



BAYESIAN MIXTURE MODEL OF \mathcal{R} -PARETO PROCESS FOR EXTREME PRECIPITATION

Haitao Gao

Supervisor: Dr. Boris Béranger, Dr. Peng Zhong and Professor Scott Sisson

School of Mathematics and Statistics
UNSW Sydney

November 2025

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF THE DEGREE OF
MASTER OF MATHEMATICS

Plagiarism statement

I declare that this thesis is my own work, except where acknowledged, and has not been submitted for academic credit elsewhere.

I acknowledge that the assessor of this thesis may, for the purpose of assessing it:

- Reproduce it and provide a copy to another member of the University; and/or,
- Communicate a copy of it to a plagiarism checking service (which may then retain a copy of it on its database for the purpose of future plagiarism checking).

I certify that I have read and understood the University Rules in respect of Student Academic Misconduct, and am aware of any potential plagiarism penalties which may apply.

By signing this declaration I am agreeing to the statements and conditions above.

Signed: _____ Date: _____

Acknowledgements

First and foremost, I am deeply grateful to my supervisor, Dr. Boris Béranger, for giving me the opportunity to undertake a research level project in statistics despite my limited formal background in mathematics. Throughout this two terms thesis journey, his steady guidance and support made the work possible. In our weekly meetings, he consistently identified where my arguments lacked rigor or precision, and his timely feedback helped me improve quickly and meaningfully.

I am also profoundly thankful to my co-supervisors, Dr. Zhong and Prof. Scott Sisson, whose patience, encouragement, and kindness sustained me throughout this process. Scott's wonderful sense of humour brought joy to our meetings and made each week's progress feel lighter. Dr. Zhong offered exceptionally detailed, practical advice clarifying the intuition behind mathematical formulations and helping with implementation which made the research both smoother and more enjoyable.

I am grateful to UNSW for its flexible curriculum and rich academic environment. That flexibility allowed me to step into the world of mathematics through courses like Bayesian Inference and Discrete Mathematics. At UNSW, I found generous mentors, inspiring peers, and supportive supervisors who encouraged me to pursue the problems that interested me most. Over the past two years, I have grown to be more rigorous, ethical, and critical in my academic work, and more open minded and outgoing in life.

Finally, my deepest thanks go to my parents. Studying in Sydney, one of the most expensive cities in the world, would not have been possible without their financial support and quiet faith in me. Their encouragement allowed me to focus on learning rather than worrying about expenses, and they have stood by my decisions to pursue a PhD and to stay in Australia in the near future. Like many Asian parents, their love is never spoken, but I feel it in every opportunity they have helped create. This thesis is as much theirs as it is mine.

Abstract

Extreme Value Theory (EVT) has been an research area of rapid growth in the last decade, which is a crucial tool for environment study. Extreme precipitation modelling is highly relied on EVT to describe the intensity and dependence structure of extreme rainfall events. \mathcal{R} -Pareto process is one of the state-of-art theory frameworks that is actively studied recently to model the extremes associated with the theoretical risk function $r(\cdot)$ that is typically from L_p norm risk function family, which provides flexibility to define extremes.

This thesis develops a Bayesian mixture model of the \mathcal{R} -Pareto process to jointly represent and classify two physically distinct types of extreme rainfall: frontal and convective events. The mixture formulation enables simultaneous inference on type membership and parameters of the dependence model providing a principled approach to both modelling and categorisation of extremes.

Through simulation studies, the proposed method accurately recovers range parameters with a fixed smoothness across scenarios with clearly separated as well as partially overlapping parameter regimes. A dependence measure diagnostic, $\chi^2(S_A, S_B)$, computed based on inferred event assignments, confirms that the two classes remain statistically distinguishable under both regimes.

Applied to extreme rainfall in Tampa Bay, Florida, USA, the model yields precise estimates of spatial range parameters and robust event-type classifications. The inferred seasonal and dynamical patterns align closely with established findings in the literature and with U.S. government geological reports.

Overall, this work extends the scope of the \mathcal{R} -Pareto framework by introducing a Bayesian mixture formulation that delivers accurate parameter estimation while flexibly capturing multiple, co-occurring extremal phenomena within a unified spatial model.

Contents

Chapter 1	Introduction	1
Chapter 2	Literature Review and Introduction to Extreme Value Theory	3
2.1	Generalised Extreme Value	3
2.2	Generalised Pareto Distribution	5
2.3	Point Process Representation	6
2.4	Max Stable Process	9
2.5	Brown-Resnick Model	10
2.6	\mathcal{R} -Pareto Process	10
Chapter 3	Bayesian Mixture Model	12
3.1	Brown-Resnick Process and Semivariogram on a Finite Grid	12
3.2	\mathcal{R} -Pareto Process on a Grid and L_1 Risk Function	13
3.3	Likelihood Function	13
3.4	Bayesian Mixture Model	14
3.4.1	Variables	14
3.4.2	Mixture Joint Distribution	15
3.4.3	Prior Distribution	15
3.4.4	Mixture Joint Distribution	17
3.4.5	Mixture Posterior Distribution	17
3.4.6	Monte Carlo Markov Chain for Fitting Model	17
Chapter 4	Simulation Study	19
4.1	Spatial Grid World	19
4.2	Distinguishable Parameters Simulation Study	19
4.3	Indistinguishable Parameters Simulation Study	24
4.4	Posterior Concentration Under Increasing Sample Size	27
4.5	Time Variant Weight Function Simulation	28
Chapter 5	Application	32
5.1	Dataset	32
5.2	Results	33
Chapter 6	Discussion	39
Chapter 7	Conclusion	41
	Appendices	43

Appendix A	Simulated Data Generation Algorithms	44
Appendix B	Bayesian MCMC in Log Space	46
B.1	Log Likelihood Function	46
B.2	Log Prior Distribution	46
B.3	Log Parameterisation	47
B.4	Log Conditional Updates	47

CHAPTER 1

Introduction

Extreme Value Theory (EVT) has emerged as one of important statistic disciplines for applied science in last eighty years. Its diverse range of modern applications has resulted in extensive development of theory in a variety of areas such as finance ([Candia and Herrera, 2024](#); [Fuentes *et al.*, 2025](#)), engineering ([Weng *et al.*, 2024](#)) and environmental science ([Hazra *et al.*, 2025](#); [Vidrio-Sahagún *et al.*, 2025](#)). The distinguish feature of EVT is the objective to quantify the stochastic behaviour of a process at unusually large or small levels. EVT explores the asymptotic distribution of the extremely large or small values, i.e. the tail distribution of a given probability function. Many Extreme Value models have been proposed to model the univariate extremes. Generalised Extreme Value (GEV) distribution is a model of maxima of large block of data ([Jenkinson, 1955](#); [De Haan and Ferreira, 2006](#)). However, GEV does not always perform well as taking block maxima leads to limited data. Generalised Pareto Distribution (GPD) is proposed to model all points over one specific threshold rather than maxima value in each block, which brings more usable data points to fit the model([Balkema and De Haan, 1974](#); [Pickands III, 1975](#); [De Haan and Ferreira, 2006](#)). The Point Process Representation (PP) is an another theory to represent the extremes. In fact, the GEV and GPD are the special cases of a Poisson Point Process (PPP) to approximate the occurrence of those values that exceed a high threshold([Ripley, 1976](#); [De Haan and Ferreira, 2006](#)).

Modelling multivariate extremes, especially the dependence structure in spatial extremes is an active area of research, with the max-stable process ([De Haan, 1984](#); [Schlather, 2002](#)) being the predominant theoretical framework. When extreme events are defined as maxima recorded over long periods, max-stable processes offer a broad class of models suitable for asymptotically dependent data. Among these, the Brown–Resnick model ([Brown and Resnick, 1977](#); [Kabluchko *et al.*, 2009](#)) is one of the most popular, due to its desirable Gaussian properties. Just as the max-stable process is a generalisation of GEV in multivariate setting, \mathcal{R} -Pareto process is the analog of GPD in multivariate setting, which is a relatively recent alternative spatial modelling framework defines extreme events as the realisations of a process for which a risk functional exceeds some high threshold([Ferreira and De Haan, 2014](#); [de Fondeville and Davison, 2018, 2022](#)) sharing the same theoretical foundations as max-stable processes. Most spatial extreme models are initially constructed as max-stable processes from their spectral representation, and their \mathcal{R} -Pareto equivalents are then derived.

However, these parametric approaches for asymptotic models face major inferential hurdles and are mostly based on the Brown–Resnick model, especially in realistic applications in high dimensions. For example, environmental processes are often observed over large spatial regions and with fine resolution, and having the dimension D in the order of thousands is common. The likelihood function of such max-stable process models is intractable in high dimensions unless low-dimensional approximations are used, such as composite likelihood (Padoan *et al.*, 2010; Whitaker *et al.*, 2020) and Vecchia approximation (Huser *et al.*, 2024), which are suitable for dimension up to 1000. However, the likelihood for \mathcal{R} -Pareto process is more tractable than for max-stable but still requires evaluation of an intractable normalising constant, which can be avoided through score match (de Fondeville and Davison, 2018; Lederer and Oesting, 2023) or choose special risk functions such as L_1 and L_∞ . Motivated by this mathematical trick, Zhong *et al.* (2024) inferences any L_p risk function by simply maximising the likelihood of \mathcal{R} -Pareto process with L_1 risk function and a large threshold, which provide a simpler and efficient inference framework than score match approach.

Although there are various research paper tackling inference difficulties in multivariate setting, they only focus on a single type of extreme when fitting and inferencing the model. However, real world applications always contain multiple types of extremes simultaneously. For example, the rainfall in a region includes extremes events as locally intense rainfall events at some locations and events with high cumulative rainfall over the whole region. There is a huge blank in the area of jointly modelling of multiple type of extremes. The purpose of this thesis, therefore, is to develop a systematic framework to model multiple extremes jointly. In particular, the proposed statistic model should model frontal and convective extreme rainfall jointly. And it can identify the categorised labels based on a latent variable.

To complete this, the thesis will explore what likelihood function and risk function might be required, and what statistic model is able to model two types of extremes jointly given the dataset. In addition, the thesis will also evaluate the performance of the proposed model in classification to check whether the inferred the membership indeed reflect the different spatial dependency.

The following thesis is organised as Chapter 2 will provide an overview of EVT, as well as a review of recent developments and the state of art theories in the modern literature. Chapter 3 and Chapter 4 will demonstrate the Bayesian mixture model derived from standard \mathcal{R} -Pareto process and verification on simulated dataset. Chapter 5 will apply this proposed model to real world rainfall dataset and show the results. Chapter 6 will provide a comprehensive discussion of this proposed model and performance on simulated and real dataset. Chapter 7 will make a conclusion of this thesis and describe where further developments or theories could be made.

CHAPTER 2

Literature Review and Introduction to Extreme Value Theory

This chapter will provide a literature review and introduction to Extreme Value Theory. It introduces the Extreme Value Theory from the univariate to multivariate and explains its strengths and weaknesses. Extreme Value Theory is a study of events which occur rarely but potentially bring high impacts. It has been highly used in modern applications such as financial risk modelling([Chaudhry et al., 2022](#)), climate change([Tabari, 2021](#)) and healthcare([Karpov et al., 2022](#)), which motivates researchers to develop more flexible, efficient and accurate Extreme Value Theory to satisfy the high demands in the real world. This chapter will summarise these fundamental theorems of Extreme Value Theory.

2.1 Generalised Extreme Value

Generalised Extreme Value (GEV) framework is built based on block maxima. The block maxima method involves dividing the total observation period into individual non-overlapping periods, where each period consists of the same number of observations, which we will call n . Let the sequence X_1, X_2, \dots, X_n denote any of these observation periods. Here, the sequence X_1, X_2, \dots, X_n are assumed to be independent and identically distributed (i.i.d) random variable with common distribution function F . In applications, X_n represents separate observations, such as hourly rainfall, sea level and stock prices. The maximum value of this sequence is denoted as

$$M_n = \max(X_1, X_2, \dots, X_n) \quad (2.1.1)$$

Assuming that n is finite and distribution function F is known, the distribution of block maximal is

$$\begin{aligned} F_{M_n}(x) &= \mathbb{P}(M_n \leq x) \\ &= \mathbb{P}(\max(X_1, X_2, \dots, X_n) \leq x) \\ &= \mathbb{P}(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) \\ &= \mathbb{P}(X \leq x)^n \\ &= F^n(X) \end{aligned}$$

However, the distribution function F of random variable X is rarely known in most of cases. Fortunately, Central Limit Theorem (CLT) would be helpful when distribution function F is unknown.

Definition 2.1.1. Let $X_i, X, n \geq 1$ be real-valued random variables with distribution function F_n and F . The sequence X_n converges in distribution (weakly) to X if and only if for all continuity point x of F

$$F_n(x) \rightarrow F(x), \quad \text{as } n \rightarrow \infty$$

Assume that there exists sequences of constant $a_n > 0, b_n \in \mathbb{R}$, and a non-degenerate random variable with distribution function G , such that

$$M_n^* \xrightarrow{d} Y$$

The distribution function G is called Extreme Value Distribution.

Theorem 2.1.2 (Fisher and Tippett (1928); Gnedenko (1943)). *The only possible limiting distribution G , up to location-scale transformation, is the Generalised Extreme Value (GEV) distribution, i.e*

$$\mathbb{P}(Y \leq y) := G_\xi(y) = \begin{cases} \exp \left\{ -[1 + \xi y]_+^{-1/\xi} \right\} & \xi \neq 0 \\ \exp \{-\exp\{-y\}\} & \xi = 0 \end{cases}, \quad \forall y \in \mathbb{R} \quad (2.1.2)$$

where $z_+ = \max(0, z)$.

The GEV distribution is characterised by its shape parameter $\xi \in \mathbb{R}$, also known as the extreme value index, unifying three limiting distribution.

Definition 2.1.3 (Type I Gumbel Distribution).

$$\Lambda(y) = G_0(y), y \in \mathbb{R}. \quad (2.1.3)$$

Definition 2.1.4 (Type II Fréchet Distribution).

$$\Phi_\alpha(y) = \begin{cases} 0 & y \leq 0, \\ \exp \{-y^{-\alpha}\} & y > 0. \end{cases} \quad (2.1.4)$$

It corresponds to $G_\xi((y-1)/\xi)$ with $\xi = 1/\alpha$

Definition 2.1.5 (Type III Weibull Distribution).

$$\psi(y) = \begin{cases} \exp \{-(-y)^\alpha\} & y \leq 0, \\ 1 & y > 0. \end{cases} \quad (2.1.5)$$

It corresponds to $G_\xi(-(y+1)/\xi)$ with $\xi = -1/\alpha < 0$

With a location $\mu \in \mathbb{R}$ and a scale $\sigma > 0$ parameter completing the parameter vector $\theta = (\mu, \sigma, \xi)$, the distribution function of GEV becomes:

$$G(y|\theta) = \begin{cases} \exp \left\{ - \left[1 + \xi \left(\frac{y-\mu}{\sigma} \right) \right]_+^{-1/\xi} \right\} & \xi \neq 0 \\ \exp \left\{ - \exp \left\{ - \frac{y-\mu}{\sigma} \right\} \right\} & \xi = 0 \end{cases}, \quad \forall y \in \mathbb{R} \quad (2.1.6)$$

Figure 2.1 shows the density of Gumbel Distribution, Fréchet Distribution and Weibull Distribution respectively.

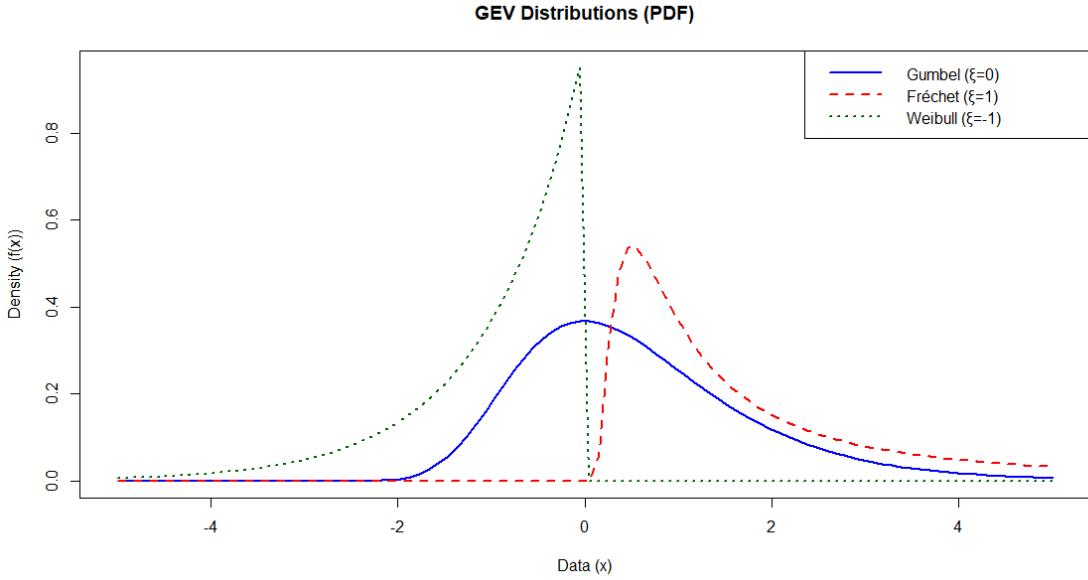


Figure 2.1: Probability Density Functions of the Gumbel, Fréchet, and Weibull Distributions. These three families are unified by the Generalised Extreme Value (GEV) distribution, where the **shape parameter**, ξ , governs the tail behaviour: the **Gumbel** ($\xi = 0$) is unbounded with an exponential tail; the **Fréchet** ($\xi = 1$) is heavy-tailed with a lower bound; and the **Weibull** ($\xi = -1$) has a finite upper bound.

2.2 Generalised Pareto Distribution

Generalised Pareto Distribution is developed based on the exceedances of a high threshold rather than the block maxima. Modelling only block maxima can be wasteful if other data on extremes are available. For example, there are more extremes in one block but only the maximum is picked but there is no extremes in another block but maximum of them is also picked. Therefore, defining extreme event based on a high threshold is naturally proposed, which provides the analysts more usable data to fit the model.

Assuming that X is a random variable with right end point x^* . For some high threshold $u < x^*$, let Y_1, \dots, Y_n represent the exceedances of u by $X_1 \dots X_n$ with

$$N_u = \text{card}\{i : i = 1, \dots, n, X_i > u\}$$

The exceedance distribution function of the random variable X (with df F) is given by:

$$F_u(y) = \mathbb{P}(X - u \leq y \mid X > u) = \mathbb{P}(Y \leq y \mid X > u), y > 0 \quad (2.2.1)$$

Theorem 2.2.1 (Pickands III (1975)). *The conditional distribution of the exceedances converges to the generalised Pareto distribution (GPD) as $u \rightarrow \infty$ under the same conditions as those for the convergence of sample maxima. The df of GPD*

is defined by:

$$H_\xi(y) = \begin{cases} 1 - (1 + \xi y)^{-1/\xi} & \xi \neq 0, \\ 1 - \exp\{-y\} & \xi = 0, \end{cases} \quad (2.2.2)$$

$$\text{where domain}(Y) = \begin{cases} y \geq 0 & \xi \geq 0, \\ 0 \leq y \leq -1/\xi & \xi < 0. \end{cases}$$

when $\xi > 0$, we obtain the ordinary Pareto distribution, when $\xi = 0$, we obtain the Exponential distribution and when $\xi < 0$, we obtain the short-tail Pareto distribution as shown in Figure 2.2

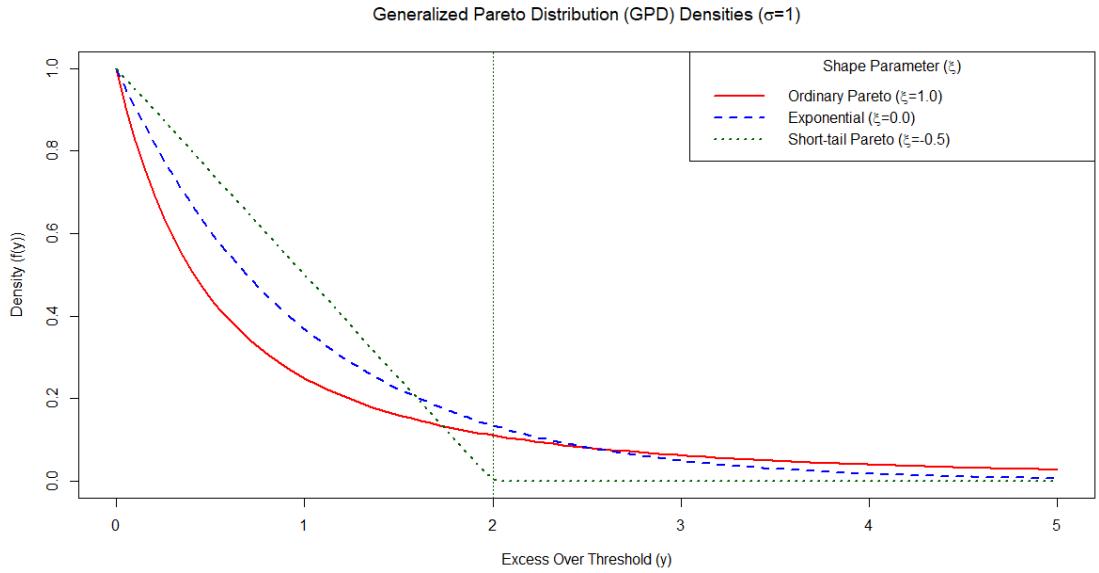


Figure 2.2: Probability Density Functions of the Generalised Pareto Distribution (GPD) for different values of the shape parameter ξ , with a fixed scale $\sigma = 1$. The shape parameter determines the tail behaviour: $\xi > 0$ yields the heavy-tailed Ordinary Pareto distribution; $\xi = 0$ yields the Exponential distribution; and $\xi < 0$ yields the Short-tail Pareto distribution, which has a finite upper limit at $y_{max} = -\sigma/\xi$.

2.3 Point Process Representation

Think of a point process simply as a random distribution of point X_i in space. Let E denote the state space where the points live and $\{X_i, i > 1\}$ be a countable collection of points in E . A point process is characterised by a set of non-negative integer valued random variables $N(A)$ such that

$$N(A) = \text{card}\{i : X_i \in A\}, \text{ for } A \in E$$

i.e. $N(A)$ counts the number of $X_i \in A$.

The feature of a point process can be characterised by

- intensity measure

$$\Lambda(A) = E(N(A))$$

- intensity (density) function

$$\lambda(\mathbf{x}) = \frac{\partial \Lambda(A)}{\partial x_1 \cdots \partial x_d}$$

where $A = [a_1, x_1] \times \cdots \times [a_d, x_d] \subset \mathbb{R}^d$.

Poisson Point Process (PPP) is one of point processes which is used widely in Spatial Extreme value theory. A point process N is called Poisson Point Process if it holds,

- For $A \subset E$ and $k \geq 0$,

$$\mathbb{P}(N(A) = k) = \begin{cases} e^{-\Lambda(A)} \frac{\Lambda(A)^k}{k!} & \text{if } \Lambda(A) < \infty, \\ 0 & \text{if } \Lambda(A) = \infty, \end{cases}$$

- For any $m \geq 1$, if A_1, \dots, A_m are mutually disjoint subsets of E , then $N(A_1), \dots, N(A_m)$ are independent random variables.

A Poisson Point Process is said to be homogenous if its intensity Λ is proportional to the length, area, volume, etc. of a set A . The intensity function is $\lambda(x) = \lambda$ where λ represents the average number of point per unit of length, area and volume. In addition, a Poisson Point Process is said to be non-homogenous if its intensity function is non-stationary, it depends on the location of space A :

$$\Lambda(A) = \int_A \lambda(x) dx$$

Similar to GEV, assume that F is unknown and that the Extremal Type Theorem applies, i.e. $\exists(a_n) > 0$, $(b_n) \in \mathbb{R}$ s.t.

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq y\right) \rightarrow G_\xi(y; \mu, \sigma) = \exp\left\{-\left[1 + \xi \left(\frac{y - \mu}{\sigma}\right)\right]_+^{-1/\xi}\right\},$$

for some $\mu, \sigma > 0$ and ξ . We construct a sequence of point processes on \mathbb{R}^2 by

$$N_n = \left\{ \left(\frac{i}{n+1}, \frac{X_i - b_n}{a_n} \right) : i = 1, \dots, n \right\} \quad (2.3.1)$$

Theorem 2.3.1 (Leadbetter et al. (2012)). *Let y_* and y^* be the lower and upper endpoints of G . On regions of the form $[0, 1] \times [u, \infty)$*

$$N_n \xrightarrow{d} N, \quad \text{as } n \rightarrow \infty, \quad (2.3.2)$$

where N is a non-homogeneous Poisson point process.

Corollary 2.3.2. *On $A = [t_1, t_2] \times [y, y^*)$, the intensity measure of the limiting Poisson point process N is*

$$\Lambda(A) = (t_2 - t_1) \left[1 + \xi \left(\frac{y - \mu}{\sigma}\right)\right]_+^{-1/\xi} \quad (2.3.3)$$

The probability of each point in N_n falling into the region $A_u = [0, 1] \times [u, \infty]$ is

$$p_n = \mathbb{P}\left(\frac{X_i - b_n}{a_n} > u\right) \approx \frac{1}{n}[1 + \xi \frac{u - \mu}{\sigma}]^{-1/\xi}$$

Let $N_n(A)$ be the number of points of N_n in region A. Because X_i 's are i.i.d, $N_n(A)$ follows a binomial distribution $\text{Bin}(n, p_n)$. As $n \rightarrow \infty$, the binomial distribution converges to a Poisson distribution with rate

$$\Lambda(A_u) = [1 + \xi \frac{u - \mu}{\sigma}]^{-1/\xi}$$

the rate is homogeneous in the time direction. Therefore, this is exactly the characterising property of a non-homogeneous Poisson process with intensity function

$$\lambda(t, v) = \frac{1}{\sigma}[1 + \xi \frac{v - u}{\sigma}]^{-1/\xi-1}$$

Since the process is homogeneous in time, we obtain the intensity measure of the limiting process on any region A of the form $[t_1, t_2] \times (u, \infty)$ with $0 \leq t_1 < t_2 \leq 1$. For all such non-overlapping set A the $N(A)$ are independent and Poisson distributed. N , therefore, is a Poisson point process.

One advantage of Poisson Point Process is that PPP can unify both GEV and GPD theories with point process representation. Let $M_n = \max_{i=1,\dots,n} X_i$ and consider the sequence of point processes

$$N_n = \left\{ \left(\frac{i}{n+1}, \frac{X_i - b_n}{a_n} \right) : i = 1, \dots, n \right\}$$

Taking $A_z = (0, 1) \times (z, \infty)$, the event

$$\left\{ \frac{M_n - b_n}{a_n} \leq z \right\} \text{ is equivalent to the event } \{N_n(A_z) = 0\}$$

and therefore

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq z\right) = \mathbb{P}(N_n(A_z) = 0) \rightarrow \mathbb{P}(N(A_z) = 0).$$

Since N follows a Poisson distribution,

$$\mathbb{P}(N(A_z) = 0) = \exp\{-\Lambda(A_z)\} = \exp\left\{-\left[1 + \xi \left(\frac{z - \mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}.$$

The limiting distribution of normalised block maxima is the GEV.

Let $N_n^i = \left(\frac{i}{n+1}, \frac{X_i - b_n}{a_n} \right)$ be the i -th point in N_n . For a sufficiently large u , we recover back GPD,

$$\begin{aligned}\mathbb{P}\left(\frac{X_i - b_n}{a_n} > u + x \mid \frac{X_i - b_n}{a_n} > u\right) &= \frac{\mathbb{P}\left(\frac{X_i - b_n}{a_n} > u + x\right)}{\mathbb{P}\left(\frac{X_i - b_n}{a_n} > u\right)} \\ &= \frac{\mathbb{P}(N_n^i \in (0, 1) \times (u + x, \infty))}{\mathbb{P}(N_n^i \in (0, 1) \times (u, \infty))} \\ &\approx \frac{\frac{1}{n} \left[1 + \xi \left(\frac{u+x-\mu}{\sigma}\right)\right]^{-1/\xi}}{\frac{1}{n} \left[1 + \xi \left(\frac{u-\mu}{\sigma}\right)\right]^{-1/\xi}} \\ &= \left[1 + \frac{\xi x}{\tilde{\sigma}}\right]^{-1/\xi}.\end{aligned}$$

Poisson Point Process will be widely utilised in next chapter where Bayesian Mixture Model is derived based on the PPP representation.

2.4 Max Stable Process

Univariant Extreme Value Theory is limited in real applications since there are more than one variable we need to consider, for example, multiple locations, multiple stocks, etc. A max-stable process extends GEV further to stochastic processes. A max-stable process, Z , with unit Fréchet margins can be characterised as

$$Z(s) = \sup_{i \geq 1} R_i W_i(s), \quad s \in \mathcal{S} \quad (2.4.1)$$

where R_1, R_2, \dots are the points of a Poisson Point Process (PPP) on $(0, \infty)$ and $W_1(s), W_2(s), \dots$ are independent copies of a stochastic process $W(s)$ with unit mean (De Haan and Ferreira, 2006; Schlather, 2002). The exponent measure restricted onto $\mathbb{R}_+^D \setminus \{0\}$, denoted by κ , is given by

$$\kappa([0, x]^c) = \mathbb{E} \left[\max_{i=1, \dots, D} \left\{ \frac{W(s_i)}{x_i} \right\} \right] = \int_0^\infty 1 - \Pr(W \in [0, xr]) dr, \quad x \in \Omega, \quad (2.4.2)$$

where c denotes the complement, $W = (W(s_1), \dots, W(s_D))^\top$ and $\Omega = \mathbb{R}_+^D \setminus \{0\}$. As such, the distribution function G can be expressed as

$$G(x) = \exp \{-\kappa([0, x]^c)\} = \exp \{-V(x)\}, \quad (2.4.3)$$

where $V(x)$ is a homogeneous function of order -1 called the exponent function. The distribution function $G(x)$ can also be viewed as the probability that no points of the PPP with mean measure $\kappa([0, x]^c)$ are greater than x at locations s_i , $i = 1, \dots, D$.

Let $B_D = \{1, \dots, D\}$ and $B_k = \{b_1, \dots, b_k\} \subset B_D$, where $b_1 < \dots < b_k$ such that $x_{B_D} \equiv x$ and $x_{B_k} = (x_{b_1}, \dots, x_{b_k})$. Let $\Omega_{B_k} = \{x \in \Omega : x_j = 0 \text{ if } j \notin B_k\}$ such that $\partial\Omega = \{\Omega_{B_k}, \forall B_k \text{ and } k = 1, \dots, D-1\}$ represents the boundaries of

Ω , i.e., the set of all subspaces of Ω . The interior of Ω is denoted by $\Omega^\circ = \{x \in \Omega : x_i > 0, \forall i \in B_D\}$, i.e., $\Omega^\circ = \Omega \setminus \partial\Omega$. Depending on the choice of W in (1), the exponent measure κ can put mass on both $\partial\Omega$ and Ω° with the intensity function on each subspace Ω_{B_k} given by

$$\lim_{x_i \rightarrow 0, i \notin B_k} -V_{B_k}(x) \quad (2.4.4)$$

where $V_{B_k} = \frac{\partial^k V}{\partial x_{b_1} \cdots \partial x_{b_k}}$. On Ω° , it can be expressed as $\kappa(x) = -V_{B_D}(x)$, where the function κ is referred to as the intensity function of the max-stable process.

2.5 Brown-Resnick Model

Brown-Resnick model is the most popular Max-Stable Process as it has desired Gaussian property ([Brown and Resnick, 1977](#); [Kabluchko et al., 2009](#)). Letting $W(s) = \exp(Y(s) - \sigma^2/2)$ in equation 2.4.1, where $Y(s)$ is Gaussian random process with zero means and variance σ^2 , yield popular Brown-Resnick Model:

$$Z(s) = \sup_{i \geq 1} R_i \exp \{W_i(s) - \delta(s)\}, \quad s \in \mathcal{S} \quad (2.5.1)$$

where:

- R_i are points from a Poisson point process (PPP) on $(0, \infty)$ with intensity $r^{-2} dr$,
- $W_i(s)$ are independent replicates of a centered Gaussian process with stationary increments,
- $\delta(s) = \frac{1}{2} \text{Var}(W(s))$ ensures unit Fréchet margins.

2.6 \mathcal{R} -Pareto Process

\mathcal{R} -Pareto process is an extension of GPD just similar to Max-Stable Process is an extension of GEV. The theory of \mathcal{R} -Pareto process ([Dombry and Ribatet, 2015](#); [de Fondeville and Davison, 2022](#)) focuses on the limit distribution of stochastic processes deemed extreme defined by some risk functional $r(\cdot)$. Assume the processes deemed extreme according to some risk function $r(\cdot)$. Assuming the process X with unit Pareto margins, i.e., $P(X(s) > u) = 1/u, u > 1$, satisfies the regular varying conditions, i.e. $\lim_{u \rightarrow \infty} u \Pr(X/u \in B) = \kappa(B)$ for all measurable sets $B \subset C^+(S)$, where $C^+(S)$ is the space of non-negative functions on S , then the limiting process

$$\tilde{Z}(s) = \lim_{u \rightarrow \infty} \frac{X(s)}{u} \Big| r(\{X(s), s \in \mathcal{S}\} > u) \quad (2.6.1)$$

defines a simple \mathcal{R} -Pareto process on $\mathcal{A}_r = \{f \in C^+(S) : r(f) > 1\}$ with the probability measure $\kappa(\cdot \cap \mathcal{A}_r)/\kappa(\mathcal{A}_r)$. Its finite dimensional density is therefore $\frac{\kappa(\mathbf{x})}{\kappa(\mathcal{A}_r^D)}$, for $\mathbf{x} \in \mathcal{A}_r^D$, where κ is the intensity function previously defined and \mathcal{A}_r^D is the set \mathcal{A}_r restricted to D dimensions.

[Zhong et al. \(2024\)](#) introduces a more flexible likelihood function for \mathcal{R} -Pareto process derived from Brown-Resnick Model adding skewness that introduces more anisotropic spatial dependency.

Formally, let $W(\mathbf{s}) = \exp\{Y(\mathbf{s}) - a(\mathbf{s})\}$ where $Y(\mathbf{s})$ is a centred skew-normal process with scale matrix Σ and slant parameter α , and $a(\mathbf{s}) = \log \mathbb{E}[\exp\{Y(\mathbf{s})\}]$. The corresponding max-stable process defined via (1) is called the skewed Brown-Resnick process and is characterized by the following exponent function:

$$V(\mathbf{x}) = \sum_{k=1}^D \Psi \left(\log \left(\frac{x_{-k}^{\circ\circ}}{x_k^{\circ\circ}} \right) + \frac{\tilde{\omega}_{-k}^2 - \tilde{\omega}_k^2}{2}; \mu_k, \Sigma_k, \hat{\alpha}_k, \tau_k \right), \quad (2.6.2)$$

where Ψ denotes the cdf of the extended skew-normal distribution ([Zhong et al., 2024](#)), $\mathbf{x}^{\circ\circ} = (x_1^{\circ\circ}, \dots, x_D^{\circ\circ})$, $x_k^{\circ\circ} = 2x_k\Phi(\tau_k; 0, 1)$, $k = 1, \dots, D$, $\tau_k = \tilde{\omega}_k\delta_k$, $\tilde{\boldsymbol{\omega}} = (\tilde{\omega}_1, \dots, \tilde{\omega}_D) = \boldsymbol{\omega}\mathbf{1}$, $\boldsymbol{\omega} = \sqrt{\text{diag}(\Sigma)}$, $\boldsymbol{\delta} = (1 + \boldsymbol{\alpha}^\top \bar{\Sigma} \boldsymbol{\alpha})^{-1/2} \bar{\Sigma} \boldsymbol{\alpha}$, $\bar{\Sigma} = \boldsymbol{\omega}^{-1} \Sigma \boldsymbol{\omega}$, $\boldsymbol{\mu}_k = \mathbf{A}_k \Sigma_k$, $\mathbf{A}_k = (e_1, \dots, e_{k-1}, -1, e_k, \dots, e_{D-1})$, e_k , $k = 1, \dots, D-1$ are standard basis vectors of dimension $(D-1)$, $\Sigma_k = \mathbf{A}_k \Sigma \mathbf{A}_k^\top$, $\hat{\boldsymbol{\alpha}}_k = (1 - \boldsymbol{\delta}^\top \boldsymbol{\omega} \mathbf{A}_k^\top \Sigma_k^{-1} \mathbf{A}_k \boldsymbol{\omega} \boldsymbol{\delta})^{-1/2} \hat{\omega}_k \Sigma_k^{-1} \mathbf{A}_k \boldsymbol{\omega} \boldsymbol{\delta}$, and $\hat{\boldsymbol{\omega}}_k = \sqrt{\text{diag}(\Sigma_k)}$.

The density function of the skewed Brown-Resnick model is given by

$$\begin{aligned} \kappa(\mathbf{x}) &= \frac{2\Phi(\tilde{\tau}; 0, 1)|\Sigma|^{-1/2}(1^\top \mathbf{q})^{-1/2}}{(2\pi)^{(D-1)/2} \prod_{k=1}^D x_k} \\ &\exp \left[-\frac{1}{2} \left[\log \mathbf{x}^{\circ\circ\top} \mathcal{M} \log \mathbf{x}^{\circ\circ} + \log \mathbf{x}^{\circ\circ\top} \left(\frac{2\mathbf{q}}{1^\top \mathbf{q}} + \tilde{\mathcal{M}}\tilde{\boldsymbol{\omega}}^2 \right) + \frac{\mathbf{q}^\top \tilde{\boldsymbol{\omega}}^{-1}}{1^\top \mathbf{q}} + \frac{1}{4}\tilde{\boldsymbol{\omega}}^{2,\top} \mathcal{M} \tilde{\boldsymbol{\omega}}^2 \right] \right] \end{aligned} \quad (2.6.3)$$

where $\mathbf{q} = \Sigma^{-1}\mathbf{1}$, $\mathcal{M} = \Sigma^{-1} - \mathbf{q}\mathbf{q}^\top/1^\top \mathbf{q}$ and

$$\tilde{\tau} = \left(1 + \frac{(\boldsymbol{\alpha}^\top \boldsymbol{\omega}^{-1} \mathbf{1})^2}{1^\top \mathbf{q}} \right)^{-1/2} \boldsymbol{\alpha}^\top \boldsymbol{\omega}^{-1} \left[\left(\mathbb{I} - \frac{1\mathbf{q}^\top}{1^\top \mathbf{q}} \right) \left(\log \mathbf{x}^{\circ\circ} + \frac{\tilde{\boldsymbol{\omega}}^2}{2} \right) + \frac{1}{1^\top \mathbf{q}} \right].$$

setting $\alpha = 0$ implies $\hat{\alpha}_k = 0$, $\tau_k = 0$ and therefore $x^{\circ\circ} = x$, recovering the Brown-Resnick model. Likelihood function measures how well a statistical model explains the observed data D by calculating the probability of seeing that data under different parameter values of the model. The likelihood function of the \mathcal{R} -Pareto process is,

$$\ell_{rP}(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n) = \sum_{i \in \{m: r(\mathbf{x}_m) > u\}} \log \left(\frac{\kappa(\mathbf{z}_i; \boldsymbol{\theta})}{K(\mathcal{A}_r; \boldsymbol{\theta})} \right) \quad (2.6.4)$$

where the numerator, $\kappa(\mathbf{z}_i; \boldsymbol{\theta})$, is the Eq (13) and the denominator $K(\mathcal{A}_r; \boldsymbol{\theta})$ is the integration on the area \mathcal{A}_r with respect to $\kappa(\mathbf{z}_i; \boldsymbol{\theta})$.

[Zhong et al. \(2024\)](#) has proposed a state-of-art model with fast inference and flexible spatial dependence in EVT field. Since [Zhong et al. \(2024\)](#)'s work, however, there is still a huge gap in which models are not able to model multiple types of extreme events jointly or classify multiple extreme events. Therefore, this gap is the position where my thesis sits. We propose a Bayesian Mixture Model based on [Zhong et al. \(2024\)](#)'s model to fit two types of rainfall extreme events: frontal and convective jointly.

CHAPTER 3

Bayesian Mixture Model

Multimodal data are common in statistics research and real applications. They often appear when observations come from two or more underlying groups or populations. The analysis of these types of data is often undertaken with mixture model. In statistics, a mixture model is a probabilistic model for representing the presence of subpopulations with an overall population, without requiring that an observed data contain the explicit information about subpopulation to which an individual observation belongs. The mixture model can be built in a Bayesian perspective, which not only estimate the value of parameters but also approximate a posterior distribution of the parameters with uncertainty that is valuable information in real world. In light of climate change, the extreme rainfall at a range of scales and durations can vary from the very small and short (roughly 1km and 5 minutes) up to thousands of square kilometres and days. Acquiring such knowledge of rainfall behaviour is critical for extreme rainfall monitoring and hazard prevention. Extreme precipitation is caused by a mixture known and well-understood processes. The primary drivers are convection, which produces spatially localised, high-intensity precipitation events over short time scales ([Schroeder et al., 2018](#)). Another processes that cause large scale extreme rainfall is called frontal, which produces stratiform precipitation of lower intensity but affecting a much larger area for a longer duration ([Catto and Pfahl, 2013](#)). Therefore, a Bayesian Mixture Model is suitable to model extreme rainfall in nature.

Building on this framework, this chapter introduces the base models Brown-Resnick Model and \mathcal{R} -Pareto Process on a finite grid first and then introduces the proposed Bayesian Mixture Model derived from an \mathcal{R} -Pareto Process. We will provide a detailed technical specification, outlining the model parameters, defining the likelihood function, justifying the choice of priors, and describing the methodology for approximating the posterior distribution.

3.1 Brown-Resnick Process and Semivariogram on a Finite Grid

Recapping the chapter 2 section 2.5, a spatial Gaussian random field is utilised to represent the spatial dependency in Brown-Resnick Model. Formally, $\{s_1, \dots, s_n\} \subset \mathbb{R}^2$ be fixed spatial sites. A Brown-Resnick (BR) max-stable process $\{Z(s) : s \in \mathbb{R}^2\}$ with semivariogram $\gamma : \mathbb{R}^2 \rightarrow [0, \infty)$ admits the spectral representation.

$$Z(s) = \max_{i \geq 1} R_i \exp\left(W_i(s) - \frac{1}{2}\gamma(s)\right), \quad (3.1.1)$$

where $\{R_i\}$ is a Poisson point process on $(0, \infty)$ with intensity $\xi^{-2} d\xi$, and $\{W_i\}$ are i.i.d. centred Gaussian processes with stationary increments such that

$$(W(s) - W(t)) = 2\gamma(s - t), \quad s, t \in \mathbb{R}^2. \quad (3.1.2)$$

On the finite grid, the induced $n \times n$ matrix is

$$\Sigma_{ij}(\gamma) = \gamma(s_i) + \gamma(s_j) - \gamma(s_i - s_j), \quad 1 \leq i, j \leq n. \quad (3.1.3)$$

The isotropic power semivariogram is utilised,

$$\gamma_{\theta, \vartheta}(h) = (\|h\|/\theta)^{\vartheta}, \quad h \in \mathbb{R}^2, \quad (3.1.4)$$

with $\vartheta = 1.5$ throughout, and write $\Sigma(\theta) \equiv \Sigma(\gamma_{\theta, 1.5})$.

3.2 \mathcal{R} -Pareto Process on a Grid and L_1 Risk Function

Let $L : \mathbb{R}^n \rightarrow [0, \infty)$ denote a risk functional, here we take the L_1 risk function

$$L_1(x) = \frac{1}{n} \sum_{i=1}^n |x_i|. \quad (3.2.1)$$

For a high threshold u , define the exceedance set

$$\mathcal{X}^{(u)} = \{x \in (0, \infty)^n : L_1(x) > u\} \quad \text{and let} \quad \{x^{(i)}\}_{i=1}^n \subset \mathcal{X}^{(u)}. \quad (3.2.2)$$

The limiting Peaks-Over-Threshold (POT) model induced by the BR process on the grid is an \mathcal{R} -Pareto law. Each exceedance admits the polar form

$$X = RV, \quad R \sim \text{Pareto}(1) \ (r \geq 1), \quad V \in \mathbb{S}_+^{n-1} := \{v > 0 : \sum_{i=1}^n v_i = 1\}, \quad (3.2.3)$$

The detailed data generation is shown in appendix A.

3.3 Likelihood Function

A Likelihood function measures how well a statistical model explains the observed data by calculating the probability of seeing that data under different parameter values of the model. [Zhong et al. \(2024\)](#) proposed the intensity function 2.6.3 for flexible \mathcal{R} -Pareto process. In this thesis, we drop the skewness such that the intensity function (angular (simplex) distribution of V in equation 3.2.3) is determined by $\Sigma(\theta)$,

$$\begin{aligned} \kappa(x; \Sigma(\theta)) = & \frac{|\Sigma|^{-1/2} (\mathbf{1}^\top \mathbf{q})^{-1/2}}{(2\pi)^{(D-1)/2} \prod_{k=1}^D w_k} \times \exp \left\{ -\frac{1}{2} \left[\log \mathbf{x}^\top \mathcal{M} \log \mathbf{x} + \log \mathbf{x}^\top \left(\frac{2\mathbf{q}}{\mathbf{1}^\top \mathbf{q}} + \mathcal{M}\boldsymbol{\omega}^2 \right) \right. \right. \\ & \left. \left. + \frac{\mathbf{q}^\top \boldsymbol{\omega}^2 - 1}{\mathbf{1}^\top \mathbf{q}} + \frac{1}{4} \boldsymbol{\omega}^{2,\top} \mathcal{M} \boldsymbol{\omega}^2 \right] \right\} \end{aligned} \quad (3.3.1)$$

where $q = \Sigma^{-1} \mathbf{1}$, $\omega = \sqrt{\text{diag}(\Sigma)}$ and $\mathcal{M} = \Sigma^{-1} - q q^\top / \mathbf{1}^\top q$.

The likelihood function, therefore, is formulated,

$$\mathcal{L}(x|\theta) = \prod_i^n \frac{\kappa(x; \theta)}{K(\mathcal{A}_r; \theta)} \quad (3.3.2)$$

As we apply L_1 risk function to define the extremes with the function, we, therefore, can drop the denominator in the likelihood function 3.3.2 as it is a constant such that the likelihood is simplified,

$$\mathcal{L}(x|\theta) = \prod_i^n \kappa(x; \theta) \quad (3.3.3)$$

3.4 Bayesian Mixture Model

The mixture r-Pareto Process is modelled in a Bayesian perspective. We assume that there are two distinct distributions ($M = 2$) in the observations: a Brown-Resnick Process with long range and a Brown-Resnick Process with small range corresponding to precipitation caused by frontal and convective respectively.

3.4.1 Variables

Formally, for observation $y = \{y_1, y_2, \dots, y_n\}$, the M -components ($M = 2$) mixture distribution is

$$f(y_i | \theta, \lambda) = \sum_{i=1}^M \lambda_m f_m(y_i | \theta_m) \quad (3.4.1)$$

where,

- $f_m(y | \theta_m)$ depends on the parameter θ_m where f_m is the intensity function 3.3.1.
- λ_m is proportion of y_i from component m , $\sum_m \lambda_m = 1$

There is an unobservable indicator variable z_{im} :

$$z_{im} = \begin{cases} 1, & \text{if } y_i \text{ is drawn from the } m^{\text{th}} \text{ component} \\ 0, & \text{otherwise} \end{cases} \quad (3.4.2)$$

Given the λ , the distribution of $z_i = (z_{i1}, \dots, z_{iM})$ is,

$$z_i | \lambda \sim \text{Multinomial}(1 : \lambda_1, \dots, \lambda_M) \quad (3.4.3)$$

where

$$\pi(z_i | \lambda) = \frac{1!}{z_{i1}! \cdots z_{iM}!} \prod_{m=1}^M \lambda_m^{z_{im}} = \prod_{m=1}^M \lambda_m^{z_{im}}, \quad (3.4.4)$$

so that

$$\Pr(z_{im} = 1) = \lambda_m, \text{ for all individual } i$$

3.4.2 Mixture Joint Distribution

We suppose that data y are conditionally independent, and if $z_{im} = 1$ then

$$f(y_i \mid z_i, \theta) = f_m(y_i \mid \theta_m) = \prod_{m=1}^M f_m(y_i \mid \theta_m, z_{im})^{z_{im}} \quad (3.4.5)$$

Hence

$$L(y \mid z, \theta) = \prod_{i=1}^n f(y_i \mid z_i, \theta) = \prod_{i=1}^n \left[\prod_{m=1}^M f_m(y_i \mid \theta_m, z_i)^{z_{im}} \right]. \quad (3.4.6)$$

Now since $z_i \mid \lambda \sim \text{Multinomial}(1 : \lambda_1, \dots, \lambda_M)$, then

$$\pi(z_i \mid \lambda) = \prod_{m=1}^M \lambda_m^{z_{im}} \quad \text{and so} \quad \pi(z \mid \lambda) = \prod_{i=1}^n \prod_{m=1}^M \lambda_m^{z_{im}}$$

Hence

$$\begin{aligned} L(y, z \mid \theta, \lambda) &= L(y \mid z, \theta) \pi(z \mid \lambda) \\ &= \prod_{i=1}^n \left[\prod_{m=1}^M f_m(y_i \mid \theta_m, z_i)^{z_{im}} \right] \prod_{i=1}^n \prod_{m=1}^M \lambda_m^{z_{im}} \\ &= \prod_{i=1}^n \prod_{m=1}^M [\lambda_m f_m(y_i \mid \theta_m)]^{z_{im}}. \end{aligned} \quad (3.4.7)$$

3.4.3 Prior Distribution

Prior Distribution is a critical component in Bayesian framework, which represents the prior knowledge or probability distribution before some evidence is taken into account. For instance, the prior could be the probability of if earth is round before we measure it actually. In Bayesian statistics, it describes our knowledge about the model parameters. Bayes's rule prescribes how to update the prior with new information, i.e. observations to obtain the posterior probability distribution, which is the conditional distribution of the uncertain quantity given new data.

Conjugate Prior means the prior distribution and corresponding posterior distribution belong to same parametric family.

Definition 3.4.1. if \mathcal{F} is a class of sampling distribution (likelihoods) $\mathcal{L}(x|\theta)$, and \mathcal{P} is a class of prior distribution $\pi(\theta)$ for θ . then the class \mathcal{P} is conjugate for \mathcal{F} if

$$\pi(\theta|x) \in \mathcal{P} \quad \text{for all } \mathcal{L}(\cdot|\theta) \in \mathcal{F} \text{ and } \pi(\cdot) \in \mathcal{P} \quad (3.4.8)$$

Here is an example to illustrate the idea behind conjugate prior,

Example 3.4.2. Suppose we have $x \sim \text{Bin}(n, \theta)$ and we wish to make inference on $0 \leq \theta \leq 1$. Here

$$L(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad x = 0, \dots, n.$$

While $\pi(\theta)$ will vary from problem to problem, we want to choose a prior family that will simplify posterior computations. Specify $\theta \sim \text{Beta}(a, b)$ so that

$$\begin{aligned}\pi(\theta) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1} \quad 0 \leq \theta \leq 1 \\ &\propto \theta^{a-1}(1-\theta)^{b-1}.\end{aligned}$$

Then, by Bayes's Theorem:

$$\begin{aligned}\pi(\theta|x) &\propto L(x|\theta)\pi(\theta) \\ &\propto \theta^x(1-\theta)^{n-x} \times \theta^{a-1}(1-\theta)^{b-1} \\ &= \theta^{a+x-1}(1-\theta)^{b+n-x-1}.\end{aligned}$$

Since we know $\pi(\theta|x)$ is a proper density function, then we must have

$$\theta|x \sim \text{Beta}(a+x, b+n-x)$$

From this example, we know that staying within Beta family means posterior computation is easy. We will use this conjugate prior to our proposed mixture model later.

There are two parameters involved in the proposed mixture model: the range of frontal rainfall and the range of convective rainfall. Let θ_1 denote the frontal extremes' range parameter and θ_2 denote the convective extremes' range parameter based on the definition of semivariogram function 3.1.4. There's an explicitly order between two parameters $\theta_1 > \theta_2$ according to the definition of frontal and convective precipitation.

Assuming that the prior for (θ, λ) is independent: $\pi(\theta, \lambda) = \pi(\theta)\pi(\lambda)$. We define the λ is from Dirichlet($\alpha_1, \dots, \alpha_M$):

$$\pi(\lambda) \propto \prod_{m=1}^M \lambda_m^{\alpha_m-1} \tag{3.4.9}$$

where

- We always give Dirichlet($1, \dots, 1$) that is a non-informative prior
- Dirichlet $\pi(\lambda)$ is conjugate with Multinomial $\pi(z | \lambda)$
- When $M = 2$, the Dirichlet distribution reduces to the Beta distribution, and the Multinomial distribution reduces to the Binomial distribution.

For range parameter θ , we give Gamma distribution $\text{Gamma}(1, 1)$. Since there is an implicit order $\theta_1 > \theta_2$, one mathematical trick we use here is that we give the prior distributions for θ_2 and the gap δ such that

$$\theta \sim \text{Gamma}(1, 1) \quad \delta \sim \text{Gamma}(1, 1) \tag{3.4.10}$$

where $\theta_1 = \theta_2 + \delta$.

3.4.4 Mixture Joint Distribution

We suppose that data y are conditionally independent, and if $z_{im} = 1$ then

$$f(y_i \mid z_i, \theta) = f_m(y_i \mid \theta_m) = \prod_{m=1}^M f_m(y_i \mid \theta_m, z_{im})^{z_{im}} \quad (3.4.11)$$

Hence

$$L(y \mid z, \theta) = \prod_{i=1}^n f(y_i \mid z_i, \theta) = \prod_{i=1}^n \left[\prod_{m=1}^M f_m(y_i \mid \theta_m, z_i)^{z_{im}} \right]. \quad (3.4.12)$$

Now since $z_i \mid \lambda \sim \text{Multinomial}(1 : \lambda_1, \dots, \lambda_M)$, then

$$\pi(z_i \mid \lambda) = \prod_{m=1}^M \lambda_m^{z_{im}} \quad \text{and so} \quad \pi(z \mid \lambda) = \prod_{i=1}^n \prod_{m=1}^M \lambda_m^{z_{im}}$$

Hence

$$\begin{aligned} L(y, z \mid \theta, \lambda) &= L(y \mid z, \theta) \pi(z \mid \lambda) \\ &= \prod_{i=1}^n \left[\prod_{m=1}^M f_m(y_i \mid \theta_m, z_i)^{z_{im}} \right] \prod_{i=1}^n \prod_{m=1}^M \lambda_m^{z_{im}} \\ &= \prod_{i=1}^n \prod_{m=1}^M [\lambda_m f_m(y_i \mid \theta_m)]^{z_{im}}. \end{aligned} \quad (3.4.13)$$

3.4.5 Mixture Posterior Distribution

The posterior probability is a type of conditional probability that results from updating the prior probability with information summarised by the likelihood via Bayes's rule. In the context of Bayesian statistics, the posterior probability distribution usually describes the epistemic uncertainty about parameters conditional on a collection of the observed data. Particularly, we update the range parameters θ and weight λ conditioned on the extreme observations.

Formally, the mixture posterior distribution is given by,

$$\pi(\theta, z, \lambda \mid y) \propto L(y \mid z, \theta) \pi(z \mid \lambda) \pi(\lambda) \pi(\theta) \quad (3.4.14)$$

where $\pi(\theta)$ denotes prior density for the parameters of the component models, commonly $\pi(\theta) = \prod_m \pi_m(\theta_m)$. So we have,

$$\pi(\theta, z, \lambda \mid y) \propto \prod_{i=1}^n \left[\prod_{m=1}^M (\lambda_m f_m(y_i \mid \theta_m))^{z_{im}} \right] \left[\prod_{m=1}^M \lambda_m^{\alpha_m - 1} \right] \pi(\theta) \quad (3.4.15)$$

3.4.6 Monte Carlo Markov Chain for Fitting Model

Markov Chain Monte Carlo (MCMC) is a class of algorithms for sampling from complicated probability distributions, especially posterior distributions in Bayesian

models, where analytical solutions are unavailable. Two of the most common MCMC techniques are:

- Gibbs sampling: update each parameter by sampling directly from its full conditional distribution.
- Metropolis–Hastings (MH): propose a new value for a parameter from some distribution and accept or reject it with a probability chosen to preserve the target posterior as the stationary distribution.

In this thesis, we utilise both techniques: Gibbs for well-defined Dirichlet–Multinomial conjugate and MH for the θ in r-Pareto process spectral density function. Based on the full mixture posterior distribution Eq 3.4.15, we are able to update the parameter θ_m, λ, z in order.

Update θ_m We denote the θ_m as the updated parameter. The full conditional distribution is

$$\pi(\theta_m | y, z, \lambda, \theta_{-m}) \propto \prod_{i=1}^n f(y_i | \theta_m)^{z_{im}} \pi_m(\theta_m) \quad (3.4.16)$$

where random walk proposal is utilised here to get new θ_m .

Update λ The full conditional distribution is

$$\pi(\lambda | y, \theta, z) \propto \prod_{i=1}^n \prod_{m=1}^M \lambda_m^{z_{im}} \lambda_m^{\alpha_m - 1} = \prod_{m=1}^M \lambda_m^{\sum_i z_{im} + \alpha_m - 1}. \quad (3.4.17)$$

where

- the distribution is proportion to $\text{Dirichlet}(\alpha_1 + n_1, \dots, \alpha_M + n_M)$
- $n_m = \sum_i z_{im}$

Update z Similarly, we denote z_i as the updated parameter and z_{-i} as the previous value. The full conditional distribution is

$$\pi(z_i | y, \theta, \lambda, z_{-i}) \propto \prod_{m=1}^M [\lambda_m f(y_i | \theta_m)]^{z_{im}}. \quad (3.4.18)$$

where,

- Each z_i is conditionally independent given λ, θ .
- $\Pr(z_{im} = 1 | \lambda, \theta) \propto \lambda_m f(y_i | \theta_m) = p_{im}$.
- Normalised probabilities $\tilde{p}_{im} = p_{im} / \sum_m p_{im}$.
- Full conditional is then $z_i \sim \text{Multinomial}(1 : \tilde{p}_{i1}, \dots, \tilde{p}_{iM})$.

When $M = 2$,

$$\tilde{p}_{i1} = \frac{\lambda f(y_i | \theta_1)}{\lambda f(y_i | \theta_1) + (1 - \lambda) f(y_i | \theta_2)}$$

We transform the whole process into log space for numerical stability. The detailed algorithm is in appendix B.

CHAPTER 4

Simulation Study

This chapter investigates the feasibility and correctness of the proposed Bayesian mixture model under different conditions. Several simulation studies are conducted due to the flexible control of the spatial dependency and number of observations. The simulation study aims to investigate whether the proposed model would converge under two distinguishable range parameters and whether the proposed model would converge under two close range parameters. In addition, one weight function depending on time t would be predefined and investigate whether the posterior distribution of latent variable z would recover the predefined weight function. The following sections introduce the spatial grid world creation, \mathcal{R} -Pareto process simulation and results analysis of posterior distribution approximations.

4.1 Spatial Grid World

In order to model a realistic precipitation scenario, a spatial grid world is essentially needed to model the size of the study region. A predefined parameter θ in semivariogram is able to control the spatial dependency based on the distance. In this thesis, we create a 10×10 square grid as our simulated region. In details, the latitude is in the range from 1 to 100 with a unit length 10. And the longitude is in the range from 1 to 100 with a unit length 10.

4.2 Distinguishable Parameters Simulation Study

We start from a easy case, which is that the subpopulations have two distinguishable range parameters $\theta_1 = 50$, $\theta_2 = 25$ and $\lambda = 0.7$. The range parameters build the spatial dependency in the \mathcal{R} -Pareto process based on Brown-Resnick Model. Figure 4.1 shows the effects of the range parameter on spatial dependency.

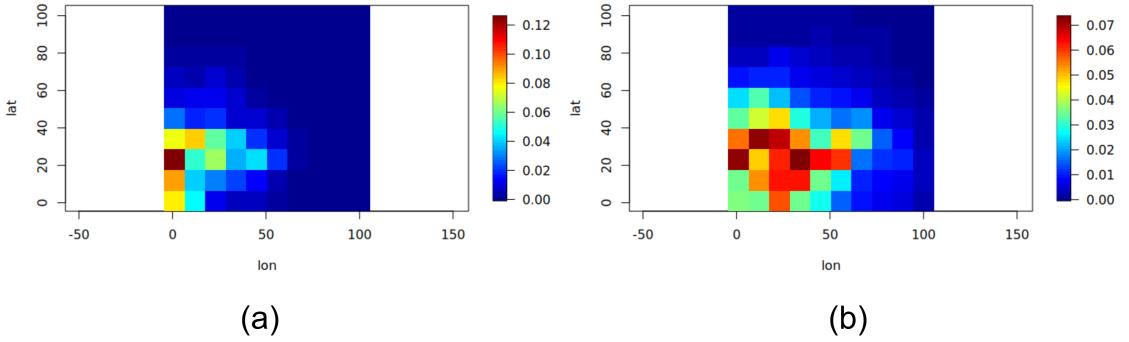


Figure 4.1: Simulated spatial-field realisations illustrating the effect of the range parameter θ on dependence. (a) $\theta = 25$ (short range): dependence decays rapidly and high values remain localised. (b) $\theta = 50$ (long range): dependence decays more slowly, producing broader, smoother regions of elevated values. Axes show grid longitude/latitude; the colour bar indicates precipitation magnitude.

Applying the algorithm from Chapter 3, Section 3.4, with the hyperparameters in Table 4.1, The posterior mean $\hat{\theta}_1 = 50.037$, $\hat{\theta}_2 = 24.954$ and $\hat{\lambda} = 0.699$ is shown in Figure 4.2. In this simple setting, the proposed mixture model accurately recovers the ground truth, and the MCMC chains exhibit good mixing and remain stable after burn-in. Additionally, we justify the posterior distribution of indicator variable z as shown in Figure 4.3, which demonstrating that the proposed model is very confident for assigning the categories about frontal or convective precipitation events.

In order to show the natural difference between convective and frontal events, we utilise $\chi_q(s_A, s_B)$ based on Equation 4.2.1.

$$\chi_q(s_A, s_B) = \Pr \left\{ Y(s_B) > F_{Y(s_B)}^{-1}(q) \mid Y(s_A) > F_{Y(s_A)}^{-1}(q) \right\}, \quad (4.2.1)$$

The upper tail index $\chi(s_A, s_B)$ with quantile q quantify extremal dependence for a process $Y(s)$ where $F_{Y(s_A)}^{-1}$ denotes the quantile function of $Y(s_A)$. The measure $\chi(s_A, s_B)$ determines the extremal dependence class; if $\chi(s_A, s_B) > 0$ or $\chi(s_A, s_B) = 0$, then $Y(s_A)$ and $Y(s_B)$ are asymptotically dependent or asymptotically independent, respectively.

We categorise the extremes by a hard-coded threshold 0.5. If the posterior probability of z is larger than the threshold, we assign label 1 which is the frontal extreme rainfall and assign label 2 otherwise which is the convective extreme rainfall. The Figure 4.4 shows that the frontal rainfall (blue points) exhibits much stronger extremal dependence than convective rainfall (red points) as values of $\chi_q(s_A, s_B)$ appears to tend to zero at a much faster rate for convective rainfall. We compute the empirical $\hat{\chi}_q(s_A, s_B)$ by equation 4.2.2

Table 4.1: MCMC settings used in the distinguishable simulation study.

Schedule	
Iterations	5000
Burn-in	1000
Thinning interval	2
Kept draws	4000
Priors	
Mixing weight λ	Beta(1, 1)
Range θ_2	Gamma(1, 1)
Offset $\delta = \theta_1 - \theta_2$	Gamma(1, 1)
Proposals & Initials	
Proposal s.d. ($\log \theta_2$)	0.01
Proposal s.d. ($\log \delta$)	0.01
Initial λ	0.5
Initial θ_2	10
Initial δ	35
Anchor site	1
Verbosity interval	50
Random seed	123

$$\hat{\chi}_q(s_A, s_B) = \frac{\sum_{t=1}^m \mathbf{1}\left\{X_t(s_A) > F_{X(s_A)}^{-1}(q), X_t(s_B) > F_{X(s_B)}^{-1}(q)\right\}}{\sum_{t=1}^m \mathbf{1}\left\{X_t(s_A) > F_{X(s_A)}^{-1}(q)\right\}} \quad (4.2.2)$$

where m denotes the number of replicates and $\mathbf{1}$ denotes the indicator function.

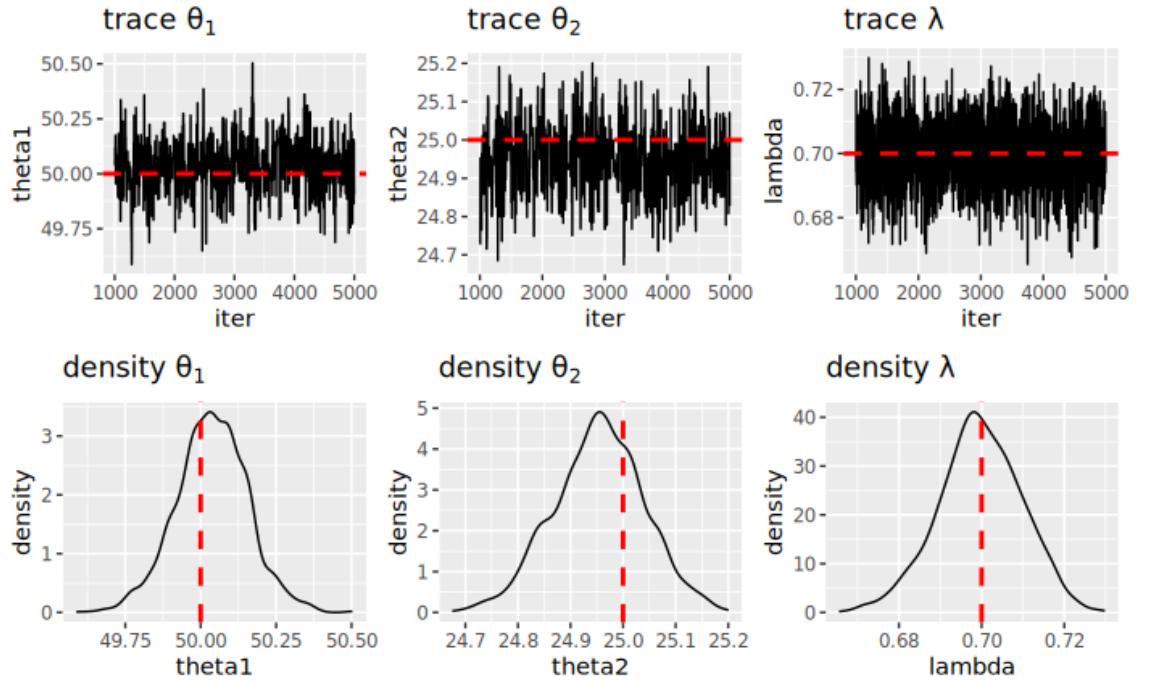


Figure 4.2: Trace (top) and marginal posterior density (bottom) for λ_1 , λ_2 , and π in the distinguishable-mixture simulation. The chains mix well and remain stable after burn-in. Red dashed lines mark the true parameter values; the posteriors concentrate near these targets, indicating accurate recovery by the proposed mixture model.

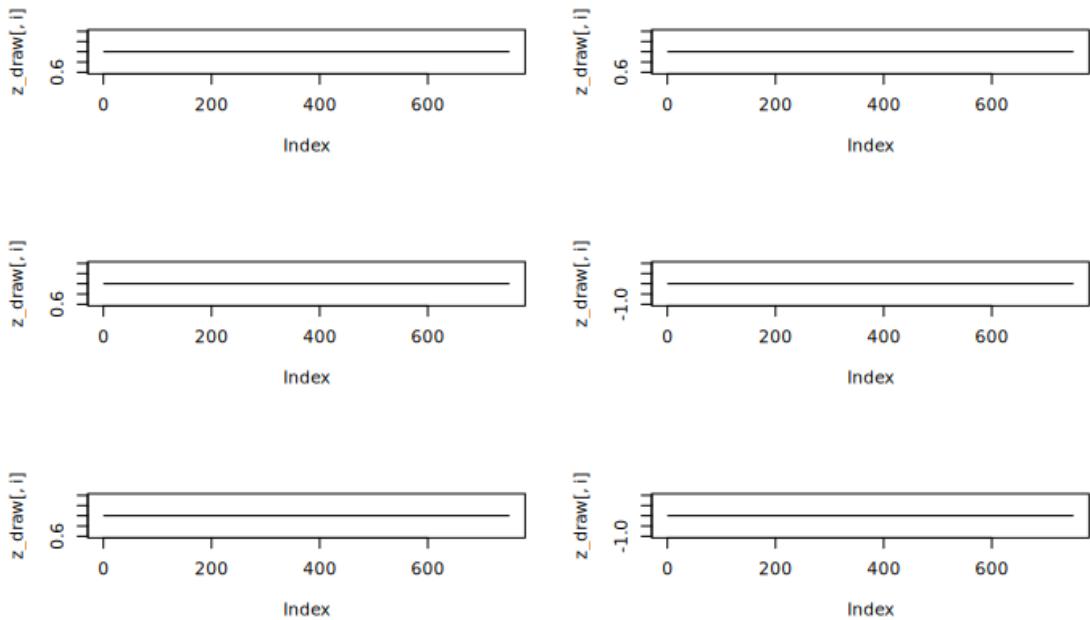


Figure 4.3: Posterior sample paths of the latent indicator z_i for six representative locations ($i \in \{1, 34, 56, 201, 344, 1002\}$). Each panel shows z_i across the kept MCMC iterations after burn-in. The traces are essentially flat at 0 or 1 with no label switching, implying posterior classification probabilities near 0 or 1 and strong separability between the frontal and convective components.

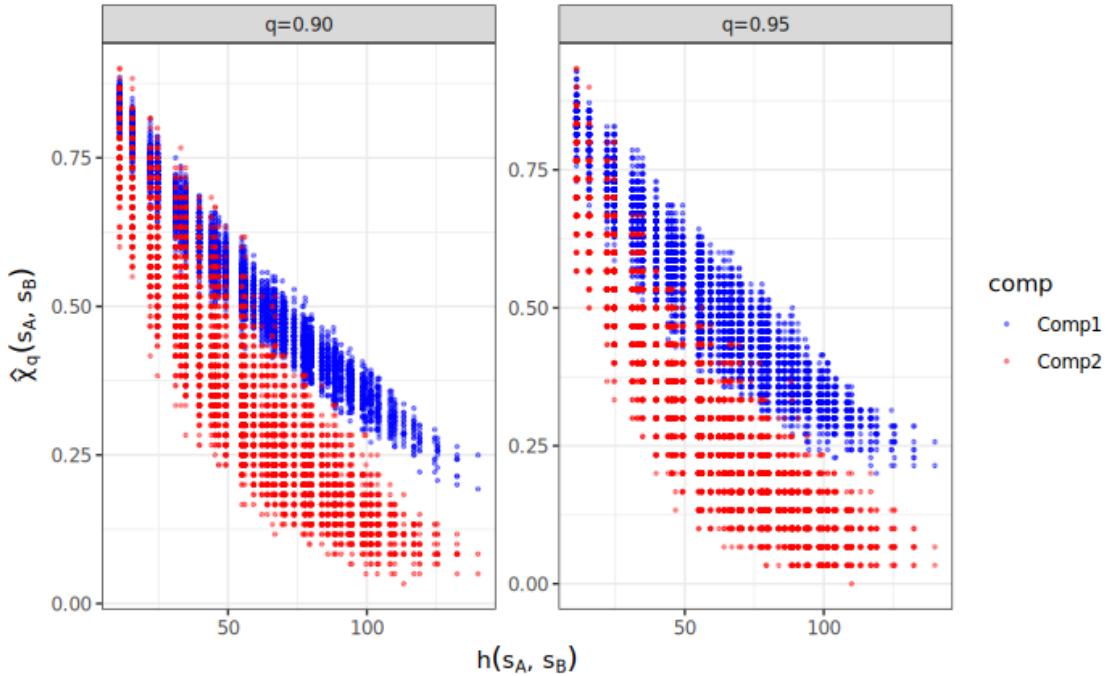


Figure 4.4: Empirical estimates of the upper tail dependence measure $\hat{\chi}_q(s_A, s_B)$ against spatial distance $h(s_A, s_B)$ for two mixture components corresponding to frontal (blue) and convective (red) rainfall events. Results are shown for quantile levels $q = 0.90$ (left) and $q = 0.95$ (right). Frontal rainfall exhibits markedly stronger extremal dependence, as $\hat{\chi}_q(s_A, s_B)$ decays more slowly with distance compared to convective rainfall, whose extremal dependence diminishes rapidly.

4.3 Indistinguishable Parameters Simulation Study

This section examines whether the proposed model can recover parameters when the two components are only weakly separated. We set the range parameters to $\theta_1 = 50$, $\theta_2 = 40$ and $\lambda = 0.7$. The model is fitted using the same algorithm as in Section 4.2, with the modified hyperparameters in Table 4.2. As similar to the previous study, Figure 4.5 presents trace plots and marginal posteriors. The posterior mean $\hat{\theta}_1 = 49.963$, $\hat{\theta}_2 = 39.914$ and $\hat{\lambda} = 0.704$. The indicator paths in Figure 4.3 exhibit frequent switches between 0 and 1, reflecting greater posterior uncertainty in class membership when component parameters are similar (i.e., label-switching behaviour). In contrast to the distinguishable setting—where classes are assigned confidently and stably—the sampler here still centres on the true values but shows increased uncertainty in allocations, as expected under weak separation. Further, the empirical $\chi_q(s_A, s_B)$ is drawn based on the classified categories as shown in Figure 4.7, which indicates that the proposed model still can capture the correct spatial dependency between different extreme rainfall events and separate them correctly.

Table 4.2: MCMC settings used in the indistinguishable simulation study.

Schedule	
Iterations	5000
Burn-in	1000
Thinning interval	2
Kept draws	4000
Priors	
Mixing weight λ	Beta(1, 1)
Range θ_2	Gamma(1, 1)
Offset $\delta = \theta_1 - \theta_2$	Gamma(1, 1)
Proposals & Initials	
Proposal s.d. ($\log \theta_2$)	0.01
Proposal s.d. ($\log \delta$)	0.01
Initial λ	0.5
Initial θ_2	35
Initial δ	10
Anchor site	1
Verbosity interval	50

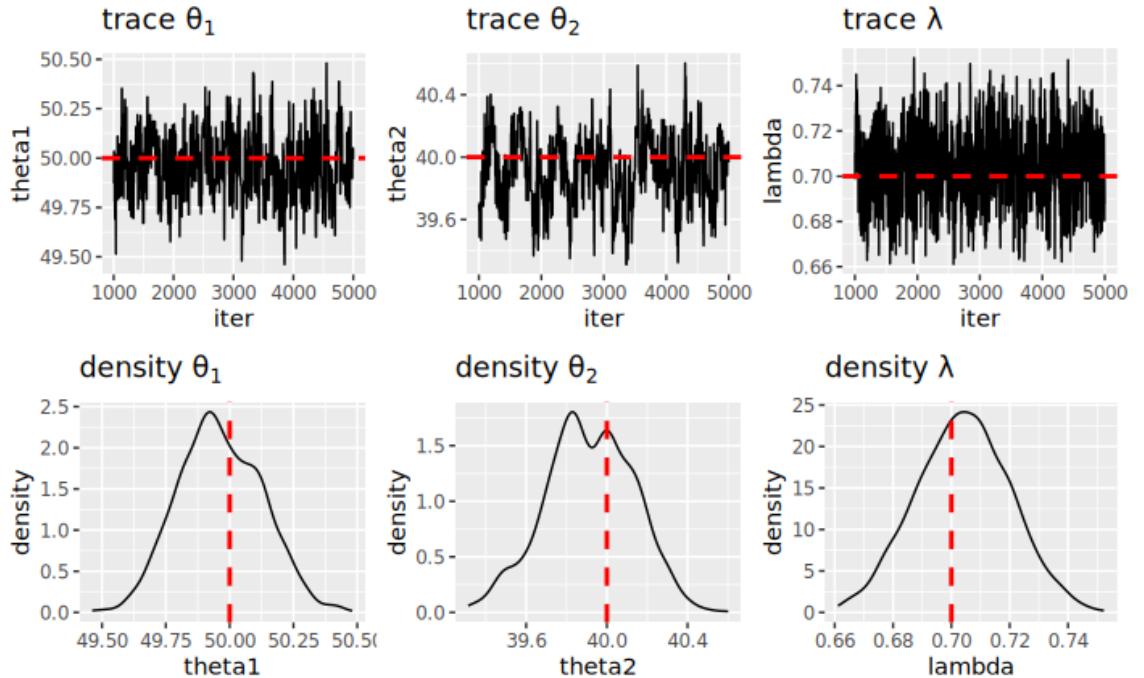


Figure 4.5: Trace (top) and marginal posterior density (bottom) for θ_1 , θ_2 , and λ in the similar-subpopulation setting. The two components have close parameter values, yielding broader and more overlapping posteriors. Red dashed lines mark the true values; the chains remain stable after burn-in and centre on these targets, indicating successful recovery despite weak separation.

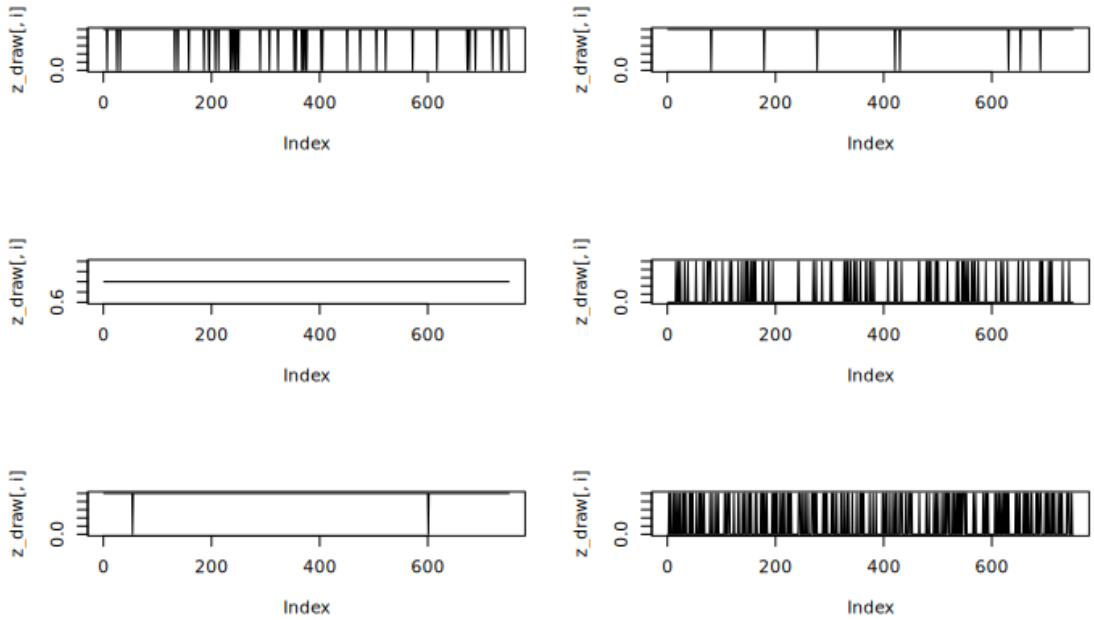


Figure 4.6: Posterior sample paths of the latent indicator z_i for six representative locations in the *similar-subpopulation* setting. Unlike the distinguishable case, several traces show frequent switches between 0 and 1, reflecting higher posterior uncertainty in class membership when component parameters are close. A few sites remain near a single state, indicating local identifiability, but overall the mixture is harder to separate under weak component separation.

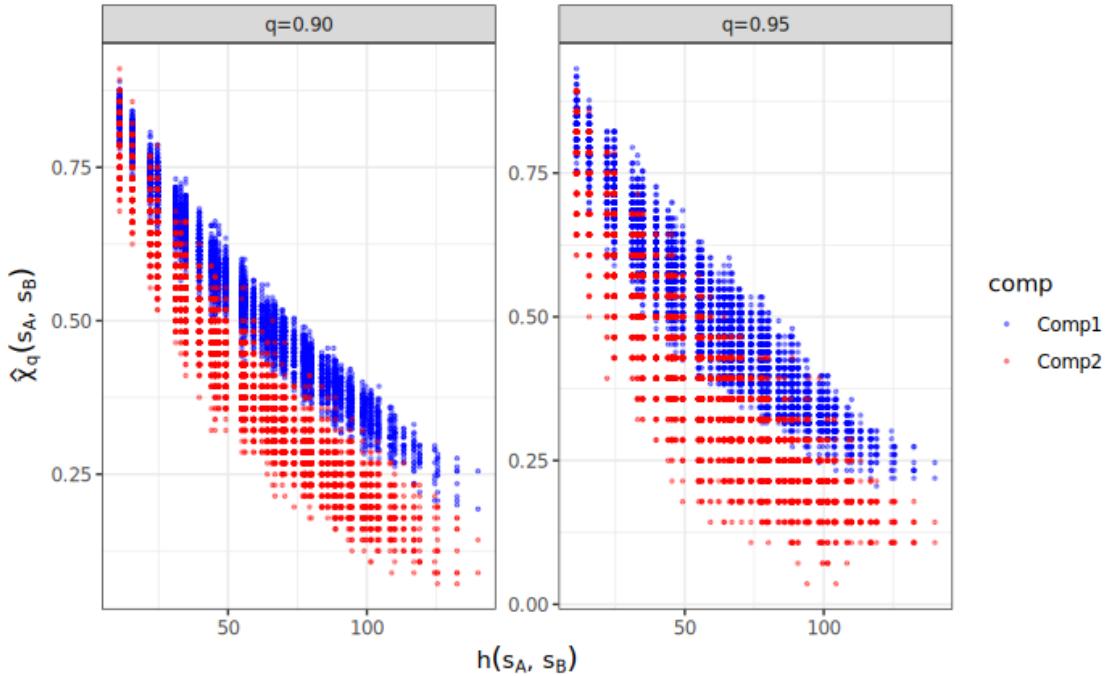


Figure 4.7: Empirical estimates of the upper tail dependence measure $\hat{\chi}_q(s_A, s_B)$ against spatial distance $h(s_A, s_B)$ for two mixture components corresponding to frontal (blue) and convective (red) rainfall events. Results are shown for quantile levels $q = 0.90$ (left) and $q = 0.95$ (right). Frontal rainfall exhibits markedly stronger extremal dependence, as $\hat{\chi}_q(s_A, s_B)$ decays more slowly with distance compared to convective rainfall, whose extremal dependence diminishes rapidly.

4.4 Posterior Concentration Under Increasing Sample Size

This section aims to investigate another property of one legit Bayesian Mixture Model, which is that the posterior distribution would be concentrated on the truth values as the sample size is increased. In this simulation, we keep previous simulation parameters setting: $\theta_1 = 50$, $\theta_2 = 40$ and $\lambda = 0.7$. For sample size, the candidates are $N = 500$, $N = 1000$ and $N = 2000$. The posterior distributions are shown in Figure 4.8. There's a clear trend of the posterior distribution centred to the ground truth values.

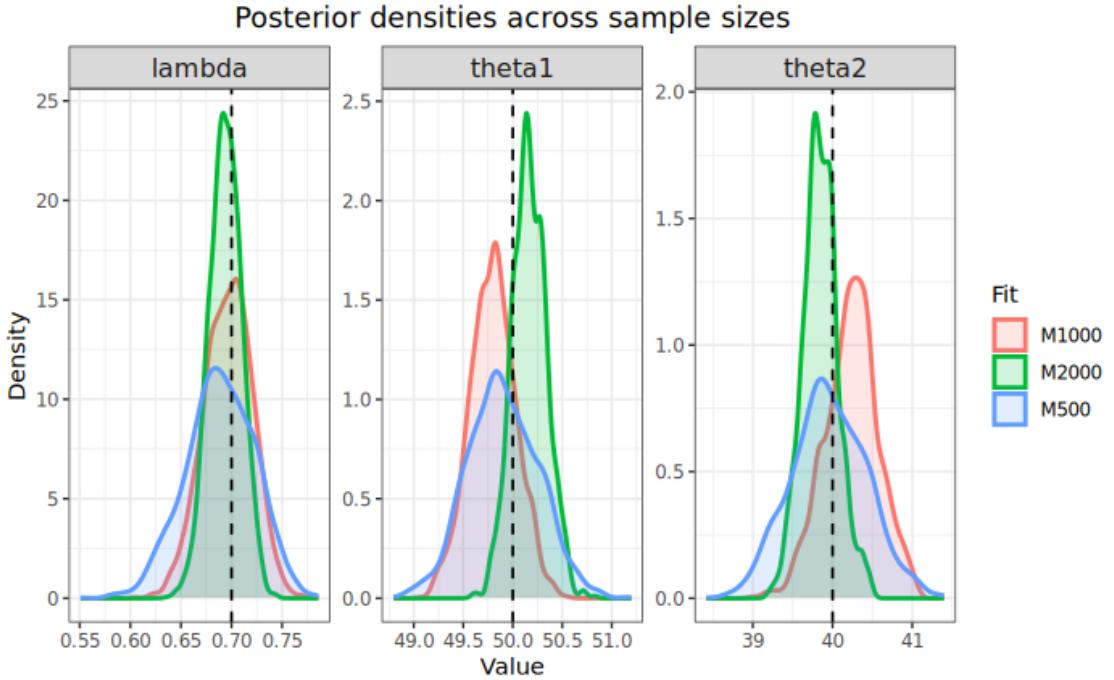


Figure 4.8: Posterior marginal densities for $(\lambda, \theta_2, \theta_2)$ under increasing sample size ($N \in \{500, 1000, 2000\}$) with true parameters $\theta_1 = 50$, $\theta_2 = 40$, and $\lambda = 0.7$. Curves correspond to fits M500 (blue), M1000 (red), and M2000 (green); vertical dashed lines mark the true values. As N grows, the posteriors contract and concentrate around the truth, illustrating posterior shrinkage and improved precision.

4.5 Time Variant Weight Function Simulation

In real-world precipitation, the probability that an event is frontal versus convective typically varies smoothly over time. Our mixture model should therefore recover this temporal pattern. To test this, we specify a continuous time-varying mixing weight $\lambda(t) \in [0, 1]$ and examine whether the posterior distribution of the indicator z (its posterior mean) tracks the predefined $\lambda(t)$ across the timeline.

Firstly, we define the time-varying mixing weight as

$$\lambda(t) = 0.5 - 0.5 \sin t \quad (4.5.1)$$

as shown in Figure 4.9.

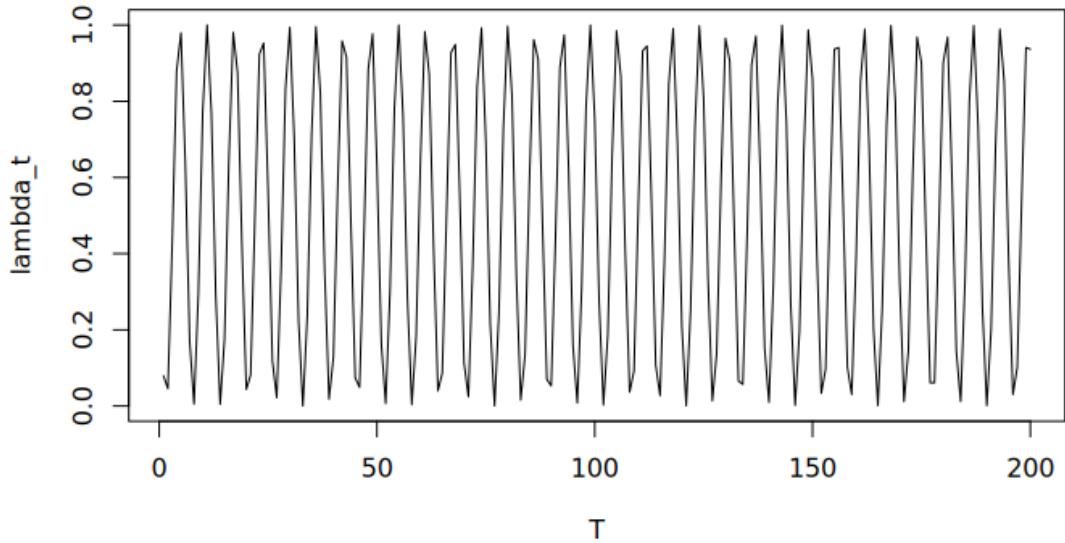


Figure 4.9: Time varying weight function

For the range parameters, we set $\theta_1 = 30$ and $\theta_2 = 20$ in this simulation. The mean posterior distribution of indicator z is plotted in Figure 4.11. The posterior mean of the indicator z fairly tracks the ground truth values with small discrepancies near rapid transitions where uncertainty appears.

Additionally, we investigate the generalisation of the proposed model to a different time-varying weight function to show the model can really capture the correct probability of frontal extremes.

$$\lambda(t) = 0.5 \sin t + 0.5 \quad (4.5.2)$$

as shown in Figure 4.10. We apply same parameters setting with previous simulation. And the posterior mean is plotted in Figure

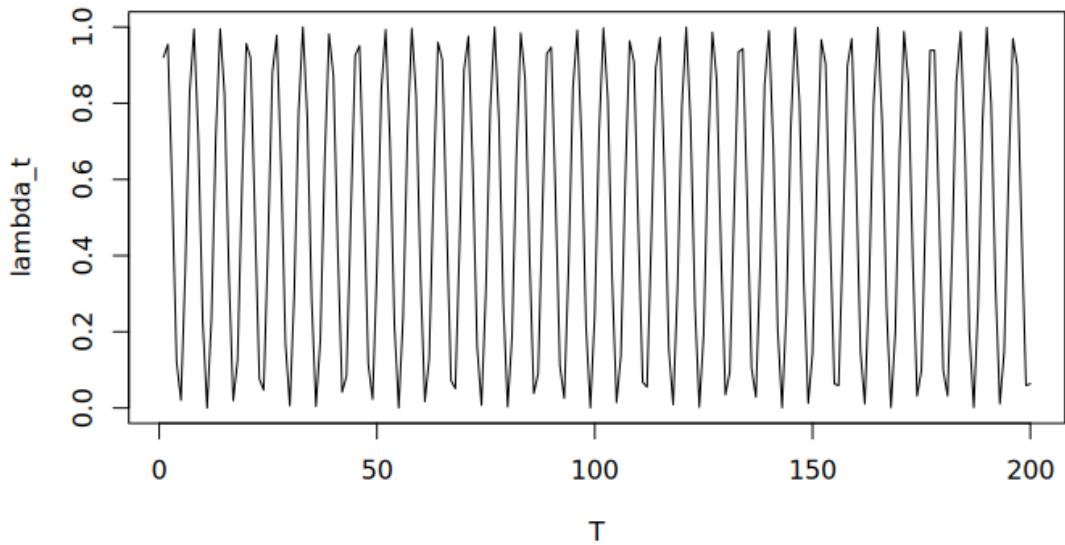


Figure 4.10: Time varying weight function

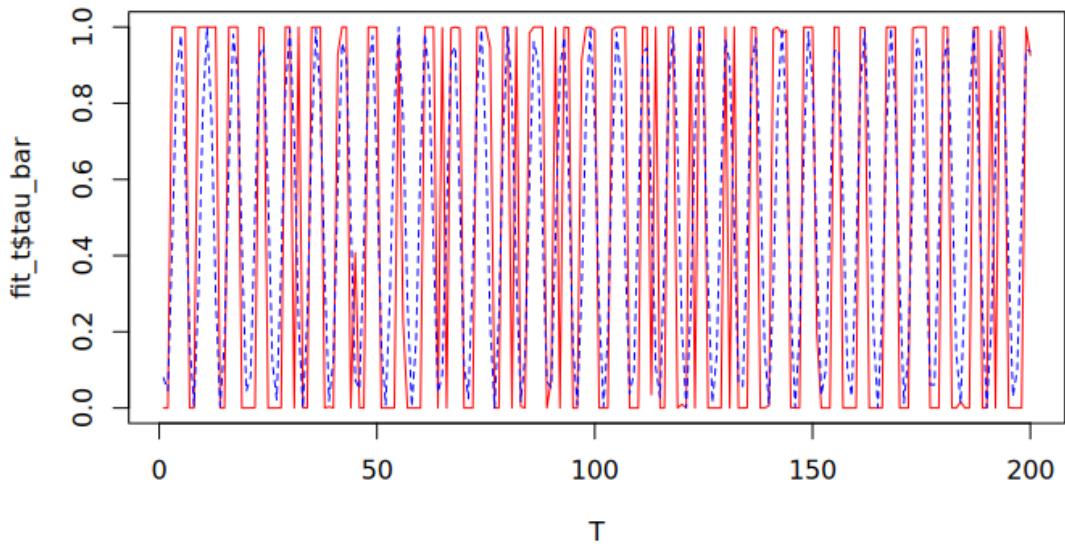


Figure 4.11: Time-varying classification: posterior mean of the indicator z_t (blue) versus the ground-truth mixing weight $\lambda(t)$ (red). The close alignment shows that the model recovers the temporal pattern of frontal-convective prevalence; small discrepancies appear mainly near rapid transitions where uncertainty is highest.

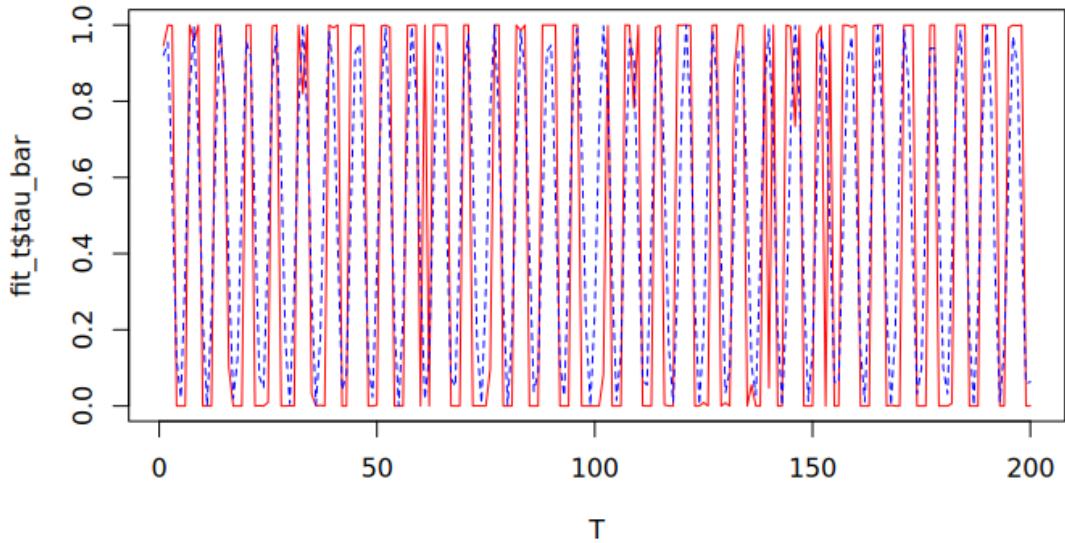


Figure 4.12: Time-varying classification with an *inverted* mixing function: posterior mean of the indicator z_t (blue) versus the inverted ground-truth curve $1 - \lambda(t)$ (red). The close correspondence indicates that the model adapts to the label swap (frontal \leftrightarrow convective), with minor deviations occurring near rapid transitions.

In summary, across all simulations, the proposed model performs well. The MCMC chains converge and remain stable around the true values; posterior classification is highly certain in the distinguishable setting (no label switching) and exhibits the expected uncertainty and occasional label switching when the component parameters are similar. As sample size increases, the marginal posteriors contract toward the ground truth (posterior concentration). Finally, the model accurately recovers the time-varying mixing pattern, with the posterior mean of the frontal–event probability closely tracking the true curve over time.

CHAPTER 5

Application

This chapter applies the proposed mixture model to precipitation data from the Tampa Bay region, Florida, USA. Our aims are to (i) apply the proposed model to fit the spatial range parameters of the extreme rainfall events in the hourly precipitation data (ii) assign the category memberships to the extreme rainfalls with frontal and convective. (iii) draw the mean posterior distribution of indicator z to show the time-varying patterns of rainfall.

5.1 Dataset

We analyse precipitation over the Tampa Bay region (Florida, USA) using weather radar measurements sampled every 15 minutes from 1995–2019, yielding 139,881 images with 4,449 spatial cells per image. Figure 5.1 shows a representative radar snapshot. For clarity and reproducibility, in this chapter we restrict attention to the full calendar year 1995.

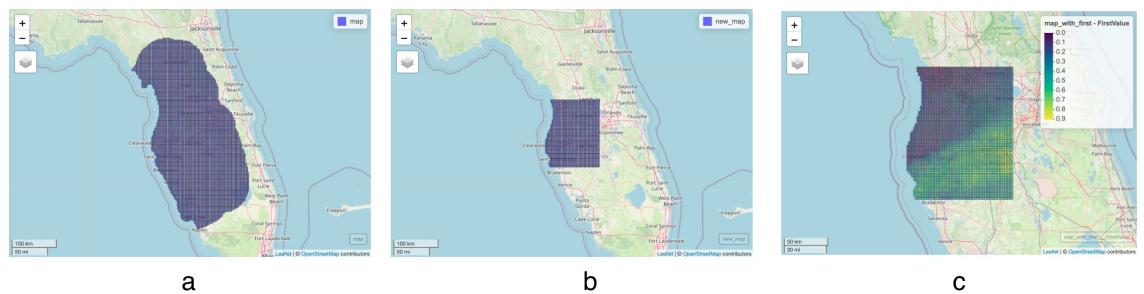


Figure 5.1: Spatial maps of the study area over Florida. (a) The full Tampa Bay coverage area, showing the complete spatial grid. (b) The restricted radar range of the Bay, focusing on the central analysis region. (c) One replicate of the daily rainfall field, mapped onto the radar grid with one replicate observed rainfall values visualised using a continuous colour scale.

In this thesis, we apply many preprocesses to the data for simplifying the computation since the whole spatial grid introduces high computational pressure to a standard computer. In terms of rainfall data, we build the dataset based on hourly time resolutions by summing them up. For spatial coordinates, one 14×14 squared grid is framed manually from the original huge area as shown in Figure 5.2. The coordinate system is EPSG:3577. EPSG:3577 (GDA94 / Australian Albers) is a projected coordinate reference system based on an Albers Equal Area Conic projection defined on the GDA94 datum (GRS80 ellipsoid), with units in metres. The latitude range is from 16,357,547 to 16,388,530 and the longitude range is from

-7,910,861 to -7,886,519. Finally, the coordinates are normalised into grid cell unit, which means the 1 unit in prior corresponding to the length of 1 unit in the grid.

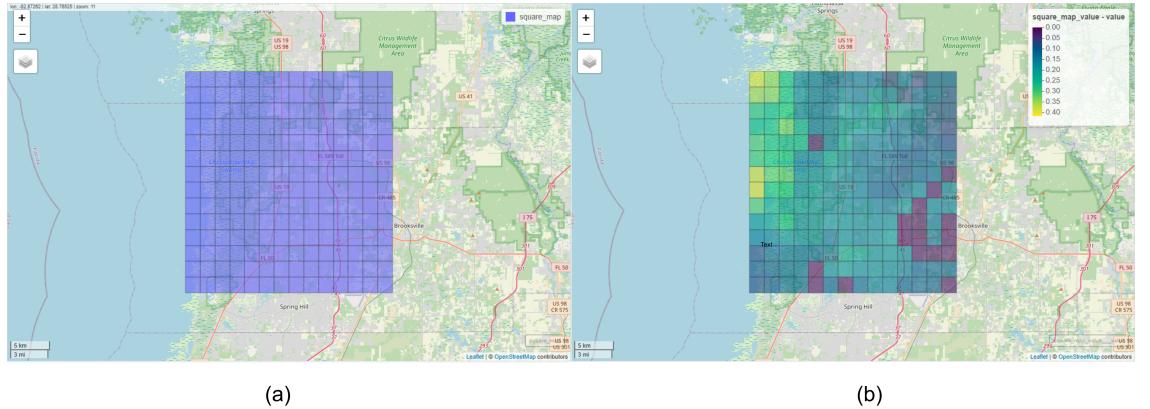


Figure 5.2: Framed map over spatial map. (a) The map is framed by 14×14 squared grid. (b) One realisation of hourly rainfall over the squared grid region.

Let $X \in \mathbb{R}^{T \times S}$ denote the rainfall field arranged as a matrix (rows = time points $t = 1, \dots, T$; columns = grid cells $s = 1, \dots, S$). We apply a sitewise marginal transform to a unit-Pareto scale using only strictly positive amounts. For each site s , let $\{x_{ts} : x_{ts} > 0\}$ be the nonzero values and set

$$u_{ts} = \frac{\text{rank}(x_{ts})}{m_s + 1} \quad \text{for } x_{ts} > 0, \quad u_{ts} = 0 \quad \text{for } x_{ts} = 0, \quad m_s = \sum_t \mathbf{1}\{x_{ts} > 0\}.$$

To avoid boundary issues, truncate $u_{ts} \in (\varepsilon, 1 - \varepsilon)$ with $\varepsilon = 10^{-10}$, then map to a heavy-tailed scale via a Generalised Pareto quantile transform with $(\xi, \beta, \mu) = (1, 1, 1)$:

$$y_{ts} = Q_{\text{GPD}}(u_{ts}; \xi = 1, \beta = 1, \mu = 1),$$

yielding the unit-Pareto matrix $Y = [y_{ts}]$. To focus on spatially extensive extremes, compute an L_1 risk score at each time point as the spatial mean,

$$R_t^{(1)} = \frac{1}{S} \sum_{s=1}^S y_{ts},$$

and define a threshold $u_{0.95}$ as the empirical 95th percentile of $\{R_t^{(1)}\}_{t=1}^T$. Time points with $R_t^{(1)} > u_{0.95}$ are retained for subsequent analysis, producing the extreme subset $Y^* = \{\mathbf{y}_t : R_t^{(1)} > u_{0.90}\}$ with $\mathbf{y}_t = (y_{t1}, \dots, y_{tS})$. We apply the same preprocess to both three datasets.

5.2 Results

For the hourly esolution analysis, we fit the model as the setting shown in tabel 5.1. Figure 5.3 displays trace plots and marginal posteriors for the range parameters and mixing weight. The posterior means are $\hat{\theta}_1 = 6.856$, $\hat{\theta}_2 = 2.094$, and $\hat{\lambda} = 0.419$. Given the working grid resolution (one grid unit ≈ 1.38 km), these correspond to an

Table 5.1: MCMC settings used in the Tempa Bay.

Schedule	
Iterations	10000
Burn-in	2000
Thinning interval	2
Kept draws	8000
Priors	
Mixing weight λ	Beta(1, 1)
Range θ_2	Gamma(1, 1)
Offset $\delta = \theta_1 - \theta_2$	Gamma(1, 1)
Proposals & Initials	
Proposal s.d. ($\log \theta_2$)	0.01
Proposal s.d. ($\log \delta$)	0.01
Initial λ	0.5
Initial θ_2	2
Initial δ	4
Anchor site	1
Verbosity interval	50

effective long range dependence of approximately 9.46 km for the frontal regime and a localised dependence of about 2.89 km for the convective regime, as illustrated in Figure 5.5. Posterior sample paths of the indicator z (Figure 5.4) show stable classifications with minimal switching, consistent with the simulation results for well separated components. The L_1 risk based spatial scale is 6.17 km, slightly larger than but within two standard deviations of the estimate reported by [Zhong et al. \(2024\)](#) as shown in table 5.2. Overall, the proposed model aligns with existing findings while jointly capturing the distinct spatial footprints of frontal and convective extremes.

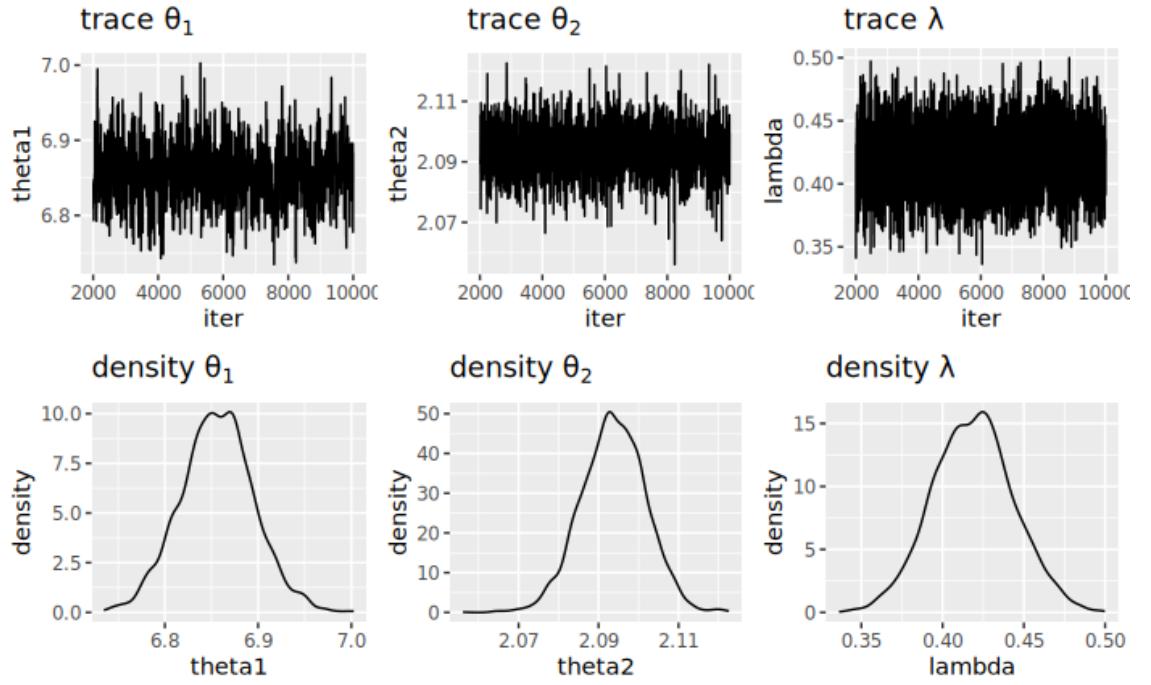


Figure 5.3: Tampa Bay application: trace (top) and marginal posterior density (bottom) for θ_1 , θ_2 , and mixing weight λ . The chains show stable mixing after burn-in with no apparent drifts, and the marginal posteriors are unimodal and reasonably concentrated, indicating well-identified spatial ranges and mixture proportion for the Tampa Bay rainfall field.

Table 5.2: Model Comparison and Proposed Mixture Results

Model	Risk Function	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}$
BR	L_1		5.91 (1.71)	
sBR	L_1		5.81 (1.74)	
Mixture Model	L_1	9.46 (0.054)	2.89 (0.011)	6.17 (0.028)

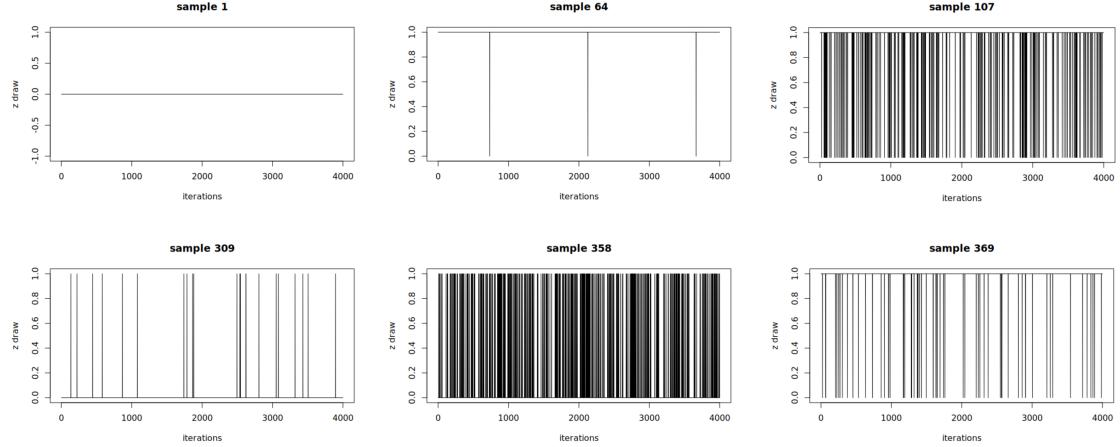


Figure 5.4: Posterior sample paths of the latent indicator z_i for six representative locations ($i \in \{1, 64, 107, 309, 385, 369\}$). Each panel shows z_i across the kept MCMC iterations after burn-in. The plots show the different confidence to assign the membership to different samples

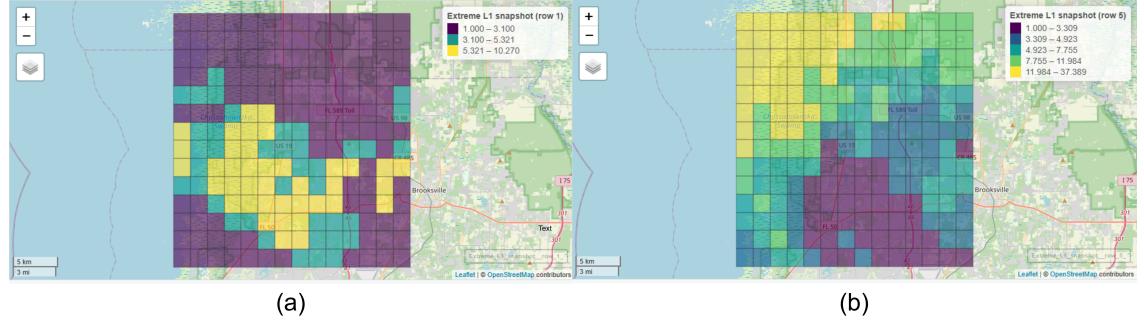


Figure 5.5: Tampa Bay (hourly resolution): two representative extreme snapshots used for regime assignment. Each grid cell side is ≈ 1.38 km; colours map intensity with yellow = higher values. (a) Localised, cellular yellow patches indicative of a short-range convective pattern (estimated range ≈ 2.76 km from $\hat{\theta}_2 = 2.093$). (b) Broad, coherent swaths of elevated intensity consistent with a long-range frontal pattern (estimated range ≈ 9.46 km from $\hat{\theta}_1 = 6.856$). These examples align with the posterior mixing weight $\hat{\lambda} = 0.420$.

We further assess the fitted spatial regimes using an empirical extremal dependence function. For each site pair (s_A, s_B) and high quantile $q \in \{0.90, 0.95\}$, we compute the empirical $\hat{\chi}_q(s_A, s_B)$ and plot it against intersite distance $h(s_A, s_B)$, colouring points by their posterior component assignment. Figure 5.6 shows a clear contrast: the convective component (red) exhibits a rapid decay of extremal dependence with distance, approaching zero at moderate separations, whereas the frontal component (blue) maintains appreciable dependence over longer ranges. These patterns are consistent with the short range versus long range interpretations of the two regimes.

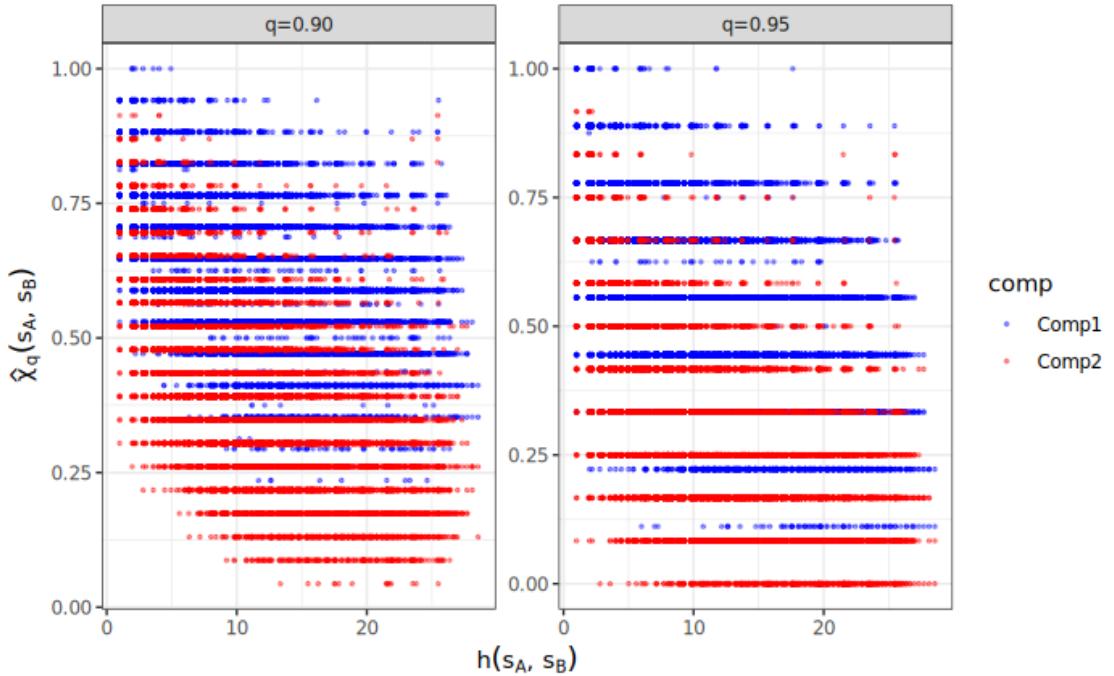


Figure 5.6: Empirical extremal dependence by distance for Tampa Bay. Each point is a site pair (s_A, s_B) with vertical coordinate $\hat{\chi}_q(s_A, s_B)$ and horizontal axis the intersite distance $h(s_A, s_B)$, shown for $q = 0.90$ (left) and $q = 0.95$ (right). Colours denote posterior regime: Comp1 (frontal) and Comp2 (convective). The convective pairs display a steep decline of $\hat{\chi}_q$ toward zero with distance, while the frontal pairs retain non-negligible extremal dependence at larger separations, corroborating the fitted short vs. long range structure.

Figure 5.7 plots the posterior mean of the indicator z over time. A clear seasonal pattern emerges: convective extremes occur much more frequently during the wet season (June–September), accompanied by greater temporal variability in rainfall intensity. In the dry season (November to early June), frontal extremes predominate in the Tampa Bay record, but during the wet season convective events are more common and conditions shift more rapidly.

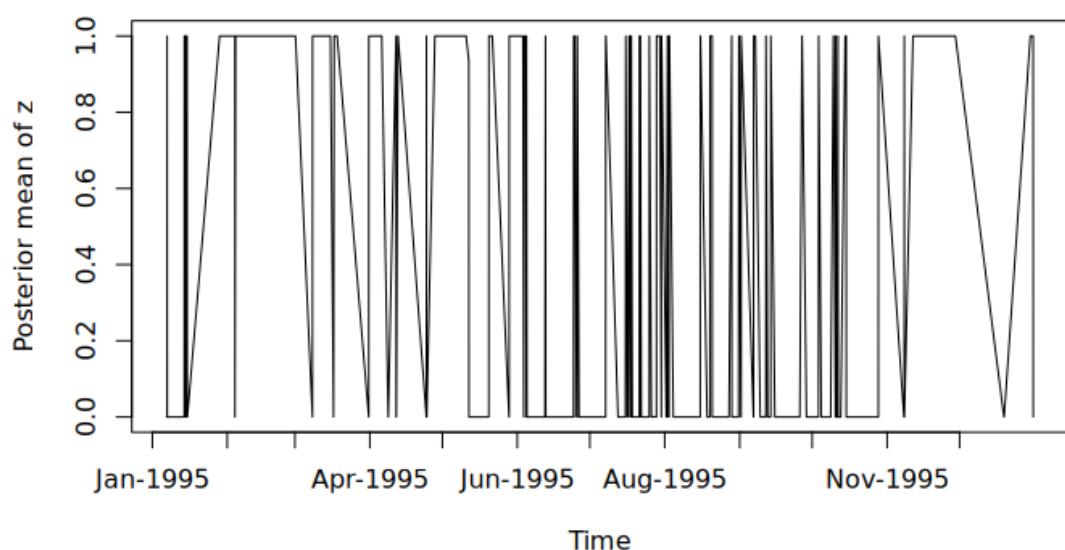


Figure 5.7: The posterior mean of indicator z : In the wet season (June - September) Tempa Bay contains more localised convective rainfall and it changes rapidly. In dry season (November - June) Tempa Bay includes more frontal rainfall and it is more stable than wet season.

CHAPTER 6

Discussion

This chapter reviewed a Bayesian mixture model built on the \mathcal{R} -Pareto process, together with simulation studies and an application to extreme precipitation. Our goal was to demonstrate how a single, coherent framework can jointly represent multiple types of extremes and deliver reliable classification alongside parameter estimation.

Over the past eighty years, Extreme Value Theory (EVT) has expanded from univariate to multivariate settings, with substantial progress on likelihood approximations in high dimensions. The \mathcal{R} -Pareto process, defined via a analyser chosen risk function $r(\cdot)$, offers a flexible route to model threshold exceedances in space. Recent advances on approximating general L_p risk functions ([Zhong et al., 2024](#)) further strengthen the case for \mathcal{R} -Pareto modelling as a state-of-the-art approach for spatial extremes.

A persistent gap in the literature is the absence of a single statistical model that jointly accommodates distinct physical drivers of extremes within one inferential pipeline. Extreme precipitation provides a natural testbed because it typically comprises two archetypes frontal and convective events with different spatial dependence ranges. We address this by proposing a Bayesian mixture of \mathcal{R} -Pareto components, where the Brown–Resnick semivariogram encodes type specific spatial ranges. The mixture structure yields simultaneous inference on range parameters and latent membership (frontal vs. convective).

In simulation, the model accurately recovers the true range dependence parameters across both well separated and partially overlapping subpopulations. The empirical separation diagnostic $\widehat{\chi^2}(S_A, S_B)$ confirms that inferred memberships are statistically distinguishable in both regimes. Applied to Tampa Bay, the fitted model delivers plausible long and short range estimates consistent with the frontal/convective dichotomy and seasonally coherent membership patterns that align with independent hydrometeorological evidence, including patterns documented by the U.S. Geological Survey ([Yates et al., 2011](#)). Together, these results indicate that the model captures salient features of extremal rainfall while providing interpretable classifications.

There are still several simplifications leave room for refinement:

- *Fixed smoothness and isotropy.* We fixed the smoothness ϑ and did not model anisotropy in the semivariogram. Both can be incorporated with additional computational budget.
- *Spatial domain size.* For computational reasons we did not fit the model over the full spatial lattice. The moedl can be fitted in full spatial region given sufficient computing resources.
- *Likelihood choice.* We employed an L_1 -based angular likelihood. While [Zhong et al. \(2024\)](#) argue that rejection sampling can approximate other L_p likelihoods at very high thresholds, practical sample sizes often fall short of asymptotic regimes; this can bias estimation relative to a “true” $L_{p \neq 1}$ data generating process.
- *Fixed number of components.* We assumed two subpopulations. Allowing an unknown number of components M would enable richer structure but requires hierarchical priors and additional identification safeguards.

Several directions appear promising:

1. *Richer spatial kernels.* Joint inference for (θ, ϑ) and anisotropy within each component.
2. *General risk functions.* Integrate L_∞ or score matching based likelihoods directly, and compare to L_1 -based approximations under finite thresholds.
3. *Adaptive model complexity.* Place a prior on M to discover multiple extremal regimes.
4. *Location mixtures.* Rather than assuming all sites share the same event type at a given time, allow site level mixture memberships so different locations can simultaneously exhibit frontal or convective extremes. This can be implemented via spatially varying mixture weights or site specific latent labels with spatial smoothing.
5. *Broader applications.* Extend to other environmental extremes (temperature, wind) and to domains such as infrastructure risk and finance, where multiple tail behaviours coexist.

Recent work explores joint estimation and clustering of multiple extremal behaviors, including precipitation type separation ([Richards et al., 2023](#)). Compared with approaches that require extensive hyperparameter tuning or external pre-classification, our mixture formulation learns memberships and spatial ranges within a single probabilistic model, reducing manual intervention and propagating uncertainty consistently. This integrated treatment is particularly valuable when regime boundaries are fuzzy or seasonally evolving.

The proposed Bayesian \mathcal{R} -Pareto mixture closes an important gap by modeling multiple extremal regimes jointly while retaining interpretability and delivering accurate parameter estimates. With the methodological extensions outlined above, the framework can become a versatile tool for analyzing complex extremal phenomena across environmental science and beyond.

CHAPTER 7

Conclusion

Spatial Extreme Value Theory (SEVT) is crucial for tackling extreme data in climate change, sea levels, and forest fires. Over the past decades, researchers have extended EVT from univariate to multivariate settings and have continuously developed better likelihood approximations for high-dimensional cases. However, there remains a significant gap in jointly modelling multiple types of extreme events. Mixture modelling offers a straightforward approach for data with multiple subpopulations. The proposed Bayesian mixture model based on the state-of-the-art \mathcal{R} -Pareto process not only yields precise parameter estimates but also models multiple extremes jointly with a single likelihood function, an approach that is both novel and insightful.

In this thesis, we build a Bayesian mixture model to handle two types of extreme rainfall: frontal extremes, which cause long range precipitation, and convective extremes, which lead to localised, high intensity precipitation. Specifically, in Chapter 2 we review EVT from classical univariate to multivariate models, tracing the development from the last century to contemporary state-of-the-art methods. In Chapter 3, we detail the proposed model's variables, parameters, likelihood, mixture structure, and the use of Bayes' rule for posterior inference. We also provide an example illustrating conjugate priors and their advantages in Bayesian computation. In Chapter 4, we present comprehensive simulations to validate the proposed model. We examine both clearly distinguishable and closely similar subpopulations, and in both scenarios the model accurately recovers the ground-truth parameters. The posterior mean of the indicator z concentrates near 0 or 1, with more frequent switching between the two when subpopulations are similar. We further compute the empirical $\hat{\chi}^2(S_A, S_B)$ between locations and show that the model assigns memberships correctly: frontal rainfall exhibits a longer range and slower decay than convective rainfall. Finally, we test whether the posterior mean of z tracks pre-defined time-varying mixture weights; the fitted posterior means closely match the target weights. In Chapter 5, we apply the model to rainfall data from Tampa Bay, Florida, USA. Our estimated range parameters align well with existing literature, and the empirical $\hat{\chi}^2(S_A, S_B)$ highlights the distinct characteristics of the two rainfall types. Surprisingly, the seasonal patterns identified independently by our model are consistent with U.S. Geological Survey findings in Florida, providing strong evidence that our approach captures key data attributes and performs well in practice.

There are several promising directions for future work. Firstly, we can relax the assumption on the number of subpopulations, since many applications may involve

more than two regimes. In addition, we can incorporate location-level mixtures rather than assuming all sites share the same event type at the same time; some locations may simultaneously exhibit frontal or convective extremes. Further, we can consider alternative risk functions L_p within the \mathcal{R} -Pareto framework to capture extremes with different characteristics.

Appendices

APPENDIX A

Simulated Data Generation Algorithms

This Chapter shows the algorithms we applied for simulating the data in our Chapter 4. We compute the semivariogram function 3.1.4 based on fixed $\vartheta = 1.5$ and the range parameter θ in algorithm 1.

Algorithm 1 `BUILDGAMMA($\mathcal{S}, \vartheta, \theta$)` (Power semivariogram with fixed $\vartheta = 1.5$)

Require: Spatial locations $\mathcal{S} = \{s_1, \dots, s_d\} \subset \mathbb{R}^q$, smoothness $\vartheta = 1.5$, range $\theta > 0$

Ensure: $\Gamma \in \mathbb{R}^{d \times d}$ with entries $\Gamma_{ij} = \gamma(s_i, s_j) = (\|s_i - s_j\|/\theta)^\vartheta$ and $\Gamma_{ii} = 0$

- 1: Compute the pairwise distance matrix $D \in \mathbb{R}^{d \times d}$, where $D_{ij} \leftarrow \|s_i - s_j\|$
- 2: $\Gamma \leftarrow (D/\theta)^\vartheta$
- 3: **for** $i = 1$ to d **do**
- 4: $\Gamma_{ii} \leftarrow 0$
- 5: **end for**
- 6: **return** Γ

Based on the semivariogram function, we can compute the covariance between each location and we store anchor site j , Cholesky decomposition L_j and semivariogram function about that site Γ_j for fast reconstruction when we build the Brown-Resnick model. The algorithm is shown in algorithm 2.

Algorithm 2 `PREPBR(Γ)`

Require: $\Gamma \in \mathbb{R}^{d \times d}$

Ensure: List $\{\mathcal{P}_j\}_{j=1}^d$ with $\mathcal{P}_j = (\text{idx}_j, L_j, \Gamma_{.j})$

- 1: $d \leftarrow \text{nrow}(\Gamma)$
- 2: **for** $j = 1$ to d **do**
- 3: $\text{idx}_j \leftarrow \{1, \dots, d\} \setminus \{j\}$
- 4: $\mathbf{g} \leftarrow \Gamma_{\text{idx}_j, j}$
- 5: $\Sigma \leftarrow \mathbf{g}\mathbf{1}^\top + \mathbf{1}\mathbf{g}^\top - \Gamma_{\text{idx}_j, \text{idx}_j}$
- 6: $\Sigma \leftarrow \Sigma + \varepsilon I_{d-1}$
- 7: $L_j \leftarrow \text{chol}(\Sigma)$
- 8: $\mathcal{P}_j \leftarrow (\text{idx}_j, L_j, \Gamma_{.j})$
- 9: **end for**
- 10: **return** $\{\mathcal{P}_j\}_{j=1}^d$

We simulate the data as same procedure described in chapter 4 as shown in algorithm 3.

Algorithm 3 SIMULATORPARETO($m, \{\mathcal{P}_j\}_{j=1}^d$)

Require: Number of samples m , prep list $\{\mathcal{P}_j\}_{j=1}^d$

Ensure: $Z \in \mathbb{R}^{m \times d}$, rows are simulated \mathcal{R} -Pareto vectors

```

1: for  $k = 1$  to  $m$  do
2:    $J \leftarrow 1$ 
3:    $(\text{idx}, L, \Gamma_{\cdot J}) \leftarrow \mathcal{P}_J$ 
4:   Draw  $\mathbf{z} \sim \mathcal{N}_{d-1}(\mathbf{0}, I)$  and set  $g_{\text{idx}} \leftarrow L^\top \mathbf{z}$ ,  $g_J \leftarrow 0$ 
5:   Form  $g \in \mathbb{R}^d$  with entries  $g_{\text{idx}}$  and  $g_J$ 
6:    $y \leftarrow \exp(g - \Gamma_{\cdot J})$ 
7:    $y \leftarrow y/y_J$ 
8:    $v \leftarrow y/\sum_{i=1}^d y_i$ 
9:   Draw  $r \sim \text{GPD}(\text{loc} = 1, \text{scale} = 1, \text{shape} = 1)$ 
10:   $Z_{k,.} \leftarrow r v^\top$ 
11: end for
12: return  $Z$ 

```

APPENDIX B

Bayesian MCMC in Log Space

This chapter will introduce the log version likelihood function, prior distribution, posterior and updating rules.

B.1 Log Likelihood Function

log-likelihood without indicators. For numerical stability, log-likelihood is a common choice. The likelihood function of this model given observations y , parameter θ and λ is

$$L(y|\theta, \lambda) = \prod_{i=1}^n f(y_i|\theta, \lambda) \quad (\text{B.1.1})$$

where $f(y_i|\theta, \lambda)$ is given by equation 3.4.1. Therefore, the log-likelihood function of this model is

$$\ell(y|\theta, \lambda) = \log L(y|\theta, \lambda) \quad (\text{B.1.2})$$

$$= \sum_{i=1}^n \log(f(y_i|\theta, \lambda)) \quad (\text{B.1.3})$$

$$= \sum_{i=1}^n \log\left(\sum_{m=1}^M \lambda_m f_m(y_i|\theta_m)\right) \quad (\text{B.1.4})$$

Log-likelihood with indicators. Let $z_{im} \in \{0, 1\}$ with $\sum_m z_{im} = 1$ and $z_i \mid \lambda \sim \text{Multinomial}(1 : \lambda_1, \dots, \lambda_M)$. The complete-data likelihood and its log form are

$$L(y, z \mid \theta, \lambda) = \prod_{i=1}^n \prod_{m=1}^M [\lambda_m f_m(y_i \mid \theta_m)]^{z_{im}}, \quad (\text{B.1.5})$$

$$\ell(y, z \mid \theta, \lambda) = \sum_{i=1}^n \sum_{m=1}^M z_{im} \left(\log \lambda_m + \log f_m(y_i \mid \theta_m) \right). \quad (\text{B.1.6})$$

B.2 Log Prior Distribution

We assume independence $\pi(\theta, \lambda) = \pi(\theta)\pi(\lambda)$.

Mixture weights. The weight vector $\lambda = (\lambda_1, \dots, \lambda_M)$ follows a Dirichlet prior:

$$\lambda \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_M), \quad \log \pi(\lambda) = \sum_{m=1}^M (\alpha_m - 1) \log \lambda_m. \quad (\text{B.2.1})$$

Component parameters. When each $\theta_m > 0$ (e.g., Brown–Resnick range parameters), we use independent Gamma priors in shape–rate form:

$$\theta_m \sim \text{Gamma}(a_m, b_m), \quad \pi(\theta_m) = \frac{b_m^{a_m}}{\Gamma(a_m)} \theta_m^{a_m-1} e^{-b_m \theta_m}, \quad (\text{B.2.2})$$

and therefore

$$\log \pi(\theta) = \sum_{m=1}^M \left[(a_m - 1) \log \theta_m - b_m \theta_m + a_m \log b_m - \log \Gamma(a_m) \right]. \quad (\text{B.2.3})$$

For vector parameters $\theta_m = (\theta_{m1}, \dots)$, take independent Gamma priors across entries:

$$\log \pi(\theta) = \sum_{m=1}^M \sum_j \left[(a_{m,j} - 1) \log \theta_{m,j} - b_{m,j} \theta_{m,j} + a_{m,j} \log b_{m,j} - \log \Gamma(a_{m,j}) \right].$$

B.3 Log Parameterisation

For mixture weights. Since λ is constrained to the simplex, it is convenient to work directly with $\log \pi(\lambda)$ as in (B.2.1).

For component parameters. For positive θ_m , we reparameterise via $\eta_m = \log \theta_m$. This ensures unconstrained proposals and symmetric random walks. By change of variables,

$$\log \pi(\eta) = \sum_{m=1}^M \left[a_m \eta_m - b_m e^{\eta_m} + a_m \log b_m - \log \Gamma(a_m) \right], \quad (\text{B.3.1})$$

where $a_m \eta_m$ comes from $(a_m - 1) \log \theta_m$ plus the Jacobian term $+\eta_m$ due to $\theta_m = e^{\eta_m}$.

Complete-data Log Posterior

Combining likelihood, priors, and indicators z , the complete-data log posterior is

$$\log \pi(\theta, \lambda, z | y) \propto \ell(y, z | \theta, \lambda) + \log \pi(\lambda) + \sum_{m=1}^M \log \pi_m(\theta_m). \quad (\text{B.3.2})$$

This expression provides the target distribution for the Metropolis–within–Gibbs sampler.

B.4 Log Conditional Updates

We update (θ, λ, z) in sequence.

Update of θ_m . Write the log full conditional kernel for θ_m as

$$\log \pi(\theta_m | y, z, \lambda, \theta_{-m}) \propto \sum_{i=1}^n z_{im} \log f_m(y_i | \theta_m) + \log \pi_m(\theta_m). \quad (\text{B.4.1})$$

If θ_m contains positive components, propose on $\eta_m = \log \theta_m$ with a symmetric random walk (e.g. $\eta'_m = \eta_m + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \Sigma)$), transform back $\theta'_m = e^{\eta'_m}$, and use

the MH log-acceptance

$$\begin{aligned} \log \alpha(\theta_m \rightarrow \theta'_m) &= \left[\sum_i z_{im} \log f_m(y_i | \theta'_m) + \log \pi_m(\theta'_m) \right] - \left[\sum_i z_{im} \log f_m(y_i | \theta_m) + \log \pi_m(\theta_m) \right] \\ &\quad + \left(\sum_j \log \theta'_{m,j} - \sum_j \log \theta_{m,j} \right), \end{aligned} \quad (\text{B.4.2})$$

where the last line is the Jacobian difference for all positive coordinates of θ_m .

Update of λ . Let $n_m = \sum_{i=1}^n z_{im}$. The log full conditional kernel is

$$\log \pi(\lambda | y, \theta, z) \propto \sum_{m=1}^M (n_m + \alpha_m - 1) \log \lambda_m, \quad (\text{B.4.3})$$

i.e. $\lambda | \cdot \sim \text{Dirichlet}(\alpha_1 + n_1, \dots, \alpha_M + n_M)$.

Update of z_i . Define the per-observation log-scores

$$a_{im} = \log \lambda_m + \log f_m(y_i | \theta_m). \quad (\text{B.4.4})$$

Then the posterior class probabilities are obtained by normalisation:

$$\tau_{im} = \Pr(z_{im} = 1 | y_i, \theta, \lambda) = \frac{\exp(a_{im})}{\sum_{k=1}^M \exp(a_{ik})} \quad (\text{B.4.5})$$

and we draw

$$z_i \sim \text{Multinomial}(1 : \tau_{i1}, \dots, \tau_{iM}). \quad (\text{B.4.6})$$

In summary, the algorithm 4 shows the overall process we actually utilise in my thesis to fit the Bayesian MCMC process.

Algorithm 4 Metropolis Hastings for 2-component r-Pareto BR mixture

Require: X , coordinates; $\lambda \sim \text{Beta}(a_\lambda, b_\lambda)$, $\theta_2 \sim \text{Gamma}(a_2, b_2)$, $\delta \sim \text{Gamma}(a_\delta, b_\delta)$; $\theta_1 = \theta_2 + \delta$; $N, B, t, s_{\eta_2}, s_{\eta_\delta}$.

Ensure: Draws of $\lambda, \theta_1, \theta_2; \bar{\tau}_i; \hat{z}_i$.

- 1: Initialize $\lambda^{(0)} \in (0, 1)$, $\theta_2^{(0)} > 0$, $\delta^{(0)} > 0$; $\theta_1^{(0)} = \theta_2^{(0)} + \delta^{(0)}$.
- 2: Compute $\ell_1^{(0)}[i] = \log f_{\text{rP}}(x^{(i)}; \Sigma(\theta_1^{(0)}))$, $\ell_2^{(0)}[i] = \log f_{\text{rP}}(x^{(i)}; \Sigma(\theta_2^{(0)}))$.
- 3: $\tau_i^{(0)} = \frac{\lambda^{(0)} e^{\ell_1^{(0)}[i]}}{\lambda^{(0)} e^{\ell_1^{(0)}[i]} + (1 - \lambda^{(0)}) e^{\ell_2^{(0)}[i]}}$, sample $z_i^{(0)} \sim \text{Bernoulli}(\tau_i^{(0)})$.
- 4: **for** $iter = 1$ **to** N **do**
- 5: **Update** $\theta \equiv (\theta_2, \delta)$:
- 6: $\eta_2 = \log \theta_2^{(iter-1)}$, $\eta_\delta = \log \delta^{(iter-1)}$.
- 7: Propose $\eta'_2 \sim \mathcal{N}(\eta_2, s_{\eta_2}^2)$, $\eta'_\delta \sim \mathcal{N}(\eta_\delta, s_{\eta_\delta}^2)$.
- 8: $\theta'_2 = \exp(\eta'_2)$, $\delta' = \exp(\eta'_\delta)$, $\theta'_1 = \theta'_2 + \delta'$.
- 9: Compute $\ell'_1[i], \ell'_2[i]$. Let

$$\mathcal{L} = \sum_i [z_i^{(iter-1)} \ell_1^{(cur)}[i] + (1 - z_i^{(iter-1)}) \ell_2^{(cur)}[i]] + \log \pi(\theta_2^{(cur)}) + \log \pi(\delta^{(cur)}) + \log \theta_2^{(cur)} + \log \delta^{(cur)}$$

$$\mathcal{L}' = \sum_i [z_i^{(iter-1)} \ell'_1[i] + (1 - z_i^{(iter-1)}) \ell'_2[i]] + \log \pi(\theta'_2) + \log \pi(\delta') + \log \theta'_2 + \log \delta'.$$

- 10: Accept with prob. $\min\{1, e^{\mathcal{L}' - \mathcal{L}}\}$; if accepted set $\theta_2^{(iter)} = \theta'_2$, $\delta^{(iter)} = \delta'$, $\theta_1^{(iter)} = \theta'_1$, $\ell_1^{(iter)} = \ell'_1$, $\ell_2^{(iter)} = \ell'_2$; else keep previous.
 - 11: **Update** λ :
 - 12: $n_1 = \sum_i z_i^{(iter-1)}$, $n_0 = m - n_1$; sample $\lambda^{(iter)} \sim \text{Beta}(a_\lambda + n_1, b_\lambda + n_0)$.
 - 13: **Update** z :
 - 14: $\tau_i = \frac{\lambda^{(iter)} e^{\ell_1^{(iter)}[i]}}{\lambda^{(iter)} e^{\ell_1^{(iter)}[i]} + (1 - \lambda^{(iter)}) e^{\ell_2^{(iter)}[i]}}$; sample $z_i^{(iter)} \sim \text{Bernoulli}(\tau_i)$.
 - 15: **if** $iter > B$ **and** $(iter - B) \bmod t = 0$ **then**
 - 16: Store $\lambda^{(iter)}, \theta_1^{(iter)}, \theta_2^{(iter)}$; update $\bar{\tau}_i$.
 - 17: **end if**
 - 18: **end for**
 - 19: **return** stored draws; MH acceptance rate; $\bar{\tau}_i; \hat{z}_i = \mathbf{1}\{\bar{\tau}_i > 0.5\}$.
-

References

- Balkema, A. A. and De Haan, L. (1974) Residual life time at great age. *The Annals of probability* pp. 792–804.
- Brown, B. M. and Resnick, S. I. (1977) Extreme values of independent stochastic processes. *Journal of Applied Probability* **14**, 732–739.
- Candia, C. and Herrera, R. (2024) An empirical review of dynamic extreme value models for forecasting value at risk, expected shortfall and expectile. *Journal of Empirical Finance* **77**, 101488.
- Catto, J. L. and Pfahl, S. (2013) The importance of fronts for extreme precipitation. *Journal of Geophysical Research: Atmospheres* **118**, 10–791.
- Chaudhry, S. M., Ahmed, R., Huynh, T. L. D. and Benjasak, C. (2022) Tail risk and systemic risk of finance and technology (fintech) firms. *Technological Forecasting and Social Change* **174**, 121191.
- De Haan, L. (1984) A spectral representation for max-stable processes. *The annals of probability* pp. 1194–1204.
- De Haan, L. and Ferreira, A. (2006) *Extreme value theory: an introduction*. Springer.
- Dombry, C. and Ribatet, M. (2015) Functional regular variations, pareto processes and peaks over threshold. *Statistics and its Interface* **8**, 9–17.
- Ferreira, A. and De Haan, L. (2014) The generalized pareto process; with a view towards application and simulation .
- Fisher, R. A. and Tippett, L. H. C. (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical proceedings of the Cambridge philosophical society*, volume 24, pp. 180–190.
- de Fondeville, R. and Davison, A. C. (2018) High-dimensional peaks-over-threshold inference. *Biometrika* **105**, 575–592.
- de Fondeville, R. and Davison, A. C. (2022) Functional peaks-over-threshold analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **84**, 1392–1422.
- Fuentes, F., Herrera, R. and Clements, A. (2025) Tail risk dynamics of banks with score-driven extreme value models. *Journal of Empirical Finance* **81**, 101593.
- Gnedenko, B. (1943) Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics* **44**, 423–453.
- Hazra, A., Huser, R. and Bolin, D. (2025) Efficient modeling of spatial extremes over large geographical domains. *Journal of Computational and Graphical Statistics* **34**, 795–811.
- Huser, R., Stein, M. L. and Zhong, P. (2024) Vecchia likelihood approximation for accurate and fast inference with intractable spatial max-stable models. *Journal of Computational and Graphical Statistics* **33**, 978–990.

- Jenkinson, A. F. (1955) The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal meteorological society* **81**, 158–171.
- Kabluchko, Z., Schlather, M. and De Haan, L. (2009) Stationary max-stable fields associated to negative definite functions .
- Karpov, O. E., Grubov, V. V., Maksimenko, V. A., Kurkin, S. A., Smirnov, N. M., Utyashev, N. P., Andrikov, D. A., Shusharina, N. N. and Hramov, A. E. (2022) Extreme value theory inspires explainable machine learning approach for seizure detection. *Scientific Reports* **12**, 11474.
- Leadbetter, M. R., Lindgren, G. and Rootzén, H. (2012) *Extremes and related properties of random sequences and processes*. Springer Science & Business Media.
- Lederer, J. and Oesting, M. (2023) Extremes in high dimensions: Methods and scalable algorithms. *arXiv preprint arXiv:2303.04258* .
- Padoan, S. A., Ribatet, M. and Sisson, S. A. (2010) Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association* **105**, 263–277.
- Pickands III, J. (1975) Statistical inference using extreme order statistics. *the Annals of Statistics* pp. 119–131.
- Richards, J., Tawn, J. A. and Brown, S. (2023) Joint estimation of extreme spatially aggregated precipitation at different scales through mixture modelling. *Spatial Statistics* **53**, 100725.
- Ripley, B. D. (1976) The second-order analysis of stationary point processes. *Journal of applied probability* **13**, 255–266.
- Schlather, M. (2002) Models for stationary max-stable random fields. *Extremes* **5**, 33–44.
- Schroeer, K., Kirchengast, G. and O, S. (2018) Strong dependence of extreme convective precipitation intensities on gauge network density. *Geophysical Research Letters* **45**, 8253–8263.
- Tabari, H. (2021) Extreme value analysis dilemma for climate change impact assessment on global flood and extreme precipitation. *Journal of Hydrology* **593**, 125932.
- Vidrio-Sahagún, C. T., He, J. and Pietroniro, A. (2025) Improved correction of extreme precipitation through explicit and continuous nonstationarity treatment and the metastatistical approach. *Water Resources Research* **61**, e2024WR037721.
- Weng, Y.-Y., Zhang, X.-Y., Lu, Z.-H. and Zhao, Y.-G. (2024) A conditional extreme value distribution method for dynamic reliability analysis of stochastic structures. *Structural Safety* **106**, 102398.
- Whitaker, T., Beranger, B. and Sisson, S. A. (2020) Composite likelihood methods for histogram-valued random variables. *Statistics and Computing* **30**, 1459–1477.
- Yates, K. K., Greening, H. and Morrison, G. (eds) (2011) *Integrating Science and Resource Management in Tampa Bay, Florida*. Number 1348 in U.S. Geological Survey Circular. Reston, Virginia: U.S. Geological Survey.
- Zhong, P., Sisson, S. A. and Beranger, B. (2024) Flexible max-stable processes for fast and efficient inference. *arXiv preprint arXiv:2407.13958* .