

DISSERTATION

TAIL DEPENDENCE MODELING FOR HIGH DIMENSIONAL SPATIAL EXTREMES

Submitted by

Muyang Shi

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2026

Doctoral Committee:

Advisor: Benjamin Shaby

Daniel Cooley

Tianjian Zhou

Maria Rugenstein

Copyright by MUYANG SHI 2026

All Rights Reserved

ABSTRACT

TAIL DEPENDENCE MODELING FOR HIGH DIMENSIONAL SPATIAL EXTREMES

This dissertation develops flexible statistical models for tail dependence in high-dimensional spatial extremes, with an emphasis on environmental applications in which rare events can exhibit substantial spatial heterogeneity and different behaviors across spatial lags. Existing spatial extremes models often impose a single extremal dependence class over an entire domain and become computationally prohibitive for high-dimensional threshold exceedance data. To address these limitations, this dissertation contributes methodology for flexible dependence modeling, scalable inference, and extensions toward multivariate tail dependence and causal structure.

To accurately characterize dependence in extreme events, we propose a mixture model that achieves flexible dependence properties and allows high-dimensional inference for extremes of spatial processes. We modify the popular random scale construction that multiplies a Gaussian random field by a single radial variable; we allow the radial variable to vary smoothly across space and add non-stationarity to the Gaussian process. As the level of extremeness increases, this single model exhibits both asymptotic independence at long ranges and either asymptotic dependence or independence at short ranges. We make joint inference on the dependence model and a marginal model using a copula approach within a Bayesian hierarchical model. Three different simulation scenarios show close to nominal frequentist coverage rates. And we apply the model to a dataset of extreme summertime precipitation over the central United States. We find that the joint tail of precipitation exhibits non-stationary dependence structure that cannot be captured by limiting extreme value models or current state-of-the-art sub-asymptotic models.

Flexible spatial extreme models based on Gaussian random scale mixtures provide a framework for capturing a broad range of extremal dependence structures. However, likelihood-based inference under the peaks-over-threshold setting is often computationally infeasible, due to the

censored likelihood requiring repeated evaluation of high-dimensional Gaussian distribution functions. We propose a multiplicative log-Laplace nugget that yields conditional independence in the censored likelihood, resulting in a joint likelihood function that is the product of univariate densities which are available in closed form. This eliminates multivariate Gaussian distribution function evaluations and thereby enables inference for threshold exceedances in high dimensions, which represents a major shift for spatial extremes modeling as the total computational cost is now primarily driven by standard spatial statistics operations. We show that under mild conditions, the spatial process that includes the proposed nugget retains the extremal dependence structure of the underlying smooth process. The approach applies broadly to several modern transformed Gaussian scale mixture models for spatial extremes, and is illustrated through simulation studies and an application to precipitation extremes.

The dissertation concludes with a future research direction on multivariate extremes. This direction aims to develop models that can represent both strong and weak forms of tail dependence among subsets of variables, use a latent indicator matrix interpretable as a directed graph to encode directional tail dependence structure, and apply these methods to critical risk analysis problems involving river discharge and multiple-pollutant health outcomes.

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to my advisors, Ben Shaby and Likun Zhang, and to our collaborator, Mark Risser. Their guidance, patience, and trust carried me through this journey. They challenged me with thoughtful questions, supported me with generosity, and gave me the space to grow as a researcher. This dissertation is, in many ways, a reflection of what I was fortunate to learn from working with them.

I am also sincerely grateful to my dissertation committee, Tianjian Zhou, Daniel Cooley, Maria Rugenstein, and Elizabeth Barnes, for the time and care they gave to this work. Their thoughtful feedback, encouragement, and perspective helped strengthen this dissertation in important ways.

Along the way, I have been fortunate to learn from many wonderful teachers, to share this experience with my fellow PhD students, and to receive kindness and support from the staff in the Department of Statistics at Colorado State University. The conversations, encouragement, and sense of community they gave me meant more than I can fully express.

Finally, I thank my parents for believing in me, especially when I didn't believe in myself.

DEDICATION

*This dissertation is dedicated to my parents,
whose love has been my shelter in doubt,
my courage in hardship,
and the reason I kept going.*

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	iv
DEDICATION	v
LIST OF TABLES	ix
LIST OF FIGURES	x
Chapter 1	Introduction 1
1.1	Background and Motivation 1
1.2	Nonstationary Modeling of Tail Dependence Structure in Spatial Extreme . 2
1.3	Efficient Inference for Flexible Spatial Extreme Models 4
1.4	Future Directions in Flexible Multivariate Extremes Models 5
Chapter 2	Spatial Scale-aware Tail Dependence Modeling for High-Dimensional Spatial Extremes 7
2.1	Introduction 7
2.2	Model 12
2.2.1	Construction 12
2.2.2	Univariate distribution of the dependence model 13
2.2.3	Joint distribution of the dependence model 14
2.2.4	Examples 17
2.3	Bayesian inference 19
2.3.1	Hierarchical model 20
2.3.2	Computation 22
2.3.3	Simulation and Coverage Analysis 23
2.4	Extreme <i>in situ</i> of Daily Precipitation 25
2.4.1	Data Analysis 25
2.4.2	Model Evaluation 29
2.4.3	Results 30
2.5	Discussion 35
Chapter 3	Log-Laplace Nuggets for Fitting Spatial Extremes Models to Threshold Exceedances 37
3.1	Introduction 37
3.1.1	Tail Dependence for Spatial Extremes 38
3.1.2	Flexible Spatial Extreme Models 40
3.1.3	The Censored Likelihood 44
3.2	Model 45
3.2.1	Construction 45
3.2.2	Computational Implications 46
3.2.2.1	Conditional Independence 46
3.2.2.2	Marginal Tractability 46

3.2.3	Tail Properties	49
3.2.3.1	Marginal tail equivalence	49
3.2.3.2	Joint tail dependence	50
3.2.4	Numerical Illustration	52
3.3	Bayesian Inference	54
3.3.1	Hierarchical Model and Computation	54
3.3.2	Simulation and Coverage Analysis	57
3.4	Extreme of <i>in situ</i> Daily Precipitation	57
3.4.1	Data Analysis	57
3.4.2	Model Evaluation	60
3.4.3	Results	62
3.5	Discussion	64
Chapter 4	Research Directions on Flexible Models and Causal Discovery for Multivariate Extremes	66
4.1	Motivation	66
4.2	Model	68
4.2.1	Construction	69
4.2.2	Pairwise Tail Dependence	70
4.2.3	Latent DAG and Causal Discovery	72
4.3	Bayesian Inference	74
4.3.1	Hierarchical Model	74
4.3.2	Simulation Study and Coverage Analysis	76
4.4	Applications	77
4.4.1	Fluvial Flooding in River Networks	77
4.4.2	Multiple Pollutants and Extreme Low Birth Weight	78
4.5	Discussion	79
Chapter 5	Conclusions	81
Appendix A	Supplemental materials for Chapter 2	90
A.1	Technical proofs	90
A.1.1	Properties of stable distribution	90
A.1.2	Proof of Proposition 1	91
A.1.3	Limiting angular measure for linear combinations of independent stable variables	92
A.1.4	Proof of Theorem 3	96
A.2	MCMC details	103
A.3	Additional Results on Simulation Study	104
A.4	Justification for Marginal Model Choices	104
A.5	Confidence Intervals for the Empirical and Model-based $\hat{\chi}$	106
Appendix B	Supplementary materials for Chapter 3	110
B.1	Log-Laplace Nugget Parameterization	110
B.2	Technical proofs	111

B.2.1	Marginal and Joint Tail Equivalence	111
B.2.1.1	Marginal Tail Equivalence	112
B.2.1.2	Joint Tail Equivalence	113
B.2.2	Proof of Theorem 7	118
B.3	MCMC Details	122

LIST OF TABLES

2.1	13 model configurations spanning different knot grids, spatial extent of basis functions, and constraints on the marginal parameters. Model naming convention is as follows: k , r , b , and m respectively denote the number of knots, radius of the compact Wendland kernel, the bandwidth of the Gaussian kernel, and restriction indicator on marginal GEV parameters. Here the effective range refers to the distance at which the Wendland kernel function becomes 0 (as it is compact) or when the Gaussian kernel function drops below 0.05. The H-W Stationary model is the stationary Huser and Wadsworth (2019) process used by Zhang et al. (2022a).	29
2.2	Posterior mean and 95% equi-tail credible interval of the k25r4b4 model. k indicates knot index.	32
3.1	Summary of marginal and bivariate tail behavior of (s_i, s_j) for random scale-mixture models (M1)–(M4). Let $W_i := g\{Z_i\}$, $\phi_i \in (0, \infty)$ be the dependence parameter, and $\eta_{ij}^Z \in (1/2, 1)$ be the residual tail dependence coefficient induced by the latent $\{Z(s)\}$ process. For (M3), $\gamma > 0$ is the Pareto tail index for $R(s)$. For (M4), α is the Stable-index. For (M3) and (M4), $\mathcal{K}_i := \{k : B_k(s_i; l) > 0\}$, the set of knots that are active for location s_i . See Section 3.1.2 for model construction details.	43
3.2	Candidate model configurations considered for the precipitation analysis. Model naming convention is as follows: k , b , and m respectively denote the number of knots, effective range of the Gaussian basis (the distance at which the kernel function drops below 0.05), and restriction indicator on marginal GPD parameters.	60

LIST OF FIGURES

1.1	Empirical estimates of $\hat{\chi}_h(u)$, “co-occurrence of extremes”, from 75 years of summertime maximum daily precipitation of central US within moving windows centered at 119 locations (marked as white ‘+’), using spatial lag distance $\approx 75\text{km}$, at three quantile levels $u = 0.9, 0.95, 0.99$. The local tail dependence behavior appears to vary across the region.	2
1.2	Two extremes frameworks. Left: Block maxima, where data are divided into blocks and only the maximum observation in each block is retained. Right: Peaks-over-threshold (POT), where all observations above a high threshold are retained as exceedances.	4
2.1	Empirical estimates of $\hat{\chi}_h(u)$ from 75 years of summertime maximum daily precipitation of central US within moving windows centered at 119 locations (marked as white ‘+’), using all pairs with separating distance $\approx 75\text{km}$, at three quantile levels $u = 0.9, 0.95, 0.99$. The local tail dependence behavior appears to vary across the region. Further exploration is shown in Figure 2.11 in Section 2.4.	9
2.2	A $\phi(s)$ surface on $[0, 10]^2$, in which the dashed line marks the transition between local AI and AD. The points with ‘+’ are centers for the Stable variables and the compact Wendland basis functions. The points with other signs/marker-styles are randomly chosen sample points that we use to verify the dependence properties in Theorem 3.	19
2.3	Empirical estimates of the dependence coefficients $\chi_{ij}(u)$ and $\eta_{ij}(u)$ between sample points i and j in Figure 2.2. The red lines mark the theoretical limit and the orange dashed and blue dotted lines respectively represent theoretical upper and lower bounds. The $\chi_{ij}(u)$ ’s are direct empirical estimates; due to numerical instability, particularly as the probability u approaches its upper limit of 1, we estimate the $\eta_{ij}(u)$ ’s using the Hill estimator with marginal transformation to standard Exponential (Leo R Belzile, 2023).	20
2.4	Dependence model parameter surfaces, $\phi(s)$ and $\rho(s)$, for the three simulation scenarios, which correspond to different levels of non-stationarity. The dashed line marks the transition between local AI and AD. The ‘+’ denotes the center of the kernel functions, and the gray circles denotes the radii.	24
2.5	Empirical coverage rates of credible intervals of the marginal parameters μ and σ (left), the dependence parameters ϕ_k (middle), and ρ_k (right), $k = 1, \dots, 9$, in simulation scenario 1.	24
2.6	(a) Location of the 590 observations/data sites and the 99 out-of-sample testing sites. (b) An illustration of the mixture component setup for model k25r4b4. The blue dots represent observation sites, red ‘+’ symbols represent the centers of the kernel functions, and the shaded circles represent areas covered by each Wendland kernel with specified radius r	26
2.7	Across 590 locations, histogram (left) of the unique dates of the yearly maxima across all sites; line charts (right) of the unique dates of the upper percentile maxima.	28

2.8	Boxplots of predictive log-likelihood for the thirteen models. Higher log-likelihood is better. “Blue” models estimate the GEV parameters in the MCMC process, while the “orange models” have some restrictions on their marginal model parameters; “blue models” perform better than their “orange counterparts”.	31
2.9	QQ-plots of four randomly selected holdout locations, comparing observed and predicted marginal quantiles for the k25r4b4 model. 95% confidence envelopes are also shown. Marginal values are transformed to a Gumbel distribution.	32
2.10	Posterior-mean surfaces for the dependence and marginal models of k25r4b4.	33
2.11	Moving-window estimates of $\chi_u(h)$ across three quantiles u and three spatial lags h . The left-hand panel shows the dataset empirical estimates of $\chi_u(h)$, and the right-hand panel shows the model-based estimates of $\chi_u(h)$, based on the k25r4b4 model.	34
3.1	A $\phi(\mathbf{s})$ surface on $[0, 10]^2$, in which the dashed line marks the transition between local AI and AD. The points with ‘•’ are centers for the Wendland basis functions. The points with other signs/marker-styles are chosen sample points that we use to illustrate the dependence properties in Theorem 9.	53
3.2	Empirical estimates of the dependence coefficients $\chi_{ij}(u)$ and $\eta_{ij}(u)$ (Leo R Belzile, 2023) between sample points i and j in Figure 3.1. The black solid, dotted, and dashed lines mark the theoretical limit, upper, and lower bounds, respectively.	55
3.3	Simulation surfaces of $\phi(\mathbf{s})$ and $\rho(\mathbf{s})$. The black dashed line marks the local AD/AI transition. The • symbols denote knot locations, and circles indicate kernel radii.	57
3.4	Posterior means (top) and standard binomial empirical coverage confidence intervals (bottom) for dependence (ϕ, ρ), marginal (σ, ξ), and kernel radius (l) parameters across 50 simulated datasets. Red dashed lines in boxplots mark the true values.	58
3.5	Location of the 590 stations (blue circle) and the 99 out-of-sample testing stations (orange triangle) in the central US used for analysis.	59
3.6	Point estimates and 95% confidence intervals for 50-year precipitation return levels at 12 randomly selected stations, estimated from annual maxima using 50-year sliding windows. No evident systematic trend is observed, supporting the assumption of temporally constant marginal parameters.	61
3.7	Predictive log scores at holdout sites for the four candidate models.	61
3.8	QQ-plots of four randomly selected holdout locations, comparing observed and predicted marginal quantiles for the k41b4 model, along with 95% confidence envelopes. All values are transformed to the Gumbel scale for ease of visualization.	62
3.9	Left: For the k41b4 model, the black dots represent the knots and circles represent areas covered by each kernel with posterior mean radius. Right: posterior mean parameter surfaces for the fitted k41b4 model.	63
3.10	Moving-window estimates of $\chi_u(h)$ across three quantiles u and three spatial lags h . The left-hand panel shows the dataset empirical estimates of $\chi_u(h)$, and the right-hand panel shows the model-based estimates of $\chi_u(h)$, based on the k41b4 model.	64
A.1	Empirical coverage rates of credible intervals of the marginal parameters μ and σ (left), the dependence parameters ϕ_k , $k = 1, \dots, 9$ (middle), and ρ_k , $k = 1, \dots, 9$ (right), in simulation scenario 2.	104

A.2	Empirical coverage rates of credible intervals of the marginal parameters μ and σ (left), the dependence parameters ϕ_k , $k = 1, \dots, 9$ (middle), and ρ_k , $k = 1, \dots, 9$ (right), in simulation scenario 3.	104
A.3	Estimates and smoothed surfaces of maximum likelihood estimates of $\log(\sigma)$ and ξ , fit to annual maxima at each station independently. The smoothed estimates of $\log(\sigma)$ show similar spatial patterns as the negative of scaled elevation at each station (right-most panel).	106
A.4	CO ₂ concentrations (and its functions) against time in black curve. Least square fit in Blue.	106
A.5	Empirical (left) and model-realized (right) χ surfaces with 95% confidence bands. . . .	108
A.6	Spatial maps displaying the difference between the empirical and modeled quantities using a diverging color scheme centered at zero. The differences are largely diminished at the most extreme quantile level $u = 0.99$	109

Chapter 1

Introduction

1.1 Background and Motivation

Accurately characterizing the spatial extent and intensity of large-scale extreme events is essential for infrastructure design, flood risk assessment, and climate adaptation (Field et al., 2014). Extreme precipitation provides a clear illustration. Intense storms can produce fluvial flooding when rivers overtop their banks and pluvial flooding when heavy rainfall overwhelms urban or overland drainage. In both cases, risk depends not only on how severely it rains at a point, but also on the spatial footprint and organization of the storm. Spatial extremes modeling is therefore necessary for estimating regional risk and designing resilient systems.

These considerations are especially important when estimating design quantities such as 100-year precipitation events over a spatial domain. Such events often exhibit spatially heterogeneous tail dependence, meaning that the tendency of extremes to co-occur varies across locations and distances; see Figure 1.1. In applications such as extreme precipitation, risk depends not only on the magnitude of an event at a single site, but also on the extent to which extremes co-occur across space. Yet many existing spatial extremes models are poorly suited to model such environmental processes.

A central challenge is that extremal dependence can be heterogeneous. The strength, range, and even asymptotic class of dependence can vary across location and spatial lag, yet many existing spatial extremes models impose a single dependence structure over an entire domain. At the same time, models that are flexible enough to reflect this heterogeneity often become computationally prohibitive when applied to high-dimensional threshold exceedance data. These twin limitations, lack of flexibility and lack of scalability, motivate this dissertation.

This dissertation develops flexible and scalable statistical methodologies for tail dependence in high dimensional extremes, motivated by environmental problems of rare but impactful extreme

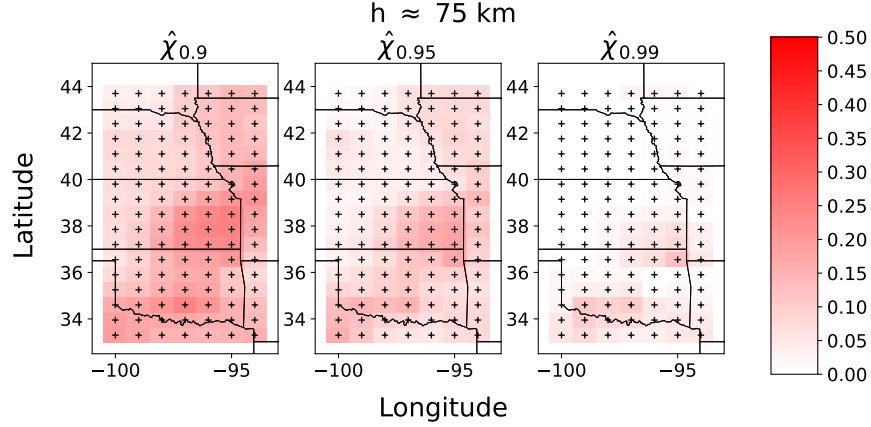


Figure 1.1: Empirical estimates of $\hat{\chi}_h(u)$, “co-occurrence of extremes”, from 75 years of summertime maximum daily precipitation of central US within moving windows centered at 119 locations (marked as white ‘+’), using spatial lag distance $\approx 75\text{km}$, at three quantile levels $u = 0.9, 0.95, 0.99$. The local tail dependence behavior appears to vary across the region.

events. A third project plans to extend this line of work to multivariate extremes, where the goal is to model heterogeneous tail dependence across variables and to introduce a formal notion of directional tail dependence, or tail causality. The chapters that follow address these challenges. Chapter 2 develops a flexible spatial model for heterogeneous tail dependence, Chapter 3 develops scalable inference for threshold exceedance data under that framework, and Chapter 4 outlines an extension to multivariate extremes and tail causality.

1.2 Nonstationary Modeling of Tail Dependence Structure in Spatial Extreme

The coefficient $\chi(u)$ summarizes how often extremes co-occur across sites at a high quantile level u .

For $u \in (0, 1)$, let $\chi(u) := \Pr(F_{Y_1}(Y_1) > u \mid F_{Y_2}(Y_2) > u)$, $\chi = \lim_{u \rightarrow 1} \chi(u)$;

then (Y_1, Y_2) is asymptotically dependent if $\chi > 0$ and is asymptotically independent

if $\chi = 0$.

Large values of $\chi(u)$ correspond to more spatially widespread joint extremes, while $\chi(u) \rightarrow 0$ indicates increasingly localized behavior in the tail. The distinction between asymptotic dependence

and asymptotic independence is fundamental in extreme value analysis because it determines how joint risk is extrapolated beyond the observed range of the data.

Misidentifying the extremal dependence class can materially distort joint tail risk estimates. In particular, using an asymptotically independent model when the data are truly asymptotically dependent can lead to severe underestimation of joint extremes, with direct consequences for risk assessment and infrastructure design. This makes flexible modeling of tail dependence especially important in environmental applications, where the most consequential events are precisely those that require extrapolation beyond the available sample.

Random scale mixture models provide one route to avoiding an a priori commitment to either dependence class. In these models, a tail index parameter governs whether the process is asymptotically dependent or asymptotically independent (Huser and Wadsworth, 2019). However, standard versions of these models still impose a global structure: for a fixed parameter value, the entire spatial domain and all spatial lags must share the same extremal regime. This is often unrealistic for large, heterogeneous domains, where short-range dependence may differ substantially from long-range dependence and where the dependence structure may vary regionally.

Recent work has introduced nonstationarity into spatial tail dependence models and has made it possible to combine short-range asymptotic dependence with long-range asymptotic independence (Castro-Camilo and Huser, 2020; Hazra et al., 2025; Majumder et al., 2024; Wadsworth and Tawn, 2022). Nonetheless, existing approaches retain important limitations. Some do not provide a fully specified joint probability model or likelihood, some offer limited analytical characterization, and some do not naturally support spatial prediction or formal uncertainty quantification. Others require boundary or degenerate parameter settings to obtain the desired dependence behavior.

The first project addresses these limitations by developing a spatial scale-aware random scale mixture model in which both the radial component and the tail dependence parameter vary over space, while the latent Gaussian process is endowed with a nonstationary kernel. This construction allows the model to exhibit long-range asymptotic independence together with either asymptotic dependence or asymptotic independence at short range within a single coherent framework. The

transition between local dependence classes occurs smoothly in the interior of the parameter space, which makes the model both interpretable and estimable.

The model is fit through a Bayesian hierarchical copula formulation that jointly estimates the marginal and dependence components, enabling spatial prediction and coherent uncertainty propagation. Simulation studies show that the framework recovers the targeted dependence behavior and yields well-calibrated inference. Applied to summertime precipitation in the central United States, the model reveals substantial spatial heterogeneity in tail dependence, including regions with qualitatively different local extremal regimes. This work forms the basis of Chapter 2.

1.3 Efficient Inference for Flexible Spatial Extreme Models

Extreme value analysis commonly proceeds through one of two frameworks: block maxima or peaks-over-threshold (POT). The block maxima approach retains only the maximum observation in each block, such as a year or season, whereas POT retains all observations above a high threshold. POT is often preferred in practice because it uses more of the available information and preserves clusters of extreme events that may occur on their natural time scale; see Figure 1.2.

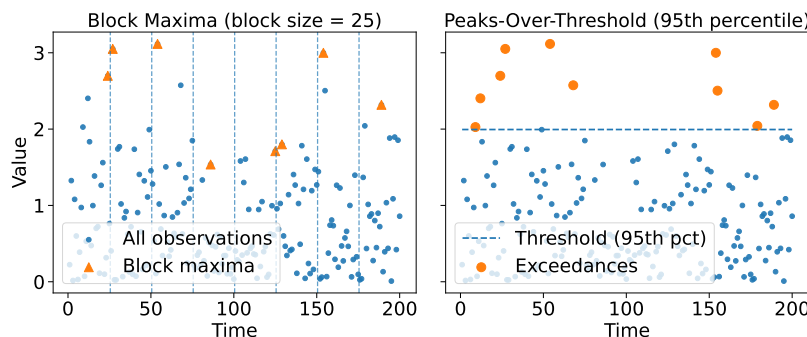


Figure 1.2: Two extremes frameworks. Left: Block maxima, where data are divided into blocks and only the maximum observation in each block is retained. Right: Peaks-over-threshold (POT), where all observations above a high threshold are retained as exceedances.

In multivariate and spatial POT models, censored likelihood is a standard inferential tool. Exceedances contribute density terms, while sub-threshold observations contribute through the distribution function, allowing estimation to focus on the tail while still using all observations. For

flexible spatial dependence models, however, this inferential advantage comes with a substantial computational cost. Repeated likelihood evaluation requires high-dimensional Gaussian distribution functions, and these quickly become infeasible as the number of sites grows.

This bottleneck is especially severe in the high-dimensional spatial setting. While the block maxima version of the model can be fit with parallelized MCMC for datasets with hundreds of locations, the POT setting introduces many more temporal replicates and therefore many more censored likelihood evaluations. As a result, direct likelihood-based inference becomes computationally prohibitive even when core calculations are implemented in C++.

Earlier work addressed this issue in the stationary setting by introducing an additive Gaussian nugget that induces conditional independence and thereby bypasses high-dimensional Gaussian CDF evaluations (Zhang et al., 2022b). However, that strategy also induces a convolution in the marginal distribution that is not available in closed form and must be evaluated numerically. In the more flexible nonstationary model considered here, this numerical integration becomes a dominant computational bottleneck.

The second project addresses this problem by introducing a multiplicative log-Laplace nugget for Gaussian random scale mixture models under a peaks-over-threshold framework. Under mild conditions, the proposed construction preserves the marginal tail behavior and extremal dependence structure of the underlying smooth process while inducing conditional independence in the censored likelihood. This removes the need for repeated evaluation of high-dimensional Gaussian distribution functions and substantially improves computational scalability. In preliminary implementation, the modified model yields roughly 150-fold higher per-iteration throughput than the original formulation. This work forms the basis of Chapter 3.

1.4 Future Directions in Flexible Multivariate Extremes Models

A natural next step is to extend these ideas and modeling frameworks to multivariate extremes, where the scientific question concerns how rare events co-occur across distinct variables

and whether these tail relationships exhibit directional structure. In many applications, joint tail risk is not symmetric. For example, extreme discharge at an upstream river gauge can increase downstream flood risk more directly than the reverse, and extreme exposure to one pollutant may be associated with adverse health outcomes in ways that differ from another pollutant. Existing multivariate extremes models are typically too rigid to represent pair-specific tail regimes within a single framework, and current approaches to tail causality remain limited, especially when some variable pairs may be asymptotically independent while others are asymptotically dependent.

This future project proposes a multivariate tail dependence model based on transformed Gaussian scale mixtures with component-specific random scaling factors. The key structural ingredient is a latent indicator matrix that can be interpreted as a directed graph among variables. By allowing the scaling structure to depend on this graph, the model is designed to accommodate both strong and weak forms of pairwise tail dependence within a single coherent framework, rather than forcing all variable pairs into a common extremal regime. The same latent graph also provides a route to defining directional features of tail dependence that are interpretable as tail causal relationships.

The motivating applications are problems in critical risk analysis where directional tail behavior is scientifically meaningful. One application concerns river discharge at multiple gauges in a river network, where flow direction induces asymmetric extremal dependence and where identifying how flood risk propagates through the network is of direct practical importance. A second application concerns health outcomes under exposure to multiple pollutants, with particular emphasis on extreme low birth weight rather than average effects. Together, these applications motivate a multivariate framework that combines flexible tail dependence, interpretable directional structure, and Bayesian uncertainty assessment. This future direction is outlined in Chapter 4.

Chapter 2

Spatial Scale-aware Tail Dependence Modeling for High-Dimensional Spatial Extremes

2.1 Introduction

Accurately characterizing the spatial extent and intensity of large-scale extreme precipitation events is a critical factor in infrastructure planning, risk mitigation, and adaptation (Field et al., 2014), particularly in our changing global climate (Milly et al., 2008). In this work, we propose a flexible tail dependence model that modulates a radial variable to introduce varying levels of dependence strength both locally and across long spatial ranges, and incorporates a nonstationary covariance function to further account for spatial heterogeneity. Our sub-asymptotic framework summarizes the tail dependence structure in a parsimonious manner and allows for efficient inference from large spatial data sets. Therefore, we can appropriately model the varying scale and intensity of extreme precipitation events.

Let $\{Y(\mathbf{s}) : \mathbf{s} \in \mathcal{S} \subseteq \mathbb{R}^2\}$ be the stochastic process of interest, and denote its realization at the j th location $\mathbf{s}_j$ by $Y_j = Y(\mathbf{s}_j)$, $j = 1, \dots, D$. In a spatial extremes context, a common approach is to model the marginal distributions of each Y_j (denoted by F_j) separately from their dependence structure via the copula. After transforming the data to the uniform scale via the probability integral transform, a carefully chosen spatial dependence model can be used to characterize the tail dependence properties of the copula while accounting for spatial non-stationarity.

Unfortunately, previous spatial extremes models have dependence structures that are too rigid for large domains and thus lead to underestimation or overestimation of the spatial extent and intensity of extreme events (Huser and Genton, 2016). Statistical modeling for spatial extremes traditionally uses either max-stable processes (when considering block maxima) or generalized Pareto processes (for exceedances of a high threshold) due to their asymptotic justifications (Davi-

son et al., 2013; Huser and Wadsworth, 2022). The copulas of both processes exhibit stability properties, wherein their dependence structures are invariant to the operations of taking maxima over larger blocks or conditioning on a higher threshold. In the bi-variate case, this means that the upper dependence measure of the process between any two locations s_i and s_j

$$\chi_{ij}(u) = P(F_j(Y_j) > u \mid F_i(Y_i) > u) \quad (2.1)$$

will always have a non-zero limit as the quantile level $u \rightarrow 1$, a property generally referred to as asymptotic dependence (AD). Max-stable and generalized Pareto models also have the property that $\chi_{ij}(u)$ becomes independent of u for increasingly extreme events. However, for sub-asymptotic modeling, both of these properties are often violated empirically: estimates of (2.1) tend to decrease with u for many observed environmental processes, including high-frequency wave height data (Huser and Wadsworth, 2019), daily fire weather indices (Zhang et al., 2022b), annual maximum temperature (Zhang et al., 2024), and extreme seasonal precipitation (see Figure 2.1). Furthermore, weakening $\chi_{ij}(u)$ in the empirical estimates as $u \rightarrow 1$ may eventually result in asymptotic independence (AI), i.e., $\lim_{u \rightarrow 1} \chi_{ij}(u) = 0$. Of course, empirically weakening $\chi_{ij}(u)$ as u approaches its observed upper bound does not necessarily correspond to asymptotic independence since estimates of $\chi_{ij}(u)$ for large u have very large uncertainty and may not suggest convergence in the limit.

To avoid having to specify the dependence class before analyzing a specific data set, a series of flexible random scale mixture models have been developed to encompass both AI and AD dependence classes (e.g., Huser et al., 2017; Huser and Wadsworth, 2019). The general form of the spatial dependence model in these approaches is

$$X(s) = R \cdot g(Z(s)), \quad R \mid \theta_R \sim F_R, \quad (2.2)$$

in which R is a random scaling factor whose distribution function F_R is parameterized by the vector θ_R containing a parameter ϕ which controls its tail heaviness, $g(\cdot)$ is a suitable link function,

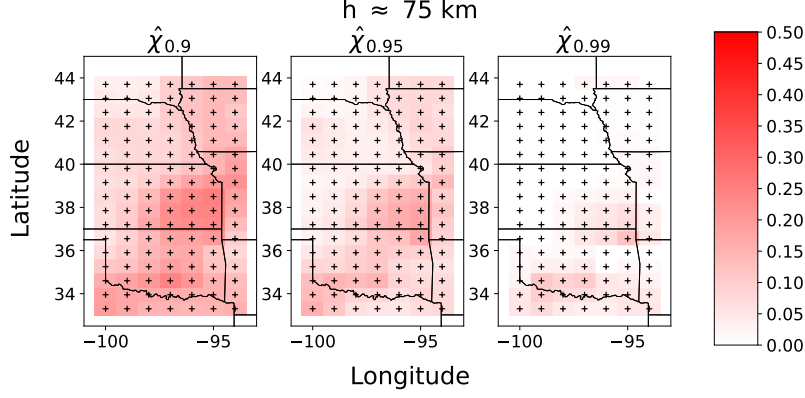


Figure 2.1: Empirical estimates of $\hat{\chi}_h(u)$ from 75 years of summertime maximum daily precipitation of central US within moving windows centered at 119 locations (marked as white ‘+’), using all pairs with separating distance ≈ 75 km, at three quantile levels $u = 0.9, 0.95, 0.99$. The local tail dependence behavior appears to vary across the region. Further exploration is shown in Figure 2.11 in Section 2.4.

and $\{Z(s)\}$ is a stationary spatial process exhibiting asymptotic independence at any two locations. Intuitively, the scale factor R can be thought of as a completely dependent spatial process in that it takes on a single value regardless of spatial location. It amplifies the co-occurrences of extreme events in $\{Z(s)\}$, and the tail behavior of $X(s)$ is controlled by the relative rate of tail decay between R and $\{g(Z(s))\}$ (Engelke et al., 2019). Consequently, varying a single tail index parameter ϕ allows random scale mixture models to be either asymptotically independent or asymptotically dependent, with the transition occurring smoothly in the interior of the parameter space. In spite of this desirable property, inference for random scale mixture models is challenging. To mitigate computational burden while preserving tail dependence properties, Zhang et al. (2022b) proposed the addition of a nugget effect which facilitates likelihood calculations for large data sets. Nevertheless, the random scale mixture models in (2.2) may lack sufficient flexibility because, for a given value of ϕ , they can only exhibit one dependence class, at all spatial lags, for the entire spatial domain. Coupled with the stationarity of the latent process $\{Z(s)\}$, this means that the dependence measure for any two locations is also stationary (i.e., $\chi_{ij}(u) \equiv \chi_{\|s_i - s_j\|}(u)$). Such assumptions of stationarity in both the dependence class of $X(\cdot)$ and the length scales of $Z(\cdot)$ are convenient but are rarely appropriate for heterogeneous real-world data sets. The assumption that the tail dependence is of the same asymptotic class at all spatial lags seems unintuitive for large

domains, where a process may have very strong tail dependence at short distances but weak tail dependence at long distances. We therefore seek tail dependence models that are “scale aware” in that they are AI for observations separated by large distances and either AI or AD for observations separated by short distances.

To introduce non-stationarity in the tail dependence structure, Castro-Camilo and Huser (2020) let the dependence parameter ϕ be a spatial process $\phi(\mathbf{s})$ that alters the amplifying effect of R over space. Their specification is in the form of random location mixture, which is equivalent to a random scale mixture through an exponential marginal transformation:

$$X(\mathbf{s}) = \phi^{-1}(\mathbf{s})R + Z(\mathbf{s}), \quad \mathbf{s} \in \mathcal{S}. \quad (2.3)$$

They furthermore specify $\{Z(\mathbf{s})\}$ to be a zero-mean Gaussian process with a nonstationary Matérn covariance function (Paciorek and Schervish, 2006), resulting in a non-stationary dependence measure χ_{ij} for locations $\mathbf{s}_i$ and $\mathbf{s}_j$. While this model yields much more flexible local dependence properties, χ_{ij} is always strictly positive unless $\max\{\phi_i, \phi_j\} = \infty$; i.e., this model still exhibits asymptotic dependence everywhere, even for locations that are very far apart.

To address this limitation, Hazra et al. (2025) generalizes the Huser et al. (2017) model by instead allowing R to be a spatial process and ϕ a spatial constant:

$$X(\mathbf{s}) = R(\mathbf{s})Z(\mathbf{s}), \quad \mathbf{s} \in \mathcal{S}, \quad (2.4)$$

where $Z(\mathbf{s})$ is a zero-mean isotropic Gaussian process, and $R(\mathbf{s})$ is a kernel-weighted sum of univariate random effects. The kernels live in same spatial domain as the data and are compact, resulting in AI of the $X(\mathbf{s})$ process for locations sufficiently far apart (see our Theorem 3 for our analogous result). Their random effects distribution contains a tail parameter $\phi \geq 0$ such that $\phi = 0$ gives AD for locations close enough to share a common kernel. The model in Hazra et al. (2025) represents a major step forward in the flexibility of spatial extremes methods in that it combines local AD with long-range AI. However, we improve upon their approach in several ways. First,

we use a different random effects distribution, which allows the transition between local AI and AD to occur in the interior of its parameter space. In contrast, whereas we are able to estimate the local tail dependence class, Hazra et al. (2025) only consider $\phi = 0$ in their analysis, which fixes the dependence class to be AD at short ranges. For highly heterogeneous weather processes like extreme precipitation, it is often far from clear that local asymptotic dependence is always realistic, as seen in Figure 2.1. Second, our random effects specification also permits a richer set of analytical results than is contained in Hazra et al. (2025), including results on sub-asymptotic tail dependence in the local AI case. Furthermore, we allow the tail parameter ϕ to vary in space, similarly to Castro-Camilo and Huser (2020). This can be important, as empirical estimates of the dependence class for our precipitation data suggests that it may exhibit both local AI and local AD in different parts of the domain; see again Figure 2.1.

Finally, the way Hazra et al. (2025) handle marginal fitting is computationally convenient but induces a few modeling disadvantages which we avoid. Specifically, they perform location-specific standardization of the data before model fitting, then assume that the standardized observations follow a location-scale transformation of their dependence model $X(\mathbf{s})$. Such a scheme does not allow for spatial prediction, since the site-specific standardization leaves no way to “de-standardize” any model-predicted $X(\mathbf{s})$ at un-observed locations. More worryingly, assuming the data follow a location-scale transformation of $X(\mathbf{s})$ forces the marginal tail weight of the observations to coincide with the marginal weight behavior of the dependence model $X(\mathbf{s})$. This inextricably couples the marginal and dependence properties, a situation that we would like to avoid. In contrast, we perform full joint inference on the marginal and dependence models using a copula approach. While this requires a good deal more implementation effort, it allows natural spatial prediction and prevents characteristics of the marginal tail weight from bleeding into the fit of the dependence parameters.

Majumder et al. (2024) also use a construction like (2.4), but $R(\mathbf{s})$ is specified as a max-stable process. This yields short-range AD and long-range AI, but their model presents challenges. First, analytical results are elusive, so tail dependence properties must be confirmed through simulation.

More importantly, likelihoods are unavailable, so Majumder et al. (2024) use a Vecchia-type approximation (Vecchia, 1988) along with neural net emulators to perform approximate Bayesian inference.

Wadsworth and Tawn (2022) use a completely different approach to obtain flexible models capable of short-range AD and long-range AI. They specify their model conditionally on observing a large value at a single reference location. Since there is usually no natural location to choose as the reference, Wadsworth and Tawn (2022) form composite likelihoods by choosing several sites in turn, and summing the log likelihoods obtained from each. This has the effect of fitting a model that requires conditioning on a single site, but uses the composite likelihood to do a kind of averaging across different choices of conditioning sites when fitting model parameters. This approach is an important advancement that generates very flexible tail dependence characteristics. However, because the resultant fits do not correspond to any well-defined joint probability model, interpretation is challenging.

2.2 Model

2.2.1 Construction

Here, we combine the desirable features of (2.3) and (2.4) by allowing both $R(\mathbf{s})$ and $\phi(\mathbf{s})$ to vary spatially by using a mixture component representation for $R(\mathbf{s})$ and a spatially-varying tail index $\phi(\mathbf{s})$ in order to introduce more flexible local and long-range tail dependence behaviors:

$$X(\mathbf{s}) = R(\mathbf{s})^{\phi(\mathbf{s})} g(Z(\mathbf{s})), \quad (2.5)$$

where $\{Z(\mathbf{s})\}$ is a spatial process with hidden regular variation (Ledford and Tawn, 1996; Hefferman and Resnick, 2005). The hidden regular variation assumption ensures

$$P(F_i^Z(Z_i) > u, F_j^Z(Z_j) > u) = L_Z\{(1-u)^{-1}\}(1-u)^{1/\eta_{ij}^Z}, \quad (2.6)$$

where $\eta_{ij}^Z \in (1/2, 1)$ so that (Z_i, Z_j) exhibits asymptotic independence, F_j^Z is the univariate distribution function of Z_j , $j = 1, \dots, D$, and L_Z is a slowly varying function, i.e., $L_Z(tz) / L_Z(z) \rightarrow 1$ as $z \rightarrow \infty$ for all fixed $t > 0$. The coefficient η_{ij}^Z was termed the *coefficient of tail dependence* for $\{Z(s)\}$ between s_i and s_j by Ledford and Tawn (1996), and it complements (2.1) in the case of asymptotic independence.

In (2.5), both the scaling factor and the tail index vary across space. This makes the model much more flexible but also adds considerable complexity to both the theoretical analysis and the computations. The theoretical tools used to analyze models where either the scaling factor (Hazra et al., 2025) or the tail index (Castro-Camilo and Huser, 2020), but not both, vary spatially, will not suffice here. However, if the random factors are weighted averages of Stable random variables (Nolan, 2020), the analysis is simplified considerably. Let

$$R(s) = \sum_{k=1}^K w_k(s, r_k) S_k, \quad S_k \stackrel{\text{indep}}{\sim} \text{Stable}(\alpha, 1, \gamma_k, \delta), \quad (2.7)$$

in which $w_k(s, r_k)$ is some compactly supported basis function over $\mathbb{R}^2$ centered at the k th knot with radius r_k , $k = 1, \dots, K$. Also, $\sum_{k=1}^K w_k(s; r_k) = 1$ for all $s \in \mathcal{S}$. Properties of the univariate and joint distribution of the constructed $\{X(s)\}$ and interpretation of its parameters are discussed in Section 2.2.2 and 2.2.3. The link function $g(\cdot)$ transforms the margins of $\{Z(s)\}$ to the margins of a type II Pareto distribution whose distribution function is

$$F^{\text{Pareto}}(x) = \{1 - (1 + x - \delta)^{-1}\} \mathbb{1}(x \geq \delta). \quad (2.8)$$

We provide justifications for these distributional choices in Section 2.3.2.

2.2.2 Univariate distribution of the dependence model

The key property of the Stable distribution which allows us to assess the tail behavior of the dependence model $X(s)$ in (2.5) is that it is closed under convolution; if $S_k \stackrel{\text{indep}}{\sim} \text{Stable}(\alpha, \beta_k, \gamma_k,$

δ_k) and constant $w_k \geq 0$ for $k = 1, \dots, K$, then

$$\sum_{k=1}^K w_k S_k \sim \text{Stable}(\alpha, \bar{\beta}, \bar{\gamma}, \bar{\delta}) \quad (2.9)$$

with $\bar{\gamma} = \{\sum_{k=1}^K (w_k \gamma_k)^\alpha\}^{1/\alpha}$, $\bar{\beta} = \sum_{k=1}^K \beta_k (w_k \gamma_k)^\alpha / \bar{\gamma}^\alpha$ and $\bar{\delta} = \sum_{k=1}^K w_k \delta_k$. To be able to examine the joint distribution of (X_i, X_j) for model (2.5), it is desirable for the mixture to have the same distributional support and rate of tail decay as each S_k . Thus in (2.7), we fixed $\beta_k \equiv 1$, $\delta_k \equiv \delta$ while imposing the constraint $\sum_{k=1}^K w_k = 1$. As a result, $\bar{\beta} = 1$ and $\bar{\delta} = \delta$, which means the univariate support of the $\{R(\mathbf{s})\}$ is $[\delta, \infty)$ everywhere. We also have tail-stationarity for $\{R(\mathbf{s})\}$ because $P(R(\mathbf{s}) > x) \sim 2\bar{\gamma}^\alpha(\mathbf{s})C_\alpha x^{-\alpha}$ for all $\mathbf{s} \in \mathcal{S}$ as $x \rightarrow \infty$, where $\bar{\gamma}(\mathbf{s}) = [\sum_{k=1}^K \{w_k(\mathbf{s}, r_k)\gamma_k\}^\alpha]^{1/\alpha}$ and $C_\alpha = \Gamma(\alpha) \sin(\alpha\pi/2)/\pi$.

With the understanding of the distributions of $\{R(\mathbf{s})\}$, the tail behavior of the univariate distributions of the dependence model (2.5) can be assessed. To avoid clutter, we denote $X_j := X(\mathbf{s}_j)$, $R_j := R(\mathbf{s}_j)$, $\phi_j := \phi(\mathbf{s}_j)$, $\bar{\gamma}_j := \bar{\gamma}(\mathbf{s}_j)$, and $W_j := g(Z(\mathbf{s}_j))$, $j = 1, \dots, D$. Proposition 1 below shows that the marginal tail behavior of the process $X(\mathbf{s}_j)$ depends on the spatially varying tail parameter ϕ_j ; the (transformed) Gaussian part dominates when ϕ_j is small, while the heavy-tailed R_j dominates when ϕ_j is large, and $\phi_j = \alpha$ marks the transition point between the two.

Proposition 1. *The univariate distribution of the process (2.5) at a location $\mathbf{s}_j$ satisfies*

$$1 - F_j(x) = \Pr\{R_j^{\phi_j} g(Z_j) > x\} \sim \begin{cases} \mathbb{E}(R_j^{\phi_j})x^{-1}, & \text{if } 0 \leq \phi_j < \alpha, \\ \frac{2C_\alpha \bar{\gamma}_j^\alpha}{1 - \alpha/\phi_j} x^{-\frac{\alpha}{\phi_j}}, & \text{if } \phi_j > \alpha, \\ 2C_\alpha \bar{\gamma}_j^\alpha x^{-1} \log x, & \text{if } \phi_j = \alpha, \end{cases} \quad (2.10)$$

as $x \rightarrow \infty$, where $C_\alpha = \Gamma(\alpha) \sin(\alpha\pi/2)/\pi$.

2.2.3 Joint distribution of the dependence model

We start by considering the joint behavior of two scaling variables, R_i and R_j (defined as kernel-weighted linear combinations of iid Stable random variables in (2.7)) at two locations $\mathbf{s}_i$ and $\mathbf{s}_j$. For notational simplicity, we denote $w_{kj} = w_k(\mathbf{s}_j, r_k)$, $k = 1, \dots, K$, and write $\mathbf{w}_j =$

$(w_{1j}, \dots, w_{Kj})$ and $\mathcal{C}_j = \{k : w_{kj} \neq 0, k = 1, \dots, K\}$, $j = 1, \dots, D$. We require that any location $\mathbf{s} \in \mathcal{S}$ is covered by at least one basis function, thus $\mathcal{C}_j$ cannot be empty for any j . The following proposition describes the dependence between two scaling variables R_i and R_j :

Proposition 2. (a) If $\mathcal{C}_i \cap \mathcal{C}_j = \emptyset$, R_i and R_j are independent and $\Pr(R_i > x, R_j > x) \sim 4C_\alpha^2 \{\bar{\gamma}_i^\alpha\} \{\bar{\gamma}_j^\alpha\} x^{-2\alpha}$ as $x \rightarrow \infty$. If $\mathcal{C}_i \cap \mathcal{C}_j \neq \emptyset$, then

$$\Pr(R_i > x, R_j > x) = 2C_\alpha C_K(\mathbf{w}_i, \mathbf{w}_j, \boldsymbol{\gamma}) x^{-\alpha}, \quad (2.11)$$

where $C_K(\mathbf{w}_i, \mathbf{w}_j, \boldsymbol{\gamma}) = \sum_{k=1}^K w_{k,\wedge}^\alpha \gamma_k^\alpha$ with $w_{k,\wedge} = \min(w_{ki}, w_{kj})$, $k = 1, \dots, K$.

(b) If $\mathcal{C}_i \cap \mathcal{C}_j \neq \emptyset$, we have for two positive constants $\phi_i < \phi_j$ and any $c_i, c_j > 0$

$$\Pr(\min(c_i R_i^{\phi_i}, c_j R_j^{\phi_j}) > x) \sim d_\wedge x^{-\alpha/\phi_i}, \quad \Pr(\max(c_i R_i^{\phi_i}, c_j R_j^{\phi_j}) > x) \sim d_\vee x^{-\alpha/\phi_j}, \quad (2.12)$$

where $d_\wedge$ and $d_\vee$ are defined to satisfy, respectively,

$$\begin{aligned} 2C_\alpha \sum_{k \in \mathcal{C}_i \cap \mathcal{C}_j} (w_{ki} \gamma_k)^\alpha c_i^{\alpha/\phi_i} &\leq d_\wedge \leq 2^{\alpha+1} C_\alpha \sum_{k \in \mathcal{C}_i} (w_{ki} \gamma_k)^\alpha c_i^{\alpha/\phi_i}, \\ 2C_\alpha \sum_{k \in \mathcal{C}_i \cap \mathcal{C}_j} (w_{kj} \gamma_k)^\alpha c_j^{\alpha/\phi_j} &\leq d_\vee \leq 2^{\alpha+1} C_\alpha \sum_{k \in \mathcal{C}_j} (w_{kj} \gamma_k)^\alpha c_j^{\alpha/\phi_j}. \end{aligned}$$

Remark 1. If $\mathbf{s}_i$ and $\mathbf{s}_j$ are so distant that $\mathcal{C}_i \cap \mathcal{C}_j = \emptyset$, R_i and R_j are completely independent and hence $\chi_{ij}^R = 0$. If there is at least one basis function that covers both $\mathbf{s}_i$ and $\mathbf{s}_j$, $\chi_{ij}^R = \lim_{x \rightarrow \infty} \Pr(R_j > x | R_i > x) = C_K(\mathbf{w}_i, \mathbf{w}_j, \boldsymbol{\gamma}) / \bar{\gamma}_i^\alpha > 0$, which means R_i and R_j are asymptotically dependent.

In the following theorem, we describe the tail dependence properties of our proposed model given in Section 2.2. It shows that the model can be simultaneously AD (or AI) at short spatial distances and AI at long spatial distances. It makes precise how the interplay between the compact kernels and the spatially varying tail parameter $\phi(\mathbf{s})$ determines the class and strength of the spatial tail dependence. Using the existing results on regular variation from Breiman (1965), Cline (1986) and Engelke et al. (2019), it presents bounds for the dependence coefficients χ_{ij} and η_{ij} , which depend on the spatial separation of the locations $\mathbf{s}_i$ and $\mathbf{s}_j$, the configuration of the compact kernels, and the value of the tail parameter $\phi(\mathbf{s})$ at $\mathbf{s}_i$ and $\mathbf{s}_j$.

Theorem 3. *Under the definitions and notation established in the previous sections, for locations s_i and s_j , we denote*

$$v_{ki} = \frac{(w_{ki}\gamma_k)^\alpha}{\sum_{k' \in \mathcal{C}_i} (w_{k'i}\gamma_k)^\alpha}, \quad v_{kj} = \frac{(w_{kj}\gamma_k)^\alpha}{\sum_{k' \in \mathcal{C}_j} (w_{k'j}\gamma_k)^\alpha},$$

and $v_{k,\wedge} = \min(v_{ki}, v_{kj})$, $v_{k,\vee} = \max(v_{ki}, v_{kj})$. Also, let $W_j = g(Z_j)$, $j = 1, \dots, D$.

(a) *If $\mathcal{C}_i \cap \mathcal{C}_j \neq \emptyset$, the dependence class of the pair $(X_i, X_j)^T$ depends on the tail index parameters ϕ_i and ϕ_j .*

(i) *If $\alpha < \phi_i < \phi_j$, the pair $(X_i, X_j)^T$ is asymptotically dependent with $\eta_{ij} = 1$ and*

$$\chi_{ij} = \mathbb{E} \left\{ \min \left(\frac{W_i^{\alpha/\phi_i}}{\mathbb{E}(W_i^{\alpha/\phi_i})}, \frac{W_j^{\alpha/\phi_j}}{\mathbb{E}(W_j^{\alpha/\phi_j})} \right) \right\} \sum_{k=1}^K v_{k,\wedge}.$$

(ii) *If $\phi_i < \phi_j < \alpha$, the pair $(X_i, X_j)^T$ is asymptotically independent with $\chi_{ij} = 0$ and*

$$\begin{cases} \eta_{ij} = \eta_{ij}^W, & \text{if } \eta_{ij}^W > \phi_j/\alpha, \\ \eta_{ij} \in [\eta_{ij}^W, \phi_j/\alpha], & \text{if } \phi_i/\alpha < \eta_{ij}^W < \phi_j/\alpha, \\ \eta_{ij} \in [\phi_i/\alpha, \phi_j/\alpha], & \text{if } \eta_{ij}^W < \phi_i/\alpha. \end{cases}$$

(iii) *If $\phi_i < \alpha < \phi_j$, the pair $(X_i, X_j)^T$ is also asymptotically independent with $\chi_{ij} = 0$ and*

$$\begin{cases} 1/\eta_{ij} \in [(1 - \phi_i/\alpha)/(2\eta_{ij}^W) + 1, 2 - \phi_i/\alpha], & \text{if } \eta_{ij}^W \leq (\phi_i/\alpha + \phi_j/\alpha)/2, \\ 1/\eta_{ij} \in [(1 + \phi_j/\alpha)/(2\eta_{ij}^W), 2 - \phi_i/\alpha], & \text{if } \eta_{ij}^W > (\phi_i/\alpha + \phi_j/\alpha)/2. \end{cases}$$

(b) *If $\mathcal{C}_i \cap \mathcal{C}_j = \emptyset$, the pair $(X_i, X_j)^T$ is asymptotically independent with $\chi_{ij} = 0$. When $\rho_{ij} := \text{Cor}(Z_i, Z_j) = 0$, X_i and X_j are completely independent. When $\rho_{ij} \neq 0$,*

(i) *If $\alpha < \phi_i < \phi_j$,*

$$\begin{cases} \eta_{ij} = 1/2, & \text{if } \phi_i/\alpha > 2, \\ \eta_{ij} \in [1/2, \alpha/\phi_i], & \text{if } \phi_i/\alpha < 2 < \phi_j/(\alpha\eta_{ij}^W), \\ \eta_{ij} \in [\alpha\eta_{ij}^W/\phi_j, \alpha/\phi_i], & \text{if } \phi_j/(\alpha\eta_{ij}^W) < 2. \end{cases}$$

(ii) *If $\phi_i < \phi_j < \alpha$,*

$$\begin{cases} \eta_{ij} = \eta_{ij}^W, & \text{if } \eta_{ij}^W > \phi_j/\alpha, \\ \eta_{ij} \in [\eta_{ij}^W, \phi_j/\alpha], & \text{if } \eta_{ij}^W < \phi_j/\alpha. \end{cases}$$

(iii) *If $\phi_i < \alpha < \phi_j$,*

$$\begin{cases} 1/\eta_{ij} \in [1/(1 + \rho_{ij}) + 1, 2], & \text{if } 2\eta_{ij}^W \leq \phi_j/\alpha, \\ 1/\eta_{ij} \in [(1 + \phi_j/\alpha)/(1 + \rho_{ij}), 2], & \text{if } 2\eta_{ij}^W > \phi_j/\alpha. \end{cases}$$

Remark 2. In Theorem 3(a), when $\mathcal{C}_i \cap \mathcal{C}_j \neq \emptyset$, both AD and AI are possible as the dependence strength is controlled by both $(\phi_i, \phi_j)^\top$ and the weights $(\mathbf{w}_i^\top, \mathbf{w}_j^\top)^\top$. Roughly speaking, larger $\min\{\phi_i, \phi_j\}$ or smaller $\|\mathbf{w}_i - \mathbf{w}_j\|$ induces stronger dependence. Additionally, when $\mathcal{C}_i \cap \mathcal{C}_j = \emptyset$, the dependence structure of $(Z_i, Z_j)^\top$ is recovered.

Remark 3. Theorem 3 suggests a natural interpretation of the spatially varying tail parameter $\phi(\mathbf{s})$. For a given location $\mathbf{s}_i$, if $\phi(\mathbf{s}_i)$ is larger than α , then $X(\mathbf{s}_i)$ is asymptotically dependent with $X(\mathbf{s}_j)$ at location $\mathbf{s}_j$, provided that $\phi(\mathbf{s}_j)$ is also larger than α and $\mathbf{s}_i$ shares a kernel with $\mathbf{s}_j$. However, if $\phi(\mathbf{s}_i)$ is smaller than α , then $X(\mathbf{s}_i)$ can never be asymptotically dependent with any other location. Therefore, the surface $\phi(\mathbf{s})$ represents the *potential strength of tail dependence* of the process at all locations $\mathbf{s}$ with any other location.

2.2.4 Examples

In this section, we provide a few concrete examples for the dependence model $\{X(\mathbf{s})\}$ defined in (2.5) and selections for the Stable variable parameters.

Example 1 (Huser and Wadsworth (2019) process). This is a scale mixture model defined in (2.2) where $\{Z(\mathbf{s})\}$ is a stationary spatial process with hidden regular variation, the link function $g(\cdot)$ transforms the margins of $\{Z(\mathbf{s})\}$ to standard Pareto (corresponding to $\delta = 1$ in (2.8)) and

$$R | \phi^H \sim \text{Pareto} \left(\frac{1 - \phi^H}{\phi^H} \right), \quad \phi^H \in [0, 1].$$

This model is tail equivalent to a special case of (2.5) obtained by fixing $K \equiv 1$, $r = \infty$ and $\phi(\mathbf{s}) \equiv \phi$, and also making $\delta = 1$, which means $R(\mathbf{s}) \equiv S_1 \sim \text{Stable}(\alpha, 1, \gamma_1, 1)$. If we denote $\alpha/\phi = (1 - \phi^H)/\phi^H$, the tail behavior in (2.10) corresponds to that of Equation (9) in Huser and Wadsworth (2019): the case when $\phi \in (0, \alpha)$ corresponds to $\phi^H \in [0, 1/2)$ which induces asymptotic independence, while $\phi \in (\alpha, \infty)$ corresponds to $\phi^H \in (1/2, 1)$ which induces asymptotic dependence. When $\phi = \alpha$, the tail decays with the same order as a Gamma random variable with rate 1 and shape 2, which is the same as the survival function of the case $\phi^H = 1/2$.

Furthermore, since $K \equiv 1$ and $r = \infty$, the interval for η_{ij} from Theorem 3(a) reduces to a singleton and we have

$$\chi_{ij} = \mathbb{E} \left[\min \left\{ \frac{W_i^{\alpha/\phi}}{\mathbb{E}(W_i^{\alpha/\phi})}, \frac{W_j^{\alpha/\phi}}{\mathbb{E}(W_j^{\alpha/\phi})} \right\} \right] \mathbb{1}(\phi < \alpha),$$

which is exactly equation (10) in Huser and Wadsworth (2019) given $\alpha/\phi = (1 - \phi^H)/\phi^H$. Similarly, we can show that η_{ij} under the stationary version of (2.5) is also equivalent to that of Huser and Wadsworth (2019, see their equation (11)).

Example 2 (Huser and Wadsworth (2019) process with spatially-varying R). Similar to the generalization from the Huser et al. (2017) process to model (2.4) by Hazra et al. (2025), we can allow $K > 1$ and $r < \infty$ while holding the parameters for the Stable variables constant at all knots. That is,

$$\phi(\mathbf{s}) \equiv \phi, S_k \stackrel{\text{iid}}{\sim} \text{Stable}(\alpha, 1, \gamma, \delta), k = 1, \dots, K.$$

To achieve local asymptotic dependence (i.e., scenario Theorem 3(a)(i)) with a constant ϕ , the value of ϕ has to be greater than α . Then, as in Hazra et al. (2025), local asymptotic independence is not possible under this construction because $\phi(\mathbf{s}) \equiv \phi > \alpha$ and (a)(i) always holds locally. However, this case achieves more flexible local dependence properties than the model specified in (2.4) as the value of $\phi \in (\alpha, \infty)$ is not fixed (recall that ϕ has to be fixed at 0 in Hazra et al. (2025) because the Huser et al. (2017) model can only achieve asymptotic dependence when $\phi = 0$).

Illustration 1 (Empirical evaluations of the bounds in Theorem 3). We simulate $N = 300,000,000$ data sets on the domain $\mathcal{S} = [0, 10] \times [0, 10]$ using a $\{\phi(\mathbf{s}) : \mathbf{s} \in \mathcal{S}\}$ surface that varies smoothly between 0.4 to 0.7; see Figure 2.2. For each simulation, we generate independent Stable variables with $\alpha = 0.5$ at 9 knots on a uniform grid and average them using compact Wendland basis functions centered at the knots (Wendland, 1995). We empirically estimate the $\chi_{ij}(u)$ and $\eta_{ij}(u)$ functions defined in (2.1) and (2.6) between pairs of the sample points on a grid of u values, using the N independent simulations of $(R_i^{\phi_i} W_i, R_j^{\phi_j} W_j)$.

Corresponding to Theorem 3(a)(i), Figure 2.3a shows the empirical estimated $\chi_{12}(u)$ and $\eta_{12}(u)$ for the 1st and 2nd sample points over a grid of u . These two points share a common Wendland kernel and $\alpha = 0.5 < \phi_2 < \phi_1$, so they are asymptotically dependent. Corresponding to Theorem 3(a)(ii), Figure 2.3b shows the empirical estimated $\chi_{34}(u)$ and $\eta_{34}(u)$ for the 3rd and 4th sample points. These two points share a common Wendland kernel, but $\phi_3 < \phi_4 < \alpha$, so they are asymptotically independent. Figure 2.3c shows the empirical estimated $\chi_{45}(u)$ and $\eta_{45}(u)$ for the 4th and 5th sample points. This corresponds to Theorem 3(a)(iii), as the two points also share a common Wendland kernel, yet $\phi_4 < \alpha < \phi_5$, so they are still asymptotically independent. Finally, Figure 2.3d shows the empirical estimated $\chi_{15}(u)$ and $\eta_{15}(u)$ for the 1st and 5th sample points. Corresponding to Theorem 3(b)(i), these two locations do not share any common Wendland kernel, so even-though $\phi_5 > \phi_1 > \alpha$, we still have asymptotically independence.

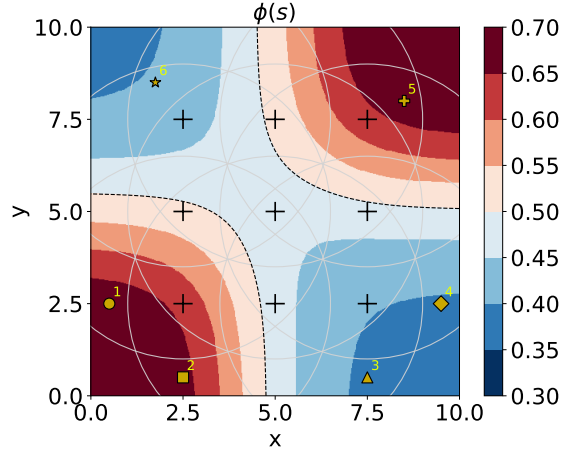


Figure 2.2: A $\phi(s)$ surface on $[0, 10]^2$, in which the dashed line marks the transition between local AI and AD. The points with ‘+’ are centers for the Stable variables and the compact Wendland basis functions. The points with other signs/marker-styles are randomly chosen sample points that we use to verify the dependence properties in Theorem 3.

2.3 Bayesian inference

We define a Bayesian hierarchical model based on the dependence model (2.5) in Section 2.2.1 and use a MCMC algorithm to fit to the data. The dependence model (2.5) is displayed again here

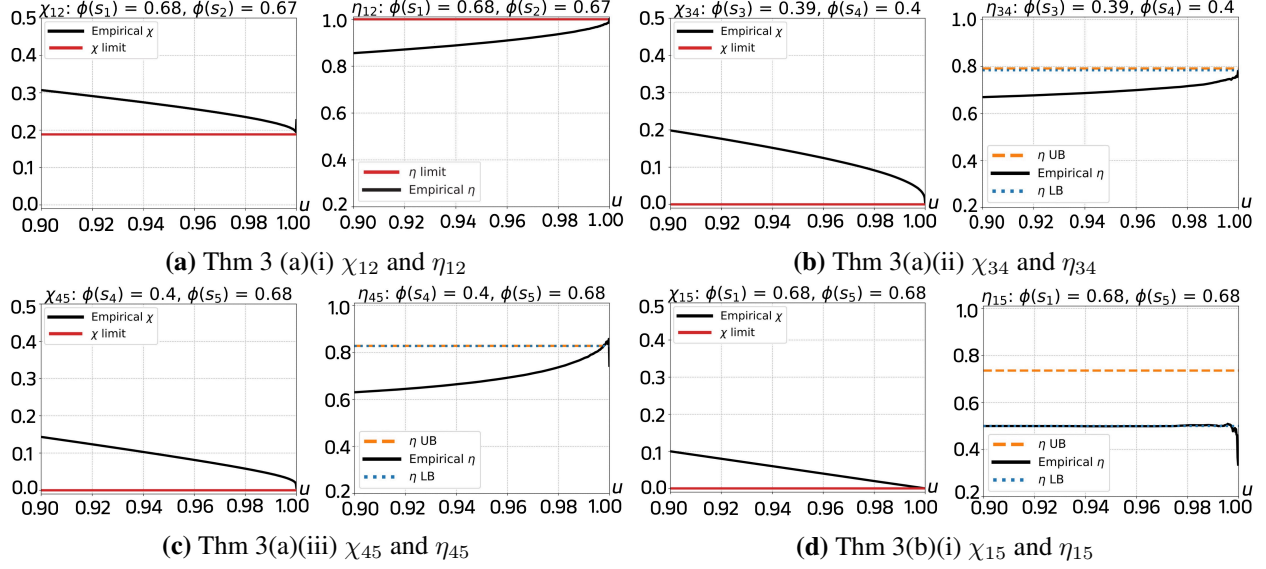


Figure 2.3: Empirical estimates of the dependence coefficients $\chi_{ij}(u)$ and $\eta_{ij}(u)$ between sample points i and j in Figure 2.2. The red lines mark the theoretical limit and the orange dashed and blue dotted lines respectively represent theoretical upper and lower bounds. The $\chi_{ij}(u)$'s are direct empirical estimates; due to numerical instability, particularly as the probability u approaches its upper limit of 1, we estimate the $\eta_{ij}(u)$'s using the Hill estimator with marginal transformation to standard Exponential (Leo R Belzile, 2023).

for convenience, now with each time replicate denoted with a subscript $t = 1, \dots, T$:

$$X_t(\mathbf{s}) = R_t(\mathbf{s})^{\phi(\mathbf{s})} g(Z_t(\mathbf{s})).$$

2.3.1 Hierarchical model

Let $\{Y_t(\mathbf{s}) : \mathbf{s} \in \mathcal{S}, t = 1, \dots, T\}$ denote the spatial process on the scale of the observations. We link this process to our dependence model using a marginal probability integral transform within the hierarchical model as

$$F_{Y|\theta_Y(\mathbf{s}),t}(Y_t(\mathbf{s})) = F_{X|\phi(\mathbf{s}),\bar{\gamma}(\mathbf{s}),t}(X_t(\mathbf{s})). \quad (2.13)$$

In principle, the dependence model can be used for any marginal distribution. Here we opt for the block-maxima approach—considering annual maxima at time t as the observed process $Y_t(\mathbf{s})$, assuming a Generalized Extreme Value (GEV) marginal distribution. We let the GEV marginal

parameters vary in space and time as

$$Y_t(\mathbf{s}) \sim \text{GEV}(\mu_t(\mathbf{s}), \sigma_t(\mathbf{s}), \xi_t(\mathbf{s})).$$

Let $\boldsymbol{\theta}_{\text{GEV},t}(\mathbf{s}) = (\mu_t(\mathbf{s}), \sigma_t(\mathbf{s}), \xi_t(\mathbf{s}))^T$ be the vector of marginal GEV parameters at location $\mathbf{s}$ and time t . Given the time-varying parameters, it is reasonable to then assume conditional temporal independence among the annual maxima.

Next, we define a hierarchical model based on the mixture (2.5). Conditioning on the scaling variables at the knots $\mathbf{S}_t$, the full conditional likelihood for the observation vector at time t is

$$\mathcal{L}(\mathbf{Y}_t | \mathbf{R}_t, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\theta}_{\text{GEV}}, \boldsymbol{\rho}) = \varphi_D(\mathbf{Z}_t) \left| \frac{\partial \mathbf{Z}_t}{\partial \mathbf{Y}_t} \right| \quad (2.14)$$

Since we have assumed conditional temporal independence by introducing the time-varying marginal parameters, likelihoods across the independent time replicates are multiplied together for the joint likelihood. In the likelihood (2.14), φ_D is the D -variate Gaussian density function with covariance matrix $\boldsymbol{\Sigma}_\rho$ and $\partial \mathbf{Z}_t / \partial \mathbf{Y}_t$ is the Jacobian. Additional details are included in the Appendix A.2.

For $\boldsymbol{\Sigma}_\rho$, we use a locally isotropic, non-stationary Matérn covariance function (Paciorek and Schervish, 2006; Risser and Calder, 2015)

$$C(\mathbf{s}, \mathbf{s}') = \zeta(\mathbf{s})\zeta(\mathbf{s}') \frac{\sqrt{\rho(\mathbf{s})\rho(\mathbf{s}')}}{\{\rho(\mathbf{s}) + \rho(\mathbf{s}')\}/2} \mathcal{M}_\nu \left(\frac{\|\mathbf{s} - \mathbf{s}'\|}{\sqrt{\{\rho(\mathbf{s}) + \rho(\mathbf{s}')\}/2}} \right), \quad (2.15)$$

where the standard deviation process $\zeta(\mathbf{s}) \equiv 1$ for $\mathbf{s} \in \mathcal{S}$, $\mathcal{M}_\nu(\cdot)$ is the Matérn correlation function with range 1 and smoothness ν , and $\rho(\mathbf{s})$ is the spatially varying range parameter.

As described in Section 2.2.1, we form a spatially varying surface $R(\mathbf{s})$ by combining K compactly supported Wendland kernel functions, each centered at a knot, each scaled by a corresponding Stable variable. In addition, we construct the $\{\phi(\mathbf{s})\}$ and $\{\rho(\mathbf{s})\}$ surfaces using Gaussian kernel functions centered at the knots.

The Wendland kernel function is parameterized as $w_k^{(S)}(\mathbf{s}) \propto \left(1 - \frac{\|\mathbf{s} - \mathbf{b}_k^{(S)}\|^2}{r}\right)_+^l$, in which $l = 2$, $\mathbf{b}_k^{(S)}$ are a grid of knots over the spatial domain $\mathcal{S}$, and r is the radius of the kernel function.

The priors for the dependence model parameters are $\phi_k \stackrel{\text{iid}}{\sim} \text{Beta}(5, 5)$ and $\rho_k \stackrel{\text{iid}}{\sim} \text{halfN}(0, 2)$, $k = 1, \dots, K$. The “halfN” refers to the positively truncated normal distribution. The Beta prior for ϕ_k is centered at the transition boundary between AI and AD, and places less mass near the edges of the support which correspond to extremely strong or weak dependence scenarios. In the following, we fix the scale parameter γ of the Stable distribution (somewhat arbitrarily) at 0.5. Varying γ does not play any role in modulating the tail dependence characteristics of the model (see Theorem 3), so fixing it at a convenient value results in almost no loss of flexibility.

2.3.2 Computation

To estimate the posterior distribution of the model parameters, we sequentially update each parameter using an adaptive random walk Metropolis (RWM) algorithm (Shaby and Wells, 2010). Since we assume independence across time, we can update R_t ’s in parallel across $t = 1, \dots, T$ at each MCMC iteration.

The probability integral transform in (2.13) and the Jacobian term in the likelihood (2.14) require evaluation of the marginal distribution and density functions of the dependence model $X(\mathbf{s})$. For general Stable variables $S_1, \dots, S_K$, this is difficult. However, under the special case of $\alpha = 1/2$, sometimes called a Lèvy distribution, we can obtain simpler analytical forms. Furthermore, fixing $\alpha = 1/2$ sacrifices no flexibility with respect to the tail dependence characteristics described in Theorem 3. When $\alpha = 1/2$, the survival function for the mixture in (2.5) with the Type II (i.e. location-shifted) Pareto link function in 2.8 is

$$1 - F_{X_j}(x) = P(R_j^{\phi_j} W_j > x) = \sqrt{\frac{\bar{\gamma}_j}{2\pi}} \int_0^\infty \frac{r^{\phi_j-3/2}}{x + r^{\phi_j}} \exp\left\{-\frac{\bar{\gamma}_j}{2r}\right\} dr. \quad (2.16)$$

Using the Type II Pareto link function, rather than the standard Pareto, does not change the tail properties but can be advantageous for MCMC sampling because it aligns the support of the ran-

dom scaling factor and the transformed Gaussian process. The trade-off is that the Type II Pareto, unlike the standard Pareto, does not give a closed form for the univariate distribution function for $X_j(s)$, and therefore requires numerical integration. Using Leibniz rule to take derivative with respect to x , we can get the univariate density function for $X_j(s)$,

$$f_{X_j}(x) = \sqrt{\frac{\bar{\gamma}_j}{2\pi}} \int_0^\infty \frac{r^{\phi_j-3/2}}{(x+r^{\phi_j})^2} \exp\left\{-\frac{\bar{\gamma}_j}{2r}\right\} dr, \quad (2.17)$$

which also requires numerical integration.

We evaluate the numerical integrals (e.g. the univariate distributions (2.16) and density functions (2.17) of the dependence model) using `GSL` libraries (Gough, 2009) in `C++`. The MCMC sampler is implemented in `python`, and the parallel updates are implemented via the `mpi4py` (Dalcin and Fang, 2021) module/interface to `OpenMPI`. After parallelization, on an AMD Milan EPYC CPU, running a chain, on datasets of 500 spatial locations and 64 time replicates, to approximately 15,000 iterations takes about 270 hours. Compared with the stationary model of Huser and Wadsworth (2019), fitting our model approximately triples the computational cost in our implementation.

2.3.3 Simulation and Coverage Analysis

We present simulation results and conduct coverage analysis to investigate whether the MCMC procedure is able to draw accurate inference on model parameters, assuming marginally GEV responses. We use $D = 500$ sites uniformly drawn from the square $\mathcal{S} = [0, 10]^2$ and $\mathcal{T} = 64$ time replicates. The latent Gaussian process $Z_t(\mathbf{s})$ is generated with a locally isotropic, non-stationary Matérn covariance function as specified in Section 2.3.1, with $\nu = 0.5$. For each time replicate, we specified $K = 9$ knot locations $\{\mathbf{b}_k^{(S)}, k = 1, \dots, K\}$ over a regularly spaced grid in the spatial domain $\mathcal{S}$. As specified in Section 2.3.1, we generate independent Lévy random variables with $\gamma_k = 0.5$ at those nine pre-specified knot locations and interpolate them to site locations using the Wendland kernel functions with radius $r = 4$ centered at those knots; see Figure 2.4 for illustration.

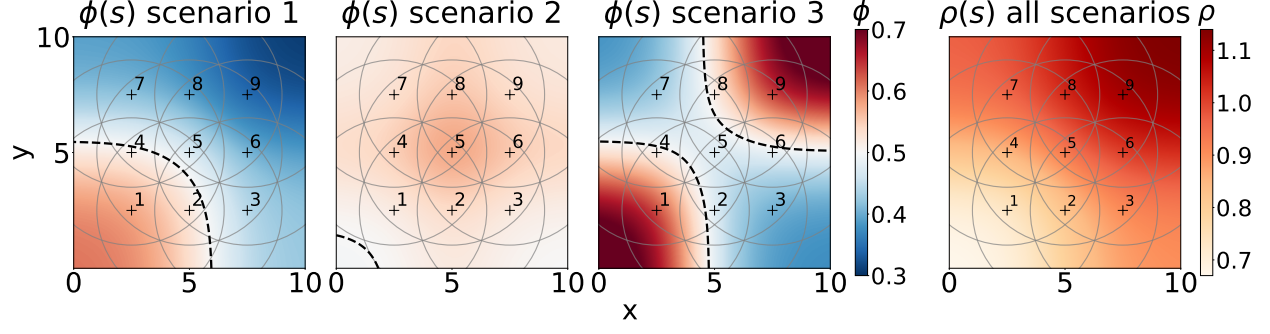


Figure 2.4: Dependence model parameter surfaces, $\phi(s)$ and $\rho(s)$, for the three simulation scenarios, which correspond to different levels of non-stationarity. The dashed line marks the transition between local AI and AD. The ‘+’ denotes the center of the kernel functions, and the gray circles denotes the radii.

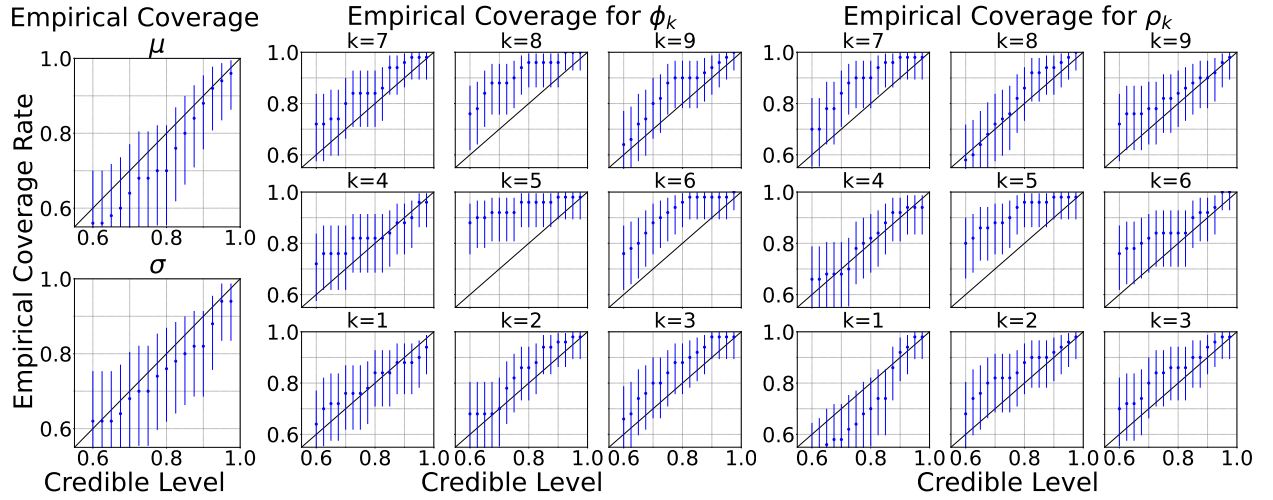


Figure 2.5: Empirical coverage rates of credible intervals of the marginal parameters μ and σ (left), the dependence parameters ϕ_k (middle), and ρ_k (right), $k = 1, \dots, 9$, in simulation scenario 1.

We consider three scenarios for the scaling parameter $\phi(s)$ and the range parameter $\rho(s)$ so that all cases in Theorem 3 are represented; we generate the nine ϕ_k and ρ_k at the same nine knots locations ($\mathbf{b}_k^{(S)} = \mathbf{b}_k^{(\phi)} = \mathbf{b}_k^{(\rho)}$), and interpolate them to the site locations using the Gaussian kernel functions with bandwidth $h^{(\phi)} = h^{(\rho)} = 4$ centered at those knots. The interpolated $\{\phi(s)\}$ and $\{\rho(s)\}$ surfaces are shown in Figure 2.4. Finally, we generated dependence model variables $X_t(s)$ under three scenarios and transformed to marginal GEV distribution with parameters $(\mu(s), \sigma(s), \xi(s))^T = (0, 1, 0.2)^T$. For the purpose of simulations, GEV parameters are set to be spatially and temporally constants with no covariate. For computational expediency in the coverage analysis, $\xi(s)$ is not updated in the simulations.

We study the coverage properties of the posterior inference based on the MCMC samples for the posterior credible intervals with 50 simulated datasets drawn under each of the scenarios. Figure 2.5 shows the empirical coverage rates of the scaling parameter ϕ and the range parameter ρ at the nine knot locations (i.e. $k = 1, \dots, 9$), as well as the location μ and scale σ of the marginal GEV parameters in simulation scenario 1. Standard binomial confidence intervals are included on the coverage plots. In all scenarios, we see that the sampler generates well-calibrated posterior inference for the GEV parameters with close to nominal frequentist coverage, and slightly over-covers for the the scaling and range parameters. The Appendix A.3 includes the additional empirical coverage rates of the parameters from simulation scenario 2 and 3 in Figures A.1 and A.2, which show similar characteristics as the coverage plots from simulation scenario 1.

2.4 Extreme *in situ* of Daily Precipitation

2.4.1 Data Analysis

In this section, we analyze extreme daily measurements from a gauged network of *in situ* weather stations from the Global Historical Climate Network (GHCN; Menne et al., 2012). A subset of the extreme summertime measurements from GHCN stations (see Figure 2.6a) over the central United States was originally analyzed in Zhang et al. (2022a) using the Huser and Wadsworth (2019) copula with a spatio-temporally varying marginal model similar to the one we use below. However, exploratory analysis shown in Figure 2.1 suggests that applying a single dependence class to such a large spatial domain is inappropriate, as there are areas in the central U.S. domain where the extreme summer precipitation appear to be locally AI and others that appear to be AD. Furthermore, when applying a single dependence class to the entire domain, Zhang et al. (2022a) found that (overall) the extreme precipitation measurements are asymptotically *independent*, meaning that the risk of concurrent extremes would be underestimated for sub-regions that exhibit asymptotic dependence. For comparison, we also approximately replicate the analysis of Zhang et al. (2022a), who used the Huser and Wadsworth (2019) copula on a similar dataset (results shown in Table 2.1 and Figure 2.8). Although we cannot make direct comparisons to the broadly

similar Hazra et al. (2025) model because it does not permit spatial prediction at un-observed locations, our results from Section 2.4.3 would suggest spatially heterogeneous transition between AD and AI (see Figure 2.10), and thus the Hazra et al. (2025) model would fit poorly because it only allows AD at short spatial lags.

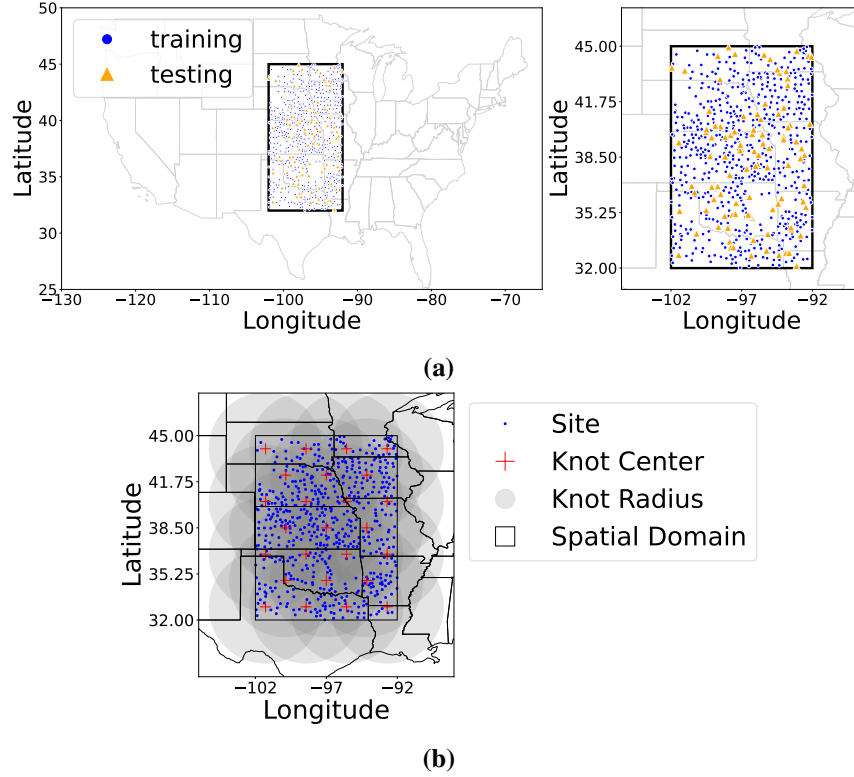


Figure 2.6: (a) Location of the 590 observations/data sites and the 99 out-of-sample testing sites. (b) An illustration of the mixture component setup for model `k25r4b4`. The blue dots represent observation sites, red '+' symbols represent the centers of the kernel functions, and the shaded circles represent areas covered by each Wendland kernel with specified radius r .

While the GHCN database contains over twenty thousand stations over the contiguous United States, following Risser et al. (2019) we analyze a quality-controlled subset of the network from a recent 75-year period, namely those stations that have a minimum of 90% non-missing daily precipitation measurements over 1949 through 2023. We then restrict our attention to summer-time daily measurements (those from June, July, and August, or JJA) in the central United States region (defined by the $[102^\circ\text{W}, 92^\circ\text{W}] \times [32^\circ\text{N}, 45^\circ\text{N}]$ longitude-latitude bounding box), resulting

in a set of $D = 590$ stations. Our focus on the central United States in the summer season is intentional, since the majority of extreme precipitation in this region and season results from severe convective storms, which can be highly localized and therefore are particularly challenging to model statistically. Let $P_{tm}(\mathbf{s}_j)$ represent the daily precipitation measurement in millimeters for day $m = 1, \dots, 92$ (the JJA season has 92 days) in year $t = 1949, \dots, 2023$ at station $\mathbf{s}_j$. We then analyze the summertime maxima, denoted $Y_t(\mathbf{s}_j) = \max_m \{P_{tm}(\mathbf{s}_j)\}$. We only record the JJA maxima at station $\mathbf{s}_j$ in year t if that season has at least two-thirds non-missing measurements, otherwise, $Y_t(\mathbf{s}_j)$ is considered missing.

As we are analyzing block maxima, one might wonder whether each field of seasonal maxima contains observations from many individual storms. Out of the 92 days in June-July-August, on average the station-specific maxima occur on approximately 75 different days (left panel of Figure 2.7). The most extreme events occur on fewer days: 99th percentile events arise from approximately 5-8 days (orange curve in the right panel of Figure 2.7). The fact that extreme precipitation arises from so many separate storms is not surprising; summertime precipitation in the central United States is primarily convective (i.e., from thunderstorms), which is characterized by highly-localized and short-lived storms.

Corresponding to the block-maxima structured data, we assume that marginally

$$Y_t(\mathbf{s}) \sim \text{GEV}(\mu_t(\mathbf{s}), \sigma_t(\mathbf{s}), \xi_t(\mathbf{s})).$$

To ensure the independence over time and account for potential systematic increase or decrease in rainfall due to global warming, we have assumed a time-varying component for the location parameter. To account for the physical features of the terrain, we incorporate spatially-varying covariate with thin plate splines to smooth over the spatial domain. That is,

$$\mu_t(\mathbf{s}) = \mu_0(\mathbf{s}) + \mu_1(\mathbf{s}) \cdot t, \quad \sigma_t(\mathbf{s}) \equiv \sigma(\mathbf{s}), \quad \xi_t(\mathbf{s}) \equiv \xi(\mathbf{s}),$$

where the spatially varying intercept and slope parameters for the GEV location is specified as a spline as

$$\mu_i(\mathbf{s}) = \beta_0 + \beta_1 \cdot \text{elev}(\mathbf{s}) + \sum_{i=1}^{11} w_i f_{tps}(\mathbf{s}), \quad i \in \{0, 1\},$$

where f_{tps} denotes the thin-plate spline kernel (using 11 degrees of freedom to smooth over the spatial domain). The GEV scale and shape parameters are specified as linear functions of elevation, as

$$\log(\sigma(\mathbf{s})) = \beta_0 + \beta_1 \cdot \text{elev}(\mathbf{s})$$

$$\xi(\mathbf{s}) = \beta_0 + \beta_1 \cdot \text{elev}(\mathbf{s}).$$

For justification of these specific marginal modeling choices, please see Appendix A.4.

We place knots in an isomorphic grid across the spatial domain. We considered thirteen different model configurations, with various combinations of different knot grids, Wendland kernel radii, Gaussian kernel bandwidths, and restricting the marginal parameters to be spatially constant. Table 2.1 describes the thirteen different setups, and Figure 2.6b illustrates the mixture component setup of the model that we eventually chose. To obtain MCMC starting values for the marginal

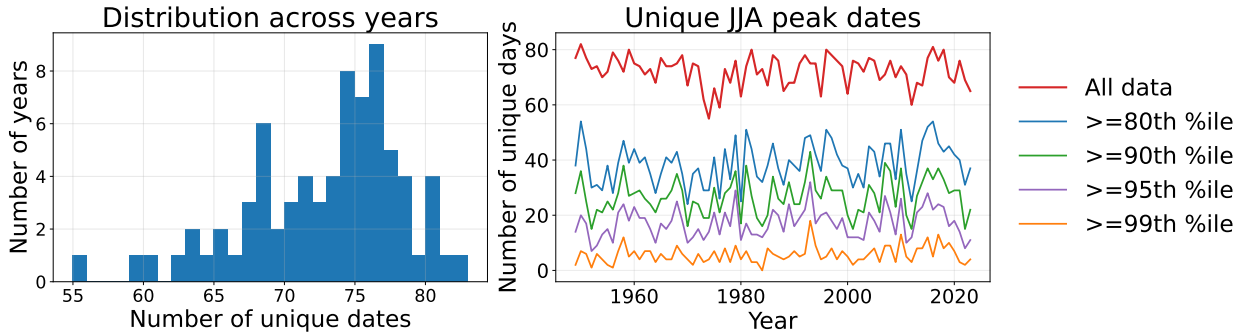


Figure 2.7: Across 590 locations, histogram (left) of the unique dates of the yearly maxima across all sites; line charts (right) of the unique dates of the upper percentile maxima.

parameters, we fit simple GEV regressions, assuming conditional independence across space and time. For the dependence model parameters, we fit empirical variograms to obtain initial values

for the range parameter ρ , set $\phi(s)$ to start at 0.5 everywhere, and used the median observation to calculate a starting point for the random scale factor R . We then ran the MCMC chains for approximately 10,000 iterations (or until convergence) such that after discarded a burn-in period, we obtain at least 5,000 posterior samples.

Table 2.1: 13 model configurations spanning different knot grids, spatial extent of basis functions, and constraints on the marginal parameters. Model naming convention is as follows: k , r , b , and m respectively denote the number of knots, radius of the compact Wendland kernel, the bandwidth of the Gaussian kernel, and restriction indicator on marginal GEV parameters. Here the effective range refers to the distance at which the Wendland kernel function becomes 0 (as it is compact) or when the Gaussian kernel function drops below 0.05. The H-W Stationary model is the stationary Huser and Wadsworth (2019) process used by Zhang et al. (2022a).

Model Name	# of Knots		Basis Effective Range			Constraint
	S, ϕ	ρ	S	ϕ	ρ	
1. H-W Stationary	1	1	∞	∞	∞	None
2. k13r4b4	13	13	4	4.89	4.89	None
3. k13r4b4m	13	13	4	4.89	4.89	Fix μ, σ, ξ
4. k25r2b0.67	25	25	2	2	2	None
5. k25r2b0.67m	25	25	2	2	2	Fix μ, σ, ξ
6. k25r2b2	25	25	2	3.46	3.46	None
7. k25r2b2m	25	25	2	3.46	3.46	Fix μ, σ, ξ
8. k25r4b4	25	25	4	4.89	4.89	None
9. k25r4b4m	25	25	4	4.89	4.89	Fix μ, σ, ξ
10. k41r1.6b0.43	41	41	1.6	1.6	1.6	None
11. k41r1.6b0.43m	41	41	1.6	1.6	1.6	Fix μ, σ, ξ
12. k41r2b0.67	41	41	2	2	2	None
13. k41r2b0.67m	41	41	2	2	2	Fix μ, σ, ξ

2.4.2 Model Evaluation

To evaluate model fit, we incorporate additional observations as out-of-sample test data. These stations have $> 85\%$ and $\leq 90\%$ non-missing daily precipitation data measurements over the same time period and spatial domain. The JJA maxima are chosen according to the same criteria

as the training dataset, enforcing at least $2/3$ non-missing values in any given year. This results in 99 stations in the test set, drawn as the yellow triangles in Figure 2.6a. We evaluate model fit and compare the thirteen models based on their predictive log-likelihoods at these 99 stations.

To obtain the log-likelihoods at the out-of-sample sites, we draw from predictive distributions of all model parameters at each MCMC iteration. For models that have fixed marginal parameters, we use the initial estimates from the GEV fit to interpolate to the testing sites. Figure 2.8 displays the boxplots of log-likelihoods at the testing sites for the thirteen models we considered. Based on its superior log-likelihood performance, we decided to use the `k25r4b4` model (Figure 2.8).

It appears that models which jointly estimate the marginal parameters within the MCMC tend to perform systematically better than models with fixed marginal parameters. This suggests that the common practice of performing the analysis in two steps—first estimating the marginal parameters, then performing the dependence analysis “downstream” by plugging in the empirical marginal estimates to transform the data to convenient margins—is prone to underperform. We believe that it is worth the extra effort to estimate the marginal model parameters together with the dependence model parameters in one unified hierarchical model.

In addition, we looked at the empirical quantile plots of the out-of-sample observations to evaluate the model’s marginal fit. We compared GEV quantiles based on posterior predictive draws of the marginal parameters at the holdout sites to the empirical quantiles at the same holdout sites. Figure 2.9 shows QQ-plots from four sites randomly selected from the 99 testing sites, for the best-performing `k25r4b4` model. The model provides a decent marginal fit, as the 95% confidence band contains the 1:1 line in each case.

2.4.3 Results

We now present results from the chosen `k25r4b4` model. First, we examine the estimates for $\phi(\mathbf{s})$ at the knot locations, specifically whether they fall within $(0, 1/2]$ or $(1/2, 1)$, corresponding to whether the data-generating process is asymptotically independent or dependent. It appears that this dataset exhibits asymptotically dependent behavior at short spatial ranges on the western

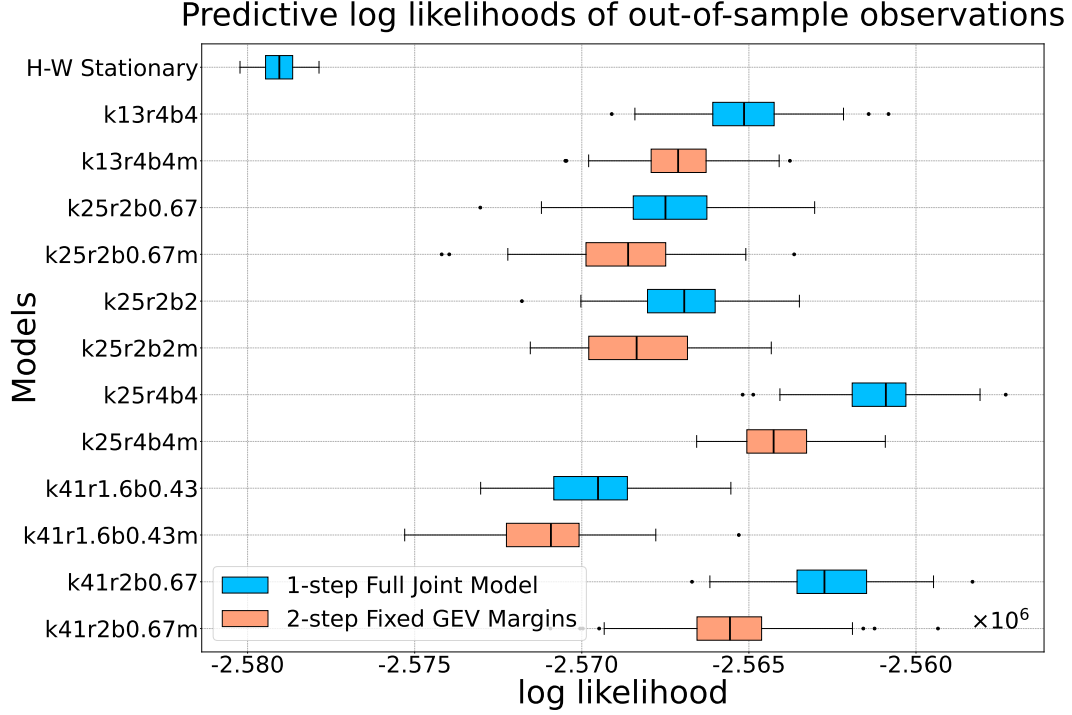


Figure 2.8: Boxplots of predictive log-likelihood for the thirteen models. Higher log-likelihood is better. “Blue” models estimate the GEV parameters in the MCMC process, while the “orange models” have some restrictions on their marginal model parameters; “blue models” perform better than their “orange counterparts”.

portion of the spatial domain, and asymptotically independent behavior at short spatial ranges on the central and eastern portions of the spatial domain. Previous spatial models, including the Hazra et al. (2025) model, cannot capture this spatially heterogeneous behavior. This highlights one of the key features of our model, which is that it simultaneously allows AI at long distances and either AI or AD at short distances. In this dataset, the tail dependence appears to be non-stationary, as the estimated value of the $\phi(s)$ changes appreciably over the spatial domain; see Figure 2.10. Next, the posterior mean $\rho(s)$ surface shows that the correlation range of the latent Gaussian process is also estimated to be variable across the spatial domain. Table 2.2 reports the posterior means and the 95% equi-tail credible intervals for the dependence parameter $\phi(s)$ and range parameter $\rho(s)$ at the kernel knot locations. Finally we include the posterior mean marginal parameter surfaces in Figure 2.10. Interestingly, the slope parameter $\mu_1(s)$ is positive in about half of the spatial domain

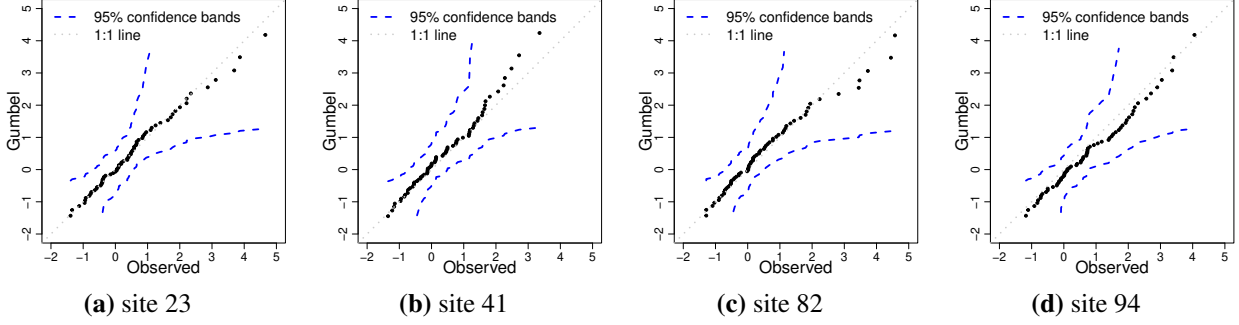


Figure 2.9: QQ-plots of four randomly selected holdout locations, comparing observed and predicted marginal quantiles for the `k25r4b4` model. 95% confidence envelopes are also shown. Marginal values are transformed to a Gumbel distribution.

and negative in the other half. The largest values correspond to a change of 7 millimeters per century in the marginal GEV location.

Table 2.2: Posterior mean and 95% equi-tail credible interval of the `k25r4b4` model. k indicates knot index.

k	ϕ_k	ρ_k	k	ϕ_k	ρ_k	k	ϕ_k	ρ_k
1	0.469 (0.271, 0.694)	0.082 (0.008, 0.168)	2	0.394 (0.184, 0.613)	0.147 (0.044, 0.228)	3	0.408 (0.199, 0.616)	0.068 (0.003, 0.165)
4	0.426 (0.218, 0.672)	0.178 (0.091, 0.269)	5	0.549 (0.316, 0.764)	0.024 (0.001, 0.066)	6	0.619 (0.355, 0.822)	0.035 (0.002, 0.095)
7	0.500 (0.276, 0.723)	0.051 (0.002, 0.145)	8	0.346 (0.168, 0.583)	0.034 (0.001, 0.100)	9	0.561 (0.321, 0.775)	0.024 (0.001, 0.072)
10	0.416 (0.212, 0.639)	0.036 (0.001, 0.114)	11	0.305 (0.121, 0.509)	0.097 (0.004, 0.254)	12	0.366 (0.172, 0.590)	0.161 (0.034, 0.337)
13	0.538 (0.301, 0.785)	0.073 (0.004, 0.187)	14	0.588 (0.319, 0.821)	0.177 (0.049, 0.297)	15	0.461 (0.224, 0.693)	0.079 (0.007, 0.185)
16	0.333 (0.179, 0.499)	0.083 (0.021, 0.146)	17	0.605 (0.376, 0.827)	0.072 (0.006, 0.162)	18	0.528 (0.289, 0.766)	0.042 (0.001, 0.100)
19	0.359 (0.161, 0.574)	0.021 (0.000, 0.062)	20	0.429 (0.213, 0.660)	0.059 (0.008, 0.108)	21	0.415 (0.195, 0.627)	0.060 (0.003, 0.137)
22	0.486 (0.251, 0.741)	0.176 (0.028, 0.301)	23	0.590 (0.323, 0.829)	0.060 (0.004, 0.146)	24	0.426 (0.208, 0.653)	0.031 (0.001, 0.094)
25	0.356 (0.147, 0.589)	0.065 (0.002, 0.173)						

To assess the tail dependence of the fitted process, we use a moving-window approach, in which we divide the spatial domain into 17×7 sub-regions and empirically estimate and visualize χ in each local window. The left-hand panel of Figure 2.11 shows the empirical $\hat{\chi}_u(h)$, while the

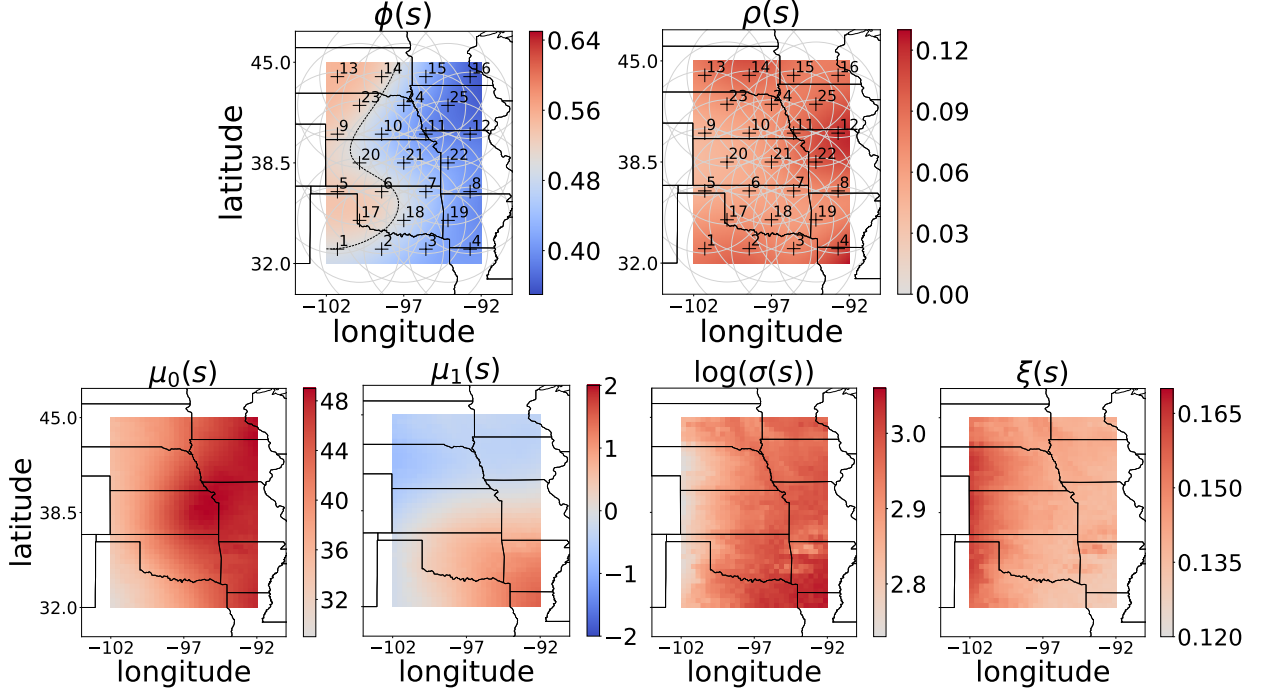


Figure 2.10: Interpolated surfaces of posterior mean for dependence and marginal models of k25r4b4. The tail dependence parameter $\phi(s)$ and range parameter $\rho(s)$ on top and the marginal parameters on bottom. For the $\phi(s)$ surface, the dashed black line marks $\phi(s) = 0.5$, the transition between AD and AI.

right hand panel of Figure 2.11 shows the analogous model-based $\hat{\chi}_u(h)$. To obtain the empirical version, we first use the k25r4b4 model-fitted marginal GEV parameters to transform the observations $Y_t(s)$ to the $X_t(s)$ scale, then empirically estimate $\hat{\chi}_u(h)$. We do this across three quantile levels, $u = 0.9, 0.95$ and 0.99 , and three distances, approximately 75km, 150km, and 225km. We obtained the analogous model-based version of $\hat{\chi}_u(h)$ by making many unconditional draws from the fitted model, then using the same empirical estimation strategy as in the left-hand panel.

The empirical and model-based estimates of $\chi_u(h)$ show similar spatial patterns at all levels. The most prominent feature is the patch of high values along the border between Texas and Oklahoma, as well as the patch of high values near the border between Nebraska and Missouri. These features appear in both the left- and right-hand panels of Figure 2.11. We also notice that as the quantile increases, the estimates $\chi_u(h)$ decrease for all spatial distances, in the central and eastern region of the spatial domain, which is the expected behavior for asymptotic independence. This is consistent with the results shown in Figure 2.10 for the regions where $\phi(s) < 0.5$. That the overall

level of some model-based $\hat{\chi}_u(h)$ surfaces is higher than that of their empirical analogues, even as the spatial patterns and trends in u are similar, is not terribly concerning; empirical estimates of $\hat{\chi}_u(h)$ are necessarily very noisy, so we would not expect perfect alignment, even if the model perfectly captured the data-generating process. Confidence bands are included in the appendix A.5.

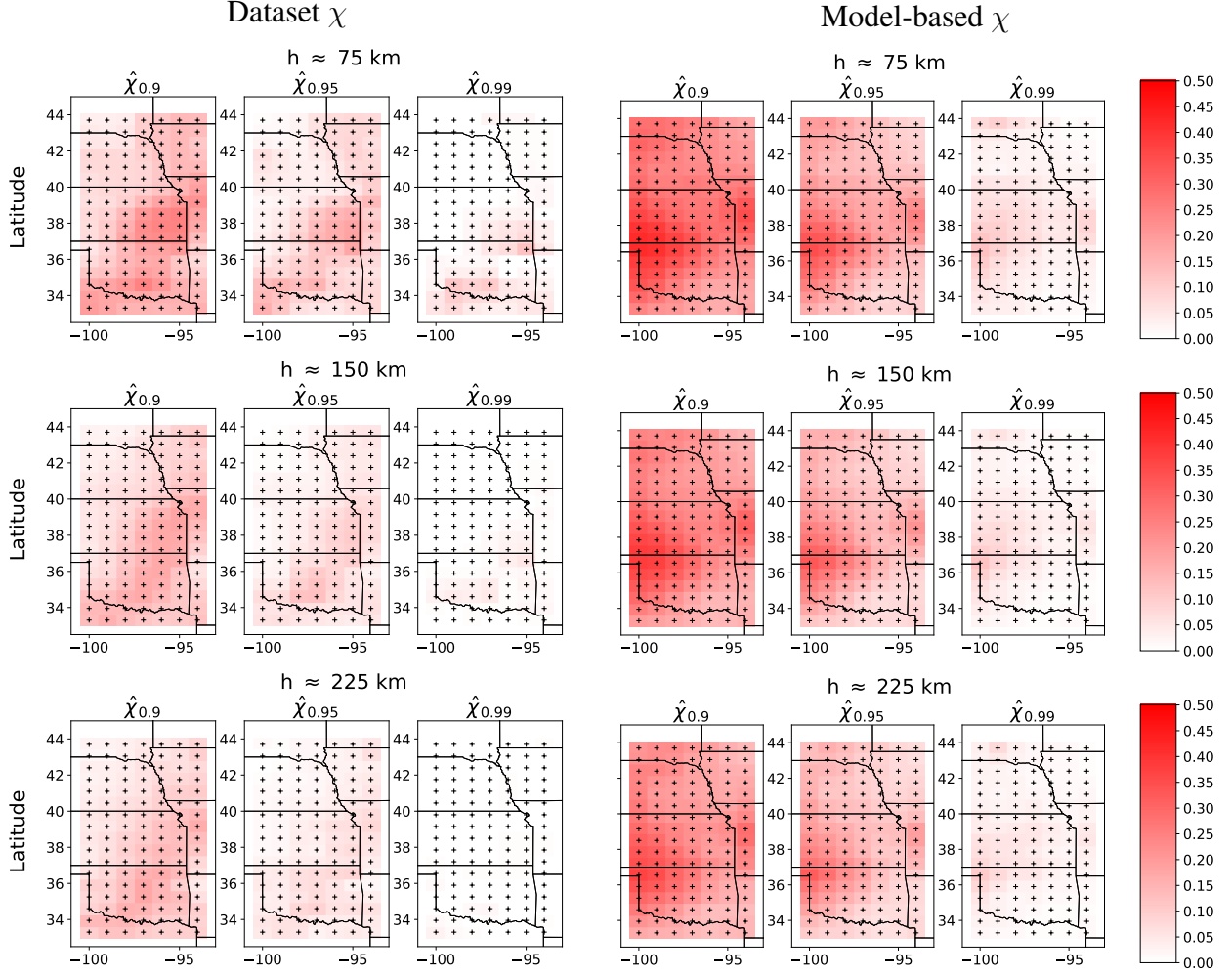


Figure 2.11: Moving-window estimates of $\chi_u(h)$ across three quantiles u and three spatial lags h . The left-hand panel shows the dataset empirical estimates of $\chi_u(h)$, and the right-hand panel shows the model-based estimates of $\chi_u(h)$, based on the k25r4b4 model.

2.5 Discussion

In this article, we have proposed a modeling approach that extends the random scale construction to obtain more flexible local and long-range tail dependence behaviors. Our proposed mixture model is capable of simultaneously exhibiting asymptotic independence at long ranges and either asymptotic dependence or independence at short ranges. The model is also able to capture non-stationary tail dependence structure with a spatially varying tail parameter that allows short-range asymptotic independence in some parts of the domain and short-range asymptotic dependence in other parts of the domain. Our model permits trivial unconditional simulation, which easily allows for direct Monte Carlo risk estimates based on complicated functionals like areal sums or extrema. It is also straightforward to do spatial prediction (i.e. interpolation) by drawing from relevant posterior predictive distributions. And because our approach is fully Bayesian, all variation in any downstream analysis is accounted for in a coherent way.

Our analysis of extreme summertime precipitation in the central US highlighted some of the model’s key features. We expected the data to exhibit short-range AI in some regions of the study area and short-range AD in other regions. As expected, the posterior mean surface of the dependence parameter showed both AI and AD at short ranges in parts of the spatial domain, and AI at long ranges. In addition, a desirable feature of a coherent flexible Bayesian model like ours is that any risk estimates based on posterior samples from our model will reflect the ambiguity in the local tail dependence regime.

An interesting side result that merits further study was that models which jointly estimated marginal and dependence parameters always performed better (in holdout set predictive log likelihoods) than analogous “two-step” models in which marginal parameters were estimated and plugged in before fitting the dependence model. Jointly estimating the marginal and dependence models as we have done incurs large costs in terms of software implementation and computational complexity. However, it seems to confer significant advantages in terms of model fit, in addition to naturally propagating variation between marginal and dependence models, and on to predictions.

Fitting our model with MCMC allows inference on spatial extreme-value datasets with relatively large numbers of locations. We have defined the model conditionally as a Bayesian hierarchical model, for which standard MCMC techniques can be used. Computation is facilitated by paralleling over time and migrating any required numerical integration to C++. Even so, the lack of closed form marginal transformations creates a significant computational challenge that scales with the total number of observations. This computational bottleneck would be even more evident if using a Generalized Pareto response for analyzing peaks-over-threshold data. While including a nugget term as in Zhang et al. (2022b) would alleviate much of the computation, it still adds an additional layer of numerical integration to the marginal transformation, which is already not in closed form.

It would be useful to find ways to alleviate the heavy computational burden required to fit our full model. Neural Bayes estimators in the style of Sainsbury-Dale et al. (2024) are likely not effective in our context because of the high dimensionality of our parameter space (i.e. thousands of parameters). Most of the computational effort in our sampler arises from the numerical integration required for joint estimation of the marginal and dependence components, which we identify as a target for future improvements, possibly including the use of neural emulators.

Finally, while many of the key features of our proposed model are also found in recent “single-site conditioning” models (Wadsworth and Tawn, 2022), to our knowledge, ours is the first fully-specified joint model to possess them. In some ways, our model can be seen as an alternative to Wadsworth and Tawn (2022)-type models. These single-site conditioning models are more parsimonious and faster to fit than what we have proposed here. However, that our model is a well-defined joint probability model gives it significant advantages in both interpretability and applicability.

Data and code for the method in this Chapter are made available at the following URL: https://github.com/muyangshi/GEV_ScaleMixture/.

Chapter 3

Log-Laplace Nuggets for Fitting Spatial Extremes

Models to Threshold Exceedances

3.1 Introduction

Flexible spatial extreme-value models based on random scale-mixture models provide a powerful framework for capturing a wide range of tail-dependence structures, including smooth transitions between different tail dependence regimes and spatially heterogeneous extremal behavior. However, despite their modeling appeal, these constructions are difficult to fit under the peaks-over-threshold (POT) framework due to severe computational bottlenecks. In this work, we slightly modify the models in a way that sidesteps the main computational challenge yet retains all of the desirable tail dependence characteristics.

Spatial extreme events such as intense precipitation, prolonged heat waves, and severe windstorms can affect broad geographic regions and trigger cascading impacts on infrastructure and interconnected services (Forzieri et al., 2018). Quantifying how such extremes co-occur across space is therefore fundamental for risk assessment, mitigation planning, and climate-resilient design (Milly et al., 2008).

In response, a large literature has emerged to address the challenge of accurately modeling dependence in the tails of spatial processes. A prominent example is the class of random scale-mixture models built on latent Gaussian processes. The preferred tool of inference for these models when applied to POT data is the censored likelihood, which retains partial information from observations below the threshold while avoiding the need to model their exact sub-threshold behavior, thereby improving efficiency without unduly increasing bias from bulk misspecification. However, evaluation of the censored likelihood requires repeated computation of high-dimensional Gaussian distribution functions, rendering inference infeasible for even moderately large numbers of loca-

tions (Zhang et al., 2022b). In this paper, we introduce a straightforward modification to this class of models by incorporating an independent multiplicative log-Laplace nugget at each spatial location. This construction yields conditional independence in the censored likelihood, resulting in a joint likelihood function that is the product of univariate densities which are available in closed form. This eliminates the need for multivariate Gaussian distribution function evaluations and enables scalable Bayesian inference for high-dimensional threshold exceedance data. We show that, under mild regular variation conditions, the spatial process that includes the proposed nugget preserves the extremal dependence structure of the underlying smooth process, with residual tail dependence coefficients unchanged and upper tail dependence coefficients modified only by multiplicative constants. We demonstrate that the proposed approach applies broadly across several modern scale-mixture models, including stationary and nonstationary constructions with spatially varying extremal dependence structures.

3.1.1 Tail Dependence for Spatial Extremes

Let $\{Y(\mathbf{s}) : \mathbf{s} \in \mathcal{S} \subseteq \mathbb{R}^2\}$ denote a spatial stochastic process of interest. Throughout, for any function $A(\mathbf{s})$ evaluated at spatial locations $\mathbf{s}_1, \dots, \mathbf{s}_D$, we use the shorthand $A_j := A(\mathbf{s}_j)$. For example, $Y_j := Y(\mathbf{s}_j)$, $j = 1, \dots, D$. Let F_{Y_j} denote the marginal distribution of Y_j . In spatial extremes modeling, it is common to separate marginal behavior from spatial dependence using a copula representation. Specifically, applying the probability integral transform yields $U_j = F_{Y_j}(Y_j)$, so that the resulting copula captures the dependence structure independently of the marginals. Interest then centers on characterize dependence in the joint upper tail of the copula.

Extremal dependence between two locations $\mathbf{s}_i, \mathbf{s}_j$ is commonly summarized by the upper tail dependence coefficient

$$\chi_{ij} = \lim_{u \rightarrow 1} \chi_{ij}(u), \quad \chi_{ij}(u) = \Pr(U_i > u \mid U_j > u). \quad (3.1)$$

The coefficient χ_{ij} quantifies how often extreme events co-occur at two sites as the quantile level u approaches one. A pair of variables exhibits asymptotic dependence (AD) if $\chi_{ij} > 0$, indicat-

ing a non-vanishing probability of joint extremes, and asymptotic independence (AI) if $\chi_{ij} = 0$, indicating that joint exceedances become increasingly rare in the limit. In the case of asymptotic independence, the coefficient χ_{ij} alone does not capture the rate at which joint exceedance probabilities decay. Additional information is provided by the residual tail dependence coefficient η_{ij} (Ledford and Tawn, 1996), defined through

$$\Pr(U_i > u \mid U_j > u) = \mathcal{L}_{ij}(1 - u)(1 - u)^{-(1-1/\eta_{ij})}, \quad (3.2)$$

where $\mathcal{L}_{ij}$ is a slowly varying function at zero, that is, $\lim_{t \rightarrow 0} \mathcal{L}(tx)/\mathcal{L}(t) = 1$ for any $x > 0$. The parameter η_{ij} summarizes the strength of extremal dependence under asymptotic independence. For a stationary isotropic process, these pairwise coefficients depend only on $h_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$, in which case we write $\chi_u(h_{ij})$, $\chi(h_{ij})$, and $\eta(h_{ij})$.

Distinguishing asymptotic dependence (AD) and asymptotic independence (AI) is a central challenge in spatial extremes modeling, especially for extrapolation beyond observed data. Incorrectly specifying the extremal dependence class can result in substantial underestimation or overestimation of the joint risk of concurrent extreme events (Huser et al., 2025). However, in practice, the tail region typically contains few observations; empirical tail diagnostics are often highly uncertain and difficult to make a definitive AD/AI classification. This motivates the need for flexible model classes that can represent both dependence regimes and allow the data to inform the dependence class.

Empirical evidence from environmental data for spatial extremes also often suggest that the conditional exceedance probability $\chi_{ij}(u)$ typically decreases as spatial separation h_{ij} increases, and also with higher quantile levels u (Shi et al., 2026). These observations motivate the development of models that can accommodate a variety of dependence features, including “scale awareness”, wherein a process may exhibit strong short-range extremal (in)dependence and long-range asymptotic independence.

3.1.2 Flexible Spatial Extreme Models

In response to these modeling needs, a growing body of work has developed flexible spatial extremes models that can bridge asymptotic dependence and asymptotic independence within a single framework, represent decreasing tail dependence with increasing spatial separation, and accommodate spatially heterogeneous dependence across large domains. An especially important and widely used class of such models is based on random scale-mixtures, which provide a unified structure for several modern constructions.

A general random scale-mixture model can be written as

$$X^*(\mathbf{s}) = R(\mathbf{s})^{\phi(\mathbf{s})} \cdot g_\phi\{Z(\mathbf{s})\} \quad (3.3)$$

where $Z(\mathbf{s})$ is a latent spatial process that is asymptotically independent at any two locations, typically a Gaussian process with covariance Σ_ρ ; $R(\mathbf{s})$ is a random scaling variable which can vary over space; $\phi(\mathbf{s})$ serves as an extremal dependence parameter that indexes how the random scaling interacts with the latent spatial process and can also vary over space; and $g_\phi(\cdot)$ is a univariate link function.

A key assumption we make is that the combination of the link function $g_\phi(\cdot)$ and scaling variable $R(\mathbf{s})$ are chosen such that the resultant process $X^*(\mathbf{s})$ has regularly varying marginal tails. That is, we assume that at every $\mathbf{s}$, for some slowly varying function $\mathcal{L}_\mathbf{s}$, $\Pr\{X^*(\mathbf{s}) > x\} = \mathcal{L}_\mathbf{s}(x) x^{-\alpha^*(\mathbf{s})}$ as $x \rightarrow \infty$ for some $\alpha^*(\mathbf{s}) > 0$, where $\alpha^*(\mathbf{s})$ is called the tail index.

Here we consider four representative constructions within the transformed Gaussian scale-mixture framework of Equation 3.3. These models differ primarily in whether the scaling mechanism is global or spatially varying and whether tail parameters are allowed to vary across space, which in turn determines the degree of scale awareness and spatial heterogeneity in extremal dependence. Together, these models illustrate the main types of extremal dependence behavior achieved by modern spatial extreme models (Huser and Wadsworth, 2019; Majumder et al., 2024;

Hazra et al., 2025; Shi et al., 2026). We briefly describe how the scale-mixture components are specified in each of the four models:

(M1) The model of Huser and Wadsworth (2019) can be written as

$$X^*(\mathbf{s}) = R^\phi g\{Z(\mathbf{s})\}^{1-\phi},$$

where R is a global scaling variable with standard Pareto distribution, $Z(\mathbf{s})$ is a stationary Gaussian process, and $g(\cdot)$ transforms $Z(\mathbf{s})$ to have standard Pareto margins. This construction yields a smooth transition between asymptotic dependence and asymptotic independence through the parameter ϕ , but it imposes the same extremal dependence structure over all spatial scales and across the entire spatial domain.

(M2) The model of Majumder et al. (2024) introduces a spatially varying random scale $R(\mathbf{s})$ derived from a Brown-Resnick max-stable process. After a monotone marginal transformation, it can be written in the transformed Gaussian scale-mixture form

$$X^*(\mathbf{s}) = R(\mathbf{s})^\phi g\{Z(\mathbf{s})\}^{1-\phi},$$

where $Z(\mathbf{s})$ is a Gaussian process and both $R(\mathbf{s})$ and $g\{Z(\mathbf{s})\}$ have standard Pareto margins. This construction yields short-range asymptotic dependence with long-range asymptotic independence.

(M3) The model of Hazra et al. (2025) also specifies spatially varying random scale through

$$X^*(\mathbf{s}) = R(\mathbf{s}) g\{Z(\mathbf{s})\}, \quad R(\mathbf{s}) = \sum_{k=1}^K B_k(\mathbf{s}; l)^{1/\gamma} R_k^*,$$

where $\{B_k(\cdot; l)\}$ are compactly supported basis functions with radius l , $k = 1, \dots, K$, $\{R_k^*\}$ are independent $\text{Pareto}(\gamma)$ variables (corresponding to $\beta = 0$ in Huser et al., 2017), and

$g(\cdot)$ is the identity link. This construction yields short-range asymptotic dependence with long-range asymptotic independence.

(M4) The model of Shi et al. (2026) can be written as

$$X^*(\mathbf{s}) = R(\mathbf{s})^{\phi(\mathbf{s})} g\{Z(\mathbf{s})\}, \quad R(\mathbf{s}) = \sum_{k=1}^K B_k(\mathbf{s}; l) S_k,$$

where $\{B_k(\cdot; l)\}$ are compactly supported basis functions with radius l , $k = 1, \dots, K$; $\{S_k\}$ are independent $\text{Stable}(\alpha, \beta, \gamma_k, \delta)$ variables; $\phi(\mathbf{s})$ is a spatially varying tail parameter surface; and $Z(\mathbf{s})$ is a Gaussian process, transformed by $g(\cdot)$ to standard Pareto margins. This construction allows both the AD and AI local dependence classes to vary across space while retaining long-range AI. Shi et al. (2026) set $\alpha = 1/2, \beta = 1, \delta = 1$ so that each $S_k \sim \text{Lévy}(0, \gamma_k)$, yielding $R(\mathbf{s}_j) \sim \text{Lévy}(0, \bar{\gamma}_j)$ with $\bar{\gamma}_j = \left[\sum_{k=1}^K \sqrt{B_k(\mathbf{s}_j; l)} \gamma_k \right]^2$.

In summary, Model (M1) achieves AD/AI flexibility through a global scaling mechanism but imposes a single stationary dependence class. Models (M2) and (M3) introduce spatially varying scaling surfaces, enabling short-range AD with long-range AI. Model (M4) further generalizes this framework by allowing the local dependence class to be *either* AD *or* AI and to vary across space through $\phi(\mathbf{s})$. Table 3.1 summarizes the marginal tail index α^* and describes the joint tail behavior through the indices χ and η of Models (M1)–(M4).

Despite their modeling appeal, these transformed Gaussian scale-mixture constructions pose substantial computational challenges under the peaks-over-threshold framework. Likelihood-based inference for multivariate threshold exceedances is typically based on a censored likelihood, and for transformed Gaussian scale-mixture models including (M1), (M3), and (M4), this likelihood requires repeated evaluation of high-dimensional Gaussian distribution functions. This quickly becomes infeasible once the number of spatial locations is just moderately large. For Model (M2), the joint likelihood is not available in closed form even in low dimensions, and Majumder et al. (2024) therefore relied on a combination of Vecchia-type approximations and neural-network emulators for approximate Bayesian inference. These computational barriers motivate the scalable

Table 3.1: Summary of marginal and bivariate tail behavior of (s_i, s_j) for random scale-mixture models (M1)–(M4). Let $W_i := g\{Z_i\}$, $\phi_i \in (0, \infty)$ be the dependence parameter, and $\eta_{ij}^Z \in (1/2, 1)$ be the residual tail dependence coefficient induced by the latent $\{Z(s)\}$ process. For (M3), $\gamma > 0$ is the Pareto tail index for $R(s)$. For (M4), α is the Stable-index. For (M3) and (M4), $\mathcal{K}_i := \{k : B_k(s_i; l) > 0\}$, the set of knots that are active for location s_i . See Section 3.1.2 for model construction details.

Model	$\overline{F}_{X^*} \in \text{RV}_{-\alpha^*}$	Joint Tail Indices
(M1)	$\alpha^* = \min\{\frac{1}{\phi}, \frac{1}{1-\phi}\}$	$\chi_{ij} = \begin{cases} \frac{2\phi-1}{\phi} \mathbb{E}[\min\{W_i, W_j\}^{(1-\phi)/\phi}] & \text{if } \delta > \frac{1}{2}, \\ 0 & \text{if } \delta \leq \frac{1}{2}. \end{cases}$ $\eta_{ij} = \begin{cases} 1, & \text{if } \delta \geq \frac{1}{2}, \\ \max\{\delta/(1-\delta), \eta_{ij}^Z\}, & \text{if } \delta < \frac{1}{2}. \end{cases}$
(M2)	$\alpha^* = \min\{\frac{1}{\phi}, \frac{1}{1-\phi}\}$	Same as (M1) (<i>novel</i> analytical results in Theorem 7)
(M3)	$\alpha^* = \gamma$	$\chi_{ij} = \sum_{k \in \mathcal{K}_i \cap \mathcal{K}_j} \left[B_k(s_i; l) \overline{T}_{\gamma+1}\{a_{ij}^{(k)}\} + B_k(s_j; l) \overline{T}_{\gamma+1}\{a_{ji}^{(k)}\} \right]$ where $a_{ij}^{(k)} = \sqrt{\gamma+1} \frac{B_k(s_i; l)^{1/\gamma} B_k(s_j; l)^{-1/\gamma - \rho(s_i, s_j)}}{\sqrt{1 - \rho(s_i, s_j)^2}}$, and $\overline{T}_{\text{df}}$ denotes Student's t survival function with df degrees of freedom $\eta_{ij} = \begin{cases} 1, & \text{if } \mathcal{K}_i \cap \mathcal{K}_j \neq \emptyset, \\ 1/2, & \text{if } \mathcal{K}_i \cap \mathcal{K}_j = \emptyset. \end{cases}$
(M4)	$\alpha_j^* = \min\{\frac{\alpha}{\phi_j}, 1\}$	$\chi_{ij} = \begin{cases} \mathbb{E} \left[\min \left\{ \frac{W_i^{\alpha/\phi_i}}{\mathbb{E}[W_i^{\alpha/\phi_i}]}, \frac{W_j^{\alpha/\phi_j}}{\mathbb{E}[W_j^{\alpha/\phi_j}]} \right\} \right] \sum_{k \in \mathcal{K}_i \cap \mathcal{K}_j} v_{k, \wedge}, & \text{if } \alpha < \phi_i < \phi_j, \\ 0 & \text{otherwise,} \end{cases}$ where $v_{k, \wedge} = \min\{v_{ki}, v_{kj}\}$ with v_{ki} and v_{kj} defined in Equation 3.11. If $\mathcal{K}_i \cap \mathcal{K}_j \neq \emptyset$, $\begin{cases} \eta_{ij} = 1 & \text{if } \alpha < \phi_i < \phi_j, \\ \eta_{ij} \in [\max(\eta_{ij}^Z, \frac{\phi_i}{\alpha}), \max(\eta_{ij}^Z, \frac{\phi_j}{\alpha})] & \text{if } \phi_i < \phi_j < \alpha, \\ 1/\eta_{ij} \in [\frac{\min\{\phi_i + \phi_j, 2\alpha\eta_{ij}^Z\} - \phi_j + \alpha}{2\alpha\eta_{ij}^Z}, 2 - \frac{\phi_i}{\alpha}] & \text{if } \phi_i < \alpha < \phi_j. \end{cases}$ If $\mathcal{K}_i \cap \mathcal{K}_j = \emptyset$, $\begin{cases} \eta_{ij} \in [\max(\frac{1}{2}, \frac{\alpha\eta_{ij}^Z}{\phi_j}), \max(\frac{1}{2}, \frac{\alpha}{\phi_i})] & \text{if } \alpha < \phi_i < \phi_j, \\ \eta_{ij} \in [\eta_{ij}^Z, \max(\eta_{ij}^Z, \frac{\phi_j}{\alpha})] & \text{if } \phi_i < \phi_j < \alpha, \\ 1/\eta_{ij} \in [\frac{\min\{\phi_j, 2\alpha\eta_{ij}^Z\} + \alpha}{2\alpha\eta_{ij}^Z}, 2] & \text{if } \phi_i < \alpha < \phi_j. \end{cases}$

approach developed in the next sections. We first review the censored likelihood formulation and identify the key bottleneck in Section 3.1.3.

3.1.3 The Censored Likelihood

In multivariate POT, inference is driven by exceedances above a high threshold, but the data also contain many sub-threshold observations. Treating observations below the threshold as censored provides an effective compromise between, on one hand, treating them as fully observed and thereby allowing them to bias estimates of properties specific to the tail, and on the other hand, removing them entirely and ignoring any information they provide (Zhang et al., 2022b; Huser and Genton, 2016). Specifically, consider observations at D locations $\mathbf{s}_1, \dots, \mathbf{s}_D$, the joint distribution of $\mathbf{X}^*$ defined in Equation 3.3 can be obtained by conditioning on R as

$$F_{\mathbf{X}^*}(\mathbf{x}^*) = \int_{\mathcal{R}} \Phi_D\{g_\phi^{-1}(\mathbf{x}^*/r^\phi); \Sigma_\rho\} f_R(r) dr, \quad (3.4)$$

where Φ_D denotes the D -variate Gaussian CDF with mean $\mathbf{0}$ and covariance Σ_ρ .

Let $\mathcal{E}$ index exceedances and $\mathcal{C}$ censored components, and partition Σ_ρ accordingly. The joint censored likelihood, obtained by differentiating Equation 3.4 with respect to $\mathbf{x}_\mathcal{E}^*$, is

$$\begin{aligned} L(\mathbf{x}^*) = \int_{\mathcal{R}} & \Phi_{|\mathcal{C}|}(\mathbf{z}_\mathcal{C} - \Sigma_{\mathcal{C}\mathcal{E}}\Sigma_{\mathcal{E}\mathcal{E}}^{-1}\mathbf{z}_\mathcal{E}; \Sigma_{\mathcal{C}|\mathcal{E}}) \phi_{|\mathcal{E}|}(\mathbf{z}_\mathcal{E}; \Sigma_{\mathcal{E}\mathcal{E}}) \\ & \times \left\{ \prod_{i \in \mathcal{E}} (g_\phi^{-1})'(x_i^*/r^\phi) \right\} r^{-\phi|\mathcal{E}|} f_R(r) dr, \end{aligned} \quad (3.5)$$

where $\mathbf{z}_\mathcal{A} = g_\phi^{-1}(\mathbf{x}_\mathcal{A}^*/r^\phi)$ and $\Sigma_{\mathcal{C}|\mathcal{E}} = \Sigma_{\mathcal{C}\mathcal{C}} - \Sigma_{\mathcal{C}\mathcal{E}}\Sigma_{\mathcal{E}\mathcal{E}}^{-1}\Sigma_{\mathcal{E}\mathcal{C}}$.

The key difficulty in evaluating Equation 3.5 is the term $\Phi_{|\mathcal{C}|}(\cdot)$, a $|\mathcal{C}|$ -dimensional Gaussian distribution function. As a result, for each time replicate, the likelihood evaluation requires repeated computation of Gaussian distribution functions in dimensions that quickly become prohibitive once $|\mathcal{C}|$ reaches even the low tens (Huser et al., 2025; Zhang et al., 2022b). The burden is further exacerbated for models with spatially varying scaling mechanisms (e.g., $R(\mathbf{s})$ and/or $\phi(\mathbf{s})$ in Models (M2), (M3) and (M4)), which introduce additional latent structure that must be repeatedly integrated or sampled within each likelihood evaluation.

This Gaussian CDF bottleneck is a major obstacle that motivates more efficient fitting of flexible Gaussian scale-mixture models to high-dimensional threshold exceedance data. Zhang et al. (2022b) circumvent this problem by supplementing $X^*(\mathbf{s})$ in Equation 3.3 with an additive Gaussian nugget, as $X(\mathbf{s}) = X^*(\mathbf{s}) + \varepsilon(\mathbf{s}_i)$, with $\varepsilon(\mathbf{s}_i) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$ for $\mathbf{s}_1, \dots, \mathbf{s}_D$. Then, while the observations themselves are left-censored below the threshold, the (now-latent) smooth process $X^*(\mathbf{s})$ is not. The resulting likelihood treats the highly-multivariate $\{X^*(\mathbf{s})\}_{i=1}^D$ as uncensored while, conditional on $\{X^*(\mathbf{s})\}_{i=1}^D$, the univariate nugget terms $\{\varepsilon(\mathbf{s})\}_{i=1}^D$ are censored. Consequently, the full likelihood requires multivariate Gaussian densities but only univariate CDFs. This eliminates the intractable CDF calculation but the addition operation introduces a *convolution* which is unavailable in closed form and must be computed numerically. For models like (M3)–(M4) with spatially varying parameters, the numerical integral must be computed at every observation location for each MCMC iteration. This renders model-fitting infeasible for more complicated models, even though the intractable high-dimensional Gaussian CDF is bypassed.

In the following sections, we introduce a *multiplicative* log-Laplace nugget that similarly yields conditional independence in the censored likelihood. This also eliminates multivariate Gaussian distribution function evaluations, but unlike the additive nugget of Zhang et al. (2022b), it gives closed-form marginal density and distribution functions, eliminating the need for any numerical integration. This enables scalable Bayesian inference for high-dimensional threshold exceedance data while, as we show below, preserving the extremal dependence properties of the underlying smooth process.

3.2 Model

3.2.1 Construction

Let $X^*(\mathbf{s})$ denote the latent smooth process that captures spatial extremal dependence. We define the modified model as

$$X(\mathbf{s}_i) = \epsilon(\mathbf{s}_i) X^*(\mathbf{s}_i) = \epsilon(\mathbf{s}_i) \cdot R(\mathbf{s}_i)^{\phi(\mathbf{s}_i)} g_\phi\{Z(\mathbf{s}_i)\}, \quad i = 1, \dots, D, \quad (3.6)$$

where the nugget terms $\epsilon(\mathbf{s}_i)$ are independent and identically distributed $\log\text{-Laplace}(0, 1/\alpha_0)$ random variables across sites, concentrated around 1.

3.2.2 Computational Implications

3.2.2.1 Conditional Independence

Under the nuggetted model Equation 3.6, the components of $\mathbf{X} = (X_1, \dots, X_D)^\top$ are independent across sites, conditional on the latent smooth process values $\mathbf{X}^* = (X_1^*, \dots, X_D^*)^\top$, because $X_i = \epsilon_i X_i^*$ with independent nugget variables $\{\epsilon_i\}_{i=1}^D$. This conditional independence dramatically simplifies the censored likelihood. Specifically, for each component i ,

$$\Pr\{X_i \leq x_{0i} \mid X_i^*\} = F_\epsilon\left(\frac{x_{0i}}{X_i^*}\right), \quad f_{X|X^*}(x_i \mid X_i^*) = f_\epsilon\left(\frac{x_i}{X_i^*}\right) \frac{1}{X_i^*},$$

where x_{0i} denotes the censoring threshold on the X -scale. Conditional on $\mathbf{X}^*$ and other model parameters, the joint censored likelihood factorizes into a product of one-dimensional terms,

$$\prod_{i \in \mathcal{C}} F_\epsilon\left(\frac{x_{0i}}{X_i^*}\right) \times \prod_{i \in \mathcal{E}} f_\epsilon\left(\frac{x_i}{X_i^*}\right) \frac{1}{X_i^*}.$$

As a result, the nuggetted construction eliminates the need for repeated evaluation of the multivariate Gaussian distribution function $\Phi_{|\mathcal{C}|}(\cdot)$ in censored likelihood computation. The remaining computational cost is the marginal evaluations of the chosen scale-mixture construction under the nugget multiplication, which we address next.

3.2.2.2 Marginal Tractability

While conditional independence removes the high-dimensional Gaussian CDF bottleneck, it does not by itself guarantee that the resulting likelihood is numerically efficient. In particular, a nugget can eliminate multivariate dependence at the likelihood level while simultaneously making marginal evaluation more expensive, as in Zhang et al. (2022b). This burden is especially severe

in settings where marginal distributions vary across space, rendering computations required for model fitting infeasible.

In contrast, the multiplicative log-Laplace nugget in Equation 3.6 avoids introducing an additional numerical integration layer. For a broad subclass of random scale-mixture models, the nugget can be absorbed into the latent scale variable, so marginal evaluation retains the same functional form as the base model and often remains available in closed form. We summarize this property for Models (M1)–(M4) below.

(M1) The base model $X^*(s)$ admits a closed-form marginal distribution,

$$F_{X^*}(x) = 1 - \frac{\phi}{2\phi - 1} x^{-1/\phi} + \frac{1 - \phi}{2\phi - 1} x^{-1/(1-\phi)}.$$

With the log-Laplace nugget, $X(s)$ also admits a closed-form marginal distribution,

$$F_X(x) = 1 - \frac{\phi}{2\phi - 1} x^{-1/\phi} A_{1/\phi}(x) + \frac{1 - \phi}{2\phi - 1} x^{-1/(1-\phi)} A_{1/(1-\phi)}(x) - \frac{1}{2} x^{-\alpha_0},$$

where

$$A_q(x) = \frac{\alpha_0}{2} \left(\frac{1}{q + \alpha_0} + \frac{x^{q-\alpha_0} - 1}{q - \alpha_0} \right) \mathbb{1}(\alpha_0 \neq q) + \left(\frac{1}{4} + \frac{\alpha_0}{2} \log x \right) \mathbb{1}(\alpha_0 = q).$$

(M2) After a monotone marginal transformation, the model of Majumder et al. (2024) can be written in the marginally equivalent scale-mixture form as Model (M1). Consequently, the nuggeted version of Model (M2) inherits the same closed-form marginal tractability as Model (M1).

(M3) The model of Hazra et al. (2025) extends from the stationary construction of Huser et al. (2017), in which the marginal distribution is known only up to a one-dimensional integral,

$$F_{X^*}(x) = \int_0^\infty \Phi(x/r) f_R(r) dr, \quad (3.7)$$

where $\Phi(\cdot)$ denotes the standard Gaussian CDF. Under the $\text{Pareto}(\gamma)$ specification of R used for analysis by Hazra et al. (2025), incorporating the log-Laplace nugget retains the same one-dimensional integral complexity required for the marginal evaluation with no additional numerical integration layer introduced:

$$F_X(x) = \int_0^\infty \Phi(x/s) f_{\tilde{R}}(s) ds, \quad (3.8)$$

where $\tilde{R} = \epsilon R$, and $f_{\tilde{R}}$ is available in closed form as

$$f_{\tilde{R}}(s) = \frac{\alpha_0 \gamma}{2} s^{-(\alpha_0+1)} \left[\frac{s^{\alpha_0-\gamma} - 1}{\alpha_0 - \gamma} \mathbb{1}(\alpha_0 \neq \gamma) + \log(s) \mathbb{1}(\alpha_0 = \gamma) \right] + \frac{\alpha_0 \gamma}{2(\gamma + \alpha_0)} s^{-(\gamma+1)}.$$

The model of Hazra et al. (2025) replaces the global $\{R(s)\}$ with a low-rank representation through deterministic weighted sum of finitely many latent scalars R_k^* (see Section 3.1.2), so marginal evaluation at each site s retains the one-dimensional integral form and the nugget does not introduce any additional numerical integration.

(M4) For the construction of Shi et al. (2026), using the standard-Pareto link $g(\cdot)$ as in Section 3.1.2, the smooth process $X^*(s)$ admits a closed-form marginal distribution expressed in terms of incomplete gamma functions:

$$F_{X_j^*}(x) = 1 - \sqrt{\frac{1}{\pi}} \gamma \left(\frac{1}{2}, \frac{\bar{\gamma}_j}{2x^{1/\phi_j}} \right) - x^{-1} \sqrt{\frac{1}{\pi}} \left(\frac{\bar{\gamma}_j}{2} \right)^{\phi_j} \Gamma \left(\frac{1}{2} - \phi_j, \frac{\bar{\gamma}_j}{2x^{1/\phi_j}} \right),$$

where $\gamma(\cdot, \cdot)$ and $\Gamma(\cdot, \cdot)$ denote the lower and upper incomplete gamma functions. Under the log-Laplace nugget, the marginal survival function of X_j remains available in closed form, with marginal distribution function

$$\begin{aligned} F_{X_j}(x) = 1 - \sqrt{\frac{1}{\pi}} & \left[\gamma \left(\frac{1}{2}, \lambda_j(x) \right) + \frac{\alpha_0^2}{\alpha_0^2 - 1} \left(\frac{\bar{\gamma}_j}{2} \right)^{\phi_j} \frac{1}{x} \Gamma \left(\frac{1}{2} - \phi_j, \lambda_j(x) \right) \right. \\ & - \frac{1}{2(\alpha_0 + 1)} x^{\alpha_0} \left(\frac{\bar{\gamma}_j}{2} \right)^{-\phi_j \alpha_0} \gamma \left(\frac{1}{2} + \phi_j \alpha_0, \lambda_j(x) \right) \\ & \left. - \frac{1}{2(\alpha_0 - 1)} x^{-\alpha_0} \left(\frac{\bar{\gamma}_j}{2} \right)^{\phi_j \alpha_0} \Gamma \left(\frac{1}{2} - \phi_j \alpha_0, \lambda_j(x) \right) \right], \end{aligned}$$

where $\lambda_j(x) = \bar{\gamma}_j / (2x^{1/\phi_j})$.

3.2.3 Tail Properties

3.2.3.1 Marginal tail equivalence

We first show that the nugget can be chosen to preserve the marginal tail of the latent smooth process. Specifically, it will have negligible impact on the tail as long as it is lighter-tailed than $X^*(\mathbf{s})$, in the sense that it possesses sufficiently high moments. Because $X^*(\mathbf{s})$ will ultimately be re-scaled to have GPD margins during inference, the marginal tail properties that we derive here are interesting only insofar as they allow us to deduce the tail dependence properties that we describe in Section 3.2.3.2.

Theorem 4 (Marginal Tail Equivalence). Assume $X^*(\mathbf{s})$ has a regularly varying upper tail with index $\alpha^*(\mathbf{s})$; that is, for slowly varying $\mathcal{L}_{\mathbf{s}}$,

$$\Pr\{X^*(\mathbf{s}) > x\} = \mathcal{L}_{\mathbf{s}}(x) x^{-\alpha^*(\mathbf{s})}, \quad x \rightarrow \infty. \quad (3.9)$$

If $\alpha_0 > \sup_{\mathbf{s} \in \mathcal{S}} \alpha^*(\mathbf{s})$, then $X(\mathbf{s}) = \epsilon(\mathbf{s})X^*(\mathbf{s})$ is marginally tail equivalent to $X^*(\mathbf{s})$:

$$\Pr\{X(\mathbf{s}) > x\} \sim \mathbb{E}\{\epsilon(\mathbf{s})^{\alpha^*(\mathbf{s})}\} \Pr\{X^*(\mathbf{s}) > x\}, \quad x \rightarrow \infty.$$

Corollary 5 (Marginal tail equivalence for models (M1)–(M4), and sufficient conditions on α_0). For each model in Section 3.1.2, the smooth process $X^*(\mathbf{s})$ has a regularly varying upper tail with index $\alpha^*(\mathbf{s})$ as in Equation 3.9. Therefore, Theorem 4 yields marginal tail equivalence whenever $\alpha_0 > \sup_{\mathbf{s} \in \mathcal{S}} \alpha^*(\mathbf{s})$. In particular:

(M1) By the marginal Pareto construction, $X^*(\mathbf{s})$ has a regularly varying tail with index

$$\alpha^* = \min\{1/\phi, 1/(1 - \phi)\},$$

Thus, it suffices to take $\alpha_0 > 2$.

- (M2) After a monotone marginal transformation, Model (M2) can be written in the equivalent form as (M1), and therefore the same condition applies and it suffices to take $\alpha_0 > 2$.
- (M3) Under the Pareto(γ) specification of Huser et al. (2017), $X^*(\mathbf{s})$ has a regularly varying upper tail with index $\alpha^* = \gamma$, since $\Pr\{RW > x\} = \frac{1}{2} \Pr\{R|W| > x\}$ given the symmetry of W and Breiman (1965, Lemma). In Hazra et al. (2025), a finite nonnegative weighted sum of the independent RV_γ variables is used to construct the spatially varying $R(\mathbf{s})$, and hence $X^*(\mathbf{s})$ retains tail index γ . Thus, it suffices to take $\alpha_0 > \gamma$.
- (M4) By construction and Proposition 1 of Shi et al. (2026), $X^*(\mathbf{s})$ has regularly varying upper tail with tail index

$$\alpha^*(\mathbf{s}) = \min \left[\frac{\alpha}{\phi(\mathbf{s})}, 1 \right].$$

Thus, it suffices to take $\alpha_0 > 1$.

3.2.3.2 Joint tail dependence

We next show that the multiplicative log-Laplace nugget preserves the extremal dependence properties of the underlying smooth process. Specifically, for any two sites $(\mathbf{s}_i, \mathbf{s}_j)$, we compare the joint upper-tail decay of $(X(\mathbf{s}_i), X(\mathbf{s}_j))$ and $(X^*(\mathbf{s}_i), X^*(\mathbf{s}_j))$ through the coefficients (χ_{ij}, η_{ij}) and $(\chi_{ij}^*, \eta_{ij}^*)$, respectively, defined as in Equation 3.1 and Equation 3.2. Under mild regular variation conditions on the bivariate tail of the smooth process and the corresponding moment condition on the nugget, the nugget leaves η_{ij} unchanged and modifies χ_{ij} only up to multiplicative constants. We formalize this in Theorem 6, and then validate its conditions for Models (M1)–(M4) in a corollary, using both existing tail dependence results and a new proposition establishing the joint tail decay for Model (M2), which was previously only assessed numerically in Majumder et al. (2024).

Theorem 6 (Joint Tail Equivalence). Consider a random scale-mixture model and suppose its underlying $\{Z(\mathbf{s})\}$ process is asymptotically independent but positively correlated ($\eta_{ij}^Z \in (1/2, 1)$). If X_i^* , X_j^* , and $\min(X_i^*, X_j^*)$ all have regularly varying tails and satisfy $\alpha_0 \geq 2 \sup_{\mathbf{s} \in \mathcal{S}} \alpha^*(\mathbf{s})$ for

any two sites s_i and s_j , then

$$\eta_{ij} = \eta_{ij}^* \quad \text{and} \quad \chi_{ij} \in [c_{ij} \chi_{ij}^*, C_{ij} \chi_{ij}^*],$$

where

$$c_{ij} = \mathbb{E} \left[\min \left(\frac{\epsilon_i^{\alpha_i^*}}{\mathbb{E}[\epsilon_i^{\alpha_i^*}]}, \frac{\epsilon_j^{\alpha_j^*}}{\mathbb{E}[\epsilon_j^{\alpha_j^*}]} \right) \right] \quad \text{and} \quad C_{ij} = \mathbb{E} \left[\max \left(\frac{\epsilon_i^{\alpha_i^*}}{\mathbb{E}[\epsilon_i^{\alpha_i^*}]}, \frac{\epsilon_j^{\alpha_j^*}}{\mathbb{E}[\epsilon_j^{\alpha_j^*}]} \right) \right]. \quad (3.10)$$

Explicit expressions of the expectations in c_{ij} and C_{ij} are provided in Appendix B.2.1.

Proposition 7 (Joint Distribution of (M2)). Consider the random scale-mixture representation (M2). For any two sites s_i, s_j , the joint exceedance probability is regularly varying in x ; as $x \rightarrow \infty$,

$$\Pr(X_i > x, X_j > x) = \begin{cases} L(x) x^{-1/\{(1-\phi)\eta_{ij}^Z\}}, & \eta_{ij}^Z \geq \frac{\phi}{1-\phi}, \\ \mathbb{E}[\min(W_i, W_j)^{(1-\phi)/\phi}] x^{-1/\phi} \{1 + o(1)\}, & \eta_{ij}^Z < \frac{\phi}{1-\phi}. \end{cases}$$

where η_{ij}^Z is the residual tail dependence coefficient induced by the Gaussian copula (Ledford and Tawn, 1996).

Corollary 8 (Joint tail equivalence for models (M1)–(M4)). For each random scale-mixture model in Section 3.1.2, assume the underlying $Z(s)$ is asymptotically independent but positively correlated. Then Theorem 6 applies. In particular:

- (M1) The regularly varying tail of $\min(X_i^*, X_j^*)$ and the corresponding joint tail behavior is given in Proposition 1 and Corollary 1 of Huser and Wadsworth (2019). Since $\sup_{s \in \mathcal{S}} \alpha^*(s) = 2$, it suffices to take $\alpha_0 \geq 4$. Thus the dependence classes and tail coefficients characterized in Huser and Wadsworth (2019) are preserved under the multiplicative log-Laplace nugget.
- (M2) Proposition 7 establishes the regularly varying tail of $\min(X_i^*, X_j^*)$ for Model (M2). Since $\sup_{s \in \mathcal{S}} \alpha^*(s) = 2$, it suffices to take $\alpha_0 \geq 4$.

- (M3) Hazra et al. (2025) established the joint tail behavior in Proposition 3 of their main paper and the Appendix, showing that (X_i, X_j) is $\text{MRV}(\gamma, a(\cdot), \nu_Z(\cdot))$ with $a(t) = t^{1/\gamma}$ and $\nu_Z(\mathcal{A}_r := [r_1, \infty) \times [r_2, \infty)) = \sum_{k=1}^K \mathbb{E} \left[\min \left\{ \frac{B_k(\mathbf{s}_i; l) Z_+^\gamma(\mathbf{s}_i)}{r_1^\gamma}, \frac{B_k(\mathbf{s}_j; l) Z_+^\gamma(\mathbf{s}_j)}{r_2^\gamma} \right\} \right]$. Since $\sup_{\mathbf{s} \in \mathcal{S}} \alpha^*(\mathbf{s}) = \gamma$, it suffices to take $\alpha_0 \geq 2\gamma$. Then, Theorem 6 gives joint tail equivalence under the log-Laplace nugget multiplication.
- (M4) Shi et al. (2026) established the regular variation of $\min(X_i^*, X_j^*)$ in Theorem 3 of their paper and Section A.4 of the Appendix. Since $\sup_{\mathbf{s} \in \mathcal{S}} \alpha^*(\mathbf{s}) = 1$, it suffices to take $\alpha_0 \geq 2$. Hence, Theorem 6 leads to the same conclusion of joint tail equivalence.

3.2.4 Numerical Illustration

We illustrate the theoretical results of Theorem 6 in the context of the most flexible available model, (M4), by providing Theorem 9 and an empirical example in Illustration 2. Theorem 6 and Theorem 8 imply that the multiplicative log-Laplace nugget leaves the residual tail dependence coefficient unchanged, i.e., $\eta_{ij} = \eta_{ij}^*$ and modifies χ_{ij} by multiplicative constants. Below, we summarize the expression for χ_{ij} , and refer to Table 3.1 and Theorem 3 of Shi et al. (2026) for the details of η_{ij} .

Corollary 9. Under the model of Shi et al. (2026) in (M4), consider two locations $\mathbf{s}_i$ and $\mathbf{s}_j$. Let $\mathcal{K}_i := \{k : B_k(\mathbf{s}_i; l) > 0\}$ and $\mathcal{K}_j := \{k : B_k(\mathbf{s}_j; l) > 0\}$ denote the sets of basis indices whose compactly supported kernels overlap $\mathbf{s}_i$ and $\mathbf{s}_j$, respectively. Define

$$v_{ki} = \frac{\{B_k(\mathbf{s}_i; l) \gamma_k\}^\alpha}{\sum_{k' \in \mathcal{K}_i} \{B_{k'}(\mathbf{s}_i; l) \gamma_{k'}\}^\alpha}, \quad \alpha_i^* = \min \left(\frac{\alpha}{\phi_i}, 1 \right). \quad (3.11)$$

Then

- (a) If $\mathcal{K}_i \cap \mathcal{K}_j \neq \emptyset$ and $\alpha < \phi_i < \phi_j$, then $(X_i, X_j)^\top$ is asymptotically dependent with $\eta_{ij} = 1$ and

$$\chi_{ij} \in \left[c_{ij} \chi_{ij}^*, C_{ij} \chi_{ij}^* \right],$$

where the nugget constants c_{ij} and C_{ij} are defined in Equation 3.10 and

$$\chi_{ij}^* = \mathbb{E} \left[\min \left\{ \frac{W_i^{\alpha/\phi_i}}{\mathbb{E}[W_i^{\alpha/\phi_i}]}, \frac{W_j^{\alpha/\phi_j}}{\mathbb{E}[W_j^{\alpha/\phi_j}]} \right\} \right] \sum_{k=1}^K \min(v_{ki}, v_{kj}).$$

where $W_i = g(Z_i)$.

- (b) Otherwise, (X_i, X_j) is asymptotically independent with $\chi_{ij} = 0$, and η_{ij} is inherited from $X^*(\mathbf{s})$, as shown in Table 3.1 and Theorem 3 of Shi et al. (2026).

Illustration 2 (Empirical evaluations of the bounds in Theorem 9). We simulate $N = 300,000,000$ draws from Model (M4) on the domain $\mathcal{S} = [0, 10] \times [0, 10]$ using a $\{\phi(\mathbf{s}) : \mathbf{s} \in \mathcal{S}\}$ surface as shown in Figure 3.1. For each simulation, we generate independent Stable variables with $\alpha = 0.5$ at 9 knots on a uniform grid and combine them using Wendland basis functions centered at the knots (Wendland, 1995). We generate independent log-Laplace nugget terms with scale parameters $\alpha_0 = 2.0, 5.0$, and 10.0 . We then empirically estimate the $\chi_{ij}(u)$ and $\eta_{ij}(u)$ functions using the N independent simulations of (X_i, X_j) .

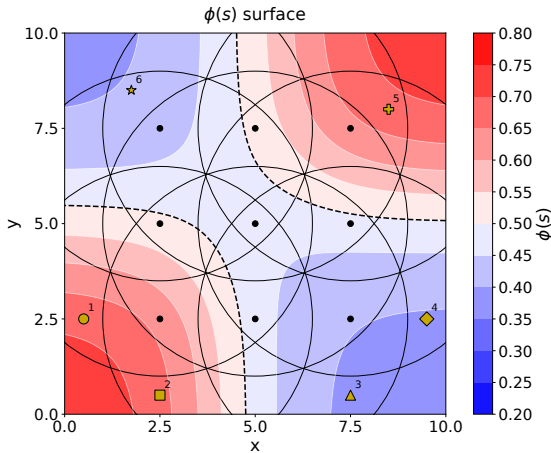


Figure 3.1: A $\phi(\mathbf{s})$ surface on $[0, 10]^2$, in which the dashed line marks the transition between local AI and AD. The points with ‘•’ are centers for the Wendland basis functions. The points with other signs/marker-styles are chosen sample points that we use to illustrate the dependence properties in Theorem 9.

Figure 3.2 shows that empirical estimates of χ_{ij} and η_{ij} fall within the theoretical bounds, for various values of the log-Laplace scale parameter α_0 . Figure 3.2a considers sample points 1 and 2, which share a common Wendland kernel and satisfy $\alpha < \phi_2 < \phi_1$, so the pair is asymptotically

dependent. Figure 3.2b considers points 3 and 4, which also share a kernel but satisfy $\phi_3 < \phi_4 < \alpha$, so the pair is asymptotically independent. Figure 3.2c considers points 4 and 5, which share a kernel yet satisfy $\phi_4 < \alpha < \phi_5$, and therefore remain asymptotically independent. Finally, Figure 3.2d considers points 1 and 5, which do not share any common Wendland kernel; consistent with the long-range case, the pair is asymptotically independent even though $\phi_5 > \phi_1 > \alpha$.

3.3 Bayesian Inference

We define a Bayesian hierarchical model based on Equation Equation 3.6 and use an MCMC algorithm to fit to the data. The dependence model Equation 3.6 is displayed again here for convenience, now with each time replicate denoted with a subscript $t = 1, \dots, T$:

$$X_t(\mathbf{s}) = \epsilon_t(\mathbf{s})X_t^*(\mathbf{s}) = \epsilon_t(\mathbf{s}) \cdot R_t(\mathbf{s})^{\phi(\mathbf{s})}g(Z_t(\mathbf{s})).$$

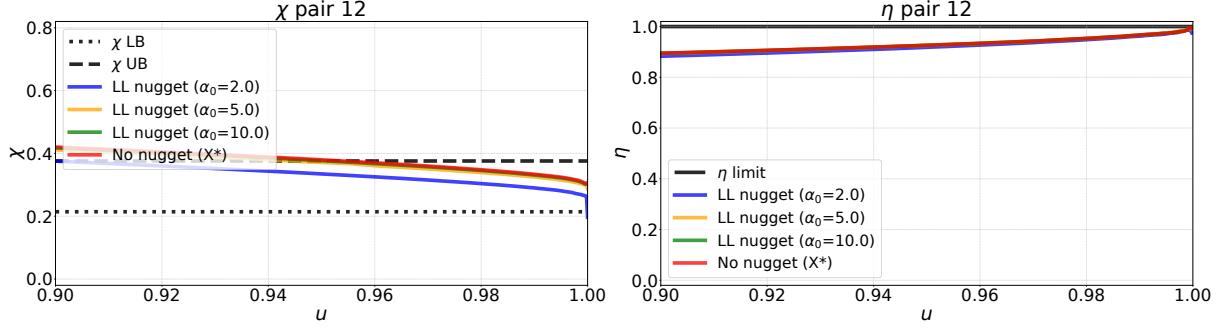
3.3.1 Hierarchical Model and Computation

We connect the dependence model in Equation 3.6 to a GPD marginal specification through a probability integral transform. For each replicate $t = 1, \dots, T$ and site $j = 1, \dots, D$, let Y_{tj} denote the observations, which are independent across t . Fix the exceedance probability $p \in (0, 1)$ and let y_{0j} denote a high threshold at site j . Define

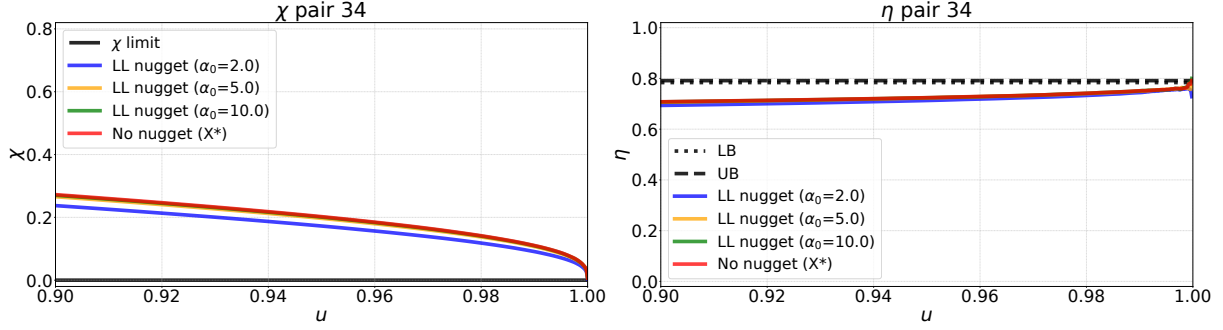
$$T(Y_{tj}) = \begin{cases} y_{0j}, & Y_{tj} \leq y_{0j}, \\ Y_{tj}, & Y_{tj} > y_{0j}. \end{cases}$$

For exceedances, assume

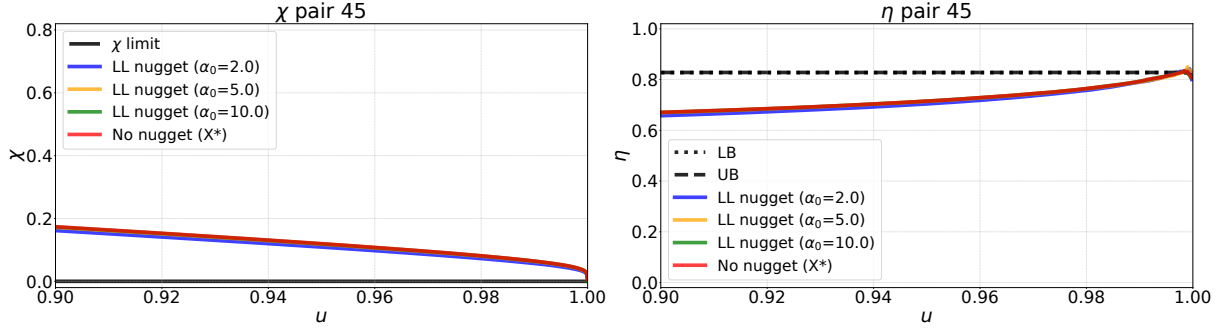
$$Y_{tj} - y_{0j} \mid (Y_{tj} > y_{0j}) \sim \text{GPD}(\sigma_j, \xi_j),$$



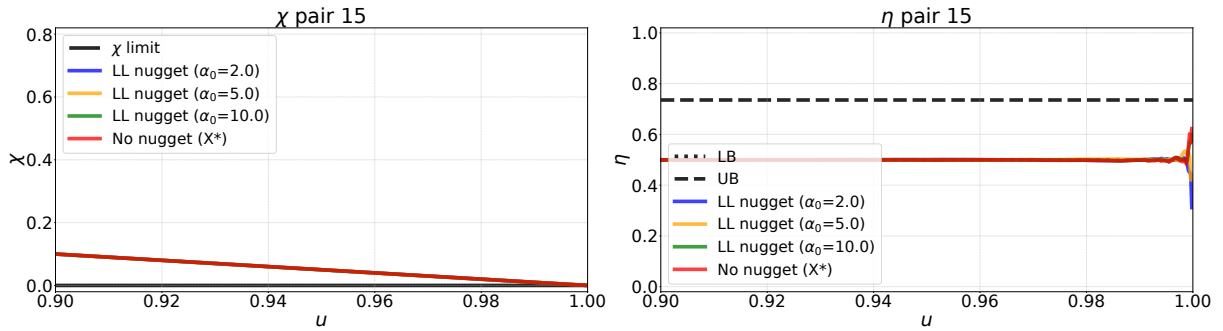
(a) Theorem 9 (a) Asymptotic Dependence: χ_{12} and η_{12}



(b) Theorem 9 (b) Asymptotic independence: χ_{34} and η_{34}



(c) Theorem 9 (b) Asymptotic independence: χ_{45} and η_{45}



(d) Theorem 9 (b) Asymptotic independence: χ_{15} and η_{15}

Figure 3.2: Empirical estimates of the dependence coefficients $\chi_{ij}(u)$ and $\eta_{ij}(u)$ (Leo R Belzile, 2023) between sample points i and j in Figure 3.1. The black solid, dotted, and dashed lines mark the theoretical limit, upper, and lower bounds, respectively.

with tail CDF

$$H_{y_{0j}}(y \mid \sigma_j, \xi_j) = 1 - \left(1 + \xi_j \frac{y - y_{0j}}{\sigma_j}\right)^{-1/\xi_j}, \quad 1 + \xi_j(y - y_{0j})/\sigma_j > 0.$$

Hence the censored marginal CDF on the Y -scale is

$$G_j(y) = \begin{cases} p, & y \leq y_{0j}, \\ p + (1 - p) H_{y_{0j}}(y \mid \sigma_j, \xi_j), & y > y_{0j}. \end{cases}$$

Let $x_{0j} := F_{X_j}^{-1}(p)$. We map observations to the latent X -scale by

$$X_{tj} = \begin{cases} x_{0j}, & Y_{tj} \leq y_{0j}, \\ F_{X_j}^{-1}\{G_j(Y_{tj})\}, & Y_{tj} > y_{0j}. \end{cases}$$

Define $\mathcal{C}_t = \{j : Y_{tj} \leq y_{0j}\}$ and $\mathcal{E}_t = \{j : Y_{tj} > y_{0j}\}$. Under the multiplicative nugget model, components are conditionally independent across sites given $\mathbf{X}_t^*$, so the censored likelihood factorizes as

$$L_t(\mathbf{Y}_t \mid \mathbf{S}_t, \mathbf{Z}_t, \boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\rho}, l, \alpha_0, \boldsymbol{\sigma}, \boldsymbol{\xi}) = \prod_{j \in \mathcal{C}_t} F_\epsilon\left(\frac{x_{0j}}{X_{tj}^*}\right) \times \prod_{j \in \mathcal{E}_t} f_{X \mid X^*}(x_{tj} \mid X_{tj}^*) \left| \frac{dX_{tj}}{dY_{tj}} \right|$$

for each $t = 1, \dots, T$. Under the temporal independence assumption, the full joint likelihood is simply the product of the likelihood for each time replicate.

Following Shi et al. (2026), we fix $\alpha = 1/2$, which simplifies the analytical formulation. Likewise, we represent $\phi(\mathbf{s})$ and $\rho(\mathbf{s})$ using Gaussian kernel basis functions centered at the knots. We model the latent Gaussian process $Z_t(\mathbf{s})$ with a locally isotropic *nonstationary* Matérn covariance (Paciorek and Schervish, 2006; Risser and Calder, 2015). We implement an adaptive random-walk Metropolis algorithm with parallel updates of replicate-specific latent blocks across t (Shaby and Wells, 2010). The full hierarchical specification and MCMC implementation details are given in section B.3.

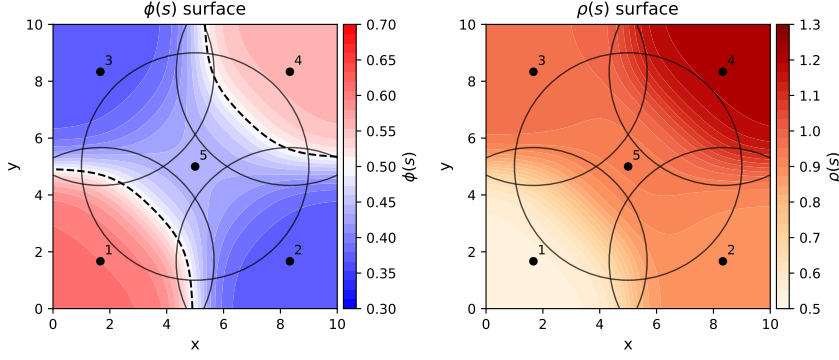


Figure 3.3: Simulation surfaces of $\phi(\mathbf{s})$ and $\rho(\mathbf{s})$. The black dashed line marks the local AD/AI transition. The $\bullet$ symbols denote knot locations, and circles indicate kernel radii.

3.3.2 Simulation and Coverage Analysis

We assessed posterior coverage using 50 independent datasets simulated from the model in Section 3.3.1. For each dataset, $D = 100$ sites were sampled uniformly over $\mathcal{S} = [0, 10]^2$, with $T = 64$ replicates. The simulation used a locally isotropic nonstationary Matérn latent field with $\nu = 1$, $K = 5$ knots on a regular grid, Wendland basis radius $l = 4$, Gaussian kernel bandwidth 4 for $\phi(\mathbf{s})$ and $\rho(\mathbf{s})$, Lévy parameter $\gamma = 1$, and log-Laplace nugget parameter $\alpha_0 = 5$; the resulting $\phi(\mathbf{s})$ and $\rho(\mathbf{s})$ surfaces are shown in Figure 3.3. Observations were then mapped to a censored peaks-over-threshold scale with $p_0 = 0.95$, threshold $y_0 = 60$, and Generalized Pareto parameters $(\sigma, \xi) = (\exp(3), 0.15)$.

For each simulated dataset, we fit the model and compute posterior credible intervals for dependence, marginal, and kernel radius parameters. Figure 3.4 reports empirical coverage rates at knot locations, with standard binomial confidence intervals around the empirical proportions. The observed coverage is close to nominal, indicating well-calibrated posterior inference.

3.4 Extreme of *in situ* Daily Precipitation

3.4.1 Data Analysis

We re-analyze the precipitation dataset from Shi et al. (2026). That work fit Model (M4) to the summer seasonal maxima using a GEV response. Here, we again fit Model (M4), but instead analyze daily exceedances over a high threshold using a GPD response. Revisiting the previous analysis of seasonal maxima is worthwhile for two reasons. First, exploratory analysis in Shi

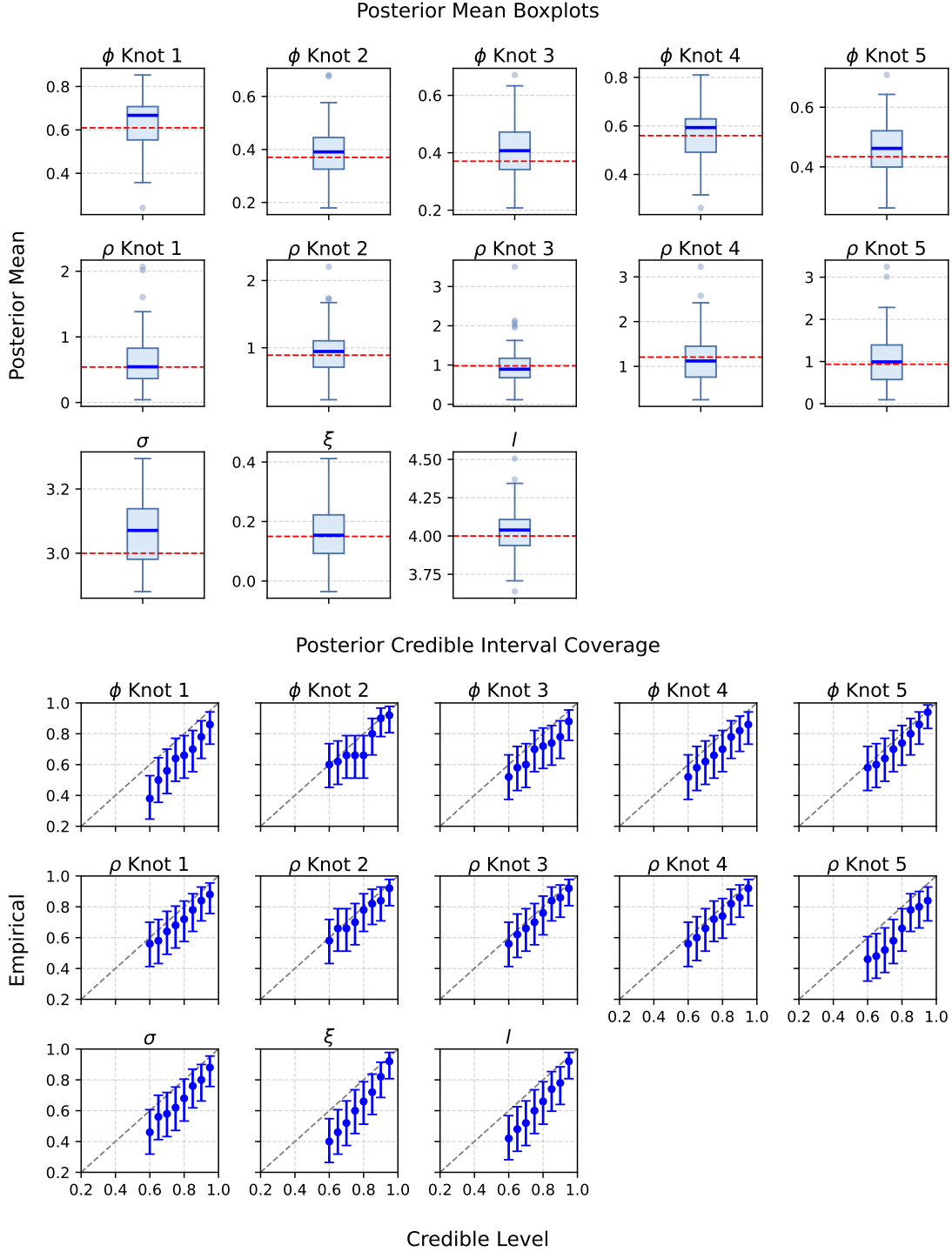


Figure 3.4: Posterior means (top) and standard binomial empirical coverage confidence intervals (bottom) for dependence (ϕ , ρ), marginal (σ , ξ), and kernel radius (l) parameters across 50 simulated datasets. Red dashed lines in boxplots mark the true values.

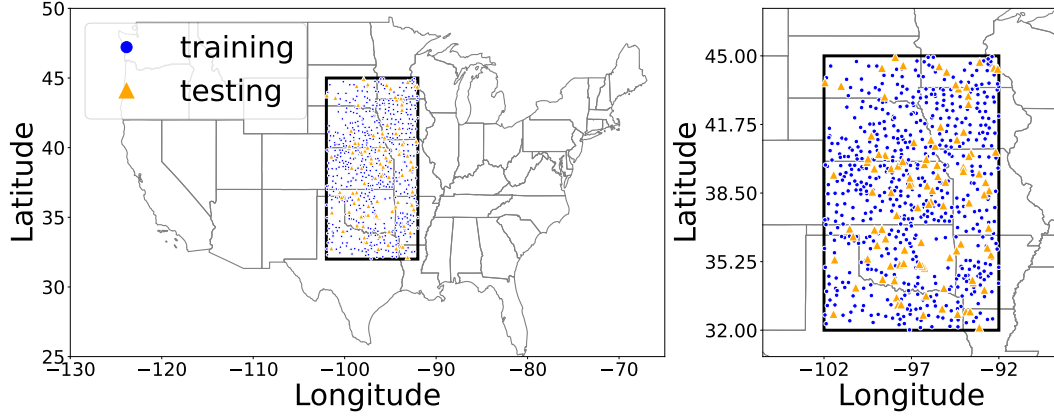


Figure 3.5: Location of the 590 stations (blue circle) and the 99 out-of-sample testing stations (orange triangle) in the central US used for analysis.

et al. (2026) found that a single field of summer maxima often combines extremes from many distinct storms; out of roughly 90 summer days, the station-wise seasonal maxima occur on about 75 different days on average. As a result, seasonal maxima may artificially inflate spatial tail dependence by pooling extremes that occur at different times. Second, the threshold-exceedance dataset is unusually large for spatial extremes, providing a stringent test of the computational feasibility of the proposed censored-likelihood approach in high dimensions.

To reduce short-range temporal dependence, we divide each summer into nine consecutive 10-day periods and retain the maximum daily precipitation within each period. This yields 675 time points at each of the 590 stations (see Figure 3.5). We fit the exceedances of a site-specific 0.95 quantile threshold.

Empirical diagnostics indicate that treating the 10-day maxima as approximately temporally independent is reasonable. Figure 3.6 shows 50-year return levels estimated from annual maxima using 50-year sliding windows at 12 randomly selected stations. The absence of systematic trends supports the assumption of temporally constant marginal parameters. Although such assumptions can be unrealistic for meteorological variables, especially temperature, they appear reasonable for precipitation extremes in this setting. To account for broad terrain effects, we model the marginal

Table 3.2: Candidate model configurations considered for the precipitation analysis. Model naming convention is as follows: k , b , and m respectively denote the number of knots, effective range of the Gaussian basis (the distance at which the kernel function drops below 0.05), and restriction indicator on marginal GPD parameters.

Model	Number of knots	Gaussian basis effective range	Constraint
k25b4	25	4	—
k25b4m	25	4	Fixed σ, ξ
k41b4	41	4	—
k41b4m	41	4	Fixed σ, ξ

GPD parameters as functions of elevation, as

$$\log \sigma(\mathbf{s}) = \beta_{\sigma,0} + \beta_{\sigma,1} \text{elev}(\mathbf{s}), \quad \xi(\mathbf{s}) = \beta_{\xi,0} + \beta_{\xi,1} \text{elev}(\mathbf{s}).$$

For the dependence model, we place knots on a regular grid over the spatial domain. Motivated by the analysis of Shi et al. (2026), we consider four candidate configurations that vary in the number of knots and in whether the marginal parameters are *fixed* at their smoothed site-wise estimates or updated *jointly* with the spatial dependence model. Across these models, the Wendland basis radius governing the latent scale surface is treated as unknown and updated within the MCMC. The model specifications are summarized in Table 3.2. We run each chain for approximately 250,000 iterations, retain every fifth draw, and after discarding burn-in obtain 15,000 posterior samples for inference. After parallelization, one chain took about 70 hours on an AMD Milan EPYC CPU, or about 1 second per iteration.

3.4.2 Model Evaluation

To assess model performance, we use an additional set of 99 holdout stations from the same spatial domain and time period (see Figure 3.5). We compare the four candidate models using predictive log scores at these holdout sites. Figure 3.7 summarizes the resulting predictive log scores. Based on this criterion, we select the k41b4 model for the remainder of the analysis.

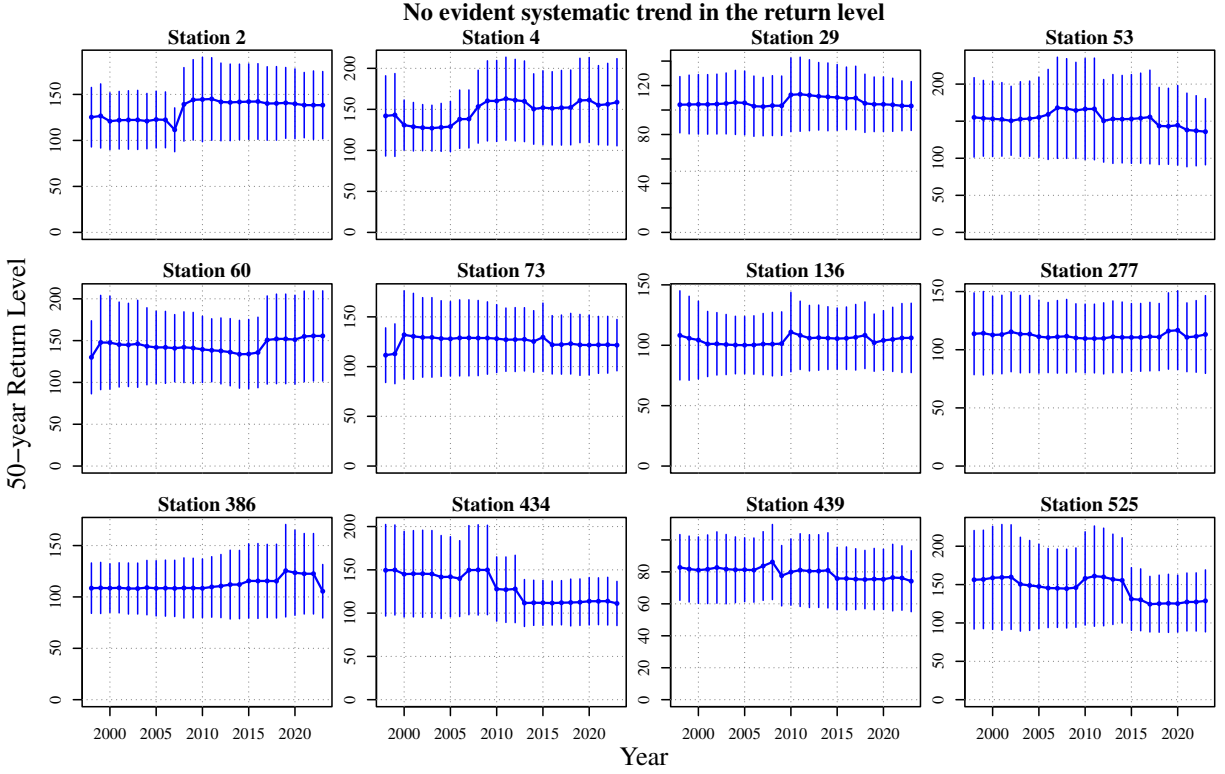


Figure 3.6: Point estimates and 95% confidence intervals for 50-year precipitation return levels at 12 randomly selected stations, estimated from annual maxima using 50-year sliding windows. No evident systematic trend is observed, supporting the assumption of temporally constant marginal parameters.

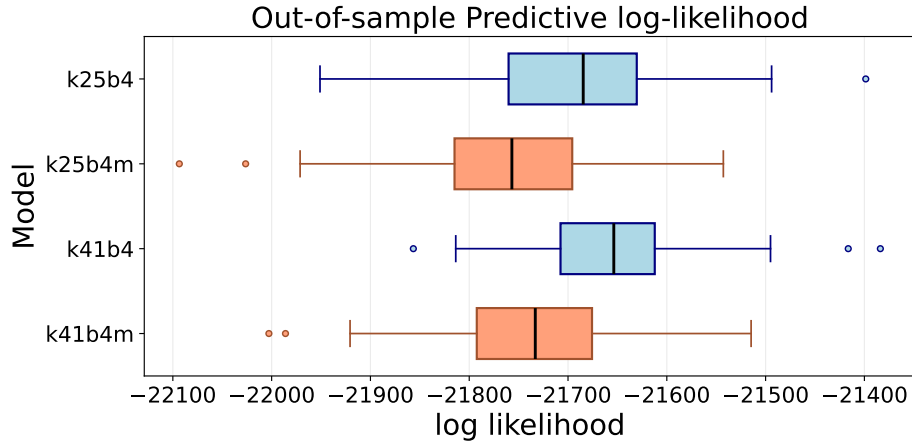


Figure 3.7: Predictive log scores at holdout sites for the four candidate models.

The comparison also shows a systematic advantage for models that update the marginal parameters *jointly* with the dependence model, relative to models that fix the marginals at pre-estimated

values. This mirrors the findings of Shi et al. (2026) and suggests that fully joint hierarchical inference can improve model performance over the more common two-step approach.

We also assess marginal fit using empirical quantile plots at the holdout sites compared to threshold exceedance draws from the posterior predictive distribution. Figure 3.8 shows QQ-plots for four randomly selected holdout stations under the selected `k41b4` model. In each case, the 95% uncertainty band covers the 1:1 line reasonably well, indicating adequate marginal fit.

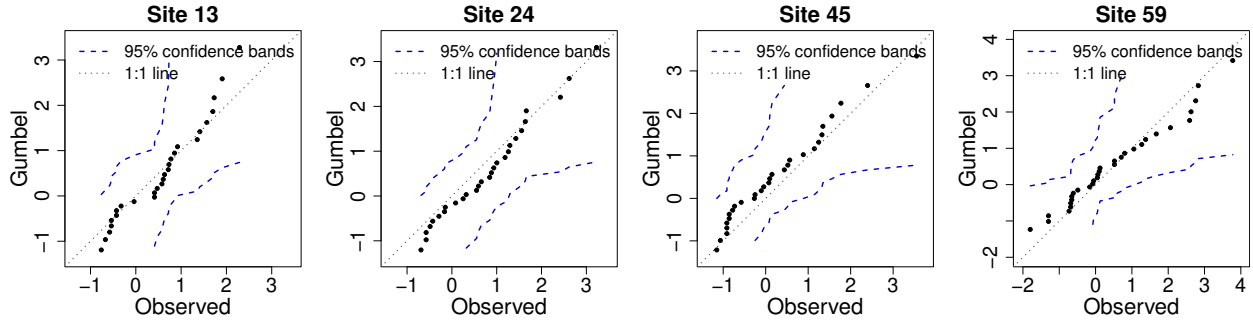


Figure 3.8: QQ-plots of four randomly selected holdout locations, comparing observed and predicted marginal quantiles for the `k41b4` model, along with 95% confidence envelopes. All values are transformed to the Gumbel scale for ease of visualization.

3.4.3 Results

We now present results from the selected `k41b4` model. Figure 3.9a visualizes the knot locations and associated basis, and Figure 3.9b–Figure 3.9e display the posterior mean dependence and marginal parameter surfaces. The posterior mean $\phi(\mathbf{s})$ surface lies below 0.5 throughout the domain, indicating short-range asymptotic independence across the study region, but with clear spatial variation in the strength of the dependence, with larger values occurring over the western part of the domain and smaller values occurring toward the east and southeast. This result resembles the spatial structure found in Shi et al. (2026). The $\rho(\mathbf{s})$ surface and the marginal surfaces for $\log \sigma(\mathbf{s})$ and $\xi(\mathbf{s})$ also vary spatially, and these patterns are broadly consistent with the spatial structure found in Shi et al. (2026).

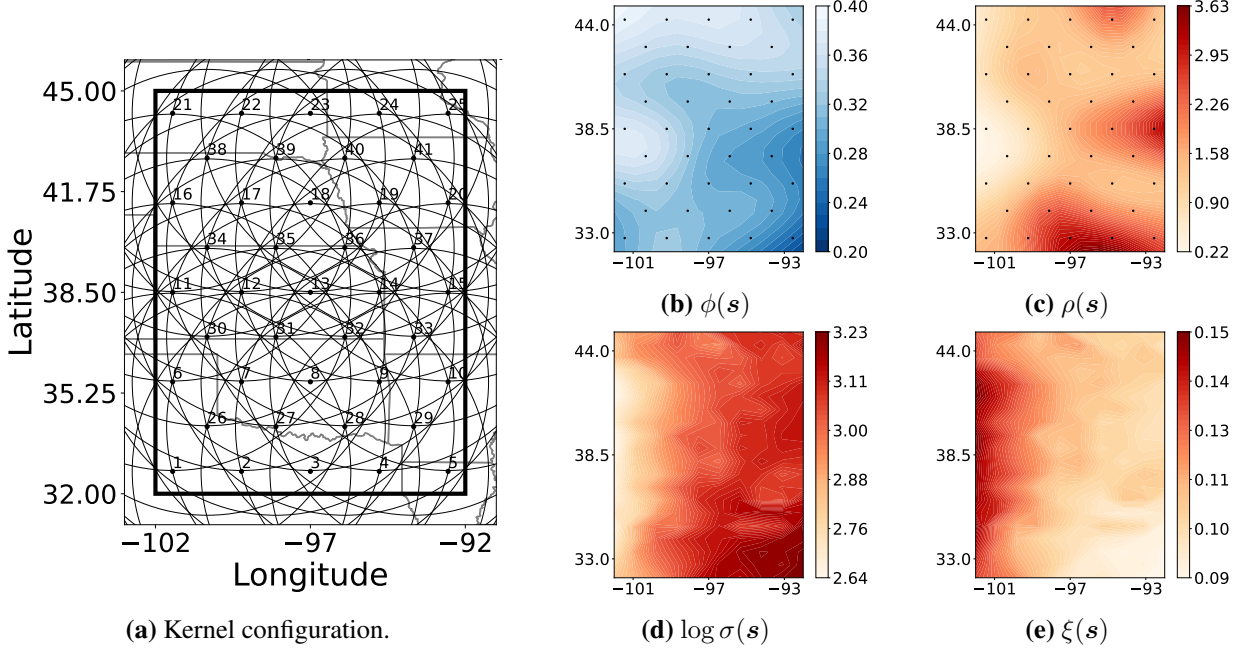


Figure 3.9: Left: For the k41b4 model, the black dots represent the knots and circles represent areas covered by each kernel with posterior mean radius. Right: posterior mean parameter surfaces for the fitted k41b4 model.

To assess extremal dependence fit, Figure 3.10 compares moving-window local empirical $\chi_u(h)$ surfaces with their model-based counterparts over several thresholds and spatial lags. The empirical surfaces in the left column are obtained by transforming the observations $Y_t(s)$ to the $X_t(s)$ scale using the fitted marginal GP parameters from the k41b4 model and then estimating $\chi_u(h)$ empirically. The fitted surfaces in the right column are obtained from conditional posterior predictive draws under the fitted model. We see good agreement between the data and the model fit, as the fitted model reproduces the main localized patches of elevated finite-threshold dependence as well as the overall weakening of dependence with increasing lag and threshold. Moreover, the empirical $\chi_u(h)$ surfaces decrease toward zero as $u \rightarrow 1$, which is consistent with the asymptotic independence implied by $\phi(s) < 0.5$. Overall, these diagnostics indicate that the k41b4 model captures both the broad spatial variation in the marginal behavior and the local subasymptotic tail-dependence structure of the precipitation extremes.

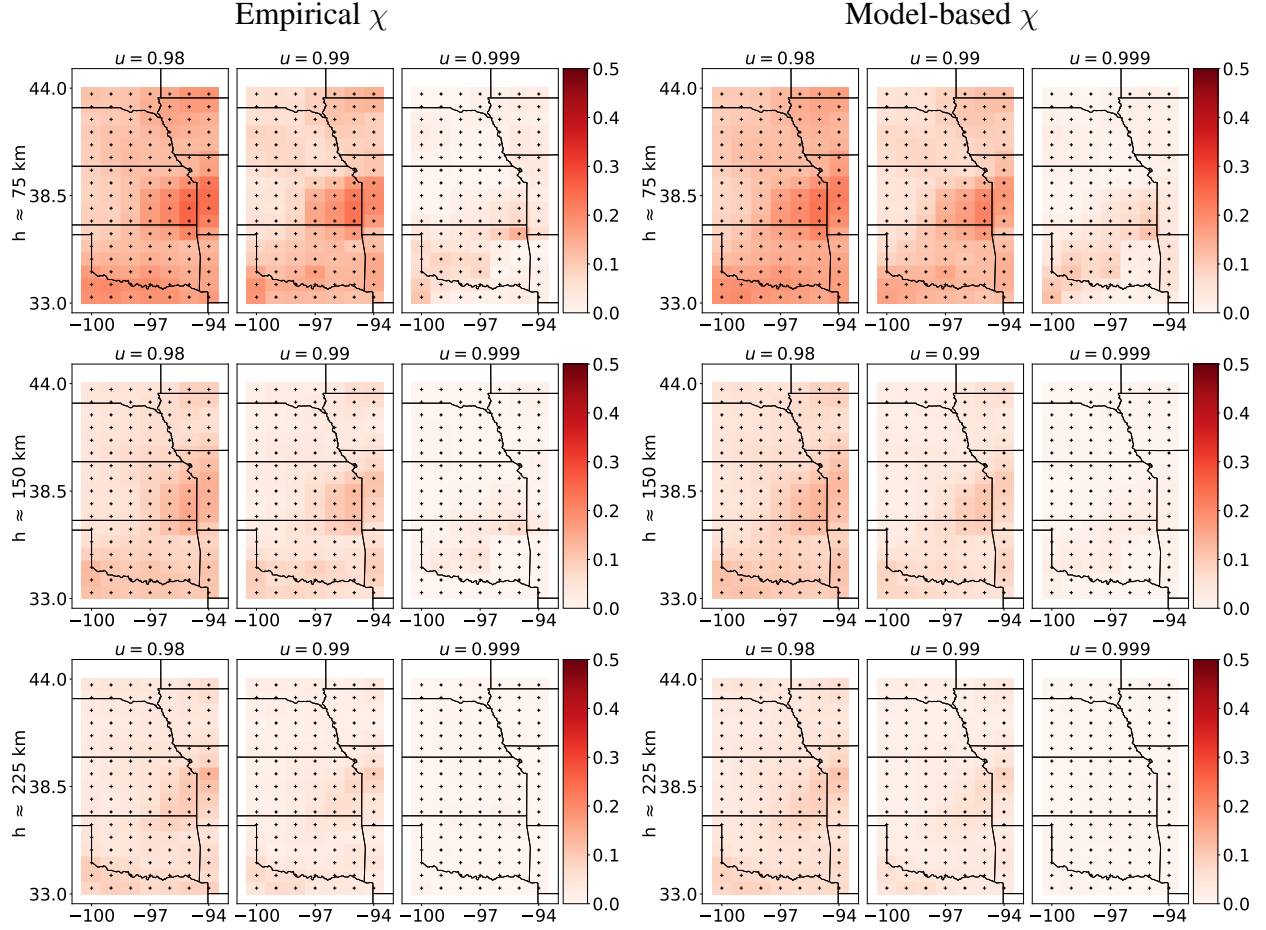


Figure 3.10: Moving-window estimates of $\chi_u(h)$ across three quantiles u and three spatial lags h . The left-hand panel shows the dataset empirical estimates of $\chi_u(h)$, and the right-hand panel shows the model-based estimates of $\chi_u(h)$, based on the k41b4 model.

3.5 Discussion

In this paper, we introduced a multiplicative log-Laplace nugget for a broad class of random scale-mixture models. For inference using the censored likelihood, augmenting the smooth process with the nugget removes the need to evaluate high-dimensional Gaussian distribution functions, substantially reducing computational cost. The particular form of the nugget results in closed forms for the marginal density and distribution functions, which eliminates the need for numerical integration and makes computation feasible for models with spatially varying parameters. At the same time, we showed analytically that including the nugget preserves the main tail properties of the underlying smooth process. The result is that the total computational cost is dominated by

standard spatial statistics operations like factorizing the covariance matrix. This represents a major shift, as heretofore likelihood-based analysis of spatial extremes has been severely limited by formidable computational obstacles, which have only been ameliorated by using bespoke approximations (e.g. Padoan et al., 2010; Shaby, 2014; Wadsworth and Tawn, 2014; de Fondeville and Davison, 2018), often entailing considerable compromises.

We illustrated the proposed modification using the model of Shi et al. (2026)—the most flexible available model to the best of our knowledge. Through simulation and data analysis, we showed that the modified model retains the ability to capture qualitatively different forms of asymptotic and subasymptotic tail dependence, while still allowing spatially varying extremal dependence across the domain.

The simulation study and precipitation application also demonstrate that the resulting nuggetted hierarchical model is scalable in practice. Inference can be carried out with standard MCMC, parallelized over time replicates, and numerical routines implemented in C++, the approach is feasible for spatial extreme datasets with both many locations and many time points. In our application, the modified model reduced computation time by a factor of around 100 per iteration relative to the GEV implementation in Shi et al. (2026), while recovering broadly similar spatial dependence patterns.

A further finding, also consistent with Shi et al. (2026), is that models that jointly estimate marginal and dependence parameters outperform analogous two-step approaches in holdout predictive log score. Although joint inference is more demanding to implement because it requires the marginal probability integral transform within the MCMC, the nugget itself does not add an additional layer of numerical integration. As a result, the extra computational cost remains manageable, while the predictive gains suggest that joint estimation is often worthwhile because it propagates uncertainty coherently between marginal and dependence components and improves out-of-sample prediction.

Chapter 4

Research Directions on Flexible Models and Causal Discovery for Multivariate Extremes

4.1 Motivation

The methods developed in Chapter 2 and Chapter 3 focus on flexible tail dependence and scalable inference for high-dimensional spatial extremes. A natural next step is to move from a single process observed over space to genuinely multivariate extremes, where the scientific questions concern how rare events co-occur across distinct variables and whether those extremal relationships exhibit directional structure. In many applications, different pairs of variables may exhibit qualitatively different tail behavior: some pairs may remain strongly dependent in the far tail, others may weaken substantially, and some may display directional extremal relationships that are not well described by symmetric dependence models.

As discussed earlier in the dissertation, the distinction between asymptotic dependence and asymptotic independence is fundamental because it governs how joint tail risk is extrapolated beyond the observed range of the data. In finite samples, however, this distinction is often difficult to diagnose empirically. Tail summaries such as $\chi_{ij}(u)$ may appear to decay with increasing threshold, but the available data rarely extend far enough into the tail to determine with confidence whether the limiting dependence remains positive or vanishes. This challenge becomes even more pronounced in multivariate settings, where different pairs of variables may lie in different tail regimes and where the strength of sub-asymptotic dependence may still vary substantially even among asymptotically independent pairs. A useful multivariate model should therefore allow the tail regime itself to be learned from the data on a pair-by-pair basis, rather than imposing a single global class in advance.

Classical multivariate extreme-value models provide the appropriate asymptotic foundation, but they are often too rigid for this purpose (Rootzén et al., 2018; Huser and Wadsworth, 2022). Limiting models such as multivariate max-stable and generalized Pareto constructions are designed to capture asymptotic behavior, yet they do not naturally accommodate the possibility that some pairs of variables exhibit strong tail dependence while others exhibit weaker forms of association within the same joint distribution. At the same time, a purely nonparametric approach is not well suited to the extrapolation problem that lies at the core of extreme-value analysis, especially as dimension increases. These considerations motivate a structured parametric model that is flexible enough to represent heterogeneous pairwise tail behavior while still supporting principled tail extrapolation.

The transformed Gaussian scale-mixture models developed in Chapter 2 and Chapter 3 suggest a route toward such a framework. In those chapters, random scaling of transformed Gaussian processes provided a mechanism for moving smoothly between asymptotic dependence and asymptotic independence within the interior of the parameter space. The multivariate setting desires a model that allow some pairs of variables to be asymptotically dependent and others asymptotically independent within a coherent joint construction. It should also capture genuinely multivariate structure (as opposed to a spatial pattern), since relationships among variables are not determined by geometric distance and need not follow a common dependence pattern.

Directional structure provides an additional motivation. In many applications, extremal relationships are not exchangeable. Extreme discharge at an upstream river gauge can increase downstream flood risk more directly than the reverse, and extreme exposure to one pollutant may be linked to severe health outcomes differently than exposure to another pollutant. These settings call for a framework that can represent shared latent extremal drivers while also encoding directionality in how those drivers propagate across variables. This future research therefore aims to develop a multivariate extremes model that can represent both strong and weak forms of pairwise tail dependence within a single coherent framework, use a latent indicator matrix to encode directional tail dependence and support causal discovery, and extend the one-stage Bayesian strategy of the

previous chapters to joint learning of marginal behavior, dependence parameters, and latent graph structure. An important application goal is to deploy this framework in settings where directional extremal relationships are scientifically consequential, especially river discharge over connected river networks and extreme low birth weight under joint exposure to multiple pollutants.

4.2 Model

The proposed model is motivated by the transformed Gaussian scale-mixture framework that underlies the spatial models in Chapter 2 and Chapter 3; see also Huser and Wadsworth (2019) and Engelke et al. (2019). In the spatial setting, one starts with a Gaussian process and applies a positive random scaling factor together with a marginal transformation to Pareto scale. The key feature of that construction is that the tail behavior of the scaling variable determines whether the resulting model is asymptotically dependent or asymptotically independent. In this way, the model can move smoothly between the two extremal regimes within the interior of the parameter space, rather than requiring a separate model family for each case.

That same idea is appealing in the multivariate setting, but the multivariate problem requires a more flexible scaling structure. If one were to carry over the spatial construction directly by using a single global scaling factor for an entire random vector, then all components would be driven by the same radial effect and the joint tail behavior would be constrained to follow a common pattern. For genuinely multivariate data, that is often unrealistic. Different variable pairs may have very different extremal relationships, and those relationships need not be determined by geometric distance or any other common organizing principle. The model therefore needs to preserve the scale-mixture mechanism that allows smooth transition between asymptotic dependence and asymptotic independence, while also allowing the tail regime to vary from one pair of variables to another.

4.2.1 Construction

To build that flexibility into the model, we begin with a Gaussian random vector and modulate each component by its own heavy-tailed scaling factor. Let $\mathbf{Z} \sim N_d(\mathbf{0}, \Sigma)$ and define

$$X_k = R_k^{\phi_k} g(Z_k), \quad k = 1, \dots, d,$$

where $g(\cdot)$ transforms each standard Gaussian margin to a standard Pareto margin, the positive scaling variables R_k are component specific, and $\phi_k > 0$ controls the tail heaviness of the scaling applied to the k th component. The vector $\phi = (\phi_1, \dots, \phi_d)^\top$ therefore governs how strongly the heavy-tailed scaling acts on each variable, while the Gaussian layer continues to account for lighter-tailed background dependence.

The remaining task is to specify the scaling vector $\mathbf{R}$ in a way that is flexible enough to encode heterogeneous pairwise extremal dependence and structured enough to admit theoretical analysis. We do this by assembling $\mathbf{R}$ from shared latent heavy-tailed building blocks:

$$\mathbf{R} = \mathbf{W}^\top \mathbf{S},$$

where $\mathbf{S} = (S_1, \dots, S_d)^\top$ has independent positive stable components, for example Lévy random variables, and $\mathbf{W}$ is a latent indicator matrix with $W_{ii} = 1$ and $W_{ji} \in \{0, 1\}$ for $i \neq j$. In this representation, $W_{ji} = 1$ indicates that the latent heavy-tailed component associated with variable j contributes to the scaling of variable i . Equivalently, the i th column of $\mathbf{W}$ identifies the predecessors whose latent factors feed into R_i .

This representation makes the roles of ϕ and $\mathbf{W}$ distinct. The entries of ϕ regulate the marginal strength of the scaling, while $\mathbf{W}$ determines which variables share latent extremal drivers. If two columns of $\mathbf{W}$ overlap, then the corresponding variables inherit at least one common heavy-tailed factor. If they do not overlap, then any extremal association between those variables must arise through the Gaussian layer. The overlap structure of $\mathbf{W}$ therefore provides the mechanism by

which some pairs of variables may be asymptotically dependent while others remain asymptotically independent within the same multivariate distribution.

The choice of positive stable, and in particular Lévy, components is useful for both probabilistic and computational reasons. Because the stable family is closed under summation, each component of $\mathbf{R}$ remains in the same heavy-tailed class as the underlying latent variables. This makes it possible to relate the joint tail behavior of $\mathbf{R}$ directly to the overlap structure of the columns of $\mathbf{W}$, which in turn provides a route to understanding the pairwise tail behavior of (X_i, X_j) . The same representation also gives $\mathbf{W}$ a natural graph interpretation, so that the scaling construction simultaneously encodes shared extremal drivers and directional structure.

4.2.2 Pairwise Tail Dependence

The main reason for introducing the scaling representation $\mathbf{R} = \mathbf{W}^\top \mathbf{S}$ is that it allows pairwise tail behavior to be controlled by two distinct ingredients: the overlap structure of the latent stable factors and the tail-heaviness parameters (ϕ_i, ϕ_j) . In other words, whether (X_i, X_j) is asymptotically dependent or asymptotically independent should depend both on whether the two variables share heavy-tailed latent drivers and on how strongly those drivers are amplified by the powers ϕ_i and ϕ_j . This is the key mechanism by which the model seeks to accommodate heterogeneous pairwise tail regimes within a single multivariate distribution.

Let $\mathbf{W}_{\cdot i}$ denote the i th column of $\mathbf{W}$, and let η_{ij}^Z denote the coefficient of tail dependence for the Gaussian layer after transformation, that is, for $(g(Z_i), g(Z_j))$. Then the central analytical task is to understand

$$\Pr(X_i > x, X_j > x) = \Pr\{R_i^{\phi_i} g(Z_i) > x, R_j^{\phi_j} g(Z_j) > x\},$$

and to determine from this joint tail probability the induced values of χ_{ij} and η_{ij} . This pair-specific flexibility is the main departure from more rigid multivariate extremes models. Rather than forcing the entire random vector into one tail regime, the proposed model allows the extremal regime itself

to vary across pairs. That is precisely the type of behavior that the motivating applications suggest is needed.

We conjecture that the pairwise tail behavior of the model is conjectured to take the following form.

Conjecture. Consider the model

$$X_k = R_k^{\phi_k} g(Z_k), \quad \mathbf{R} = \mathbf{W}^T \mathbf{S},$$

and suppose, without loss of generality, that $\phi_i \leq \phi_j$.

(a) If $\mathbf{W}_{\cdot i}^T \mathbf{W}_{\cdot j} = 0$, then X_i and X_j do not share latent Lévy components, and

$$\chi_{ij} = 0.$$

If $\text{Cov}(Z_i, Z_j) = 0$, then X_i and X_j are independent. If $\text{Cov}(Z_i, Z_j) \neq 0$, then

$$\eta_{ij} \approx \max\{\min(\phi_i, \phi_j, 1/2), \eta_{ij}^Z\}.$$

(b) If $\mathbf{W}_{\cdot i}^T \mathbf{W}_{\cdot j} > 0$, then R_i and R_j share a non-empty set of Lévy variables.

(i) If $1/2 < \phi_i < \phi_j$, then (X_i, X_j) is asymptotically dependent with

$$\eta_{ij} = 1$$

and

$$\chi_{ij} \approx (\mathbf{W}_{\cdot i}^T \mathbf{W}_{\cdot j}) \mathbb{E} \left[\min \left\{ \frac{g(Z_i)^{1/(2\phi_i)}}{\mathbb{E}[g(Z_i)^{1/(2\phi_i)}]}, \frac{g(Z_j)^{1/(2\phi_j)}}{\mathbb{E}[g(Z_j)^{1/(2\phi_j)}]} \right\} \right].$$

(ii) If $\phi_i < 1/2 < \phi_j$, then (X_i, X_j) is asymptotically independent with

$$\chi_{ij} = 0$$

and

$$\eta_{ij} \approx (1 + 2\phi_j)^{-1} \max(2\eta_{ij}^Z, 2\phi_i + 2\phi_j).$$

(iii) If $\phi_i < \phi_j < 1/2$, then (X_i, X_j) is asymptotically independent with

$$\chi_{ij} = 0$$

and

$$\eta_{ij} \approx \max(\eta_{ij}^Z, 2\phi_j).$$

The content of this conjecture is what makes the model compelling. When two variables do not share latent stable factors, their extremal association is inherited only through the Gaussian layer, so the pair remains asymptotically independent. When they do share latent factors, the shared heavy-tailed scaling can dominate the joint tail, but whether that produces asymptotic dependence or only stronger sub-asymptotic association depends on the magnitudes of ϕ_i and ϕ_j . In this way, the model is designed to allow some pairs of variables to be asymptotically dependent while others are asymptotically independent, all within a single coherent multivariate construction.

4.2.3 Latent DAG and Causal Discovery

The indicator matrix $\mathbf{W}$ does more than determine which variable pairs share latent heavy-tailed drivers. It also induces a directed graph on the variables, and that graph provides the basis for causal discovery in the tail. The i th column of $\mathbf{W}$ identifies the predecessor variables whose latent stable components contribute to the scaling of X_i . In that sense, $\mathbf{W}$ records how extremal drivers propagate across the multivariate system.

This interpretation closely resembles recursive structural models on directed acyclic graphs. In those models, each variable is built from its ancestors and an idiosyncratic noise term. The present formulation plays an analogous role in the tail: because the components of $\mathbf{S}$ are regularly varying and stable under summation, ancestor variables contribute shared heavy-tailed factors to their descendants through the scaling vector $\mathbf{R}$. When the graph implied by $\mathbf{W}$ is acyclic, poste-

rior learning of $\mathbf{W}$ therefore amounts to learning a latent DAG that organizes the propagation of extremal behavior across variables.

A simple two-variable illustration makes this concrete. Suppose the variables are precipitation P and temperature T , ordered as (P, T) . A physically motivated directional relationship is $T \rightarrow P$ but not $P \rightarrow T$: unusually high temperature can contribute to atmospheric conditions favorable for intense precipitation, whereas precipitation tends to cool the air and therefore does not induce the same extremal effect on temperature. In that case one may take

$$\mathbf{W}_0 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{R} = \mathbf{W}_0^\top \mathbf{S} = \begin{pmatrix} S_1 + S_2 \\ S_2 \end{pmatrix}.$$

Under this construction, the temperature-related latent factor S_2 contributes to both components, while the precipitation-specific factor S_1 contributes only to precipitation. Large values of the temperature-related latent factor therefore propagate into the precipitation scaling, but large precipitation-specific latent effects need not propagate back into temperature. This is the sense in which the edge $T \rightarrow P$ is encoded in the latent DAG, as ancestors contribute their heavy-tailed drivers to descendants, creating an asymmetric implication in the tail.

More generally, if variable j is an ancestor of variable i in the latent graph, then the heavy-tailed factors associated with j appear in the scaling of X_i . Extreme values of X_j therefore tend to induce extreme values of X_i more directly than the reverse, unless the descendant has sufficiently strong idiosyncratic latent factors of its own. The latent DAG thus provides an interpretable mechanism for directional tail dependence.

This perspective is especially natural in the motivating applications. In a river network, upstream gauges can act as ancestors of downstream gauges, so the graph structure reflects flow direction and identifies the dominant pathways through which flood risk propagates. In multiple-pollutant health studies, the graph can encode which exposures share latent extremal drivers and how those drivers feed into an extreme health outcome. In both settings, the latent DAG offers a structured way to connect directional scientific knowledge with multivariate tail dependence.

4.3 Bayesian Inference

Inference is most naturally carried out through a hierarchical copula formulation that separates marginal behavior from joint tail behavior while estimating both parts simultaneously. A standard copula analysis would proceed in two stages: first fit marginal models for each variable and transform the data to a common scale, then estimate the dependence parameters conditionally on those transformed values. For the present problem, that approach is not ideal, because uncertainty in the marginal transformations should propagate directly into inference on the tail dependence parameters and the latent graph. As with the earlier chapters, we therefore adopt a one-stage Bayesian strategy in which the marginal and dependence components are learned jointly.

4.3.1 Hierarchical Model

Let $\mathbf{Y} = (Y_1, \dots, Y_d)^\top$ denote the observed variables on their original scales, and let $\mathbf{X} = (X_1, \dots, X_d)^\top$ denote the corresponding latent variables on the multivariate tail-model scale. Write $F_{Y_i}(\cdot \mid \boldsymbol{\theta}_Y)$ for the marginal distribution of Y_i and $F_{X_i}(\cdot \mid \boldsymbol{\theta}_X)$ for the marginal distribution induced by the dependence model, with corresponding densities f_{Y_i} and f_{X_i} . The pointwise transformation between the observed and latent scales is then

$$Y_i = F_{Y_i}^{-1}\{F_{X_i}(X_i)\}, \quad i = 1, \dots, d,$$

or equivalently

$$X_i = F_{X_i}^{-1}\{F_{Y_i}(Y_i)\}.$$

With this construction, the joint likelihood for $\mathbf{Y}$ can be written as

$$L(\mathbf{Y} \mid \boldsymbol{\theta}_Y, \boldsymbol{\theta}_X) = f_{\mathbf{X}}(\mathbf{X} \mid \boldsymbol{\theta}_X) \left| \frac{\partial \mathbf{X}}{\partial \mathbf{Y}} \right|,$$

where $f_{\mathbf{X}}$ is the joint density of the latent multivariate tail model and the Jacobian matrix is diagonal, with entries

$$\frac{\partial X_i}{\partial Y_i} = \frac{f_{Y_i}(Y_i)}{f_{X_i}(X_i)}, \quad i = 1, \dots, d.$$

This hierarchical representation allows the marginal parameters and the dependence parameters to be estimated in a single posterior analysis rather than treating the marginal transformation as fixed.

Under the proposed multivariate model, the dependence parameter block includes the Gaussian dependence matrix Σ , the tail-heaviness parameters ϕ , and the latent indicator matrix $\mathbf{W}$. The first two are continuous parameters, while $\mathbf{W}$ introduces a discrete structure-learning component. This creates the main new inferential feature relative to the earlier chapters: posterior computation must learn both a flexible tail dependence model and a latent directed graph. Since the graph enters through the scaling representation $\mathbf{R} = \mathbf{W}^T \mathbf{S}$ rather than through a Gaussian precision matrix, standard Gaussian DAG methods do not apply directly. Instead, inference can be based on the likelihood of the full multivariate tail model.

A practical prior specification is to assign each unordered pair of variables one of three possible edge states: no edge, $i \rightarrow j$, or $j \rightarrow i$. Equivalently, for each pair $i \neq j$, the prior support for (W_{ij}, W_{ji}) is

$$(W_{ij}, W_{ji}) \in \{(0, 0), (1, 0), (0, 1)\}.$$

Posterior computation can then proceed by combining updates for the continuous parameters (ϕ, Σ) with Metropolis-Hastings proposals for the off-diagonal entries of $\mathbf{W}$. When causal interpretation is desired, the graph can be restricted to be acyclic, in which case any proposed move that creates a directed cycle is rejected automatically. This is especially reasonable in the low- to moderate-dimensional settings of interest here, where the number of variables is small enough that full posterior exploration of the graph remains feasible.

The resulting MCMC algorithm targets the joint posterior of the marginal parameters, the dependence parameters, and the latent graph. In principle, this yields coherent uncertainty quantification not only for pairwise tail dependence summaries and joint tail probabilities, but also for

directional features encoded by $\mathbf{W}$. A preliminary implementation of the multivariate sampler already provides the foundation for this inferential framework. The next stage is to integrate the computation into software for simulation, posterior fitting, and model diagnostics.

4.3.2 Simulation Study and Coverage Analysis

To evaluate the proposed multivariate model and the associated Bayesian inferential procedure, we will conduct a simulation study accompanied by a coverage analysis. The goal is to assess whether the model reproduces the pairwise tail behavior it is designed to capture, including settings with mixed asymptotic dependence classes and directional extremal structure, and whether posterior credible intervals provide reliable uncertainty quantification for both the continuous dependence parameters and the latent graph components.

The simulation design should include several distinct scenarios. One class of scenarios should contain pairs of variables that do not share latent stable factors, so that any extremal association arises only through the Gaussian layer. A second should contain shared latent factors with tail parameters large enough to induce asymptotic dependence. A third should mix these behaviors within the same random vector, so that some pairs are asymptotically dependent while others are asymptotically independent. Since the latent graph is itself an inferential target, additional scenarios should vary the complexity of the directed structure, including sparse graphs, graphs with multiple ancestors, and settings in which directional relationships are weak and therefore difficult to recover. This is the setting in which the proposed model should offer its clearest advantage over more rigid multivariate tail models.

The simulation analysis should examine both parameter recovery and induced tail behavior. At the parameter level, posterior interval coverage can be studied for the entries of ϕ , summaries of Σ , and the edge indicators in $\mathbf{W}$. At the dependence level, one can assess whether posterior samples recover the correct pairwise regime through χ_{ij} and η_{ij} , and whether they reproduce joint exceedance probabilities and related risk summaries. In scenarios with known latent graph structure, posterior recovery of ancestor-descendant relationships can also be evaluated directly.

Model assessment should also compare the proposed framework with simpler alternatives that impose a single global extremal regime. Posterior predictive checks can then be used to assess pairwise upper-tail probabilities, empirical estimates of $\chi_{ij}(u)$ over a range of thresholds, and summaries of sub-asymptotic dependence for pairs that remain asymptotically independent. In this way, the simulation study evaluates both the asymptotic properties targeted by the theory and the finite-threshold behavior for practical risk analysis.

4.4 Applications

The methodological development in this chapter is motivated by problems in which extremal dependence is both heterogeneous and directional. One also wants to understand which components of the system share extremal drivers, which relationships weaken in the far tail, and how extreme risk propagates across variables. The proposed multivariate framework is especially well suited to applications where some pairs of variables may be tightly linked in the tail while others are only weakly associated, and where the direction of those relationships is itself scientifically meaningful. Two natural applications are: river discharge over connected river networks and extreme low birth weight under joint exposure to multiple pollutants.

4.4.1 Fluvial Flooding in River Networks

River discharge observed at multiple gauges provides a particularly natural application because river flow induces a clear directional structure from upstream to downstream locations. At the same time, extremal dependence need not be uniform across a network. Gauges on the same tributary may share strong upper-tail behavior, gauges separated by confluences may exhibit more complicated relationships, and gauges on disconnected branches may be only weakly related in the tail. A useful statistical model in this setting must therefore do more than represent network connectivity. It must also allow the extremal regime itself to vary across gauge pairs while preserving the directionality imposed by flow.

The Mississippi River system is an especially compelling target because the scientific questions concern not only whether high discharge occurs jointly at several gauges, but also how flood stress propagates through the basin. Under changing climatic conditions, historical assumptions of stationarity are increasingly difficult to maintain (Milly et al., 2008). At the same time, land-use change, channel modification, levees, reservoirs, and other engineered controls can alter both the magnitude of flood extremes and the way those extremes are transmitted through the network. For this reason, it is important to identify not only where tail dependence is strong, but also how its structure changes across tributaries and through time.

The proposed model is designed to address these questions directly. The pairwise tail dependence component allows some gauge pairs to be asymptotically dependent and others asymptotically independent within the same joint model, while the latent graph structure provides a statistical representation of how extremal drivers move through the river system. In a fitted analysis, the graph could identify dominant pathways of extremal transmission, distinguish broad basin-scale drivers from more localized branch-specific effects, and reveal portions of the network where downstream vulnerability is amplified by upstream extremes. In this way, the model links multivariate tail dependence to flood-risk propagation and infrastructure planning.

4.4.2 Multiple Pollutants and Extreme Low Birth Weight

A second application concerns multiple-pollutant exposure during pregnancy and extreme low birth weight. Much of the existing literature on pollution and birth outcomes focuses on mean effects, asking how exposure shifts the average level of birth weight. However, from a public-health perspective, the more consequential outcomes may lie in the lower tail of the birth-weight distribution, where unusually adverse outcomes occur. The scientific goal is therefore to understand how extreme exposure profiles across several pollutants are related to extreme low birth weight, rather than to average birth weight alone.

This setting also motivates a genuinely multivariate extremes approach. Pollutants need not share a common upper-tail regime, as some may co-occur strongly because they arise from similar

sources or meteorological conditions, while others may exhibit much weaker joint tail behavior. A model that imposes the same extremal class on all pollutant pairs would therefore miss an important feature of the data. The proposed framework is designed to separate pollutant pairs whose extreme exposures co-occur strongly from those whose dependence weakens in the tail, while still embedding those relationships in a coherent multivariate model.

The latent graph structure is also scientifically useful in this application. It provides a way to represent how extreme exposures and extreme health outcomes are organized jointly, allowing the analysis to move beyond marginal exposure-response summaries toward a model for the joint tail of the exposure vector and the outcome. In practice, this would make it possible to study whether particular pollutants or combinations of pollutants are more directly associated with the most adverse birth outcomes, while accounting for shared extremal drivers among the exposures themselves. The resulting analysis would focus squarely on the rare outcomes of greatest clinical concern and would provide a more natural framework for studying mixture effects in the tail.

4.5 Discussion

This chapter outlines a natural extension of the high-dimensional spatial extremes modeling to multivariate extremes with heterogeneous and directional dependence structure. The central methodological idea is to extend the transformed Gaussian scale-mixture framework developed in Chapter 2 and Chapter 3 from a spatially indexed scaling construction to a multivariate setting with component-specific scaling variables built from shared latent stable factors. This construction is intended to allow some variable pairs to be asymptotically dependent while others remain asymptotically independent within the same joint model, while the latent indicator matrix provides a mechanism for representing directional extremal relationships and supporting causal discovery through an estimated DAG.

The chapter carries forward the main themes of flexible tail dependence modeling, coherent Bayesian inference, and interpretable risk assessment for rare events, by extending methodologies developed in Chapter 2 and Chapter 3 to multivariate systems in which tail behavior may vary

across pairs of variables and may exhibit directionality. The resulting framework would contribute a new class of multivariate tail models together with a Bayesian strategy for learning both dependence parameters and latent graph structure. Applications such as river networks and pollution-health studies illustrate the scientific utility of this framework, since in both settings the organization of the joint tail is scientifically meaningful. More broadly, this research direction would extend the scope of flexible tail dependence modeling to a wider class of multivariate risk problems in which heterogeneous and directional extremal dependence are themselves primary objects of inference.

Chapter 5

Conclusions

Extreme events rarely occur in isolation, and understanding how they co-occur across space is central to realistic risk assessment. This dissertation develops flexible statistical methodology for tail dependence in high-dimensional extremes, motivated by the need for models that can reflect the spatial heterogeneity observed in environmental data while remaining feasible to fit over large spatial domains. It introduces a transformed Gaussian scale-mixture constructions that allow extremal dependence to vary across space and spatial lags, and it develops computational tools that make threshold exceedance inference for the broader generic class of Gaussian scale-mixture models tractable. These ideas further suggest a natural extension to multivariate extremes with directional structure.

Chapter 2 introduced a spatial scale-aware model fitted to marginal block maxima data that allows extremal dependence to vary across both location and distance. By allowing the random scaling mechanism and the tail dependence parameter to evolve smoothly over space, the model can exhibit long-range asymptotic independence together with either short-range asymptotic dependence or short-range asymptotic independence within a single joint framework. This flexibility is important because environmental extremes may not follow a single dependence class over a large and heterogeneous domain. The simulation study showed that the model can recover the intended dependence behavior with well-calibrated inference, and the application to summertime precipitation in the central United States revealed substantial spatial heterogeneity in tail dependence. In particular, the fitted model suggested that local extremal behavior differs across subregions, demonstrating that a single global dependence class would miss scientifically meaningful spatial structure.

Chapter 3 addressed the main computational barrier to fitting a class of transformed Gaussian scale-mixture models under the peaks-over-threshold framework. For threshold exceedance data, censored likelihood inference for Gaussian scale-mixture models typically requires repeated

evaluation of high-dimensional Gaussian distribution functions, which quickly becomes infeasible as the number of sites grows. The multiplicative log-Laplace nugget developed in that chapter removes this bottleneck by inducing conditional independence in the censored likelihood while preserving the marginal tail behavior and extremal dependence structure of the underlying smooth process. This is important because it improves scalability without giving up the appealing flexible tail behaviors of the original models. In the precipitation application, the nuggeted model made Bayesian inference feasible for a large monitoring network, recovered spatial dependence patterns broadly consistent with the heterogeneity identified in the earlier analysis, and delivered substantial computational gains. Taken together, Chapter 2 and Chapter 3 show that flexible tail dependence modeling and practical high-dimensional inference can be developed within the same framework rather than treated as separate goals.

A broader theme connecting these two chapters is the value of one-stage Bayesian inference. In both the block-maxima and threshold-exceedance settings, models that jointly estimated marginal and dependence parameters performed better than analogous two-step approaches based on plug-in marginal estimates. Beyond improved predictive performance, the joint hierarchical formulation propagates uncertainty coherently from the marginal model into inference on dependence, return levels, and other risk summaries. For environmental applications, where decisions often depend on the joint behavior of extremes rather than on marginal tails alone, this coherence is a substantial practical advantage.

Chapter 4 then outlined a natural next step of extending these ideas from spatial extremes to genuinely multivariate extremes. The proposed framework replaces the spatially indexed scaling construction with component-specific scaling variables built from shared latent stable factors, allowing some variable pairs to be asymptotically dependent while others remain asymptotically independent within the same joint model. A latent indicator matrix provides both a mechanism for heterogeneous pairwise tail behavior and an interpretable directed graph for representing how extremal drivers propagate across variables. This extension carries forward the central themes of the dissertation on flexible tail dependence modeling, coherent Bayesian inference, and interpretable

risk assessment for rare events. Applications such as river discharge over connected river networks and multiple-pollutant exposure with extreme low birth weight illustrate the scientific potential of this framework and the importance of modeling heterogeneous and directional extremal relationships directly.

Overall, this dissertation develops a flexible framework for modeling tail dependence in high-dimensional extremes. The first two projects establish this framework in the spatial setting by introducing a spatial scale-aware model that captures heterogeneous extremal dependence and by developing scalable inference on threshold exceedance data for the Gaussian scale-mixture class. The proposed multivariate extension then carries these ideas beyond spatial extremes to settings where tail dependence may vary across variable pairs and may also exhibit directional structure. Taken together, these contributions broaden the scope of extreme-value modeling toward methods that are interpretable and practically feasible for complex high-dimensional systems.

Bibliography

- Raphaël Huser and Jennifer L Wadsworth. Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, 114(525):434–444, 2019.
- Likun Zhang, Mark D Risser, Edward M Molter, Michael F Wehner, and Travis A O’Brien. Accounting for the spatial structure of weather systems in detected changes in precipitation extremes. *Weather and Climate Extremes*, 38:100499, 2022a.
- Leo R Belzile. *mev: Modelling of Extreme Values*, 2023. URL <https://lbelzile.github.io/mev/>.
- Christopher B Field, Vicente R Barros, Michael D Mastrandrea, Katharine J Mach, MA-K Abdrabo, N Adger, Yury A Anokhin, Oleg A Anisimov, Douglas J Arent, Jonathon Barnett, et al. Summary for policymakers. In *Climate change 2014: impacts, adaptation, and vulnerability. Part A: global and sectoral aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 1–32. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2014. doi: 10.1017/CBO9781107415379.
- Daniela Castro-Camilo and Raphaël Huser. Local likelihood estimation of complex tail dependence structures, applied to us precipitation extremes. *Journal of the American Statistical Association*, 115(531):1037–1054, 2020.
- Arnab Hazra, Raphaël Huser, and David Bolin. Efficient modeling of spatial extremes over large geographical domains. *Journal of Computational and Graphical Statistics*, 34(3):795–811, 2025.
- Reetam Majumder, Brian J Reich, and Benjamin A Shaby. Modeling extremal streamflow using deep learning approximations and a flexible spatial process. *The Annals of Applied Statistics*, 18(2):1519–1542, 2024.

- JL Wadsworth and JA Tawn. Higher-dimensional spatial extremes via single-site conditioning. *Spatial Statistics*, 51:100677, 2022.
- Likun Zhang, Benjamin A. Shaby, and Jennifer L. Wadsworth. Hierarchical transformed scale mixtures for flexible modeling of spatial extremes on datasets with many locations. *Journal of the American Statistical Association*, 117(539):1357–1369, 2022b.
- Paul CD Milly, Julio Betancourt, Malin Falkenmark, Robert M Hirsch, Zbigniew W Kundzewicz, Dennis P Lettenmaier, and Ronald J Stouffer. Stationarity is dead: Whither water management? *Science*, 319(5863):573–574, 2008.
- Raphaël Huser and Marc G Genton. Non-stationary dependence structures for spatial extremes. *Journal of agricultural, biological, and environmental statistics*, 21(3):470–491, 2016.
- A. C. Davison, R. Huser, and E. Thibaud. Geostatistics of Dependent and Asymptotically Independent Extremes. *Mathematical Geosciences*, 45(5):511–529, 2013. ISSN 1874-8953.
- Raphaël Huser and Jennifer L. Wadsworth. Advances in statistical modeling of spatial extremes. *WIREs Computational Statistics*, 14(1):e1537, 2022. ISSN 1939-0068. doi: 10.1002/wics.1537.
- Likun Zhang, Mark D. Risser, Michael F. Wehner, and Travis A. O’Brien. Leveraging extremal dependence to better characterize the 2021 pacific northwest heatwave. *Journal of Agricultural, Biological and Environmental Statistics*, June 2024. ISSN 1537-2693.
- Raphaël Huser, Thomas Opitz, and Emeric Thibaud. Bridging asymptotic independence and dependence in spatial extremes using gaussian scale mixtures. *Spatial Statistics*, 21:166–186, 2017.
- Sebastian Engelke, Thomas Opitz, and Jennifer Wadsworth. Extremal dependence of random scale constructions. *Extremes*, 22(4):623–666, 2019.
- Christopher J Paciorek and Mark J Schervish. Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17(5):483–506, 2006.

- Aldo V Vecchia. Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 50(2):297–312, 1988.
- Anthony W Ledford and Jonathan A Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187, 1996.
- Janet Heffernan and Sidney Resnick. Hidden regular variation and the rank transform. *Advances in Applied Probability*, 37(2):393–414, 2005.
- John P Nolan. *Univariate Stable Distributions: Models for Heavy Tailed Data*. Springer Nature, 2020.
- Leonard Breiman. On some limit theorems similar to the arc-sin law. *Theory of Probability & Its Applications*, 10(2):323–331, 1965.
- Daren BH Cline. Convolution tails, product tails and domains of attraction. *Probability Theory and Related Fields*, 72(4):529–557, 1986.
- Holger Wendland. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in Computational Mathematics*, 4(1):389–396, December 1995.
- Mark D Risser and Catherine A Calder. Regression-based covariance functions for nonstationary spatial modeling. *Environmetrics*, 26(4):284–297, 2015.
- Benjamin Shaby and Martin Wells. Exploring an adaptive Metropolis algorithm. Technical Report 1011-14, Duke University Department of Stastical Science, 2010.
- Brian Gough. *GNU scientific library reference manual*. Network Theory Ltd., 2009.
- Lisandro Dalcin and Yao-Lung L. Fang. mpi4py: Status update after 12 years of development. *Computing in Science & Engineering*, 23(4):47–54, 2021.

- Matthew J Menne, Imke Durre, Russell S Vose, Byron E Gleason, and Tamara G Houston. An overview of the global historical climatology network-daily database. *Journal of atmospheric and oceanic technology*, 29(7):897–910, 2012.
- Mark D Risser, Christopher J Paciorek, Michael F Wehner, Travis A O’Brien, and William D Collins. A probabilistic gridded product for daily precipitation extremes over the united states. *Climate Dynamics*, 53(5):2517–2538, 2019.
- Matthew Sainsbury-Dale, Andrew Zammit-Mangion, and Raphaël Huser. Likelihood-free parameter estimation with neural Bayes estimators. *Amer. Statist.*, 78(1):1–14, 2024. ISSN 0003-1305,1537-2731. doi: 10.1080/00031305.2023.2249522. URL <https://doi.org/10.1080/00031305.2023.2249522>.
- Giovanni Forzieri, Alessandra Bianchi, Filipe Batista e Silva, Mario Alberto Marin Herrera, Antoine Leblois, Carlo Lavallo, Jeroen Aerts, and Luc Feyen. Escalating impacts of climate extremes on critical infrastructures in europe. *Global Environmental Change*, 48:97–107, 2018.
- Raphaël Huser, Thomas Opitz, and Jennifer L. Wadsworth. Modeling of spatial extremes in environmental data science: time to move away from max-stable processes. *Environmental Data Science*, 4:e3, 2025.
- Muyang Shi, Likun Zhang, Mark Risser, and Benjamin A Shaby. Spatial scale-aware tail dependence modeling for high-dimensional spatial extremes. *Journal of the American Statistical Association*, 2026. In press.
- Simone A Padoan, Mathieu Ribatet, and Scott A Sisson. Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489):263–277, 2010.
- Benjamin A Shaby. The open-faced sandwich adjustment for mcmc using estimating functions. *Journal of Computational and Graphical Statistics*, 23(3):853–876, 2014.
- Jennifer L Wadsworth and Jonathan A Tawn. Efficient inference for spatial extreme value processes associated to log-gaussian random functions. *Biometrika*, 101(1):1–15, 2014.

- Raphaël de Fondeville and Anthony C Davison. High-dimensional peaks-over-threshold inference. *Biometrika*, 105(3):575–592, 2018.
- Holger Rootzén, Johan Segers, and Jennifer L Wadsworth. Multivariate generalized pareto distributions: Parametrizations, representations, and properties. *Journal of Multivariate Analysis*, 165:117–131, 2018.
- Daniel Cooley and Emeric Thibaud. Decompositions of dependence for high-dimensional extremes. *Biometrika*, 106(3):587–604, 2019.
- Anthony G Pakes. Convolution equivalence and infinite divisibility. *Journal of Applied Probability*, pages 407–424, 2004.
- Vladimir M Zolotarev. *One-dimensional stable distributions*, volume 65. American Mathematical Soc., 1986.
- Sidney I Resnick. *Extreme values, regular variation and point processes*. Springer, 2008.
- Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.
- Seth Westra, Lisa V. Alexander, and Francis W. Zwiers. Global increasing trends in annual maximum daily precipitation. *Journal of Climate*, 26(11):3904–3918, 2013.
- Mark D Risser and Michael F Wehner. Attributable human-induced changes in the likelihood and magnitude of the observed extreme precipitation during hurricane harvey. *Geophysical Research Letters*, 44(24):12–457, 2017.

- M. Etminan, G. Myhre, E. J. Highwood, and K. P. Shine. Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing. *Geophysical Research Letters*, 43(24):12,614–12,623, 2016. doi: <https://doi.org/10.1002/2016GL071930>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GL071930>.
- Mark D. Risser, William D. Collins, Michael F. Wehner, Travis A. O’Brien, Huanping Huang, and Paul A. Ullrich. Anthropogenic aerosols mask increases in us rainfall by greenhouse gases. *Nature Communications*, 15(1):1318, 2024.
- Mark D. Risser, Michael F. Wehner, John P. O’Brien, Christina M. Patricola, Travis A. O’Brien, William D. Collins, Christopher J. Paciorek, and Huanping Huang. Quantifying the influence of natural climate variability on in situ measurements of seasonal total and extreme daily precipitation. *Climate Dynamics*, 56(9):3205–3230, 2021.
- Siu Kwan Lam, Antoine Pitrou, and Stanley Seibert. Numba: a llvm-based python jit compiler. In *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC, LLVM ’15*, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450340052.

Appendix A

Supplemental materials for Chapter 2

A.1 Technical proofs

A.1.1 Properties of stable distribution

The Stable distribution is important in both theory and application because it is the generalized central limit of random variables without second (or even first) order moments. Assume $S \sim \text{Stable}(\alpha, \beta, \gamma, \delta)$ under the 1-parameterization (Nolan, 2020) where $\alpha \in (0, 2]$ is the concentration parameter, $\beta \in [-1, 1]$ is the skewness parameter, $\gamma > 0$ is the scale parameter and $\delta \in \mathbb{R}$ is the location parameter. Then S has the characteristic function

$$E \exp(iuS) = \exp[-\gamma^\alpha |u|^\alpha \{1 - i\beta \omega(u) \text{sign}(u)\} + i\delta u], \quad u \in \mathbb{R},$$

where $\text{sign}(u)$ is the sign of u and $\omega(u) = \tan(\pi\alpha/2)\mathbb{1}(\alpha \neq 1) - 2\pi^{-1} \log |u|\mathbb{1}(\alpha = 1)$. If $\alpha < 1$ and $\beta = 1$, S has support $[\delta, \infty]$. Moreover, if $\alpha < 2$ and $0 < \beta \leq 1$, the tail of S is Pareto-like and satisfies $\Pr(S > x) \sim \gamma^\alpha (1 + \beta) C_\alpha x^{-\alpha}$ with $C_\alpha = \Gamma(\alpha) \sin(\alpha\pi/2)/\pi$ as $x \rightarrow \infty$.

More importantly, the Stable distributions are closed under convolution; sums of α -Stable variables (Stable variables with concentration parameter α) are still α -Stable. If $S_k \stackrel{\text{indep}}{\sim} \text{Stable}(\alpha, \beta_k, \gamma_k, \delta_k)$ and constant $w_k \geq 0$ for $k = 1, \dots, K$, then

$$\sum_{k=1}^K w_k S_k \sim \text{Stable}(\alpha, \bar{\beta}, \bar{\gamma}, \bar{\delta}) \tag{A.1}$$

with $\bar{\gamma} = \{\sum_{k=1}^K (w_k \gamma_k)^\alpha\}^{1/\alpha}$, $\bar{\beta} = \sum_{k=1}^K \beta_k (w_k \gamma_k)^\alpha / \bar{\gamma}^\alpha$ and $\bar{\delta} = \sum_{k=1}^K w_k \delta_k$.

To be able to examine the joint distribution of (X_i, X_j) for model (2.5), it is desirable for the mixture to have the same distributional support and rate of tail decay as each S_k . Thus in (2.7), we fixed $\beta_k \equiv 1$, $\delta_k \equiv \delta$ while imposing the constraint $\sum_{k=1}^K w_k = 1$. As a result, $\bar{\beta} = 1$ and $\bar{\delta} = \delta$,

which means the univariate support of the $\{R(\mathbf{s})\}$ is $[\delta, \infty)$ everywhere. In consequence, we have $\Pr(R(\mathbf{s}) > x) \sim 2\bar{\gamma}^\alpha(\mathbf{s})C_\alpha x^{-\alpha}$ for all $\mathbf{s} \in \mathcal{S}$ as $x \rightarrow \infty$ with $\bar{\gamma}(\mathbf{s}) = [\sum_{k=1}^K \{w_k(\mathbf{s}, r_k)\gamma_k\}^\alpha]^{1/\alpha}$, which means the process $\{R(\mathbf{s})\}$ has tail-stationarity.

A.1.2 Proof of Proposition 1

We begin by recalling a couple of useful theoretical results.

Lemma 10 (Breiman (1965)). Assume X_1 and X_2 are two independent random variables that are both supported by $\mathbb{R}^+$, and that $\Pr(X_1 > x) \in RV_{-\alpha}, \alpha \geq 0$.

(a) If there exists $\epsilon > 0$ such that $E(X_2^{\alpha+\epsilon}) < \infty$, then

$$\Pr(X_1 X_2 > x) \sim E(X_2^\alpha) \Pr(X_1 > x). \quad (\text{A.2})$$

(b) Under the assumptions of part ((a)), we have

$$\sup_{x \geq y} \left| \frac{\Pr(X_1 X_2 > x)}{\Pr(X_1 > x)} - E(X_2^\alpha) \right| \rightarrow 0, \quad y \rightarrow \infty.$$

(c) If $\Pr(X_1 > x) \sim cx^{-\alpha}$, (A.2) holds under $E(X_2^\alpha) < \infty$.

(d) If $\Pr(X_2 > x) = o(\Pr(X_1 > x))$, then $\Pr(X_1 X_2 > x) \in RV_{-\alpha}$.

Lemma 11 (Theorem 4(v) of Cline (1986)). Let $X_1 \sim F_1$ and $X_2 \sim F_2$ be two random variables that are both exponential tailed with the same rate, i.e. $F_i \in \text{ET}_{\alpha, \beta_i}, \alpha > 0, \beta_i > -1, i = 1, 2$. Then

$$\Pr(X_1 + X_2 > x) \sim \alpha \frac{\Gamma(\beta_1 + 1)\Gamma(\beta_2 + 1)}{\Gamma(\beta_1 + \beta_2 + 1)} l(x) x^{\beta_1 + \beta_2 + 1} \exp(-\alpha x),$$

where $l(\cdot)$ is slowly varying.

Proof of Proposition 1. When $0 < \phi_j < \alpha$, the ϕ_j th moment of R_j exists: $E(R_j^{\phi_j}) < \infty$. This is sometimes called the fractional lower order moment. Recall that marginally $\Pr(W_j > x) = x^{-1}$. Thus applying Lemma 10 part (c) to $R_j W_j$ yields $\Pr(R_j^{\phi_j} W_j > x) = E(R_j^{\phi_j}) x^{-1}$.

When $\phi_j > \alpha$, $E(R_j^{\phi_j}) = \infty$. However, the tail of R_j is still regularly varying:

$$\Pr(R_j^{\phi_j} > x) \sim 2C_\alpha \bar{\gamma}_j^\alpha x^{-\frac{\alpha}{\phi_j}}$$

and

$$E\left(W_j^{\frac{\alpha}{\phi_j}}\right) = \int_1^\infty w^{\frac{\alpha}{\phi_j}-2} dw = \frac{1}{1 - \alpha/\phi_j}.$$

From Lemma 10(c),

$$\Pr(R_j^{\phi_j} W_j > x) \sim E\left(W_j^{\frac{\alpha}{\phi_j}}\right) \Pr(R_j^{\phi_j} > x),$$

from which the result follows.

When $\phi_j = \alpha$, $R_j^{\phi_j}$ and W_j are regularly varying with the same index -1 . Therefore, $\log R_j^{\phi_j}$ and $\log W_j$ are both in $ET_{1,0}$. By Lemma 11,

$$\Pr(R_j^{\phi_j} W_j > x) \sim 2C_\alpha \bar{\gamma}_j^\alpha x^{-1} \log x.$$

□

Remark 4. By Theorem 3.8 and Lemma 3.12 of Nolan (2020), we can write out the exact form of the fractional lower order moment of $R_j \sim \text{Stable}(\alpha, 1, \bar{\gamma}_j, \delta)$ if $\delta = 0$:

$$E(R_j^{\phi_j}) = \bar{\gamma}_j^{\phi_j} \cos^{-\frac{\phi_j}{\alpha}} \left(\frac{\pi\alpha}{2} \right) \frac{\Gamma(1 - \phi_j/\alpha)}{\Gamma(1 - \phi_j)}.$$

A.1.3 Limiting angular measure for linear combinations of independent stable variables

Consider

$$R_1 = w_{11}S_1 + \dots + w_{K1}S_K,$$

$$R_2 = w_{12}S_1 + \dots + w_{K2}S_K,$$

$\vdots$

$$R_m = w_{1m}S_1 + \dots + w_{Km}S_K,$$

with deterministic coefficient vectors $\mathbf{w}_i = (w_{1i}, \dots, w_{Ki})^T$, $i = 1, \dots, m$. The random vector $\mathbf{S} = (S_1, \dots, S_K)^T$ is composed of regularly varying variables. In matrix notation, we write $\Psi = (\mathbf{w}_1, \dots, \mathbf{w}_m)^T \in \mathbb{R}^{m \times K}$ and

$$(R_1, \dots, R_m)^T = \Psi \mathbf{S}.$$

To show Proposition 2, we first slightly reformulate Corollary 1 of Cooley and Thibaud (2019) where we consider only non-negative weights in Ψ .

Lemma 12 (Cooley and Thibaud, 2019). Assume the independently and identically distributed random variables $\tilde{S}_k$, $k = 1, \dots, K$ are regularly varying at infinity, i.e.,

$$n \Pr(\tilde{S}_k/b_n > x) \rightarrow x^{-\alpha}, \quad n \rightarrow \infty, \quad (\text{A.3})$$

where $b_n > 0$ and $\alpha > 0$. Also, the normalizing sequence $b_n \rightarrow \infty$, and Ψ is a matrix with m rows, K columns and has only non-negative entries and at least one positive entry. Then the random vector $\Psi \tilde{\mathbf{S}}$ is regularly varying at infinity with tail index α and it has angular measure (in dimension K and with respect to any chosen norm $\|\cdot\|$) given by

$$H_{\Psi \tilde{\mathbf{S}}}(\cdot) = \sum_{k=1}^K \|\boldsymbol{\omega}_k\|^\alpha \times \delta_{\boldsymbol{\omega}_k/\|\boldsymbol{\omega}_k\|}(\cdot), \quad (\text{A.4})$$

where δ is the Dirac mass function, and we set the sum term for k to 0 if $\|\boldsymbol{\omega}_k\| = 0$.

Since $\Psi \tilde{\mathbf{S}}$ exhibits regularly variation, the scaling property holds for a given norm $\|\cdot\|$. Define the unit ball $\mathbb{S}_{m-1}^+ = \{\mathbf{x} \in \mathbb{R}_+^m : \|\mathbf{x}\| = 1\}$. For a set $C(r, B) = \{\mathbf{x} \in \mathbb{R}_+^m : \|\mathbf{x}\| > r, \|\mathbf{x}\|^{-1}\mathbf{x} \in B\}$

$B\}$ with $r > 0$ and $B \subset \mathbb{S}_{m-1}^+$ being a Borel set. Then we have

$$n\Pr(b_n^{-1}\Psi\tilde{\mathbf{S}} \in C(r, B)) \sim r^{-\alpha}H_{\Psi\tilde{\mathbf{S}}}(B), \text{ as } n \rightarrow \infty.$$

If we let $B = \mathbb{S}_{m-1}^+$, the event $b_n^{-1}\Psi\tilde{\mathbf{S}} \in C(r, B)$ is equivalent to $\|\Psi\tilde{\mathbf{S}}\|/b_n > r$. Also note that the angular measure $H_{\Psi\tilde{\mathbf{S}}}$ is usually not a probability measure since its total measure is $m_{\Psi\tilde{\mathbf{S}}} = H_{\Psi\tilde{\mathbf{S}}}(\mathbb{S}_{m-1}^+) = \sum_{k=1}^K \|\omega_k\|^\alpha$. Therefore,

$$n \Pr\left(\|\Psi\tilde{\mathbf{S}}\|/b_n > r\right) \sim m_{\Psi\tilde{\mathbf{S}}}r^{-\alpha}, \text{ as } n \rightarrow \infty. \quad (\text{A.5})$$

Using this result, marginal and joint upper tail behavior of $(R_1, \dots, R_m)^\text{T}$ can be derived when the coefficients in $\mathbf{S}_t$ are regularly varying. Proposition 2 constitutes a special case, focusing on two linear combinations of stable distributions, which are a specific example of regularly varying random variables.

Proof of Proposition 2. (a) The case of $\mathcal{C}_i \cap \mathcal{C}_j = \emptyset$ is clear because independence holds and individually $\Pr(R_i > x) \sim 2\bar{\gamma}_i^\alpha C_\alpha x^{-\alpha}$ and $\Pr(R_j > x) \sim 2\bar{\gamma}_j^\alpha C_\alpha x^{-\alpha}$.

When $\mathcal{C}_i \cap \mathcal{C}_j = \emptyset$, we show Expression (2.11) by applying Lemma 12. First we notice that the upper tails of the stable distributions with the same concentration α and skewness $\beta = 1$ are equivalent up to a positive scaling constant. More specifically, we denote $\tilde{S}_k = S_k/\gamma_k$, $k = 1, \dots, K$, and then all $\tilde{S}_k$'s are iid $\text{Stable}(\alpha, 1, 1, \delta)$ variables because the S_k 's share the location δ in Equation (2.7).

Therefore, Lemma 12 is applicable for the pair $(R_i, R_j)^\text{T}$ under the new weights

$$(w_{1i}\gamma_1, \dots, w_{Ki}\gamma_K)^\text{T} \quad \text{and} \quad (w_{1j}\gamma_1, \dots, w_{Kj}\gamma_K)^\text{T}.$$

Since $\Pr(\tilde{S}_k > x) \sim 2C_\alpha x^{-\alpha}$ as $x \rightarrow \infty$, we may therefore choose

$$b_n = (2nC_\alpha)^{1/\alpha}$$

so $n\Pr(\tilde{S}_k/b_n > x) \rightarrow x^{-\alpha}$ as $n \rightarrow \infty$. Choosing the componentwise min-operator as the norm, Expression (A.5) can be re-written as:

$$n\Pr(\min(R_i, R_j)/b_n > r) \sim r^{-\alpha} \sum_{k=1}^K \min(w_{ki}\gamma_k, w_{kj}\gamma_k)^\alpha = r^{-\alpha} \sum_{k=1}^K w_{k,\wedge}^\alpha \gamma_k^\alpha,$$

in which $w_{k,\wedge} = \min(w_{ki}, w_{kj})$. The previous display immediately induces Equation (2.11).

(b) We only prove the inequality for $\min(c_i R_i^{\phi_i}, c_j R_j^{\phi_j})$. The proof for $\max(c_i R_i^{\phi_i}, c_j R_j^{\phi_j})$ is analogous.

If $\mathcal{C}_i = \mathcal{C}_j$, then R_i and R_j share the same nonzero indices. Since $1/\phi_i - 1/\phi_j > 0$, for sufficiently large x ,

$$\frac{c_i^{1/\phi_i} w_{ki}}{x^{1/\phi_i - 1/\phi_j}} < c_j^{1/\phi_j} w_{kj}, \quad k = 1, \dots, K.$$

Therefore

$$\begin{aligned} \Pr\{\min(c_i R_i^{\phi_i}, c_j R_j^{\phi_j}) > x\} &= P\left(\frac{c_i^{1/\phi_i}}{x^{1/\phi_i - 1/\phi_j}} R_i > x^{1/\phi_j}, c_j^{1/\phi_j} R_j > x^{1/\phi_j}\right) \\ &= P\left(\frac{c_i^{1/\phi_i}}{x^{1/\phi_i - 1/\phi_j}} R_i > x^{1/\phi_j}\right) \sim 2C_\alpha \sum_{k \in \mathcal{C}_i} (w_{ki}\gamma_k)^\alpha c_i^{\alpha/\phi_i} x^{-\alpha/\phi_i}. \end{aligned}$$

If either $\mathcal{C}_i \setminus \mathcal{C}_j$ or $\mathcal{C}_j \setminus \mathcal{C}_i$ is non-empty, we define

$$R_{i,\setminus} = \sum_{k \in \mathcal{C}_i \setminus \mathcal{C}_j} w_{ki} S_k, \quad R_{i,\cap} = \sum_{k \in \mathcal{C}_i \cap \mathcal{C}_j} w_{ki} S_k, \quad (\text{A.6})$$

and $R_i = R_{i,\setminus} + R_{i,\cap}$. Similarly we define $R_{j,\setminus}$ and $R_{j,\cap}$. Then the classic c_r -inequality gives us $R_{i,\cap}^{\phi_i} \leq R_i^{\phi_i} \leq 2^{\phi_i} (R_{i,\cap}^{\phi_i} + R_{i,\setminus}^{\phi_i})$. Similarly, it is also true that $R_{j,\cap}^{\phi_j} \leq R_j^{\phi_j} \leq 2^{\phi_j} (R_{j,\cap}^{\phi_j} + R_{j,\setminus}^{\phi_j})$. Therefore,

$$\begin{aligned} \Pr\{\min(c_i R_i^{\phi_i}, c_j R_j^{\phi_j}) > x\} &\leq P\{\min(c_i R_i^{\phi_i}, c_j R_j^{\phi_j}) > x\} \leq \\ &P\{2^{\phi_i} c_i (R_{i,\cap}^{\phi_i} + R_{i,\setminus}^{\phi_i}) > x, 2^{\phi_j} c_j (R_{j,\cap}^{\phi_j} + R_{j,\setminus}^{\phi_j}) > x\}. \end{aligned} \quad (\text{A.7})$$

The lower bound in (A.7) can be approximated using the first case $\mathcal{C}_i = \mathcal{C}_j$. The upper bound in (A.7) can be further bounded by

$$\Pr(\min(2^{\phi_i} c_i R_{i,\cap}^{\phi_i}, 2^{\phi_j} c_j R_{j,\cap}^{\phi_j}) + \max(2^{\phi_i} c_i R_{i,\setminus}^{\phi_i}, 2^{\phi_j} c_j R_{j,\setminus}^{\phi_j}) > x),$$

which can be approximated using Lemma 13 because

$$\begin{aligned} \Pr\{\min(2^{\phi_i} c_i R_{i,\cap}^{\phi_i}, 2^{\phi_j} c_j R_{j,\cap}^{\phi_j}) > x\} &\sim 2C_\alpha \sum_{k \in \mathcal{C}_i \cap \mathcal{C}_j} (w_{ki} \gamma_k)^\alpha c_i^{\alpha/\phi_i} x^{-\alpha/\phi_i}, \\ \Pr\{\max(2^{\phi_i} c_i R_{i,\setminus}^{\phi_i}, 2^{\phi_j} c_j R_{j,\setminus}^{\phi_j}) > x\} &\sim 2^{\alpha+1} C_\alpha \sum_{k \in \mathcal{C}_i \setminus \mathcal{C}_j} (w_{ki} \gamma_k)^\alpha c_i^{\alpha/\phi_i} x^{-\alpha/\phi_i}. \end{aligned}$$

The second inequality holds due to the independence between $R_{i,\setminus}$ and $R_{j,\setminus}$. Combining the approximations of the two bounds yields the stated range of the constant $d_\wedge$.

□

A.1.4 Proof of Theorem 3

To examine the joint tail of (R_i, R_j) and (X_i, X_j) , we begin by recalling the useful results from the literature. The first result is an easy but useful inequality

$$\begin{aligned} \Pr(\min(R_1, R_2) + \min(W_1, W_2) > x) &\leq \Pr(R_1 + W_1 > x, R_2 + W_2 > x) \leq \\ &\Pr(\min(R_1, R_2) + \max(W_1, W_2) > x). \end{aligned} \tag{A.8}$$

The second relates to the tail behavior on the convolution with a CE_α distribution.

Lemma 13 (Theorem 1 of Cline (1986) and Lemma 5.1 of Pakes (2004)). Let $Y_1 \sim F_1$ and $Y_2 \sim F_2$ be random variables. If distribution function $F_1 \in \text{CE}_\alpha$ with $\alpha \geq 0$ and $E(e^{\alpha Y_2}) < \infty$ while $\Pr(Y_2 > x)/\Pr(Y_1 > x) \rightarrow c \geq 0$ as $x \rightarrow \infty$, then

$$\Pr(Y_1 + Y_2 > x)/\Pr(Y_1 > x) \rightarrow E(e^{\alpha Y_2}) + cE(e^{\alpha Y_1}), \quad x \rightarrow \infty.$$

Lemma 14 (Proposition 5 in Engelke et al. (2019)). Suppose $\bar{F}_W \in RV_{-\alpha_W}$ with $\alpha_W \geq 0$, and $\bar{F}_R \in RV_{-\alpha_R}$ with $\alpha_R > \alpha_W$. Let $R \sim F_R$ and $(W_1, W_2) \stackrel{d}{=} F_W$ marginally while $R \perp\!\!\!\perp (W_1, W_2)$. Denote $(X_1, X_2) = R(W_1, W_2)$. If the extremal dependence of (W_1, W_2) is summarized by (χ_W, η_W) , then $\chi_X = \chi_W$ and

$$\eta_X = \begin{cases} \alpha_W/\alpha_R, & \text{if } \alpha_R < \alpha_W/\eta_W, \\ \eta_W, & \text{if } \alpha_R > \alpha_W/\eta_W. \end{cases}$$

Proof of Theorem 3. (a) Since $\phi_i > \alpha$ and $\phi_j > \alpha$, (2.10) ensures

$$F_{X_i}^{-1}(1-q) \sim \{(1 - \alpha/\phi_i)\theta_i\}^{-\phi_i/\alpha} q^{-\phi_i/\alpha} \text{ with } \theta_i = \frac{1}{2C_\alpha \bar{\gamma}_i^\alpha} > 0.$$

Similar result holds for $F_{X_j}^{-1}(1-q)$. Therefore

$$\begin{aligned} & P(X_i \geq F_{X_i}^{-1}(1-q), X_j \geq F_{X_j}^{-1}(1-q)) \\ &= P\left(R_i^{\phi_i} W_i > \{(1 - \alpha/\phi_i)\theta_i\}^{-\phi_i/\alpha} q^{-\phi_i/\alpha}, R_j^{\phi_j} W_j > \{(1 - \alpha/\phi_j)\theta_j\}^{-\phi_j/\alpha} q^{-\phi_j/\alpha}\right) \\ &= P\left(\theta_i R_i^\alpha \frac{W_i^{\alpha/\phi_i}}{E(W_i^{\alpha/\phi_i})} > q^{-1}, \theta_j R_j^\alpha \frac{W_j^{\alpha/\phi_j}}{E(W_j^{\alpha/\phi_j})} > q^{-1}\right). \end{aligned} \tag{A.9}$$

Using Expression (A.8) and Lemma 10 part (c) again, we deduce that the right-hand side of (A.9) is bounded with the range

$$E\left\{\min\left(\frac{W_i^{\alpha/\phi_i}}{E(W_i^{\alpha/\phi_i})}, \frac{W_j^{\alpha/\phi_j}}{E(W_j^{\alpha/\phi_j})}\right)\right\} \left[\Pr(\min(\theta_i R_i^\alpha, \theta_j R_j^\alpha) > q^{-1}), \Pr(\max(\theta_i R_i^\alpha, \theta_j R_j^\alpha) > q^{-1})\right].$$

Meanwhile, we know from Proposition 2(a) that

$$\begin{aligned} \Pr(\min(\theta_i R_i^\alpha, \theta_j R_j^\alpha) > q^{-1}) &= \Pr(\theta_i^{1/\alpha} R_i > q^{-1/\alpha}, \theta_j^{1/\alpha} R_j > q^{-1/\alpha}) \\ &= 2C_\alpha C_K(\theta_i^{1/\alpha} \mathbf{w}_i, \theta_j^{1/\alpha} \mathbf{w}_j, \gamma)q, \end{aligned} \tag{A.10}$$

in which

$$\begin{aligned}
C_K(\theta_i^{1/\alpha} \mathbf{w}_i, \theta_j^{1/\alpha} \mathbf{w}_j, \gamma) &= \sum_{k=1}^K \min\{\theta_i \gamma_k^\alpha w_{ki}^\alpha, \theta_j \gamma_k^\alpha w_{kj}^\alpha\} \\
&= \frac{1}{2C_\alpha} \sum_{k=1}^K \min\left\{\frac{\gamma_k^\alpha w_{ki}^\alpha}{\bar{\gamma}_i^\alpha}, \frac{\gamma_k^\alpha w_{kj}^\alpha}{\bar{\gamma}_j^\alpha}\right\} \\
&= \frac{1}{2C_\alpha} \sum_{k=1}^K \min\left\{\frac{(w_{ki} \gamma_k)^\alpha}{\sum_{\ell \in \mathcal{C}_i} (w_{\ell i} \gamma_\ell)^\alpha}, \frac{(w_{kj} \gamma_k)^\alpha}{\sum_{\ell \in \mathcal{C}_j} (w_{\ell j} \gamma_\ell)^\alpha}\right\} \\
&= \frac{1}{2C_\alpha} \sum_{k \in \mathcal{C}_i \cap \mathcal{C}_j} v_{k, \wedge}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\Pr(\max(\theta_i R_i^\alpha, \theta_j R_j^\alpha) > q^{-1}) &= 1 - \Pr(\max(\theta_i R_i^\alpha, \theta_j R_j^\alpha) \leq q^{-1}) \\
&= \Pr(\theta_i R_i^\alpha > q^{-1}, \theta_j R_j^\alpha < q^{-1}) \\
&\quad + \Pr(\theta_i R_i^\alpha < q^{-1}, \theta_j R_j^\alpha > q^{-1}) \\
&\quad + \Pr(\theta_i R_i^\alpha > q^{-1}, \theta_j R_j^\alpha > q^{-1}).
\end{aligned}$$

By Proposition 2(a) and Expression (A.10), the previous display becomes

$$\Pr(\max(\theta_i R_i^\alpha, \theta_j R_j^\alpha) > q^{-1}) = 2C_\alpha C_K(\theta_i^{1/\alpha} \mathbf{w}_i, \theta_j^{1/\alpha} \mathbf{w}_j, \gamma)q + o(q).$$

Therefore, both $\Pr(\min(\theta_i R_i^\alpha, \theta_j R_j^\alpha) > q^{-1})$ and $\Pr(\max(\theta_i R_i^\alpha, \theta_j R_j^\alpha) > q^{-1})$ are dominated by $2C_\alpha C_K(\theta_i^{1/\alpha} \mathbf{w}_i, \theta_j^{1/\alpha} \mathbf{w}_j, \gamma)q$. Consequently, we get $\eta_X = 1$ and

$$\chi_{ij} = E \left\{ \min \left(\frac{W_i^{\alpha/\phi_i}}{E(W_i^{\alpha/\phi_i})}, \frac{W_j^{\alpha/\phi_j}}{E(W_j^{\alpha/\phi_j})} \right) \right\} \sum_{k=1}^K v_{k, \wedge}.$$

(b) When $0 < \phi_i < \phi_j < \alpha$, write $\lambda_i = E(R_i^{\phi_i})$ and $\lambda_j = E(R_j^{\phi_j})$. By (2.10),

$$F_{X_i}^{-1}(1 - q) \sim \lambda_i q^{-1}, \quad F_{X_j}^{-1}(1 - q) \sim \lambda_j q^{-1},$$

and

$$\Pr(X_i \geq F_{X_i}^{-1}(1-q), X_j \geq F_{X_j}^{-1}(1-q)) = P\left(\frac{R_i^{\phi_i}}{\lambda_i} W_i > q^{-1}, \frac{R_j^{\phi_j}}{\lambda_j} W_j > q^{-1}\right).$$

If $\phi_i/\alpha < \phi_j/\alpha < \eta_W$, we first regard $R^* = \max(R_i^{\phi_i}/\lambda_i, R_j^{\phi_j}/\lambda_j)$ as a radial variable. From Proposition 2(b), we have $\Pr(R^* > x) \in RV_{-\alpha/\phi_j}$ and $\alpha_{R^*} = \alpha/\phi_j > \alpha_W = 1$. Since $\alpha/\phi_j > 1/\eta_W$, we know from Lemma 14 that

$$\lim_{q \rightarrow 0} \frac{\log \Pr(R^* W_i > q^{-1}, R^* W_j > q^{-1})}{\log \Pr(R^* W_i > q^{-1})} = \eta_W.$$

Then we regard $R_* = \min(R_i^{\phi_i}/\lambda_i, R_j^{\phi_j}/\lambda_j)$ as a radial variable. Since $\alpha_{R_*} = \alpha/\phi_i > 1/\eta_W$, Lemma 14 again gives

$$\lim_{q \rightarrow 0} \frac{\log \Pr(R_* W_i > q^{-1}, R_* W_j > q^{-1})}{\log \Pr(R_* W_i > q^{-1})} = \eta_W. \quad (\text{A.11})$$

Moreover, $E(R^*) < \infty$ and $E(R_*) < \infty$ due to $\phi_i < \phi_j < \alpha$. We can apply Lemma 10 part (c) again to show $\log \Pr(R^* W_i > q^{-1}) \sim \log \Pr(R_* W_i > q^{-1}) \sim \log \Pr(R_i^{\phi_i} W_i > \lambda_i q^{-1})$. By sandwich limit theorem and (A.8),

$$\eta_X = \lim_{q \rightarrow 0} \frac{\log \Pr(R_i^{\phi_i} W_i > \lambda_i q^{-1}, R_i^{\phi_i} W_j > \lambda_i q^{-1})}{\log \Pr(R_i^{\phi_i} W_i > \lambda_i q^{-1})} = \eta_W,$$

and $\chi_X = \chi_W = 0$.

If $\eta_W < \phi_i/\alpha < \phi_j/\alpha$, we have $\alpha/\phi_j < \alpha/\phi_i < 1/\eta_W$. Lemma 14 ensures

$$\lim_{q \rightarrow 0} \frac{\log \Pr(R^* W_i > q^{-1}, R^* W_j > q^{-1})}{\log \Pr(R^* W_i > q^{-1})} = \alpha_W/\alpha_{R^*} = \phi_j/\alpha, \quad (\text{A.12})$$

and

$$\lim_{q \rightarrow 0} \frac{\log \Pr(R_* W_i > q^{-1}, R_* W_j > q^{-1})}{\log \Pr(R_* W_i > q^{-1})} = \alpha_W/\alpha_{R_*} = \phi_i/\alpha.$$

Therefore, $\eta_X \in [\phi_i/\alpha, \phi_j/\alpha]$ and $\chi_X = \chi_W = 0$.

If $\phi_i/\alpha < \eta_W < \phi_j/\alpha$, we have $\alpha/\phi_j < 1/\eta_W < \alpha/\phi_i$. Therefore, (A.12) holds for R^* and (A.11) holds for R_* , which proves $\eta_X \in [\eta_W, \phi_j/\alpha]$ and $\chi_X = 0$.

(c) When $\phi_i < \alpha < \phi_j$, write $\lambda_i = E(R_i^{\phi_i})$ and $\psi_j = (1 - \alpha/\phi_j)/\{2C_\alpha \bar{\gamma}_i^\alpha\}$. By (2.10),

$$F_{X_i}^{-1}(1 - q) \sim \lambda_i q^{-1}, \quad F_{X_j}^{-1}(1 - q) \sim (\psi_j q)^{-\phi_j/\alpha}.$$

First, we assume $\mathbf{w}_i = \mathbf{w}_j$, i.e., $R_i = R_j$. Denote the distribution and density function of R_i and R_j as F_R and f_R , and then

$$F_R = \text{Stable}(\alpha, 1, \bar{\gamma}, 0), \quad \bar{\gamma} = \left\{ \sum_{k \in \mathcal{C}_i} (w_{ki} \gamma_k)^\alpha \right\}^{\frac{1}{\alpha}}.$$

Thus,

$$\begin{aligned} \Pr(X_i \geq F_{X_i}^{-1}(1 - q), X_j \geq F_{X_j}^{-1}(1 - q)) &= \Pr(R_i^{\phi_i} W_i > \lambda_i q^{-1}, R_j^{\phi_j} W_j > (\psi_j q)^{-\phi_j/\alpha}) \\ &= \int_0^\infty P(r^{\phi_i} W_i > \lambda_i q^{-1}, r^{\phi_j} W_j > (\psi_j q)^{-\phi_j/\alpha}) f_R(r) dr. \end{aligned} \quad (\text{A.13})$$

Since $q^{-1/\phi_j} > q^{-1/\alpha}$ for sufficiently small q , we can split the limits of the integral in (A.13) into $(0, (\theta_j^* q)^{-1/\alpha})$, $((\theta_j^* q)^{-1/\alpha}, (\lambda_i q^{-1})^{1/\phi_i})$, and $((\lambda_i q^{-1})^{1/\phi_i}, \infty)$.

(i) When $r \in ((\lambda_i q^{-1})^{1/\phi_i}, \infty)$, we have $\lambda_i q^{-1}/r^{\phi_i} < 1$ and $(\psi_j q)^{-\phi_j/\alpha}/r^{\phi_j} < 1$. Thus,

$$\begin{aligned} &\int_{(\lambda_i q^{-1})^{1/\phi_i}}^\infty P(r^{\phi_i} W_i > \lambda_i q^{-1}, r^{\phi_j} W_j > (\psi_j q)^{-\phi_j/\alpha}) f_R(r) dr \\ &= \Pr(R_i > (\lambda_i q^{-1})^{1/\phi_i}) \sim 2C_\alpha \bar{\gamma}^\alpha \lambda_i^{-\alpha/\phi_i} q^{\alpha/\phi_i}. \end{aligned} \quad (\text{A.14})$$

(ii) When $r \in ((\theta_j^* q)^{-1/\alpha}, (\lambda_i q^{-1})^{1/\phi_i})$, $\lambda_i q^{-1}/r^{\phi_i} > 1$ and $(\psi_j q)^{-\phi_j/\alpha}/r^{\phi_j} < 1$. Thus,

$$\int_{(\theta_j^* q)^{-1/\alpha}}^{(\lambda_i q^{-1})^{1/\phi_i}} P(r^{\phi_i} W_i > \lambda_i q^{-1}, r^{\phi_j} W_j > (\psi_j q)^{-\phi_j/\alpha}) f_R(r) dr$$

$$\begin{aligned}
&= \int_{(\theta_j^* q)^{-1/\alpha}}^{(\lambda_i q^{-1})^{1/\phi_i}} \Pr(r^{\phi_i} W_i > \lambda_i q^{-1}) f_R(r) dr = \lambda_i^{-1} q \int_{(\theta_j^* q)^{-1/\alpha}}^{(\lambda_i q^{-1})^{1/\phi_i}} r^{\phi_i} f_R(r) dr \quad (\text{A.15}) \\
&= \lambda_i^{-1} q \int_{(\theta_j^* q)^{-1/\alpha}}^{(\lambda_i q^{-1})^{1/\phi_i}} \frac{1}{\pi \bar{\gamma}} \sum_{m=1}^{\infty} \frac{\Gamma(m\alpha + 1) \sin(m\pi\alpha) (-1)^{m+1}}{m!} \\
&\quad \times \left(\cos \frac{\pi\alpha}{2} \right)^{-m} \frac{r^{\phi_i - m\alpha - 1}}{\bar{\gamma}^{-m\alpha - 1}} dr \\
&\sim 2\Gamma(\alpha) \sin(\pi\alpha/2) \bar{\gamma}^\alpha \frac{\lambda_i^{-1} \psi_j^{1 - \phi_i/\alpha}}{\pi(1 - \phi_i/\alpha)} q^{2 - \phi_i/\alpha} = \frac{\lambda_i^{-1} \psi_j^{-\phi_i/\alpha} (1 - \alpha/\phi_j)}{1 - \phi_i/\alpha} q^{2 - \phi_i/\alpha}.
\end{aligned}$$

The penultimate line uses the series expansion for the stable density when $\alpha \neq 1$; see Zolotarev (1986, Chapter 2).

(iii) When $r \in (0, (\theta_j^* q)^{-1/\alpha})$, we have

$$\begin{aligned}
&\int_0^{(\theta_j^* q)^{-1/\alpha}} P(r^{\phi_i} W_i > \lambda_i q^{-1}, r^{\phi_j} W_j > (\psi_j q)^{-\phi_j/\alpha}) f_R(r) dr \\
&\leq \int_0^{(\theta_j^* q)^{-1/\alpha}} \exp \left[-\frac{\log(\lambda_i q^{-1}/r^{\phi_i}) + \log\{(\psi_j q)^{-\phi_j/\alpha}/r^{\phi_j}\}}{1 + \rho} \right] f_R(r) dr \quad (\text{A.16}) \\
&= (\psi_j^{-\phi_j/\alpha} \lambda_i)^{-1/(1+\rho)} q^{(1+\phi_j/\alpha)/(1+\rho)} \int_0^{(\theta_j^* q)^{-1/\alpha}} r^{(\phi_i + \phi_j)/(1+\rho)} f_R(r) dr.
\end{aligned}$$

Note that

$$r^{(\phi_i + \phi_j)/(1+\rho)} f_R(r) \rightarrow 0, \quad r \rightarrow 0,$$

and

$$r^{(\phi_i + \phi_j)/(1+\rho)} f_R(r) \sim r^{(\phi_i + \phi_j)/(1+\rho) - \alpha - 1}, \quad r \rightarrow \infty.$$

By Karamata's Theorem (Resnick, 2008, p.17),

$$\int_0^x r^{(\phi_i + \phi_j)/(1+\rho)} f_R(r) dr \in RV_{(\phi_i + \phi_j)/(1+\rho) - \alpha}$$

when $(\phi_i + \phi_j)/(1 + \rho) \geq \alpha$. When $(\phi_i + \phi_j)/(1 + \rho) < \alpha$, $\int_0^x r^{(\phi_i + \phi_j)/(1 + \rho)} f_R(r) dr < E\{R_i^{(\phi_i + \phi_j)/(1 + \rho)}\} < \infty$. Apply this result to the right-hand side of (A.16) to get

$$\begin{aligned} & \int_0^{(\theta_j^* q)^{-1/\alpha}} P(r^{\phi_i} W_i > \lambda_i q^{-1}, r^{\phi_j} W_j > (\psi_j q)^{-\phi_j/\alpha}) f_R(r) dr \\ & \leq \begin{cases} L_R(q^{-1}) q^{(1 - \phi_i/\alpha)/(1 + \rho) + 1}, & \text{if } (\phi_i + \phi_j)/(2\eta_W) \geq \alpha, \\ C_R q^{(1 + \phi_j/\alpha)/(1 + \rho)}, & \text{if } (\phi_i + \phi_j)/(2\eta_W) < \alpha, \end{cases} \end{aligned}$$

where $L_R \in RV_0$ and $C_R = (\psi_j^{-\phi_j/\alpha} \lambda_i)^{-1/(1 + \rho)} E\{R_i^{(\phi_i + \phi_j)/(1 + \rho)}\}$.

On the other hand,

$$\begin{aligned} & \int_0^{(\theta_j^* q)^{-1/\alpha}} P(r^{\phi_i} W_i > \lambda_i q^{-1}, r^{\phi_j} W_j > (\psi_j q)^{-\phi_j/\alpha}) f_R(r) dr \\ & \geq \int_0^{(\theta_j^* q)^{-1/\alpha}} P(r^{\phi_i} W_i > \lambda_i q^{-1}) P(r^{\phi_j} W_j > (\psi_j q)^{-\phi_j/\alpha}) f_R(r) dr \quad (\text{A.17}) \\ & = \psi_j^{\phi_j/\alpha} \lambda_i^{-1} q^{1 + \phi_j/\alpha} \int_0^{(\theta_j^* q)^{-1/\alpha}} r^{\phi_i + \phi_j} f_R(r) dr = \tilde{L}_R(q^{-1}) q^{2 - \phi_i/\alpha}, \end{aligned}$$

in which $\tilde{L}_R \in RV_0$.

Combine the results from (A.14) - (A.17) under the assumption that $\mathbf{w}_i = \mathbf{w}_j$, and we can obtain lower and upper bounds of the regularly varying index for $\Pr(X_i \geq F_{X_i}^{-1}(1 - q), X_j \geq F_{X_j}^{-1}(1 - q))$, which is also known as η_X .

When $\mathbf{w}_i \neq \mathbf{w}_j$ and $\mathcal{C}_i \cap \mathcal{C}_j \neq \emptyset$,

$$\begin{aligned} & P \left[\left(\sum_k w_{k, \wedge} S_k \right)^{\phi_i} W_i > \lambda_i q^{-1}, \left(\sum_k w_{k, \wedge} S_k \right)^{\phi_j} W_j > (\psi_j q)^{-\phi_j/\alpha} \right] \\ & \leq \Pr(R_i^{\phi_i} W_i > \lambda_i q^{-1}, R_j^{\phi_j} W_j > (\psi_j q)^{-\phi_j/\alpha}) \\ & \leq P \left[\left(\sum_k w_{k, \vee} S_k \right)^{\phi_i} W_i > \lambda_i q^{-1}, \left(\sum_k w_{k, \vee} S_k \right)^{\phi_j} W_j > (\psi_j q)^{-\phi_j/\alpha} \right], \end{aligned}$$

whose bounds can be dealt with the results we just obtained while assuming $\mathbf{w}_i = \mathbf{w}_j$. Since the bounds for η_X did not depend on the weights, they stay the same for when $\mathbf{w}_i \neq \mathbf{w}_j$. $\square$

A.2 MCMC details

Conditioning on the scaling variable $\mathbf{S}_t$ at the knots, we define the hierarchical model as

$$\begin{aligned}\mathcal{L}(\mathbf{Y}_t | \boldsymbol{\theta}_{\mathbf{Y}_t}, \mathbf{S}_t, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\rho}) &= \varphi_D(\mathbf{Z}_t) \left| \frac{\partial \mathbf{Z}_t}{\partial \mathbf{Y}_t} \right| \\ \mathbf{S}_t | \boldsymbol{\gamma} &\sim \text{Stable}(\alpha = 0.5, 1, \boldsymbol{\gamma}, \delta = 0) \\ \phi_k &\stackrel{\text{iid}}{\sim} \text{Beta}(5, 5), \\ \rho_k &\stackrel{\text{iid}}{\sim} \text{halfNormal}(0, 2), \\ k &= 1, \dots, K\end{aligned}$$

in which φ_D is the D -variate Gaussian density function with covariance matrix $\boldsymbol{\Sigma}_\rho$ and $\partial \mathbf{Z}_t / \partial \mathbf{Y}_t$ is the Jacobian. Using the inverse function theorem that the derivative of $F^{-1}(t)$ is equal to $1/F'(F^{-1}(t))$ as long as $F'(F^{-1}(t)) \neq 0$, the Jacobian is a diagonal matrix with elements

$$\begin{aligned}\frac{\partial Z_{tj}}{\partial Y_{tj}} &= \frac{1}{\varphi \left\{ \Phi^{-1} \left(1 - \frac{1}{X_{tj}/R_{tj}^{\phi_j} + 1} \right) \right\}} \cdot \frac{1}{(X_{tj}/R_{tj}^{\phi_j} + 1)^2 R_{tj}^{\phi_j}} \cdot \frac{\partial X_{tj}}{\partial Y_{tj}} \\ &= \frac{1}{\varphi(Z_{tj})} \cdot \frac{1}{(X_{tj}/R_{tj}^{\phi_j} + 1)^2 R_{tj}^{\phi_j}} \cdot \frac{f_Y(Y_{tj})}{f_X(X_{tj})},\end{aligned}$$

in which φ and Φ are, respectively, the density and the distribution function of a univariate standard Gaussian, $X_{tj} = F_X^{-1} \circ F_Y(Y_{tj})$, $Z_{tj} = g^{-1}(X_{tj}/R_{tj}^{\phi_j}) = \Phi^{-1}(1 - 1/(X_{tj}/R_{tj}^{\phi_j} + 1))$, f_Y is the marginal density of the observed data distribution, and f_X is the univariate density of the dependence model derived in (2.17). Then, as we have assumed temporal independence by introducing marginal temporal parameter, likelihoods across the independent time replicates are multiplied together for the joint likelihood.

A.3 Additional Results on Simulation Study

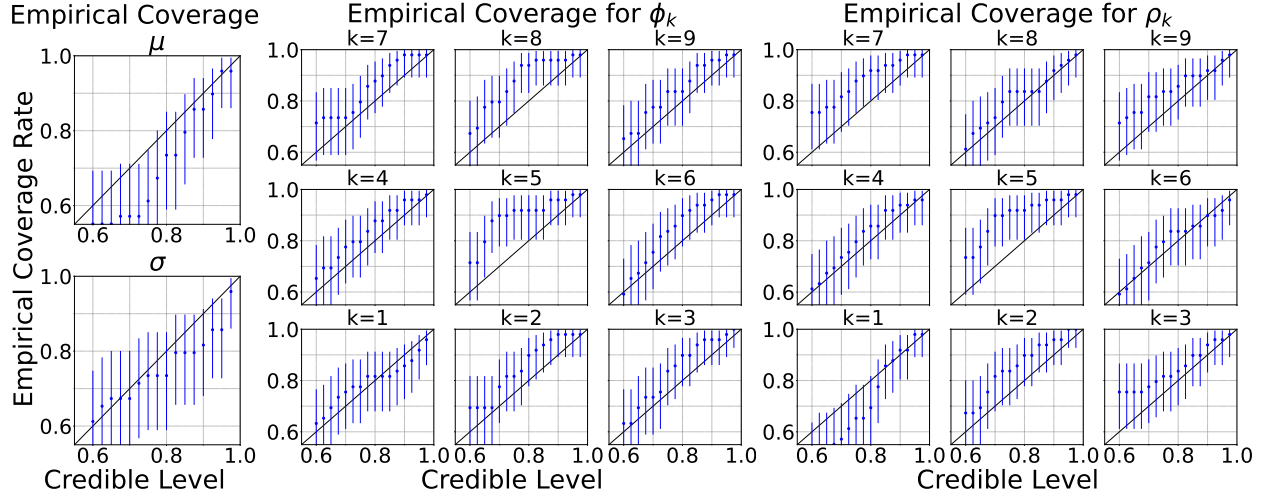


Figure A.1: Empirical coverage rates of credible intervals of the marginal parameters μ and σ (left), the dependence parameters ϕ_k , $k = 1, \dots, 9$ (middle), and ρ_k , $k = 1, \dots, 9$ (right), in simulation scenario 2.

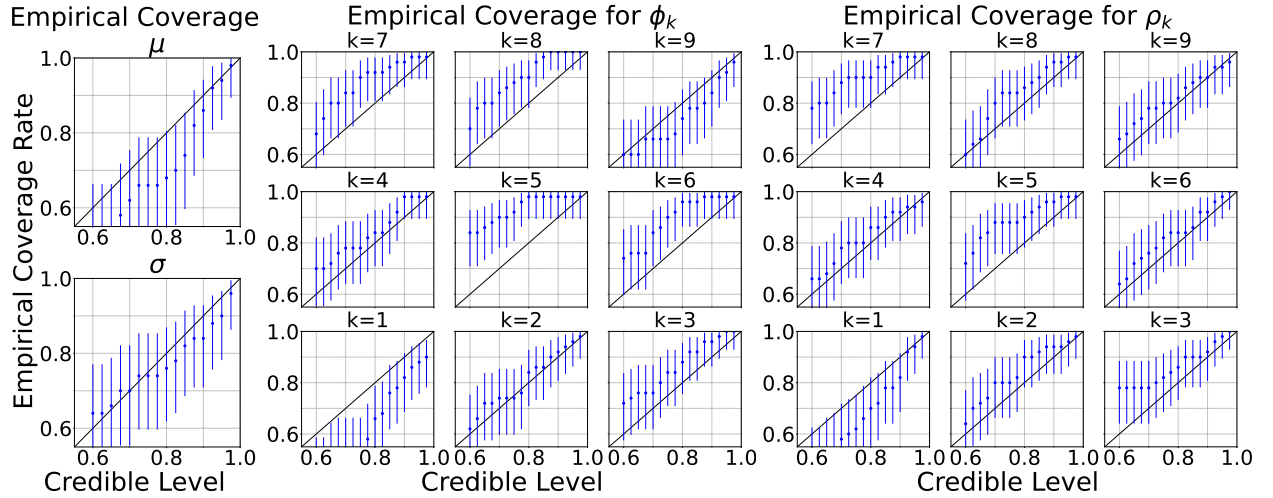


Figure A.2: Empirical coverage rates of credible intervals of the marginal parameters μ and σ (left), the dependence parameters ϕ_k , $k = 1, \dots, 9$ (middle), and ρ_k , $k = 1, \dots, 9$ (right), in simulation scenario 3.

A.4 Justification for Marginal Model Choices

To build our model for the marginal GEV parameters μ , σ , and ξ , we perform a suite of exploratory analyses. First, we fit GEV distributions by maximum likelihood to the annual maxima

at each station, independently. Because the resulting fits are very noisy, to reveal any underlying trends we next smooth the estimated parameters using a bivariate cubic spline as implemented in `Scipy's SmoothBivariateSpline` (Virtanen et al., 2020).

In Figure A.3, we compare the smoothed surfaces of the MLEs of $\log(\sigma)$ and ξ with the negative of elevation at each observation location. We observe similar spatial structure between smoothed estimates of $\log(\sigma)$ and elevation (there is not much spatial structure in the smoothed estimates of ξ at all), so we conclude that elevation is sufficient to use as the sole covariate for $\log(\sigma)$ and ξ . Furthermore, consistent with previous work on precipitation in this region showing no gain from allowing σ and ξ to vary temporally (e.g. Westra et al., 2013; Risser et al., 2019), we specify a time trend in only the location parameter μ .

To model the temporal trend in μ , previous studies have used either the natural logarithm of CO_2 concentrations (Risser and Wehner, 2017), the radiative forcing associated with CO_2 concentrations (which is logarithmic in concentrations; see Etminan et al., 2016), or the sum-total radiative forcing from the five primary greenhouse gases (Risser et al., 2024), making any one of these a natural choice. However, each of these variables is nearly linear over our study period of 1949–2023 (correlations in excess of 0.987; see the figure below). Therefore, our results would be the same regardless of the specific covariate used.

Additionally, we could have used other covariates in the time trend model that describe the effect of large-scale climate drivers on extreme precipitation in the central US, such as ENSO. However, previous work has shown that these drivers have a minimal effect on seasonal extreme precipitation in the central US (see Figure 1 of Risser et al., 2021)

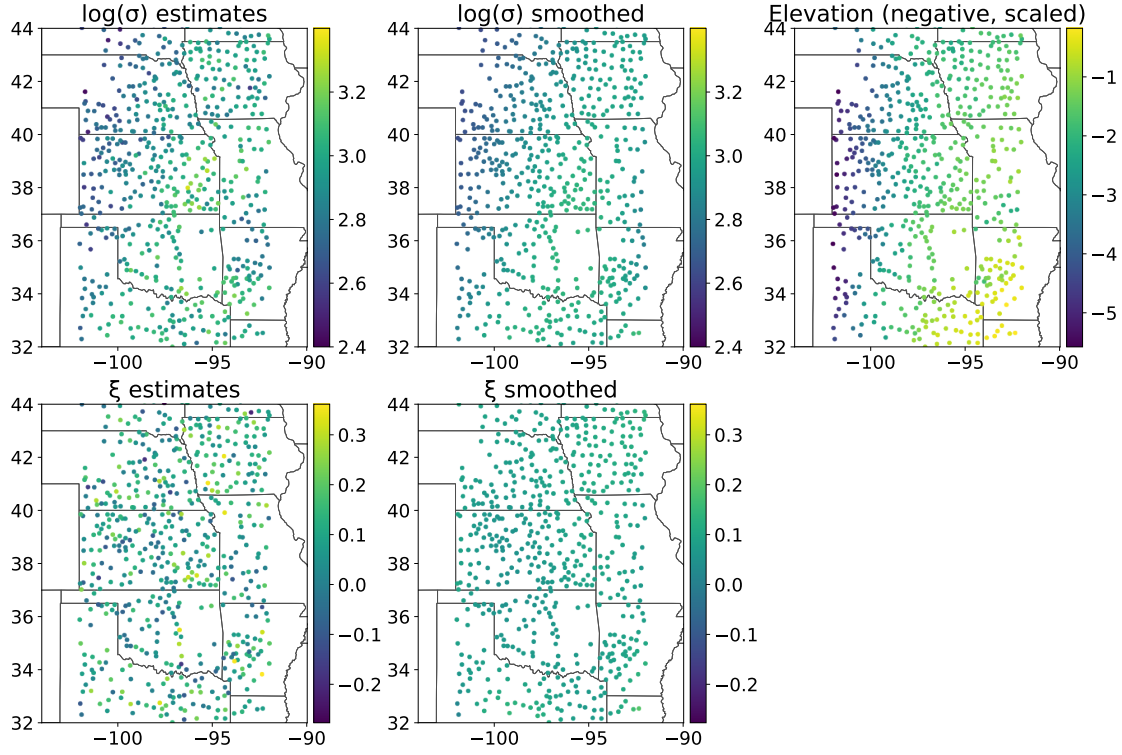


Figure A.3: Estimates and smoothed surfaces of maximum likelihood estimates of $\log(\sigma)$ and ξ , fit to annual maxima at each station independently. The smoothed estimates of $\log(\sigma)$ show similar spatial patterns as the negative of scaled elevation at each station (rightmost panel).

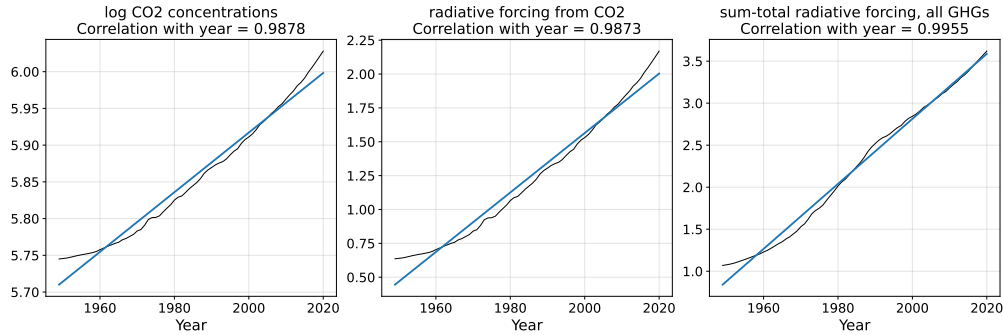
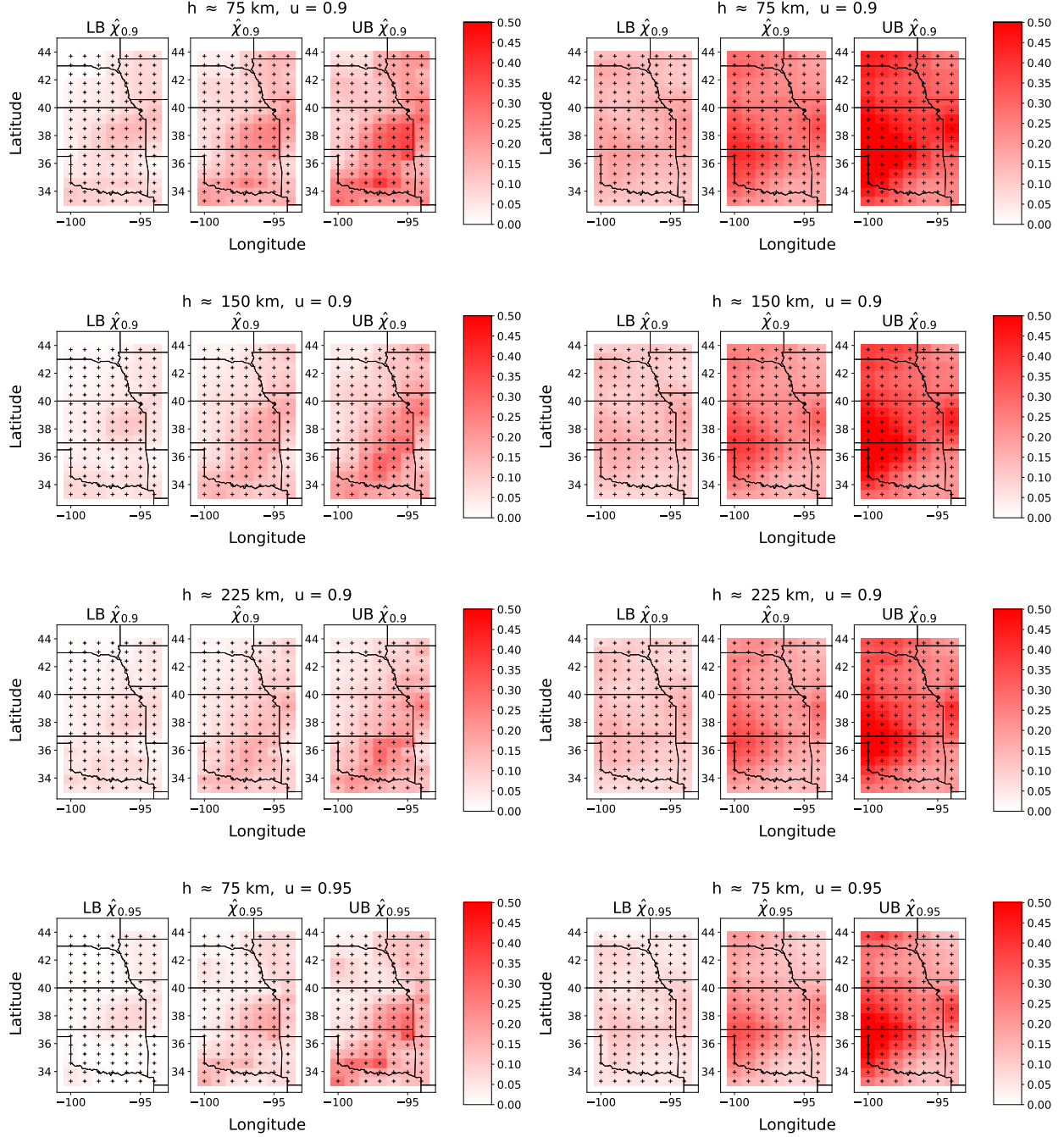


Figure A.4: CO₂ concentrations (and its functions) against time in black curve. Least square fit in Blue.

A.5 Confidence Intervals for the Empirical and Model-based $\hat{\chi}$

The figures below show approximate confidence intervals for the empirical and model realized χ surfaces. Across spatial lags, the model's χ surfaces generally fall within the empirical confidence bands at higher quantiles (e.g., $u=0.99$), suggesting good agreement in the upper tail. At

lower quantiles (e.g., $u=0.9$), however, the model does seem to estimate slightly higher dependence. The model's 95% confidence interval for χ does seem to overlap with the intervals from the empirical estimates. Empirical $\hat{\chi}$ plots are in the *left* column, and those from model realizations are in the *right* column.



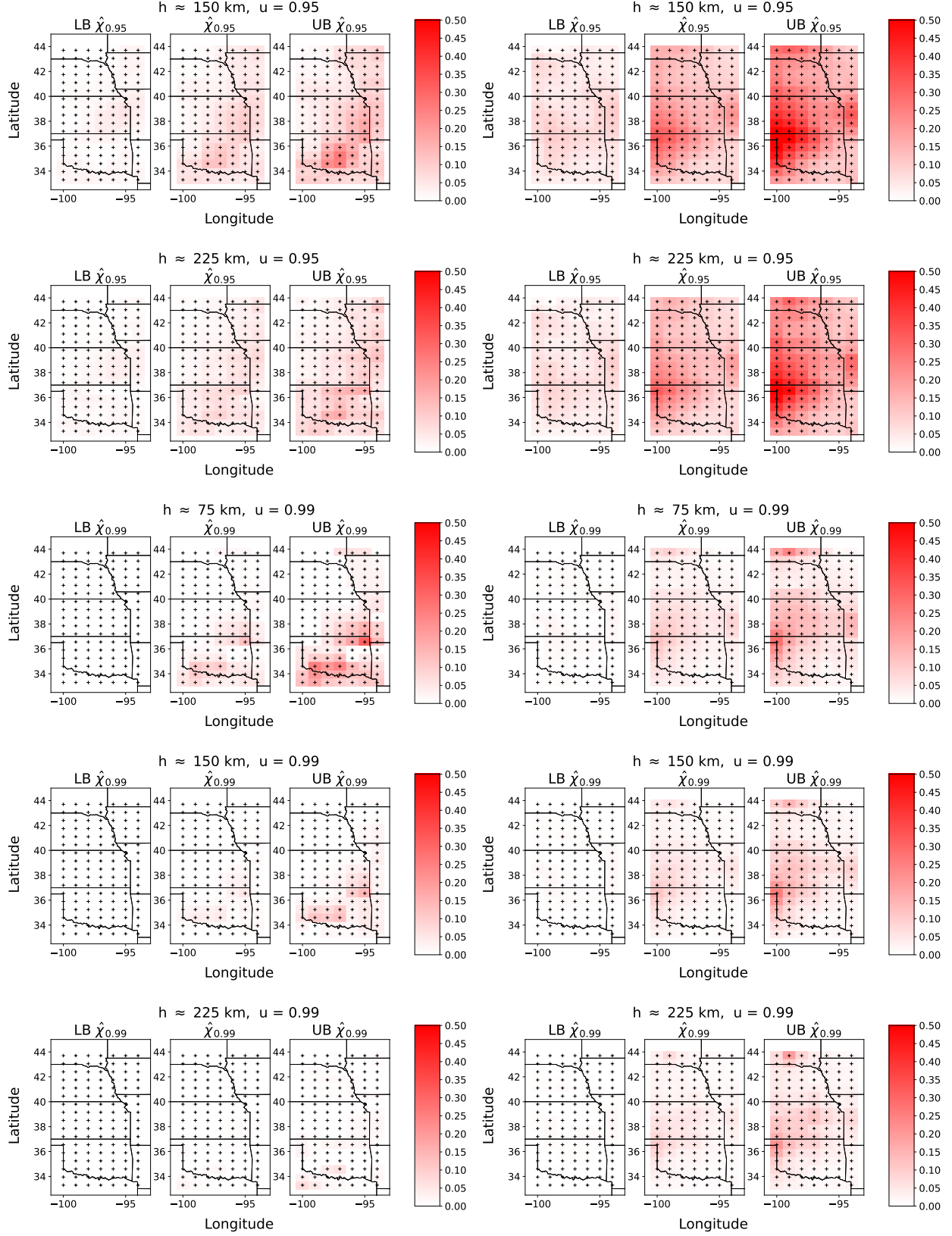


Figure A.5: Empirical (left) and model-realized (right) χ surfaces with 95% confidence bands.

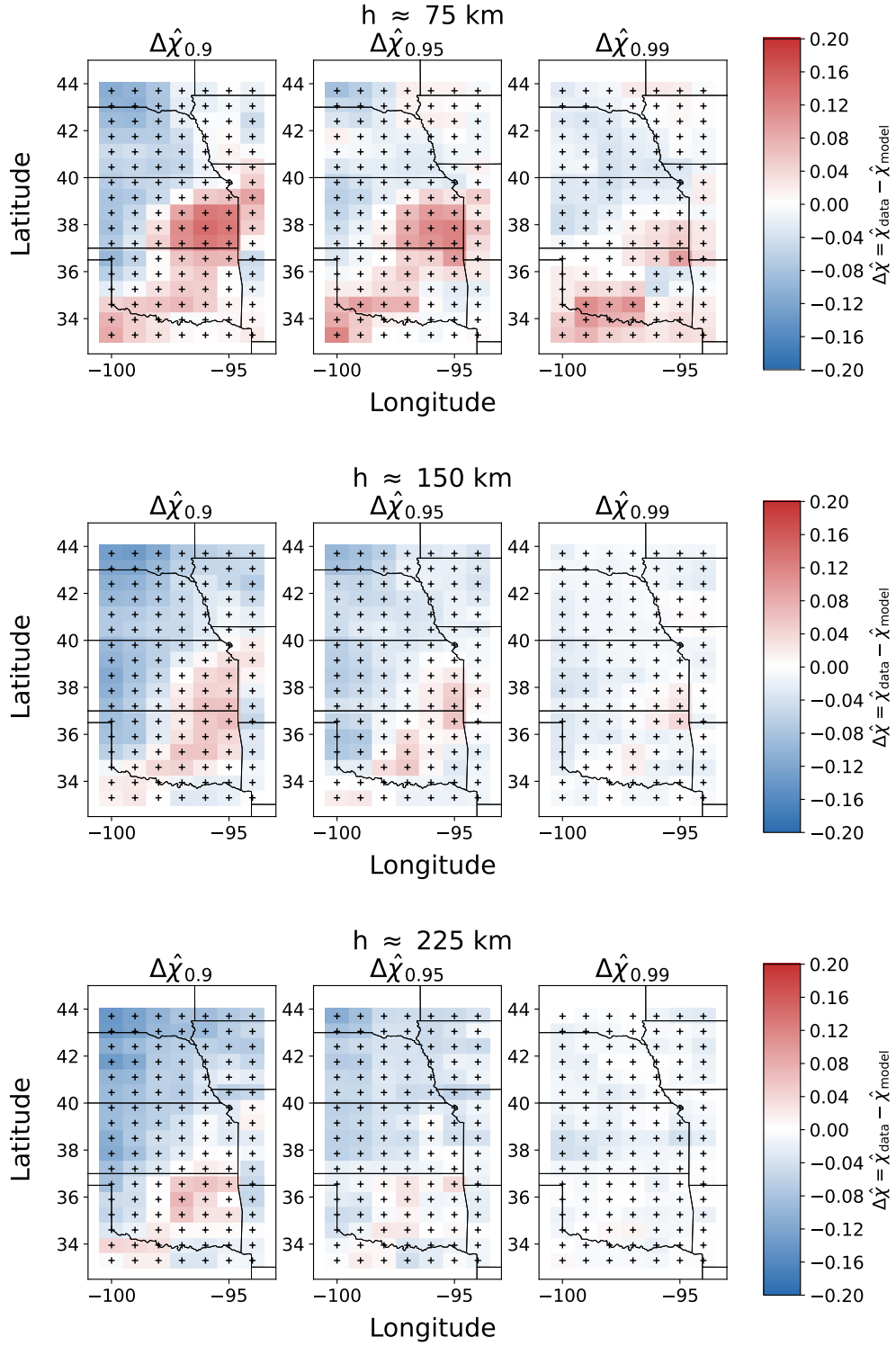


Figure A.6: Spatial maps displaying the difference between the empirical and modeled quantities using a diverging color scheme centered at zero. The differences are largely diminished at the most extreme quantile level $u = 0.99$.

Appendix B

Supplementary materials for Chapter 3

B.1 Log-Laplace Nugget Parameterization

We record the parameterization of the log-Laplace nugget used throughout the appendix. Let ϵ be a Log-Laplace($0, 1/\alpha_0$) variable, parametrized as

$$\Pr\{\epsilon \leq x\} = \begin{cases} \frac{1}{2} \exp\{\alpha_0 \log x\}, & 0 < x \leq 1, \\ 1 - \frac{1}{2} \exp\{-\alpha_0 \log x\}, & x > 1, \end{cases}$$

where $x > 0$ and $\alpha_0 > 0$; the corresponding density function is

$$f_\epsilon(x) = \begin{cases} \frac{\alpha_0}{2} x^{\alpha_0-1}, & 0 < x \leq 1 \\ \frac{\alpha_0}{2} x^{-\alpha_0-1}, & 1 < x \end{cases}$$

For intuition, α_0 controls the tail heaviness of the log-Laplace nugget ϵ , where larger α_0 gives us lighter tails.

B.2 Technical proofs

This section provides the proofs of Theorem 4, Theorem 6, and Theorem 7.

B.2.1 Marginal and Joint Tail Equivalence

Recall that

$$X(\mathbf{s}) = \epsilon(\mathbf{s})X^*(\mathbf{s}),$$

where $X^*(\mathbf{s})$ is the latent smooth process with regularly varying tail and $\epsilon(\mathbf{s})$ is the log-Laplace nugget. For any pair of sites $\mathbf{s}_i, \mathbf{s}_j \in \mathcal{S}$, let (χ_{ij}, η_{ij}) denote the upper tail dependence and residual tail dependence coefficients of the nuggeted pair (X_i, X_j) , and let $(\chi_{ij}^*, \eta_{ij}^*)$ denote the corresponding coefficients of the smooth pair (X_i^*, X_j^*) , with both pairs of coefficients defined through Equation 3.1 and Equation 3.2. We also write $\alpha^*(\mathbf{s})$ for the marginal tail index of $X^*(\mathbf{s})$.

Under the condition $\alpha_0 > \sup_{\mathbf{s} \in \mathcal{S}} \alpha^*(\mathbf{s})$, we first show that, for any site $\mathbf{s} \in \mathcal{S}$,

$$\Pr\{X(\mathbf{s}) > x\} \sim \mathbb{E}\{\epsilon(\mathbf{s})^{\alpha^*(\mathbf{s})}\} \Pr\{X^*(\mathbf{s}) > x\}, \quad x \rightarrow \infty.$$

We then show that, for a random scale-mixture model whose underlying $\{Z(\mathbf{s})\}$ process is asymptotically independent but positively correlated (i.e., $\eta_{ij}^Z \in (1/2, 1)$), for any pair of sites $\mathbf{s}_i, \mathbf{s}_j \in \mathcal{S}$, if X_i^*, X_j^* , and $\min\{X_i^*, X_j^*\}$ all have regularly varying tails and satisfy $\alpha_0 \geq 2 \sup_{\mathbf{s} \in \mathcal{S}} \alpha^*(\mathbf{s})$, then

$$\eta_{ij} = \eta_{ij}^* \quad \text{and} \quad \chi_{ij} \in [c_{ij}\chi_{ij}^*, C_{ij}\chi_{ij}^*].$$

The proof uses Breiman's lemma, which we recall below.

Lemma 15 (Breiman (1965)). Let $X \geq 0$ and $Y \geq 0$ be independent. If X has a regularly varying tail with index $\alpha > 0$, that is,

$$\Pr(X > x) = x^{-\alpha} L(x), \quad x \rightarrow \infty,$$

for some slowly varying function $L(\cdot)$, and if

$$\mathbb{E}(Y^{\alpha+\delta}) < \infty \quad \text{for some } \delta > 0,$$

then

$$\Pr(XY > x) \sim \mathbb{E}(Y^\alpha) \Pr(X > x), \quad x \rightarrow \infty.$$

B.2.1.1 Marginal Tail Equivalence

Fix a site $\mathbf{s} \in \mathcal{S}$. Write $\epsilon(\mathbf{s}) = e^{\zeta(\mathbf{s})}$, where $\zeta(\mathbf{s}) \stackrel{\text{iid}}{\sim} \text{Laplace}(0, b = 1/\alpha_0)$. The moment generating function of $\zeta(\mathbf{s})$ is

$$M_\zeta(t) = \mathbb{E}[e^{t\zeta(\mathbf{s})}] = \frac{1}{1 - b^2 t^2}, \quad |t| < \frac{1}{b} = \alpha_0.$$

Hence, for any p such that $|p| < \alpha_0$,

$$\mathbb{E}\{\epsilon(\mathbf{s})^p\} = \mathbb{E}[e^{p\zeta(\mathbf{s})}] = M_\zeta(p) = \frac{1}{1 - (p/\alpha_0)^2}.$$

Now suppose that $\alpha_0 > \alpha^*(\mathbf{s})$. Then we may choose $\delta > 0$ such that $\alpha^*(\mathbf{s}) + \delta < \alpha_0$, and therefore

$$\mathbb{E}\{\epsilon(\mathbf{s})^{\alpha^*(\mathbf{s})+\delta}\} < \infty.$$

Since $X^*(\mathbf{s})$ has a regularly varying upper tail with index $\alpha^*(\mathbf{s})$ by assumption, Breiman's lemma applies to $X(\mathbf{s}) = \epsilon(\mathbf{s})X^*(\mathbf{s})$ and yields

$$\Pr\{X(\mathbf{s}) > x\} = \Pr\{\epsilon(\mathbf{s})X^*(\mathbf{s}) > x\} \sim \mathbb{E}\{\epsilon(\mathbf{s})^{\alpha^*(\mathbf{s})}\} \Pr\{X^*(\mathbf{s}) > x\}, \quad x \rightarrow \infty.$$

Thus, to satisfy $\alpha^*(\mathbf{s}) < \alpha_0$ at any site $\mathbf{s}$, we impose $\alpha_0 > \sup_{\mathbf{s} \in \mathcal{S}} \alpha^*(\mathbf{s})$.

B.2.1.2 Joint Tail Equivalence

For two sites s_i, s_j , we write

$$\alpha_i^* = \alpha^*(s_i), \quad \alpha_j^* = \alpha^*(s_j), \quad m_i = \mathbb{E}(\epsilon_i^{\alpha_i^*}), \quad m_j = \mathbb{E}(\epsilon_j^{\alpha_j^*}).$$

We make use of the following “sandwich” inequality

$$\begin{aligned} \Pr(\min(\epsilon_i, \epsilon_j) \times \min(X_i^*, X_j^*) > x) &\leq \Pr(\epsilon_i X_i^* > x, \epsilon_j X_j^* > x) \\ &\leq \Pr(\max(\epsilon_i, \epsilon_j) \times \min(X_i^*, X_j^*) > x). \end{aligned} \quad (\text{Sandwich})$$

For $\ell \in \{i, j\}$, by definition,

$$\overline{F}_{X_\ell}(F_{X_\ell}^{-1}(u)) = 1 - u,$$

and by marginal tail equivalence,

$$\overline{F}_{X_\ell}(x) \sim m_\ell \overline{F}_{X_\ell^*}(x).$$

Write the tail of the smooth process as

$$\overline{F}_{X_\ell^*}(x) \sim \tilde{c}_\ell x^{-\alpha_\ell^*}.$$

Hence

$$1 - u = \overline{F}_{X_\ell}(F_{X_\ell}^{-1}(u)) \sim m_\ell \tilde{c}_\ell (F_{X_\ell}^{-1}(u))^{-\alpha_\ell^*},$$

so that

$$F_{X_\ell}^{-1}(u) \sim m_\ell^{1/\alpha_\ell^*} \tilde{c}_\ell^{1/\alpha_\ell^*} (1 - u)^{-1/\alpha_\ell^*}.$$

Hence, as $u \rightarrow 1$,

$$\Pr \left[X_i > F_{X_i}^{-1}(u), X_j > F_{X_j}^{-1}(u) \right]$$

$$\begin{aligned}
&= \Pr \left[X_i > m_i^{1/\alpha_i^*} \tilde{c}_i^{1/\alpha_i^*} (1-u)^{-1/\alpha_i^*}, X_j > m_j^{1/\alpha_j^*} \tilde{c}_j^{1/\alpha_j^*} (1-u)^{-1/\alpha_j^*} \right] \\
&= \Pr \left[\frac{\epsilon_i^{\alpha_i^*}}{m_i} \frac{(X_i^*)^{\alpha_i^*}}{\tilde{c}_i} > (1-u)^{-1}, \frac{\epsilon_j^{\alpha_j^*}}{m_j} \frac{(X_j^*)^{\alpha_j^*}}{\tilde{c}_j} > (1-u)^{-1} \right] \tag{Δ}
\end{aligned}$$

Applying Equation Sandwich, we have

$$\begin{aligned}
\Delta \in & \left[\Pr \left[\min \left(\frac{\epsilon_i^{\alpha_i^*}}{m_i}, \frac{\epsilon_j^{\alpha_j^*}}{m_j} \right) \times \min \left(\frac{(X_i^*)^{\alpha_i^*}}{\tilde{c}_i}, \frac{(X_j^*)^{\alpha_j^*}}{\tilde{c}_j} \right) > (1-u)^{-1} \right], \right. \\
& \left. \Pr \left[\max \left(\frac{\epsilon_i^{\alpha_i^*}}{m_i}, \frac{\epsilon_j^{\alpha_j^*}}{m_j} \right) \times \min \left(\frac{(X_i^*)^{\alpha_i^*}}{\tilde{c}_i}, \frac{(X_j^*)^{\alpha_j^*}}{\tilde{c}_j} \right) > (1-u)^{-1} \right] \right].
\end{aligned}$$

We now apply Breiman's lemma to the two bounds. First note that, as $u \rightarrow 1$,

$$\begin{aligned}
&\Pr \left[\min \left(\frac{(X_i^*)^{\alpha_i^*}}{\tilde{c}_i}, \frac{(X_j^*)^{\alpha_j^*}}{\tilde{c}_j} \right) > (1-u)^{-1} \right] \\
&= \Pr \left[X_i^* > \tilde{c}_i^{1/\alpha_i^*} (1-u)^{-1/\alpha_i^*}, X_j^* > \tilde{c}_j^{1/\alpha_j^*} (1-u)^{-1/\alpha_j^*} \right], \\
&= \Pr \left[X_i^* > F_{X_i^*}^{-1}(u), X_j^* > F_{X_j^*}^{-1}(u) \right],
\end{aligned}$$

which is regularly varying by assumption. Denote its tail exponent by κ_{ij} , that is,

$$\Pr \left[\min \left(\frac{(X_i^*)^{\alpha_i^*}}{\tilde{c}_i}, \frac{(X_j^*)^{\alpha_j^*}}{\tilde{c}_j} \right) > t \right] \in \text{RV}_{-\kappa_{ij}}.$$

We next show that

$$\min \left(\frac{\epsilon_i^{\alpha_i^*}}{m_i}, \frac{\epsilon_j^{\alpha_j^*}}{m_j} \right) \quad \text{and} \quad \max \left(\frac{\epsilon_i^{\alpha_i^*}}{m_i}, \frac{\epsilon_j^{\alpha_j^*}}{m_j} \right)$$

are lighter tailed.

Indeed,

$$\begin{aligned}
\Pr \left[\min \left(\frac{\epsilon_i^{\alpha_i^*}}{m_i}, \frac{\epsilon_j^{\alpha_j^*}}{m_j} \right) > t \right] &= \Pr \left[\frac{\epsilon_i^{\alpha_i^*}}{m_i} > t, \frac{\epsilon_j^{\alpha_j^*}}{m_j} > t \right] \\
&= \Pr \left[\frac{\epsilon_i^{\alpha_i^*}}{m_i} > t \right] \Pr \left[\frac{\epsilon_j^{\alpha_j^*}}{m_j} > t \right] \quad \text{as } \epsilon_i \perp\!\!\!\perp \epsilon_j
\end{aligned}$$

$$\sim L_i(t)L_j(t)t^{-[(\alpha_0/\alpha_i^*)+(\alpha_0/\alpha_j^*)]},$$

and

$$\begin{aligned} \Pr \left[\max \left(\frac{\epsilon_i^*}{m_i}, \frac{\epsilon_j^*}{m_j} \right) > t \right] &= 1 - \Pr \left[\frac{\epsilon_i^*}{m_i} < t, \frac{\epsilon_j^*}{m_j} < t \right] \\ &= 1 - \Pr \left[\frac{\epsilon_i^*}{m_i} < t \right] \Pr \left[\frac{\epsilon_j^*}{m_j} < t \right] \quad \text{as } \epsilon_i \perp\!\!\!\perp \epsilon_j \\ &= \Pr \left[\frac{\epsilon_i^*}{m_i} > t \right] + \Pr \left[\frac{\epsilon_j^*}{m_j} > t \right] - \Pr \left[\frac{\epsilon_i^*}{m_i} > t \right] \Pr \left[\frac{\epsilon_j^*}{m_j} > t \right] \\ &\sim L_i(t)t^{-\alpha_0/\alpha_i^*} + L_j(t)t^{-\alpha_0/\alpha_j^*} - L_i(t)L_j(t)t^{-[(\alpha_0/\alpha_i^*)+(\alpha_0/\alpha_j^*)]} \\ &\sim L(t)t^{-\min(\alpha_0/\alpha_i^*, \alpha_0/\alpha_j^*)}. \end{aligned}$$

For example, the transformed Gaussian scale-mixture models with the underlying Gaussian process having positive pairwise correlations,

$$\eta_{ij}^Z = \frac{1 + \rho_{ij}}{2} \in \left(\frac{1}{2}, 1 \right).$$

Since we assumed in general, the smooth-pair residual tail dependence coefficient satisfies $\eta_{ij}^* > 1/2$,

$$\kappa_{ij} = \frac{1}{\eta_{ij}^*} < 2,$$

while in the AD case $\kappa_{ij} = 1$. Therefore, the assumption

$$\alpha_0 \geq 2 \sup_{\mathbf{s} \in \mathcal{S}} \alpha^*(\mathbf{s})$$

implies

$$\min \left(\frac{\alpha_0}{\alpha_i^*}, \frac{\alpha_0}{\alpha_j^*} \right) > \kappa_{ij},$$

so the required moments are finite and Breiman's lemma applies. Therefore, for some $\delta > 0$,

$$\mathbb{E} \left[\min \left(\frac{\epsilon_i^*}{m_i}, \frac{\epsilon_j^*}{m_j} \right)^{\kappa_{ij} + \delta} \right] < \infty, \quad \mathbb{E} \left[\max \left(\frac{\epsilon_i^*}{m_i}, \frac{\epsilon_j^*}{m_j} \right)^{\kappa_{ij} + \delta} \right] < \infty.$$

Thus Breiman's lemma applies to both bounds. The lower bound can therefore be written as

$$\begin{aligned} & \Pr \left[\min \left(\frac{\epsilon_i^*}{m_i}, \frac{\epsilon_j^*}{m_j} \right) \times \min \left(\frac{(X_i^*)^{\alpha_i^*}}{\tilde{c}_i}, \frac{(X_j^*)^{\alpha_j^*}}{\tilde{c}_j} \right) > (1-u)^{-1} \right] \\ & \sim \mathbb{E} \left[\min \left(\frac{\epsilon_i^*}{m_i}, \frac{\epsilon_j^*}{m_j} \right) \right] \Pr \left[\min \left(\frac{(X_i^*)^{\alpha_i^*}}{\tilde{c}_i}, \frac{(X_j^*)^{\alpha_j^*}}{\tilde{c}_j} \right) > (1-u)^{-1} \right]. \end{aligned}$$

Similarly, the upper bound can be written as

$$\begin{aligned} & \Pr \left[\max \left(\frac{\epsilon_i^*}{m_i}, \frac{\epsilon_j^*}{m_j} \right) \times \min \left(\frac{(X_i^*)^{\alpha_i^*}}{\tilde{c}_i}, \frac{(X_j^*)^{\alpha_j^*}}{\tilde{c}_j} \right) > (1-u)^{-1} \right] \\ & \sim \mathbb{E} \left[\max \left(\frac{\epsilon_i^*}{m_i}, \frac{\epsilon_j^*}{m_j} \right) \right] \Pr \left[\min \left(\frac{(X_i^*)^{\alpha_i^*}}{\tilde{c}_i}, \frac{(X_j^*)^{\alpha_j^*}}{\tilde{c}_j} \right) > (1-u)^{-1} \right]. \end{aligned}$$

Since the constant factor does not change the tail exponent, we obtain

$$\eta_{ij} = \eta_{ij}^*.$$

To identify the constants in the bound for χ_{ij} , we calculate

$$c_{ij} = \mathbb{E} \left[\min \left(\frac{\epsilon_i^*}{m_i}, \frac{\epsilon_j^*}{m_j} \right) \right] \quad \text{and} \quad C_{ij} = \mathbb{E} \left[\max \left(\frac{\epsilon_i^*}{m_i}, \frac{\epsilon_j^*}{m_j} \right) \right].$$

Denote

$$U := \frac{\epsilon_i^*}{m_i}, \quad V := \frac{\epsilon_j^*}{m_j},$$

and we have

$$\begin{aligned}
\mathbb{E}[\min(U, V)] &= \int_0^\infty \Pr(U > t, V > t) dt \\
&= \int_0^\infty \Pr\left[\frac{\epsilon_i^{\alpha_i^*}}{m_i} > t\right] \Pr\left[\frac{\epsilon_j^{\alpha_j^*}}{m_j} > t\right] dt \\
&= \int_0^\infty \Pr[\epsilon_i > (m_i t)^{1/\alpha_i^*}] \Pr[\epsilon_j > (m_j t)^{1/\alpha_j^*}] dt.
\end{aligned}$$

Without loss of generality, assume $m_i > m_j$, so $1/m_i \leq 1/m_j$. Define

$$t_i = 1/m_i, \quad t_j = 1/m_j,$$

and split the integral into

$$\int_0^\infty \dots = \underbrace{\int_0^{t_i} \dots}_{I_1} + \underbrace{\int_{t_i}^{t_j} \dots}_{I_2} + \underbrace{\int_{t_j}^\infty \dots}_{I_3},$$

because for $\ell \in \{i, j\}$ we have

$$\Pr[\epsilon_\ell > (m_\ell t)^{1/\alpha_\ell^*}] = \begin{cases} 1 - \frac{1}{2}(m_\ell t)^{p_\ell} & t < 1/m_\ell, \\ \frac{1}{2}(m_\ell t)^{-p_\ell} & t \geq 1/m_\ell, \end{cases}$$

where $p_\ell = \alpha_0/\alpha_\ell^*$. Then,

$$\begin{aligned}
I_1 &= \int_0^{t_i} \left[1 - \frac{1}{2}(m_i t)^{p_i}\right] \left[1 - \frac{1}{2}(m_j t)^{p_j}\right] dt \\
&= \frac{1}{m_i} - \frac{1}{2} \frac{1}{(p_i + 1)m_i} - \frac{1}{2} \frac{m_j^{p_j}}{(p_j + 1)m_i^{p_j+1}} + \frac{1}{4} \frac{m_j^{p_j}}{(p_i + p_j + 1)m_i^{p_j+1}}, \\
I_2 &= \int_{t_i}^{t_j} \left[\frac{1}{2}(m_i t)^{-p_i}\right] \left[1 - \frac{1}{2}(m_j t)^{p_j}\right] dt \\
&= \frac{1}{2} \frac{1}{p_i - 1} \left(\frac{1}{m_i} - \frac{m_j^{p_i-1}}{m_i^{p_i}}\right) - \frac{1}{4} \frac{1}{1 + p_j - p_i} \left(\frac{m_j^{p_i-1}}{m_i^{p_i}} - \frac{m_j^{p_j}}{m_i^{p_j+1}}\right),
\end{aligned}$$

$$\begin{aligned}
I_3 &= \int_{t_j}^{\infty} \left[\frac{1}{2} (m_i t)^{-p_i} \right] \left[\frac{1}{2} (m_j t)^{-p_j} \right] dt \\
&= \frac{1}{4} \frac{m_i^{-p_i} m_j^{p_i-1}}{p_i + p_j - 1}.
\end{aligned}$$

Thus,

$$c_{ij} = \mathbb{E}[\min(U, V)] = I_1 + I_2 + I_3.$$

Since $\mathbb{E}[\max(U, V)] + \mathbb{E}[\min(U, V)] = \mathbb{E}[U + V] = 2$, we have

$$C_{ij} = \mathbb{E}[\max(U, V)] = 2 - (I_1 + I_2 + I_3).$$

Therefore,

$$\chi_{ij} \in [c_{ij}\chi_{ij}^*, C_{ij}\chi_{ij}^*].$$

B.2.2 Proof of Theorem 7

The model of Majumder et al. (2024) is

$$\tilde{X}^*(\mathbf{s}) = \delta \tilde{R}(\mathbf{s}) + (1 - \delta) \tilde{W}(\mathbf{s})$$

where $\tilde{R}(\mathbf{s})$ and $\tilde{W}(\mathbf{s})$ have $\text{Exp}(1)$ margins. Consequently, $\tilde{X}^*(\mathbf{s})$ has marginal distribution

$$F_{\tilde{X}^*}(x) = 1 - \frac{1 - \delta}{1 - 2\delta} \exp \left\{ -\frac{x}{1 - \delta} \right\} + \frac{\delta}{1 - 2\delta} \exp \left\{ -\frac{x}{\delta} \right\}.$$

Because the copula is invariant under strictly increasing marginal transformations, it is more convenient to work with

$$X^*(\mathbf{s}) = \exp(\tilde{X}^*(\mathbf{s})) = R(\mathbf{s})^\delta W(\mathbf{s})^{1-\delta}$$

where $R(\mathbf{s})$ and $W(\mathbf{s})$ are marginally standard Pareto.

We now derive the joint tail decay rate of this model. Let $x := F_{X^*}^{-1}(u)$. Then

$$\begin{aligned} \lim_{u \rightarrow 1} \Pr[X_1^* > F_{X_1^*}^{-1}(u), X_2^* > F_{X_2^*}^{-1}(u)] \\ = \lim_{u \rightarrow 1} \Pr[R_1^\delta W_1^{1-\delta} > x, R_2^\delta W_2^{1-\delta} > x], \end{aligned}$$

so we study the asymptotic behavior of the right-hand side as $u \rightarrow 1$.

To bound this probability, we use the sandwich inequality Equation Sandwich:

$$\begin{aligned} \Pr[\min(R_1^\delta, R_2^\delta) \cdot \min(W_1^{1-\delta}, W_2^{1-\delta}) > x] \\ \leq \Pr[X_1^* > F_{X_1^*}^{-1}(u), X_2^* > F_{X_2^*}^{-1}(u)] \\ \leq \Pr[\max(R_1^\delta, R_2^\delta) \cdot \min(W_1^{1-\delta}, W_2^{1-\delta}) > x] \end{aligned}$$

Hence, we determine the tail behavior of $\Pr\{\min(R_1^\delta, R_2^\delta) > x\}$, $\Pr\{\min(W_1^{1-\delta}, W_2^{1-\delta}) > x\}$, and $\Pr\{\max(R_1^\delta, R_2^\delta) > x\}$.

We first analyze $\Pr\{\min(R_1^\delta, R_2^\delta) > x\}$. In Majumder et al. (2024), $R(s)$ is induced by a Brown–Resnick process satisfying

$$\Pr[\tilde{R}_1 < r_1, \tilde{R}_2 < r_2] = \exp[-\Lambda(r_1, r_2)]$$

where

$$\Lambda(r_1, r_2) = \frac{1}{r_1} \Phi \left\{ \frac{a}{2} - \frac{1}{a} \log \left(\frac{r_1}{r_2} \right) \right\} + \frac{1}{r_2} \Phi \left\{ \frac{a}{2} - \frac{1}{a} \log \left(\frac{r_2}{r_1} \right) \right\}$$

in which

$$a = [2\gamma(h)]^{1/2}, \quad \gamma(h) = (h/\rho_R)^{\alpha_R}, \quad \Phi\{\cdot\} \text{ is standard Gaussian distribution.}$$

After transforming to standard Pareto margins, we obtain

$$\Pr[R_1 < r_1, R_2 < r_2]$$

$$\begin{aligned}
&= \Pr \left[-\frac{1}{\log(1 - \frac{1}{R_1})} < -\frac{1}{\log(1 - \frac{1}{r_1})}, -\frac{1}{\log(1 - \frac{1}{R_2})} < -\frac{1}{\log(1 - \frac{1}{r_2})} \right] \\
&= \exp \left[-\Lambda \left(-\frac{1}{\log(1 - \frac{1}{r_1})}, -\frac{1}{\log(1 - \frac{1}{r_2})} \right) \right]
\end{aligned}$$

Therefore, as $x \rightarrow \infty$,

$$\begin{aligned}
&\Pr[\min(R_1^\delta, R_2^\delta) > x] \\
&= \Pr[R_1 > x^{1/\delta}, R_2 > x^{1/\delta}] \\
&= 1 - \Pr[R_1 \leq x^{1/\delta}] - \Pr[R_2 \leq x^{1/\delta}] + \Pr[R_1 \leq x^{1/\delta}, R_2 \leq x^{1/\delta}] \\
&= \frac{2}{x^{1/\delta}} - \left\{ 1 - \exp \left[-\Lambda \left(-\frac{1}{\log(1 - \frac{1}{x^{1/\delta}})}, -\frac{1}{\log(1 - \frac{1}{x^{1/\delta}})} \right) \right] \right\} \\
&\sim \frac{2}{x^{1/\delta}} - \Lambda \left(-\frac{1}{\log(1 - \frac{1}{x^{1/\delta}})}, -\frac{1}{\log(1 - \frac{1}{x^{1/\delta}})} \right) \\
&= \frac{2}{x^{1/\delta}} + 2\Phi \left(\frac{a}{2} \right) \left[\log \left(1 - \frac{1}{x^{1/\delta}} \right) \right] \\
&\sim L_1(x)x^{-1/\delta}
\end{aligned}$$

By the same argument, we also obtain

$$\begin{aligned}
&\Pr[\max(R_1^\delta, R_2^\delta) > x] = 1 - \Pr[\max(R_1^\delta, R_2^\delta) < x] \\
&\sim L_2(x)x^{-1/\delta}
\end{aligned}$$

For the Gaussian-copula component, Ledford and Tawn (1996) show that the joint survival function with unit Pareto margins satisfies

$$\Pr[\min(W_1^{1-\delta}, W_2^{1-\delta}) > x] \sim L_3(x)x^{-1/\{\eta_W(1-\delta)\}}$$

where $\eta_W = (1 + \rho)/2$ and $\rho = \text{Cor}(Z_1, Z_2)$.

Combining these with Breiman's lemma and the sandwich inequality yields

$$\Pr \left[X_1^* > F_{X_1^*}^{-1}(u), X_2^* > F_{X_2^*}^{-1}(u) \right] \sim \begin{cases} L(x)x^{-1/\{\eta_W(1-\delta)\}} & \eta_W \geq \delta/(1-\delta), \\ \mathbb{E} [\min(W_1, W_2)^{(1-\delta)/\delta}] x^{-1/\delta} & \eta_W < \delta/(1-\delta). \end{cases}$$

Thus the joint exceedance probability is regularly varying.

B.3 MCMC Details

This section gives the full hierarchical specification used for posterior sampling. Here $t = 1, \dots, T$ indexes replicates, $j = 1, \dots, D$ indexes sites, and $k = 1, \dots, K$ indexes knots. The observed series Y_{tj} is treated as an anomaly process. We assume temporal independence across replicates over time. We restate the censored likelihood from Section 3.3.1. Let $\mathcal{C}_t = \{j : Y_{tj} \leq u_j\}$ and $\mathcal{E}_t = \{j : Y_{tj} > u_j\}$, with $j = 1, \dots, D$ and $x_{0j} = F_{X_j}^{-1}(p)$. Then

$$L_t(\mathbf{Y}_t \mid \mathbf{S}_t, \mathbf{Z}_t, \boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\rho}, l, \alpha_0, \boldsymbol{\sigma}, \boldsymbol{\xi}) = \prod_{j \in \mathcal{C}_t} F_\epsilon \left(\frac{x_{0j}}{X_{tj}^*} \right) \times \prod_{j \in \mathcal{E}_t} f_{X|X^*}(x_{tj} \mid X_{tj}^*) \left| \frac{dX_{tj}}{dY_{tj}} \right|.$$

For $j \in \mathcal{E}_t$,

$$\left| \frac{dX_{tj}}{dY_{tj}} \right| = \frac{g_{Y_{tj}}(Y_{tj})}{f_{X_{tj}}(x_{tj})}, \quad g_{Y_{tj}}(y) = (1 - p) h_{u_j}(y \mid \sigma_j, \xi_j),$$

where h_{u_j} is the GP density associated with H_{u_j} . Under temporal independence of the anomaly replicates, the full likelihood is the product of L_t at each replicate.

The latent process is

$$X_t(\mathbf{s}) = \epsilon_t(\mathbf{s}) X_t^*(\mathbf{s}), \quad X_t^*(\mathbf{s}) = R_t(\mathbf{s})^{\phi(\mathbf{s})} g\{Z_t(\mathbf{s})\},$$

with

$$R_t(\mathbf{s}) = \sum_{k=1}^K B_k(\mathbf{s}; l) S_{tk},$$

where $B_k(\cdot; l)$ are Wendland (1995) basis functions with radius l , and

$$S_{tk} \mid \gamma_k \sim \text{Stable}(\alpha = 1/2, \beta = 1, \gamma_k, \delta = 0) \equiv \text{Lévy}(0, \gamma_k),$$

$$l \sim \text{Half-}t_{\nu=1}(0, 3).$$

Similar to Shi et al. (2026), we fix the scale parameter γ at 1, as varying it does not play any role in modulating the tail dependence characteristics of the model.

The nugget terms are i.i.d. across sites, and we enforce $\alpha_0 > 1$ by

$$\epsilon_t(\mathbf{s}_j) \stackrel{\text{iid}}{\sim} \log\text{-Laplace}(0, 1/\alpha_0), \quad \alpha_0 = 1 + \exp(\vartheta), \quad \vartheta \sim N(3, 0.5^2).$$

The latent Gaussian field satisfies

$$\mathbf{Z}_t = (Z_t(\mathbf{s}_1), \dots, Z_t(\mathbf{s}_D))^\top \mid \boldsymbol{\rho} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\rho),$$

where $\boldsymbol{\Sigma}_\rho$ is a locally isotropic nonstationary Matérn covariance (Paciorek and Schervish, 2006; Risser and Calder, 2015), with entries

$$C(\mathbf{s}_i, \mathbf{s}_j) = \zeta(\mathbf{s}_i)\zeta(\mathbf{s}_j) \frac{\sqrt{\rho(\mathbf{s}_i)\rho(\mathbf{s}_j)}}{\{\rho(\mathbf{s}_i) + \rho(\mathbf{s}_j)\}/2} \mathcal{M}_\nu\left(\frac{\|\mathbf{s}_i - \mathbf{s}_j\|}{\sqrt{\{\rho(\mathbf{s}_i) + \rho(\mathbf{s}_j)\}/2}}\right),$$

where $\zeta(\mathbf{s}) \equiv 1$ and ν is fixed in our implementation.

The surfaces $\phi(\mathbf{s})$ and $\rho(\mathbf{s})$ are represented using Gaussian kernel basis functions centered at the K knots. Similar to Shi et al. (2026), the prior for ϕ is centered at the AI - AD transition boundary and assigns relatively little mass near the edges of its support, which correspond to extremely strong or extremely weak tail dependence. Knot-level priors are

$$\phi_k \sim \text{Beta}(2, 2), \quad \rho_k \sim \text{Half-Normal}(0, 10^2), \quad k = 1, \dots, K.$$

Marginal GP parameters are modeled as

$$\log \sigma_j = \mathbf{c}_j^\top \boldsymbol{\beta}_\sigma, \quad \xi_j = \mathbf{d}_j^\top \boldsymbol{\beta}_\xi,$$

with

$$\boldsymbol{\beta}_\sigma \mid \sigma_{\beta_\sigma} \sim \text{MVN}(\mathbf{0}, \sigma_{\beta_\sigma}^2 \mathbf{I}), \quad \boldsymbol{\beta}_\xi \mid \sigma_{\beta_\xi} \sim \text{MVN}(\mathbf{0}, \sigma_{\beta_\xi}^2 \mathbf{I}),$$

$$\sigma_{\beta_\sigma} \sim \text{Half-}t_{\nu=2}(0, 1), \quad \sigma_{\beta_\xi} \sim \text{Half-}t_{\nu=2}(0, 1).$$

Let

$$\Theta = \{\phi, \rho, l, \alpha_0, \beta_\sigma, \beta_\xi, \sigma_{\beta_\sigma}, \sigma_{\beta_\xi}\}.$$

The posterior target is

$$p(\Theta, \{\mathbf{S}_t, \mathbf{Z}_t\}_{t=1}^T \mid \mathbf{Y}_{1:T}) \propto \prod_{t=1}^T L_t(\mathbf{Y}_t \mid \mathbf{S}_t, \mathbf{Z}_t, \Theta, \gamma) p(\mathbf{S}_t \mid \gamma) p(\mathbf{Z}_t \mid \rho) p(\Theta).$$

We sample from the posterior using an adaptive random-walk Metropolis-within-Gibbs algorithm (Shaby and Wells, 2010). Parameters shared across all replicates are updated sequentially, and replicate-specific latent blocks $(\mathbf{S}_t, \mathbf{Z}_t)$, $t = 1, \dots, T$, are updated in parallel.

The sampler is implemented in `Python`, with parallelization via `mpi4py` and `OpenMPI` (Dalcin and Fang, 2021). For Model (M4), evaluations of F_X and F_X^{-1} are implemented in `C++` using `GSL` (Gough, 2009), and likelihood evaluations are JIT-compiled with `Numba` (Lam et al., 2015).