General Stationary Observations', *The Annals of Statistics* **17**(3), 1217–1241.

Leadbetter, M. (1991), 'On a basis for Peaks over Threshold modeling', *Statistics & Probability Letters* **12**(4), 357–362.

Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983), *Extremes and Related Properties of Random Sequences and Processes*, Springer Verlag, New York.

Ledford, A. W. and Tawn, J. A. (1996), 'Statistics for near independence in multivariate extreme values', *Biometrika* **83**(1), 169–187.

Ledford, A. W. and Tawn, J. A. (1997), 'Modelling dependence within joint tail regions.', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **59**(2), 475–499.

Ledford, A. W. and Tawn, J. A. (2003), 'Diagnostics for dependence within time series extremes', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**(2), 521–543.

Lee, M. L. T. and Whitmore, G. A. (1993), 'Stochastic processes directed by randomized time', *Journal of Applied Probability* **30**(2), 302–314.

Leobacher, G. and Ngare, P. (2011), 'On modelling and pricing rainfall derivatives with seasonality', *Applied Mathematical Finance* **18**(1), 71–91.

Lewis, A. L. (2001), A simple option formula for general jump-diffusion and other exponential Lévy processes. Available from `http://optioncity.net/pubs/ExpLevy.pdf`.

Lindgren, F., Rue, H. and Lindstrøm, J. (2011), 'An explicit link between Gaussian fields and Gaussian Markov random fields: The SPDE approach', *Journal of the Royal Statistical Society Series B* **73**, 423–498.

Liu, J. S. and Sabatti, C. (2000), 'Generalised Gibbs sampler and multigrid Monte Carlo for Bayesian computation', *Biometrika* **87**(2), 353–369.

Liu, J. S., Wong, W. H. and Kong, A. (1994), 'Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes', *Biometrika* **81**(1), 27–40.

López Cabrera, B., Odening, M. and Ritter, M. (2013), 'Pricing rainfall futures at the CME', *Journal of Banking & Finance* **37**(11), 4286–4298.

Matérn, B. (1960), 'Spatial variation: Stochastic models and their application to some problems in forest surveys and other sampling investigations', *Meddelanden fran Statens Skogsforskningsinstitut* **49**(5).

Møller, J. and Torrisi, G. L. (2005), 'Generalised shot noise Cox processes', *Advances in Applied Probability* **37**(1), 48–74.

Noven, R. C., Veraart, A. E. D. and Gandy, A. (2015), 'A Levy-driven rainfall model with applications to futures pricing', *AStA Advances in Statistical Analysis* (99), 403–432.

O'Brien, G. L. (1987), 'Extreme Values for Stationary and Markov Sequences', *The Annals of Probability* **15**(1), 281–291.

Øksendal, B. (2000), *Stochastic Differential Equations: an introduction with applications*, second edn, Springer-Verlag Berlin.

Onof, C., Chandler, R. E., Kakou, A., Northrop, P., Wheater, H. S. and Isham, V. (2000), 'Rainfall modelling using Poisson-cluster processes: a review of developments', *Stochastic Environmental Research and Risk Assessment* **14**(6), 384 – 411.

Papaspiliopoulos, O. and Roberts, G. O. (2003), 'Non-Centered Parameterisations for Hierarchical Models and Data Augmentation', *Bayesian Statistics* **7**, 307–326.

Pedersen, J. (2003), The Lévy-Itô decomposition of an independently scattered random measure. Available from `http://www.maphysto.dk/publications/MPS-RR/2003/2.pdf`.

Politis, D. N. and Romano, J. P. (1994), 'The Stationary Bootstrap', *Journal of the American Statistical Association* **89**(428), pp. 1303–1313.

Protter, P. (2005), *Stochastic Integration and Differential Equations*, second edn, Version 2.1, Springer-Verlag Berlin.

Rajput, B. S. and Rosinski, J. (1989), 'Spectral representations of infinitely divisible processes', *Probability Theory and Related Fields* **82**(3), 451–487.

Robbins, H. and Monro, S. (1951), 'A Stochastic Approximation Method', *The Annals of Mathematical Statistics* **22**(3), 400–407.

Roberts, G. O., Gelman, A. and Gilks, W. R. (1997), 'Weak convergence and optimal scaling of random walk Metropolis algorithms', *Annals of Applied Probability* **7**(1), 110–120.

Rodriguez-Iturbe, I., Cox, D. R. and Isham, V. (1987), 'Some models for rainfall based on stochastic point processes', *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **410**(1839), 269–288.

Samorodnitsky, G. and Taqqu, M. S. (1994), *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapman and Hall/CRC.

Samuel, C. R. (1999), Stochastic rainfall modelling of convective storms in Walnut Gulch, Arizona., PhD thesis, Imperial College London.

Sato, K. I. (1999), *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press.

Smith, R. L., Tawn, J. A. and Coles, S. G. (1997), 'Markov chain models for threshold exceedances', *Biometrika* **84**(2), 249–268.

Tanner, M. and Wong, W. H. (1987), 'The Calculation of Posterior Distributions by Data Augmentation', *Journal of the American Statistical Association* **82**(398), 528–540.

Tsai, H. and Chan, K. S. (2005), 'A note on non-negative continuous time processes', *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **67**(4), 589–597.

UK Meteorological Office (2012), 'Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853-current)'. NCAS British Atmospheric Data Centre. Available from `http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__dataent_ukmo-midas`.

Varin, C. (2008), 'On composite marginal likelihoods', *AStA Advances in Statistical Analysis* **92**(1), 1–28.

Varin, C., Reid, N. and Firth, D. (2011), 'An overview of composite likelihood methods', *Statistica Sinica* **21**, 5–42.

Wallin, J. and Bolin, D. (2015), 'Geostatistical Modelling Using Non-Gaussian Matérn Fields', *Scandinavian Journal of Statistics* **42**(3), 872–890.

Wilks, D. (1998), 'Multisite generalization of a daily stochastic precipitation generation model', *Journal of Hydrology* **210**(1), 178–191.

Wolpert, R. L. and Taqqu, M. S. (2005), 'Fractional Ornstein-Uhlenbeck Lévy processes and the Telecom process: Upstairs and downstairs', *Signal Processing* **85**(8), 1523–1545.

Woolhiser, D. A. and Roldán, J. (1982), 'Stochastic daily precipitation models: 2. A comparison of distributions of amounts', *Water Resources Research* **18**(5), 1461–1468.

Yan, J., Cowles, M. K., Wang, S. and Armstrong, M. P. (2007), 'Parallelizing MCMC for Bayesian spatiotemporal geostatistical models', *Statistics and Computing* **17**(4), 323–335.

LIST OF SYMBOLS

This list shows selected symbols that recur in the text.

δ	Time increment for rainfall model	10
(μ, ρ, κ)	Parameters of Tweedie distribution	11
$\triangle Y(t_i)$	Discrete-time rainfall process	12
$a(x)$	AR polynomial of CARMA process	13
$b(x)$	MA polynomial of CARMA process	13
$\{w_i, \lambda_i\}$	Parameters of CARMA process	15
S	Domain of random measure L	44
λ^{Leb}	Lebesgue measure	44
L	Lévy basis	46
L'	Lévy seed	48
Λ_t	Trawl process	52
A_t	Trawl set	52
$d_{exp}(s)$	Exponential trawl set function	53
$C(u; X)$	Cumulant function of X	53
$\mathcal{T}(t_i)$	Slices in A_{t_i}	57
$\mathcal{K}_s$	Cross-section of trawl set	60
$(\rho, \alpha, \beta, \kappa)$	Exceedance model parameters	75
$(\rho, \kappa, \xi, \sigma)$	Transformed exceedance model parameters	79
$\triangle$	Separation in pairwise likelihood	82
$\varphi(u_1, u_2)$	Conditional tail dependence coefficient	84
θ_y	Observational distribution parameters	96
θ_s	Lévy seed distribution parameters	97
θ_t	Trawl set distribution parameters	97
S_i	Slice variable and intersection set	105
$g_u^F(\theta, x)$	Quantile move	114
c_i	Cumulant scaling factor	114