On statistical inference for spatial and spatio-temporal extreme processes

Abdul-Fattah Abu-Awwad

N° d'ordre NNT : 2019LYSE1079

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

École Doctorale 512
InfoMaths

Spécialité de doctorat : Mathématiques Appliquées
Discipline : Statistiques

Soutenue publiquement le 20 / 06 / 2019, par :
Abdul-Fattah ABU-AWWAD

Sur l'inférence statistique pour des processus spatiaux et spatio-temporels extrêmes
On statistical inference for spatial and spatio-temporal extreme processes

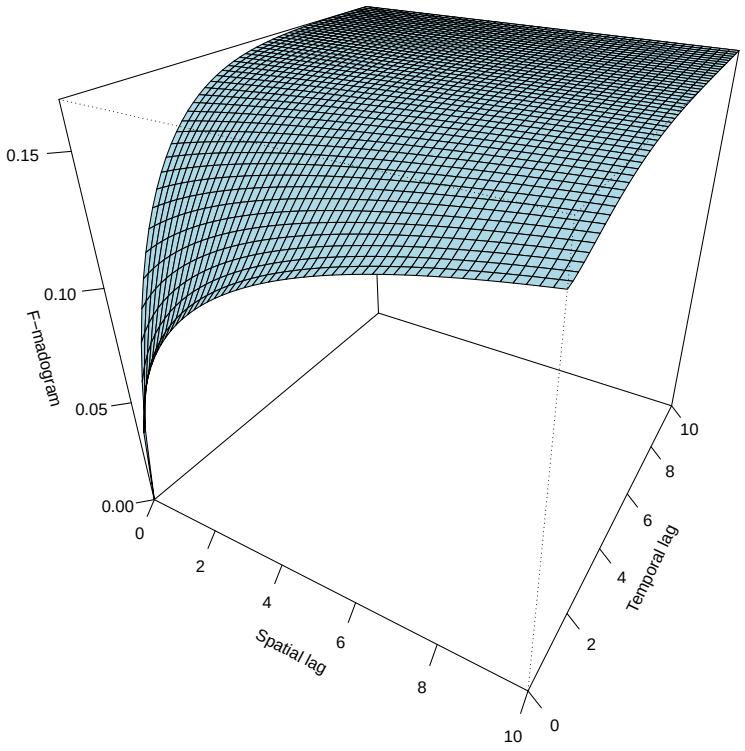
---

Devant le jury composé de :

Guillou Armelle, Professeure, Université de Strasbourg	Rapporteure
Naveau Philippe, Directeur de recherche CNRS, LSCE	Rapporteur
Fougères Anne-Laure, Professeure, Université Claude Bernard Lyon 1	Examinatrice
Girard Stéphane, Directeur de recherche, Inria Grenoble Rhône-Alpes	Examinateur
Toulemonde Gwladys, Maître de conférences, Université de Montpellier	Examinatrice
Maume-Deschamps Véronique, Professeure, Université Claude Bernard Lyon 1	Directrice de thèse
Ribereau Pierre, Maître de conférences, Université Claude Bernard Lyon 1	Co-directeur de thèse

Institut Camille Jordan
UMR 5208 du CNRS

Sur l'inférence statistique pour des processus spatiaux et spatio-temporels extrêmes



Abdul-Fattah ABU-AWWAD

Thèse de doctorat

Université Claude Bernard Lyon 1

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université	**M. le Professeur Frédéric FLEURY**
Président du Conseil Académique	M. le Professeur Hamda BEN HADID
Vice-président du Conseil d'Administration	M. le Professeur Didier REVEL
Vice-président du Conseil Formation et Vie Universitaire	M. le Professeur Philippe CHEVALIER
Vice-président de la Commission Recherche	M. Fabrice VALLÉE
Directeur Général des Services	M. Alain HELLEU

COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude Bernard	Directeur : M. le Professeur J. ETIENNE
Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux	Directeur : Mme la Professeure C. BURILLON
Faculté d'Odontologie	Directeur : M. le Professeur D. BOURGEOIS
Institut des Sciences Pharmaceutiques et Biologiques	Directeur : Mme la Professeure C. VINCIGUERRA
Institut des Sciences et Techniques de la Réadaptation	Directeur : M. le Professeur Y. MATILLON
Département de formation et Centre de Recherche en Biologie Humaine	Directeur : Mme la Professeure A-M. SCHOTT

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies	Directeur : M. F. DE MARCHI
Département Biologie	Directeur : M. le Professeur F. THEVENARD
Département Chimie Biochimie	Directeur : Mme C. FELIX
Département GEP	Directeur : M. Hassan HAMMOURI
Département Informatique	Directeur : M. le Professeur S. AKKOUCHE
Département Mathématiques	Directeur : M. le Professeur G. TOMANOV
Département Mécanique	Directeur : M. le Professeur H. BEN HADID
Département Physique	Directeur : M. le Professeur J-C PLENET
UFR Sciences et Techniques des Activités Physiques et Sportives	Directeur : M. Y.VANPOULLE
Observatoire des Sciences de l'Univers de Lyon	Directeur : M. B. GUIDERDONI
Polytech Lyon	Directeur : M. le Professeur E.PERRIN
Ecole Supérieure de Chimie Physique Electronique	Directeur : M. G. PIGNAULT
Institut Universitaire de Technologie de Lyon 1	Directeur : M. le Professeur C. VITON
Ecole Supérieure du Professorat et de l'Education	Directeur : M. le Professeur A. MOUGNIOTTE
Institut de Science Financière et d'Assurances	Directeur : M. N. LEBOISNE

Résumé

Les catastrophes naturelles comme les canicules, les tempêtes ou les précipitations extrêmes, proviennent de processus physiques et ont, par nature, une dimension spatiale ou spatio-temporelle. Le développement de modèles et de méthodes d'inférences pour ces processus est un domaine de recherche très actif. Cette thèse traite de l'inférence statistique pour les événements extrêmes dans le cadre spatial et spatio-temporel. En particulier, nous nous intéressons à deux classes de processus stochastique: les processus spatiaux max-mélange et les processus max-stable spatio-temporels. Nous illustrons les résultats obtenus sur des données de précipitations dans l'Est de l'Australie et dans une région de la Floride aux Etats-Unis.

Dans la partie spatiale, nous proposons deux tests sur le paramètre de mélange a d'un processus spatial max-mélange: le test statistique Z_a et le rapport de vraisemblance par paire LR_a. Nous comparons les performances de ces tests sur simulations. Nous utilisons la vraisemblance par paire pour l'estimation. Dans l'ensemble, les performances des deux tests sont satisfaisantes. Toutefois, les tests rencontrent des difficultés lorsque le paramètre a se situe à la frontière de l'espace des paramètres, i.e., $a \in \{0, 1\}$, dues à la présence de paramètre de "nuisance" qui ne sont pas identifiés sous l'hypothèse nulle. Nous appliquons ces tests dans le cadre d'une analyse d'excès au delà d'un grand seuil pour des données de précipitations dans l'Est de l'Australie. Nous proposons aussi une nouvelle procédure d'estimation pour ajuster des processus spatiaux max-mélanges lorsqu'on ne connait pas la classe de dépendance extrêmal. La nouveauté de cette procédure est qu'elle permet de faire de l'inférence sans spécifier au préalable la famille de distributions, laissant ainsi parle les données et guider l'estimation. En particulier, la procédure d'estimation utilise un ajustement par la méthode des moindres carrés sur l'expression du F^λ-madogramme d'un modèle max-mélange qui contient les paramètres d'intérêt. Nous montrons la convergence de l'estimateur du paramètre de mélange a. Une indication sur la normalité asymptotique est donnée numériquement. Une étude sur simulation montrent que la méthode proposée améliore les coefficients empiriques pour la classe de modèles max-mélange. Nous implémentons notre procédure d'estimations sur des données de maximas mensuels de précipitations en Australie dans un but exploratoire et confirmatoire.

Dans la partie spatio-temporelle, nous proposons une méthode d'estimation semi-paramétrique pour les processus max-stables spatio-temporels en nous basant sur une expression explicite du F-madogramme spatio-temporel. Cette partie permet de faire le pont entre la géostatistique et la théorie des valeurs extrêmes. En particulier, pour des observations sur grille régulière, nous estimons le F-madogramme spatio-temporel par sa version empirique et nous appliquons une procédure basée sur les moments pour obtenir les estimations des paramètres d'intérêt. Nous

illustrons les performances de cette procédure par une étude sur simulations. Ensuite, nous appliquons cette méthode pour quantifier le comportement extrêmal de maximum de données radar de précipitations dans l'Etat de Floride. Cette méthode peut être une alternative ou une première étape pour la vraisemblance composite. En effet, les estimations semi-paramétriques pourrait être utilisées comme point de départ pour les algorithmes d'optimisation utilisés dans la méthode de vraisemblance par paire, afin de réduire le temps de calcul mais aussi d'améliorer l'efficacité de la méthode.

Mots-clé: Dépendance/Indépendance asymptotique, vraisemblance composite, événement extrême, F^{λ}-madogramme, processus max-stable, processus max-mélange, précipitations, estimation semi-paramétrique, processus max-stable spatio-temporel.

Abstract

Natural hazards such as heat waves, extreme wind speeds, and heavy rainfall, arise due to physical processes and are spatial or spatio-temporal in extent. The development of models and inference methods for these processes is a very active area of research. This thesis deals with the statistical inference of extreme and rare events in both spatial and spatio-temporal settings. Specifically, our contributions are dedicated to two classes of stochastic processes: spatial max-mixture processes and space-time max-stable processes. The proposed methodologies are illustrated by applications to rainfall data collected from the East of Australia and from a region in the State of Florida, USA.

In the spatial part, we consider hypothesis testing for the mixture parameter a of a spatial max-mixture model using two classical statistics: the Z-test statistic Z_a and the pairwise likelihood ratio statistic LR_a. We compare their performance through an extensive simulation study. The pairwise likelihood is employed for estimation purposes. Overall, the performance of the two statistics is satisfactory. Nevertheless, hypothesis testing presents some difficulties when a lies on the boundary of the parameter space, i.e., $a \in \{0, 1\}$, due to the presence of additional nuisance parameters which are not identified under the null hypotheses. We apply this testing framework in an analysis of exceedances over a large threshold of daily rainfall data from the East of Australia. We also propose a novel estimation procedure to fit spatial max-mixture processes with unknown extremal dependence class. The novelty of this procedure is to provide a way to make inference without specifying the distribution family prior to fitting the data. Hence, letting the data speak for themselves. In particular, the estimation procedure uses nonlinear least squares fit based on a closed form expression of the so-called F^λ-madogram of max-mixture models which contains the parameters of interest. We establish the consistency of the estimator of the mixing parameter a. An indication for asymptotic normality is given numerically. A simulation study shows that the proposed procedure improves empirical coefficients for the class of max-mixture models. In an analysis of monthly maxima of Australian daily rainfall data, we implement the proposed estimation procedure for diagnostic and confirmatory purposes.

In the spatio-temporal part, based on a closed form expression of the spatio-temporal F-madogram, we suggest a semi-parametric estimation methodology for space-time max-stable processes. This part provides a bridge between geostatistics and extreme value theory. In particular, for regular grid observations, the spatio-temporal F-madogram is estimated nonparametrically by its empirical version and a moment-based procedure is applied to obtain parameter estimates. The performance of the method is investigated through an extensive simulation study. Afterward, we apply this method to quantify the extremal behavior of radar daily rainfall maxima

data from a region in the State of Florida. This approach could serve as an alternative or a pre-requisite to pairwise likelihood estimation. Indeed, the semi-parametric estimates could be used as starting values for the optimization algorithm used to maximize the pairwise log-likelihood function in order to reduce the computational burden and also to improve the statistical efficiency.

Keywords: Asymptotic dependence/independence, composite likelihood, extreme event, F^{λ}-madogram, max-stable process, max-mixture process, rainfall data, semi-parametric estimation, space-time max-stable process.

Acknowledgments

I would foremost like to thank my supervisors Prof. Maume-Deschamps Véronique and Asst. Prof. Ribereau Pierre for their outstanding guidance, remarkable knowledge, numerous inspiring discussions, patience and encouragement, and for giving me the opportunity to travel to several conferences, where I was able to present our work and meet famous researchers from all over the world. I have been very fortunate to work with these excellent professors.

My sincere thanks and gratitude extend to my thesis committee: the referees Prof. Guillou Armelle and Prof. Naveau Philippe and the examiners Prof. Fougères Anne-Laure, Prof. Girard Stéphane and Prof. Toulemonde Gwladys—not only for the careful reading of my thesis and their extreme patience but for their helpful intellectual contributions and valuable comments.

Many thanks also to all my colleagues at ICJ. Special thanks to my friend Manaf Ahmad for his accompaniment during my first year and also for the many discussions that we had about work.

Further, I owe my gratitude to the French government for supporting Palestinian students to complete their graduate studies in France. In particular, I want to express my gratefulness to the French Consulate in Jerusalem, for funding my research during the past three years within the French consulate scholarship program.

I would also like to thank Margit Crowell from the Southwest Florida Water Management District (SWFWMD) for providing the Florida rainfall dataset analyzed in this thesis.

Last, but not least, I would like to express my deep gratitude to the most important people in my life: my parents, my grandmother, my sisters, my brothers, my wife, and my children Omar and Jana. Many thanks for love and unlimited support at any time.

Contents

Chapter 1

Introduction

This thesis aims to develop novel statistical inference methods that allow quantifying the extremal dependence structure of both spatial and spatio-temporal processes. Specifically, we focus on two families of extreme stochastic processes: the spatial max-mixture processes and the space-time max-stable processes. These inference methods comprise parametric, semi-parametric, as well as nonparametric procedures. In addition, we show all methods at work in both simulation studies and in an analysis of extremal rainfall data collected in two regions: the East of Australia and the State of Florida, USA.

In this introductory chapter, we present some general statements on spatial and spatio-temporal modeling of extremes, with a highlight on the existing statistical inference methods in this context. The latter motivated us to conduct this thesis. Afterward, we present the main contributions and outline of the thesis.

Chapter 2 will recall the mathematical background on spatial and spatio-temporal max-stable processes as well as spatial max-mixture processes. We refer the reader to that chapter for precise definitions of the concepts and tools used in this chapter.

1.1 General introduction and motivation

Many extreme value phenomena are inherently spatial, or spatio-temporal, in nature. For example, [81] provided a bridge between classical geostatistics and extreme value theory to analyze annual maxima of daily precipitation measurement in Bourgogne, France. [18] fitted spatial max-stable models to extreme snowfall data in the Alpine region of Switzerland. [44] modeled the annual temperature maxima in Switzerland, which is a relatively large-scale phenomenon compared to rainfall, based on spatial max-stable processes derived from underlying Gaussian random fields. [64] extended the latter to spatio-temporal modeling of hourly rainfall measurements in western Switzerland. [11] (respectively, [101]) used spatial max-mixture models to analyze daily precipitations over the East of Australia based on block maxima over the observation period and exceedances over a large threshold (respectively, extremes of the winter observations of a hindcast dataset of significant wave height, a measure of ocean energy, from the North Sea). [24] used a particular class of space-time max-stable models in an analysis of radar rainfall measurements in a region located in Florida.

In the context of climate change, some extreme events tend to be more and more frequent. For example, heatwaves and heavy rainfall are predicted to become more frequent in some regions of Europe, see [16]. These natural disasters (meteorological and more generally environmental disasters) have a considerable impact on societies since they leave destruction and chaos behind while passing over certain areas. Hence, the statistical modeling of extremes constitutes a crucial challenge.

Typically, extremes of environmental and climate processes such as extreme wind speeds or heavy precipitation are modeled using extreme value theory. For an introduction to the univariate extreme value theory, see, e.g., [32] and in the multivariate case, see, e.g., [14]. Max-stable processes are ideally suited for the statistical modeling of spatial extremes as they form the natural extension of multivariate extreme value distributions to infinite dimensions, see [49]. Various families of max-stable models have been proposed for extremal data, see [69, 82, 90, 94]. Within the class of max-stable models, only two types of dependence structures are possible; either the process is asymptotically dependent or it is exactly independent, see, e.g., [11, 45, 96]. In particular, fitting asymptotically dependent models to asymptotically independent data may lead to mis-estimation of probabilities of extreme joint events, since it is wrongfully assumed that the most extreme marginal events may occur simultaneously [33].

Alternatively, [101] introduced a class of max-mixture models, which are able to capture both asymptotic dependence and asymptotic independence. The basic idea is to mix max-stable and asymptotically independent processes (i.e., Gaussian and inverted max-stable processes). For statistical inference on max-stable and max-mixture models, the common approach has been to maximize the composite likelihood, see, e.g., [11, 26, 63, 64, 84, 101]. The composite likelihood is obtained by multiplying likelihoods of marginal or conditional events [73, 98]. Specifically, the pairwise likelihood is often used because it only requires the specification of bivariate events and typically leads to substantial computational gains compared to the full likelihood approach. These advantages come at the cost of a loss of efficiency which depends mainly on the underlying true model. Usually, the composite likelihood information criterion (CLIC) [100] is used for model selection.

The inference on spatial and spatio-temporal extreme processes is an open field that is still in development. Various techniques have been proposed for parameter estimation for these processes. Each technique has its pros and cons. For instance, a semi-parametric estimation procedure was proposed by [5] for spatial max-mixture processes as an alternative or a prerequisite to the widely used pairwise likelihood inference which has gained much popularity due to its theoretical properties. Nevertheless, parameter estimation using pairwise likelihood suffers from some defects:

(i) it can be onerous, since the computation and subsequent optimization of the objective function is time-consuming, see, e.g., [23, 101],

(ii) the choice of good initial values for optimization of the pairwise likelihood is essential, see, e.g., [23, 26],

(iii) the resulting estimates of the asymptotically independent process parameters in the max-mixture model are unsatisfactory in some situations, see, e.g., [5, 11],

(iv) it is a model-based procedure which means that a parametric distribution family has to be specified prior to fitting the data.

On the basis of point (iv), fitting max-mixture models with this approach can be laborious, owing to the large number of possible combinations that can be constructed from asymptotically dependent and asymptotically independent processes. Moreover, within the same family of max-stable processes, a large variety of behaviors can be obtained depending on the used dependence model. For instance, different options of the semivariogram model for a Brown-Resnick (BR) process (see [20, 69]) lead to different behaviors. As an illustration, Figure 1.1 displays the theoretical behaviors of some valid isotropic semivariogram models and the associated extremal coefficient functions for the BR process. Using power semivariogram models (e.g., quadratic or linear) leads to independence as the distance tends to ∞ (i.e., the extremal dependence coefficient tends to 2). While using bounded semivariogram models (e.g., exponential or Cauchy), independence cannot be reached even when the distance tends to ∞ (i.e., the extremal dependence coefficient is less than 2). In other words, when the semivariogram model is bounded, then the resulting BR process will be dependent even at very long ranges. Analogously, for Schlather processes (see [90]), the correlation function can be modeled by different valid parametric families allowing for considerable diversity of spatial behavior. In other words, the smoothness of these processes can be controlled by the choice of the correlation function.

Similarly, the semi-parametric estimation approach proposed in [5] for fitting these processes is also a model-based approach. Indeed, in [5], the estimation is based on a least squares minimization between the empirical F-madogram (see [35]) and its theoretical counterpart, computed for several parametric models. However, in real-world applications, asymptotic properties are always difficult to infer, see [67]. Accordingly, it could be interesting to fit spatial models encompassing both asymptotic dependence classes, and let the data speak for themselves. Furthermore, there is a need to develop efficient exploratory tools that may guide the model choice in existing model-based inference methods.

In addition, the observations at spatial locations are often assumed to be independent in time, see, e.g., [45, 46, 84]. However, many extreme environmental processes observations show: (i) a spatial dependence structure, which means that geographically close locations show similar patterns (ii) a temporal dependence, which can be noticed from similar high values for two successive time moments such as within hours. As an illustration, Figure 1.2 depicts the daily rainfall maxima for the wet seasons (June−September) from the years 2007−2012 at one fixed grid location in Florida. We observe that it is likely that a high value is followed by a value of a similar magnitude. So, the temporal dependence may be present. Accordingly, the temporal dependence structure should be considered in an appropriate way. Currently, space-time models are still taking up little space in the literature. Max-stable processes have been expanded to quantify extremal dependence in spatio-temporal data, see [24, 40, 54, 64]. As with spatial max-stable processes, pairwise likelihood estimation has been found useful to estimate the parameters of these processes, see, e.g., [41, 54, 64]. Alternatively, [23] introduced a semi-parametric estimation procedure based on a closed form expression of the so-called extremogram [42] to estimate the parameters of space-time max-stable BR process. The extremogram has been estimated nonparametrically by its empirical version, where space and time are separated.

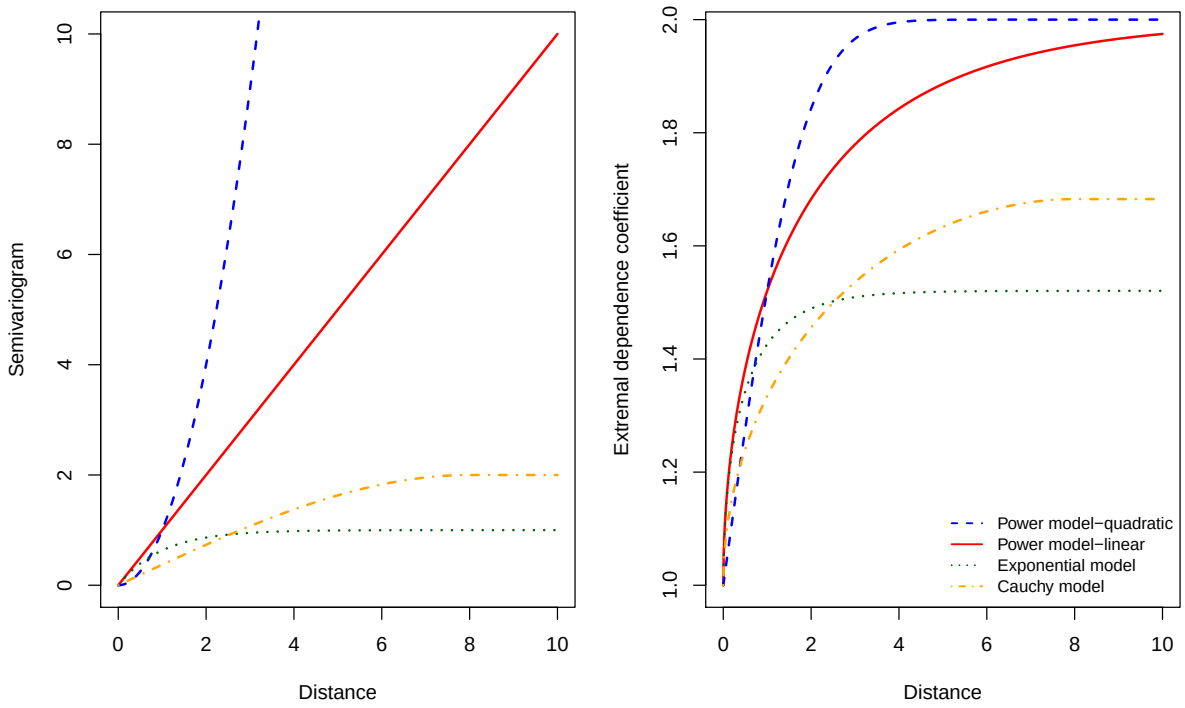


Figure 1.1: Different semivariogram models for a BR process plotted as functions of distance (left panel). The associated extremal dependence functions (right panel).

A constrained weighted linear regression is then applied in order to produce parameter estimates. A big advantage of this semi-parametric method is the substantial reduction of computation time compared to the pairwise likelihood estimation used so far. However, an implicit difficulty in any extreme value analysis is the limited amount of data for model estimation, see, e.g., [32]. Hence, inference on the extremogram is difficult because few observations are available as the threshold approaches 1. Consequently, the semi-parametric estimates obtained by [23] showed a larger bias than pairwise likelihood estimates and are sensitive to the choice of the threshold used for the extremogram. Accordingly, the surrogates of existing estimation techniques should be welcomed.

1.2 Main realizations of the thesis

1.2.1 Hypothesis testing for the mixture parameter a of a spatial max-mixture model

Many efforts has been previously placed on determining the appropriate dependence class for modeling spatial extremes. For instance, in [9], a madogram-based test has been proposed to test the asymptotic independence of bivariate maxima vectors with a generalization to the spatial context (Gaussian and max-stable processes). Such a test could facilitate the modeling of spatial data by a random field with appropriate extremal behavior. In this context, we provide

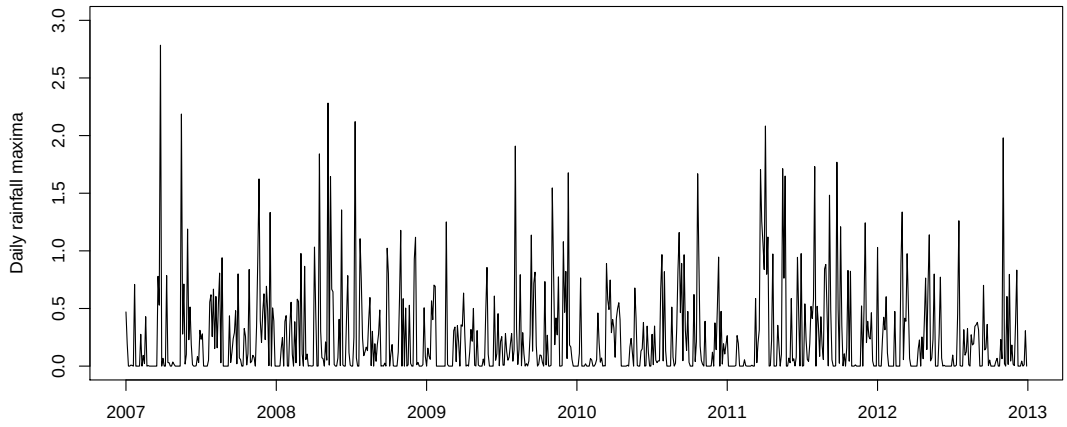


Figure 1.2: Daily rainfall maxima in inches taken over hourly accumulated measurements from 2007−2012 for a fixed location in Florida, USA.

a parametric model-based approach for testing hypothesis on the mixture parameter of a max-mixture process. The latter is defined as $Z(\boldsymbol{s}) = \max\{aX(\boldsymbol{s}), (1-a)Y(\boldsymbol{s})\}$, $\boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^2$, with $\{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ is a max-stable process, $\{Y(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ is an asymptotically independent process and the mixture parameter $a \in [0,1]$. Accordingly, a controls the level of the asymptotically dependent part present in this process.

To test the hypothesis $H_0 : a = a_0$ versus $H_1 : a \neq a_0$, for some specified value $a_0 \in$[0,1], we consider two statistics: the composite likelihood ratio statistic LR_a and the Z-test statistic Z_a. Pairwise likelihood is adopted for estimation purposes. To that aim, we consider the following usual testing framework. Assume that we observe the spatial max-mixture process Z at D locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_D$ and T times $t_1, \ldots, t_T$, where the observations are assumed to be independent in time. Suppose that the parameters of a max-mixture model $\boldsymbol{\vartheta} \in \mathbb{R}^q$ is partitioned as $\boldsymbol{\vartheta} = (\boldsymbol{\gamma}, \boldsymbol{\eta}) \in \mathbb{R}^{q_1} \times \mathbb{R}^{q_2}$, with $q_1 + q_2 = q$, and that we want to test whether the null hypothesis $H_0 : \boldsymbol{\gamma} = \boldsymbol{\gamma}^*$ holds (i.e., reject it, or fail to reject it). The parameter $\boldsymbol{\gamma} \in \mathbb{R}^{q_1}$ is the parameter of interest, while $\boldsymbol{\eta} \in \mathbb{R}^{q_2}$ acts as a nuisance parameter. Let $\widehat{\boldsymbol{\vartheta}} = (\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\eta}})$ denotes the unrestricted maximum pairwise likelihood estimator. Under some regularity conditions, for large T, $\widehat{\boldsymbol{\vartheta}}$ is asymptotically normally distributed (see, e.g., [27, 99]), i.e., $\widehat{\boldsymbol{\vartheta}} \xrightarrow{\mathcal{D}} \mathcal{N}_q\left(\boldsymbol{\vartheta}, \mathcal{G}^{-1}(\boldsymbol{\vartheta})\right)$, where $\mathcal{N}_q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the q-dimensional normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$ and $\xrightarrow{\mathcal{D}}$ denotes the convergence in distribution. The asymptotic variance is given by

$$\mathcal{G}^{-1}(\boldsymbol{\vartheta}) = \mathcal{H}^{-1}(\boldsymbol{\vartheta})\mathcal{J}(\boldsymbol{\vartheta})\mathcal{H}^{-1}(\boldsymbol{\vartheta}),$$

where $\mathcal{G}(\boldsymbol{\vartheta})$ is the Godambe information matrix, $\mathcal{H}(\boldsymbol{\vartheta}) = \mathbb{E}\left\{-\nabla^2\ell(\boldsymbol{\vartheta})\right\}$ is called the sensitivity matrix, $\mathcal{J}(\boldsymbol{\vartheta}) = \mathbb{V}ar\{\nabla\ell(\boldsymbol{\vartheta})\} = \mathbb{E}\{\nabla\ell(\boldsymbol{\vartheta})\nabla^t\ell(\boldsymbol{\vartheta})\}$ is called the variability matrix, and $\ell(.)$ is the (weighted) pairwise log-likelihood. We denote by $\widehat{\boldsymbol{\vartheta}}^* = (\boldsymbol{\gamma}^*, \widehat{\boldsymbol{\eta}}^*)$ the maximum pairwise likelihood estimator under the null hypothesis, i.e., $\widehat{\boldsymbol{\eta}}^*$ is the maximum pairwise likelihood

estimator of $\boldsymbol{\eta}$ when $\boldsymbol{\gamma}$ is held fixed to the value $\boldsymbol{\gamma}^*$. A two-sided pairwise likelihood ratio test may be based on the statistic (see, e.g. [70, 99])

$$LR = 2\left\{\ell(\widehat{\boldsymbol{\vartheta}}) - \ell(\widehat{\boldsymbol{\vartheta}}^*)\right\} \xrightarrow{\mathcal{D}} \sum_{j=1}^{q_1} c_j W_j,$$

where the W_j's are independent χ^2_1 random variables, and the c_j's are the eigenvalues of the matrix $\{\boldsymbol{\mathcal{H}}^{\gamma\gamma}\}^{-1}\boldsymbol{\mathcal{G}}^{\gamma\gamma}$ evaluated under the null hypothesis, where $\boldsymbol{\mathcal{H}}^{\gamma\gamma}(\boldsymbol{\vartheta})$ and $\boldsymbol{\mathcal{G}}^{\gamma\gamma}(\boldsymbol{\vartheta})$ denote, respectively, the $q_1 \times q_1$ submatrices of the inverse of $\boldsymbol{\mathcal{H}}(\boldsymbol{\vartheta})$ and $\boldsymbol{\mathcal{G}}(\boldsymbol{\vartheta})$ with elements corresponding to $\boldsymbol{\gamma}$. Many different adjustments have been proposed in the literature to recover an asymptotic chi-squared distribution $\chi^2_{q_1}$ when $q_1 > 1$, see, e.g., [28, 83, 88]. Nevertheless, simulation-based techniques could be used to approximate the quantiles of the limit $\sum_{j=1}^{q_1} c_j W_j$. Accordingly, to test $H_0 : a = a_0$ versus $H_1 : a \neq a_0$, $a_0 \in [0,1]$, we set $\boldsymbol{\gamma} = a$. For this special case, we adopt the following two statistics:

(i) The pairwise likelihood ratio statistic with $q_1 = 1$ can be expressed as

$$LR_a = c^{-1} LR \xrightarrow{\mathcal{D}} \chi^2_1,$$

where the constant c is computed by the same manner described above, that is, $c = \{\boldsymbol{\mathcal{H}}^{aa}\}^{-1}\boldsymbol{\mathcal{G}}^{aa}$, where $\boldsymbol{\mathcal{H}}^{aa}(\boldsymbol{\vartheta})$ and $\boldsymbol{\mathcal{G}}^{aa}(\boldsymbol{\vartheta})$ denote, respectively, the 1×1 submatrices of the inverse of $\boldsymbol{\mathcal{H}}(\boldsymbol{\vartheta})$ and $\boldsymbol{\mathcal{G}}(\boldsymbol{\vartheta})$ with elements corresponding to a.

(ii) The Z-test statistic which is straightforwardly derived from the central limit theorem (CLT) for maximum pairwise likelihood estimators.

$$Z_a = \frac{\widehat{a} - a}{\sqrt{\boldsymbol{\mathcal{G}}^{aa}(\widehat{\boldsymbol{\vartheta}})}} \xrightarrow{\mathcal{D}} N\{0,1\},$$

where $\boldsymbol{\mathcal{G}}^{aa}(\widehat{\boldsymbol{\vartheta}})$ denotes a 1×1 submatrix of the inverse of $\boldsymbol{\mathcal{G}}(\widehat{\boldsymbol{\vartheta}})$ pertaining to a.

In a simulation study, we obtain the power curves when the corresponding null hypotheses are false (i.e. the proportion of null hypotheses rejected) for the two statistics: LR_a and Z_a. Generally, we notice that the performance of the two statistics is satisfactory. As anticipated, the power to reject asymptotic dependence, i.e., $H_0 : a = 1$ (respectively, asymptotic independence, i.e., $H_0 : a = 0$) improves as $a_0 \to 0$ (respectively, $a_0 \to 1$), although the tests seem not very powerful in these two cases. Furthermore, these likelihood-based tests statistics can control the type I error rate α. Finally, to illustrate the benefits of the suggested testing approach, we analyze daily rainfall data (in millimeters) recorded during 1972−2014 at 38 monitoring stations (which are shown by black cross-symbols in Figure 1.3) in the East of Australia using exceedances over a large threshold. We show that our testing procedure could be an effective tool for model validation on the mixing parameter a. Testing when all model parameters are of interest is also discussed.

1.2.2 F^{λ}-madogram for a spatial max-mixture model

Based on the definition of F^{λ}-madogram [15], we derive a closed form expression of the F^{λ}-madogram for max-mixture model containing the parameters of interest.

Proposition 1.1. *Let $\{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$, $\mathcal{S}\subset\mathbb{R}^2$ be a simple max-stable process, with extremal dependence coefficient function θ_X, and $\{Y(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ be an inverted max-stable process with tail dependence coefficient $\eta = 1/\theta_Y$, where θ_Y is the extremal dependence coefficient function of the latent max-stable process. Then, for any spatial lag $\boldsymbol{h} = \boldsymbol{s}_1 - \boldsymbol{s}_2$, $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathbb{S}$, the F^{λ}-madogram of the spatial max-mixture process $\{Z(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathbb{S}}$ is given for $a \neq 1$ by*

$$\nu_{F^{\lambda}}(\boldsymbol{h}) = \frac{\lambda}{1+\lambda} - \frac{2\lambda}{a(\theta_X(\boldsymbol{h})-1)+1+\lambda} + \frac{\lambda}{a\theta_X(\boldsymbol{h})+\lambda} - \frac{\lambda\theta_Y(\boldsymbol{h})}{(1-a)\theta_Y(\boldsymbol{h})+a\theta_X(\boldsymbol{h})+\lambda}\beta\left(\frac{a\theta_X(\boldsymbol{h})+\lambda}{1-a}, \theta_Y(\boldsymbol{h})\right), \quad \lambda \geq 0$$

where $\beta(.,.)$ is the beta function.

As a consequence of the above formula for $\nu_{F^{\lambda}}(\boldsymbol{h})$, we easily recover the expressions of the following special cases:

(i) By letting $a \to 1$, the F^{λ}-madogram corresponds to a simple max-stable process $\{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ and switches to the extremal dependence coefficient $\theta_X(\boldsymbol{h})$, i.e.,

$$\nu_{F^{\lambda}}(\boldsymbol{h}) = \frac{\lambda}{\lambda+1}\frac{\theta_X(\boldsymbol{h})-1}{\lambda+\theta_X(\boldsymbol{h})},$$

where $\nu_{F^{\lambda}}(\boldsymbol{h}) \in \left[0, \frac{\lambda}{(1+\lambda)(2+\lambda)}\right]$, see [15].

(ii) By setting $a = 0$, the F^{λ}-madogram corresponds to an inverted max-stable process $\{Y(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ and switches to the extremal dependence coefficient $\theta_Y(\boldsymbol{h})$, i.e.,

$$\nu_{F^{\lambda}}(\boldsymbol{h}) = \frac{1}{1+\lambda} - \frac{\lambda\theta_Y(\boldsymbol{h})}{\lambda+\theta_Y(\boldsymbol{h})}\beta\left(\lambda, \theta_Y(\boldsymbol{h})\right).$$

(iii) By setting $\lambda = 1$, we obtain formula of F-madogram for max-mixture process $\{Z(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$, see [5].

Accordingly, the F^{λ}-madogram can detect both dependence structures: asymptotic dependence and independence.

1.2.3 F^{λ}-madogram estimation procedure for a spatial max-mixture model parameters

When modeling dependence in spatial extremes, imposing a specific type of asymptotic behavior has important consequences on the estimation of return levels for spatial functionals. So, it could be interesting to fit spatial models encompassing both asymptotic dependence classes, and let the data speak for themselves. Thus, we develop a novel estimation procedure based on the

F^{λ}-madogram to fit spatial max-mixture models, which combine max-stable and inverted max-mixture processes into a single family, in order to capture both asymptotic dependence and independence in spatial extremes.

The estimation procedure consists of using a moment-based approach to estimate the mixture parameter a and the bivariate dependence summaries $(\theta_X(\boldsymbol{h}), \theta_Y(\boldsymbol{h}))$ for the two processes (max-stable, inverted max-stable). More precisely, we split the estimation procedure into two steps: we first estimate the extremal dependence coefficient functions $\theta_X(\boldsymbol{h})$ and $\theta_Y(\boldsymbol{h})$, with a fixed mixing parameter a, and then we estimate a with the former estimates of $\theta_X(\boldsymbol{h})$ and $\theta_Y(\boldsymbol{h})$. In other words, we may write the F^{λ}-madogram as a function of a, λ, θ_X and θ_Y; that is, $\nu_{F^{\lambda}}(\boldsymbol{h}) = g(a, \lambda, \theta_X(\boldsymbol{h}), \theta_Y(\boldsymbol{h}))$. The idea beyond our procedure is that θ_X and θ_Y may be estimated by $\widehat{\theta}_X$ and $\widehat{\theta}_Y$, minimizing the square difference between $g(a, \lambda, \theta_X(\boldsymbol{h}), \theta_Y(\boldsymbol{h}))$ and its empirical counterpart. A similar approach has been proposed by [15] to estimate the extremal dependence coefficient function in a max-stable setting. Then we can estimate a by $\widehat{a}$ such that the empirical version of the $F^{\lambda'}$-madogram is the closest to $g(a, \lambda', \widehat{\theta}_X(\boldsymbol{h}), \widehat{\theta}_Y(\boldsymbol{h}))$.

The proposed estimation procedure does not provide a full characterization of the max-mixture process. It allows estimating the mixture parameter a and bivariate summaries of the two components in the mixture $(\theta_X(\boldsymbol{h}), \theta_Y(\boldsymbol{h}))$. While the likelihood-based procedures can be used to estimate a full generative spatial model and not only bivariate dependence summaries. So, the proposed approach may be viewed as an intermediate approach between a completely nonparametric approach based on empirical extremal coefficients, which have high variability and do not yield a valid spatial model, and a fully parametric approach, which is very efficient and yields a valid spatial model. Therefore, we employ our procedure for selection and diagnostic purposes.

We establish the consistency of the resulting estimates from the proposed F^{λ}-madogram estimation procedure. Then, in a simulation study, we assess the performance of the resulting estimates $\left(\widehat{a}, \widehat{\theta}_X(\boldsymbol{h}), \widehat{\theta}_Y(\boldsymbol{h})\right)$. It shows that the proposed estimation procedure performs well, even when considering the boundary values for $a \in \{0, 1\}$. In addition, an indication for asymptotic normality is given numerically. Finally, we apply our inference framework in an analysis of monthly maxima of daily rainfall data collected in the East of Australia. Our procedure works as a guiding tool for pairwise likelihood estimation. It offers an exploratory assessment for model selection in a preliminary step and model checking in the validation step. Furthermore, we perform a small prediction application at ungauged locations. Denoting by $\{Z_M(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$, $\mathcal{S} \subset \mathbb{R}^2$ the site-wise monthly maxima process of the max-mixture rainfall process $Z(\boldsymbol{s})$. An interesting question might be "given that the process $Z_M(\boldsymbol{s})$ exceeds a high threshold z at location $\boldsymbol{s}_2$, how likely is it high at a (nearby) location $\boldsymbol{s}_1$?". More precisely, how can we predict

$$\mathcal{P}(z) = \mathbb{P}[Z_M(\boldsymbol{s}_1) > z | Z_M(\boldsymbol{s}_2) > z]?,$$

where z denotes a high threshold. In the following lemma, we express the bivariate conditional exceedance probability $\mathcal{P}(z)$ for the max-mixture process $Z(\boldsymbol{s})$ in terms of the parameters: a, θ_X and θ_Y.

Lemma 1.1. *The bivariate conditional exceedance probability of a max-mixture process* $Z(\boldsymbol{s})$

is given by

$$\mathcal{P}(z) = \frac{1 - 2e^{-\frac{1}{z}} + e^{-\frac{a\theta_X(\boldsymbol{h})}{z}} \left\{ -1 + 2e^{-\frac{1-a}{z}} + \left[1 - e^{-\frac{1-a}{z}} \right]^{\theta_Y(\boldsymbol{h})} \right\}}{1 - e^{-\frac{1}{z}}}, \ a \in [0,1].$$

where $\boldsymbol{h} = \boldsymbol{s}_1 - \boldsymbol{s}_2$ *is the distance vector between the locations* $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{S}$.

Since the aim of spatial analysis is usually to enable prediction at unobserved location, we predict $\mathcal{P}(z)$ at three "unused" sites $\boldsymbol{s}_1$ using our proposed estimation methodology and pairwise likelihood method with various values of the threshold z. The geographical locations of the unused sites $\boldsymbol{s}_1$ are shown by colored numbers in Figure 1.3. Afterward, the goodness-of-fit is assessed based on empirical counterparts of $\mathcal{P}(z)$ at the unused locations. The results are promising with a good agreement overall.

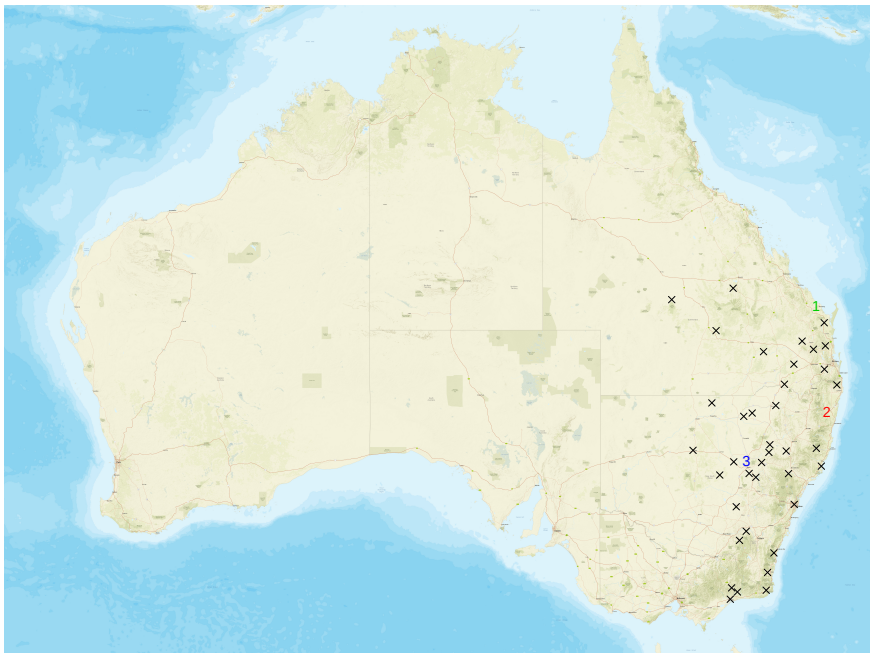


Figure 1.3: Geographical locations of 41 meteorological stations located in the East of Australia. The 38 stations with cross symbols are used for model inference. The other 3 stations labeled by colored numbers {1,2,3} are put aside for assessing the goodness-of-fit.

1.2.4 Semi-parametric estimation for space-time max-stable processes

We propose two novel semi-parametric estimation schemes for space-time max-stable processes based on the spatio-temporal F-madogram $\nu_F(\boldsymbol{h}, l)$. Let $X := \{X(\boldsymbol{s},t) : (\boldsymbol{s},t) \in \mathcal{S} \times \mathcal{T}\}$ be a sta-

tionary space-time max-stable random process, then the spatio-temporal version of F-madogram (originally due to [35]) is defined by

$$\nu_F(\boldsymbol{h}, l) := \nu_F(\boldsymbol{s}_1 - \boldsymbol{s}_2, t_1 - t_2) = \frac{1}{2}\mathbb{E}[|F(X(\boldsymbol{s}_1, t_1)) - F(X(\boldsymbol{s}_2, t_2))|],$$

where $\boldsymbol{h}$ is the spatial lag, l is the temporal lag and F is the standard Fréchet probability distribution function. We assume that the locations lie on a regular grid and that time points are equidistant. Using the link between the extremal coefficient and the F-madogram in the context of a space-time max-stable process, we suggest the following estimation schemes:

(i) ***Scheme 1***: we estimate spatial and temporal parameters separately. So, we consider the spatial evolution of $X(\boldsymbol{s}, t)$ at a given time of reference (a merely spatial process) or its evolution over time at a given location of reference (a merely temporal process). Based on nonlinear least squares, we minimize the squared difference between the empirical estimates of purely spatial/temporal F-madograms and their model-based counterparts. [24] proposed a similar approach for fitting isotropic space-time max-stable BR process based on the extremogram which is the natural extreme analog of the correlation function for stationary processes, see [42].

(ii) ***Scheme 2***: we generalize the nonlinear least squares to estimate spatial and temporal parameters simultaneously. Thus, we consider how $X(\boldsymbol{s}, t)$ evolves in both space and time.

Let us remark that similar estimation fashions can be found in the geostatistics literature, which are based on the basic tool in geostatistics, the semivariogram, see Section 2.1.2. In a simulation study, we evaluate our new semi-parametric estimation procedure. We compute performance metrics: the mean estimate, the root mean squared error (RMSE) and the mean absolute error (MAE). Altogether, we observe that the estimates are close to the true values. In addition, we remark that our estimation methodology outperforms the one suggested by [23]. Finally, we apply our semi-parametric estimation procedure to quantify the extremal behavior of daily rainfall maxima in a region of the State of Florida, see Figure 1.4. We consider various space-time max-stable models, where the Akaike Information Criterion (AIC) under the framework of least squares estimation is used to select the best-fitting model, see [13]. Lower values of AIC indicate a better fit. Moreover, permutation tests are used to determine the range of clear spatial and temporal dependence.

1.3 Outline of the thesis

The remainder of the thesis is organized as follows. ***Chapter 2*** introduces theoretical foundations on spatial and spatio-temporal modeling of extremes, which will be used in the subsequent chapters. In ***Chapter 3***, we consider hypothesis testing for the mixture parameter a of a spatial max-mixture model using two classical statistics: the Z-test statistic Z_a and the pairwise likelihood ratio statistic LR_a. A novel estimation procedure in the family of spatial max-mixture models based on a closed form expression of the F^λ-madogram is described and analyzed in

Chapter 4. In ***Chapter 5***, we develop a novel semi-parametric inference approach for space-time max-stable processes based on the spatio-temporal F-madogram. Concluding remarks are discussed in ***Chapter 6***. In addition, we mention some remaining problems which are subject to future research.

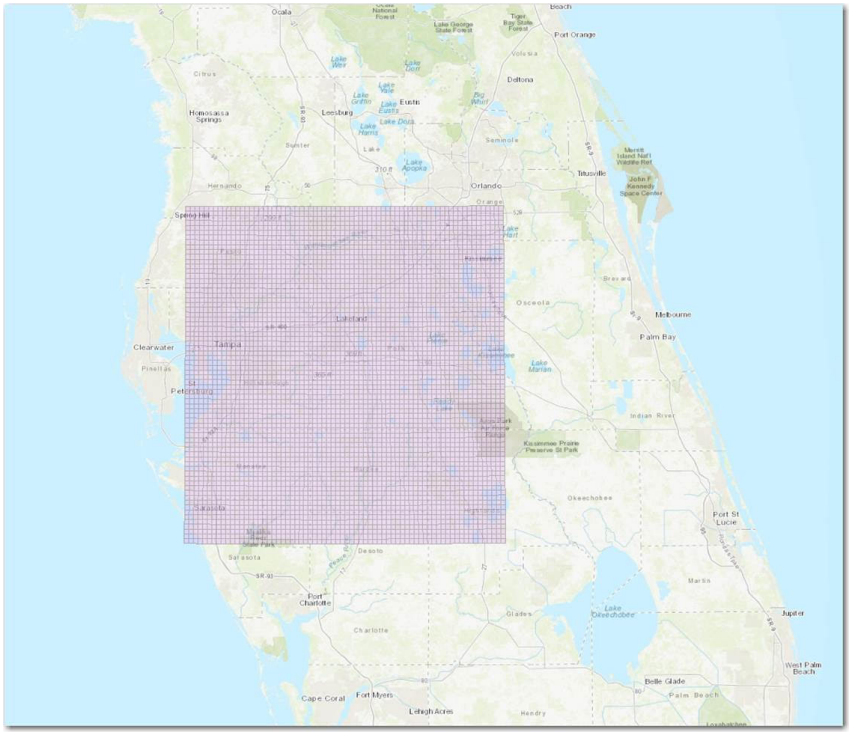


Figure 1.4: Radar rainfall observation area in the State of Florida, USA. Source: Southwest Florida Water Management District (SWFWMD).

Chapter 2

Statistical modeling for spatial and spatio-temporal extremes

In this chapter, we introduce theoretical foundations, which will be used in the subsequent chapters. In Section 2.1, we recall some well-known results on spatial stochastic processes. In Section 2.2, we define spatial extremes processes: max-stable, inverted max-stable and max-mixture processes. Furthermore, some classical extremal dependence measures are presented. In Section 2.3, we state some general definitions for extreme space-time processes.

2.1 Fundamentals of spatial stochastic processes

In this section, we provide some basics on spatial random processes: definitions, notations, examples and related properties. For a detailed description, see, e.g., [36, 37, 56, 95] and the references therein.

2.1.1 Definitions, notations and important properties

Definition 2.1. ***(A spatial stochastic process)*** *A spatial stochastic process X is a collection of random variables $\{X(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\} =: \{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ defined on a common probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and indexed by a parameter $\boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^d$. The points $\boldsymbol{s}$ denote the spatial coordinates and are called "sites" or "locations" or "stations". Different notations can be adopted in specific frameworks:*

(i) When X is a purely temporal process, it is common to replace $\boldsymbol{s} \in \mathcal{S}$ by $t \in \mathcal{T} \subset \mathbb{R}^+$ to indicate time, where the points t denote the temporal coordinates and are called "times" or "moments". The process is usually referred to as a time series.

(ii) When X is a spatio-temporal process, we usually write $\left\{X(\boldsymbol{s},t) : (\boldsymbol{s},t) \in \mathcal{S}\times\mathcal{T} \subset \mathbb{R}^d \times \mathbb{R}^+\right\}$, where the space $\mathcal{S}\times\mathcal{T}$ is the spatio-temporal domain.

Usually, a spatial process is observed at a set of locations. For example, the rainfall measurements in a region are observed at the points where the monitoring stations are located. However, the rainfall process is distributed continuously in space. A spatial process in dimension $d \geq 2$ is

also called a random field. Generally, throughout this thesis, we will consider spatial processes when $d = 2$.

The first two moments of a stochastic process $\{X(\boldsymbol{s})\}_{s\in\mathcal{S}}$ are useful to describe it, that is,

- the expectation $\mu(\boldsymbol{s}) = \mathbb{E}\{X(\boldsymbol{s})\},\ \boldsymbol{s} \in \mathcal{S}$,
- the covariance function $\mathbb{C}ov(\boldsymbol{s}_1, \boldsymbol{s}_2) = \mathbb{E}\{X(\boldsymbol{s}_1)X(\boldsymbol{s}_2)\} - \mu(\boldsymbol{s}_1)\mu(\boldsymbol{s}_2),\ \boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{S}$, also called covariogram.

The variance is defined as $\mathbb{V}ar(\boldsymbol{s}) = \mathbb{C}ov(\boldsymbol{s}, \boldsymbol{s})$ and the correlation function is given by $\rho(\boldsymbol{s}_1, \boldsymbol{s}_2) = \mathbb{C}ov(\boldsymbol{s}_1, \boldsymbol{s}_2)\{\mathbb{V}ar(\boldsymbol{s}_1)\mathbb{V}ar(\boldsymbol{s}_2)\}^{-1/2},\ \boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{S}$. An important class of stochastic processes is formed by Gaussian processes, which are fully characterized by $\mu(\boldsymbol{s})$ and $\mathbb{C}ov(\boldsymbol{s}_1, \boldsymbol{s}_2)$.

Definition 2.2. ***(Gaussian process)*** *The stochastic process $\{X(\boldsymbol{s})\}_{s\in\mathcal{S}}$ is a Gaussian process if for any $k \in \mathbb{N}$, $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k \in \mathcal{S}$, the joint distribution of of the random vector $\boldsymbol{X} = (X(\boldsymbol{s}_1), \ldots, X(\boldsymbol{s}_k))^t$ is a k-dimensional Gaussian distribution with mean vector $\boldsymbol{\mu} = \mathbb{E}(\boldsymbol{X})$ and covariance matrix $\boldsymbol{\Sigma} = \left(\mathbb{C}ov(\boldsymbol{s}_i, \boldsymbol{s}_j)\right)_{i,j},\ i, j = 1, \ldots, k$.*

Provided the covariance matrix $\boldsymbol{\Sigma}$ is nonsingular, $\boldsymbol{X}$ has a Gaussian probability density function (p.d.f.) given by

$$f(\boldsymbol{x}) = \frac{\partial^k}{\partial_1, \ldots, \partial_k} F(\boldsymbol{x}) = (2\pi)^{-k/2} (\det \boldsymbol{\Sigma})^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right\}, \tag{2.1}$$

where ∂_i is the differentiation with respect to the variable x_i, F is the Gaussian cumulative distribution function (c.d.f.) of $\boldsymbol{X}$ and $\boldsymbol{x}$ are possible values of $\boldsymbol{X}$. Gaussian processes have been used extensively in classical geostatistics due to their appealing theoretical properties, tractability in high dimensions, explicit conditional distributions, and simulation facility, see [57].

When modeling spatial phenomena, various forms of a stationary process may be assumed. Definition 2.3 states three "levels" of stationarity. In what follows, $\overset{\mathcal{D}}{=}$ stands for equality in distribution.

Definition 2.3. ***(Stationarity)***

(i) ***(Strict stationarity)*** *A spatial stochastic process $\{X(\boldsymbol{s})\}_{s\in\mathcal{S}}$ is called strictly/strongly stationary if for any $k \in \mathbb{N}$, $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k \in \mathcal{S}$, $\boldsymbol{s}_1 + \boldsymbol{h}, \ldots, \boldsymbol{s}_k + \boldsymbol{h} \in \mathcal{S}$ with $\boldsymbol{h} \in \mathbb{R}^d$,*

$$(X(\boldsymbol{s}_1), \ldots, X(\boldsymbol{s}_k))^t \overset{\mathcal{D}}{=} (X(\boldsymbol{s}_1 + \boldsymbol{h}), \ldots, X(\boldsymbol{s}_k + \boldsymbol{h}))^t,$$

In other words, strict stationarity implies that the finite-dimensional distribution is unaffected by the translation of an arbitrary quantity $\boldsymbol{h} \in \mathbb{R}^d$.

(ii) ***(Weak stationarity)*** *A spatial stochastic process $\{X(\boldsymbol{s})\}_{s\in\mathcal{S}}$ is called weakly/second-order stationary or stationary in the broad sense, if its mean is constant, that is $\mu(\boldsymbol{s}) = \mu,\ \boldsymbol{s} \in \mathcal{S}$ and if its covariance function depends only on the separation vector, that is $\mathbb{C}ov(\boldsymbol{s}_1, \boldsymbol{s}_2) = \mathbb{C}ov(\boldsymbol{h})$, where $\boldsymbol{h} = \boldsymbol{s}_1 - \boldsymbol{s}_2$, and $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{S}$. In particular, the variance takes the constant value $\mathbb{V}ar(\boldsymbol{s}) = \sigma^2,\ \boldsymbol{s} \in \mathcal{S}$, and the correlation function equals $\rho(\boldsymbol{h}) = \mathbb{C}ov(\boldsymbol{h})/\sigma^2$.*

(iii) ***(Intrinsic stationarity and semivariogram)*** *A spatial stochastic process* $\{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ *is called intrinsically stationary, if its increments are weakly stationary; in particular, there exists a function* $\gamma(\boldsymbol{h})$*, called the semivariogram, for which* $2\gamma(\boldsymbol{h}) = \mathbb{V}ar\left(X(\boldsymbol{s}) - X(\boldsymbol{s}+\boldsymbol{h})\right)$*, provided that* $\boldsymbol{s}, \boldsymbol{s}+\boldsymbol{h} \in \mathcal{S}$*. The function* $2\gamma(\boldsymbol{h})$ *is called the variogram.*

These three types of stationarity relate to shift-invariance property. Nevertheless, they are not equivalent. The strict stationarity implies weak stationarity, which in turn entails intrinsic stationarity. However, the converse is not necessarily true. In the case of Gaussian processes, second-order stationarity is equivalent to strict stationarity. In what follows, unless specified otherwise, we shall say "stationary" for strictly stationary.

The concept of isotropy is closely linked to stationarity, where both are geometric invariance properties of stochastic processes. Similarly, as stationarity is a shift-invariance property, isotropy is a rotation-invariance property.

Definition 2.4. ***(Isotropy)*** *A spatial stochastic process* $\{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ *is called isotropic, if for any* $\mathbb{R}^d$ *isometry* m *with* $m(\boldsymbol{s}_i) \in \mathcal{S}$*,* $i = 1, \ldots, k$*,*

$$(X(\boldsymbol{s}_1), \ldots, X(\boldsymbol{s}_k))^t \overset{\mathcal{D}}{=} (X(m(\boldsymbol{s}_1)), \ldots, X(m(\boldsymbol{s}_k)))^t .$$

If the stochastic process $\{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ is weakly stationary and isotropic, then its covariance function $\mathbb{C}ov(\boldsymbol{s}_1, \boldsymbol{s}_2)$ depends only on $\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$ not on its orientation, where $\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$ denotes the Euclidean distance between sites $\boldsymbol{s}_1, \boldsymbol{s}_2$, that is, $\mathbb{C}ov(\boldsymbol{s}_1, \boldsymbol{s}_2) = \mathbb{C}ov(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|)$.

2.1.2 Correlation functions and semivariograms

In geostatistics, correlation functions and semivariograms are used to express the structure of the spatial dependence present in the realization observed. In a weakly stationary framework, both the correlation functions and the semivariograms are theoretically equivalent. More precisely, we have

$$\gamma(\boldsymbol{s}_1 - \boldsymbol{s}_2) = \sigma^2 \{1 - \rho(\boldsymbol{s}_1 - \boldsymbol{s}_2)\} . \tag{2.2}$$

Therefore, $\gamma(\boldsymbol{s}_1 - \boldsymbol{s}_2)$ can be easily recovered from $\rho(\boldsymbol{s}_1 - \boldsymbol{s}_2)$. Note that if the semivariogram γ is unbounded, a relationship of the form (2.2) with some correlation function ρ cannot hold, (see, e.g., [56]). Correlation functions and semivariograms cannot be an arbitrary functions and must verify specific constraints. More precisely, for any given set $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k$ of arbitrary locations in $\mathcal{S}$ and for any given set of real numbers $a_1, \ldots, a_k \in \mathbb{R}$, correlation functions are nonnegative-definite, that is,

$$\sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j \rho(\boldsymbol{s}_i - \boldsymbol{s}_j) \geq 0,$$

whilst semivariograms are conditionally negative-definite, that is,

$$\sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) \leq 0, \ \sum_{i=1}^{k} a_i = 0.$$

Examples of correlation functions

The following examples are isotropic parametric families of correlation functions on $\mathbb{R}^d$ traditionally used in geostatistics, where $\rho(\boldsymbol{s}_1, \boldsymbol{s}_2) = \rho(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|), \forall \boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^d$. Other models are presented in [1, 56, 103]. Below, c_1 denotes a smoothness parameter (also called the shape parameter), c_2 denotes a range parameter (sometimes called the scale parameter), $\Gamma(.)$ is the gamma function, J_{c_1} and K_{c_1} denote, respectively, the Bessel and the modified Bessel functions of the second kind with order c_1 and d is the dimension of the spatial process.

- ***Bessel***: $\rho(\boldsymbol{s}_1, \boldsymbol{s}_2) = \left(\frac{2c_2}{\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|}\right)^{c_1} \Gamma(c_1 + 1) J_{c_1}\left(\frac{\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|}{c_2}\right),\ c_2 > 0,\ c_1 \geq \frac{d-2}{2},$
- ***Cauchy***: $\rho(\boldsymbol{s}_1, \boldsymbol{s}_2) = \left[1 + \left(\frac{\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|}{c_2}\right)^2\right]^{-c_1},\ c_1, c_2 > 0,$
- ***Powered exponential (stable model)***: $\rho(\boldsymbol{s}_1, \boldsymbol{s}_2) = \exp\left[-\left(\frac{\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|}{c_2}\right)\right]^{c_1},\ 0 < c_1 \leq 2,\ c_2 > 0,$
- ***Whittle-Matérn***: $\rho(\boldsymbol{s}_1, \boldsymbol{s}_2) = \frac{2^{1-c_1}}{\Gamma(c_1)}\left(\frac{\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|}{c_2}\right)^{c_1} K_{c_1}\left(\frac{\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|}{c_2}\right),\ c_1, c_2 > 0,$
- ***Spherical***: $\rho(\boldsymbol{s}_1, \boldsymbol{s}_2) = \left[1 - 1.5\left(\frac{\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|}{c_2}\right) + 0.5\left(\frac{\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|}{c_2}\right)^3\right] \mathbb{1}_{\{\|\boldsymbol{s}_1 - \boldsymbol{s}_2\| \leq c_2\}},\ c_2 > 0.$

In what follows, we will call the range the smallest distance $\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$ for which $\rho(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|) = 0$ (if it exists). Also, let us remark that setting $c_1 = 1$ (respectively, $c_1 = 2$) in the powered exponential model, leads to exponential (respectively, Gaussian) models. Figure 2.1 depicts the correlation functions for the first four parametric families introduced above with various smoothness parameters c_1. For example, with this particular setting for the correlation parameters, the slope of the powered exponential correlation function near the origin is steeper than Whittle-Matérn, leading to a rougher dependence structure. Complete positive (respectively, negative) dependence is reached when $\rho(\boldsymbol{s}_1, \boldsymbol{s}_2) = 1$ (respectively, -1), whereas independence occurs when $\rho(\boldsymbol{s}_1, \boldsymbol{s}_2) = 0$. Most parametric correlation families do not allow negative values, see [86].

Examples of semivariograms

In many applications the semivariogram is taken to be stationary, i.e., it depends only on $\boldsymbol{s}_1 - \boldsymbol{s}_2$, and isotropic, i.e., it depends only on the length $\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$. As we mentioned above, bounded semivariograms can be constructed from correlation functions according to relation (2.2). For instance, the powered exponential model is defined by

$$\gamma(\boldsymbol{s}_1, \boldsymbol{s}_2) = \sigma^2 \left\{1 - \exp\left[-\left(\frac{\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|}{c_2}\right)\right]^{c_1}\right\},\ 0 < c_1 \leq 2,\ c_2, \sigma^2 > 0,$$

where c_1 and c_2 are defined as above, and σ^2 denotes the sill; the value that the semivariogram model attains at the range. Other bounded models are defined analogously. Another useful class of models, the semivariograms without a sill. These models are unbounded and correspond to an intrinsically stationary stochastic process that is not weakly stationary. For instance, the fractal Brownian motion (FBM) semivariogram (also called the power model) is defined by

$$\gamma(\boldsymbol{s}_1, \boldsymbol{s}_2) = \left(\frac{\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|}{c_2}\right)^{c_1},\ 0 < c_1 < 2,\ c_2 > 0,$$

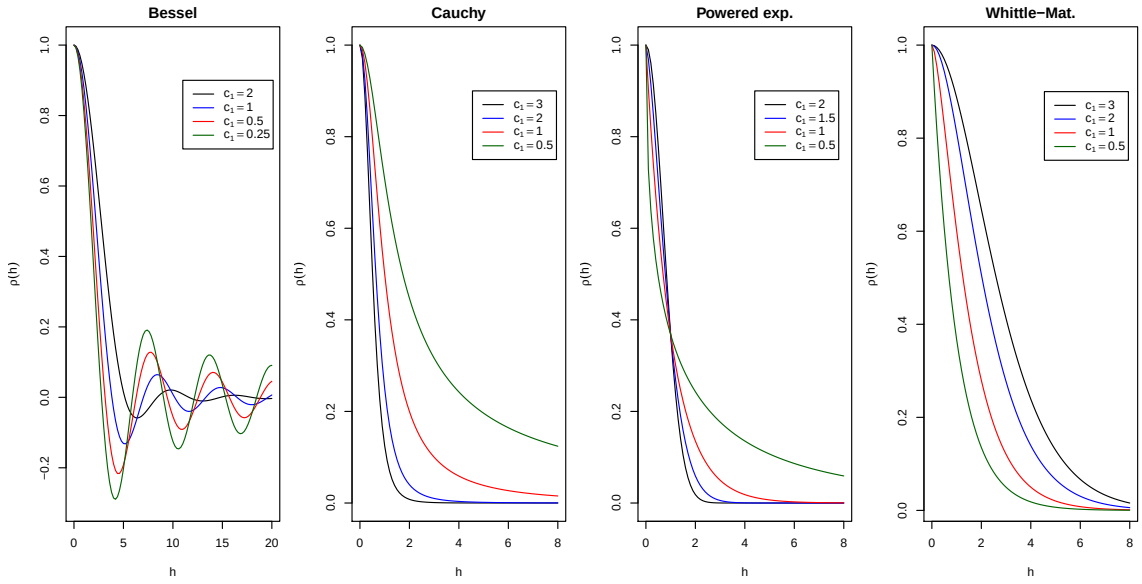


Figure 2.1: Dependence structure of the Bessel, Cauchy, powered exponential and Whittle-Matérn correlation functions - from left to right. The range parameter $c_2 = 1$ while the smoothness parameter c_1 values are given in the legends. The distance $h = \|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$.

where the case $c_1 = 1$ results in the linear model. Figure 2.2 displays the dependence behavior of isotropic Cauchy, powered exponential, Whittle-Matérn and FBM semivariograms. The first three semivariograms are bounded and their counterparts of correlation functions are displayed in Figure 2.1. In addition, we can see that depending on the value of c_1, the FBM model exhibits a large variety of behaviors near the origin.

Empirical semivariogram

In practice, the empirical (or experimental) semivariogram/covariogram is the instrument used to estimate the structure of spatial variability existing in the phenomenon of interest. The empirical classical semivariogram of a stationarity spatial is obtained by implementing the method of moments (MoM).

Definition 2.5. ***(Spatial empirical semivariogram, [76])*** *Let $\{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ be a stationary spatial process, then the empirical classical semivariogram is defined by*

$$\widehat{\gamma}(\boldsymbol{h}) = \frac{1}{2|\mathcal{B}_{\boldsymbol{h}}|} \sum_{(\boldsymbol{s}_i,\boldsymbol{s}_i+\boldsymbol{h})\in\mathcal{B}_{\boldsymbol{h}}} (X(\boldsymbol{s}_i) - X(\boldsymbol{s}_i+\boldsymbol{h}))^2, \tag{2.3}$$

where $|.|$ denotes the cardinality, $X(\boldsymbol{s})$ are the values of the phenomenon of interest at locations $\boldsymbol{s}$ and $\mathcal{B}_{\boldsymbol{h}}$ is the set of pair of locations whose pairwise distances $\boldsymbol{h}$.

Under the intrinsic stationarity hypothesis, the empirical semivariogram $\widehat{\gamma}(\boldsymbol{h})$ is an unbiased estimator of the semivariogram $\gamma(\boldsymbol{h})$, that is, $\mathbb{E}\{\widehat{\gamma}(\boldsymbol{h})\} = \gamma(\boldsymbol{h})$, see Section 5.1.2 in [56].

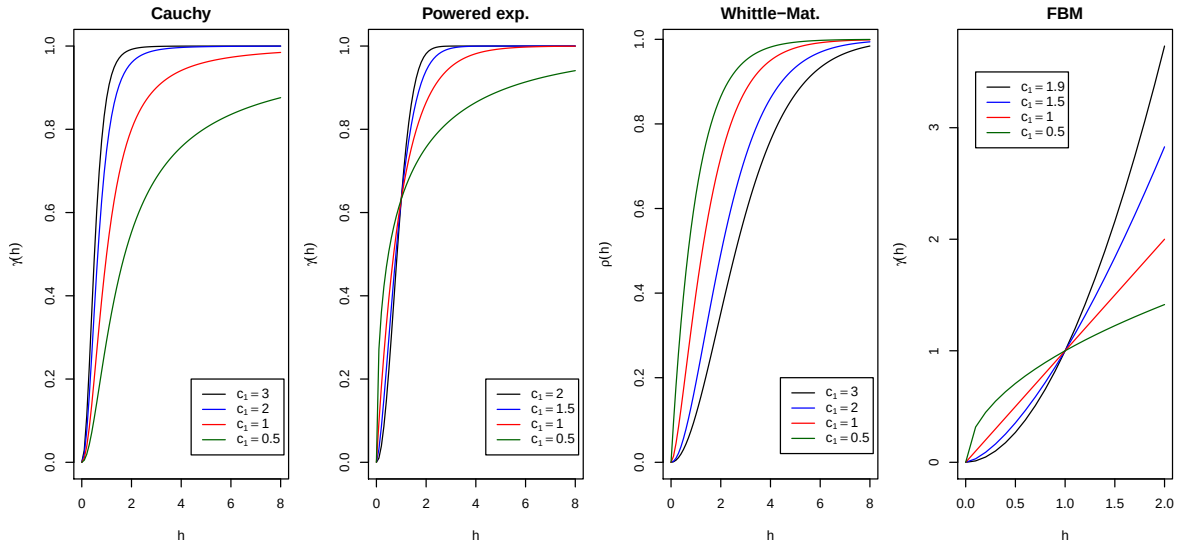


Figure 2.2: Cauchy, powered exponential, Whittle-Matérn and FBM semivariograms - from left to right. For all models, we set the sill and range parameters $\sigma^2 = c_2 = 1$ while the smoothness parameter c_1 values are given in the legends. The distance $h = \|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$.

2.2 Spatial extremes processes: models

Although Gaussian processes are convenient for modeling some phenomena (see, e.g., [56]). However, they are unrealistic for many environmental data, e.g., for precipitation data. Moreover, they have been heavily criticized as being unable to capture asymptotic dependence, see [45]. Indeed, the Gaussian p.d.f., recall 2.1, has exceptionally light tails and therefore can badly underestimate probabilities associated to extreme events. In addition, Gaussian processes are asymptotically independent, meaning that the dependence strength between events observed at two distinct spatial sites vanishes as their extremeness increases. However, when modeling the joint occurrence of extremes over a region, different forms of extremal dependence may arise: asymptotic dependence and asymptotic independence. Therefore, there is a need to develop more suitable models for analyzing spatial extremes. Extreme value theory provides powerful statistical tools for this purpose. Max-stable processes form an extension of extreme value theory to the level of stochastic processes.

The following definition extends the concept of asymptotic independence and dependence between a pair of random variables into the framework of spatial processes, see [33, 93].

Definition 2.6. ***(Asymptotic independence and dependence)*** *Let $\{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ be a stationary spatial process with univariate common marginal distribution function F. The upper tail dependence function χ is defined by*

$$\chi(\boldsymbol{h}) = \lim_{x\to\infty} \mathbb{P}\{X(\boldsymbol{s}) > x | X(\boldsymbol{s}+\boldsymbol{h}) > x\},\ \boldsymbol{s}, \boldsymbol{s}+\boldsymbol{h} \in \mathcal{S}. \tag{2.4}$$

- *If $\chi(\boldsymbol{h}) = 0$, the pair $(X(\boldsymbol{s}), X(\boldsymbol{s}+\boldsymbol{h}))$ is said to be **asymptotically independent**.*
- *If $\chi(\boldsymbol{h}) \neq 0$, the pair $(X(\boldsymbol{s}), X(\boldsymbol{s}+\boldsymbol{h}))$ is said to be **asymptotically dependent**.*

- *The process X is said to be asymptotically independent (respectively, asymptotically dependent) process if $\chi(\boldsymbol{h}) = 0$ (respectively, $\chi(\boldsymbol{h}) \neq 0$) for any $\boldsymbol{h} \in \mathcal{S}$.*

Accordingly, a max-stable process $X(\boldsymbol{s})$ (see Definition 2.7) defined on a spatial domain $\mathcal{S}$ is asymptotically dependent in the sense that $\chi(\boldsymbol{h}) > 0$, for all $\boldsymbol{h} \in \mathcal{S}$. Accordingly, a large event at $\boldsymbol{s}+\boldsymbol{h}$ leads to a non zero probability of a similarly large event at $\boldsymbol{s}$. However, in some applications, it seems that $\chi(\boldsymbol{h}) = 0$ and then asymptotic independence models may be preferred. As mentioned above, Gaussian processes are asymptotically independent. However, they are too restrictive in the bulk of extremal applications, so broader classes of models are needed to allow flexible modeling. For instance, the inverted max-stable processes, see Definition 2.15.

In the sequel, we present the theory of extreme spatial processes: max-stable, inverted max-stable and max-mixture processes. In addition, we describe measures of dependence that are suitable either for asymptotic dependence or independence models, or to help to discriminate between them.

2.2.1 Models for asymptotic dependence: max-stable models

Classical extreme value theory provides support for the use of max-stable models for block-maxima (e.g., annual maxima of daily temperature or precipitation). Indeed, max-stable models form the natural extension of the multivariate extreme value distribution to infinite dimensions. We start with the definition of max-stable processes, see [49, 50, 51].

Definition 2.7. ***(Spatial max-stable process)*** *Let $\{X_i(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S}\}$, $i = 0, 1, 2, \ldots$, denote an independent and identically distributed (i.i.d.) replicates of a random process $X(\boldsymbol{s})$. The process X is called max-stable if there exist sequences of continuous functions $\{a_n(\boldsymbol{s}) > 0\}$ and $\{b_n(\boldsymbol{s})\}$ such that,*

$$\left\{ \bigvee_{i=1}^{n} \frac{X_i(\boldsymbol{s}) - b_n(\boldsymbol{s})}{a_n(\boldsymbol{s})} \right\}_{\boldsymbol{s}\in\mathcal{S}} \overset{\mathcal{D}}{=} \{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}, \; n = 1, 2, \ldots, \tag{2.5}$$

where $\bigvee$ denotes the max-operator.

This definition of max-stable processes offers a natural choice for modeling spatial extremes. Univariate extreme value theory implies that the marginal distributions of $\{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ are ***Generalized Extreme Value GEV***$_{\mu(\boldsymbol{s}),\sigma(\boldsymbol{s}),\xi(\boldsymbol{s})}$ distributed, i.e.,

$$F(x) := \mathbb{P}(X(\boldsymbol{s}) \leq x) = \begin{cases} \exp\left\{ -\left(1 + \xi(\boldsymbol{s})\frac{x-\mu(\boldsymbol{s})}{\sigma(\boldsymbol{s})}\right)^{-1/\xi(\boldsymbol{s})} \right\}, & \xi(\boldsymbol{s}) \neq 0, \\ \exp\left\{ -e^{-\frac{x-\mu(\boldsymbol{s})}{\sigma(\boldsymbol{s})}} \right\}, & \xi(\boldsymbol{s}) = 0, \end{cases} \tag{2.6}$$

provided that $1 + \xi(\boldsymbol{s})\frac{x-\mu(\boldsymbol{s})}{\sigma(\boldsymbol{s})} > 0$, for some location $\mu(\boldsymbol{s}) \in \mathbb{R}$, scale $\sigma(\boldsymbol{s}) > 0$ and shape $\xi(\boldsymbol{s}) \in \mathbb{R}$. The parameter ξ determines the tail behavior of the distribution. Consequently, it divides the GEV into three standardized families of distributions as follows.

- Type I (***Gumbel***, $\xi = 0$): $F(x) = \exp\{-e^{-x}\}$, $x \in \mathbb{R}$.

- Type II (***Fréchet***, $\xi = \alpha^{-1} > 0$): $F(x) = \begin{cases} 0, & x < 0, \\ \exp\{-x^{-\alpha}\}, & x \geq 0. \end{cases}$
- Type III (***Reversed Weibull***, $\xi = -\alpha^{-1} < 0$): $F(x) = \begin{cases} \exp\{-(-x)^{\alpha}\}, & x < 0, \\ 1, & x \geq 0. \end{cases}$

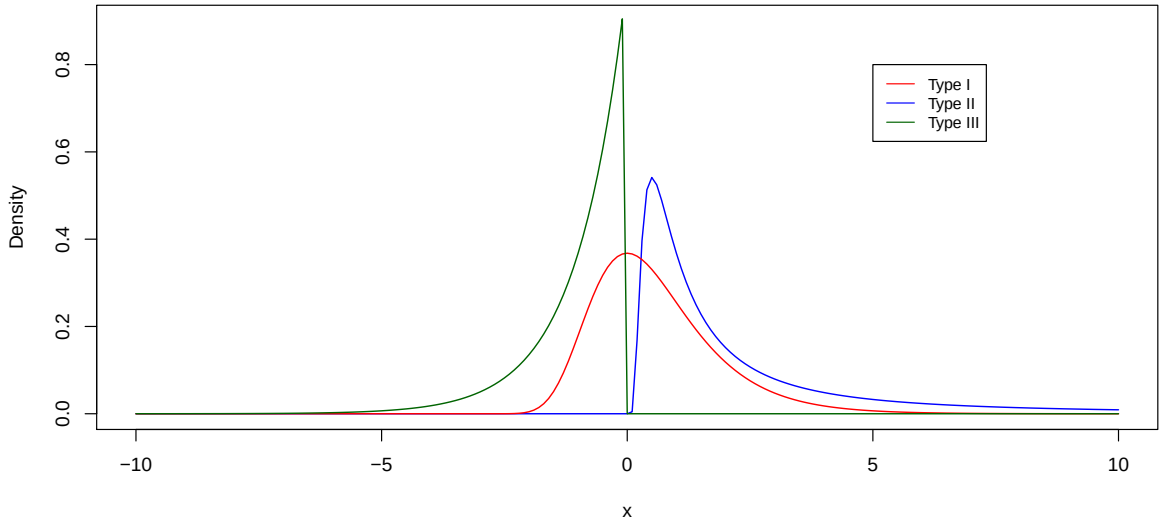


Figure 2.3: Density functions of the GEV distribution families.

Figure 2.3 visualizes the density curves of the three GEV distribution families. Theoretically, there is no loss of generality in transforming the margins to have unit Fréchet distribution, that is, $F(x) =: \mathrm{GEV}_{1,1,1}(x) = \exp\{-x^{-1}\}$, $x > 0$, see [55]. With this choice, the normalizing functions are $a_n = n$ and $b_n = 0$ and the process is called ***simple max-stable process***. For a detailed overview of extreme value theory, we refer the reader to [14, 32, 50]. A simple max-stable process can be constructed using the spectral representation introduced by [49, 90].

Theorem 2.1. ***(Spectral representation for spatial simple max-stable processes)*** *Let X be a simple max-stable process on $\mathcal{S}$. Then there exists $\{\zeta_i\}_{i\geq 1}$ that generates a Poisson point process on $\mathbb{R}^+$ with intensity $\zeta^{-2}d\zeta$ and a sequence $\{U_i(\boldsymbol{s}), \boldsymbol{s} \in \mathcal{S}\}_{i\geq 1}$ of i.i.d. copies of a positive stochastic process $U(\boldsymbol{s})$ that satisfies $\mathbb{E}[U(\boldsymbol{s})] = 1$ for all $\boldsymbol{s} \in \mathcal{S}$ such that*

$$X(\boldsymbol{s}) \overset{\mathcal{D}}{=} \left\{ \bigvee_{i=1}^{\infty} \{\zeta_i U_i(\boldsymbol{s})\} \right\}_{\boldsymbol{s}\in\mathcal{S}}. \tag{2.7}$$

Theorem 2.2. ***(Multivariate maxima)*** *For $D \in \mathbb{N} \setminus \{0\}$, $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_D \in \mathcal{S}$, and $x_1, \ldots, x_D > 0$, the finite D-dimensional distributions of the max-stable process X (2.7) is*

$$\mathbb{P}(X(\boldsymbol{s}_1) \leq x_1, \ldots, X(\boldsymbol{s}_D) \leq x_D) = \exp\left\{ -\mathbb{E}\left[\bigvee_{j=1}^{D} \left\{ \frac{U(\boldsymbol{s}_j)}{x_j} \right\} \right] \right\} =: \exp\{-V_{\boldsymbol{s}_1,\ldots,\boldsymbol{s}_D}(x_1, \ldots, x_D)\}, \tag{2.8}$$

where $V_{s_1,\ldots,s_D}(.)$ is called the ***D-dimensional exponent function****.*

The function $V_{s_1,\ldots,s_D}(.)$ summarizes the extremal dependence structure and satisfies

- homogeneity of order -1, that is, $V_{s_1,\ldots,s_D}(tx_1,\ldots,tx_D) = t^{-1}V_{s_1,\ldots,s_D}(x_1,\ldots,x_D)$, $t > 0$.
- $\max\left\{x_1^{-1},\ldots,x_D^{-1}\right\} \leq V_{s_1,\ldots,s_D}(x_1,\ldots,x_D) \leq x_1^{-1}+\cdots+x_D^{-1}$, where the lower (respectively, upper) bound corresponds to complete dependence (respectively, independence).
- $V_{s_1,\ldots,s_D}(x,\infty,\ldots,\infty) = 1/x$ for any permutation of the D arguments.
- setting $x_i = x$ for all $i = 1,\ldots,D$ yields $V_{s_1,\ldots,s_D}(1,\ldots,1) = \theta_D(\boldsymbol{s}_1,\ldots,\boldsymbol{s}_D)$, where the function $\theta_D(.)$ is known as the ***D-dimensional extremal dependence function***, see [92].

Unlike for univariate case, the function $V_{s_1,\ldots,s_D}(.)$ has no explicit form. Thus, various approaches are exist for estimating it. For instance, a nonparametric estimation procedure has been proposed by [25] in the bivariate case. Also to that aim, the authors in [81] tied together the tools used in classical geostatistics (in particular, the madogram) and extreme value theory. The corresponding p.d.f. can be obtained by computing the derivative of (2.8) with respect to all x_i, $i = 1,\ldots,D$. Namely,

$$f(x_1,\ldots,x_D) = \exp\{-V_{s_1,\ldots,s_D}(x_1,\ldots,x_D)\}\sum_{\eta\in\mathscr{B}_D}(-1)^{|\eta|}\prod_{i=1}^{|\eta|}\frac{\partial^{|\eta_i|}}{\partial x_{\eta_i}}V_{s_1,\ldots,s_D}(x_1,\ldots,x_D), \quad (2.9)$$

where $\mathscr{B}_D$ denotes the set of all possible partitions of the set $\{\boldsymbol{s}_1,\ldots,\boldsymbol{s}_D\}$, $\eta = (\eta_1,\ldots,\eta_t)$, $|\eta| = t$ is the size of the partition η and $\frac{\partial^{|\eta_i|}}{\partial x_{\eta_i}}$ denotes the mixed partial derivatives with respect to the elements of the i-th element of the partition η. For example, with $D = 3$ there are five partitions. If $\eta = \{(\boldsymbol{s}_1,\boldsymbol{s}_2)(\boldsymbol{s}_3)\}$, we have

$$\frac{\partial^{|\eta_1|}}{\partial x_{\eta_1}}V_{s_1,\ldots,s_D}(x_1,\ldots,x_D) = \frac{\partial^2}{\partial x_1\partial x_2}V_{s_1,s_2,s_3}(x_1,x_2,x_3) =: V_{12},$$

$$\frac{\partial^{|\eta_2|}}{\partial x_{\eta_2}}V_{s_1,\ldots,s_D}(x_1,\ldots,x_D) = \frac{\partial}{\partial x_3}V_{s_1,s_2,s_3}(x_1,x_2,x_3) =: V_3,$$

and so forth for the other four partitions. The p.d.f. of the max-stable process X in this case is

$$f(x_1,x_2,x_3) = \exp\{-V\}\{V_1V_{23}+V_2V_{13}+V_3V_{12}-V_{123}-V_1V_2V_3\}.$$

Definition 2.8. ***(Extremal dependence function; [92])*** *Let X be a simple max-stable process on $\mathcal{S}$. The D-dimensional extremal coefficient function is defined for $D \in \mathbb{N}\setminus\{0\}$, $\boldsymbol{s}_1,\ldots,\boldsymbol{s}_D \in \mathcal{S}$, and $x > 0$, by*

$$\theta_D(\boldsymbol{s}_1,\ldots,\boldsymbol{s}_D) = -x\log\mathbb{P}(X(\boldsymbol{s}_1)\leq x,\ldots,X(\boldsymbol{s}_D)\leq x) \in [1,D]. \quad (2.10)$$

The function θ_D varies from $\theta_D = 1$ when the observations X_i, $i \in \{1,\ldots,D\}$ are completely dependent to $\theta_D = D$ when they are completely independent. In practice, this coefficient provides

a summary of the degree of dependence on the diagonal (i.e., $x_1 = \ldots = x_D = x$). In this sense, it does not give an exhaustive description of extremal dependence, especially for large D.

Throughout this thesis, we are particularly interested with the bivariate case (i.e., $D = 2$), where the corresponding pairwise extremal dependence function $\theta_2(\boldsymbol{s}_1, \boldsymbol{s}_2) : \mathbb{R}^+ \mapsto [1, 2]$ satisfies

$$\theta_2(\boldsymbol{s}_1, \boldsymbol{s}_2) = V_{\boldsymbol{s}_1,\boldsymbol{s}_2}(1, 1) = -x \log \mathbb{P}[\max\{X(\boldsymbol{s}_1), X(\boldsymbol{s}_2)\} \leq x],\ x > 0. \tag{2.11}$$

Several interesting properties for $\theta_2(.)$ have been proved in [92]. From now on, we will consider stationary processes such that $\theta_2(\boldsymbol{s}_1, \boldsymbol{s}_2)$ depends only on $\boldsymbol{h} = \boldsymbol{s}_1 - \boldsymbol{s}_2$. For the sake of notational simplicity, we will write $\theta(\boldsymbol{h})$ for $\theta_2(\boldsymbol{s}_1, \boldsymbol{s}_2)$ and $V_{\boldsymbol{h}}$ for $V_{\boldsymbol{s}_1,\boldsymbol{s}_2}$.

An important property for stationary max-stable processes defined on $\mathbb{Z}$, intrinsically related to the extremal dependence function $\theta(\boldsymbol{h})$, has been discussed in [68]. Their result is summarized in the next theorem.

Theorem 2.3. ***(Mixing property of stationary max-stable process); [68]*** *Let X be a stationary simple max-stable process on $\mathcal{S} \subset \mathbb{Z}$, then X is mixing if and only if* $\lim_{\|\boldsymbol{h}\| \to \infty} \theta(\boldsymbol{h}) \to 2$.

Stationary parametric max-stable models

Based on the spectral representation (recall Theorem 2.1), different choices for the process $U(s)$ lead to more or less flexible models for spatial maxima. We now provide some examples of well-known max-stable models, where different sequences of U_i are considered. These models will be frequently used in the subsequent chapters. In the following, let $\{\zeta_i\}_{i \geq 1}$ be a Poisson point process on $(0, \infty)$ with intensity $\zeta^{-2} d\zeta$.

(i) ***Smith model (Guassian extreme value model)***
Let $\phi_d(., \boldsymbol{\Sigma})$ denote the d-variate Gaussian p.d.f. with mean 0 and covariance matrix $\boldsymbol{\Sigma}$. For all $\boldsymbol{s} \in \mathbb{R}^d$, define $U_i(\boldsymbol{s}) = \phi_d(\boldsymbol{s} - \boldsymbol{s}_i; \boldsymbol{\Sigma})$, where $\{\boldsymbol{s}_i\}_{i \geq 1}$ be a Poisson point process on $\mathbb{R}^d$ with intensity $d\boldsymbol{s}$. The spatial Smith model [94] is defined as

$$X(\boldsymbol{s}) = \bigvee_{i=1}^{\infty} \{\zeta_i \phi_d(\boldsymbol{s} - \boldsymbol{s}_i; \boldsymbol{\Sigma})\}. \tag{2.12}$$

The $\boldsymbol{\Sigma}$ matrix controls the dependence range and the degree of anisotropy of the realized random field. In the general case, the storm events have an elliptical shape. If $\boldsymbol{\Sigma}$ is diagonal and all the diagonal elements are identical, the ellipsoids change to spheres (that is, the random process is isotropic). However, due to the deterministic storm shapes of this model, it is not very flexible for modeling environmental data.

(ii) ***Schlather model (extremal Gaussian model)***
Let $\{U_i\}_{i \geq 1}$ be independent replications of the stochastic process $U(\boldsymbol{s}) = \sqrt{2\pi} \max\{0, \varepsilon(\boldsymbol{s})\}$ for $\boldsymbol{s} \in \mathcal{S}$, where ε is a standard Gaussian process with correlation function $\rho(.)$. The spatial Schlather process [90] is defined as

$$X(\boldsymbol{s}) = \bigvee_{i=1}^{\infty} \left\{\sqrt{2\pi} \zeta_i \max\{0, \varepsilon_i(\boldsymbol{s})\}\right\}. \tag{2.13}$$

Accordingly, the random shape of the storms may provide more realistic realizations. As for Gaussian processes, the smoothness of the process constructed from (2.13) can be controlled by the choice of correlation function ρ, recall Figure 2.1. However, this model cannot take into account extremes that become independent when the distance becomes large. Indeed, $\lim_{\|\boldsymbol{h}\|\to\infty}\theta(\boldsymbol{h}) \leq 1.707$, where this upper boundary value of $\theta(\boldsymbol{h})$ corresponding to $\rho(\boldsymbol{h}) = 0$, see Table 2.1 and Figure 2.4. So, this model is not mixing, recall Theorem 2.3. To get around this issue, a truncated version of $U(\boldsymbol{s})$ can be used. Let $\{q_i\}_{i\geq 1}$ be a homogeneous Poisson point process of unit rate on $\mathcal{S}$. Then, the spatial ***truncated extremal Gaussian model (TEG)*** (originally due to [90]) is defined as

$$X(\boldsymbol{s}) = \bigvee_{i=1}^{\infty} \left\{\sqrt{2\pi}(\mathbb{E}\{|\mathcal{A}|\})^{-1}\zeta_i \max\{0, \varepsilon_i(\boldsymbol{s})\}\mathbb{1}_{\mathcal{A}_i}(\boldsymbol{s} - q_i)\right\}, \tag{2.14}$$

where $\mathbb{1}_{\mathcal{A}}$ denotes an indicator function of a compact random set $\mathcal{A} \subset \mathcal{S}$, $\mathcal{A}_i$ are i.i.d. copies of $\mathcal{A}$, $|.|$ is used to denote the volume of a set and $\varepsilon(\boldsymbol{s})$ is defined as above. According to this construction, the short-range dependence is largely determined by the correlation function $\rho(\boldsymbol{h})$, whereas the long-range dependence is regulated by the geometry of the set $\mathcal{A}$. Thus, in the bivariate case, we can choose $\mathcal{A}$ so that $\alpha(\boldsymbol{h}) \to 0$ and thus $\lim_{\|\boldsymbol{h}\|\to\infty}\theta(\boldsymbol{h}) \to 2$, where $\alpha(\boldsymbol{h})$ is the expected volume of overlap between the random set $\mathcal{A}$ and $\mathcal{A} + \boldsymbol{h}$. So, this model is mixing, see Table 2.1 and Figure 2.4. Accordingly, when modeling spatial/spatio-temporal phenomena, a suitable choice for $\mathcal{A}$ has to be considered, so that independence can be captured. The defect of this model is that realizations are not continuous due to the set $\mathcal{A}$. Nevertheless, it can be adapted to model processes with very local effects, or phenomena that reflect perfect independence after some fixed lag (e.g., spatio-temporal applications). The TEG model has been fitted to annual temperature maxima in Switzerland (respectively, daily rainfall in the East of Australia) in [44] (respectively, [11]). In addition, a spatio-temporal version of it has been fitted satisfactorily to hourly rainfall extremes in [64].

(iii) ***Brown-Resnick model (BR)***

Sometimes called the geometric Gaussian model [20, 69]. Let $U_i(\boldsymbol{s}) = \exp\{\varepsilon_i(\boldsymbol{s}) - \gamma(\boldsymbol{s})\}$, where $\varepsilon_i(\boldsymbol{s})$ are independent copies of an intrinsically stationary centered Gaussian process and with dependence function $\gamma(.)$ which is termed the semivariogram (recall Definition 2.3). The spatial BR model is defined as

$$X(\boldsymbol{s}) = \bigvee_{i=1}^{\infty} \{\zeta_i \exp\{\varepsilon_i(\boldsymbol{s}) - \gamma(\boldsymbol{s})\}\}, \ \boldsymbol{s} \in \mathcal{S}. \tag{2.15}$$

This construction may provide more realistic event realizations. At large distances the maxima are close to independence, see Figure 2.4. The function γ is often considered as the unbounded FBM model, i.e., $\gamma(\boldsymbol{h}) = \left(\frac{\|\boldsymbol{h}\|}{\phi}\right)^{\kappa}$, $\phi > 0$, $\kappa \in (0, 2]$. In this case, the process is mixing and ϕ is a range parameter, whereas κ is a smoothness parameter with higher values indicating smoother processes. Furthermore, small and large values of $\gamma(\boldsymbol{h})$ respectively correspond to strong and to weak dependence; as $\gamma(\boldsymbol{h}) \to 0$ and $\gamma(\boldsymbol{h}) \to \infty$, we see that

$V_{\boldsymbol{h}}(x_1, x_2) \to \max\{1/x_1, 1/x_2\}$, $V_{\boldsymbol{h}}(x_1, x_2) \to 1/x_1 + 1/x_2$, corresponding respectively to complete dependence of $X(\boldsymbol{s})$ and $X(\boldsymbol{s}+\boldsymbol{h})$ and to their independence, see Table 2.1.

(iv) ***Extremal-t model***

Another interesting max-stable model, the extremal-t model, is a generalization of Schlather model, see [82, 87]. Let $U_i(\boldsymbol{s}) = c_\nu \max\{0, \varepsilon_i(\boldsymbol{s})\}^\nu$, $c_\nu = \sqrt{\pi} 2^{-(\nu-2)/2} \Gamma\left(\frac{\nu+1}{2}\right)^{-1}$, $\nu \geq 1$, where $\varepsilon(\boldsymbol{s})$ is defined as in Schlather model, Γ is the Gamma function and ν is the degrees of freedom (df) parameter. The spatial extremal-t model is defined as

$$X(\boldsymbol{s}) = \bigvee_{i=1}^{\infty} \{\zeta_i c_\nu \max\{0, \varepsilon_i(\boldsymbol{s})\}^\nu\}, \ \boldsymbol{s} \in \mathcal{S}. \tag{2.16}$$

As for BR processes, the process realizations are realistic and close to independence for distances of going to ∞, see Figure 2.4. In the bivariate case, if $\rho(\boldsymbol{h}) \to 0$, we have $\lim_{\|\boldsymbol{h}\| \to \infty} \theta(\boldsymbol{h}) \to 2T_{\nu+1}\left(\sqrt{\nu+1}\right)$, where T_ν is the c.d.f. of a Student random variable with df ν. So, the process is not mixing (i.e., independence cannot be captured), unless $\nu \to \infty$, see Table 2.1. Accordingly, the df parameter ν controls the degree of long-range dependence.

In the following, we will focus on bivariate max-stable models. Table 2.1 gives a brief summary of the above-mentioned max-stable processes by their bivariate exponent function $V_{\boldsymbol{h}}(.)$ and extremal coefficient function $\theta(.)$. Figure 2.4 plots the extremal coefficient function for isotropic versions of the max-stable models introduced in Table 2.1. Clearly, the Smith model covers the whole range of dependence while Schlather's model has an upper bound of 1.707 if the covariance function only takes positive values. More generally, the properties of isotropic positive-definite functions imply that $\theta(.) \leq 1.838$ in $\mathbb{R}^2$ and $\theta(.) \leq 1.781$ in $\mathbb{R}^3$, see [75, 86]. In other words, the asymptotic dependence persists even at distances going to ∞. In the TEG model, extremes separated by a distance $\|\boldsymbol{h}\| \geq 2r$ are (exactly) independent. For extremal-t and BR processes, the pairs of sites separated by a large distance are close to independence, where the df parameter ν controls the upper bound of the extremal coefficient function for extremal-t processes, that is, $\theta(\boldsymbol{h}) \to 2T_{\nu+1}\left(\sqrt{\nu+1}\right)$, $\|\boldsymbol{h}\| \to \infty$. Note that general expressions of the c.d.f. for some popular max-stable models can be found in the literature. For instance, [58] provided a closed form expression of the c.d.f. for the Smith max-stable model indexed by $\mathbb{R}^d$ for $D \leq d+1$ sites and the generalization of this to the BR process can be found in [63].

Remark 2.1. *From Table 2.1, it is easy to verify the following special cases:*

- *Setting* $2\gamma(\boldsymbol{h}) = \sqrt{\boldsymbol{h}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{h}}$ *in BR model, for some covariance matrix* $\boldsymbol{\Sigma}$*, we recover Smith model. In particular, isotropic Smith process is recovered when the FBM semivariogram is* $\gamma(\boldsymbol{h}) = \left(\frac{\|\boldsymbol{h}\|}{\phi}\right)^2$, $\phi > 0$.
- *Setting* $\alpha(\boldsymbol{h}) = 1$ *in TEG model, we recover Schlather model.*
- *Setting the df* $\nu = 1$ *in extremal-t model, we recover Schlather model.*

Table 2.1: Bivariate marginal distributions for spatial max-stable models.

Model	$V_{\boldsymbol{h}}(x_1, x_2)$	$\theta(\boldsymbol{h})$
Smith	$\frac{1}{x_1}\Phi\left(\frac{b(\boldsymbol{h})}{2}+\frac{\log\left(\frac{x_2}{x_1}\right)}{b(\boldsymbol{h})}\right)+\frac{1}{x_2}\Phi\left(\frac{b(\boldsymbol{h})}{2}+\frac{\log\left(\frac{x_1}{x_2}\right)}{b(\boldsymbol{h})}\right)$ $b(\boldsymbol{h})=\sqrt{\boldsymbol{h}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{h}}$ and $\Phi(.)$ denotes the standard normal distribution	$2\Phi\left(\frac{b(\boldsymbol{h})}{2}\right)$
Schlather	$\frac{1}{2}\left(\frac{1}{x_1}+\frac{1}{x_2}\right)\left[1+\sqrt{1-\frac{2(\rho(\boldsymbol{h})+1)x_1x_2}{(x_1+x_2)^2}}\right]$ $\rho(\boldsymbol{h})$ is the correlation coefficient function	$1+\left(\frac{1-\rho(\boldsymbol{h})}{2}\right)^{1/2}$
TEG	$\left(\frac{1}{x_1}+\frac{1}{x_2}\right)\left[1-\frac{\alpha(\boldsymbol{h})}{2}\left(1-\sqrt{1-\frac{2(\rho(\boldsymbol{h})+1)x_1x_2}{(x_1+x_2)^2}}\right)\right]$ $\alpha(\boldsymbol{h})=\mathbb{E}[\lvert\mathcal{A}\cap\{\boldsymbol{h}+\mathcal{A}\}\rvert]/\mathbb{E}(\lvert\mathcal{A}\rvert)$, $\lvert.\rvert$: volume of the random set $\mathcal{A}$ If $\mathcal{A}$ is a disk of fixed radius r, $\alpha(\boldsymbol{h})\simeq(1-\lVert\boldsymbol{h}\rVert/2r)\mathbb{1}_{[0,2r]}$	$2-\alpha(\boldsymbol{h})\left[1-\left(\frac{1-\rho(\boldsymbol{h})}{2}\right)^{1/2}\right]$
BR	$\frac{1}{x_1}\Phi\left(\sqrt{\frac{\gamma(\boldsymbol{h})}{2}}+\frac{\log\left(\frac{x_2}{x_1}\right)}{\sqrt{2\gamma(\boldsymbol{h})}}\right)+\frac{1}{x_2}\Phi\left(\sqrt{\frac{\gamma(\boldsymbol{h})}{2}}+\frac{\log\left(\frac{x_1}{x_2}\right)}{\sqrt{2\gamma(\boldsymbol{h})}}\right)$ $\gamma(.)$ denotes the semivariogram	$2\Phi\left[\left(\frac{\gamma(\boldsymbol{h})}{2}\right)^{1/2}\right]$
Extremal-t	$\frac{1}{x_1}T_{\nu+1}\left(-b\rho(\boldsymbol{h})+b\left(\frac{x_2}{x_1}\right)^{1/\nu}\right)+\frac{1}{x_2}T_{\nu+1}\left(-b\rho(\boldsymbol{h})+b\left(\frac{x_1}{x_2}\right)^{1/\nu}\right)$ T_ν is the c.d.f. of a Student random variable with df $\nu\geq 1$. $b^2=\{\nu+1\}/\{1-\rho^2(\boldsymbol{h})\}$	$2T_{\nu+1}\left(\sqrt{(\nu+1)\frac{1-\rho(\boldsymbol{h})}{1+\rho(\boldsymbol{h})}}\right)$

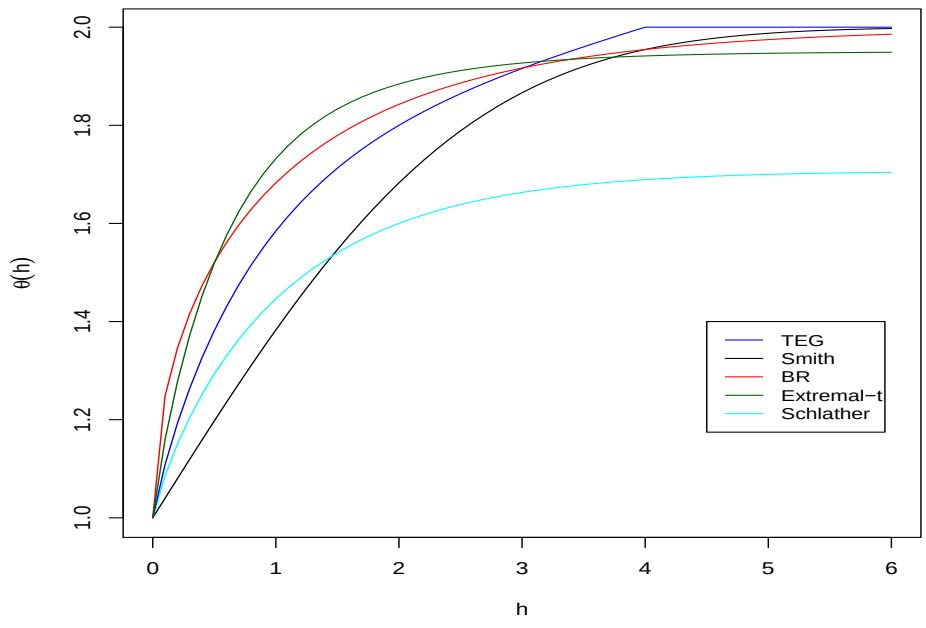


Figure 2.4: Extremal coefficient function $\theta(.)$ for isotropic max-stable families given in Table 2.1 plotted as a function of the distance $h=\lVert\boldsymbol{h}\rVert\geq 0$. Covariance matrix: $\boldsymbol{\Sigma}=\mathrm{Id}_2$, Id_2 is 2 by 2 identity matrix. Correlation function: Whittle-Matérn with $c_1=c_2=1$. Compact set: $\mathcal{A}$ is a disk of fixed radius 2; $\alpha(\lVert\boldsymbol{h}\rVert)\simeq(1-\lVert\boldsymbol{h}\rVert/4)\mathbb{1}_{[0,4]}$. Semivariogram: $\gamma(\lVert\boldsymbol{h}\rVert)=2\lVert\boldsymbol{h}\rVert$. Degrees of freedom: $\nu=5$.

Remark 2.2. *Another well-known representation for bivariate distribution of a max-stable process is given by the* ***Pickands dependence function*** $A(.) : [0,1] \to [1/2,1]$ *[85]. It turns out that we can rewrite the exponent measure as*

$$V_{\boldsymbol{h}}(x_1,x_2) = \left(x_1^{-1} + x_2^{-2}\right) A_{\boldsymbol{h}}\left(\frac{x_1}{x_1+x_2}\right). \tag{2.17}$$

Setting $\varphi = x_1/(x_1+x_2)$, $\varphi \in (0,1)$ in (2.17), we obtain

$$A_{\boldsymbol{h}}(\varphi) = \varphi(1-\varphi)V_{\boldsymbol{h}}(\varphi,1-\varphi), \tag{2.18}$$

where the Pickands dependence function $A_{\boldsymbol{h}}(.)$ is a convex function satisfying: $\max(t,1-t) \leq A_{\boldsymbol{h}}(t) \leq 1$, $t \in [0,1]$ and $A_{\boldsymbol{h}}(0) = A_{\boldsymbol{h}}(1) = 1$, where the lower (respectively, upper) bound corresponds to complete dependence (respectively, independence). Accordingly, the Pickands dependence function A is related to the extremal dependence function $\theta(.)$ through a simple link $\theta(\boldsymbol{h}) = V_{\boldsymbol{h}}(1,1) = 2A_{\boldsymbol{h}}(1/2)$. For a detailed overview on $A_{\boldsymbol{h}}(.)$, refer to [14]. Analogously, an extension of $A_{\boldsymbol{h}}(.)$ to spatio-temporal framework can be found in [40], see Section 2.3.1.

2.2.2 Summary measures for extremal dependence

When dealing with non-extremal data, a common tool is the semivariogram, recall Section 2.1. However, when extreme values are of interest, the semivariogram is no longer a useful tool, as it may even not exist. Therefore, there is a need to develop more suitable tools for analyzing the spatial dependence of extremes. Different measures of extremal dependence have been proposed in the literature. The extremal dependence function θ, described in Section 2.2.1, is suitable for asymptotically dependent processes (i.e., max-stable processes). It provides sufficient information about extremal dependence for many problems although it does not characterize the full distribution. In what follows, we shortly recall functions that may be used to describe the extremal dependence behavior for spatial processes. A concise review can be found in [12, 45].

Two model-free diagnostic measures: $\chi_u(\boldsymbol{h})$ and $\bar{\chi}_u(\boldsymbol{h})$ have been suggested in [33] for distinguishing among different types of tail dependence, see also Section 8.4 from [32].

Definition 2.9. ***(Upper tail dependence function; [33])*** *For a stationary spatial process* $\{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ *with univariate margin cumulative distribution function* F*. The upper tail dependence function* χ *in (2.4) may be rewritten as*

$$\chi(\boldsymbol{h}) = \lim_{u\to 1^-} \mathbb{P}\{F(X(\boldsymbol{s})) > u | F(X(\boldsymbol{s}+\boldsymbol{h})) > u\},\ \boldsymbol{s}, \boldsymbol{s}+\boldsymbol{h} \in \mathcal{S}. \tag{2.19}$$

Recall that If $\chi(\boldsymbol{h}) > 0$, $\forall \boldsymbol{h} \in \mathcal{S}$, the process X is said to be asymptotically dependent. On the other hand, a process is asymptotically independent if $\chi(\boldsymbol{h}) = 0$ for any $\boldsymbol{h} \in \mathcal{S}$. The best-known example is the Gaussian model which is asymptotically independent for all correlations $\rho(\boldsymbol{h}) \neq 1$, see [93]. Alternatively, for $u \in [0,1]$, $\chi(\boldsymbol{h})$ can be expressed as

$$\chi_u(\boldsymbol{h}) = 2 - \frac{\log \mathbb{P}\{F(X(\boldsymbol{s})) < u, F(X(\boldsymbol{s}+\boldsymbol{h})) < u\}}{\log \mathbb{P}\{F(X(\boldsymbol{s}+\boldsymbol{h})) < u\}} \quad \text{and} \quad \chi(\boldsymbol{h}) = \lim_{u\to 1^-} \chi_u(\boldsymbol{h}). \tag{2.20}$$

The function $\chi_u(\boldsymbol{h})$ can be viewed as a measure of the dependence at the u-quantile level. Note also for a max-stable process X the function $\chi_u(\boldsymbol{h})$ is constant with respect to u for a fixed $\boldsymbol{h}$, i.e., $\chi_u(\boldsymbol{h}) = 2-\theta(\boldsymbol{h})$, see Figure 2.7. The dependence measure $\chi(\boldsymbol{h})$ is related to the pairwise extremal function $\theta(\boldsymbol{h})$ of a max-stable process through a simple link: $\chi(\boldsymbol{h}) = 2-\theta(\boldsymbol{h})$. Thus, both $\theta(.)$ and $\chi(.)$ functions provide simple measures of extremal dependence within the class of asymptotically dependent processes, whereas, under asymptotic independence, they are uninformative and of limited interest.

In order to measure the strength of dependence for asymptotically independent processes, a complementary dependence measure known as lower tail dependence function $\bar{\chi}(.)$ has been introduced by [33].

Definition 2.10. ***(Lower tail dependence function; [33])*** *Let $\{X(\boldsymbol{s})\}_{s\in\mathcal{S}}$ be a stationary spatial process with univariate margin F. The lower tail dependence function $\bar{\chi}(.)$ is defined as*

$$\bar{\chi}_u(\boldsymbol{h}) = \frac{2\log\mathbb{P}(F(X(\boldsymbol{s})) > u)}{\log\mathbb{P}(F(X(\boldsymbol{s})) > u, F(X(\boldsymbol{s}+\boldsymbol{h})) > u)} - 1 \quad \textit{and} \quad \bar{\chi}(\boldsymbol{h}) = \lim_{u\to 1^-} \bar{\chi}_u(\boldsymbol{h}). \tag{2.21}$$

Accordingly, asymptotic dependence (respectively, asymptotic independence) is achieved if $\bar{\chi}(\boldsymbol{h}) = 1$ (respectively, $\bar{\chi}(\boldsymbol{h}) < 1$). Moreover, $\bar{\chi}_u(\boldsymbol{h}) \in (0,1)$ (respectively, $(-1,0)$) implies positive (respectively, negative) association at distance $\boldsymbol{h}$. Hence, $|\bar{\chi}_u(\boldsymbol{h})|$ increases with the dependence. For bivariate max-stable processes with extremal dependence function $\theta(\boldsymbol{h})$, it is easy to verify that

$$\bar{\chi}_u(\boldsymbol{h}) = \frac{2\log(1-u)}{\log\left(1-2u+u^{\theta(\boldsymbol{h})}\right)} - 1,\ u \in [0,1] \text{ and } \theta(\boldsymbol{h}) \in [1,2]. \tag{2.22}$$

So, in practice, the two indicators χ and $\bar{\chi}$ should be considered together. For a fixed lag $\boldsymbol{h}$, the pair of diagnostics $\{\chi_u(\boldsymbol{h}), \bar{\chi}_u(\boldsymbol{h})\}$ can be used as a tool to detect asymptotic independence when u approaches 1. Alternatively, for a fixed extreme level u, it can provide a "*correlogram*" for extreme events, when considered as a function of $\boldsymbol{h}$. Note that these dependence measures may be estimated by their empirical rank-based counterparts. Furthermore, [71] proposed a model that smoothly links the asymptotic dependence and independence based on the value of another summary measure known as the ***coefficient of tail dependence*** $\eta(\boldsymbol{h})$. This measure determines the rate of tail decay towards independence, where asymptotic dependence being a particular case. More precisely, let $\{X(\boldsymbol{s})\}_{s\in\mathcal{S}}$ be a stationary spatial process with unit Fréchet margin, their approach relies on the assumption that the bivariate joint tail (survivor function) and the conditional upper tail behave respectively as

$$\mathbb{P}(X(\boldsymbol{s}) > x, X(\boldsymbol{s}+\boldsymbol{h}) > x) \sim \mathcal{L}_{\boldsymbol{h}}(x)x^{-1/\eta(\boldsymbol{h})},\ x\to\infty, \tag{2.23}$$

$$\mathbb{P}(X(\boldsymbol{s}) > x | X(\boldsymbol{s}+\boldsymbol{h}) > x) \sim \mathcal{L}_{\boldsymbol{h}}(x)x^{1-1/\eta(\boldsymbol{h})},\ x\to\infty, \tag{2.24}$$

where $\mathcal{L}_{\boldsymbol{h}}(.)$ is a slowly varying function at ∞; that is, $\lim_{y\to\infty}\frac{\mathcal{L}_{\boldsymbol{h}}(yt)}{\mathcal{L}_{\boldsymbol{h}}(y)} = 1$, $t > 0$, and $\eta(\boldsymbol{h}) \in (0,1]$. Indeed, several dependence scenarios can be handled within this simple modeling framework (see [62, 71, 72] for more details):

- $\eta(\boldsymbol{h}) < 1$ and $\mathcal{L}_{\boldsymbol{h}}(x) > 0$, corresponds to asymptotic independence, where $\eta(\boldsymbol{h}) \in \left(\frac{1}{2}, 1\right)$ $\left(\text{respectively, } \left(0, \frac{1}{2}\right)\right)$ implies a positive (respectively, negative) association at distance $\boldsymbol{h}$.
- $\eta(\boldsymbol{h}) = \frac{1}{2}$ and $\mathcal{L}_{\boldsymbol{h}}(x) \geq 1$, corresponds to near independence.
- $\eta(\boldsymbol{h}) = 1$ and $\mathcal{L}_{\boldsymbol{h}}(x) > 0$, corresponds to asymptotic dependence. For instance, simple max-stable processes have $\eta(\boldsymbol{h}) = 1$ for all $\boldsymbol{h}$.

Moreover, under condition (2.23), one can easily deduce that $\bar{\chi}(\boldsymbol{h}) = 2\eta(\boldsymbol{h}) - 1$.

Example 2.1. ***(Gaussian copula model)*** *As an illustration example, let us consider the stationary Gaussian process $\{\varepsilon(\boldsymbol{s})\}_{s\in\mathcal{S}}$ with zero mean, unit variance and correlation function $\rho(\boldsymbol{h})$, then the spatial process $Q(\boldsymbol{s}) = -1/\log\{\Phi(\varepsilon(\boldsymbol{s}))\}$ is an asymptotically independent process with unit Fréchet margins and verifies,*

$$\mathbb{P}(Q(\boldsymbol{s}) > x, Q(\boldsymbol{s}+\boldsymbol{h}) > x) \sim c_\rho x^{-2/\{1+\rho(\boldsymbol{h})\}} (\log x)^{-\rho(\boldsymbol{h})/\{1+\rho(\boldsymbol{h})\}},$$

with $c_\rho = \{1+\rho(\boldsymbol{h})\}^{3/2}\{1-\rho(\boldsymbol{h})\}^{-1/2}(4\pi)^{-\rho(\boldsymbol{h})/\{1+\rho(\boldsymbol{h})\}}$, where Φ denotes the c.d.f. of a standard normal distribution. Obviously, according to (2.23), $\eta(\boldsymbol{h}) = (1+\rho(\boldsymbol{h}))/2$ and $\bar{\chi}(\boldsymbol{h}) = \rho(\boldsymbol{h})$, see [71]. Accordingly, the process $Q(\boldsymbol{s})$ is asymptotically independent when $\rho < 1$. In addition, if $\rho \to 1$ (complete dependence), we have $\mathbb{P}(Q(\boldsymbol{s}) > x, Q(\boldsymbol{s}+\boldsymbol{h}) > x) \sim x^{-1}$. Whereas, if $\rho \to 0$ (exact independence), we have $\mathbb{P}(Q(\boldsymbol{s}) > x, Q(\boldsymbol{s}+\boldsymbol{h}) > x) \sim x^{-2}$.

The natural extreme analog of the correlation function of a stationary process which is known as the ***extremogram*** was introduced by [42] for the use particularly with time series. It has been extended to a spatial setting in [31].

Definition 2.11. ***(Spatial extremogram)*** *The extremogram for a stationary spatial process $\{X(\boldsymbol{s})\}_{s\in\mathcal{S}}$ is defined by*

$$\rho_{\mathscr{A}_1,\mathscr{A}_2}(\boldsymbol{h}) = \lim_{x\to\infty} \frac{\mathbb{P}\left\{x^{-1}X(\boldsymbol{s}) \in \mathscr{A}_1, x^{-1}X(\boldsymbol{s}+\boldsymbol{h}) \in \mathscr{A}_2\right\}}{\mathbb{P}\left\{x^{-1}X(\boldsymbol{s}) \in \mathscr{A}_1\right\}}, \ \boldsymbol{h} \in \mathbb{R}^d, \tag{2.25}$$

where $\mathscr{A}_1$ and $\mathscr{A}_2$ are Borel sets bounded away from 0.

Setting $\mathscr{A}_1 = \mathscr{A}_2 = (1,\infty)$, we recover the tail dependence coefficient $\chi(\boldsymbol{h}) = \rho_{(1,\infty),(1,\infty)}$, recall (2.4). So, if $\rho_{(1,\infty),(1,\infty)} = 0$ then $X(\boldsymbol{s})$ and $X(\boldsymbol{s}+\boldsymbol{h})$ are asymptotically independent, and in case $0 < \rho_{(1,\infty),(1,\infty)} \leq 1$ they are asymptotically dependent. The spatio-temporal version of the extremogram is defined analogously, see Section 5.2.

Lastly, we present the madogram-based tools that are especially well-adapted to extremes. Various definitions of madograms are exist: ***F-madogram***, ***F^λ-madogram*** and ***λ-madogram***. Extensions of these summary measures to spatio-temporal framework are given in Section 2.3.1.

(i) [35] proposed the F-madogram as a summary statistics for the spatial extremal dependence structure.

Definition 2.12. ***(Spatial F-madogram; [35])*** *Let $\{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ be a stationary spatial process with marginal distribution function F. The F-madogram is defined by*

$$\nu_F(\boldsymbol{h}) = \frac{1}{2}\mathbb{E}[|F(X(\boldsymbol{s})) - F(X(\boldsymbol{s}+\boldsymbol{h}))|],\ 0 \le \nu_F(\boldsymbol{h}) \le 1/6. \tag{2.26}$$

Complete dependence (respectively, complete independence) is achieved when $\nu_F(\boldsymbol{h}) = 0$ (respectively, $\nu_F(\boldsymbol{h}) = 1/6$). For a max-stable process, the F-madogram is related to the extremal dependence function $\theta(.)$ through a simple link

$$\nu_F(\boldsymbol{h}) = \frac{1}{2} - \frac{1}{\theta(\boldsymbol{h})+1}. \tag{2.27}$$

The $\nu_F(\boldsymbol{h})$ may be estimated nonparametrically by its empirical version (see, e.g., [35, 5]),

$$\widehat{\nu}_F(\boldsymbol{h}) = \frac{1}{2T}\sum_{i=1}^{T}\left[|\widehat{F}(X_i(\boldsymbol{s})) - \widehat{F}(X_i(\boldsymbol{s}+\boldsymbol{h}))|\right], \tag{2.28}$$

where $\widehat{F}$ denotes the empirical distribution function, i.e., $\widehat{F}(x) = \frac{1}{T+1}\sum_{i=1}^{T}\mathbb{1}_{\{X_i(\boldsymbol{s})<x\}}$, where $\mathbb{1}_{\{\mathcal{B}\}}$ corresponds to the indicator function of $\mathcal{B}$, $X_i(\boldsymbol{s})$ the i-th observations of the spatial process X at location $\boldsymbol{s}$ and T is the total number of observations. In some applications, it might be preferable to use a "*binned*" version of estimator (2.28), that is,

$$\widehat{\nu}_F(\boldsymbol{h}) = \frac{1}{2T|\mathcal{B}_{\boldsymbol{h}}|}\sum_{(\boldsymbol{s},\boldsymbol{s}+\boldsymbol{h})\in\mathcal{B}_{\boldsymbol{h}}}\sum_{i=1}^{T}\left[|\widehat{F}(X_i(\boldsymbol{s})) - \widehat{F}(X_i(\boldsymbol{s}+\boldsymbol{h}))|\right], \tag{2.29}$$

where $\mathcal{B}_{\boldsymbol{h}}$ is a lag class of pairs $(\boldsymbol{s}, \boldsymbol{s}+\boldsymbol{h})$ at distance $\boldsymbol{h} \in \mathbb{R}^d$ within a certain tolerance Δ.

(ii) The F^λ-madogram has been introduced in [15] as a generalization of F-madogram.

Definition 2.13. ***(Spatial F^λ-madogram; [15])*** *For a stationary spatial process $\{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ with marginal distribution function F. For any $\lambda > 0$, the F^λ-madogram is defined as*

$$\nu_{F^\lambda}(\boldsymbol{h}) = \frac{1}{2}\mathbb{E}\left[|F^\lambda\{X(\boldsymbol{s})\} - F^\lambda\{X(\boldsymbol{s}+\boldsymbol{h})\}|\right]. \tag{2.30}$$

(iii) The extremal function θ and thus the madogram functions ν_F and ν_{F^λ} do not fully characterize the spatial dependence of a random field. Indeed, they only consider $\mathbb{P}(X(\boldsymbol{s}) \le x_1, X(\boldsymbol{s}+\boldsymbol{h}) \le x_2)$ where $x_1 = x_2 = x$. To bypass this issue, [81] introduced the λ-madogram.

Definition 2.14. ***(Spatial λ-madogram; [81])*** *Let $\{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ be a stationary spatial process with marginal distribution function F. The λ-madogram is defined as*

$$\nu_\lambda(\boldsymbol{h}) = \frac{1}{2}\mathbb{E}\left[|F^\lambda(X(\boldsymbol{s})) - F^{1-\lambda}(X(\boldsymbol{s}+\boldsymbol{h}))|\right],\ \lambda \in [0,1]. \tag{2.31}$$

The idea beyond this is to explore the whole space. By varying λ, we focus on $\mathbb{P}(X(\boldsymbol{s}) \le x_1, X(\boldsymbol{s}+\boldsymbol{h}) \le x_2)$ with $x_1 = \lambda x$ and $x_2 = (1-\lambda)x$. Hence, this measure can provide information on $V_{\boldsymbol{h}}(x_1, x_2)$ for $x_1 \neq x_2$.

Moreover, in [60], the madogram has been adapted for the asymptotic independence framework instead of asymptotic dependence only, whereas, [74] extended the concept of the λ-madogram from a bivariate set up to a multivariate one.

2.2.3 Models for asymptotic independence: inverted max-stable models

In what follows, for any stationary process X, we denote by $F^X_{\boldsymbol{h}}(x_1, x_2)$ the bivariate c.d.f. of the pair $(X(\boldsymbol{s}), X(\boldsymbol{s}+\boldsymbol{h}))$.

Processes derived from Gaussian fields by a nondecreasing transformation of the marginals are examples of asymptotically independent processes, recall Example 2.1. Alternatively, the class of inverted max-stable processes has been proposed by [101]. It was found to be more flexible than Gaussian derivated processes in some applications, see [96, 45]. Indeed, the inverted max-stable process is obtained by simply inverting a max-stable process as follows.

Definition 2.15. ***(Spatial inverted max-stable process, [101])*** *Let $\{X'(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ be a simple max-stable process with extremal dependence coefficient function $\theta'(\boldsymbol{h})$. Define $t(x) : (0,\infty) \mapsto (0,\infty)$ by $t(x) = -1/\log[1-\exp\{-1/x\}]$, then the process $Y(\boldsymbol{s}) = t(X'(\boldsymbol{s}))$ is an inverted max-stable process. It has unit Fréchet margins and satisfies (2.23) with tail dependence coefficient $\eta(\boldsymbol{h}) = 1/\theta'(\boldsymbol{h})$.*

With this construction, each max-stable process may be transformed into an asymptotically independent counterpart. The stochastic process $Y(\boldsymbol{s})$ has bivariate c.d.f. of the form

$$F^Y_{\boldsymbol{h}}(x_1, x_2) = -1 + \exp(-x_1^{-1}) + \exp(-x_2^{-1}) + \exp\left\{-V^{X'}_{\boldsymbol{h}}\{t(x_1), t(x_2)\}\right\}, \tag{2.32}$$

where $V^{X'}_{\boldsymbol{h}}$ is the exponent function of the bivariate c.d.f. corresponding to $(X'(\boldsymbol{s}), X'(\boldsymbol{s}+\boldsymbol{h}))$. From now on, with a slight abuse of notations, we will write $V^Y_{\boldsymbol{h}}$ for $V^{X'}_{\boldsymbol{h}}$ and $\theta_Y(\boldsymbol{h})$ for $\theta'(\boldsymbol{h})$. Figure 2.5 displays two realizations of isotropic asymptotically independent processes over the $[0,10]^2$ square. The corresponding η is also represented as a function of $h = \|\boldsymbol{h}\| \geq 0$. According to [72], the case $\eta(h) = 1/2$ corresponds to near independence (recall Section 2.2.2). So, the asymptotically independent process constructed from an isotropic Smith process allows asymptotic independence but tends to near independence for long distances, whereas the asymptotically independent process constructed from extremal-t process presents a stronger dependence in the asymptotic independence when h is sufficiently large.

The conditional exceedance probabilities in (2.24), $P(u) =: \mathbb{P}(F(X(\boldsymbol{s}_1)) > u | F(X(\boldsymbol{s}_2)) > u)$, $u \in [0,1]$, for a max-stable (respectively, an inverted max-stable) process $\left(P(u) = \frac{1-2u+u^{\theta_X(\boldsymbol{h})}}{1-u}\right)$ $\left(\text{respectively, } P(u) = (1-u)^{\theta_Y(\boldsymbol{h})-1}\right)$ and for various levels of dependence are displayed in Figure 2.6. Obviously, the conditional probabilities for max-stable processes are convex in u and converge to a positive value, whereas they are concave in u and tend to zero for inverted max-stable processes. Furthermore, the probabilities for inverted max-stable processes tend very slowly to zero.

Figure 2.7 depicts the summary dependence measures $\chi_u(\boldsymbol{h})$ and $\bar{\chi}_u(\boldsymbol{h})$ in (2.20) and (2.21) as functions of the extremal level u for max-stable and inverted max-stable processes. Clearly, for max-stable processes $\chi_u(\boldsymbol{h})$ takes the constant value $2-\theta_X(\boldsymbol{h})$ and $\bar{\chi}_u(\boldsymbol{h})$ increases to one. While $\chi_u(\boldsymbol{h})$ decreases to zero and $\bar{\chi}_u(\boldsymbol{h})$ takes the constant value $2/\theta_Y(\boldsymbol{h}) - 1$ in the case of inverted max-stable processes. For the latter, it is easy to verify that

$$\chi_u(\boldsymbol{h}) = 2 - \frac{\log\left(-1+2u+(1-u)^{\theta_Y(\boldsymbol{h})}\right)}{\log(u)}, \; u \in [0,1] \text{ and } \theta_Y(\boldsymbol{h}) \in [1,2]. \tag{2.33}$$

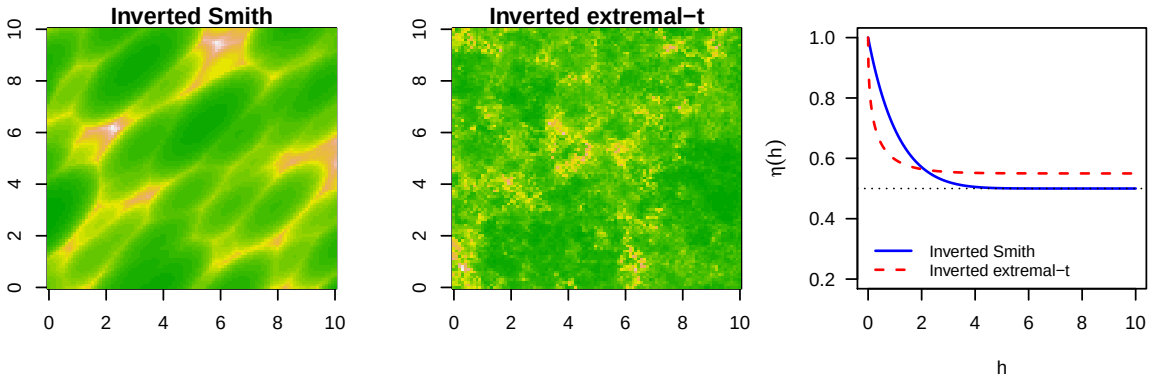


Figure 2.5: Simulations of two inverted max-stable processes on the logarithm scale. Left panel: simulated image of an asymptotically independent process constructed by inverting an isotropic Smith process with $\mathbf{\Sigma} = 3/4\ \mathrm{Id}_2$, Id_2 is 2 by 2 identity matrix. Middle panel: simulated image of an asymptotically independent process constructed by inverting an isotropic extremal-t process with df $\nu = 2$ and exponential correlation function $\rho(\boldsymbol{h}) = \exp(-h/\phi)$ with range $\phi = 1$. On the right panel: the associated function η plotted against $h = \|\boldsymbol{h}\|$.

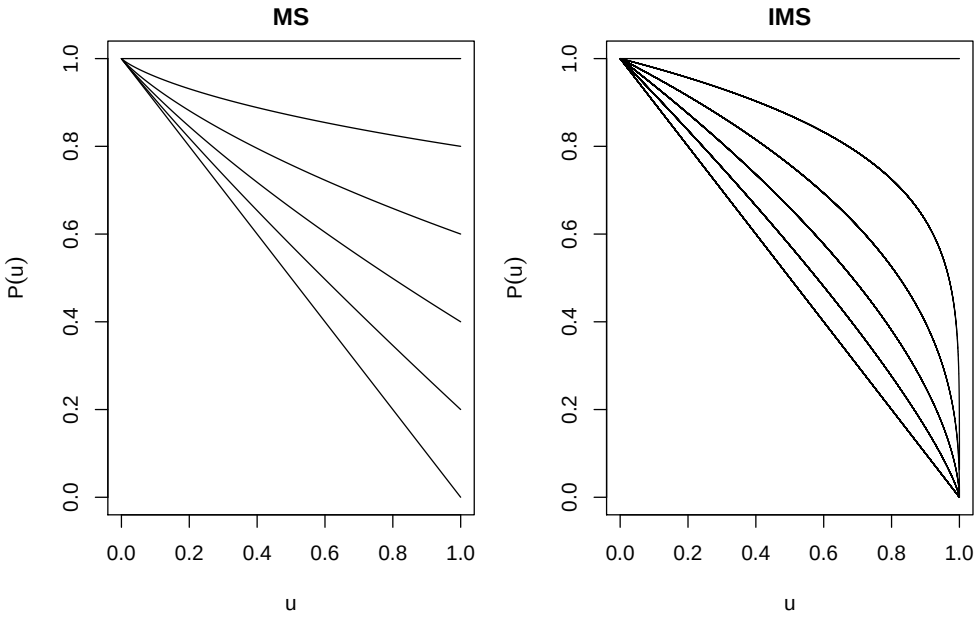


Figure 2.6: Conditional exceedance probabilities for random variables $(X(\boldsymbol{s}_1), X(\boldsymbol{s}_2))$ with uniform margins on $[0, 1]$. Left panel: a max-stable process with $\theta_X(\boldsymbol{h}) = 1, 1.2, 1.4, 1.6, 1.8, 2$ (from top to bottom). Right panel: an inverted max-stable process with $\theta_Y(\boldsymbol{h}) = 2, 1.8, 1.6, 1.4, 1.2, 1$ (also from bottom to top).

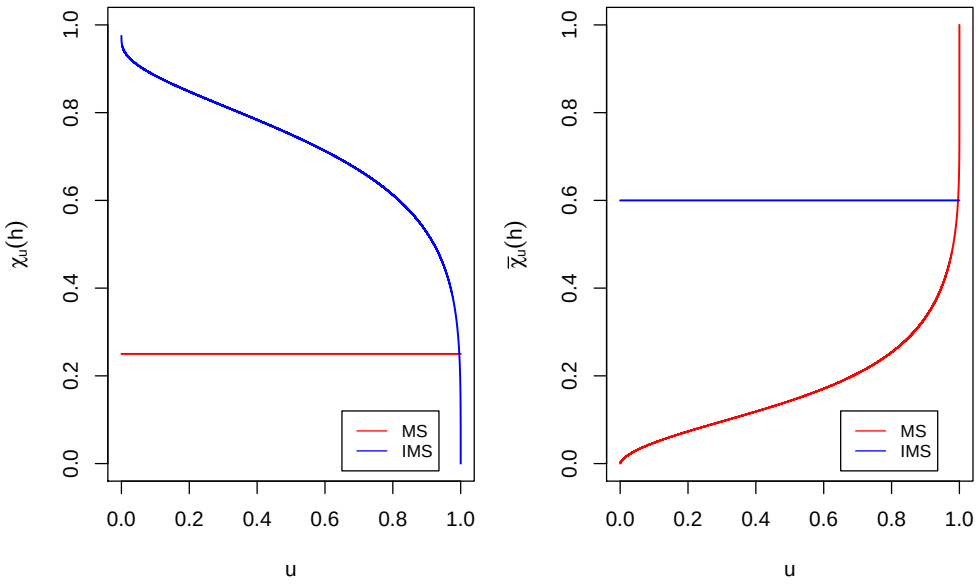


Figure 2.7: $\chi_u(\boldsymbol{h})$ and $\bar{\chi}_u(\boldsymbol{h})$ for a max-stable process with $\theta_X(\boldsymbol{h}) = 1.75$ (red) and an inverted max-stable process with $\theta_Y(\boldsymbol{h}) = 1.25$ (blue).

2.2.4 Max-mixture models of spatial extremal dependence

Max-mixture processes [101] are an interesting alternative to max-stable and asymptotically independent (e.g., Gaussian and inverted max-stable) processes, especially when the regions monitored are wide. These processes are obtained by mixing max-stable and asymptotically independent processes. In what follows, we will consider inverted max-stable processes for the asymptotically independent component.

Definition 2.16. ***(Spatial max-mixture process, [101])*** *Let $\{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ be a simple max-stable process with extremal coefficient $\theta_X(\boldsymbol{h})$, and $\{Y(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ be an inverted max-stable process with coefficient of tail dependence $\eta(\boldsymbol{h}) = 1/\theta_Y(\boldsymbol{h})$. Assume X and Y are independent. Then the process*

$$Z(\boldsymbol{s}) = \max\{aX(\boldsymbol{s}), (1-a)Y(\boldsymbol{s})\},\ 0 \le a \le 1, \tag{2.34}$$

is a max-mixture process. It has unit Fréchet margins.

Clearly, models that are only asymptotically dependent (respectively, independent) are sub-models of Z, obtained for $a = 1$ (respectively, $a = 0$). Figure 2.8 displays four realizations of two isotropic max-mixture processes over the $[0, 10]^2$ square according to different values of the mixing parameter a. In order to show the role of the mixing parameter, the plots are obtained by considering values between $a = 1$ (max-stable process) and $a = 0$ (inverted max-stable process). The mixing parameter a represents the proportion of asymptotic dependence in the max-mixture process Z. The simulations have been carried out using the function *rmaxstab* of the R package SpatialExtremes [86].

The bivariate c.d.f. and the bivariate conditional upper tail distribution for a pair of sites $(Z(\boldsymbol{s}), Z(\boldsymbol{s}+\boldsymbol{h}))$ are straightforwardly obtained for $0 < a < 1$ by the independence between $X(\boldsymbol{s})$

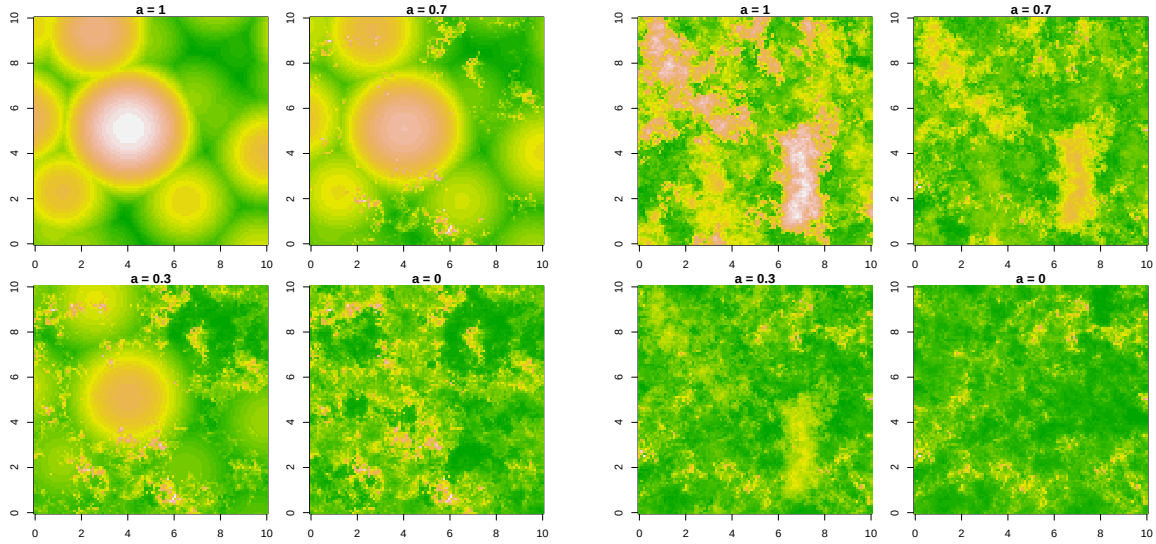


Figure 2.8: Simulations of the max-mixture model (2.34) on the logarithm scale according different values of mixing coefficient $a \in \{1, 0.7, 0.3, 0\}$. Left panel: X is an isotropic Smith process with $\boldsymbol{\Sigma} = \mathrm{Id}_2$, and Y is an isotropic inverted extremal-t process with df $\nu = 2$ and exponential correlation function $\rho(\boldsymbol{h}) = \exp(-\|\boldsymbol{h}\|/\phi)$ with range $\phi = 1$. Right panel: X is an isotropic extremal-t process with df $\nu_1 = 2$ and exponential correlation function $\rho_X(\boldsymbol{h}) = \exp(-\|\boldsymbol{h}\|/\phi_X)$ with range $\phi_X = 1$, and Y is an isotropic inverted extremal-t process with df $\nu_2 = 5$ and $\rho_Y(\boldsymbol{h}) = \exp(-\|\boldsymbol{h}\|/\phi_Y)$ with range $\phi_Y = 1$.

and $Y(\boldsymbol{s})$ as (see, [101, 11])

$$\begin{aligned} F_{\boldsymbol{h}}^{Z}(z_1, z_2) =& F_{\boldsymbol{h}}^{X}\left(\frac{z_1}{a}, \frac{z_2}{a}\right) F_{\boldsymbol{h}}^{Y}\left(\frac{z_1}{1-a}, \frac{z_2}{1-a}\right), \\ =& e^{-aV_{\boldsymbol{h}}^{X}(z_1,z_2)} \left\{ e^{\frac{-(1-a)}{z_1}} + e^{\frac{-(1-a)}{z_2}} - 1 + e^{-V_{\boldsymbol{h}}^{Y}(t_a(z_1), t_a(z_2))} \right\}, \end{aligned} \tag{2.35}$$

where $t_a(z) = t\left(\frac{z}{1-a}\right)$.

$$\mathbb{P}(Z(\boldsymbol{s}) > z | Z(\boldsymbol{s}+\boldsymbol{h}) > z) \sim a\{2 - \theta_X(\boldsymbol{h})\} + (1-a)^{\theta_Y(\boldsymbol{h})} \frac{\mathcal{L}_{\boldsymbol{h}}\{z/(1-a)\}}{z^{\theta_Y(\boldsymbol{h})-1}}, \; z \to \infty. \tag{2.36}$$

According to (2.36), we can deduce the following interesting remarks:

- max-stable models ($a = 1$), may be too restrictive since they have only the first term, recall (2.4) and (2.19).
- inverted max-stable models ($a = 0$), may be unreliable since they are left with the second term only, recall (2.24).
- If $a = 0$ or $\theta_X(\boldsymbol{h}) = 2$, then the first term on the right-hand side (which corresponds to the max-stable part of Z) vanishes. Hence, in the case where $a \neq 0$ and if there exists finite $\boldsymbol{h}' = \inf\{\boldsymbol{h} : \theta_X(\boldsymbol{h}) = 2\}$, then the process Z is asymptotically dependent up to distance $\boldsymbol{h}'$ and asymptotically independent for longer distances. This is in contrast with the

exact independence that would arise at distances $\boldsymbol{h} > \boldsymbol{h}'$ under the max-stable process X. For currently known max-stable processes, $\boldsymbol{h}'$ is not attained at any finite level except the TEG process, where $\mathcal{A}$ is often assumed to be a disk of fixed radius r, so, $\alpha(\boldsymbol{h})$ can be approximated by $\alpha(\boldsymbol{h}) \simeq (1 - \|\boldsymbol{h}\|/2r)\mathbb{1}_{[0,2r]}$, refer to [44]. If the approximation holds, it turns out that pairs of sites separated by a distance $\|\boldsymbol{h}\| < 2r$ are asymptotically dependent and asymptotically independent otherwise. More precisely, these findings can be summarized into the formula (see [11])

$$\bar{\chi}(\boldsymbol{h}) = \mathbb{1}_{[0,2r]}(\|\boldsymbol{h}\|) + (2/\theta_Y(\boldsymbol{h}) - 1)\mathbb{1}_{[2r,\infty)}(\|\boldsymbol{h}\|). \tag{2.37}$$

In [11], various instances of max-mixture processes have been considered by combining the TEG process with asymptotically independent processes. So, asymptotic dependence (respectively, independence) is present at short (respectively, intermediate) lags and possibly exact independence at larger ones, see Example 2.2. Analogously to max-stable models, max-mixture models have been fitted using composite likelihood and compared using the CLIC criterion. Details are deferred to Chapter 3.

- Recalling (2.4), it is easy to show that for the max-mixture process Z

$$\chi(\boldsymbol{h}) = \lim_{z\to\infty} \mathbb{P}(Z(\boldsymbol{s}) > z | Z(\boldsymbol{s}+\boldsymbol{h}) > z) = a\{2 - \theta_X(\boldsymbol{h})\} \tag{2.38}$$

Finally, the summary dependence measures $\chi_u(\boldsymbol{h})$ and $\bar{\chi}_u(\boldsymbol{h})$ in (2.20) and (2.21) can be easily computed for bivariate max-mixture processes. Namely, for $u \in [0,1]$ and $\theta_X(\boldsymbol{h}), \theta_Y(\boldsymbol{h}) \in [1,2]$, we have

$$\chi_u(\boldsymbol{h}) = 2 - a\theta_X(\boldsymbol{h}) - \frac{\log\left(-1 + 2u^{(1-a)} + \left(1 - u^{(1-a)}\right)^{\theta_Y(\boldsymbol{h})}\right)}{\log(u)}, \text{ and} \tag{2.39}$$

$$\bar{\chi}_u(\boldsymbol{h}) = \frac{2\log(1-u)}{\log\left\{1 - 2u + \left[\left(u^{a\theta_X(\boldsymbol{h})}\right)\left(-1 + 2u^{(1-a)} + (1-u^{(1-a)})^{\theta_Y(\boldsymbol{h})}\right)\right]\right\}} - 1. \tag{2.40}$$

Figure 2.9 displays the summary dependence measures $\chi_u(\boldsymbol{h})$ and $\bar{\chi}_u(\boldsymbol{h})$ in (2.39) and (2.40) for max-mixture processes. Obviously, for $a \in (0,1)$, $\chi_u(\boldsymbol{h})$ decreases from $a(2 - \theta_X(\boldsymbol{h}))$ to zero, whereas $\bar{\chi}_u(\boldsymbol{h})$ increases to unity. For $a \in \{0,1\}$, recall Figure 2.7.

Example 2.2. ***(Detecting asymptotic dependence classes for max-mixture models using the pair of diagnostics:*** $\{\chi_u(\boldsymbol{h}), \bar{\chi}_u(\boldsymbol{h})\}$***)*** *As we mentioned in Section 2.2.2, the two indicators* $\{\chi_u(\boldsymbol{h}), \bar{\chi}_u(\boldsymbol{h})\}$ *can be used as a tool to detect asymptotic dependence classes when u approaches unity. As an illustration, let us consider the following isotropic max-mixture model:*

M_1*: is a max-mixture model (2.34) in which X is a TEG process (see Table 2.1) with* $\mathcal{A}_X$ *a disk of fixed radius* $r_X = 0.25$ *and a stationary isotropic correlation function* $\rho_X(\boldsymbol{h}) = \exp\{-(\|\boldsymbol{h}\|/\phi_X)^{\kappa_X}\}$ *with range parameter* $\phi_X = 0.1$ *and smoothness* $\kappa_X = 1$*. The asymptotically independent process Y is an inverted TEG process with* $\mathcal{A}_Y$ *a disk of fixed radius* $r_Y = 0.75$ *and a stationary isotropic correlation function* $\rho_Y(\boldsymbol{h}) = \exp\{-(\|\boldsymbol{h}\|/\phi_Y)^{\kappa_Y}\}$ *with range parameter* $\phi_Y = 1.2$ *and smoothness* $\kappa_Y = 1$*. Recall that* $\|\boldsymbol{h}\| = \|\boldsymbol{s}_1 - \boldsymbol{s}_2\| \geq 0$ *denotes the Euclidean distance between sites* $\boldsymbol{s}_1, \boldsymbol{s}_2$*.*

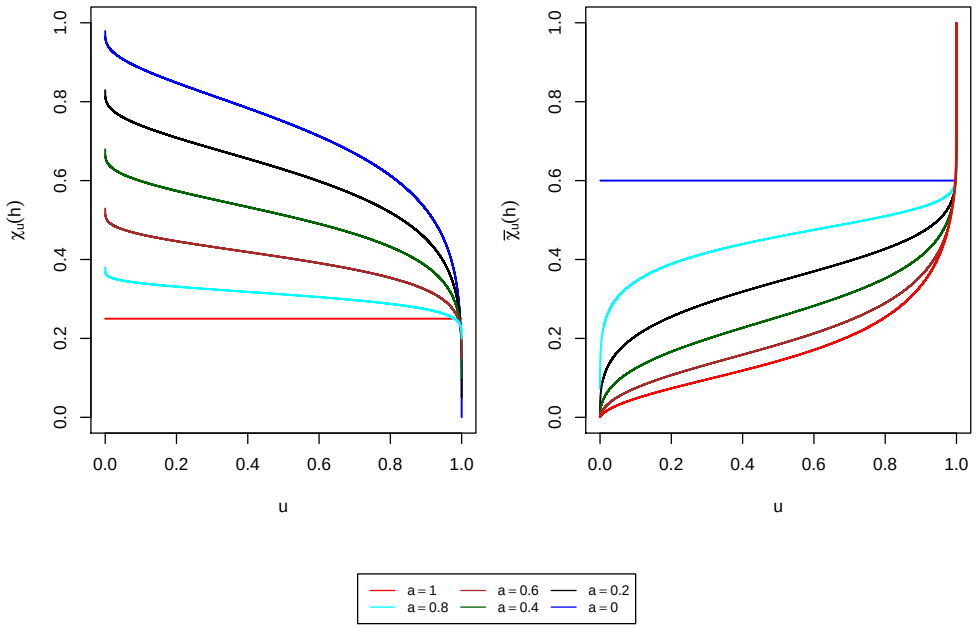


Figure 2.9: $\chi_u(\boldsymbol{h})$ and $\bar{\chi}_u(\boldsymbol{h})$ for a max-mixture process with extremal dependence summary measures $(\theta_X(\boldsymbol{h}), \theta_Y(\boldsymbol{h})) = (1.75, 1.25)$ and various mixing parameters $a \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$.

Fig 2.10 displays the pair of diagnostics $\{\chi_{0.97}(\boldsymbol{h}), \bar{\chi}_{0.97}(\boldsymbol{h})\}$ *in (2.39) and (2.40) for model* M_1. *Clearly, with this particular setting for the model parameters, the model is asymptotically dependent up to distance* $\|\boldsymbol{h}\| = 0.5$ $(\chi_{0.97}(\boldsymbol{h}) > 0$, *for all* $\|\boldsymbol{h}\| \leq 0.5)$ *and asymptotically independent at* $\|\boldsymbol{h}\| \in (0.5, 2.4)$ $(\bar{\chi}_{0.97}(\boldsymbol{h}) \in (0, 1)$, *for all* $\|\boldsymbol{h}\| \in (0.5, 2.4))$. *Furthermore, for* $\|\boldsymbol{h}\| \geq 2.4$, *the model is exactly independent* $(\bar{\chi}_{0.97}(\boldsymbol{h}) = 0$, *for all* $\|\boldsymbol{h}\| \geq 2.4)$.

2.3 Space-time max-stable models

Max-stable processes have been expanded to quantify extremal dependence in spatio-temporal data. Different spectral representations of stationary space-time max-stable processes have been developed, see, e.g., [40, 54]. In what follows, $\{X(\boldsymbol{s}, t) : (\boldsymbol{s}, t) \in \mathcal{S} \times \mathcal{T}\}$, $\mathcal{S} \times \mathcal{T} \subset \mathbb{R}^d \times \mathbb{R}^+$ (generally, $d = 2$) is space-time max-stable process. The space index $\boldsymbol{s}$ and time index t will respectively belong to the sets $\mathcal{S}$ and $\mathcal{T}$. In addition, we will denote by $\boldsymbol{h} = \boldsymbol{s}_1 - \boldsymbol{s}_2 \in \mathbb{R}^2$ (respectively, $l = t_1 - t_2 \in \mathbb{R}$) the spatial (respectively, temporal) lag.

2.3.1 Space-time max-stable models without spectral separability

The next two theorems extend the spectral representation and the D-finite dimensional distribution of spatial max-stable process to spatio-temporal setting, recall Theorem 2.1 and Theorem 2.2.

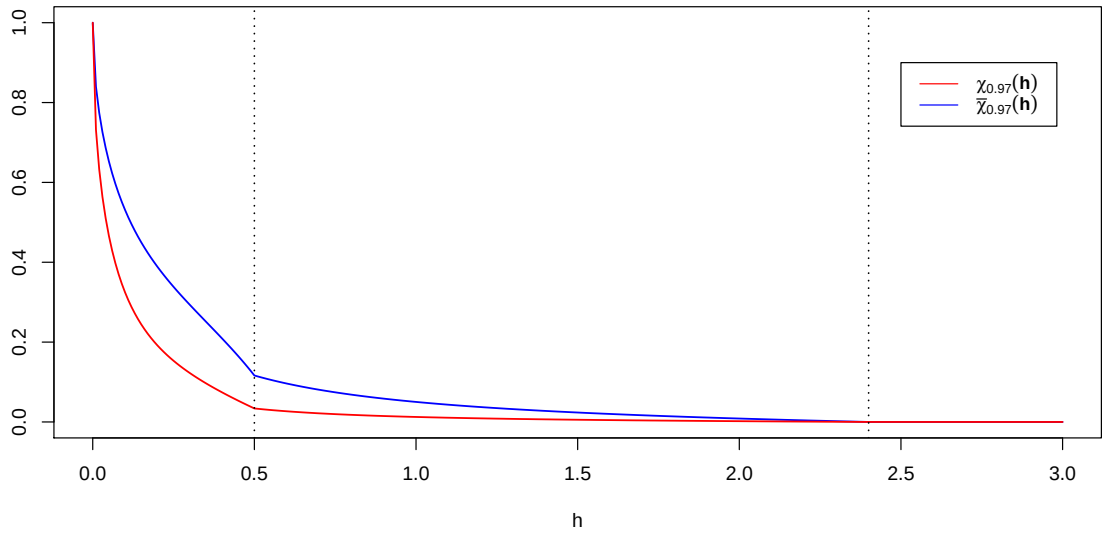


Figure 2.10: Theoretical behavior of $\chi_{0.97}(\boldsymbol{h})$ and $\bar{\chi}_{0.97}(\boldsymbol{h})$ as functions of the distance $h = \|\boldsymbol{h}\|$ derived from model M_1. We set $\phi_X = 0.1$, $r_X = 0.25$, $\phi_Y = 0.75$, $r_Y = 1.2$, $\kappa_X = \kappa_Y = 1$ and $a = 0.5$. The dotted vertical lines represent the diameters of disks in the TEG processes.

Theorem 2.4. ***(Spectral representation for space-time max-stable processes, [49])*** *Let $X(\boldsymbol{s},t)$ be a simple space-time max-stable process on $\mathcal{S}\times\mathcal{T}$. Then there exists $\{\zeta_i\}_{i\geq 1}$ i.i.d. points of a Poisson process on $(0,\infty)$ with intensity $\zeta^{-2}d\zeta$ and a sequence $\{U_i(\boldsymbol{s},t)\}_{i\geq 1}$ of independent replications of some space-time process $\{U(\boldsymbol{s},t),(\boldsymbol{s},t)\in\mathcal{S}\times\mathcal{T}\}$ with $\mathbb{E}\{U(\boldsymbol{s},t)\}<\infty$ for each $(\boldsymbol{s},t)\in\mathcal{S}\times\mathcal{T}$, and $U(\boldsymbol{s},t)\geq 0$, which are also independent of ζ_i, such that*

$$X(\boldsymbol{s},t) \overset{\mathcal{D}}{=} \bigvee_{i=1}^{\infty} \zeta_i U_i(\boldsymbol{s},t),\ (\boldsymbol{s},t)\in\mathcal{S}\times\mathcal{T}. \tag{2.41}$$

Remember that simple means that the margins are unit Fréchet, that is, $F(x) := \mathbb{P}(X(\boldsymbol{s},t)\leq x) = \exp\{-x^{-1}\}$, $x>0$. Analogously to spatial max-stable processes, this representation is in particular useful for the simulation of space-time max-stable processes as well as for providing examples of such processes.

Theorem 2.5. ***(Multivariate maxima)*** *For $D\in\mathbb{N}\setminus\{0\}$, $\boldsymbol{s}_1,\ldots,\boldsymbol{s}_D\in\mathcal{S}$, $t_1,\ldots,t_D\in\mathcal{T}$ and $x_1,\ldots,x_D>0$, the finite D-dimensional distributions of the space-time max-stable process X are given by*

$$\begin{aligned}\mathbb{P}(X(\boldsymbol{s}_1,t_1)\leq x_1,\ldots,X(\boldsymbol{s}_D,t_D)\leq x_D) &= \mathbb{P}\left\{\zeta_i \bigvee_{j=1}^{D}\frac{U_i(\boldsymbol{s}_j,t_j)}{x_j}\leq 1,\forall i=1,2,\ldots\right\} \\ &= \exp\left\{-\mathbb{E}\left(\bigvee_{j=1}^{D}\frac{U(\boldsymbol{s}_j,t_j)}{x_j}\right)\right\}.\end{aligned} \tag{2.42}$$

Hence, all finite-dimensional distributions are multivariate extreme value distributions with unit Fréchet margins. In particular, for $x_1, x_2 > 0$, the bivariate c.d.f. F_{s_1,t_1,s_2,t_2} of the space-time max-stable process $X(\boldsymbol{s},t)$ in (2.41) can be expressed in terms of the underlying bivariate spatio-temporal exponent function V_{s_1,t_1,s_2,t_2} as

$$-\log F_{s_1,t_1;s_2,t_2}(x_1,x_2) = -\log \mathbb{P}[X(t_1,\boldsymbol{s}_1) \le x_1, X(t_2,\boldsymbol{s}_2) \le x_2] =: V_{s_1,t_1;s_2,t_2}(x_1,x_2). \quad (2.43)$$

Below, we will consider stationary space-time processes such that $V_{s_1,t_1;s_2,t_2}$ depends on $\boldsymbol{h} = \boldsymbol{s}_1 - \boldsymbol{s}_2$ and $l = t_1 - t_2$, so that we can write $F_{\boldsymbol{h},l}$ for $F_{s_1,t_1;s_2,t_2}$ and $V_{\boldsymbol{h},l}$ for $V_{s_1,t_1;s_2,t_2}$.

Spatio-temporal extremal dependence summary measures

In order to measure the spatio-temporal extremal dependence, we provide in the next Definition, extensions to the spatio-temporal setting of quantities that have been introduced in the spatial context, recall Section 2.2.2.

Definition 2.17. ***(Summary measures of spatio-temporal dependence)*** *For a stationary spatio-temporal max-stable process $\{X(\boldsymbol{s},t) : (\boldsymbol{s},t) \in \mathcal{S}\times\mathcal{T}\}$ with univariate margin c.d.f. F, we define*

(i) ***(Spatio-temporal extremal dependence function)***

$$\theta(\boldsymbol{h},l) = -x\log\mathbb{P}(X(\boldsymbol{s},t) \le x, X(\boldsymbol{s}+\boldsymbol{h},t+l) \le x) \in [1,2],\ x > 0. \quad (2.44)$$

(ii) ***(Spatio-temporal upper tail dependence function)***

$$\chi_u(\boldsymbol{h},l) = 2 - \frac{2\log\mathbb{P}\{F(X(\boldsymbol{s},t)) < u, F(X(\boldsymbol{s}+\boldsymbol{h},t+l)) < u\}}{\log\mathbb{P}\{F(X(\boldsymbol{s}+\boldsymbol{h},t+l)) < u\}} \quad (2.45)$$

and $\chi(\boldsymbol{h},l) = \lim_{u\to 1^-}\chi_u(\boldsymbol{h},l)$, $u \in [0,1]$. Similarly to spatial setting, $\chi(\boldsymbol{h},l) = 2-\theta(\boldsymbol{h},l)$.

(iii) ***(Spatio-temporal lower tail dependence function)***

$$\bar{\chi}_u(\boldsymbol{h},l) = \frac{2\log\mathbb{P}(F(X(\boldsymbol{s},t)) > u)}{\log\mathbb{P}(F(X(\boldsymbol{s},t)) > u, F(X(\boldsymbol{s}+\boldsymbol{h},t+l)) > u)} - 1, \quad (2.46)$$

and $\bar{\chi}(\boldsymbol{h},l) = \lim_{u\to 1^-}\bar{\chi}_u(\boldsymbol{h},l)$.

(iv) ***(Spatio-temporal F-madogram)***

$$\nu_F(\boldsymbol{h},l) = \frac{1}{2}\mathbb{E}[|F(X(\boldsymbol{s},t)) - F(X(\boldsymbol{s}+\boldsymbol{h},t+l))|] \in [0,1/6]. \quad (2.47)$$

Furthermore, the F^λ-madogram and λ-madogram can be defined analogously, recall Definitions 2.14 and 2.13.

(v) ***(Spatio-temporal Pickands dependence function)***

$$A_{\boldsymbol{h},l}(\varphi) = \varphi(1-\varphi)V_{\boldsymbol{h},l}(\varphi,1-\varphi),\ \varphi \in (0,1). \quad (2.48)$$

Setting $\varphi = \frac{1}{2}$, it easy to verify that $A_{\boldsymbol{h},l}\left(\frac{1}{2}\right) = \frac{1}{2}\theta(\boldsymbol{h},l)$.

Stationary parametric space-time max-stable models without spectral separability

Below, we briefly review three examples of space-time max-stable processes satisfying (2.41). Recall that $\{\zeta_i\}_{i\geq 1}$ are points of a Poisson process on $(0,\infty)$ with intensity $\zeta^{-2}d\zeta$.

(i) ***BR process in space and time***

[40] introduced the spatial BR model (2.15) in space and time. A strictly stationary spatio-temporal BR process X has the following spectral representation

$$X(\boldsymbol{s},t) = \bigvee_{i=1}^{\infty} \zeta_i \exp\{\varepsilon_i(\boldsymbol{s},t) - \gamma(\boldsymbol{s},t)\},\ (\boldsymbol{s},t) \in \mathcal{S}\times\mathcal{T}, \tag{2.49}$$

where the processes $\{\varepsilon_i(\boldsymbol{s},t) : (\boldsymbol{s},t) \in (\mathcal{S}\times\mathcal{T})\}$ are independent replications of a Gaussian process $\{\varepsilon(t,\boldsymbol{s})\}$ with stationary increments, $\varepsilon(\mathbf{0},0) = 0$, $\mathbb{E}[\varepsilon(\boldsymbol{s},t)] = 0$ and covariance function $\mathbb{C}ov(\varepsilon(\boldsymbol{s}_1,t_1),\varepsilon(\boldsymbol{s}_2,t_2)) = \gamma(\boldsymbol{s}_1,t_1) + \gamma(\boldsymbol{s}_2,t_2) - \gamma(\boldsymbol{s}_1-\boldsymbol{s}_2,t_1-t_2)$, for all $(\boldsymbol{s}_1,t_1),(\boldsymbol{s}_2,t_2) \in \mathcal{S}\times\mathcal{T}$. The dependence function γ which is termed the spatio-temporal semivariogram of the process $\{\varepsilon(\boldsymbol{s},t)\}$, is nonnegative and conditionally negative definite, that is, $\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j \gamma\left(\boldsymbol{s}_i-\boldsymbol{s}_j, t_i-t_j\right) \leq 0$, $\sum_{i=1}^{k} a_i = 0$, for any $k \in \mathbb{N}$, $(\boldsymbol{s}_1,t_1),\ldots,(\boldsymbol{s}_k,t_k) \in \mathcal{S}\times\mathcal{T}$ and $a_1,\ldots,a_k \in \mathbb{R}$.

The process $X(\boldsymbol{s},t)$ in (2.49) is fully characterized by the dependence function γ. The function γ is given by $\gamma(\boldsymbol{s}_1-\boldsymbol{s}_2,t_1-t_2) = \frac{1}{2}\mathbb{V}ar\left(\varepsilon(\boldsymbol{s}_1,t_1)-\varepsilon(\boldsymbol{s}_2,t_2)\right)$ (recall Section 2.1.2). Let Φ denote the standard normal distribution function. For $x_1,x_2>0$, the bivariate exponent function of $(X(\boldsymbol{s}_1,t_1),X(\boldsymbol{s}_2,t_2))$ in the stationary case is given by

$$V_{\boldsymbol{h},l}(x_1,x_2) = \frac{1}{x_1}\Phi\left(\sqrt{\frac{\gamma(\boldsymbol{h},l)}{2}} + \frac{\log\left(\frac{x_2}{x_1}\right)}{\sqrt{2\gamma(\boldsymbol{h},l)}}\right) + \frac{1}{x_2}\Phi\left(\sqrt{\frac{\gamma(\boldsymbol{h},l)}{2}} + \frac{\log\left(\frac{x_1}{x_2}\right)}{\sqrt{2\gamma(\boldsymbol{h},l)}}\right). \tag{2.50}$$

Recall that if γ is assumed to depend only on the norm of $\boldsymbol{s}_1-\boldsymbol{s}_2$, the associated process is spatially isotropic. The pairwise spatio-temporal extremal dependence function for this model is $\theta(\boldsymbol{h},l) = 2\Phi\left\{\sqrt{\gamma(\boldsymbol{h},l)/2}\right\}$ and the spatio-temporal Pickands function is

$$A_{\boldsymbol{h},l}(\varphi) = \varphi\Phi\left(\sqrt{\frac{\gamma(\boldsymbol{h},l)}{2}} + \frac{\log\left(\frac{\varphi}{1-\varphi}\right)}{\sqrt{2\gamma(\boldsymbol{h},l)}}\right) + (1-\varphi)\Phi\left(\sqrt{\frac{\gamma(\boldsymbol{h},l)}{2}} + \frac{\log\left(\frac{1-\varphi}{\varphi}\right)}{\sqrt{2\gamma(\boldsymbol{h},l)}}\right).$$

This model has been used in [23] to quantify the extremal behavior of radar rainfall data in a region in Florida, where a new semi-parametric procedure based on the extremogram [42] is applied to estimate the model parameters.

(ii) ***Smith's storm profile in space-time domain***

An extension of the spatial Smith's storm profile model (2.12) to a space-time setting has been driven in [40], where extremes are observed at certain locations through time. The characterization of a space-time max-stable Smith process is

$$X(\boldsymbol{s},t) = \bigvee_{i=1}^{\infty} \zeta_i \phi(z_i,\eta_i;\boldsymbol{s},t),\ (\boldsymbol{s},t) \in \mathcal{S}\times\mathcal{T}, \tag{2.51}$$

where $\{z_i, \eta_i\}_{i\geq 1}$ denote the points of a Poisson point process on $E_1 \times E_2$ with intensity $\mu_1(dz) \times \mu_2(d\eta)$, $E_1 \times E_2 \subset \mathcal{S} \times \mathcal{T}$, μ_i denotes Lebesgue measure on $\mathcal{S}$ for $i = 1, 2$ and the function ϕ represents the shape of the storm with $\int_{E_1\times E_2} \phi(z, \eta; s, t)\mu_1(dz)\mu_2(d\eta) = 1$. A potential physical interpretation of the construction (2.51) is as follows: think of the product $\zeta_i \phi(z_i, \eta_i; t, s)$ as the wind speed at location s and moment t from the i-th storm with intensity ζ_i, where the wind speed has its maximum at the center of the spatial location z_i at time η_i. According to Theorem 2.6 from [40], a closed form of the bivariate c.d.f. of Smith's model in spatio-temporal setting has been calculated. More precisely, setting the function ϕ as a trivariate Gaussian density with mean (z, η) and covariance matrix $\mathbf{\Sigma}'$, i.e., $\phi(z, \eta; s, t) = \phi_3(z - s, \eta - t)$, where ϕ_3 is a trivariate Gaussian density with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}' = \begin{pmatrix} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \sigma_{33} \end{pmatrix}$, where the spatial dependence is modeled through the matrix $\mathbf{\Sigma}$ and the temporal dependence is given by σ_{33}. With this particular setting, the bivariate exponent function of this model in the stationary case can be expressed as

$$V_{\boldsymbol{h},l}(x_1, x_2) = \frac{1}{x_1}\Phi\left(\frac{2\sigma_{33}\log\left(\frac{x_2}{x_1}\right) + \sigma_{33}b(\boldsymbol{h})^2 + l^2}{2\sqrt{\sigma_{33}^2 b(\boldsymbol{h})^2 + \sigma_{33}l^2}}\right) + \frac{1}{x_2}\Phi\left(\frac{2\sigma_{33}\log\left(\frac{x_1}{x_2}\right) + \sigma_{33}b(\boldsymbol{h})^2 + l^2}{2\sqrt{\sigma_{33}^2 b(\boldsymbol{h})^2 + \sigma_{33}l^2}}\right), \tag{2.52}$$

where $b(\boldsymbol{h})^2 = \boldsymbol{h}^T\mathbf{\Sigma}^{-1}\boldsymbol{h}$. As a consequence of (2.52), by setting time lag $l = 0$, we easily recover spatial max-stable field as calculated in [84], recall also Table 2.1. The extremal dependence function for this model has the form $\theta(\boldsymbol{h}, l) = 2\Phi\left\{\sqrt{\sigma_{33}b(\boldsymbol{h})^2 + l^2}/2\sqrt{\sigma_{33}}\right\}$ and the spatio-temporal Pickands dependence function is

$$A_{\boldsymbol{h},l}(\varphi) = \varphi\Phi\left(\frac{2\sigma_{33}\log\left(\frac{\varphi}{1-\varphi}\right) + \sigma_{33}b(\boldsymbol{h})^2 + l^2}{2\sqrt{\sigma_{33}^2 b(\boldsymbol{h})^2 + \sigma_{33}l^2}}\right) + (1-\varphi)\Phi\left(\frac{2\sigma_{33}\log\left(\frac{1-\varphi}{\varphi}\right) + \sigma_{33}b(\boldsymbol{h})^2 + l^2}{2\sqrt{\sigma_{33}^2 b(\boldsymbol{h})^2 + \sigma_{33}l^2}}\right).$$

Lastly, it is easy to show that the bivariate distribution of Smith's storm model (2.52) is a special case of BR (2.50) when $\gamma(\boldsymbol{h}, l) = \frac{1}{2}\left\{b(\boldsymbol{h})^2 + \frac{l^2}{\sigma_{33}}\right\}$.

(iii) ***Schlather TEG space-time model***

[64] considered an extension of the spatial Schlather TEG model (2.14) that comprises a truncated Gaussian process, so that storm shapes are stochastic, and includes a compact random set which allows independent extremes. More precisely, the space-time max-stable TEG process is defined by

$$X(s,t) = \sqrt{2\pi}(\mathbb{E}\{|\mathcal{A}|\})^{-1}\bigvee_{i=1}^{\infty}\zeta_i W_i^+(s,t)\mathbb{1}_{\mathcal{A}_i}\{(s,t) - P_i\},\ (s,t) \in \mathcal{S}\times\mathcal{T}, \tag{2.53}$$

here $|.|$ is used to denote the volume of a set, $W_i^+(s,t) = \max\{0, \varepsilon_i(s,t)\}$, $\{\varepsilon_i(s,t)\}_{i\geq 1}$ are independent replicates of a Gaussian random field with space-time correlation function ρ, $\mathbb{1}_{\mathcal{A}}$ is the indicator function of a compact random set $\mathcal{A} \subset \mathcal{S}\times\mathcal{T}$, $\mathcal{A}_i$ are independent

replicates of $\mathcal{A}$ and P_i are points of a unit Poisson process on $\mathcal{S}\times\mathcal{T}$, independent of ε_i. The bivariate exponent function of this model in the stationary case has the form

$$V_{\boldsymbol{h},l}(x_1,x_2)=\left(\frac{1}{x_1}+\frac{1}{x_2}\right)\times\left[1-\frac{\alpha(\boldsymbol{h},l)}{2}\left(1-\sqrt{1-\frac{2\{\rho(\boldsymbol{h},l)+1\}x_1x_2}{(x_1+x_2)^2}}\right)\right], \tag{2.54}$$

where $\alpha(\boldsymbol{h},l)=\mathbb{E}[|\mathcal{A}\cap\{(\boldsymbol{h},l)+\mathcal{A}\}|]/\mathbb{E}(|\mathcal{A}|)$. Accordingly, the pairwise spatio-temporal extremal coefficient is $\theta(\boldsymbol{h},l)=2-\alpha(\boldsymbol{h},l)\left\{1-\sqrt{(1-\rho(\boldsymbol{h},l))/2}\right\}$ and the spatio-temporal Pickands dependence function is $A_{\boldsymbol{h},l}(\varphi)=2-\alpha(\boldsymbol{h},l)\left\{1-\sqrt{\frac{(1-\rho(\boldsymbol{h},l))\varphi(1-\varphi)}{2}}\right\}$. In [64], this model has been fitted to hourly rainfall extremes in western Switzerland with the adoption of pairwise likelihood for statistical inference, see Section 3.1.2.

The fundamental advantages of the spectral representation in (2.41) are (i) the construction of spatio-temporal models from widely studied max-stable processes (ii) the large literature available on spatio-temporal correlation functions for Gaussian processes, allowing for considerable diversity of spatio-temporal behavior. However, time has no specific role but is equivalent to an additional spatial dimension (the spatial and temporal distributions belong to a similar family of models). To overcome this defect, a new class of space-time max-stable models has been suggested in [54]. It allows distinct roles for space and time.

2.3.2 Space-time max-stable models with spectral separability

Definition 2.18. ***(Space-time max-stable models with spectral separability, [54])*** *The class of space-time max-stable models with spectral separability is defined by the following spectral decomposition:*

$$X(\boldsymbol{s},t)=\bigvee_{i=1}^{\infty}\zeta_i U_t(Q_i)U_{\mathcal{R}(t,Q_i)\boldsymbol{s}}(W_i),\quad \textit{where} \tag{2.55}$$

- $\{\zeta_i,Q_i,W_i\}_{i\geq 1}$ *are the points of a Poisson process on* $(0,\infty)\times E_1\times E_2$*, and with intensity* $\zeta^{-2}d\zeta\times\mu_1(dq)\times\mu_2(dw)$ *for some Polish measure spaces* $(E_1,\mathcal{E}_1,\mu_1)$ *and* $(E_2,\mathcal{E}_2,\mu_2)$*,*
- *the spectral function* $U_t:E_1\to(0,\infty)$ *is measurable such that* $\int_{E_1}U_t(q)\mu_1(dq)=1$ *for each* $t\in\mathcal{T}$ *and contributes to the temporal dynamic of the process, whereas the spectral function* $U_{\boldsymbol{s}}:E_2\to(0,\infty)$ *is measurable such that* $\int_{E_2}U_{\boldsymbol{s}}(w)\mu_2(dw)=1$ *for each* $\boldsymbol{s}\in\mathcal{S}$ *and drives the shape of the main spatial patterns,*
- *the operators* $\mathcal{R}(t,q)$ *are bijective from* $\mathcal{S}$ *to* $\mathcal{S}$ *for each* $(t,q)\in\mathcal{T}\times E_1$ *and describes how the spatial patterns move in space.*

The construction (2.55) allows one to deal with the temporal and spatial aspects separately. So, the estimation procedure can be simplified by estimating in a first step the spatial parameters independently from the temporal ones. Several examples of subclasses of the general class of space-time process X (2.55) were introduced by [54], where the operator is either a translation or a rotation. The authors in that paper focused mainly on a special case of models where

the function corresponding to the time in the spectral representation is the exponential density (continuous-time case) or the probability values of a geometric random variable (discrete-time case). So, the corresponding models become Markovian and have a useful max-autoregressive representation, i.e.,

$$X(\boldsymbol{s},t) = \max\{\delta X(\boldsymbol{s}-\boldsymbol{\tau},t-1),(1-\delta)H(\boldsymbol{s},t)\},\ (\boldsymbol{s},t)\in\mathcal{S}\times\mathcal{T}, \tag{2.56}$$

where the parameter $\delta\in(0,1)$ measures the influence of the past, the parameter $\boldsymbol{\tau}\in\mathbb{R}^2$ represents some type of specific direction of propagation/contagion in space and $H =: \{H(\boldsymbol{s},t), \boldsymbol{s}\in\mathcal{S}, t\in\mathcal{T}\}$ is a time-independent process that is derived from independent replications of a spatial max-stable process $\{H(\boldsymbol{s}), \boldsymbol{s}\in\mathcal{S}\}$.

Indeed, the model in (2.56) can be seen as an extension of the real-valued max-autoregressive moving-average process MARMA(1,0) to the spatial context, see [43]. The value at location $\boldsymbol{s}$ and time t is either related to the value at location $\boldsymbol{s}-\boldsymbol{\tau}$ at time $t-1$ or to the value of another process (the innovation), H, that characterizes a new event happening at location $\boldsymbol{s}$. This model may be useful for phenomena that propagate in space.

Definition 2.19. ***(max-autoregressive moving-average process, [43])*** *The real-valued process $\{M(t)\}_{t\in\mathbb{Z}}$ follows the max-autoregressive moving-average process of orders p and q, i.e., MARMA(p,q), if it satisfies the recursion*

$$M(t) = \max\left\{a_1M(t-1),\ldots,a_pM(t-p),A(t),b_1A(t-1),\ldots,b_qA(t-q)\right\},$$

where $a_i,b_j\geq 0$ for $i=1,\ldots,p$, $j=1,\ldots,q$, and $\{A(t)\}_{t\in\mathbb{Z}}$ are i.i.d. max-stable random variables.

In the following, we will focus on the processes satisfying (2.56). Let $V_{\mathbf{0},\boldsymbol{h}-l\boldsymbol{\tau}}$ denote the exponent function characterizing the spatial distribution of the process $H(\boldsymbol{s},t)$, then the bivariate c.d.f. $F_{\boldsymbol{h},l}$ of $(X(\mathbf{0},0),X(\boldsymbol{h},l))$ can be expressed for $x_1,x_2>0$ as

$$-\log F_{\boldsymbol{h},l}(x_1,x_2) = V_{\mathbf{0},\boldsymbol{h}-l\boldsymbol{\tau}}\left(x_1,\frac{x_2}{\delta^l}\right)+\frac{1-\delta^l}{x_2}. \tag{2.57}$$

Moreover, the spatio-temporal extremal dependence function in (2.44) can be easily deduced in this case by setting $x_1=x_2=x$ in (2.57),

$$\theta(\boldsymbol{h},l) = V_{\mathbf{0},\boldsymbol{h}-l\boldsymbol{\tau}}\left(1,\delta^{-l}\right)+1-\delta^l. \tag{2.58}$$

Clearly, space and time are not fully separated in the extremal dependence function, even if $\boldsymbol{\tau}=\mathbf{0}$ (space and time are completely separated in the spectral representation, where U_t depends on time and $U_{\mathcal{R}(t,q)\boldsymbol{s}}=U_{\boldsymbol{s}}$ depends only on space). Asymptotic time independence is achieved when $(\lim_{l\to\infty}\theta(\boldsymbol{h},l)\to 2)$. Furthermore, X is called strongly mixing in space (respectively, time) if and only if $\lim_{\|\boldsymbol{h}\|\to\infty}\theta(\boldsymbol{h},0)\to 2$ (respectively, $\lim_{l\to\infty}\theta(\mathbf{0},l)\to 2$), recall Theorem 2.3. In the sequel, we give two examples of a bivariate space-time max-stable process satisfying (2.56).

(i) ***Spectrally separable space-time max-stable Smith process***

If the innovation process H is derived from independent replications of a spatial Smith process with a covariance matrix $\boldsymbol{\Sigma}$, recall (2.12). Then the bivariate c.d.f. $F_{\boldsymbol{h},l}$ of the resulting spatio-temporal model in (2.56) has the form

$$-\log F_{\boldsymbol{h},l}(x_1,x_2) = \frac{1}{x_1}\Phi\left(\frac{b(\boldsymbol{h},l)}{2}+\frac{1}{b(\boldsymbol{h},l)}\log\left(\frac{x_2}{\delta^l x_1}\right)\right) + \frac{\delta^l}{x_2}\Phi\left(\frac{b(\boldsymbol{h},l)}{2}+\frac{1}{b(\boldsymbol{h},l)}\log\left(\frac{\delta^l x_1}{x_2}\right)\right)+\frac{1-\delta^l}{x_2}, \tag{2.59}$$

where $b(\boldsymbol{h},l)=\sqrt{(\boldsymbol{h}-l\boldsymbol{\tau})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{h}-l\boldsymbol{\tau})}$. The associated spatio-temporal extremal dependence function with this model is

$$\theta(\boldsymbol{h},l)=\Phi\left(\frac{b(\boldsymbol{h},l)}{2}+\frac{1}{b(\boldsymbol{h},l)}\log\left(\delta^{-l}\right)\right)+\delta^l\Phi\left(\frac{b(\boldsymbol{h},l)}{2}+\frac{1}{b(\boldsymbol{h},l)}\log\left(\delta^{l}\right)\right)+1-\delta^l. \tag{2.60}$$

(ii) ***Spectrally separable space-time max-stable Schlather process***

The spatio-temporal model in (2.56) when the innovation process H is derived from independent replications of a spatial Schlather process with correlation function $\rho(.)$, recall (2.13), has a bivariate c.d.f. $F_{\boldsymbol{h},l}$ of the form

$$-\log F_{\boldsymbol{h},l}(x_1,x_2)=\frac{1}{2}\left(\frac{1}{x_1}+\frac{\delta^l}{x_2}\right)\times\left[\left(1+\sqrt{1-\frac{2\delta^l(\rho(\boldsymbol{h},l)+1)x_1x_2}{(\delta^l x_1+x_2)^2}}\right)\right]+\frac{1-\delta^l}{x_2}, \tag{2.61}$$

where $\rho(\boldsymbol{h},l)$ is the spatio-temporal correlation function associated to this model. The associated spatio-temporal extremal coefficient with this model is

$$\theta(\boldsymbol{h},l)=\frac{1}{2}(1+\delta^l)\left[\left(1+\sqrt{1-\frac{2\delta^l(\rho(\boldsymbol{h},l)+1)}{(1+\delta^l)^2}}\right)\right]+1-\delta^l. \tag{2.62}$$

If the time lag $l=0$, the formulas in (2.59) and (2.61) reduce to the bivariate distributions of the max-stable spatial fields given in Table 2.1.

2.3.3 Space-time Gaussian correlation functions

Recently, a significant growth of interest in spatio-temporal correlation models has emerged. This section introduces some basic concepts for space-time Gaussian correlation functions. For a detailed review of spatio-temporal correlation functions, see [48, 80] and the references therein.

Definition 2.20. ***(Space-time correlation function)*** *The correlation function of a Gaussian space-time process $\{\varepsilon(\boldsymbol{s},t):(\boldsymbol{s},t)\in\mathcal{S}\times\mathcal{T}\}$ is defined by*

$$\rho(\boldsymbol{s}_1,t_1;\boldsymbol{s}_2,t_2)=\frac{\mathbb{C}ov(\varepsilon(\boldsymbol{s}_1,t_1),\varepsilon(\boldsymbol{s}_2,t_2))}{\sqrt{\mathbb{V}ar(\varepsilon(\boldsymbol{s}_1,t_1))\mathbb{V}ar(\varepsilon(\boldsymbol{s}_2,t_2)))}}.$$

For simplicity, we assume that $\mathbb{V}ar(\varepsilon(\boldsymbol{s},t))=1$ for all $\boldsymbol{s}\in\mathcal{S}$ and $t\in\mathcal{T}$.

Definition 2.21. ***(Basic concepts for space-time correlation functions)*** *We call the space-time correlation function ρ*

(i) ***stationary****, if ρ only depends on the spatial and the temporal lags. More precisely,*

$$\rho(\boldsymbol{s}_1, t_1) = \rho(\boldsymbol{s}_1 + \boldsymbol{h}, t_1 + l) =: \rho(\boldsymbol{h}, l),\ \boldsymbol{h} \in \mathbb{R}^d,\ l \in \mathbb{R}.$$

(ii) ***isotropic****, if the stationary correlation function only depends on the norm of spatial lag and the absolute value of the temporal lag, that is, $\rho(\boldsymbol{h}, l) =: \rho(\|\boldsymbol{h}\|, |l|)$. Note that if the stationary correlation function is isotropic in space and time, then it is also* ***fully symmetric****, that is, $\rho(\boldsymbol{h}, l) = \rho(-\boldsymbol{h}, l) = \rho(\boldsymbol{h}, -l) = \rho(-\boldsymbol{h}, -l)$, $\forall (\boldsymbol{h}, l) \in \mathbb{R}^d \times \mathbb{R}$.*

(iii) ***separable****, if ρ can be separated into a purely temporal correlation function ρ_t and a purely spatial correlation function ρ_s. For instance,*

- *the product model (factorizing): $\rho(\boldsymbol{s}_1, t_1; \boldsymbol{s}_2, t_2) = \rho_s(\boldsymbol{s}_1, \boldsymbol{s}_2)\rho_t(t_1, t_2)$.*
- *the sum model (additive): $\rho(\boldsymbol{s}_1, t_1; \boldsymbol{s}_2, t_2) = \rho_s(\boldsymbol{s}_1, \boldsymbol{s}_2) + \rho_t(t_1, t_2)$.*

Figure 2.11 displays two examples of parametric families of isotropic separable space-time correlation functions under the product model. The plots have been obtained using the function *drape.plot* of the R package fields. Separable models do not allow for spatio-temporal interaction. This is unrealistic in many applications, see, e.g., [22, 21]. Therefore, ***non-separable*** model constructions have been developed. For instance, a very general flexible class of spatio-temporal correlation models has been proposed by [59].

Definition 2.22. ***(Gneiting's class of correlation functions)*** *Let $\phi\colon \mathbb{R}^+ \to \mathbb{R}$ be a completely monotone function and let $\psi\colon \mathbb{R}^+ \to \mathbb{R}$ be a positive function with completely monotone derivative. Then, the function*

$$\rho(\|\boldsymbol{h}\|, |l|) = \frac{1}{\psi(|l|^2)} \phi\left(\frac{\|\boldsymbol{h}\|^2}{\psi(|l|^2)}\right), \tag{2.63}$$

is a non-separable valid isotropic spatio-temporal correlation function.

Table 2.2 lists some known completely monotone functions, whereas Table 2.3 provides some examples of positive functions with completely monotone derivatives. Note that a continuous function $f(x)$, $x \geq 0$ is said to be completely monotonic if it satisfies: $(-1)^n \frac{d^n}{dx^n} f(x) \geq 0$, $n = 1, 2, \ldots$, $x > 0$. Several families of valid space-time correlation functions can be constructed using the entries in Tables 2.2 and 2.3, allowing for considerable diversity of spatio-temporal behavior.

Example 2.3. ***(Non-separable flexible spatio-temporal correlation function)*** *By combining the first entries in Tables 2.2 and 2.3 with expression (2.63) and then multiplying by the purely temporal correlation function $\{1 + (|l|/b_t)^{a_t}\}^{-1}$, we obtain*

$$\rho(\|\boldsymbol{h}\|, |l|) = \left\{1 + \left(\frac{|l|}{b_t}\right)^{a_t}\right\}^{-(\beta+1)} \exp\left[-\frac{\left(\frac{\|\boldsymbol{h}\|}{b_s}\right)^{a_s}}{\left\{1 + \left(\frac{|l|}{b_t}\right)^{a_t}\right\}^{b_s \beta/2}}\right], \tag{2.64}$$

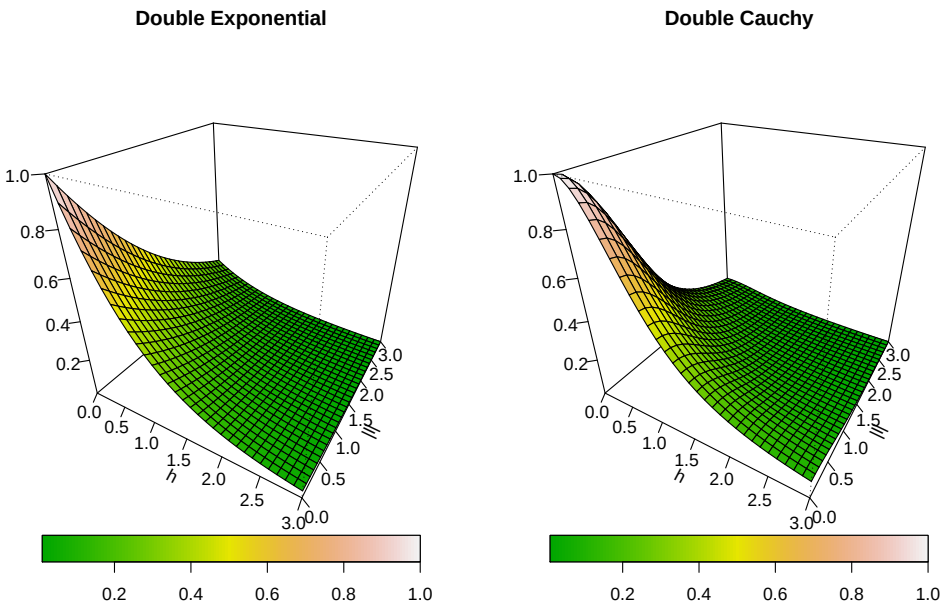


Figure 2.11: Isotropic separable space-time correlation functions. Double exponential correlation function $\rho(\|\boldsymbol{h}\|,|l|) = \exp\{-(0.5\|\boldsymbol{h}\| + |l|)\}$ (left panel). Double cauchy correlatio function $\rho(\|\boldsymbol{h}\|,|l|) = \{1+\|\boldsymbol{h}\|^2\}^{-1}\{1+|l|^2\}^{-1}$ (right panel), where $h = \|\boldsymbol{h}\|$.

where $b_t, b_s > 0$ determine spatial and temporal scale parameters, $a_t, a_s \in (0,2]$ are temporal and spatial smoothness parameters, and $\beta \in [0,1]$ is the separability parameter quantifying the space-time interactions, see Section 5.2 in [64]. Accordingly, when $\beta = 0$, the correlation function (2.64) is separable, whereas, as $\beta \to 1$, the spatial and temporal components become increasingly entwined.

Table 2.2: Some completely monotone functions $\phi(r)$, $r > 0$. K_c denotes the modified Bessel function of order c.

Function $\phi(r)$	Parameters
$\exp\{-(\sqrt{r}/b)^a\}$	$b > 0,\ a \in (0,2]$
$\{1+(\sqrt{r}/b)^a\}^{-c}$	$b,\ c > 0,\ a \in (0,2]$
$\{2^{c-1}\Gamma(c)\}^{-1}(2\sqrt{cr}b^{-1})^c K_c(2\sqrt{cr}b^{-1})$	$c, b > 0$

Table 2.3: Some positive functions $\psi(t)$, $t > 0$, with a completely monotone derivative.

Function $\psi(t)$	Parameters
$\{(\sqrt{t}/b)^a + 1\}^{\zeta}$	$b > 0,\ a \in (0,2],\ \zeta \in [0,1]$
$\log\{(\sqrt{t}/b)^a + c\}/\log(c)$	$b > 0,\ c > 1,\ a \in (0,2]$
$\{(\sqrt{t}/b)^a + c\}/[c\{(\sqrt{t}/b)^a + 1\}]$	$b > 0,\ c \in (0,1],\ a \in (0,2]$

Chapter 3

Pairwise likelihood-based tests for mixture parameter of spatial max-mixture models

This chapter is based on our results from the paper: A. Abu-Awwad, V. Maume-Deschamps and P. Ribereau—Censored pairwise likelihood-based tests for mixture parameter of spatial max-mixture models. Accepted for publication to *Revista de Investigacion Operacional*.

Testing of hypotheses is one of the main tools in statistics and crucial in many applications. For instance, within the spatial framework, a madogram statistical testing procedure has been proposed by [9] to discriminate between asymptotic independence and asymptotic dependence at various lags. More precisely, for a stationary isotropic spatial process $\{X(\boldsymbol{s})\}_{s\in\mathcal{S}}$ with marginal distribution F, the asymptotic independence of the pairs $(X(\boldsymbol{s}_1), X(\boldsymbol{s}_2))$ such that $\|\boldsymbol{s}_1 - \boldsymbol{s}_2\| = h$ can be checked by testing the null hypothesis $H_0 : \nu_F(h) = \frac{1}{6}$ against $H_1 : \nu_F(h) < \frac{1}{6}$ using the following straightforward statistic

$$\sqrt{T}\frac{\widehat{\nu}_F(h) - \frac{1}{6}}{\widehat{\sigma}(h)} \xrightarrow{\mathcal{D}} N(0,1), \text{ as } T \to \infty, \tag{3.1}$$

where $\widehat{\nu}_F(.)$ is the empirical version of $\nu_F(.)$ over T independent observations and $\sigma^2(h) = \mathbb{V}ar\left\{\frac{1}{2}|F(X(\boldsymbol{s}_1)) - F(X(\boldsymbol{s}_2))|\right\}$. In that study, the test (3.1) has been applied to examine the spatial asymptotic independence of two climatological datasets: daily maxima temperature in France and daily maxima precipitations in the French Burgundy region. Although its simplicity, the major advantage of this test is that it is a model-free test. Needless to specify a distribution family prior to the analysis.

In the spatial max-mixture framework, the mixture parameter a controls the level of the asymptotic dependence component present in the max-mixture process $\{Z(\boldsymbol{s})\}_{s\in\mathcal{S}}$, recall (2.34). In this chapter, we focus on two model-based statistical tests for the mixing parameter a. Pairwise likelihood has employed for statistical inference.

The remainder of the chapter is organized as follows. Pairwise likelihood estimation for max-mixture processes is presented in Section 3.1. The proposed pairwise statistics and their main properties are detailed in Section 3.2. In Section 3.3, we show by various simulation studies the performance of the proposed statistics. In Section 3.4, we apply the proposed testing approach to daily precipitation from the East of Australia. Concluding remarks and limitations are addressed in Section 3.5.

3.1 Inference for max-mixture processes

Modeling joint occurrence of spatial extremes over a region usually comprises the following two steps: (i) estimating the marginal distribution and (ii) characterizing the dependence via a model issued from the multivariate extreme value theory. So, a proper inferential approach requires to fit marginal and dependence parameters. To that aim, the spatial extreme process is supposed to be a plausible model for block maxima or large threshold exceedances. The block maxima approach is based on dividing the observation period into non-overlapping periods of equal size and restrict attention to the maximum observation in each period (e.g., seasonal maxima). The choice of block size can be critical. Too small blocks may lead to bias in the estimation and extrapolation and large ones may lead to large variance in estimation. On the other hand, this approach becomes wasteful if other data on extremes are available. Then, it is natural to regard as extreme events that exceed some high threshold. However, the issue of threshold choice is analogous to the choice of block size in the block maxima approach, implying a trade-off between bias and variance. In this case, too low thresholds are likely to violate the asymptotic support of the model, leading to bias in the estimation, while too high thresholds will generate few excesses, leading to high variance in estimation.

3.1.1 Marginal fitting

The max-mixture model in (2.34) assumes a unit marginal Fréchet distributions. Since, in practical applications, this is not the case, marginal laws have to be transformed to unit Fréchet margins. Consequently, the marginal distribution F has to be estimated, and this can be done in the following ways

(i) nonparametrically by the empirical distribution function, i.e., $\widehat{F}(z) = \frac{1}{T+1}\sum_{i=1}^{T} \mathbb{1}_{Z_i(\boldsymbol{s})<z}$, where $\mathbb{1}$ is the indicator function, see, e.g., [81, 15, 5],

(ii) parametrically by fitting the GEV distribution in (2.6) separately to each site $\boldsymbol{s}$, which is usually applied to block maxima approach, i.e., $\widehat{F}(z) = \exp\left\{-\left(1+\widehat{\xi}(\boldsymbol{s})\frac{z-\widehat{\mu}(\boldsymbol{s})}{\widehat{\sigma}(\boldsymbol{s})}\right)^{-1/\widehat{\xi}(\boldsymbol{s})}\right\}$, see, e.g., [11, 101],

(iii) semi-parametrically, F may approximated empirically below the marginal threshold u and a site-wise fitted Generalized Pareto Distribution (GPD) above the threshold u, which is the standard way for threshold exceedance approach, see [34, 67], i.e.,

$$\widehat{F}(z) = \begin{cases} \widetilde{F}(z), & z \le u, \\ 1-\widehat{\zeta}_u(\boldsymbol{s})\left\{1+\widehat{\xi}(\boldsymbol{s})(z-u)/\widehat{\sigma^*}(\boldsymbol{s})\right\}^{-1/\widehat{\xi}(\boldsymbol{s})}, & z > u, \end{cases} \tag{3.2}$$

where $\widetilde{F}(z)$ is the empirical distribution function, $\widehat{\zeta}_u$ is the estimated probability of exceeding the threshold u and $\left(\widehat{\sigma^*}, \widehat{\xi}\right)$ are estimates of the GPD parameters (σ^*, ξ), which has the form (see [47])

$$\mathrm{GPD}_{\sigma^*(\boldsymbol{s}),\xi(\boldsymbol{s})}(z) = 1-\left(1+\frac{\xi(\boldsymbol{s})}{\sigma^*(\boldsymbol{s})}z\right)_+^{-1/\xi(\boldsymbol{s})}, \; z>0, \tag{3.3}$$

where the scale parameter is linked to that of the GEV distribution by $\sigma^* = \sigma + \xi(u - \mu)$ and the shape parameter ξ is the same as for GEV distribution.

Then according to the probability integral transformation: $t(z) = -\frac{1}{\log(\widehat{F}(z))}$, the observations are approximately standardizes to have unit Fréchet distribution. Sometimes it may be useful to transform to other marginal distributions such as Gumbel $\left(t(z) = -\log[-\log(\widehat{F}(z))]\right)$ or Weibull $\left(t(z) = \log(\widehat{F}(z))\right)$.

3.1.2 Dependence parameters fitting: composite likelihood approach

Likelihood inference for multivariate or spatial extremes is computationally challenging. In a max-stable framework, the c.d.f. can be easily written as (2.8), whereas the density for a single vector $\boldsymbol{x} = (x_1, \ldots, x_D)^t$, D is the number of monitoring stations, is much more complicated when D is large, since it corresponds to the derivative of (2.8) with respect to all components of $\boldsymbol{x}$, recall (2.9). For example, with $D = 10$ there are around 10^5 terms in the density. To bypass this difficulty, the usual procedure has been to maximize the composite likelihoods constructed from small subsets of data (e.g., [17, 73, 98, 99]), on the basis of pairs (e.g., [84, 44]) or triples (e.g., [58, 63, 89]). Under some regularity conditions, these misspecified likelihoods enable consistent but generally less efficient estimation of parameters and are computationally convenient. However, high-order composite likelihood inference for max-stable processes is sometimes possible. For instance, full likelihood inference for BR process seems to be restricted to dimension $D = 12$ or $D = 13$ with current technologies. However, the computations are very time-consuming and need a very powerful computer, see [26].

Before focusing on max-mixture processes, let us introduce the notion of composite likelihood. Let $\boldsymbol{Z}$ be an N-dimensional random vector with p.d.f. $g(\boldsymbol{z}; \boldsymbol{\psi})$, for some q-dimensional parameter $\boldsymbol{\psi} \in \boldsymbol{\Psi}$. Let $\{\mathcal{E}_1, \ldots, \mathcal{E}_K\}$ denote a set of marginal or conditional events with associated likelihoods $g(\boldsymbol{z} \in \mathcal{E}_k; \boldsymbol{\psi})$. Following [73], the composite likelihood for n independent replicates of $\boldsymbol{Z}$ can be constructed as

$$\mathcal{L}_c(\boldsymbol{\psi}) = \prod_{i=1}^{n} \prod_{k=1}^{K} g(\boldsymbol{z}_i \in \mathcal{E}_k; \boldsymbol{\psi})^{\omega_k}, \tag{3.4}$$

where ω_k are a nonnegative deterministic weights. The composite log-likelihood is

$$\ell_c(\boldsymbol{\psi}) = \sum_{i=1}^{n} \sum_{k=1}^{K} \omega_k \log g(\boldsymbol{z}_i \in \mathcal{E}_k; (\boldsymbol{\psi}). \tag{3.5}$$

The maximum composite likelihood estimator $\widehat{\boldsymbol{\psi}}_c$ is obtained by maximizing (3.5), i.e., $\widehat{\boldsymbol{\psi}}_c = \arg\max_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} \ell_c(\boldsymbol{\psi})$. Marginal likelihoods (e.g., pseudolikelihood, pairwise likelihood, triple-wise likelihood) which are constructed from lower-order marginal densities form a subclass of composite likelihood.

For our purpose of fitting max-mixture processes, we will adopt the widely used pairwise likelihood approach. Let z_{ik} denote the site-wise block maximum of the max-mixture process of interest $\{Z(\boldsymbol{s})\}_{s \in \mathcal{S}}$, observed at site $\boldsymbol{s}_i$, $i = 1, \ldots, D$ and at time t_k, $k = 1, \ldots, T$, where the

observations are assumed to be independent in time. We denote by $\boldsymbol{\vartheta}$ the vector gathering the parameters to be estimated using pairwise likelihood. Then the (weighted) pairwise log-likelihood is

$$\ell_p(\boldsymbol{\vartheta}) = \sum_{k=1}^{T}\sum_{i=1}^{D-1}\sum_{j>i}^{D} \omega_{ij} \log \mathcal{L}_p(z_{ik}, z_{jk}; \boldsymbol{\vartheta}) =: \sum_{k=1}^{T} \ell_{pk}(\boldsymbol{\vartheta}), \tag{3.6}$$

where $\mathcal{L}_p(z_{ik}, z_{jk}; \boldsymbol{\vartheta})$ is the likelihood of the pair (z_{ik}, z_{jk}), and ω_{ij} are nonnegative weights specifying the contribution of each pair. Then, the maximum pairwise likelihood estimator is $\widehat{\boldsymbol{\vartheta}}_p = \arg\max \ell_p(\boldsymbol{\vartheta})$.

Unfortunately there is no existing theory to define optimal weights. However, a careful choice of the weights ω_{ij} could improve the estimation efficiency as well as the computational one. Various weighting strategies have been proposed in this context, see, e.g., [61, 26, 19, 64, 89]. For example, a simple weighting choice is to let

$$\omega_{ij} = \mathbb{1}_{\{\|\boldsymbol{s}_i - \boldsymbol{s}_j\| \leq h^*\}}, \tag{3.7}$$

for some specified value h^*, where $\mathbb{1}_{\{\mathcal{B}\}}$ represents the indicator function of $\mathcal{B}$, see, e.g., [61]. With this approach, the selection of the threshold h^* is critical, where the complete pairwise likelihood is recovered when $h^* \to \infty$. Complete pairwise likelihood has been adopted by [18] (respectively, [44]) to fit max-stable processes to annual maxima of snow depth in the Alpine region (respectively, annual temperature maxima in Switzerland). Generally, neighboring sites are more informative about dependence parameters than distant ones. Thus, in [10, 11], it is suggested to choose h^* as the q-quantile of the distribution of the distances between pairs of sites, $q \in [0, 1]$. Another possibility is to downweight the distant pairs such that all $\omega_{ij} \neq 0$ $\Big($e.g., $\omega_{ij} = 1/\|\boldsymbol{s}_i - \boldsymbol{s}_j\|$ or $\exp\left\{-\tau\|\boldsymbol{s}_i - \boldsymbol{s}_j\|\right\}$, $\tau > 0$ $\Big)$, this approach might improve the statistical efficiency, but no computational gains can be expected with respect to complete pairwise likelihood. Furthermore, as already pointed out by [26], when the dimensionality is too high, the use of a truncated composite likelihood (or also called tapered composite likelihood, see [89]), that is, using binary weights $\omega_{ij} = 0$ or 1, is absolutely needed.

On the other hand, according to [101], under asymptotic independence, a more efficient inference is feasible by modeling the extremes of original events that exceed a large threshold u rather than site-wise maxima. More precisely, let $F^Z_{\boldsymbol{h}}$ denote the bivariate c.d.f. of the max-mixture process Z, recall (2.35), and $u \in \mathbb{R}$ a high threshold; the censored pairwise likelihood contribution is defined as follows

$$\mathcal{L}^u_p(z_{ik}, z_{jk}; \boldsymbol{\vartheta}) = \begin{cases} F^Z_{\boldsymbol{h}}(u, u; \boldsymbol{\vartheta}), & \text{if } \max(z_{ik}, z_{jk}) \leq u, \\ \partial^2_{12} F^Z_{\boldsymbol{h}}(z_{ik}, z_{jk}; \boldsymbol{\vartheta}), & \text{if } \max(z_{ik}, z_{jk}) > u, \end{cases} \tag{3.8}$$

where ∂_i is the differentiation with respect to the variable z_i, and z_{ik} is the observed value at the i-th site and k-th time. So, the maximum censored pairwise likelihood estimator is given by $\widehat{\boldsymbol{\vartheta}}^u_p = \text{argmax } \ell^u_p(\boldsymbol{\vartheta})$, where $\ell^u_p(.)$ is the (weighted) censored pairwise log-likelihood. The choice of the threshold u, which is sometimes aided by diagnostic plots, is crucial. A common choice is to set u corresponding to the empirical q-quantile at each site, provided that q is sufficiently large, see, e.g., [64, 11]. The estimation methodologies in (3.6) and (3.8) have been

used satisfactorily for fitting spatial max-mixture processes by [101] (respectively, [11]) in an analysis of extremes of the winter observations of a hindcast dataset of significant wave height from the North Sea (respectively, daily precipitations over the East of Australia).

Remark 3.1. *Different censoring approaches of the pairwise likelihood have been introduced. In addition to (3.8), the censored pairwise likelihood contribution $\mathcal{L}_p^u(z_{ik}, z_{jk}; \boldsymbol{\vartheta})$ of a pair (z_{ik}, z_{jk}) can be taken as follows*

$$\mathcal{L}_p^u(z_{ik}, z_{jk}; \boldsymbol{\vartheta}) = \begin{cases} F_{\boldsymbol{h}}^Z(u, u; \boldsymbol{\vartheta}), & \text{if } max\ (z_{ik}, z_{jk}) \leq u, \\ \partial_1 F_{\boldsymbol{h}}^Z(z_{ik}, u; \boldsymbol{\vartheta}), & \text{if } z_{ik} > u, z_{jk} \leq u, \\ \partial_2 F_{\boldsymbol{h}}^Z(u, z_{jk}; \boldsymbol{\vartheta}), & \text{if } z_{ik} \leq u, z_{jk} > u, \\ \partial_{12}^2 F_{\boldsymbol{h}}^Z(z_{ik}, z_{jk}; \boldsymbol{\vartheta}), & \text{if } min(z_{ik}, z_{jk}) > u. \end{cases} \tag{3.9}$$

This approach has proved to be useful for the statistical inference of spatial/spatio-temporal extremes, see, e.g., [64, 102, 97, 65].

3.1.3 Asymptotics and assessing uncertainties

For notational simplicity, we will write $\widehat{\boldsymbol{\vartheta}}$ for $\widehat{\boldsymbol{\vartheta}}_p$ / $\widehat{\boldsymbol{\vartheta}}_p^u$ and $\ell(.)$ for $\ell_p(.)$ / $\ell_p^u(.)$. Asymptotics properties of $\widehat{\boldsymbol{\vartheta}}$ are available from [11, 84, 70, 73], which we now summarize. For large T and under some regularity conditions (see, Section 9.2.2 from [79]), $\widehat{\boldsymbol{\vartheta}}$ is asymptotically Gaussian with asymptotic variance $\mathcal{G}^{-1}(\boldsymbol{\vartheta}) = \mathcal{H}^{-1}(\boldsymbol{\vartheta})\mathcal{J}(\boldsymbol{\vartheta})\mathcal{H}^{-1}(\boldsymbol{\vartheta})$, where $\mathcal{G}(\boldsymbol{\vartheta})$ is the ***Godambe information matrix***, $\mathcal{H}(\boldsymbol{\vartheta}) = \mathbb{E}\left\{-\nabla^2\ell(\boldsymbol{\vartheta})\right\} =: \mathbb{E}\{-\nabla_{\boldsymbol{\vartheta}} U(\boldsymbol{\vartheta})\}$ is called the ***sensitivity matrix***, $\mathcal{J}(\boldsymbol{\vartheta}) = \mathbb{V}ar\{\nabla\ell(\boldsymbol{\vartheta})\} = \mathbb{E}\{\nabla\ell(\boldsymbol{\vartheta})\nabla\ell(\boldsymbol{\vartheta})^t\} =: \mathbb{E}\{U(\boldsymbol{\vartheta})U(\boldsymbol{\vartheta})^t\}$ is called the ***variability matrix*** and $U(\boldsymbol{\vartheta}) = \nabla_{\boldsymbol{\vartheta}}\ell(\boldsymbol{\vartheta}) =: (\partial/\partial\boldsymbol{\vartheta})\,\ell(\boldsymbol{\vartheta})$ is called the ***pairwise score function***.

Reporting the standard errors $\widehat{\boldsymbol{\vartheta}}$ requires consistent estimation of the Godambe matrix and its components. Analytical expressions for $\mathcal{H}$ and $\mathcal{J}$ are difficult to obtain in most realistic applications. So, they are usually estimated either empirically, exploiting groups of independent or almost independent data, or through simulation. The empirical (sample) estimate of $\mathcal{H}$ is given by

$$\widehat{\mathcal{H}}^E(\boldsymbol{\vartheta}) = -T^{-1}\sum_{k=1}^{T}\nabla^2\ell_k(\widehat{\boldsymbol{\vartheta}}), \tag{3.10}$$

which corresponds to minus the Hessian matrix. So, $\mathcal{H}$ can be estimated from the Hessian matrix computed in the optimization algorithm. The $\mathcal{J}$ matrix for large independent or pseudoindependent replicates T of the data may be estimated empirically by the outer product of the pairwise scores computed at $\widehat{\boldsymbol{\vartheta}}$,

$$\widehat{\mathcal{J}}^E(\boldsymbol{\vartheta}) = T^{-1}\sum_{k=1}^{T}\nabla\ell_k(\widehat{\boldsymbol{\vartheta}})\nabla\ell_k(\widehat{\boldsymbol{\vartheta}})^t. \tag{3.11}$$

When T is not sufficiently large compared to the dimension of $\widehat{\boldsymbol{\vartheta}}$, the above empirical estimates of $\mathcal{H}$ and $\mathcal{J}$ may be imprecise. On the other hand, when it is easy to simulate data from the

full model, Monte-Carlo simulations can be used as an alternative method to estimate these matrices, see, e.g., Section 5 in [99] and Section 3 in [27].

For model selection, the composite likelihood information criterion (CLIC) [100], an analog of the Akaike information criterion (AIC), can be used to choose the best-fitting model. It is defined as CLIC$= -2\left[\ell(\widehat{\boldsymbol{\vartheta}}) - \mathrm{tr}\left\{\mathcal{J}(\widehat{\boldsymbol{\vartheta}})\mathcal{H}^{-1}(\widehat{\boldsymbol{\vartheta}})\right\}\right]$, where tr(.) denotes the trace operator. Lower values of CLIC indicate a better fit. A rescaled version of CLIC has been proposed by [44] to be comparable with AIC in the independence case, i.e., CLIC$^* = (D-1)^{-1}$ CLIC when pairwise likelihood is adopted.

3.2 Pairwise likelihood statistics for testing $H_0 : a = a_0$

In order to propose a testing procedure on the mixing parameter a of a max-mixture process $\{Z(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$, we use pairwise likelihood for parametric estimation. Suppose that the parameters of a max-mixture model $\boldsymbol{\vartheta} \in \mathbb{R}^q$ is partitioned as $\boldsymbol{\vartheta} = (\boldsymbol{\gamma}, \boldsymbol{\eta}) \in \mathbb{R}^{q_1} \times \mathbb{R}^{q_2}$, with $q_1 + q_2 = q$, and that we want to test whether the null hypothesis $H_0 : \boldsymbol{\gamma} = \boldsymbol{\gamma}^*$ holds (that is, reject it, or fail to reject it). In this testing framework, the parameter $\boldsymbol{\gamma} \in \mathbb{R}^{q_1}$ is the parameter of interest, while $\boldsymbol{\eta} \in \mathbb{R}^{q_2}$ acts as a nuisance parameter. With the partition $\boldsymbol{\vartheta} = (\boldsymbol{\gamma}, \boldsymbol{\eta})$, the pairwise score function is partitioned as $U(\boldsymbol{\vartheta}) = \left(U_{\boldsymbol{\gamma}}(\boldsymbol{\vartheta}), U_{\boldsymbol{\eta}}(\boldsymbol{\vartheta})\right) =: \left((\partial/\partial\boldsymbol{\gamma})\ell(\boldsymbol{\vartheta}), (\partial/\partial\boldsymbol{\eta})\ell(\boldsymbol{\vartheta})\right)$. Also, considering the further partitions

$$\mathcal{H} = \begin{pmatrix} \mathcal{H}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} & \mathcal{H}_{\boldsymbol{\gamma}\boldsymbol{\eta}} \\ \mathcal{H}_{\boldsymbol{\eta}\boldsymbol{\gamma}} & \mathcal{H}_{\boldsymbol{\eta}\boldsymbol{\eta}} \end{pmatrix}, \quad \mathcal{H}^{-1} = \begin{pmatrix} \mathcal{H}^{\boldsymbol{\gamma}\boldsymbol{\gamma}} & \mathcal{H}^{\boldsymbol{\gamma}\boldsymbol{\eta}} \\ \mathcal{H}^{\boldsymbol{\eta}\boldsymbol{\gamma}} & \mathcal{H}^{\boldsymbol{\eta}\boldsymbol{\eta}} \end{pmatrix}, \tag{3.12}$$

where, $\mathcal{H}_{\boldsymbol{\gamma}\boldsymbol{\eta}} = \mathbb{E}\left\{-(\partial/\partial\boldsymbol{\eta})U_{\boldsymbol{\gamma}}(\boldsymbol{\vartheta})\right\}$ and so forth. Analogously for $\mathcal{G}$ and $\mathcal{G}^{-1}$. Let $\widehat{\boldsymbol{\vartheta}} = (\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\eta}})$ denote the unrestricted maximum pairwise likelihood estimator, and $\widehat{\boldsymbol{\vartheta}}^* = (\boldsymbol{\gamma}^*, \widehat{\boldsymbol{\eta}}^*)$ denote the maximum pairwise likelihood estimator under the null hypothesis, i.e., $\widehat{\boldsymbol{\eta}}^*$ is the maximum pairwise likelihood estimator of $\boldsymbol{\eta}$ when $\boldsymbol{\gamma}$ is held fixed to the value $\boldsymbol{\gamma}^*$. A two-sided pairwise likelihood ratio test may be based on the statistic

$$LR = 2\left\{\ell(\widehat{\boldsymbol{\vartheta}}) - \ell(\widehat{\boldsymbol{\vartheta}}^*)\right\} \xrightarrow{\mathcal{D}} \sum_{j=1}^{q_1} c_j W_j, \tag{3.13}$$

where the W_j's are independent χ^2_1 random variables, and the c_j's are the eigenvalues of the matrix $\{\mathcal{H}^{\boldsymbol{\gamma}\boldsymbol{\gamma}}\}^{-1}\mathcal{G}^{\boldsymbol{\gamma}\boldsymbol{\gamma}}$ evaluated under the null hypothesis. This follows from Theorem 3.1 in [70]. The form of asymptotic distribution in (3.13) is mainly due to the failure of the second Bartlett identity, which refers to the equality between the expected negative Hessian of the log-likelihood and the covariance of the score function, i.e., $\mathcal{H}(\boldsymbol{\vartheta}) \neq \mathcal{J}(\boldsymbol{\vartheta})$, and the problem is that generally the distribution of $\sum_{j=1}^{q_1} c_j W_j$ is not known exactly when $q_1 > 1$. The other pairwise likelihood versions of Wald and score statistics for testing $H_0 : \boldsymbol{\gamma} = \boldsymbol{\gamma}^*$ are easily constructed, and they have the usual asymptotic $\chi^2_{q_1}$. The Wald-type statistic for the parameter of interest is

$$W = (\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*)^t \left\{\mathcal{G}^{\boldsymbol{\gamma}\boldsymbol{\gamma}}\left(\widehat{\boldsymbol{\vartheta}}^*\right)\right\}^{-1} (\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*)\,. \tag{3.14}$$

While the score-type statistic is

$$S = U_{\boldsymbol{\gamma}}\left(\widehat{\boldsymbol{\vartheta}}^*\right)^t H^{\boldsymbol{\gamma}\boldsymbol{\gamma}}\left(\widehat{\boldsymbol{\vartheta}}^*\right)\left\{G^{\boldsymbol{\gamma}\boldsymbol{\gamma}}\left(\widehat{\boldsymbol{\vartheta}}^*\right)\right\}^{-1}\boldsymbol{\mathcal{H}}^{\boldsymbol{\gamma}\boldsymbol{\gamma}}\left(\widehat{\boldsymbol{\vartheta}}^*\right)U_{\boldsymbol{\gamma}}\left(\widehat{\boldsymbol{\vartheta}}^*\right). \tag{3.15}$$

However, as is well known, these two statistics suffer from practical limitations: Wald-type statistics are not invariant by reparametrization, while score-type statistics suffer from numerical instability. In this respect, a likelihood ratio-type statistic would be more appealing since it often has better finite-sample performance, see, e.g., [83]. So, various adjustments have been proposed in the literature to recover an asymptotic chi-squared distribution $\chi^2_{q_1}$ when $q_1 > 1$. For instance, [88] considered $q_1 LR/\sum_{j=1}^{q_1} c_j W_j$ as a $\chi^2_{q_1}$ random variable. While [28] proposed to adjust the composite likelihood through a vertical scaling, instead of adjusting the asymptotic likelihood ratio statistic ditribution, in the following way

$$LR_I = \frac{(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})^t\left\{\boldsymbol{\mathcal{G}}^{\boldsymbol{\gamma}\boldsymbol{\gamma}}\left(\widehat{\boldsymbol{\vartheta}}\right)\right\}^{-1}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})}{(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})^t\,\boldsymbol{\mathcal{H}}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\left(\widehat{\boldsymbol{\vartheta}}\right)(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})}LR. \tag{3.16}$$

Above, recall that $\boldsymbol{\mathcal{H}}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}$ and $\boldsymbol{\mathcal{G}}^{\boldsymbol{\gamma}\boldsymbol{\gamma}}$ denote, respectively, the $q_1 \times q_1$ submatrices of $\boldsymbol{\mathcal{H}}$ and the inverse of $\boldsymbol{\mathcal{G}}$ pertaining to $\boldsymbol{\gamma}$. The resulting statistic LR_I has asymptotic $\chi^2_{q_1}$ distribution but it lacks invariance under reparameterization. To get around this issue, [83] proposed another adjustment based on the composite score function, i.e.,

$$LR_{II} = \frac{S}{U_{\boldsymbol{\gamma}}\left(\widehat{\boldsymbol{\vartheta}}^*\right)^t\boldsymbol{\mathcal{H}}^{\boldsymbol{\gamma}\boldsymbol{\gamma}}\left(\widehat{\boldsymbol{\vartheta}}^*\right)U_{\boldsymbol{\gamma}}\left(\widehat{\boldsymbol{\vartheta}}^*\right)}LR, \tag{3.17}$$

again LR_{II} has an asymptotic $\chi^2_{q_1}$ null distribution. A regularity condition underlying these adjustments is that the parameter $\boldsymbol{\gamma}^*$ is interior to its parameter space, so that $U(\widehat{\boldsymbol{\vartheta}}) = 0$. However, this condition is usually not satisfied in the case where $\boldsymbol{\gamma}^*$ lies on the boundary. So, existing adjustments may not yield an asymptotic $\chi^2_{q_1}$. In addition, the limiting distribution for "*regular*" boundary problems (i.e., the same number of nuisance parameters under H_0 and H_1) is a complicated mixture of weighted χ^2 distributions. The computation of the mixing weights is challenging. It depends on a partition of the parameter space, where the partition relies on the decomposition of the $\boldsymbol{\mathcal{G}}$ matrix, see [29, 30]. On the other hand, simulation-based techniques could be used to approximate the quantiles of the limit $\sum_{j=1}^{q_1} c_j W_j$. In the present chapter, our purpose is to test the hypothesis $H_0 : a = a_0$ versus $H_1 : a \neq a_0$, for some specified value $a_0 \in [0,1]$. Therefore, we propose using following two pairwise likelihood-based statistics:

(i) the pairwise likelihood ratio statistic, which can be easily deduced from (3.13) with $\boldsymbol{\gamma} = a$ and $q_1 = 1$, that is,

$$LR_a = c^{-1}LR \xrightarrow{\mathcal{D}} \chi^2_1, \tag{3.18}$$

here the constant $c = \{\boldsymbol{\mathcal{H}}^{aa}\}^{-1}\boldsymbol{\mathcal{G}}^{aa}$, where $\boldsymbol{\mathcal{H}}^{aa}(\boldsymbol{\vartheta})$ and $\boldsymbol{\mathcal{G}}^{aa}(\boldsymbol{\vartheta})$ denote, respectively, the 1×1 submatrices of the inverse of $\boldsymbol{\mathcal{H}}(\boldsymbol{\vartheta})$ and $\boldsymbol{\mathcal{G}}(\boldsymbol{\vartheta})$ with elements corresponding to a,

(ii) the Z-test statistic which is straightforwardly derived from the central limit theorem for maximum pairwise log-likelihood estimator $\widehat{\boldsymbol{\vartheta}}$, recall that $\widehat{\boldsymbol{\vartheta}} \xrightarrow{\mathcal{D}} \mathcal{N}_q\left(\boldsymbol{\vartheta}, \mathcal{G}^{-1}(\boldsymbol{\vartheta})\right)$ when $T \to \infty$, i.e.,

$$Z_a = \frac{\widehat{a} - a}{\sqrt{\mathcal{G}^{aa}(\widehat{\boldsymbol{\vartheta}})}} \xrightarrow{\mathcal{D}} N(0,1), \tag{3.19}$$

where $\mathcal{G}^{aa}(\widehat{\boldsymbol{\vartheta}})$ denotes a 1×1 submatrix of the inverse of $\mathcal{G}(\widehat{\boldsymbol{\vartheta}})$ pertaining to a.

Testing at boundary points $H_0 : a = 1$ (asymptotic dependence) or $H_0 : a = 0$ (asymptotic independence) is nonstandard since there are additional nuisance parameters that are present only under the alternative hypothesis. [38, 39] studied the situations were the nuisance parameters are present only under H_1. To the best of our knowledge, there is no systematic theoretical study dealing with "*non-regular*" boundary problems with an unequal number of nuisance parameters under H_0 and H_1. So, much more theoretical research has to be undertaken to determine the limiting distribution for this case. For this reason, we apply our LR_a test at points close to the boundaries, i.e., $a_0 = 0.99$ or $a_0 = 0.01$.

Remark 3.2. *If the mixing parameter a is interior to its parameter space (i.e., $0 < a < 1$), one may tests the global hypothesis $H_0 : \boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$ using the statistics in (3.2), (3.17), (3.14) and (3.15). All of them having the usual χ^2_q, where q is the of dimension of $\boldsymbol{\vartheta}$.*

3.3 Simulation study

This section studies the performance of the proposed test statistics LR_a and Z_a via several simulation studies. To that aim, we consider testing $H_0 : a = a_0$ against $H_1 : a \neq a_0$, where a_0 varies from 0.01 to 0.99 by steps of 0.01. Let $\boldsymbol{\vartheta}$ denote the vector gathering the model parameters. Throughout this section, we consider the following two max-mixture models:

(i) M_1: is a max-mixture model (2.34) which combines TEG and inverted TEG processes (recall Table 2.1) with isotropic exponential correlation functions $\rho(\boldsymbol{h}) = \exp\left(-\|\boldsymbol{h}\|/\phi\right)$, for some range parameter $\phi > 0$. The compact set $\mathcal{A}$ is a disk of fixed radius r. The model parameters are $\boldsymbol{\vartheta} = (\phi_X, r_X, a, \phi_Y, r_Y)^t$. Recall that $\|\boldsymbol{h}\| = \|\boldsymbol{s}_1 - \boldsymbol{s}_2\| \geq 0$ denotes the Euclidean distance between sites $\boldsymbol{s}_1, \boldsymbol{s}_2$. With this construction of model M_1, the pairs of sites separated by a distance $\|\boldsymbol{h}\|$ smaller than $2r_X$ (respectively, greater than $2r_Y$) are asymptotically dependent (respectively, exactly independent). Whereas, at intermediate distances the pairs exhibit asymptotic independence, recall Example 2.2.

(ii) M_2: is a max-mixture model, which combines a TEG process as in M_1 and an inverted BR process with a stationary isotropic semivariogram $\gamma_Y(\boldsymbol{h}) = \left(\|\boldsymbol{h}\|/\phi_Y\right)^{\kappa_Y}$, the range $\phi_Y > 0$ and smoothness $\kappa_Y \in (0, 2]$. The model parameters are $\boldsymbol{\vartheta} = (\phi_X, r_X, a, \phi_Y, \kappa_Y)^t$. In this model, the pairwise max-mixture processes $(Z(\boldsymbol{s}), Z(\boldsymbol{s}+\boldsymbol{h}))$ are asymptotically dependent at distance $\|\boldsymbol{h}\|$ up to $2r_X$ and asymptotically independent for all $\|\boldsymbol{h}\| > 2r_X$, see Figure 3.1.

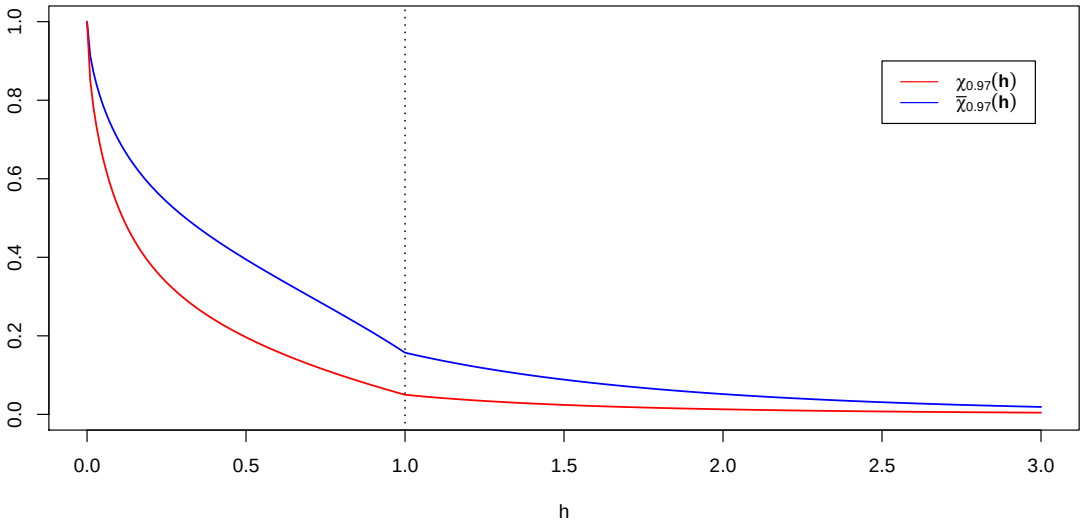


Figure 3.1: Theoretical behavior of $\chi_{0.97}(\boldsymbol{h})$ and $\bar{\chi}_{0.97}(\boldsymbol{h})$ derived from model M_2 as functions of the distance $h = \|\boldsymbol{h}\|$. We set $\phi_X = 0.1$, $r_X = 0.5$, $\phi_Y = 1$, $\kappa_Y = 1.5$ and $a = 0.5$. The dotted vertical line represents the diameter of the disk in the TEG process.

3.3.1 Results on model M_1

Estimation performance

The pairwise likelihood contribution (3.8) is used for estimation, where the threshold u is taken corresponding to the 0.9 empirical quantiles at each site. The weights are ignored, i.e., $\omega_{ij} = 1$. Indeed, choosing equal weights in the pairwise likelihood may not be ideal both in terms of computational efficiency and statistical efficiency, see, e.g., [26]. Nevertheless, the equally weighted pairwise likelihood has been adopted in many applications, see, e.g., [18, 44]. The matrices $\mathcal{H}(\boldsymbol{\vartheta})$ and $\mathcal{J}(\boldsymbol{\vartheta})$ are estimated empirically. The fitting is done by the code which was used by [11] with some appropriate modifications. Note that the pairwise likelihood function was coded in C and used as a function in R in order to reduce computational burden.

We start by presenting the results on model M_1. To assess the quality of the pairwise likelihood estimation procedure, a simulation study has been carried out. We simulate $T = 1000$ independent copies of the model M_1 at $D = 50$ sites randomly and uniformly distributed in the square $[0,3]^2$. We consider several mixing parameters $a \in \{0, 0.25, 0, 50, 0.75$ and $1\}$. The parameters used are $\phi_X = 0.10, r_X = 0.25, \phi_Y = 0.75, r_Y = 1.20$. Each experiment was repeated $J = 200$ times to obtain boxplots of the estimated parameters and compute performance metrics, i.e., the mean estimate, the root mean squared error (RMSE), and the mean absolute error (MAE). Denote by $\widehat{\boldsymbol{\vartheta}}_j$ the jth estimation, then

$$\text{RMSE} = \left[J^{-1} \sum_{j=1}^{J} (\widehat{\boldsymbol{\vartheta}}_j - \boldsymbol{\vartheta})^2 \right]^{1/2} \quad \text{and} \quad \text{MAE} = J^{-1} \sum_{j=1}^{J} |\widehat{\boldsymbol{\vartheta}}_j - \boldsymbol{\vartheta}| \tag{3.20}$$

The boxplots of the errors of the estimated parameters on the J samples are displayed in

Figure 3.2. Table 3.1 reports the mean estimate, RMSE, and MAE of the estimated parameters. Generally, the estimation procedure appears to work well although the variability of some estimates was relatively large, especially for the asymptotic independence parameters $\{\phi_Y, r_Y\}$. This probably stems from the fact that asymptotic independence is difficult to estimate, see [5, 45]. Moreover, we observe that contrary to asymptotic independence parameters $\{\phi_Y, r_Y\}$, the estimation of asymptotic dependence parameters $\{\phi_X, r_X\}$ becomes more accurate as the mixing parameter value increases (the RMSE and MAE are lower). Using (3.9) instead of (3.8) as censored pairwise likelihood leads to similar results, see Figure 3.3 and Figure 3.4.

Testing performance with true a being a non-boundary point in $(0, 1)$

We evaluate the performance of the proposed pairwise likelihood test statistics (3.18) and (3.19). We test whether the null hypothesis $H_0 : a = a_0$ holds for all a_0 values in the set $\{0.01, 0.02, \ldots, 0.99\}$. Here we examine three cases with a true mixing parameter $a \in \{0.25, 0.5, 0.75\}$, based on $J = 150$ simulation replicates from $T = 1000$ independent copies simulated at $D = 50$ randomly and uniformly sampled locations in the square $[0, 3]^2$ from the model M_1 with parameters $\{\phi_X = 0.10, r_X = 0.25, \phi_Y = 0.75, \text{ and } r_Y = 1.2\}$. We compute the empirical probabilities of rejecting H_0 denoted by P. In other words, P represents the power of the test when H_0 is false (i.e., the proportion of null hypotheses rejected). Decisions obtained at three significance levels $\alpha \in \{0.01, 0.05, 0.10\}$.

For both statistics LR_a and Z_a, our results are summarized in Figure 3.5 and Table 3.2. Overall, the results show a reasonable performance for the two statistics. In particular, both tests seem to be unbiased (the power is greater than the sensitivity level α). Moreover, the type I errors are close to the nominal level size $\alpha = 0.01, 0.05$ and 0.1 (recall that the type I error is the probability of rejecting a true null hypothesis H_0). Consequently, these pairwise likelihood statistics can control the type I error rate α. We also remark that the LR_a test seems to be more powerful than the Z_a one as can be seen on Table 3.2.

Testing performance with true a being a boundary point (i.e., $a = 0$ or $a = 1$)

The pairwise likelihood ratio test statistics LR_a cannot be applied directly for testing $H_0 : a = a_0$ with a_0 being the boundary point (i.e., $a_0 = 0$ or $a_0 = 1$) since there are additional nuisance parameters which are present only under the alternative hypothesis H_1. To circumvent this, we apply our statistics at some points very close to the boundaries, i.e., $a_0 = 0.01$ or $a_0 = 0.99$. For this purpose, we simulate $T = 1000$ independent observations of a TEG process (asymptotic dependence case) (respectively, an inverted TEG process (asymptotic independence case)) with parameters $\{\phi_X = 0.10, r_X = 0.25\}$ (respectively, $\{\phi_Y = 0.75, r_Y = 1.2\}$) at $D = 50$ sites uniformly generated in $[0, 3]^2$. We repeat this experiment $J = 150$ times. For both cases, we test whether the null hypothesis $H_0 : a = a_0$ holds, where a_0 varies from 0.01 to 0.99 by steps of 0.01. Here, the global model M_1 that mixes both processes (TEG and inverted TEG) is used to perform testing for each step on the basis of the simulated data from TEG (asymptotic dependence case) and inverted TEG (asymptotic independence case). Similarly, as before, the empirical probabilities (P) are computed for both cases. Decisions are obtained at three significance levels

$\alpha \in \{0.01, 0.05, 0.10\}$.

Figure 3.6 and Table 3.3 compare the empirical probabilities (P) obtained for both statistics LR_a and Z_a. We note that the performance of the two statistics is satisfactory. As anticipated, the power to reject asymptotic dependence, i.e., $H_0 : a = 1$ (respectively, asymptotic independence, i.e., $H_0 : a = 0$) improves as $a_0 \to 0$ (respectively, $a_0 \to 1$). Moreover, we observe that the values of P at a_0 close to the boundaries, i.e., $a_0 = 0.01$ or $a_0 = 0.99$ are close to the nominal level size $\alpha = 0.01, 0.05, 0.1$. Therefore, these pairwise likelihood statistics can control the type I error rate α. So, these statistics can provide a strong indication for both asymptotic dependence and independence. Let us remark that near the boudaries ($a = 0$ or $a = 1$), the LR_a test seems to be less powerful than the Z_a one (contrary to what we observed when a is interior to its parameter space).

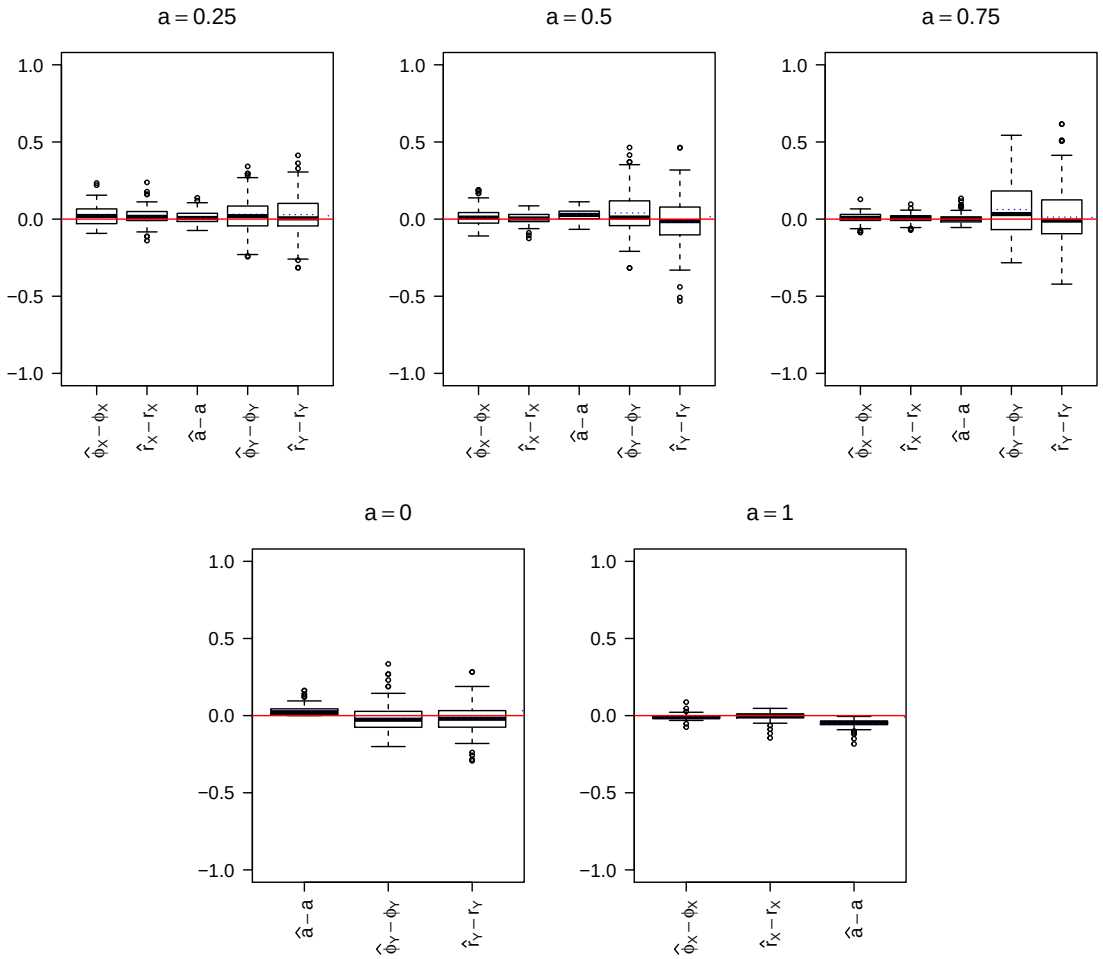


Figure 3.2: Boxplots display $\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}$ of the estimated parameters using pairwise likelihood contribution (3.8) based on 200 simulation replicates from 1000 independent copies of the model M_1 with parameters $\phi_X = 0.10, r_X = 0.25, \phi_Y = 0.75, r_Y = 1.20$, and mixture parameter $a \in \{0, 0.25, 0.5, 0.75, 1\}$. Blue dotted/red solid horizontal lines show the estimated errors means/zero value.

Table 3.1: Performance of the estimation for 200 simulated M_1 models with parameters $\phi_X = 0.10, r_X = 0.25, \phi_Y = 0.75, r_Y = 1.20$, and various mixing parameter values $a \in \{0, 0.25, 0.5, 0.75, 1\}$. The mean estimate, RMSE, and MAE of the estimated paramters.

True	Performance metrics		
	Mean estimate	RMSE	MAE
$a = 0$	0.031	0.043	0.031
$\phi_Y = 0.75$	0.731	0.092	0.072
$r_Y = 1.2$	1.183	0.095	0.073
$\phi_X = 0.1$	0.121	0.064	0.053
$r_X = 0.25$	0.269	0.053	0.039
$a = 0.25$	0.261	0.037	0.029
$\phi_Y = 0.75$	0.782	0.115	0.088
$r_Y = 1.2$	1.229	0.125	0.091
$\phi_X = 0.1$	0.115	0.058	0.044
$r_X = 0.25$	0.257	0.036	0.028
$a = 0.5$	0.529	0.043	0.032
$\phi_Y = 0.75$	0.789	0.143	0.105
$r_Y = 1.2$	1.183	0.168	0.126
$\phi_X = 0.1$	0.111	0.030	0.023
$r_X = 0.25$	0.255	0.027	0.021
$a = 0.75$	0.751	0.029	0.020
$\phi_Y = 0.75$	0.812	0.186	0.144
$r_Y = 1.2$	1.213	0.203	0.153
$\phi_X = 0.1$	0.093	0.018	0.014
$r_X = 0.25$	0.231	0.024	0.017
$a = 1$	0.951	0.053	0.049

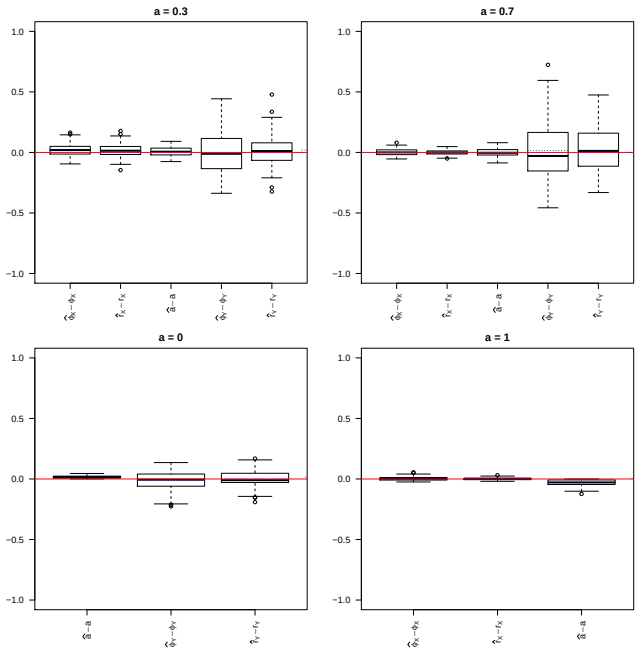


Figure 3.3: Boxplots display $\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}$ of the estimated parameters using pairwise likelihood contribution (3.9) based on 100 simulation replicates from 1000 independent copies of the model M_1 with parameters $\phi_X = 0.10, r_X = 0.25, \phi_Y = 0.75, r_Y = 1.20$, and mixture parameter $a \in \{0, 0.3, 0.7, 1\}$. Blue dotted/red solid horizontal lines show the estimated errors means/zero value.

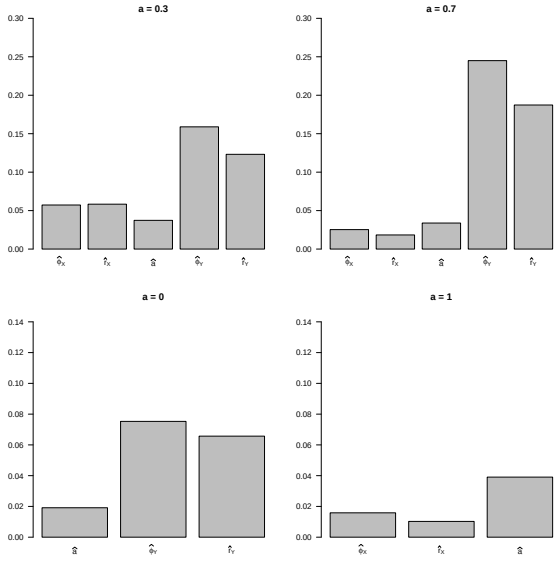


Figure 3.4: Barplots display RMSE of $\widehat{\boldsymbol{\vartheta}} = \{\widehat{\phi}_X, \widehat{r}_X, \widehat{a}, \widehat{\phi}_Y, \widehat{r}_Y\}$ for each estimated parameter using pairwise likelihood contribution (3.9) based on 100 simulation replicates from 1000 independent copies of the model M_1 with parameters $\phi_X = 0.10, r_X = 0.25, \phi_Y = 0.75, r_Y = 1.20$, and mixture parameter $a \in \{0, 0.3, 0.7, 1\}$.

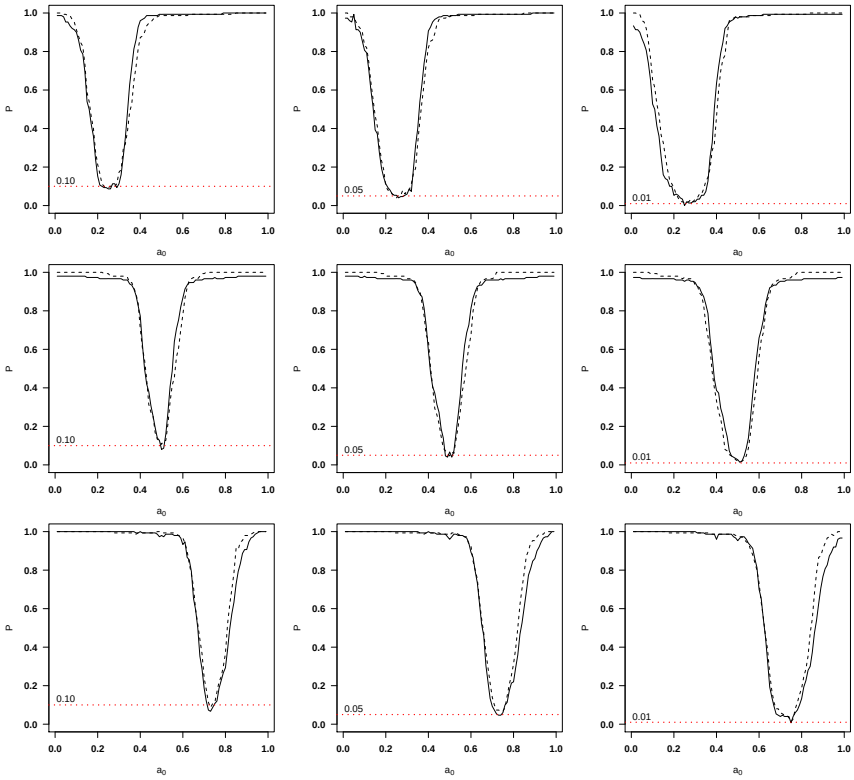


Figure 3.5: Empirical probabilities (P) based on 150 replicates simulation study of the model M_1 with $\phi_X = 0.10, r_X = 0.25, \phi_Y = 0.75, r_Y = 1.20$. The mixing coefficients (top row: $a = 0.25$, middle row: $a = 0.50$ and bottom row: $a = 0.75$). (Solid line: Z_a test and dashed line: LR_a test). Horizontal red dotted lines show the nominal level size $\alpha = 0.1, 0.05, 0.01$.

Table 3.2: Empirical probabilities (P) for testing $H_0 : a = a_0$ against $H_1 : a \neq a_0$ based on 150 simulation replicates from 1000 independent copies of the model M_1 with parameters $\phi_X = 0.10,, r_X = 0.25, \phi_Y = 0.75, r_Y = 1.2$ and mixing coefficients $a \in \{0.25, 0.5, 0.75\}$ at three significance levels $\alpha \in \{0.01, 0.05, 0.10\}$.

	True $a = 0.25$					
	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
a_0	LR_a	Z_a	LR_a	Z_a	LR_a	Z_a
0.05	0.960	0.867	0.980	0.993	0.993	0.953
0.10	0.653	0.527	0.833	0.807	0.913	0.900
0.15	0.327	0.160	0.467	0.393	0.587	0.540
0.20	0.093	0.053	0.100	0.113	0.213	0.153
0.25	0.013	0.000	0.040	0.053	0.093	0.087
0.30	0.033	0.020	0.080	0.060	0.173	0.113
0.35	0.127	0.073	0.327	0.400	0.487	0.627
0.40	0.460	0.633	0.820	0.907	0.873	0.960
0.50	0.987	0.973	0.987	0.980	0.987	0.993
0.80	0.993	0.993	0.993	0.993	0.993	1.000
	True $a = 0.5$					
	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
a_0	LR_a	Z_a	LR_a	Z_a	LR_a	Z_a
0.10	0.993	0.967	1.000	0.980	1.000	0.980
0.25	0.973	0.953	0.980	0.967	0.993	0.967
0.35	0.727	0.833	0.913	0.907	0.947	0.940
0.40	0.333	0.387	0.647	0.593	0.707	0.773
0.45	0.073	0.147	0.213	0.273	0.273	0.307
0.50	0.020	0.027	0.047	0.067	0.107	0.080
0.55	0.087	0.127	0.280	0.360	0.413	0.487
0.60	0.560	0.660	0.673	0.813	0.780	0.867
0.65	0.887	0.920	0.960	0.933	0.973	0.947
0.75	0.973	0.960	1.000	0.960	1.000	0.973
	True $a = 0.75$					
	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
a_0	LR_a	Z_a	LR_a	Z_a	LR_a	Z_a
0.40	0.987	0.960	0.993	1.000	0.993	1.000
0.50	0.973	0.953	0.980	0.967	0.993	0.980
0.60	0.733	0.700	0.927	0.893	0.967	0.933
0.65	0.260	0.193	0.547	0.520	0.713	0.653
0.70	0.060	0.040	0.140	0.093	0.240	0.187
0.75	0.013	0.007	0.067	0.060	0.113	0.107
0.80	0.187	0.133	0.313	0.220	0.367	0.293
0.85	0.547	0.373	0.787	0.607	0.913	0.680
0.90	0.913	0.767	0.953	0.860	0.980	0.907
0.95	0.980	0.920	1.000	0.973	1.000	0.993

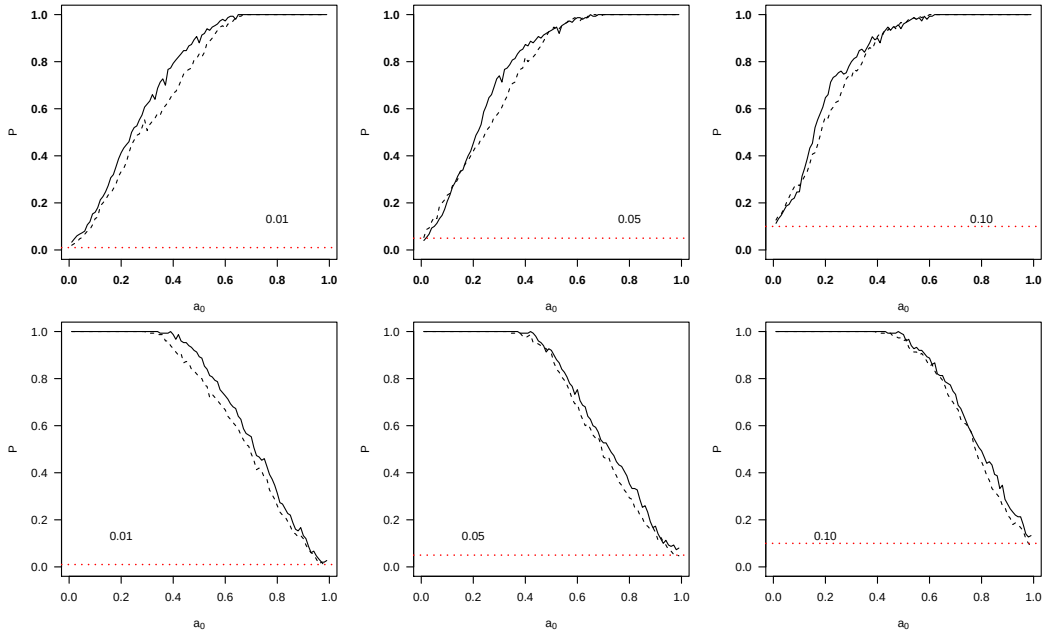


Figure 3.6: Top row: empirical probabilities (P) for testing $H_0 : a = 0$ (asymptotic independence case) based on 150 simulation replicates from 1000 independent copies from an inverted TEG process with parameters $\{\phi_Y = 0.75,\ r_Y = 1.2\}$). Bottom row: empirical probabilities (P) for testing $H_0 : a = 1$ (asymptotic dependence case) based on 150 simulation replicates from 1000 independent copies from a TEG process with parameters $\{\phi_X = 0.1,\ r_X = 0.25\}$). (Solid line: Z_a test and dashed line: LR_a test). Horizontal red dotted lines show the nominal level size $\alpha = 0.1, 0.05, 0.01$.

Table 3.3: Empirical probabilities (P) for testing $H_0 : a = a_0$ against $H_1 : a \neq a_0$ based on 150 simulation replicates from 1000 independent copies of an inverted TEG process (true $a = 0$) with parameters $\{\phi_Y = 0.75, r_Y = 1.2\}$ (left table) and a TEG process (true $a = 1$) with parameters $\{\phi_X = 0.10, r_X = 0.25\}$ (right table). Significance levels $\alpha \in \{0.01, 0.05, 0.10\}$.

	True $a = 0$					
	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
a_0	LR_a	Z_a	LR_a	Z_a	LR_a	Z_a
0.01	0.020	0.033	0.053	0.040	0.127	0.113
0.05	0.053	0.073	0.140	0.100	0.193	0.187
0.10	0.133	0.160	0.233	0.207	0.273	0.247
0.20	0.333	0.413	0.420	0.453	0.567	0.647
0.30	0.507	0.620	0.587	0.740	0.727	0.800
0.40	0.667	0.793	0.820	0.873	0.907	0.893
0.50	0.833	0.880	0.933	0.940	0.960	0.967
0.60	0.933	0.973	0.980	0.987	1.000	0.993
0.70	1.000	1.000	1.000	1.000	1.000	1.000

	True $a = 1$					
	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
a_0	LR_a	Z_a	LR_a	Z_a	LR_a	Z_a
0.30	0.993	1.000	1.000	1.000	1.000	1.000
0.40	0.927	0.987	0.973	0.993	0.933	1.000
0.50	0.807	0.893	0.907	0.920	0.960	0.987
0.60	0.667	0.727	0.693	0.753	0.853	0.887
0.70	0.480	0.553	0.467	0.527	0.680	0.733
0.80	0.260	0.313	0.293	0.353	0.447	0.493
0.90	0.120	0.133	0.147	0.173	0.220	0.267
0.95	0.033	0.047	0.073	0.107	0.167	0.213
0.99	0.000	0.027	0.047	0.080	0.087	0.133

3.3.2 Results on model M_2

Using the same procedure introduced in Section 3.3.1, we carry out another simulation study based on model M_2 with $\phi_X = 0.1$, $r_X = 0.5$, $\phi_Y = 1$ and $\kappa_Y = 1.5$. The obtained results are summarized in Figures 3.7, 3.8 and 3.9. Figures 3.7 and 3.8 show the performance of the resulting pairwise likelihood estimates when $a \in \{0, 0.3, 0.7, 1\}$, whereas Figure 3.9 displays the proportion of null hypotheses rejected (i.e., the power curves when the corresponding null hypotheses are false), estimated using the 150 simulations and plotted as a function of $a_0 \in (0, 1)$. Similarly to model M_1, we obtain an equally satisfying results.

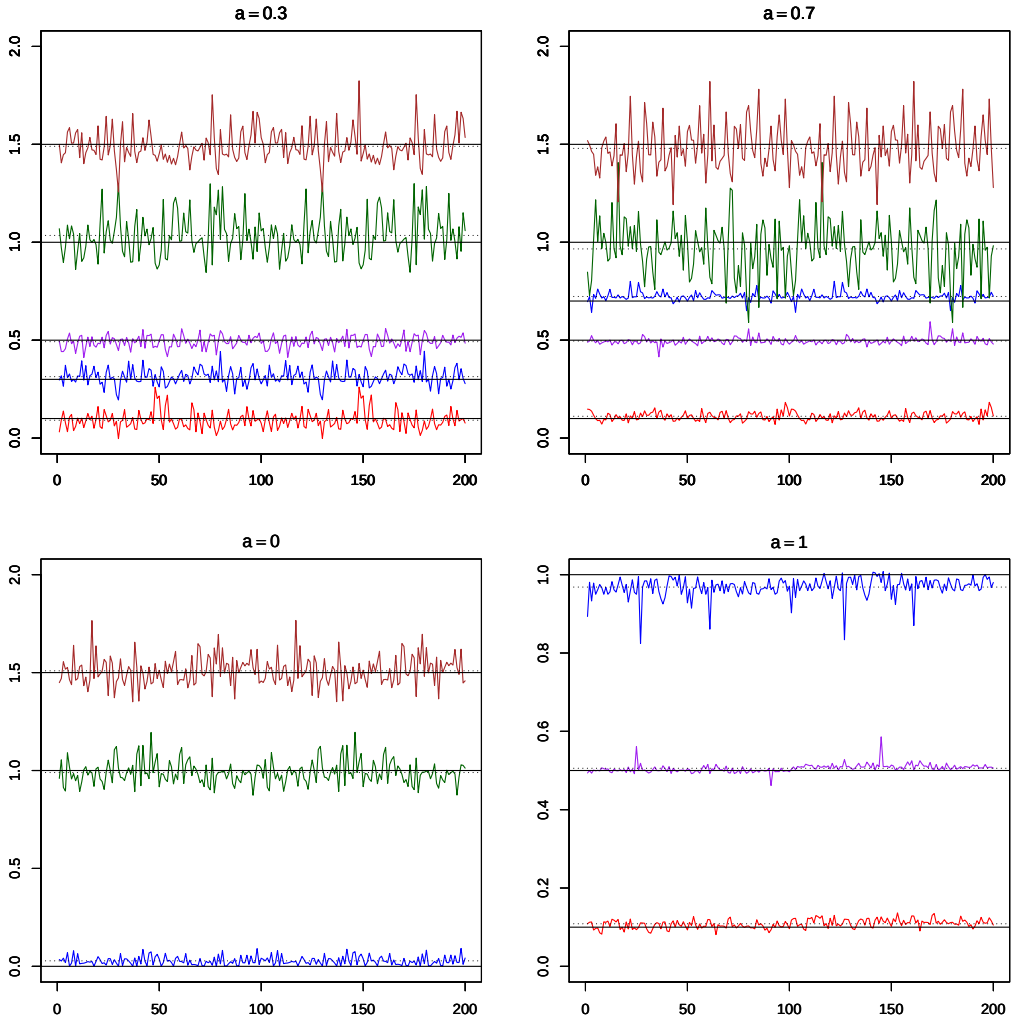


Figure 3.7: Plots display $\widehat{\boldsymbol{\vartheta}}$ of the estimated parameters using pairwise likelihood contribution (3.8) based on 200 simulation replicates from 1000 independent copies of the model M_2 with parameters $\phi_X = 0.10, r_X = 0.5, \phi_Y = 1, \kappa_Y = 1.5$ and mixture parameter $a \in \{0, 0.3, 0.7, 1\}$. Black dotted/solid horizontal lines show the estimated means/true values. $\widehat{\phi_X}$: red, $\widehat{r_X}$: purple, $\widehat{a}$: blue, $\widehat{\phi_Y}$: green and $\widehat{\kappa_Y}$: brown.

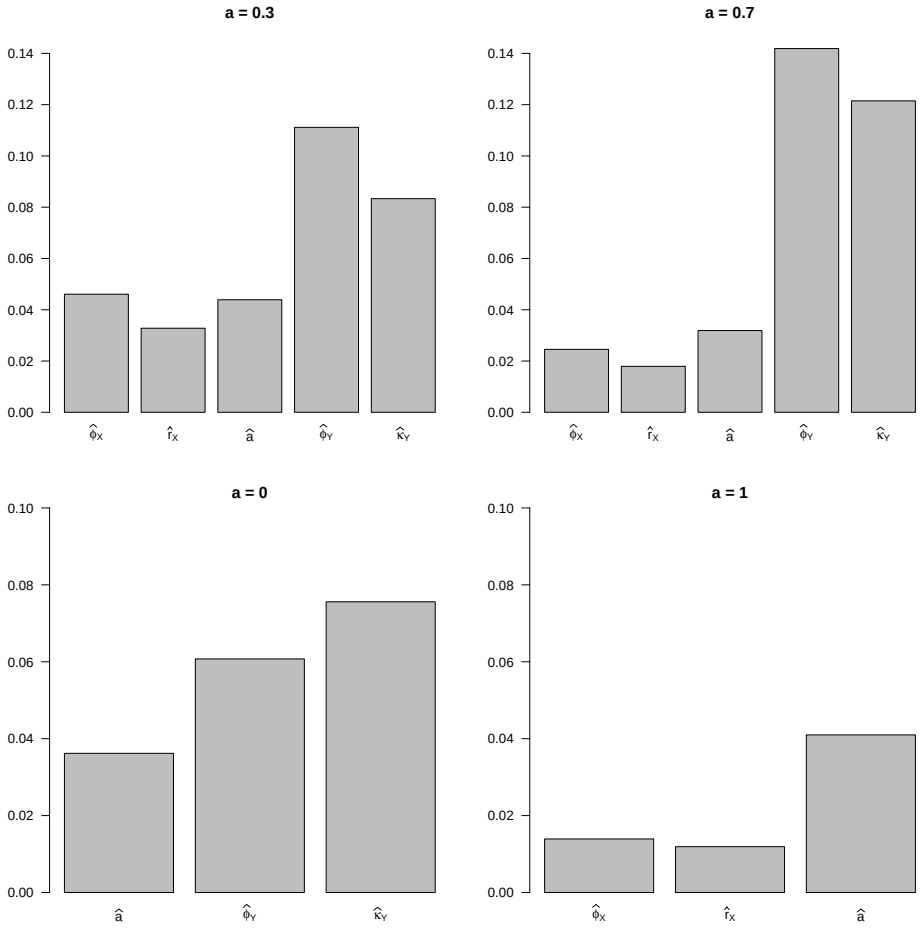


Figure 3.8: Barplots display RMSE of $\widehat{\boldsymbol{\vartheta}} = \{\widehat{\phi}_X, \widehat{r}_X, \widehat{a}, \widehat{\phi}_Y, \widehat{\kappa}_Y\}$ for each estimated parameter using pairwise likelihood contribution (3.8) based on 200 simulation replicates from 1000 independent copies of the model M_2 with parameters $\phi_X = 0.10, r_X = 0.5, \phi_Y = 1, \kappa_Y = 1.5$, and mixture parameter $a \in \{0, 0.3, 0.7, 1\}$.

3.4 Rainfall data example: Australian Rainfall data

3.4.1 Description of the dataset

The dataset analyzed in this section is composed of daily rainfall measurements (in millimeters), totals for the 24 h (measured at 9 am), recorded from 1972 to 2014 at 38 monitoring stations on the south east coast of Australia, where it has a temperate climate, oceanic, or Mediterranean. The locations of the monitoring stations are shown in Figure 3.10. Being in the Southern Hemisphere, Australia has, of course, reversed seasons in comparison with Europe or North America. Only the winter period (April−September) is considered. The entire dataset comprises $183 \times 43 \times 38 = 299,022$ measurements, with 7869 data points per station, where 183 is the number of days in the wet season considered in one particular year. The altitude of the stations varies from 4 to 552 meters above mean sea level. The stations are separated by distances ranging from 34 km to 1383 km. Australian rainfall data are freely available on http://www.

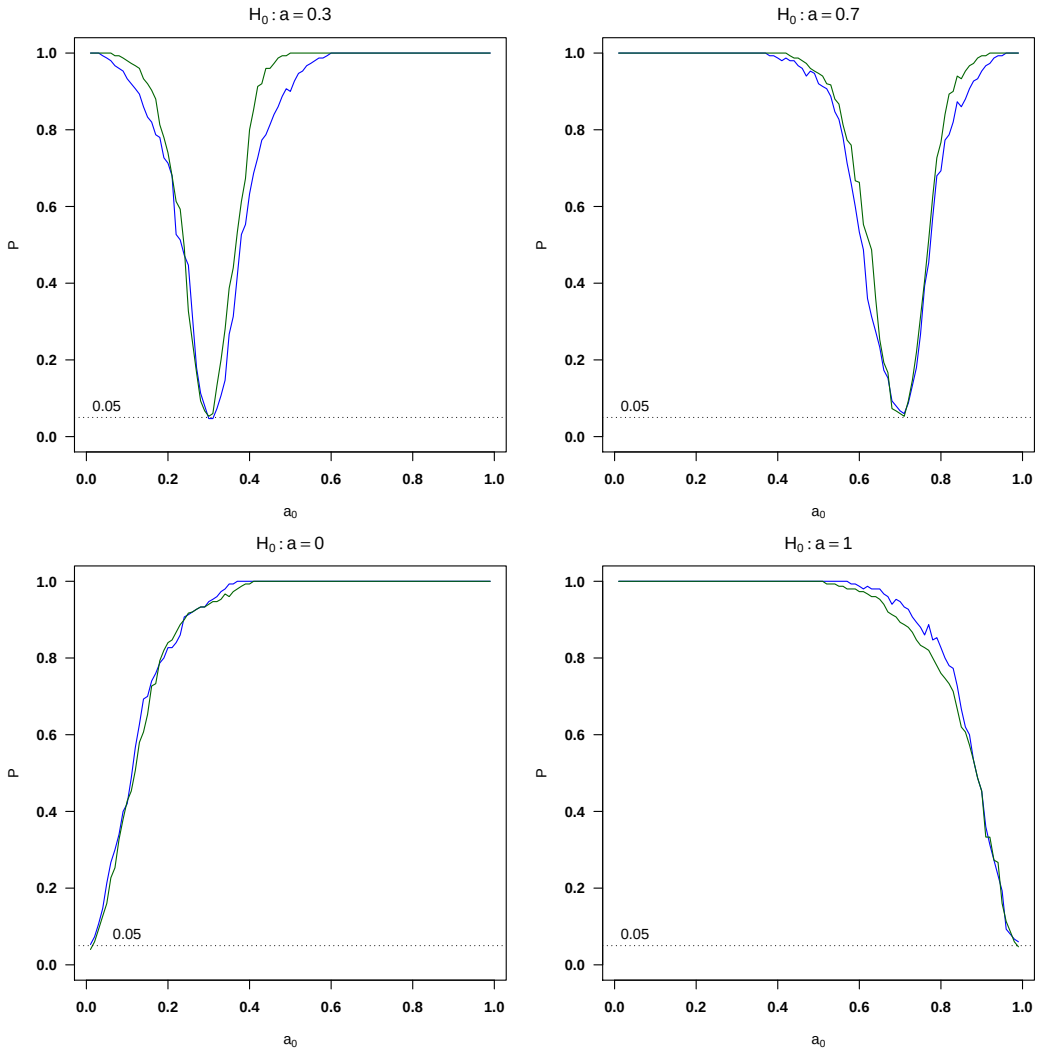


Figure 3.9: Empirical probabilities (P) for testing $H_0 : a = a_0$, plotted against a_0, where a_0 varies from 0.01 to 0.99 by steps of 0.01. P computed based on 150 simulations from $T = 1000$ independent copies simulated at $D = 50$ randomly and uniformly sampled locations in the square $[0, 3]^2$. Top left: data simulated from the model M_2 with parameters $\phi_X = 0.10$, $r_X = 0.5$, $a = 0.3$, $\phi_Y = 1$, and $\kappa_Y = 1.5$. Top right: as in top left with $a = 0.7$. Bottom left: data simulated from the asymptotically independent component of M_2 (i.e., the inverted BR process) with parameters $\phi_Y = 1$, $\kappa_Y = 1.5$. Bottom right: data simulated from the max-stable component of M_2 (i.e., the TEG process) with parameters $\phi_X = 0.1$, $r_X = 0.5$. Blue/green lines show results from Z_a/LR_a tests. The horizontal dotted line at 0.05 shows the nominal level α used for the tests.

bom.gov.au/climate/data/. This dataset and more details about the monitoring stations can be found in the online supplementary material: http://math.univ-lyon1.fr/homes-www/abuawwad/Australia_Rainfalldata/.

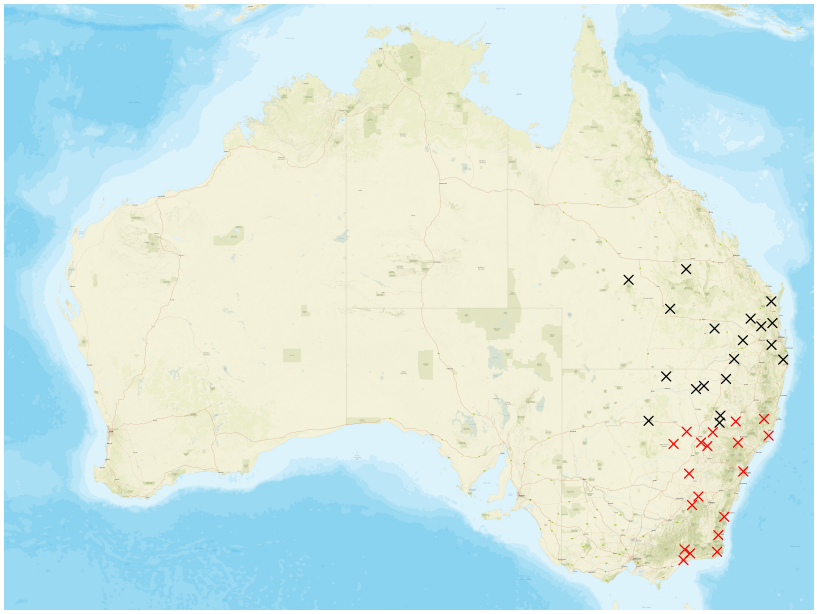


Figure 3.10: Geographical locations of the 38 meteorological stations located in the East of Australia. Black/Red cross-symbols show the stations in A/B zones. Stations in zone A are used for model inference, and the other stations in zone B are put aside for hypothesis testing.

3.4.2 Exploratory analysis

For given data, it might be important to check whether the extremes in space have directional dependence. There exist various graphical diagnostic approaches in the literature. They are based on pairwise summary measures, recall Section 2.2.2. For example, [101] explored possible anisotropy based on the estimates of the coefficient of tail dependence function $\eta(\boldsymbol{h})$. [11] proposed to use the summary measures $\chi_u(\boldsymbol{h})$ and $\bar{\chi}_u(\boldsymbol{h})$. Another possibility would be to consider the F-madogram approach, see [5]. Following the approach in [11], a graphical test based on the empirical versions of $\chi_u(\boldsymbol{h})$ in (2.20) and $\bar{\chi}_u(\boldsymbol{h})$ in (2.21) is used to explore possible anisotropy of the spatial dependence. Consider Z_t, $t = 1,\ldots,T$, T copies of a max-mixture process Z with unit Fréchet margin F. It is easy to compute the empirical estimates of $\chi_u(\boldsymbol{h})$ and $\bar{\chi}_u(\boldsymbol{h})$ from the empirical univariate and bivariate distributions as follows:

$$\widehat{\chi}_u(\boldsymbol{h}) = 2 - \frac{\log\left(T^{-1}\sum_{t=1}^{T}\mathbb{1}_{\{U_t(\boldsymbol{s})<u,U_t(\boldsymbol{s}+\boldsymbol{h})<u\}}\right)}{\log\left(T^{-1}\sum_{t=1}^{T}\mathbb{1}_{\{U_t(\boldsymbol{s})<u\}}\right)}, \quad \boldsymbol{s},\boldsymbol{s}+\boldsymbol{h}\in\mathcal{S}, \tag{3.21}$$

$$\widehat{\bar{\chi}}_u(\boldsymbol{h}) = \frac{2\log\left(T^{-1}\sum_{t=1}^{T}\mathbb{1}_{\{U_t(\boldsymbol{s})>u\}}\right)}{\log\left(T^{-1}\sum_{t=1}^{T}\mathbb{1}_{\{U_t(\boldsymbol{s})>u,U_t(\boldsymbol{s}+\boldsymbol{h})>u\}}\right)} - 1, \quad \boldsymbol{s},\boldsymbol{s}+\boldsymbol{h}\in\mathcal{S}, \tag{3.22}$$

where $U_t = F(Z_t)$ and $\mathbb{1}(.)$ represents the indicator function.

The dataset has been divided according to four different directional sectors: $(-\pi/8,\pi/8]$, $(\pi/8,3\pi/8]$, $(3\pi/8,5\pi/8]$, and $(5\pi/8,7\pi/8]$, where 0 represents the northing direction. On the basis of observed data, we construct the empirical estimates $\widehat{\chi}_u(\boldsymbol{h})$ and $\widehat{\bar{\chi}}_u(\boldsymbol{h})$. Figure 3.11

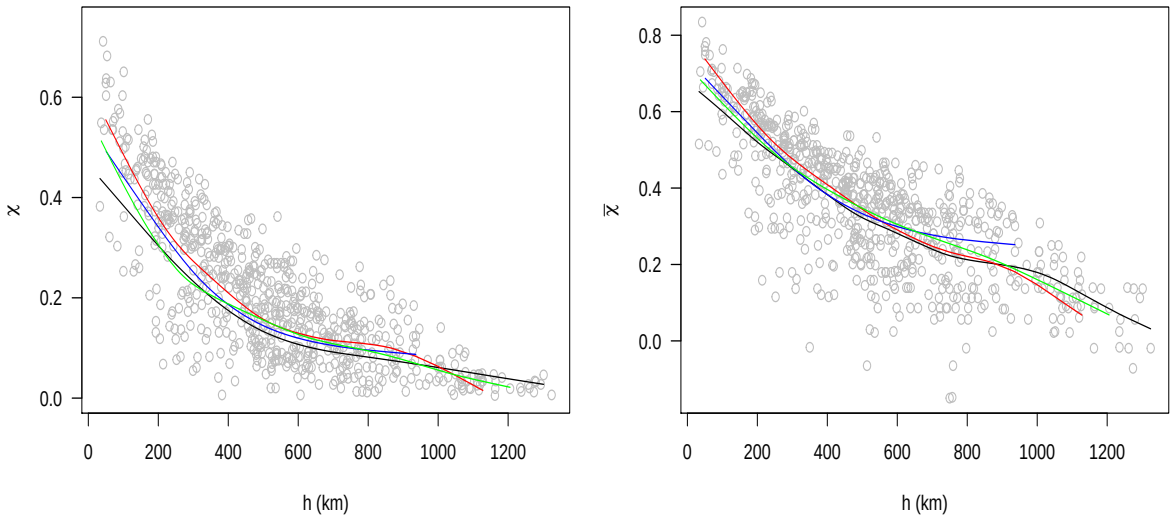


Figure 3.11: Pairwise empirical estimates of $\chi(.)$ (left panel) and $\bar{\chi}(.)$ (right panel) versus distance $h = \|\boldsymbol{h}\|$ at threshold $u = 0.970$. Gray points are empirical pairwise estimates for all data pairs. Colored curves are the loess smoothed values of the empirical estimates in different directional sectors: black curve $(-\pi/8, \pi/8]$, red curve $(\pi/8, 3\pi/8]$, blue curve $(3\pi/8, 5\pi/8]$, and green curve $(5\pi/8, 7\pi/8]$.

displays the directional loess smoothing of the empirical estimates $\widehat{\chi}_u(\boldsymbol{h})$ and $\widehat{\bar{\chi}}_u(\boldsymbol{h})$ at $u = 0.97$ with respect to distance $h = \|\boldsymbol{h}\|$. Based on these estimates, there is no strong evidence against isotropy. On the other hand, as mentioned in Section 2.2.2, the empirical estimates of $\chi_u(\boldsymbol{h})$ and $\bar{\chi}_u(\boldsymbol{h})$ can be useful in distinguishing between asymptotic dependence and independence. By means of visual inspection, Figure 3.11 provides an indication that asymptotic dependence between stations seems to be present up to a distance of 500 km, where $\widehat{\chi}_{0.97}(h) > 0$ for $h \leq 500$ km, whereas asymptotic independence could be conjectured between 500 km and 1000 km, where $\widehat{\bar{\chi}}_{0.97}(h) < 1$ for $h \in (500, 1000)$ km. Moreover, the pairs of sites separated by a distance $h \geq 1000$ km are (exactly) independent as $\widehat{\bar{\chi}}_{0.97}(h) \simeq 0$ for $h \geq 1000$ km, recall Example 2.2.

3.4.3 Testing procedure

The two statistics LR_a and Z_a have emerged as an effective tool for testing hypothesis on the mixing parameter of a spatial max-mixture model by the simulation study. These tests are model-based approaches. The first step is to adopt a family of models within which the true distribution of the data is assumed to lie. However, it is difficult to assess in practice whether a data set should be modeled using an asymptotically dependent or asymptotically independent model. So, the common way to perform statistical inference is to fit several parametric models. Afterward, selecting the best-fitting model based on a suitable criterion, see, e.g., [5, 11, 101]. Nevertheless, many dependence structures can be obtained by varying the max-stable and asymptotically independent components (X and Y) of a max-mixture model Z. In addition, various dependence structures can be obtained by varying the dependence function

(e.g., correlation functions and semivariograms) for each component.

We now describe a testing procedure for the mixing parameter of a max-mixture process based on the observed rainfall dataset. We split the daily rainfall dataset from the 38 sites into two zones; A and B, see Figure 3.10. The 19 sites in zone A are used for model inference and the other sites in zone B are put aside for testing. For this purpose, we consider nine arbitrary models that belong to the three classes: max-mixture, max-stable, and inverted max-stable. The models are

(i) Class A: consists of max-mixture models A_1-A_3.

- A_1: a max-mixture model which combines TEG and inverted TEG processes with an exponential correlation functions $\rho(\boldsymbol{h}) = \exp(-\|\boldsymbol{h}\|/\phi)$, $\phi > 0$. The compact set $\mathcal{A}$ is a disk of fixed and unknown radius r.
- A_2: a max-mixture model which combines a TEG and inverted BR processes. The correlation function for the TEG process is of type powered exponential defined by $\rho_X(\boldsymbol{h}) = \exp[-(\|\boldsymbol{h}\|/\phi_X)^{\kappa_X}]$, $\phi_X > 0$ and $0 < \kappa_X < 2$, where ϕ_X and κ_X denote, respectively, the range and the smoothing parameters. While the semivariogram for the inverted BR process is modeled by a spherical model

$$\gamma_Y(\boldsymbol{s}_1, \boldsymbol{s}_2) = \begin{cases} \alpha_Y \left\{1.5\,(\|\boldsymbol{h}\|/\phi_Y) - 0.5\,(\|\boldsymbol{h}\|/\phi_Y)^3\right\} & , \|\boldsymbol{h}\| \le \phi_Y \\ \alpha_Y & , \|\boldsymbol{h}\| > \phi_Y, \end{cases}$$

  where $\alpha_Y > 0$ is the value of the semivariogram when it reaches its sill, and $\phi_Y > 0$ is the range, recall Section 2.1.2.
- A_3: a max-mixture model which combines TEG process as in A_1 and an inverted Smith process with a diagonal covariance matrix $\boldsymbol{\Sigma}_Y = \phi_Y\,\mathrm{Id}_2$.

(ii) Class B: consists of max-stable models B_1-B_3.

- B_1: the TEG process described in A_1.
- B_2: a BR process with a spherical semivariogram.
- B_3: a Smith process with a diagonal covariance matrix.

(iii) Class C: consists of models C_1-C_3 which are the inverted max-stable components in models A_1-A_3.

For all the considered models, the marginal distribution is assumed to be unit Fréchet. Therefore, we transform the dataset to Fréchet by adopting a nonparametric transformation using the empirical distribution function, recall Section 3.1.1. Afterward, we estimate the dependence parameters using the censored pairwise approach (3.8), where the threshold u in (3.8) is taken as the 0.9 empirical quantiles at each site. So, with this censoring mechanism, we focus on threshold exceedances. Equal weights are assumed. The matrices $\boldsymbol{\mathcal{H}}(\boldsymbol{\vartheta})$ and $\boldsymbol{\mathcal{J}}(\vartheta)$ and the related quantities, CLIC and standard errors (S.E.) are estimated empirically.

Our results are summarized in Table 3.4. The best-fitting model for zone A, as judged by CLIC, is the hybrid dependence model A_2, for which pairs of sites separated by a distance h smaller than $2\widehat{r}_X \simeq 720$ km (respectively, greater than 720 km) are asymptotically max-stable dependent (respectively, asymptotically independent).

Table 3.4: Summary of the fitted models based on the daily exceedances from the Australian data at zone A. The CLIC and S.E. values are reported. ^{a} indicates to the lower CLIC.

Class A	$\widehat{\phi}_X$	κ_X	$\widehat{r}_X$	$\widehat{a}$	$\widehat{\phi}_Y$	$\widehat{r}_Y$	$\widehat{\alpha}_Y$	CLIC
A_1	342.87	-	713.08	$\simeq 1$	2235.51	986.40	-	1952371
S.E.	58.39	-	210.74	0.03	498.65	271.34	-	-
A_2	206.22	1.93	361.71	0.39	1014.54	-	2.46	1952188^a
S.E.	81.17	1.04	199.62	0.11	306.55	-	1.24	-
A_3	32.19	-	108.55	0.28	1154.96	-	-	1952402
S.E.	13.76	-	70.02	0.16	217.88	-	-	-
Class B	$\widehat{\phi}_X$	$\widehat{r}_X$	$\widehat{\alpha}_X$					CLIC
B_1	311.92	761.84	-					1952378
S.E.	68.52	189.10	-					-
B_2	422.89	-	4.77					1952266
S.E.	78.29	-	2.93					-
B_3	189.59	-	-					1960850
S.E.	41.66	-	-					-
Class C	$\widehat{\phi}_Y$	$\widehat{r}_Y$	$\widehat{\alpha}_Y$					CLIC
C_1	18.05	707.12	-					1960726
S.E.	6.81	115.41	-					-
C_2	533.04	-	5.63					1952411
S.E.	88.15	-	2.27					-
C_3	316.63	-	-					1963215
S.E.	207.85	-	-					-

Considering the best-fitting model in zone A (i.e., A_2), we perform the proposed statistical tests LR_a and Z_a to examine whether model A_2 can be an appropriate model to quantify the extremal dependence structures for rainfall data in zone B as in zone A with the same mixing parameter $a_0 = 0.39$. Simply, on the basis of observed data in zone B, we want to test whether the null hypothesis $H_0 : a = a_0$ holds for $a_0 \in \{0.01, 0.1, 0.2, 0.3, 0.39, 0.5, 0.6, 0.99\}$. In addition, to assess the validity of our results, the corresponding CLIC under the null hypothesis is computed. Our results are summarized in Table 3.5. In summary, both statistics retain the null hypothesis $H_0 : a = 0.39$. Moreover, we observe the agreement between the test findings and CLIC values. Thus, the max-mixture model A_2 may be used to model the dependence structures of daily precipitation in both zones A and B with the same a. These conclusions seem to be reasonable and could be expected since both studied zones are located on the east coast of Australia and may have a homogeneous precipitation pattern. Equivalently, an independent two-samples Z-test is performed to compare the mixture parameter in both zones. Denote by a_A (respectively, a_B the mixing parameter for the best-fitting model in zone A (respectively, zone B). We want to test the

null hypothesis $H_0 : a_A = a_B$. We compute the statistic

$$Z_a^* = \frac{\widehat{a}_A - \widehat{a}_B}{SE_{\widehat{a}_A - \widehat{a}_B}}, \tag{3.23}$$

where SE stands for the estimated standard error. We found $|Z_a^*| = 0.42$, and p-value $= 0.67$, $\alpha = 0.05$. Similarly, we conclude that there is no significant difference between the mixing parameters in both zones. On the other hand, for testing $H_0 : a_A = a_B$ by the LR statistics, a reparameterization may be required, i.e., to define a new parameter $a' = a_A - a_B$ and treat the rest as nuisance parameters. Investigation about the latter is deferred to future work.

Table 3.5: Testing results of the null hypothesis $H_0 : a = a_0$, with $a_0 \in \{0.01, 0.1, 0.2, 0.3, 0.39, 0.5, 0.6, 0.99\}$ on the basis of observed data in zone B. $\alpha = 0.05$. The corresponding CLIC values under H_0 are reported.

	LR_a		$\lvert Z_a \rvert$		
a_0	statistic	p-value	statistic	p-value	CLIC
0.01	16.07	6.10×10^{-5}	4.42	9.87×10^{-6}	2085837
0.10	10.98	9.21×10^{-4}	3.63	2.83×10^{-4}	1980814
0.20	9.53	2.02×10^{-3}	2.75	5.96×10^{-3}	1955458
0.30	6.34	1.18×10^{-2}	2.19	2.85×10^{-2}	1952356
0.39	1.21	2.71×10^{-1}	0.92	3.58×10^{-1}	1952194
0.50	3.78	5.18×10^{-2}	1.83	6.72×10^{-2}	1952227
0.60	10.14	1.45×10^{-3}	2.68	7.36×10^{-3}	1957503
0.99	21.46	3.61×10^{-6}	4.69	2.73×10^{-6}	2213019

Finally, by the introduced testing scheme, the statistics LR_a and Z_a can be viewed as an efficient parametric approach for model validation on the mixing parameter a. Of course, since $\widehat{a} \in (0, 1)$, testing the global hypothesis $H_0 : \boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$, where $\boldsymbol{\vartheta}_0 = (206.22, 1.93, 361.71, 0.39, 1014.54, 2.46)$ is possible with a limiting distribution χ^2_6, recall Remark 3.2. This test might provide a parametric validation tool for all model parameters. However, in the case of rejection H_0, an individual test has to be applied to each parameter.

3.5 Conclusions

In summary, we considered hypothesis testing for the mixture parameter of a max-mixture model using two statistics the Z_a and LR_a. A censored pairwise likelihood has employed for statistical estimation. A simulation study showed that both statistics perform well, even when we considered testing at values that are very close to the boundary points. In addition, we observed the following:

- Generally, the resulting pairwise likelihood estimates perform well, although the variability in some estimates is relatively large, especially for the asymptotic independence parameters. It also shows some bias in the estimation of asymptotic independence model parameters.

- The two statistics LR_a and Z_a show a plausible performance. Both tests seem to be unbiased. Furthermore, these tests can control the type I error rate α.
- The LR_a test seems to be more powerful than the Z_a one when the true mixing parameter a being a non-boundary point (i.e., $a \in (0,1)$), whereas, the converse is true when a being a boundary point (i.e., $a = 0$ or 1).
- As expected, the power to reject asymptotic dependence, i.e., $H_0 : a = 1$ (respectively, asymptotic independence, i.e., $H_0 : a = 0$) improves as the tested value $a_0 \to 0$ (respectively, $a_0 \to 1$). Nevertheless, these tests do not seem to be very powerful in this case. One possibility that might improve the performance is to consider a truncated high-order composite likelihood, which results in a better estimation performance, see [26].

In the real data example, we showed that our testing procedure could be a performant model validation tool.

From a theoretical point of view, traditional methods for deriving hypothesis tests at boundaries (i.e., $H_0 : a = 0$ or $a = 1$) do not work in this non-regular situation due to the presence of additional nuisance parameters which are not identified under the null hypotheses. The subject of non-regular likelihood-based inference for composite likelihoods is not well studied and might be very difficult, where the usual asymptotic theory doesn't work. So, deriving hypothesis tests for this case remains an open question.

Finally, let us remark that the composite LR statistic could also be useful when comparing nested models. For instance, it could be interesting to test whether it is more appropriate to use a powered exponential correlation function $\rho(\boldsymbol{h}) = \exp\left\{-\left(\frac{\|\boldsymbol{h}\|}{\phi}\right)^{\kappa}\right\}$, for $\phi > 0$ and $0 < \kappa < 2$, or an exponential correlation function $\rho(\boldsymbol{h}) = \exp\left\{-\frac{\|\boldsymbol{h}\|}{\phi}\right\}$ in spatial TEG model, i.e., $H_0 : \kappa = 1$ versus $H_1 : \kappa \neq 1$.

Chapter 4

Fitting spatial max-mixture processes with unknown extremal dependence class: an exploratory analysis tool

This chapter is based on our results from the paper: A. Abu-Awwad, V. Maume-Deschamps and P. Ribereau—Fitting spatial max-mixture processes with unknown extremal dependence class: an exploratory analysis tool. Published online to *TEST journal.*

Typically, the classes of models fitted to extreme data are based on asymptotic arguments that do not apply to the data themselves. Hence, the statistical estimation for extreme values is always subject to mis-specification bias. So, there is still a need to develop tools for initial analysis and for checking the validity of fitted models.

In the max-stable framework, on the basis of the infinite possible choices of λ for F^{λ}-madogram (2.30), the authors in [15] proposed an estimation procedure for the extremal dependence function $\theta(.)$ of a max-stable process. They first estimate the F^{λ}-madogram nonparametrically by its empirical version; then, a nonlinear least squares minimization between the empirical F^{λ}-madogram and its theoretical counterpart is applied to obtain the estimates. The appropriate choice of λ has been determined via some numerical trials. In a simulation study, this procedure seems to improve the empirical estimates of $\theta(.)$ and may outperform other known estimators: the nonparametric estimator [25], the maximum likelihood estimator [92] and the madogram-based estimator see [35]. In this chapter, we broaden the use of F^{λ}-madogram to fit the class of spatial max-mixture processes, recall Definition 2.16.

Fitting spatial max-mixture processes using existing inference approaches such as the widely adopted pairwise likelihood estimation and the semi-parametric estimation method in [5] have some disadvantages, recall Section 1.1. Specifically, they are model-based approaches. However, in practice, asymptotic properties are always difficult to infer, see [67]. Hence, it would be interesting to fit spatial max-mixture models which encompassing both asymptotic dependence classes and let the data speak for themselves. In this chapter, we develop a new estimation procedure in the family of max-mixture models. Unlike existing approaches, there is no necessity to make subjective a priori judgments about which distribution family to adopt. In this sense, this approach can be seen as "*a model-free*" approach in the class of max-mixture models. The main motivation for developing this approach is to offer a better model selection at an exploratory

stage and better model checking at the validation stage. Our approach uses a least squares fit based on the F^λ-madogram. It has some similarities with the semi-parametric estimation approach used in [5]. However, there are important fundamental differences between the two procedures in the inferential methodology and target, see Section 4.2.

The remainder of the chapter is organized as follows. The F^λ-madogram for max-mixture models is calculated in Section 4.1. The proposed F^λ-madogram inference procedure and its main properties are detailed in Section 4.2. Section 4.3 illustrates the performance of the proposed estimation procedure through various simulation studies. In Section 4.4, we apply our inference approach to analyze monthly maxima of daily rainfall over the East of Australia. Concluding remarks are given in Section 4.5.

4.1 F^λ-madogram for spatial max-mixture model

The $\chi(.)$ (respectively, $\bar{\chi}(.)$) measures defined in Section 2.2.2 are suited to quantify asymptotic dependence (respectively, asymptotic independence). Nevertheless, inference based on these measures is difficult because few observations are available as u approaches 1, see, e.g., [9, 23]. Therefore, in the present chapter, we propose an appropriate dependence measure for the class of max-mixture processes that may provide information on several extremal dependence structures.

The F^λ-madogram is defined in (2.30). We derive a closed-form expression of the F^λ-madogram for max-mixture models containing the parameters of interest. Afterward, we develop F^λ-madogram procedure to estimate the model parameters: θ_X, θ_Y and a.

Proposition 4.1. *Let $\{X(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ be a simple max-stable process, with extremal dependence coefficient function θ_X, and $\{Y(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ be an inverted max-stable process with extremal dependence coefficient function θ_Y. Then, for any spatial lag $\boldsymbol{h} = \boldsymbol{s}_1 - \boldsymbol{s}_2$, $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{S}$, the F^λ-madogram of the spatial max-mixture process $\{Z(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ is given for $a \neq 1$ by*

$$\nu_{F^\lambda}(\boldsymbol{h}) = \frac{\lambda}{1+\lambda} - \frac{2\lambda}{a(\theta_X(\boldsymbol{h})-1)+1+\lambda} + \frac{\lambda}{a\theta_X(\boldsymbol{h})+\lambda} - \frac{\lambda\theta_Y(\boldsymbol{h})}{(1-a)\theta_Y(\boldsymbol{h})+a\theta_X(\boldsymbol{h})+\lambda}\beta\left(\frac{a\theta_X(\boldsymbol{h})+\lambda}{1-a}, \theta_Y(\boldsymbol{h})\right), \tag{4.1}$$

where $\beta(.,.)$ is the beta function.

As a consequence of Proposition 4.1, we rediscover the expressions of F^λ-madograms for max-stable and inverted max-stable processes, that is,

- letting $a \to 1$, the F^λ-madogram of a simple max stable process X with extremal dependence coefficient θ_X is
$$\nu_{F^\lambda}(\boldsymbol{h}) = \frac{\lambda}{\lambda+1}\frac{\theta_X(\boldsymbol{h})-1}{\lambda+\theta_X(\boldsymbol{h})} \tag{4.2}$$
and we have $\nu_{F^\lambda}(\boldsymbol{h}) \in [0, \frac{\lambda}{(1+\lambda)(2+\lambda)}]$, see [15],

- setting $a = 0$, the F^λ-madogram of an inverted max-stable process Y with extremal dependence coefficient θ_Y is

$$\nu_{F^\lambda}(\boldsymbol{h}) = \frac{1}{1+\lambda} - \frac{\lambda\theta_Y(\boldsymbol{h})}{\lambda+\theta_Y(\boldsymbol{h})}\beta(\lambda,\theta_Y(\boldsymbol{h})), \tag{4.3}$$

- the choice $\lambda = 1$ gives the formula for F-madogram for max-mixture models, see [5].

Proof. Proposition 4.1. Following the idea of the proof from [60, 81], we use that for any $x, y \in \mathbb{R}$, $|x-y| = 2\max\{x,y\} - (x+y)$ with $x = F^\lambda\{Z(\boldsymbol{s})\}$ and $y = F^\lambda\{Z(\boldsymbol{s}+\boldsymbol{h})\}$. Moreover, recall that $\mathbb{E}[F^\alpha\{Z(\boldsymbol{s})\}] = 1/(1+\alpha)$ for any $\alpha > 0$. We have

$$\begin{aligned}\nu_{F^\lambda}(\boldsymbol{h}) &= \mathbb{E}\left[\max\left\{F^\lambda\{Z(\boldsymbol{s})\}, F^\lambda\{Z(\boldsymbol{s}+\boldsymbol{h})\}\right\}\right] - \frac{1}{2}\mathbb{E}\left[F^\lambda\{Z(\boldsymbol{s})\} + F^\lambda\{Z(\boldsymbol{s}+\boldsymbol{h})\}\right] \\ &= \mathbb{E}\left[\max\left\{F^\lambda\{Z(\boldsymbol{s})\}, F^\lambda\{Z(\boldsymbol{s}+\boldsymbol{h})\}\right\}\right] - \frac{1}{(1+\lambda)}.\end{aligned}$$

Considering the random variable, $W = \max\left\{F^\lambda\{Z(\boldsymbol{s})\}, F^\lambda\{Z(\boldsymbol{s}+\boldsymbol{h})\}\right\}$. The probability distribution function G of W satisfies (recall (2.32) and (2.35))

$$\begin{aligned}G(z) &= \mathbb{P}[W \le z] \\ &= \mathbb{P}\left[\max\left\{F^\lambda\{Z(\boldsymbol{s})\}, F^\lambda\{Z(\boldsymbol{s}+\boldsymbol{h})\}\right\} \le z\right] \\ &= \mathbb{P}\left[Z(\boldsymbol{s}) \le F^{-1}(z^{1/\lambda}), Z(\boldsymbol{s}+\boldsymbol{h}) \le F^{-1}(z^{1/\lambda})\right] \\ &= \mathbb{P}\left[Z(\boldsymbol{s}) \le -\frac{\lambda}{\log(z)}, Z(\boldsymbol{s}+\boldsymbol{h}) \le -\frac{\lambda}{\log(z)}\right] \\ &= \exp\left\{-V_{\boldsymbol{h}}^X\left(-\frac{\lambda}{a\log(z)}, -\frac{\lambda}{a\log(z)}\right)\right\} \cdot \left[z^{\frac{1-a}{\lambda}} + z^{\frac{1-a}{\lambda}} - 1 + \right. \\ &\left. \exp\left\{-V_{\boldsymbol{h}}^Y\left(-\frac{1}{\log\left(1-z^{\frac{1-a}{\lambda}}\right)}, -\frac{1}{\log\left(1-z^{\frac{1-a}{\lambda}}\right)}\right)\right\}\right].\end{aligned}$$

This may be rewritten as

$$\begin{aligned}G(z) &= z^{\frac{a\theta_X(\boldsymbol{h})}{\lambda}} \cdot \left[2z^{\frac{(1-a)}{\lambda}} - 1 + \left(1 - z^{\frac{(1-a)}{\lambda}}\right)^{\theta_Y(\boldsymbol{h})}\right] \\ &= 2z^{\frac{a(\theta_X(\boldsymbol{h})-1)+1}{\lambda}} - z^{\frac{a}{\lambda}\theta_X(\boldsymbol{h})} + z^{\frac{a}{\lambda}\theta_X(\boldsymbol{h})}\left(1 - z^{\frac{(1-a)}{\lambda}}\right)^{\theta_Y(\boldsymbol{h})}.\end{aligned}$$

Thus, one has

$$\begin{aligned}\mathbb{E}[W] = \int_0^1 z dG(z) &= zG(z)\Big|_0^1 - \int_0^1 G(z)dz \\ &= 1 - \left[\frac{2\lambda}{a(\theta_X(\boldsymbol{h})-1)+1+\lambda} - \frac{\lambda}{a\theta_X(\boldsymbol{h})+\lambda} + I\right]\end{aligned}$$

and

$$
\begin{aligned}
I :&= \int_0^1 z^{\frac{a}{\lambda}\theta_X(\boldsymbol{h})} \left(1 - z^{\frac{(1-a)}{\lambda}}\right)^{\theta_Y(\boldsymbol{h})} dz \\
&= \frac{\lambda}{(1-a)} \beta\left(\frac{a\theta_X(\boldsymbol{h}) + \lambda}{1-a}, \theta_Y(\boldsymbol{h}) + 1\right) \\
&= \frac{\lambda\theta_Y(\boldsymbol{h})}{(1-a)\theta_Y(\boldsymbol{h}) + a\theta_X(\boldsymbol{h}) + \lambda} \beta\left(\frac{a\theta_X(\boldsymbol{h}) + \lambda}{1-a}, \theta_Y(\boldsymbol{h})\right).
\end{aligned}
$$

This proves (4.1). □

Figures 4.1 and 4.2 visualize the theoretical behavior of the F^λ-madogram as a function of the scalar distance $h = \|\boldsymbol{h}\| =: \|\boldsymbol{s}_1 - \boldsymbol{s}_2\| \geq 0$, $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{S}$ for isotropic max-stable, inverted max-stable and max-mixture models. Different values of λ are used, i.e., $\lambda \in \{0.5, 1, 1.5, 3\}$. Obviously, depending on the value of λ, the F^λ-madogram exhibits a large variety of behaviors away from the origin, where $\lambda = 1.5$ corresponds to the largest values of the F^λ-madogram.

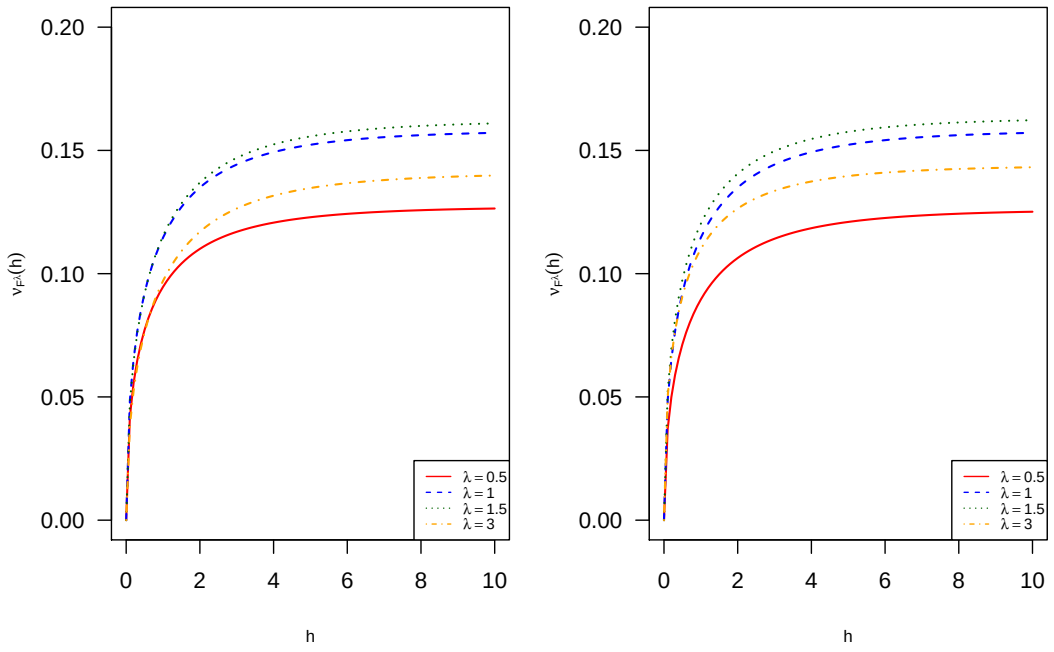


Figure 4.1: F^λ-madogram plotted as a function of the distance $h = \|\boldsymbol{h}\|$ for an extremal-t process with degrees of freedom (df) $\nu = 4$ and exponential correlation function $\rho(\boldsymbol{h}) = \exp(-\|\boldsymbol{h}\|/3)$ (left panel) and for an inverted extremal-t process with the same parameters (right panel).

4.2 Estimation of max-mixture models using F^λ-madogram

4.2.1 Estimation methodology

We now describe an estimation scheme for the max-mixture process Z (2.34). Due to computational reasons, we split the estimation scheme into two steps: we first estimate the extremal

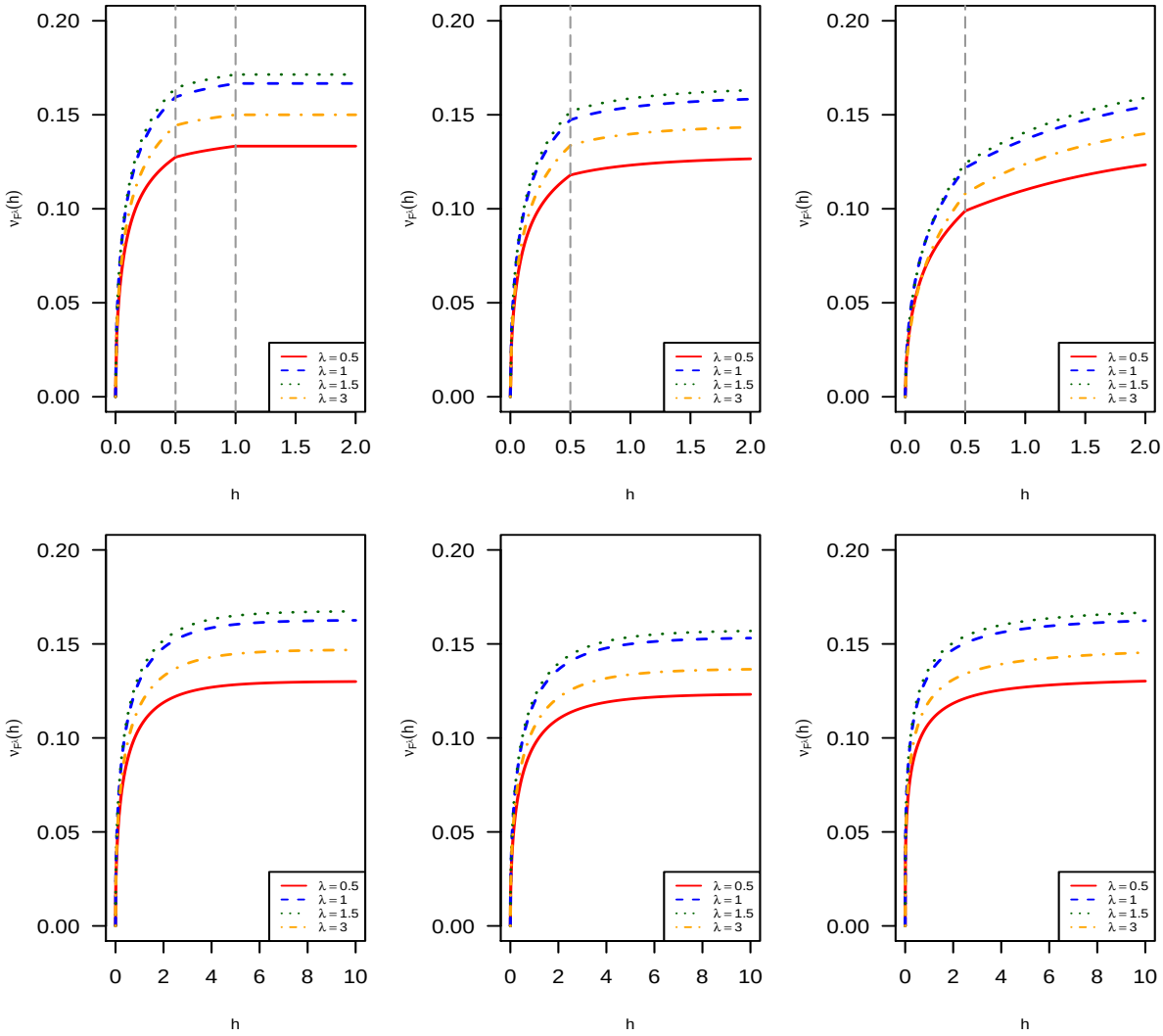


Figure 4.2: Theoretical F^λ-madogram functions (4.1) for $\lambda = \{0.5, 1, 1.5, 3\}$. Top row: (left panel) a max-mixture model in which X is a TEG process with $\mathcal{A}_X$ a disk of fixed radius $r_X = 0.25$ and $\rho_X(\boldsymbol{h}) = \exp(-\|\boldsymbol{h}\|/0.2)$. The asymptotically independent process Y is an inverted TEG process with $\mathcal{A}_Y$ a disk of fixed radius $r_Y = 0.5$ and $\rho_Y(\boldsymbol{h}) = \exp(-\|\boldsymbol{h}\|/0.4)$. (Middle panel) a max-mixture model in which X is a TEG process as before. The asymptotically independent process Y is an inverted extremal-t process with df $\nu = 2$ and $\rho_Y(\boldsymbol{h}) = \exp(-\|\boldsymbol{h}\|)$. (Right panel) a max-mixture model in which X is a TEG process as before. The asymptotically independent process Y is an inverted Smith process with $\boldsymbol{\Sigma} = \mathrm{Id}_2$. The gray vertical lines represent the diameters of the disks for TEG processes. Bottom row: (left panel) a max-mixture model in which X is a BR process with an intrinsically stationary isotropic semivariogram model $\gamma(\boldsymbol{h}) = 2\|\boldsymbol{h}\|$. The asymptotically independent process Y is an inverted extremal-t process with df $\nu = 3$ and exponential correlation function $\rho_Y(\boldsymbol{h}) = \exp(-\|\boldsymbol{h}\|/2)$. (Middle panel) a max-mixture model in which X is an extremal-t process with df $\nu_1 = 2$ and exponential correlation function $\rho_X(\boldsymbol{h}) = \exp(-\|\boldsymbol{h}\|/2)$. The asymptotically independent process Y is an inverted extremal-t process with df $\nu_2 = 4$ and $\rho_Y(\boldsymbol{h}) = \exp(-\|\boldsymbol{h}\|/3)$. (Right panel) a max-mixture model in which X is a BR process with semivariogram $\gamma(\boldsymbol{h}) = 2\|\boldsymbol{h}\|^{1/2}$. The asymptotically independent process Y is an inverted BR process with semivariogram $\gamma(\boldsymbol{h}) = 2\|\boldsymbol{h}\|$. The mixing parameter is fixed to $a = 0.5$ in the six models.

dependence coefficient functions $\theta_X(\boldsymbol{h})$ and $\theta_Y(\boldsymbol{h})$, with a fixed mixing parameter a, and then we estimate a with the former estimates of $\theta_X(\boldsymbol{h})$ and $\theta_Y(\boldsymbol{h})$. From (4.1), we may write the F^{λ}-madogram as a function of a, λ, θ_X and θ_Y, that is, $\nu_{F^\lambda}(\boldsymbol{h}) = g(a, \lambda, \theta_X(\boldsymbol{h}), \theta_Y(\boldsymbol{h}))$. The idea beyond our procedure is that θ_X and θ_Y may be estimated by $\widehat{\theta}_X$ and $\widehat{\theta}_Y$, minimizing the square difference between $g(a, \lambda, \theta_X(\boldsymbol{h}), \theta_Y(\boldsymbol{h}))$ and its empirical counterpart, then we can estimate a by $\widehat{a}$ such that the empirical version of the $F^{\lambda'}$-madogram is the closest to $g(a, \lambda', \widehat{\theta}_X(\boldsymbol{h}), \widehat{\theta}_Y(\boldsymbol{h}))$. Formally, for $Z_i, i = 1, \ldots, T$, independent copies of Z, let

$$Q_i(\boldsymbol{h}, \lambda) = \frac{1}{2}|F^{\lambda}(Z_i(\boldsymbol{s})) - F^{\lambda}(Z_i(\boldsymbol{s}+\boldsymbol{h}))|,$$

where F denotes the unit Fréchet distribution function. From the definition of the F^{λ}-madogram, we have $\mathbb{E}[Q_i(\boldsymbol{h}, \lambda)] = \nu_{F^\lambda}(\boldsymbol{h})$. In real-world applications, the marginal laws are usually not unit Fréchet and thus have to be transformed to unit Fréchet. In that case, the empirical distribution function $\widehat{F}$ is used instead of F. More precisely, we have:

$$\widehat{Q}_i(\boldsymbol{h}, \lambda) = \frac{1}{2}|\widehat{F}^{\lambda}(Z_i(\boldsymbol{s})) - \widehat{F}^{\lambda}(Z_i(\boldsymbol{s}+\boldsymbol{h}))| \text{ and } \widehat{\nu}_{F^\lambda}(\boldsymbol{h}) = T^{-1}\sum_{i=1}^{T}\widehat{Q}_i(\boldsymbol{h}, \lambda),$$

where $\widehat{\nu}_{F^\lambda}(\boldsymbol{h})$ denotes the empirical version of $\nu_{F^\lambda}(\boldsymbol{h})$. Denote by $\Lambda \subset (0, \infty)$ a finite set of some possible choices for λ; then, for a given value of a, a semi-parametric nonlinear least squares minimization procedure for estimating the summary functions $\boldsymbol{\theta}(\boldsymbol{h}) = (\theta_X(\boldsymbol{h}), \theta_Y(\boldsymbol{h}))^t$ is

$$\widehat{\boldsymbol{\theta}}^{a}_{\boldsymbol{NLS}}(\boldsymbol{h}) = \underset{(\theta_X, \theta_Y)\in[1,2]}{\arg\min}\; T^{-1}\sum_{\lambda\in\Lambda}\sum_{i=1,\ldots,T}\left[\widehat{Q}_i(\boldsymbol{h}, \lambda) - g(a, \lambda, \theta_X(\boldsymbol{h}), \theta_Y(\boldsymbol{h}))\right]^2. \tag{4.4}$$

The suggested minimizing criterion (4.4) has some similarities with the semi-parametric estimation approach of [5]. Indeed, in [5], the estimation is based on a least squares minimization between the empirical F-madogram (i.e., $\lambda = 1$) and its theoretical counterpart, computed for several parametric models. However, one major difference here is that by (4.4) the estimates of θ_X and θ_Y do not rely on given forms for $\theta_X(\boldsymbol{h})$ and $\theta_Y(\boldsymbol{h})$ (recall Table 2.1); hence, it provides a way to make inference without specifying a specific distribution family before fitting the model.

Next, we estimate a by profiling the values of a goodness-of-fit criterion with respect to a. Assume that the Z_i's are observed at locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_D$ and let $\boldsymbol{h}$ be the pairwise distances between the $\boldsymbol{s}_j$'s. For fixed a, let $\widehat{\boldsymbol{\theta}}^{a}_{\boldsymbol{NLS}}(\boldsymbol{h}) = \left(\widehat{\theta}^{a}_X(\boldsymbol{h}), \widehat{\theta}^{a}_Y(\boldsymbol{h})\right)^t$ be estimated as above with some chosen distinct values $\lambda \in \Lambda$. Let $\lambda' \in [0, \infty)/\Lambda$ and denote by $\mathcal{H} \subset [0, \infty)$ the finite set of spatial lags $\boldsymbol{h}$, i.e.,

$$\mathcal{H} = \left\{\boldsymbol{h}_k = \boldsymbol{s}_i - \boldsymbol{s}_j,\ i = 1, \ldots, D-1,\ j = i+1, \ldots, D,\ k = 1, \ldots, \frac{D(D-1)}{2}\right\}.$$

Moreover, let $\widetilde{\nu}_{F^\lambda}(\boldsymbol{h}) = g\left(a, \lambda, \widehat{\theta}_X(\boldsymbol{h}), \widehat{\theta}_Y(\boldsymbol{h})\right)$. We define the following decision criterion (DC)

$$\mathrm{DC}(a) = \sum_{\boldsymbol{h}\in\mathcal{H}}\omega(\boldsymbol{h})\,[R-1]^2, \tag{4.5}$$

where $R = \dfrac{\widehat{\nu}_{F^{\lambda'}}(\boldsymbol{h})}{\widehat{\nu}_{F^{\lambda'}}(\boldsymbol{h})}$ and $\omega(\boldsymbol{h})$'s are nonnegative weights which can be used for example to reduce the number of pairs included in the estimation in order to decrease the computational burden and improving statistical efficiency, recall Section 3.1.2. Lastly, we estimate a by $\widehat{a}$ that minimizes $\mathrm{DC}(a)$.

Remark 4.1. *Equivalently, instead of (4.4), one may solve the following optimization problem*

$$\widehat{\boldsymbol{\theta}}^{a}_{\boldsymbol{NLS}}(\boldsymbol{h}) = \underset{(\theta_X,\theta_Y)\in[1,2]}{arg\ min} \sum_{\lambda\in\Lambda} \left[\widehat{\nu}_{F^{\lambda}}(\boldsymbol{h}) - g(a,\lambda,\theta_X(\boldsymbol{h}),\theta_Y(\boldsymbol{h}))\right]^2. \tag{4.6}$$

A similar approach has been proposed by [15] to estimate the extremal dependence coefficient function in a max-stable setting.

Of course, the likelihood-based procedures can be used to estimate a full generative spatial model and not only bivariate dependence summaries. However, the motivation beyond the construction of this inference strategy is to gain flexibility, where the data have the chance to speak for themselves without relying on a parametric model. This property is interesting from a statistical point of view, it may reduce the model risk and can be used as a preliminary step toward parametric modeling.

4.2.2 Consistency results

Below we establish the consistency of the resulting estimates from the F^{λ}-madogram estimation approach. We will assume the set $\Lambda = \{\lambda_1, \lambda_2\}$, see Section 4.3. Accordingly, we rewrite the criterion in (4.4) as

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}^{a}_{\boldsymbol{NLS}}(\boldsymbol{h}) &= \underset{(\theta_X,\theta_Y)\in[1,2]^2}{\arg\min}\ T^{-1} \sum_{i=1,\dots,T} \left[\widehat{Q}_i(\boldsymbol{h},\lambda_1) - g(a,\lambda_1,\theta_X(\boldsymbol{h}),\theta_Y(\boldsymbol{h})\right]^2 \\
&\quad + \left[\widehat{Q}_i(\boldsymbol{h},\lambda_2) - g(a,\lambda_2,\theta_X(\boldsymbol{h}),\theta_Y(\boldsymbol{h})\right]^2, \\
&=: \underset{\boldsymbol{\theta}\in[1,2]^2}{\arg\min}\ \mathcal{L}_T(\boldsymbol{h}, a, \lambda_1, \lambda_2, \boldsymbol{\theta}(\boldsymbol{h})).
\end{aligned}$$

In order to prove the consistency of the estimators, we need the two following assumptions:

(i) I_1: for any $a \in [0,1]$, $\exists\{\lambda_1,\lambda_2\} \in \Lambda$ with $\lambda_1 \neq \lambda_2$, such that the mapping

$$\begin{aligned}
[1,2]^2 &\longrightarrow \mathbb{R}^2 \\
(x,y) &\mapsto (g(a,\lambda_1,x,y), g(a,\lambda_2,x,y))
\end{aligned}$$

is injective.

(ii) I_2: let θ_X (respectively, θ_Y), θ'_X (respectively, θ'_Y) be the extremal coefficients of max-stable (respectively, inverted max-stable) processes. Let λ be fixed, if for all $\boldsymbol{h}$,

$$g(a,\lambda,\theta_X(\boldsymbol{h}),\theta_Y(\boldsymbol{h})) = g(a',\lambda,\theta'_X(\boldsymbol{h}),\theta'_Y(\boldsymbol{h})),$$

then $a = a'$.

Remark 4.2. *The hypotheses I_1 and I_2 are identifiability hypotheses. Numerical tests on several models seem to indicate that they are satisfied for various max-mixture models, but we have no formal proof.*

Theorem 4.1. *Assume that $\left(Z_i(\boldsymbol{s}_j)\right)_{i=1,\dots,T}$ are i.i.d. copies of $Z(\boldsymbol{s}_j)$, $j = 1,\dots,D$ where Z is a max-mixture spatial process with mixing parameter $a_0 \in [0,1]$. Assume that the injectivity conditions I_1 and I_2 are verified. Then the estimator $\widehat{a}$ of a is consistent in the sense that*

$$\widehat{a} \longrightarrow a_0 \text{ in probability as } T \to \infty.$$

Proof. of Theorem 4.1. We first begin with standard arguments on least squares to prove the consistency of $\widehat{\theta}_X^{a_0}$ and $\widehat{\theta}_Y^{a_0}$. Consider $\{\lambda_1, \lambda_2\} \in \Lambda$, with $\lambda_1 \neq \lambda_2$ satisfying I_1. For any $\boldsymbol{h} \in \mathcal{H}$, let $\boldsymbol{\theta}^0(\boldsymbol{h}) = \left(\theta_X^0(\boldsymbol{h}), \theta_X^0(\boldsymbol{h})\right)^t \in [1,2]^2$ denote the unknown true parameter of extremal coefficients, write

$$\varepsilon_{\boldsymbol{h},i}^1 = \widehat{Q}_i(\boldsymbol{h}, \lambda_1) - g\left(a_0, \lambda_1, \boldsymbol{\theta}^0(\boldsymbol{h})\right) \text{ and}$$
$$\varepsilon_{\boldsymbol{h},i}^2 = \widehat{Q}_i(\boldsymbol{h}, \lambda_2) - g\left(a_0, \lambda_2, \boldsymbol{\theta}^0(\boldsymbol{h})\right).$$

So, we have

$$\mathcal{L}_T\left(\boldsymbol{h}, a_0, \lambda_1, \lambda_2, \boldsymbol{\theta}^0(\boldsymbol{h})\right) = \frac{1}{T}\sum_{i=1}^{T}(\varepsilon_{\boldsymbol{h},i}^1)^2 + \frac{1}{T}\sum_{i=1}^{T}(\varepsilon_{\boldsymbol{h},i}^2)^2 \text{ and}$$

$$\mathcal{L}_T\left(\boldsymbol{h}, a_0, \lambda_1, \lambda_2, \widehat{\boldsymbol{\theta}}_{\boldsymbol{NLS}}^{\boldsymbol{a_0}}(\boldsymbol{h})\right) = \frac{1}{T}\sum_{i=1}^{T}(\varepsilon_{\boldsymbol{h},i}^1)^2 + \frac{1}{T}\sum_{i=1}^{T}(\varepsilon_{\boldsymbol{h},i}^2)^2 + \left[g(a_0, \lambda_1, \boldsymbol{\theta}^0(\boldsymbol{h})) - g(a_0, \lambda_1, \widehat{\boldsymbol{\theta}}_{\boldsymbol{NLS}}^{\boldsymbol{a_0}}(\boldsymbol{h}))\right]^2$$
$$+\frac{2}{T}\left[g(a_0, \lambda_1, \boldsymbol{\theta}^0(\boldsymbol{h})) - g(a_0, \lambda_1, \widehat{\boldsymbol{\theta}}_{\boldsymbol{NLS}}^{\boldsymbol{a_0}}(\boldsymbol{h}))\right]\sum_{i=1}^{T}\varepsilon_{\boldsymbol{h},i}^1 + \left[g(a_0, \lambda_2, \boldsymbol{\theta}^0(\boldsymbol{h})) - g(a_0, \lambda_2, \widehat{\boldsymbol{\theta}}_{\boldsymbol{NLS}}^{\boldsymbol{a_0}}(\boldsymbol{h}))\right]^2$$
$$+\frac{2}{T}\left[g(a_0, \lambda_2, \boldsymbol{\theta}^0(\boldsymbol{h})) - g(a_0, \lambda_2, \widehat{\boldsymbol{\theta}}_{\boldsymbol{NLS}}^{\boldsymbol{a_0}}(\boldsymbol{h}))\right]\sum_{i=1}^{T}\varepsilon_{\boldsymbol{h},i}^2.$$

Since $\mathcal{L}_T(\boldsymbol{h}, a_0, \lambda_1, \lambda_2, \widehat{\boldsymbol{\theta}}_{\boldsymbol{NLS}}^{\boldsymbol{a_0}}(\boldsymbol{h})) \leq \mathcal{L}_T(\boldsymbol{h}, a_0, \lambda_1, \lambda_2, \boldsymbol{\theta}^0(\boldsymbol{h}))$, we get

$$\left[g(a_0, \lambda_1, \boldsymbol{\theta}^0(\boldsymbol{h})) - g(a_0, \lambda_1, \widehat{\boldsymbol{\theta}}_{\boldsymbol{NLS}}^{\boldsymbol{a_0}}(\boldsymbol{h}))\right]^2 + \frac{2}{T}\left[g(a_0, \lambda_1, \boldsymbol{\theta}^0(\boldsymbol{h})) - g(a_0, \lambda_1, \widehat{\boldsymbol{\theta}}_{\boldsymbol{NLS}}^{\boldsymbol{a_0}}(\boldsymbol{h}))\right]\sum_{i=1}^{T}\varepsilon_{\boldsymbol{h},i}^1$$
$$+\left[g(a_0, \lambda_2, \boldsymbol{\theta}^0(\boldsymbol{h})) - g(a_0, \lambda_2, \widehat{\boldsymbol{\theta}}_{\boldsymbol{NLS}}^{\boldsymbol{a_0}}(\boldsymbol{h}))\right]^2 + \frac{2}{T}\left[g(a_0, \lambda_2, \boldsymbol{\theta}^0(\boldsymbol{h})) - g(a_0, \lambda_2, \widehat{\boldsymbol{\theta}}_{\boldsymbol{NLS}}^{\boldsymbol{a_0}}(\boldsymbol{h}))\right]\sum_{i=1}^{T}\varepsilon_{\boldsymbol{h},i}^2 \leq 0.$$

Using the convergence results from [81] (Proposition 3), we have that

$$\frac{1}{T}\sum_{i=1}^{T}\varepsilon_{\boldsymbol{h},i}^k \to 0 \text{ in probability, } k = 1,2.$$

It follows that $g(a_0, \lambda_k, \widehat{\boldsymbol{\theta}}_{\boldsymbol{NLS}}^{\boldsymbol{a_0}}(\boldsymbol{h})) \to g(a_0, \lambda_k, \boldsymbol{\theta}^0(\boldsymbol{h}))$ in probability, $k = 1,2$. Consider $\boldsymbol{\theta}^*(\boldsymbol{h}) = \limsup \widehat{\boldsymbol{\theta}}_{\boldsymbol{NLS}}^{\boldsymbol{a_0}}(\boldsymbol{h})$ and $\boldsymbol{\theta}_*(\boldsymbol{h}) = \liminf \widehat{\boldsymbol{\theta}}_{\boldsymbol{NLS}}^{\boldsymbol{a_0}}(\boldsymbol{h})$ (exist and are random variables), leading to

$$g(a_0, \lambda_k, \boldsymbol{\theta}^*(\boldsymbol{h})) = g(a_0, \lambda_k, \boldsymbol{\theta}^0(\boldsymbol{h})) \text{ and } g(a_0, \lambda_k, \boldsymbol{\theta}_*(\boldsymbol{h})) = g(a_0, \lambda_k, \boldsymbol{\theta}^0(\boldsymbol{h})), \ k = 1,2.$$

Using the injectivity hypothesis I_1, we conclude that for any $\boldsymbol{h} \in \mathcal{H}, \boldsymbol{\theta}^*(\boldsymbol{h}) = \boldsymbol{\theta}^0(\boldsymbol{h})$ and $\boldsymbol{\theta}_*(\boldsymbol{h}) = \boldsymbol{\theta}^0(\boldsymbol{h})$ and thus $\widehat{\boldsymbol{\theta}}^{a_0}_{\boldsymbol{NLS}}(\boldsymbol{h}) \to \boldsymbol{\theta}^0(\boldsymbol{h})$ (in probability).

Now, since $\widehat{a}$ reaches the minimum of DC, we have $\mathrm{DC}(\widehat{a}) \leq \mathrm{DC}(a_0)$. The convergence of $\widehat{\boldsymbol{\theta}}^{a_0}_{\boldsymbol{NLS}}(\boldsymbol{h})$ and $\widehat{\nu}_{F^{\lambda'}}$ from [81] (Proposition 3) implies that $\lim_{T\to\infty} \mathrm{DC}(a_0) = 0$. By the same arguments as above and using the injectivity condition I_2, we conclude that $\widehat{a} \to a_0$ (in probability). □

4.3 Simulation study

In this section, we provide some simulation results to assess the F^{λ}-madogram estimation approach introduced in Section 4.2. Denote by $\boldsymbol{\vartheta}$ the vector gathering the parameters to be estimated. Three examples of isotropic max-mixture models are studied, namely

(i) M_1: is the max-mixture model described in Section 3.3, which combines TEG and inverted TEG processes with stationary isotropic exponential correlation functions. The model parameters are $\boldsymbol{\vartheta} = (a, \phi_X, r_X, \phi_Y, r_Y)^t$.

(ii) M_2: is a max-mixture model where X is a TEG process as in M_1 and Y is an isotropic inverted extremal-t process with df ν_Y and exponential correlation function $\rho_Y(\boldsymbol{h}) = \exp(-\|\boldsymbol{h}\|/\phi_Y)$, $\phi_Y > 0$. The model parameters are $\boldsymbol{\vartheta} = (a, \phi_X, r_X, \phi_Y, \kappa_Y, \nu_Y)^t$.

(iii) M_3: is a max-mixture model, which combines extremal-t process X (recall 2.16) and inverted extremal-t process Y. Both X and Y have a stationary isotropic correlation function $\rho(\boldsymbol{h}) = \exp\{-(\|\boldsymbol{h}\|/\phi)^{\kappa}\}$, with range $\phi > 0$ and smoothness $\kappa \in (0, 2]$ and df $\nu \geq 1$. The model parameters are $\boldsymbol{\vartheta} = (a, \phi_X, \phi_Y, \kappa_X, \kappa_Y, \nu_X, \nu_Y)^t$.

4.3.1 Setup for simulation study

(i) For each experiment, we consider a moderately sized dataset from models M_1, M_2, and M_3 with a true mixing parameter a_0, $D = 50$ sites randomly and uniformly distributed in the square $[0, L]^2$, $L \in \mathbb{N}$ and $T = 1000$ independent replications at each site. The simulations have been carried out using the function *rmaxstab* of the R package SpatialExtremes (see [86]), except the TEG process (not implemented in SpatialExtremes) which has been simulated as in [44]. Each experiment was repeated $M = 100$ times.

(ii) For each dataset in (i), we estimate the extremal dependence functions $\theta_X(\boldsymbol{h})$ and $\theta_Y(\boldsymbol{h})$ using the nonlinear least squares estimation procedures (4.4) and (4.6). We perform this step with a set of different mixing parameters including a_0.

(iii) We calculate $\mathrm{DC}(a)$ with the estimated $\widehat{\theta}^a_X(\boldsymbol{h})$ and $\widehat{\theta}^a_Y(\boldsymbol{h})$ functions from (ii). It is expected that lower values of DC are likely related to the estimates of $\theta_X(\boldsymbol{h})$ and $\theta_Y(\boldsymbol{h})$ under a_0 in (ii). Equal weights $\omega(.)$ are assumed.

4.3.2 Results for F^{λ}-madogram estimation approach

Before showing some simulation studies results, as in [15], we have to determine the range of λ and λ'. After some exploratory analysis based on the considered models and using different sets of λ and λ', we found that the lower values of DC are likely related to the choice of $\Lambda = \{1, 3\}$ and $\lambda' = 1.5$. Furthermore, this choice seems a reasonable compromise between accuracy and computation time.

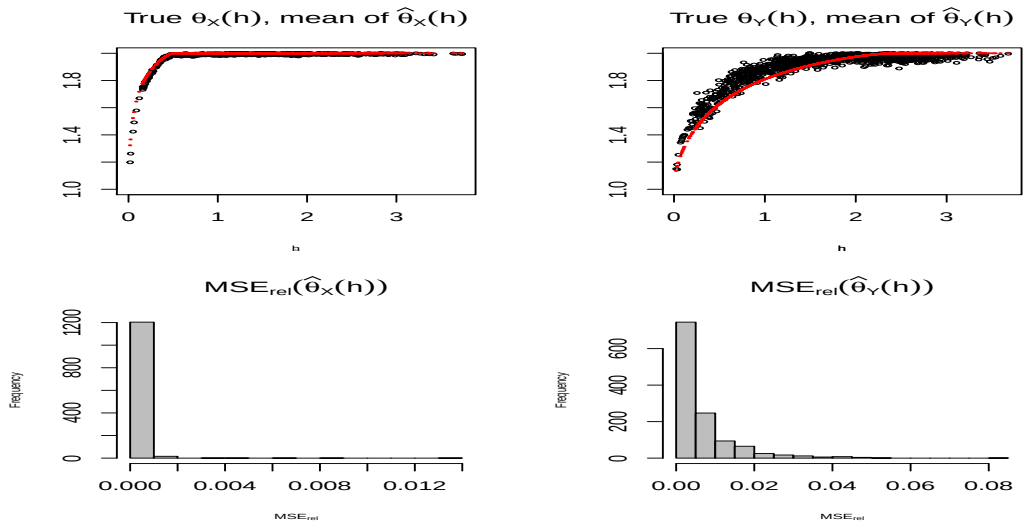
To assess the performance of the nonlinear least squares estimator $\widehat{\boldsymbol{\theta}}^{a}_{NLS}(\boldsymbol{h})$, we simulate data from models M_1, M_2 and M_3 as mentioned in (i). The performance of the estimators is given by the relative mean square error $\mathrm{MSE}_{\mathrm{rel}}$, for the $D(D-1)/2$ pairwise distances $h_k = \|\boldsymbol{h}_k\|$, see [15] (page 171), for a similar definition in the multivariate context

$$\mathrm{MSE}_{\mathrm{rel}}(h_k) = M^{-1} \sum_{i=1}^{M} \frac{(\widehat{\theta}^{a}_{i}(h_k) - \theta(h_k))^2}{\theta(h_k)}, \; k = 1, \ldots, D(D-1)/2. \tag{4.7}$$

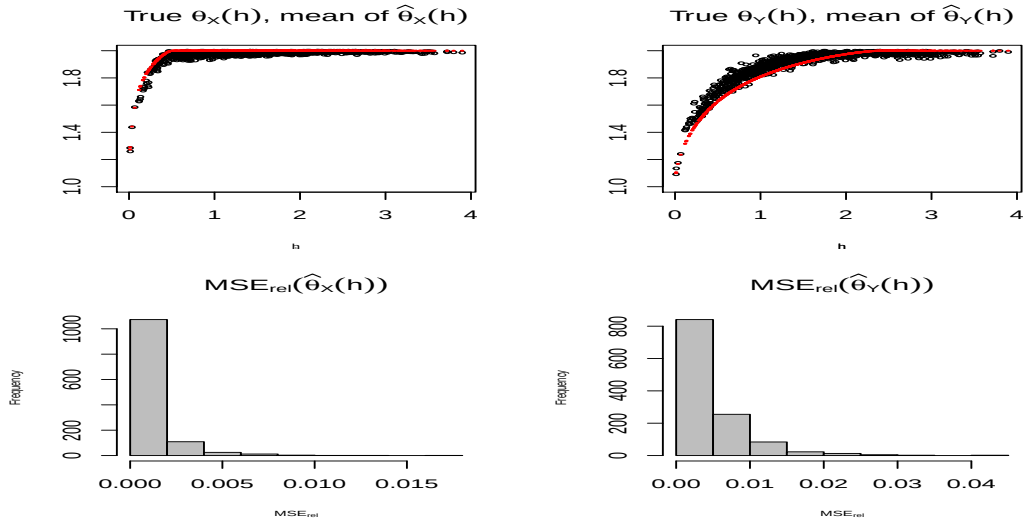
Figures 4.3 and 4.4 display the estimation performance of $\widehat{\theta}_X(\boldsymbol{h})$ and $\widehat{\theta}_Y(\boldsymbol{h})$ for models M_1 and M_2 as functions of the distance $h = \|\boldsymbol{h}\|$ for mixing parameters $a \in \{0.75, 0.5, 0.25\}$ based on the minimizing criterion (4.4), while Figure 4.5 shows the performance for model M_3 on the basis of the minimizing criterion (4.6). Generally, our estimation procedure appears to work well. It improves the empirical dependence summaries, which have high variability and do not yield a valid spatial model. Moreover, we observe that contrary to $\theta_Y(\boldsymbol{h})$, the estimation of $\theta_X(\boldsymbol{h})$ becomes more accurate as the mixing parameter value increases. The latter conclusion is consistent with the results obtained by the parametric estimation using pairwise likelihood, where the estimation of asymptotic dependence parameters becomes more accurate as the mixing parameter value increases (the RMSE and MAE are lower) and the converse seems to hold for asymptotic independence parameters, recall Section 3.3.

Next, we turn to the proposed decision criterion (DC) about the mixing parameter a. We perform a number of simulation studies using the above-mentioned max-mixture models. The boxplots in Figure 4.6, 4.7, and 4.8 show the values of DC against different mixing parameters $a \in \{0, 0.25, 0.5, 0.75, 1\}$ for models M_1, M_2, and M_3. Clearly, the lower values of DC are likely related with the true mixing parameter a_0.

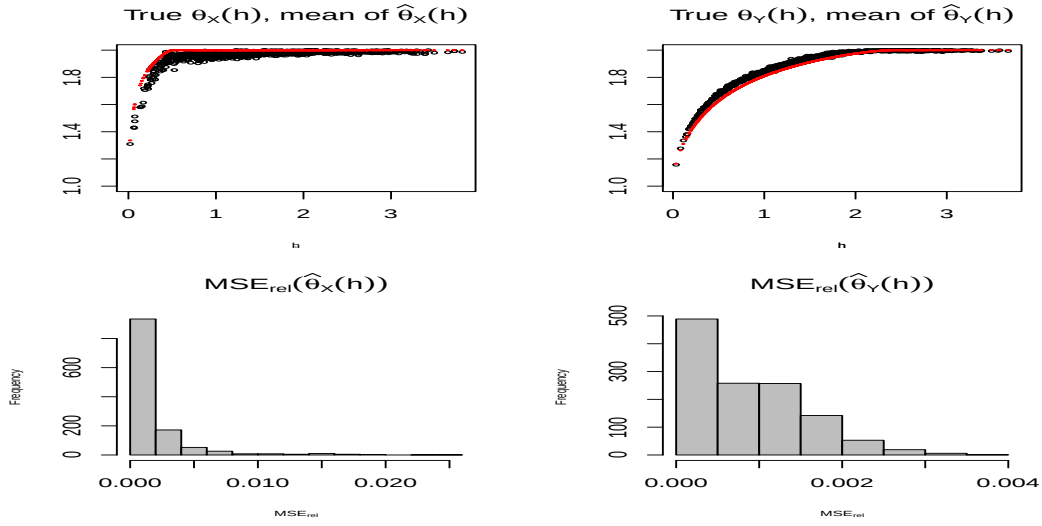
Lastly, we assess whether the resulting estimates of the mixing parameter a from the proposed F^{λ}-madogram estimation procedure are asymptotically Gaussian distributed. We apply our inferential methodology for all $a \in (0, 1)$ by step 0.01. The data have been simulated from the max-mixture model M_1 with a true mixing parameter value $a_0 \in \{0.25, 0.5, 0.75\}$. Each experiment was repeated 500 times. Figure 4.9 displays the histograms and density curves of the obtained errors: $\widehat{a} - a_0$. Clearly, the density of the errors is close to a centered Gaussian distribution.



(a) $a = 0.75$

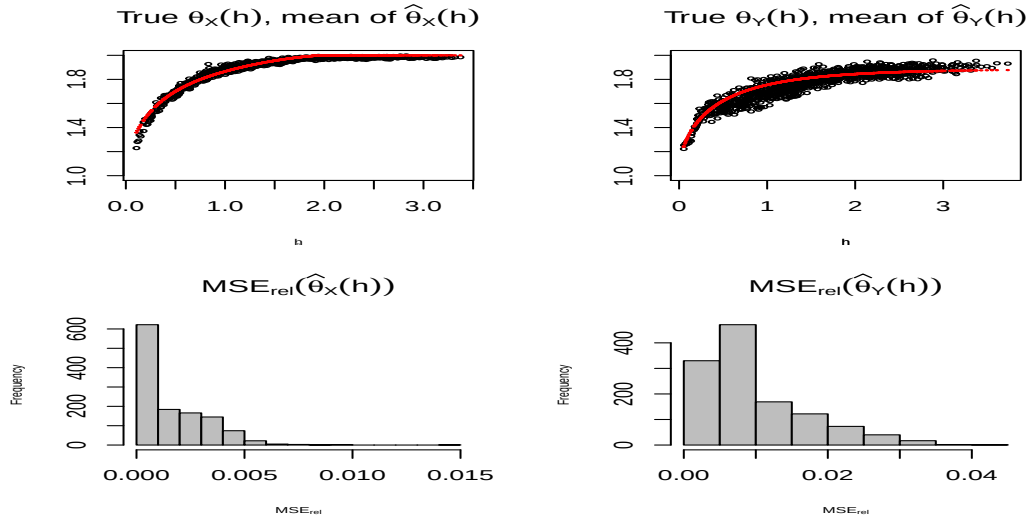


(b) $a = 0.5$

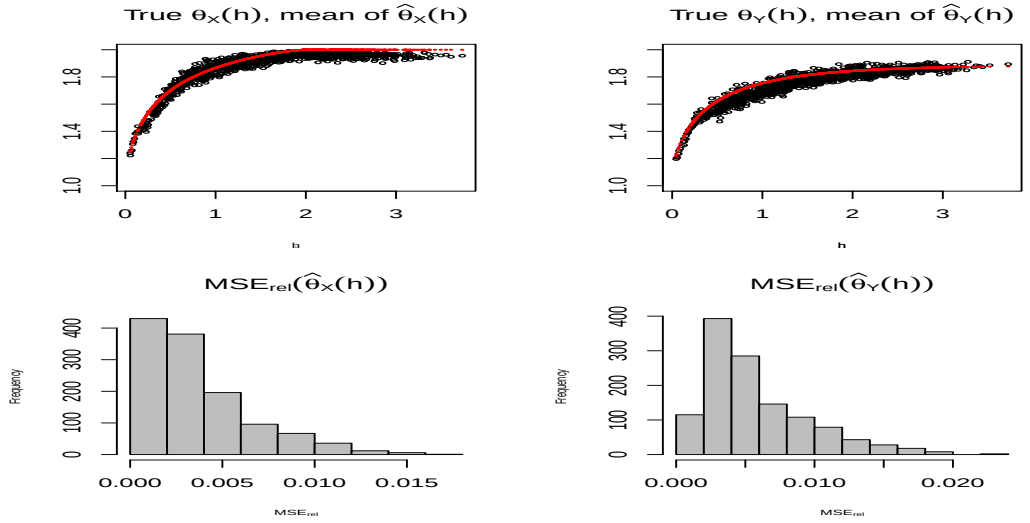


(c) $a = 0.25$

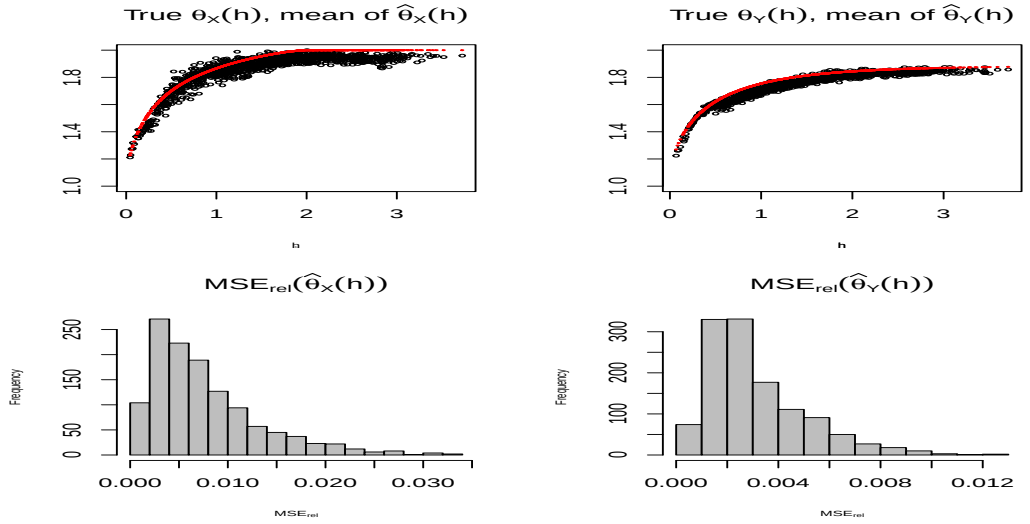
Figure 4.3: Estimation performance of $\hat{\boldsymbol{\theta}}_{NLS}(\boldsymbol{h})$. Data simulated from model M_1. We set $\phi_X = 0.1$, $r_X = 0.25$, $\phi_Y = 0.75$, and $r_Y = 1.2$ over a square $[0,3]^2$. Black points/red points: nonlinear least squares mean estimates/true extermal coefficients. The histograms display the MSE_{rel} of the nonlinear least squares estimates.



(a) $a = 0.75$

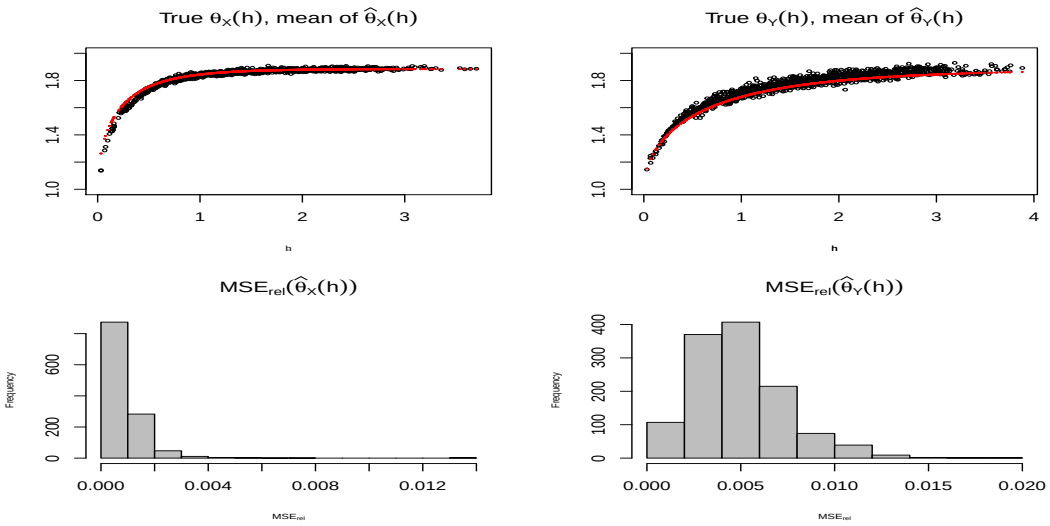


(b) $a = 0.5$

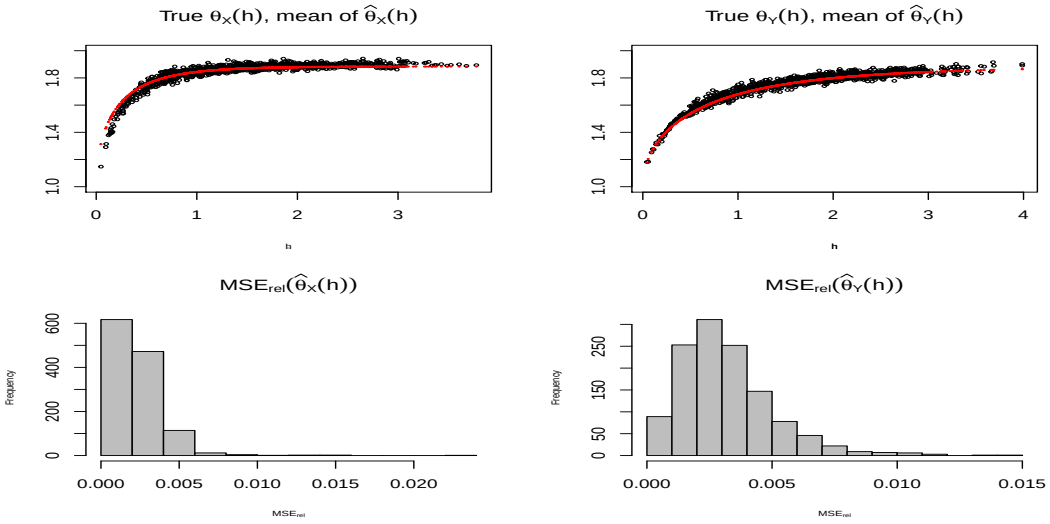


(c) $a = 0.25$

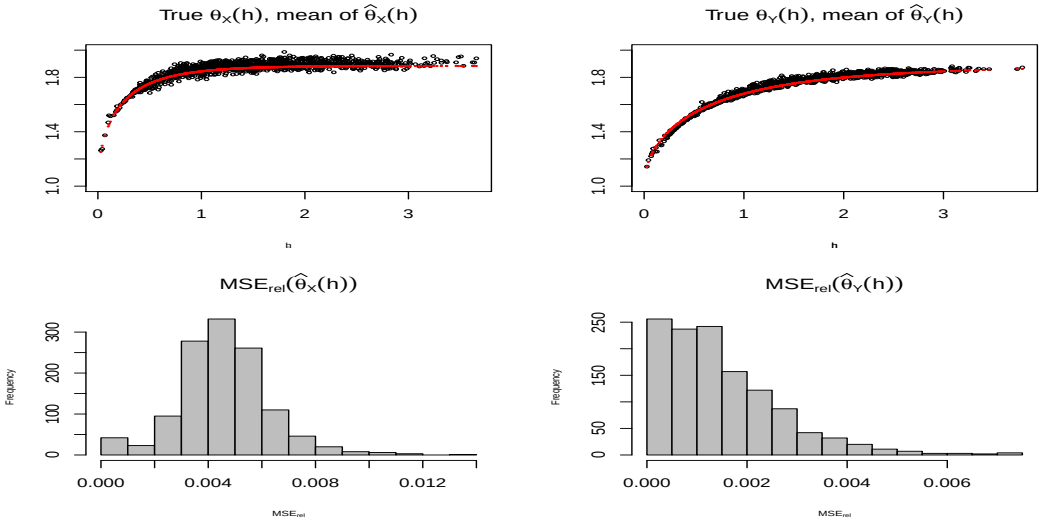
Figure 4.4: Estimation performance of $\hat{\boldsymbol{\theta}}_{NLS}(\boldsymbol{h})$. Data simulated from model M_2. We set $\phi_X = 0.5$, $r_X = 1$, $\phi_Y = 1$, and $\nu = 3$ over a square $[0,3]^2$. Black points/red points: nonlinear least squares mean estimates/true extremal coefficients. The histograms display the MSE_{rel} of the nonlinear least squares estimates.



(a) $a = 0.75$



(b) $a = 0.5$



(c) $a = 0.25$

Figure 4.5: Estimation performance of $\hat{\boldsymbol{\theta}}_{NLS}(\boldsymbol{h})$. Data simulated from M_3 using the same setup in Section 4.3.1. We set $\phi_X = 0.5$, $\phi_Y = 1.5$, $\kappa_X = \kappa_Y = 1$, and $\nu_X = \nu_Y = 3$ over a square $[0, 3]^2$. Black points/red points: nonlinear least squares mean estimates/true extermal coefficients. The histograms display the MSE_{rel} of the nonlinear least squares estimates.

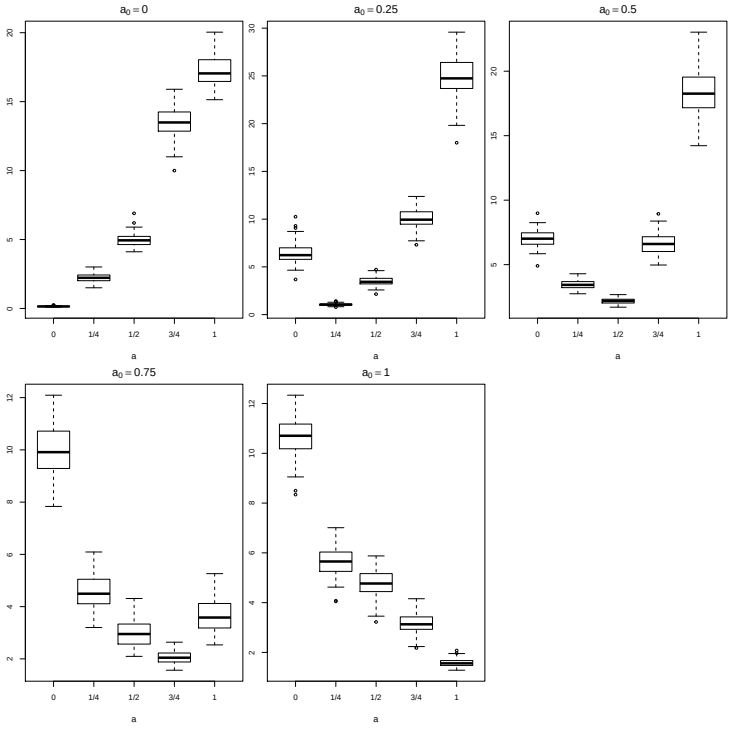


Figure 4.6: Boxplots for decision criterion (DC) of 100 data replications of 1000 independent copies from M_1 with parameters $\phi_X = 0.1$, $r_X = 0.25$, $\phi_Y = 0.75$ and $r_Y = 1.2$ over the square $[0,3]^2$. a_0 corresponds to the true mixing parameter.

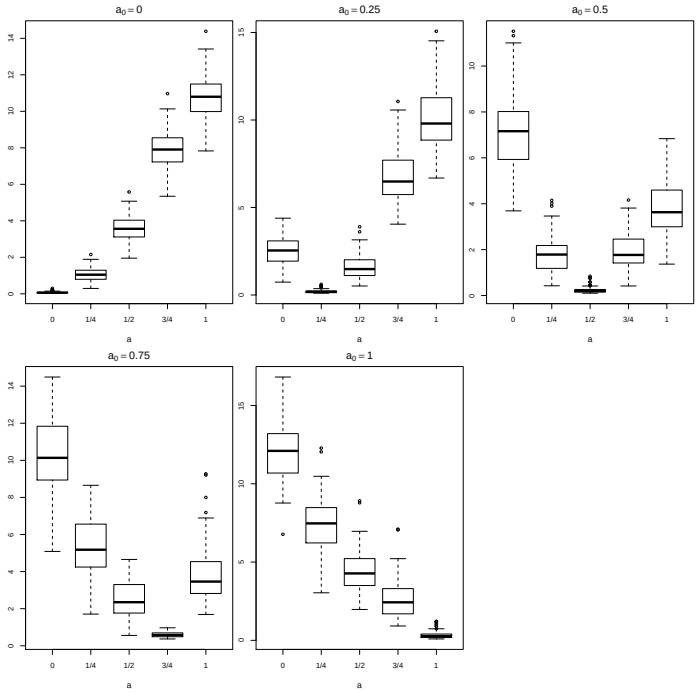


Figure 4.7: Boxplots for decision criterion (DC) of 100 data replications of 1000 independent copies from M_2 with parameters $\phi_X = 0.5$, $r_X = 1$, $\phi_Y = 1$ and $\nu = 3$ over the square $[0,3]^2$. a_0 corresponds to the true mixing parameter.

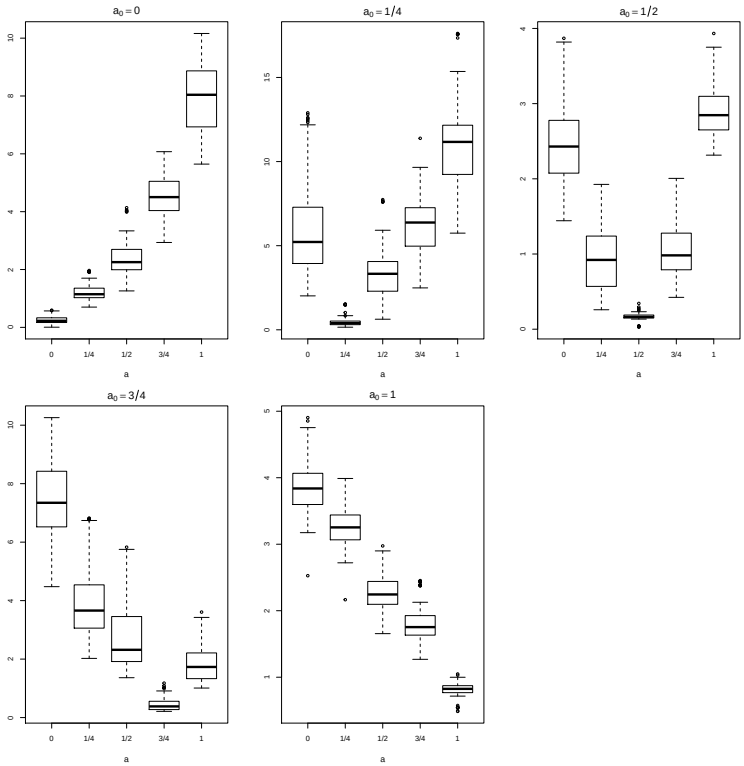


Figure 4.8: Boxplots for decision criterion (DC) of 100 data replications of 1000 independent copies from M_3 with parameters $\phi_X = 0.5$, $\phi_Y = 1.5$, $\kappa_X = \kappa_Y = 1$ and $\nu_X = \nu_Y = 3$ over a square $[0,3]^2$. a_0 corresponds to the true mixing parameter.

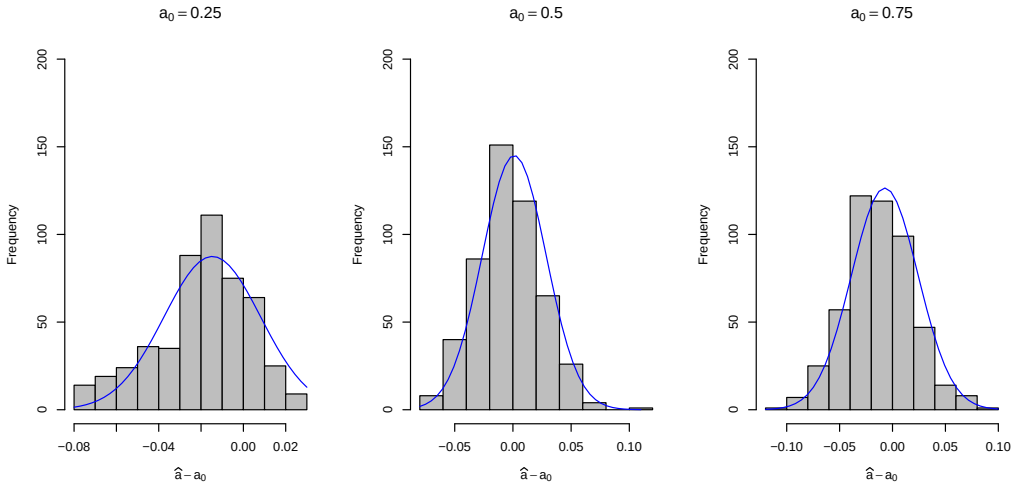


Figure 4.9: Histograms and density curves (blue) of the errors between the estimated mixing coefficient $\widehat{a}$ and the true value a_0 from model M_1 with parameters $\phi_X = 0.1$, $r_X = 0.25$, $\phi_Y = 0.75$, and $r_Y = 1.2$ over the square $[0,3]^2$.

4.4 Australian rainfall data revisited

In this section, we discuss a spatial extreme application to illustrate the benefits of our F^{λ}-madogram estimation approach. We revisit the Australian daily rainfall example analyzed in Section 3.4. Three additional monitoring stations are added to assess the goodness-of-fit. They are labeled by colored numbers {1,2,3}, see Figure 4.10.

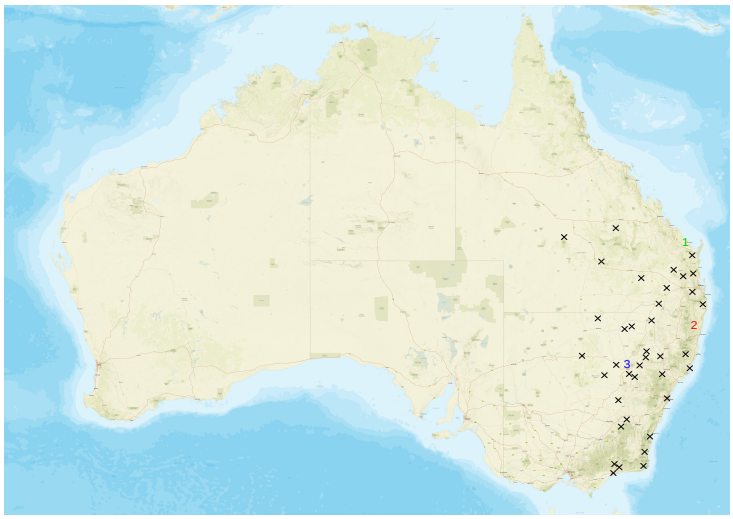


Figure 4.10: Geographical locations of 41 meteorological stations located in the East of Australia. The 38 stations with a cross-label are used for model inference. The other 3 stations labeled by colored numbers {1,2,3} are put aside for assessing the goodness-of-fit.

4.4.1 Monthly maxima data

Modeling only the seasonal winter maxima might lead to large variances in estimating empirical F^{λ}-madogram, which probably can be justified by the small number of independent replicates over the years 1972−2014. On the other hand, the potential problem with taking weekly maxima is that there will be some (even many) zero values. In addition, weekly maxima, which might correspond to just one wet day, might not be extreme. So, here we analyze block maxima of east Australia rainfall data by taking monthly maxima of daily rainfall in the winter season over the 43-year period. This yields a dataset comprising a total of $6 \times 43 = 258$ monthly maxima at each monitoring station. The entire dataset comprises 9804 measurements. About 2.6% of the measurements equal zero. Figure 4.11 depicts the distribution for the entire dataset of monthly maxima measurements, while Table 4.1 reports basic statistics for these measurements. Clearly, they suggest that the distribution of monthly maxima measurements is highly right-skewed and heavy-tailed. Furthermore, by inspecting autocorrelation plots of monthly maxima with a fixed location, we conclude that there is no temporal dependence in the time series.

In the literature, e.g., [60] analyzed monthly maxima of hourly precipitation measured at two French stations located in the west of Paris from February 1987 to December 2002. In that study, the authors proposed a simple inference tool inspired by geostatistics, the madogram, to

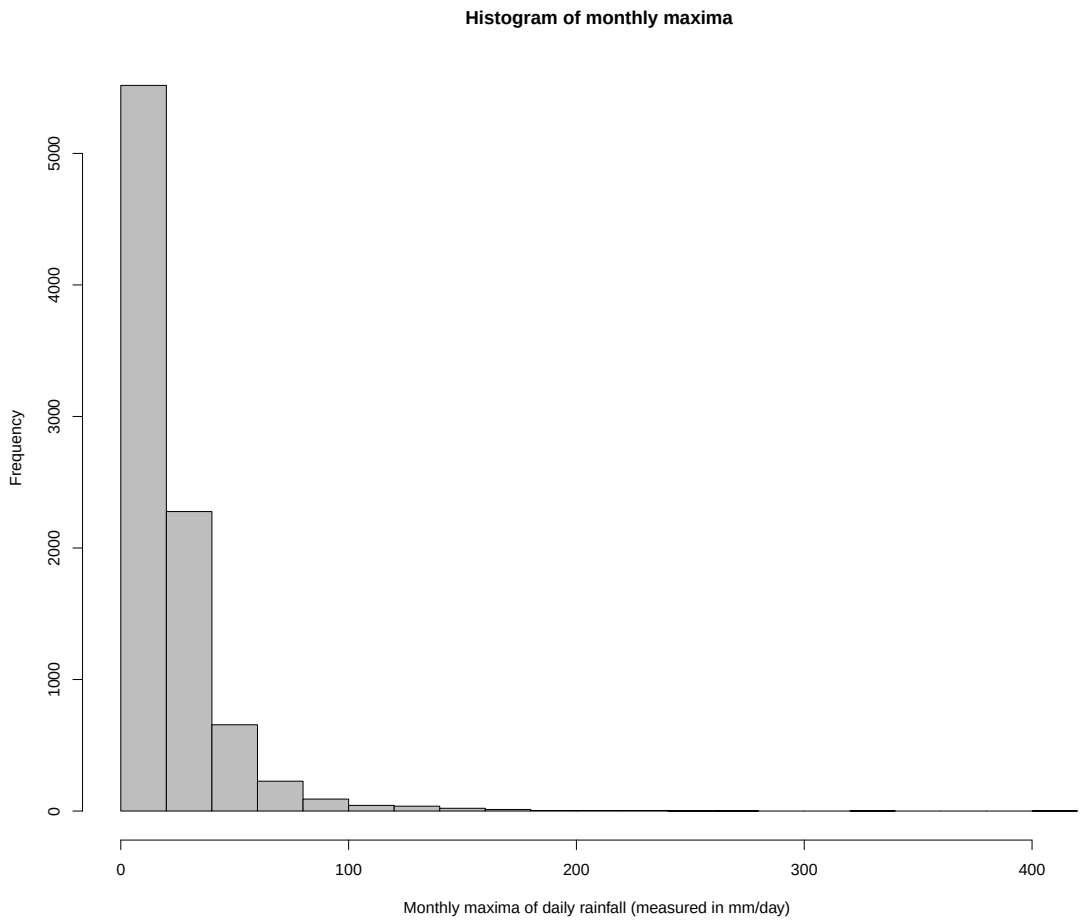


Figure 4.11: Australian rainfall example—histogram of winter monthly maxima of daily rainfall (measured in mm/day), combined for the 38 monitoring stations over the 43-year period (1972−2014).

Table 4.1: Australian rainfall example—basic statistics for winter monthly maxima of daily rainfall (measured in mm/day), combined for the 38 monitoring stations over the 43-year period (1972−2014). The central moments (first row) and empirical quantiles (second row).

Central moments	Mean	Variance	Skewness	Kurtosis				
	21.3	535.4	3.8	30.5				
Empirical quantiles	Minimum	25%	50%	75%	90%	95%	99%	Maximum
	0	7.2	15.4	27.4	45.0	60.2	118.0	410.7

the context of asymptotic independence between pairwise block maxima. Finally, let us remark that the daily rainfall in this region was also analyzed in two other studies [11, 5]. In [11], the pairwise likelihood estimation was adopted in the analysis of site-wise winter maxima and exceedances over a large threshold, while, in [5], a semi-parametric fit based on F-madogram was applied to analyze all daily rainfall data.

4.4.2 Marginal fitting

An appropriate inferential approach requires to fit marginal and dependence parameters. For marginal fitting, the generalized extreme value distribution $\text{GEV}_{\mu(\boldsymbol{s}),\sigma(\boldsymbol{s}),\xi(\boldsymbol{s})}$ in (2.6) is usually fitted to block maxima, recall Section 3.1. So, we fit this distribution separately to each station. Figure 4.12 displays the shape parameter estimates for each monitoring station. They are positive with positive confidence intervals. This may suggest a Fréchet distribution to be the most appropriate for the data at hand.

We assess the goodness of the marginal fits by quantile-quantile plots (QQ-plots) of the winter monthly maxima observations against fitted GEV distribution for each monitoring station. For example, Figure 4.13 depicts the QQ-plots at four monitoring stations. All plots show a reasonable fit. Afterward, data are transformed to unit Fréchet scale according to the probability integral transformation: $z \to \frac{-1}{\log(\widehat{F}(z))}$, where $\widehat{F}(.)$ is the estimated generalized extreme value distribution, that is, $\text{GEV}_{\widehat{\mu}(\boldsymbol{s}),\widehat{\sigma}(\boldsymbol{s}),\widehat{\xi}(\boldsymbol{s})}$.

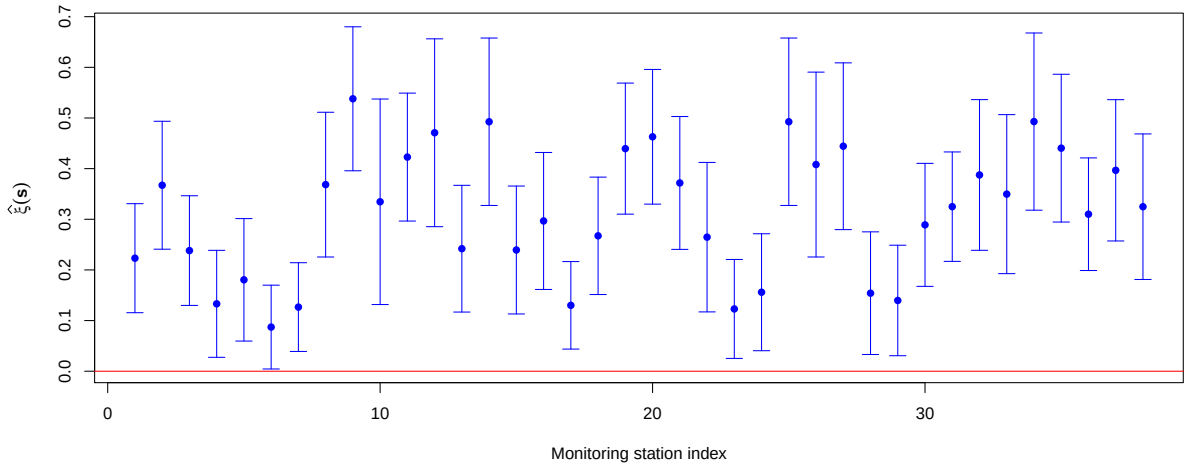


Figure 4.12: Estimated GEV shape parameter $\widehat{\xi}(\boldsymbol{s})$ at all monitoring stations with 95% confidence intervals.

4.4.3 Exploratory data analysis: F^{λ}-madogram estimation approach

The composite likelihood estimation serves as a reliable method to estimate parameters for extreme processes, see, e.g., [84]. So much effort has been placed on determining the appropriate dependence class using this method, where the standard way to perform statistical inference is based on fitting a list of models. Afterward, select the best-fitting model based on a suitable

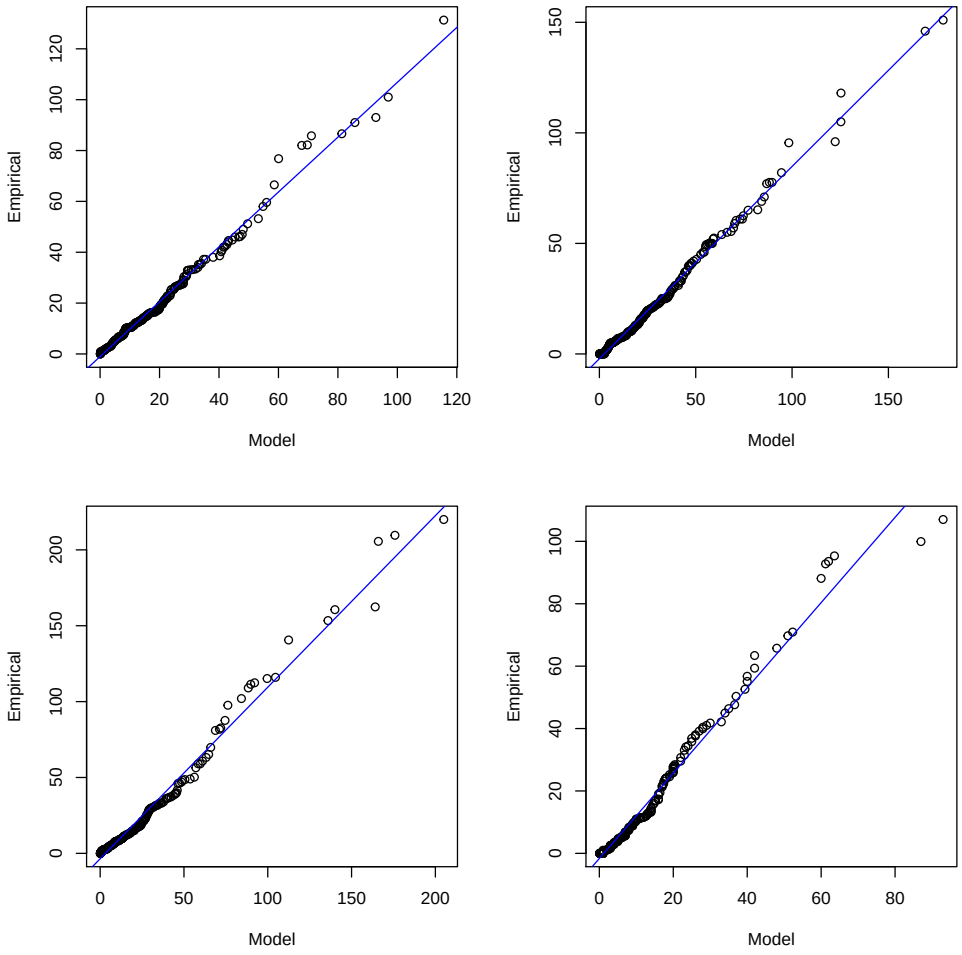


Figure 4.13: QQ-plots of winter monthly maxima of daily rainfall versus the fitted GEV distribution with $\widehat{\mu}(\boldsymbol{s})$, $\widehat{\sigma}(\boldsymbol{s})$ and $\widehat{\xi}(\boldsymbol{s})$ based on the time series corresponding to four monitoring stations.

criterion. Sometimes (even often) the list of fitted models seems arbitrary. The major advantage of the F^{λ}-madogram inference approach for spatial extremes over existing model-based ones is needless to explicitly determine parametric families on which we should work. Since we do not want to give an unfair advantage to our approach, it could serve as an exploratory tool to assess (in a preliminary step) what kind of parametric model would be appropriate for fitting the data.

In a preliminary analysis, we apply the F^{λ}-madogram estimation procedure described in Section 4.2 to estimate the max-mixture model parameters: a and the summary measures $\theta_X(.)$ and $\theta_Y(.)$. We use the following two steps.

(i) We first estimate a, where a varies from 0 to 1 by steps of 0.01. Similarly as mentioned in Section 4.3, we fixed $\Lambda = \{1, 3\}$ and $\lambda' = 1.5$. The $a \mapsto \mathrm{DC}(a)$ function is plotted in Figure 4.14. The best-fitting max-mixture model as judged by the DC criterion has an estimated mixing parameter of $\widehat{a} = 0.64$.

(ii) Next, for estimating the functions $\theta_X(\boldsymbol{h})$ and $\theta_Y(\boldsymbol{h})$, we refit the dataset again by the

nonlinear least squares procedure (4.4) on the basis of the resulting estimate of the mixing parameter in (i). The results are summarized in Figure 4.15.

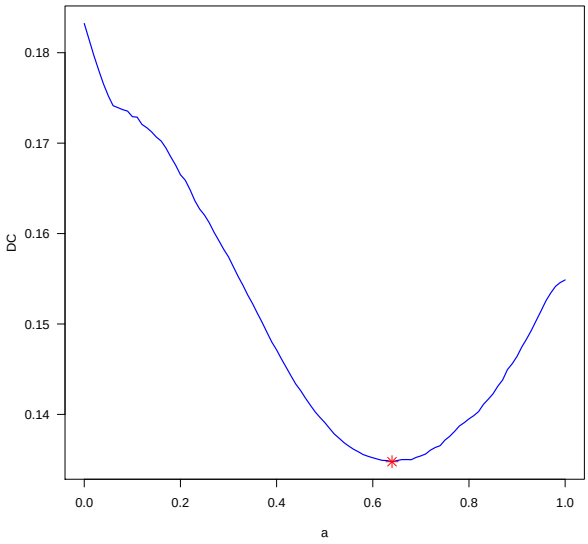


Figure 4.14: $a \mapsto \mathrm{DC}(a)$ for real data example on the interval $a \in (0, 1)$ by steps 0.01. Red star corresponds to the minimum DC value which occurs at $a = 0.64$.

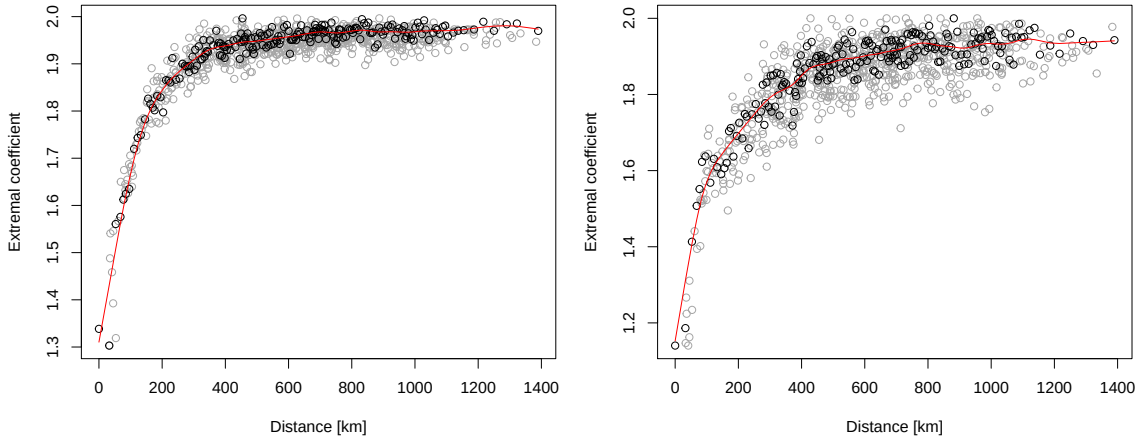


Figure 4.15: Nonlinear least squares estimates for $\theta_X(\boldsymbol{h})$ (left panel) and $\theta_Y(\boldsymbol{h})$ (right panel) as a function of distance $\|\boldsymbol{h}\|$. Gray points are pairwise estimates; black ones are binned estimates—the number of bins is 200. Red lines represent smoothed values of the binned estimates.

Although the estimates of $\theta_Y(\boldsymbol{h})$ appear to be relatively more "random" than those of $\theta_X(\boldsymbol{h})$ which might be expected since $\widehat{a} > 0.5$, this approach improves the empirical extremal summary measures for the class of max-mixture models. Hence, this approach could be used to guide the likelihood-based procedure: perform better model selection at this exploratory stage and better model checking at the validation stage.

In the left panel of Figure 4.15, asymptotic dependence between monitoring stations seems to be present up to a distance of 400 km and asymptotic independence could be conjectured

after. While the right panel seems to suggest the same for a distance of 600 km. Only a few number of pairs appear to be completely independent, i.e., $\widehat{\theta}_X(\boldsymbol{h})$ or/and $\widehat{\theta}_Y(\boldsymbol{h}) \to 2$. This means that for large distances there is still a weak extremal dependence. Such a process might be modeled by a BR or extremal-t processes with their inverted versions as asymptotically independent processes. The distribution of a BR process depends only on the semivariogram $\gamma(\boldsymbol{h})$. Although a wide range of $\gamma(\boldsymbol{h})$ can be employed, a common choice is the so-called stable semivariogram, sometimes called fractional Brownian motion (FBM) semivariogram, $\gamma(\boldsymbol{h}) = (\|\boldsymbol{h}\|/\phi)^{\kappa}$, $\phi > 0$, and $\kappa \in (0, 2]$. The variation in κ and ϕ yields max-stable processes with quite different roughnesses and scales for spatial dependence. The corresponding correlation function is $\rho(\boldsymbol{h}) = \exp(-\gamma(\|\boldsymbol{h}\|))$ which can be employed when modeling data with an extremal-t process. Under certain conditions, the extremal-t process converges to a BR process as the degrees of freedom $\nu \to \infty$. So, an extremal-t process will typically fit data at least as well as a BR process.

On the other hand, it appears that the multivariate distribution of monthly maxima of rainfall in this region might be poorly modeled by the Smith, Schlather, and TEG models. The Smith models are not flexible enough to cope with pairwise extremal coefficients that increase quickly around the origin but increase much more slowly for larger distances, whereas Schlather model is unable to capture the long-range independence, where the extremal dependence function is bounded above by $1 + \sqrt{1/2}$. Finally, the TEG model which can be adapted for modeling very local effects and sudden changes, or phenomena that reflect complete independence after some fixed lag, might be too restrictive for this application.

Besides performing better model selection, providing good starting values to maximize composite likelihoods may decrease the computational time and also improves the statistical efficiency, see [26]. So, the resulting estimate of the mixing parameter a could serve as a starting value for the optimization routine used to maximize the pairwise log-likelihood function. Also, some indications about range parameters could be obtained.

4.4.4 Fitting dependence parameters: pairwise likelihood estimation

To examine our conjecture in the preliminary analysis step about the suitable max-mixture class for this application, we consider four stationary isotropic max-mixture processes to assess the spatial dependence of monthly maxima rainfall events over the entire study region. In addition, we take into account the max-stable and inverted max-stable components of these models. The models are

(i) Class A: consists of max-mixture A_1−A_4 models.

- A_1 : a mixture which combines BR and an inverted BR processes with semivariograms $\gamma(\boldsymbol{h}) = (\|\boldsymbol{h}\|/\phi)^{\kappa}$, $\phi > 0$ and $\kappa \in (0, 2]$. The model parameters are $\boldsymbol{\vartheta} = (a, \phi_X, \kappa_X, \phi_Y, \kappa_Y)^t$.
- A_2 : a mixture which combines extremal-t and an inverted extremal-t processes with powered exponential correlation functions $\rho(\boldsymbol{h}) = \exp(-\gamma(\|\boldsymbol{h}\|))$. The model parameters are $\boldsymbol{\vartheta} = (a, \phi_X, \kappa_X, \nu_X, \phi_Y, \kappa_Y, \nu_Y)^t$.

- A_3 : a mixture which combines TEG and an inverted TEG processes with powered exponential correlation functions and with a compact random set $\mathcal{A}$ is a disk of fixed radius r. So, the expected volume of overlap between the random set $\mathcal{A}$ and itself shifted by the space lag $\boldsymbol{h} = \boldsymbol{s}_1 - \boldsymbol{s}_2$ is approximately $\alpha(\boldsymbol{h}) \simeq (1 - \|\boldsymbol{h}\|/2r)\mathbb{1}_{[0,2r]}$. The model parameters are $\boldsymbol{\vartheta} = (a, \phi_X, \kappa_X, r_X, \phi_Y, \kappa_Y, r_Y)^t$.
- A_4 : a mixture which combines Smith and an inverted Smith processes with Gaussian density kernels defined through the diagonal covariance matrices $\boldsymbol{\Sigma} = \phi^2\,\mathrm{Id}_2$, which corresponds to max-stable BR process when $\kappa = 2$, recall that Id_2 denotes a 2 by 2 identity matrix. The model parameters are $\boldsymbol{\vartheta} = (a, \phi_X, \phi_Y)^t$.

(ii) Class B: consists of models B_1−B_4 which are, respectively, the max-stable components in models A_1−A_4.

(iii) Class C: consists of models C_1−C_4 which are, respectively, the inverted max-stable components in models A_1−A_4.

The dependence parameters are estimated using the pairwise likelihood contribution (3.6). Equal weights are assumed. For evaluating the CLIC and the standard errors (S.E.), we assume that the monthly maxima are independent in time. In that case, we estimate $\boldsymbol{\mathcal{H}}$ and $\boldsymbol{\mathcal{J}}$ empirically, recall Section 3.1.3. Graphical diagnostics based on nonlinear least squares estimates of $\theta_X(\boldsymbol{h})$ and $\theta_Y(\boldsymbol{h})$ are also used to examine the suitability of the fitted models.

Table 4.2 reports the parameter estimates along with CLIC and S.E. values. As expected from the exploratory analysis, the hybrid dependence model A_2 has the lowest CLIC value and therefore would be considered the best of the candidate models for this dataset, closely followed by hybrid model A_1. Figure 4.16 depicts the fitted $\chi_{0.97}(\boldsymbol{h})$ and $\bar{\chi}_{0.97}(\boldsymbol{h})$ derived from model A_2. It seems that pairs of sites separated by a distance d smaller than 400 km are asymptotically dependent. At larger distances, the monthly maxima exhibit asymptotic independence. We also notice that there is a good agreement between the mixing parameter estimate $\widehat{a}$ obtained from these two models and the proposed F^{λ}-madogram approach. Furthermore, the Smith models (A_4, B_4 and C_4) are the worst ones. They have the largest values of CLIC.

For assessing the goodness-of-fit, Figures 4.17 and 4.18 compare the nonlinear pairwise extremal dependence estimates $\left(\widehat{\theta}_X(\boldsymbol{h}),\ \widehat{\theta}_Y(\boldsymbol{h})\right)$ as a function of distance $\|\boldsymbol{h}\|$ and their model-based counterparts derived from the fitted models. These figures confirm the conclusions we draw from Table 4.2. Models A_1 and A_2 provide plausible fits, the models A_4, B_4 and C_4 are too rigid to appropriately capture the decay of dependence with distance and model A_3 seems to overestimate the strength of short-range dependence and underestimate the strength of long-range dependence by its max-stable component (similar to model B_3), while it tends to overestimate the strength of dependence over the entire range by the inverted max-stable part (similar to model C_3).

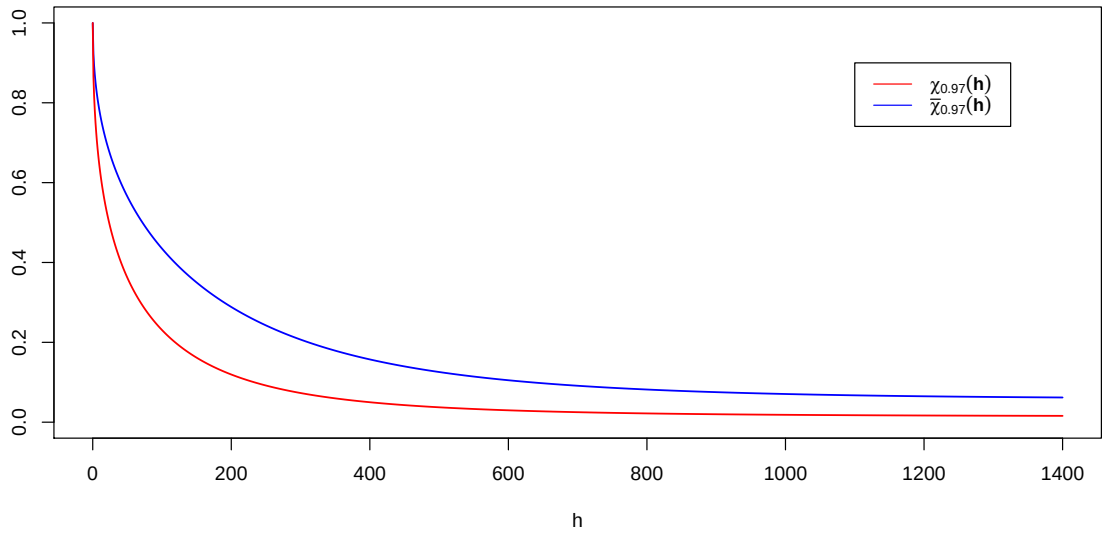


Figure 4.16: Fitted $\chi_{0.97}(\boldsymbol{h})$ and $\bar{\chi}_{0.97}(\boldsymbol{h})$ derived from model A_2 as functions of the distance $h = \|\boldsymbol{h}\|$.

Table 4.2: Estimated dependence parameters for all selected models fitted to Australian winter rainfall monthly maxima. The CLIC and S.E. values are reported. ^{a} indicates to the lower CLIC.

Class A	$\widehat{a}$	$\widehat{\phi}_X$ [km]	$\widehat{\kappa}_X$	$\widehat{\nu}_X$	$\widehat{r}_X$ [km]	$\widehat{\phi}_Y$ [km]	$\widehat{\kappa}_Y$	$\widehat{\nu}_Y$	$\widehat{r}_Y$ [km]	CLIC
A_1	0.61	31.42	0.72	-	-	81.07	0.84	-	-	62926
S.E.	0.04	5.93	0.11	-	-	42.86	0.20	-	-	-
A_2	0.73	358.14	1.09	7.31	-	796.13	1.65	5.58	-	62902^a
S.E.	0.14	81.05	0.29	2.52	-	123.44	0.29	3.06	-	-
A_3	0.18	116.37	0.74	-	222.15	367.48	1.68	-	709.73	64463
S.E.	0.09	24.99	0.40	-	96.33	82.00	0.56	-	63.65	-
A_4	0.42	29.72	-	-	-	83.50	-	-	-	68859
S.E.	0.12	23.60	-	-	-	35.19	-	-	-	-
Class B	$\widehat{\phi}_X$	$\widehat{\kappa}_X$	$\widehat{\nu}_X$	$\widehat{r}_X$						CLIC
B_1	44.14	0.86	-	-						63109
S.E.	16.03	0.26	-	-						-
B_2	269.60	0.82	7.23	-						62985
S.E.	41.62	0.10	1.95	-						-
B_3	212.49	1.08	-	263.64						68705
S.E.	17.12	0.37	-	116.00						-
B_4	77.02	-	-	-						68721
S.E.	9.39	-	-	-						-
Class C	$\widehat{\phi}_Y$	$\widehat{\kappa}_Y$	$\widehat{\nu}_Y$	$\widehat{r}_Y$						CLIC
C_1	137.55	1.13	-	-						63650
S.E.	60.45	0.07	-	-						-
C_2	461.87	1.58	3.99	-						63672
S.E.	165.08	0.47	0.89	-						-
C_3	288.33	1.65	-	672.05						64413
S.E.	75.70	0.01	-	54.06						-
C_4	69.86	-	-	-						68838
S.E.	26.94									-

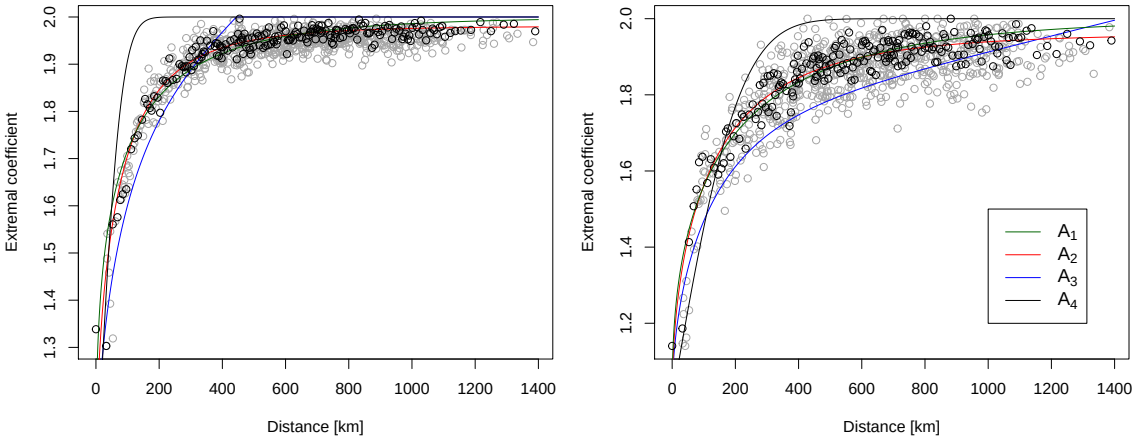


Figure 4.17: Site-wise winter monthly maxima analysis (Class A): nonlinear least squares and model-based estimates of the extremal dependence functions $\theta_X(\boldsymbol{h})$ (left panel) and $\theta_Y(\boldsymbol{h})$ (right panel). Gray points are nonlinear least squares pairwise estimates, black ones are binned estimates—the number of bins is 200.

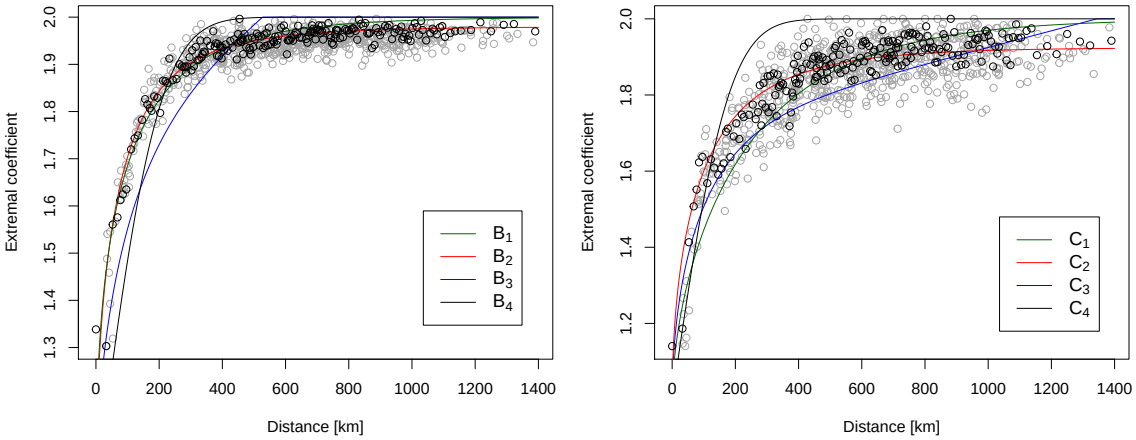


Figure 4.18: Site-wise winter monthly maxima analysis (Classes B and C): nonlinear least squares and model-based estimates of the extremal dependence functions $\theta_X(\boldsymbol{h})$ (left panel) and $\theta_Y(\boldsymbol{h})$ (right panel). Gray points are nonlinear least squares pairwise estimates, black ones are binned estimates—the number of bins is 200.

Finally, on the basis of the proposed F^{λ}-madogram inference approach, we estimate the full process by contrasting the estimated nonlinear pairwise summaries $\widehat{\theta}_X(\boldsymbol{h})$ and $\widehat{\theta}_Y(\boldsymbol{h})$ as a function of $\|\boldsymbol{h}\|$ with their parametric counterparts from the best-fitting model A_2. More precisely, we apply the following nonlinear least squares minimization procedure:

$$\begin{pmatrix}\widehat{\phi}\\ \widehat{\kappa}\\ \widehat{\nu}\end{pmatrix} = \underset{\substack{\phi>0,\ \kappa\in(0,2],\\ \nu\geq 1}}{\arg\min} \sum_{\|\boldsymbol{h}_k\|} \omega(\boldsymbol{h}_k)\left[\widehat{\theta}(\boldsymbol{h}_k) - 2T_{\nu+1}\left(\sqrt{(\nu+1)\frac{1-\rho(\boldsymbol{h}_k)}{1+\rho(\boldsymbol{h}_k)}}\right)\right]^2,\ k=1,\ldots,703. \tag{4.8}$$

The resulting estimates are: $\widehat{\phi}_X = 350.69$ km, $\widehat{\kappa}_X = 1.22$, $\widehat{\nu}_X = 6.08$, $\widehat{\phi}_Y = 738.35$ km, $\widehat{\kappa}_Y = 1.77$ and $\widehat{\nu}_Y = 4.46$. Overall, there is a good agreement between these estimates and the ones obtained by model A_2.

4.4.5 Threshold exceedances probability

One of the most important weather hazards is heavy rainfall. Extreme rainfall may cause flooding or landslides which in turn threatens the lives of organisms, disrupt transportation and communication, damage infrastructure and so forth. Extreme rainfall observations show a spatial dependence structure, where neighboring sites within some distance exhibit similar patterns. For example, Figure 4.19 displays the scatterplot of winter monthly maxima measured over the period 1972-2014 between two adjacent monitoring stations (distance $d \approx 37$ km). It seems that the dependence structure is strong (the points are strongly clustered). Furthermore, large values of winter monthly maxima rainfall tend to occur together. The aim of spatial analysis is usually to enable prediction at an unobserved location. Hence, quantifying the dependence structure across a sample of stations may be useful in estimating the extreme value behavior at other unobserved stations.

As the last step in this study, in order to illustrate the behavior of the best-fitting model A_2 and the F^{λ}-madogram estimation approach, we perform a small prediction in the context of the rainfall example. Denote by $Z_M(\boldsymbol{s})$ the site-wise monthly maxima process of the max-mixture rainfall process $Z(\boldsymbol{s})$. An interesting question might be "given that the process $Z_M(\boldsymbol{s})$ at location $\boldsymbol{s}_2$ exceeds a high threshold z, how likely is it high at a 'nearby' location $\boldsymbol{s}_1$?". In other words, how can we predict

$$\mathcal{P}(z) = \mathbb{P}[Z_M(\boldsymbol{s}_1) > z | Z_M(\boldsymbol{s}_2) > z]?$$

Indeed, the probability $\mathcal{P}$ can be viewed as an alarm for the risk of extreme rainfall. In the following lemma, we derive the probability $\mathcal{P}$ for the max-mixture process $Z(\boldsymbol{s})$.

Lemma 4.1. *The bivariate conditional exceedance probability of a max-mixture process* $\{Z(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ *is given by*

$$\mathcal{P}(z) = \frac{1 - 2e^{-\frac{1}{z}} + e^{-\frac{a\theta_X(\boldsymbol{h})}{z}}\left\{-1 + 2e^{-\frac{1-a}{z}} + \left[1 - e^{-\frac{1-a}{z}}\right]^{\theta_Y(\boldsymbol{h})}\right\}}{1 - e^{-\frac{1}{z}}}, \quad a \in [0,1]. \tag{4.9}$$

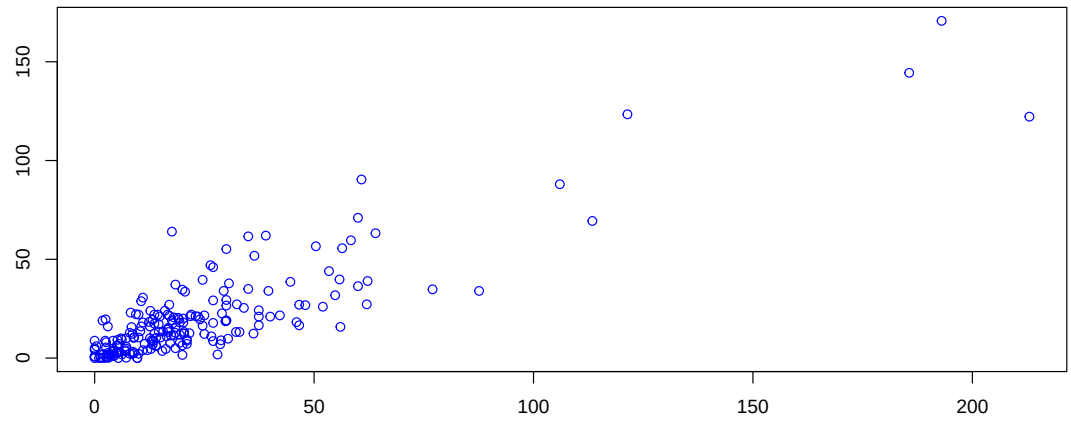


Figure 4.19: Australian rainfall example—scatterplot of winter monthly maxima (measured in mm) recorded during 1972−2014 between two adjacent monitoring stations. The two stations are separated by a distance $d \approx 37$ km.

Proof. We may write the joint survivor function of the spatial max-mixture process Z in (2.34)

$$\begin{aligned}\mathbb{P}[Z(\boldsymbol{s}_1) > z_1, Z(\boldsymbol{s}_2) > z_2] &= \bar{F}_Z(z_1, z_2) \\ &= 1 - F_Z(z_1) - F_Z(z_2) + F_Z(z_1, z_2),\end{aligned}$$

where $F_Z(z)$ and $F_Z(z_1, z_2)$ denote, respectively, the marginal probability distribution function (i.e., unit Fréchet) and the bivariate c.d.f. of the max-mixture process $Z(.)$. Applying (2.35) and setting $z_1 = z_2 = z$, we obtain

$$\begin{aligned}\mathbb{P}[Z(\boldsymbol{s}_1) > z | Z(\boldsymbol{s}_2) > z] &= \frac{\mathbb{P}[Z(\boldsymbol{s}_1) > z, Z(\boldsymbol{s}_2) > z]}{\mathbb{P}[Z(\boldsymbol{s}_2) > z]} \\ &= \frac{1 - 2\mathbb{P}(Z(\boldsymbol{s}_1) < z) + F^X_{\boldsymbol{h}}\left(\frac{z}{a}, \frac{z}{a}\right) F^Y_{\boldsymbol{h}}\left(\frac{z}{1-a}, \frac{z}{1-a}\right)}{1 - \mathbb{P}[Z(s_2) \leq z]} \\ &= \frac{1 - 2e^{-\frac{1}{z}} + e^{-aV^X_{\boldsymbol{h}}(z,z)}\left\{-1 + 2e^{-\frac{1-a}{z}} + e^{-V^Y_{\boldsymbol{h}}\left[t\left(\frac{1-a}{z}\right), t\left(\frac{1-a}{z}\right)\right]}\right\}}{1 - e^{-\frac{1}{z}}},\end{aligned}$$

where $t\left(\frac{1-a}{z}\right) = -1/\log\left[1 - e^{-\frac{1-a}{z}}\right]$, $F^X_{\boldsymbol{h}}\left(\frac{z}{a}, \frac{z}{a}\right) = \mathbb{P}\left(X(\boldsymbol{s}_1) < \frac{z}{a}, X(\boldsymbol{s}_2) < \frac{z}{a}\right)$ and $F^Y_{\boldsymbol{h}}\left(\frac{z}{1-a}, \frac{z}{1-a}\right) = \mathbb{P}\left(Y(\boldsymbol{s}_1) < \frac{z}{1-a}, Y(\boldsymbol{s}_2) < \frac{z}{1-a}\right)$.

As a consequence of the homogeneity property of the exponent functions $V^X_{\boldsymbol{h}}(.)$ and $V^Y_{\boldsymbol{h}}(.)$, we have $V^X_{\boldsymbol{h}}(z, z) = \theta_X / z$ and $e^{-V^Y_{\boldsymbol{h}}\left[t\left(\frac{1-a}{z}\right), t\left(\frac{1-a}{z}\right)\right]} = \left[1 - e^{-\frac{1-a}{z}}\right]^{\theta_Y}$, and this gives (4.9). □

In what follows, let $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ denote, respectively, the unused stations (which are shown by the numbers $\{1, 2, 3\}$, see Figure 4.10) and the closest adjacent (nearby) stations to $\boldsymbol{s}_1$ and let

$d = \|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$. Taking site-wise monthly maxima data, we estimate $\mathcal{P}$ by model A_2 and the F^λ-madogram approach. Then, the obtained results are compared with the empirical counterparts of $\mathcal{P}$. The threshold z in (4.9) is taken corresponding to the q-empirical quantile at station $\boldsymbol{s}_2$, $q \in (0,1)$. Specifically, $q \in \{0.75, 0.80, 0.85, 0.90, 0.925, 0.95\}$. For the F^λ-madogram approach, under the isotropy hypothesis, one may extrapolate $\theta_X(\boldsymbol{h})$ and $\theta_Y(\boldsymbol{h})$ when $\|\boldsymbol{h}\| = d$ from the (smoothed) binned versions of the nonlinear least squares pairwise estimates by distance classes.

The diagnostic probability plots in Figure 4.20 compare the resulting estimates of $\mathcal{P}$ using model A_2 and the F^λ-madogram approach with their empirical counterparts, where the empirical estimates of $\mathcal{P}$ are computed on the basis of the observed data at the unused monitoring stations. There is a good agreement overall, as the points are sufficiently close to linearity. We also observe a slight outperformance for model A_2.

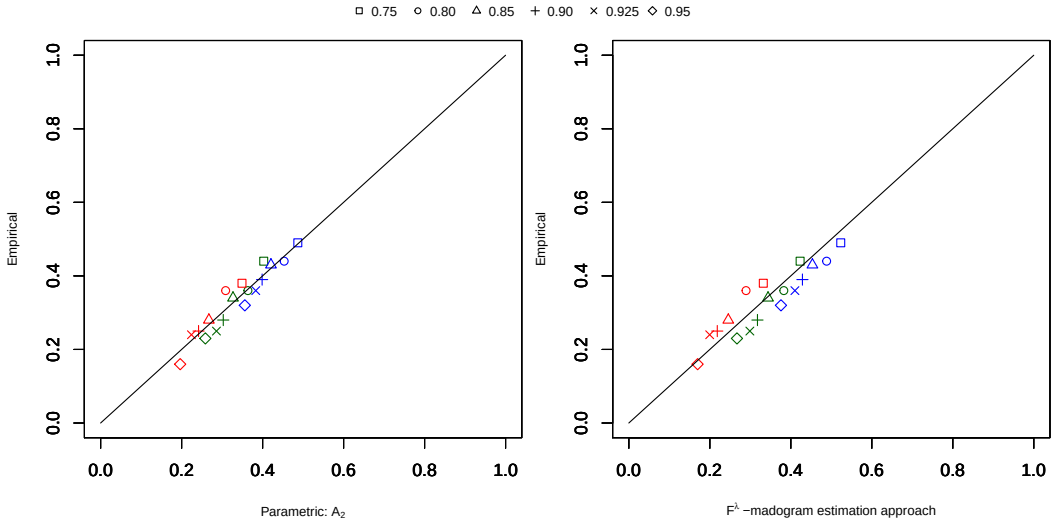


Figure 4.20: Comparison between empirical estimates of $\mathcal{P}$ at the unused monitoring stations for Australian monthly maxima rainfall data example and their model-based counterparts on the basis of the best-fitting parametric model A_2 (left panel) and their F^λ-madogram approach counterparts (right panel). A perfect agreement would place all points on the black diagonal line. Green: station 1; red: station 2; blue: station 3 , see Figure 4.10.

4.5 Conclusions

To sum up, semi/nonparametric estimation methods are valuable for the initial analysis and for checking the adequacy of fitted models. Most approaches are based on pairs of extremal observations, which provide direct insight into the dependence structure of the data. In this chapter, we developed an estimation approach for max-mixture models, which combine max-stable processes and inverted max-stable processes into a single family, in order to capture both asymptotic dependence and independence in spatial extremes. We suggested using F^λ-madogram approach to gain in flexibility.

The proposed approach can be viewed as a compromise (or an intermediate approach) between a completely nonparametric approach based on empirical extremal dependence functions (which have high variability and do not yield a valid spatial model) and a fully parametric approach (which is very efficient and yields a valid spatial model).

A simulation study has shown that our estimation procedure performs well, even when we have considered the boundary values for a. It improves empirical dependence functions for the class of max-mixture models. In the rainfall example, we used our estimation approach in order to perform better model selection at an exploratory stage and better model checking at the validation stage. Furthermore, we estimated the bivariate conditional exceedance probability of monthly maxima rainfall at ungauged locations.

Finally, isotropic models have been widely studied. Owing to our motivating isotropic rainfall dataset, we focused on fitting isotropic max-mixture processes. However, anisotropic situations are often observed in environmental phenomena, especially when the study regions are very large. So, it could be interesting to introduce anisotropy to our inference framework, see Section 6.2.

Chapter 5

Semi-parametric estimation for space-time max-stable processes

This chapter is based on our results from the submitted paper: A. Abu-Awwad, V. Maume-Deschamps and P. Ribereau—Semi-parametric estimation for space-time max-stable processes: F-madogram-based estimation approach.

In the literature, the observations at spatial locations are often assumed to be temporally independent. So, only the spatial structure is studied, see, e.g., [84]. This may be unrealistic. For example, hourly and even daily measurements of rainfall are rarely independent in time, where the chance of a day being free of rain may be substantially greater if the previous few days were dry rather than wet. Thus, space-time max-stable processes have been developed to study the extremal dependence of spatio-temporal data, see [40, 41, 64, 54].

In the spatial setting, the λ-madogram (recall Definition 2.31) has been proposed by [81] to estimate the extremal dependence function $V_{\boldsymbol{h}}(x_1, x_2)$. Its simplest version F-madogram [35] (recall Definition 2.26) is used to estimate the extremal dependence function $\theta(\boldsymbol{h}) = V_{\boldsymbol{h}}(1,1)$. Recently, a nonlinear least squares approach based on the F-madogram has been applied satisfactorily for statistical inference on spatial max-mixture models, see [5]. In this chapter, we broaden the use of the F-madogram as an inferential tool to the spatio-temporal setting. Specifically, we propose the following two novel semi-parametric estimation schemes based on spatio-temporal F-madogram to fit space-time max-stable processes:

(i) ***Scheme 1:*** we estimate spatial and temporal parameters separately. Based on nonlinear least squares, we minimize the squared difference between the empirical estimates of spatial/temporal F-madograms and their model-based counterparts. Our inferential methodology is close to the one that has been proposed by [23] as an alternative or a preliminary analysis to the pairwise likelihood approach in [41], where only isotropic space-time max-stable BR process (2.49) has been fitted.

(ii) ***Scheme 2:*** we generalize the nonlinear least squares to estimate spatial and temporal parameters simultaneously.

These methods can be viewed as similar in spirit to those used in the geostatistics literature in the context of nonlinear least squares fitting of the semivariogram, see, e.g., [80]. For instance, in

the spatial setting, let $\widehat{\gamma}(\boldsymbol{h}) = (\widehat{\gamma}(\boldsymbol{h}_1), \ldots, \widehat{\gamma}(\boldsymbol{h}_k))^t$ denotes the $k \times 1$ vector containing the values of the spatial empirical semivariogram (recall Definition 2.5) at the k lag distances considered and let $\gamma(\boldsymbol{h}, \boldsymbol{\psi})$ be the vector of the values derived by the theoretical model from the same lags, where $\boldsymbol{\psi}$ denotes the vector gathering the parameters to be estimated. Then, one may write

$$\widehat{\gamma}(\boldsymbol{h}) = \gamma(\boldsymbol{h}, \boldsymbol{\psi}\}) + \epsilon(\boldsymbol{h}, \boldsymbol{\psi}), \tag{5.1}$$

where $\epsilon(.)$ has $\mathbf{0}$ mean, provided that $\widehat{\gamma}(\boldsymbol{h})$ is unbiased estimator of $\gamma(\boldsymbol{h})$. Its covariance matrix typically depends on $\boldsymbol{\psi}$. Let

$$Q(\boldsymbol{\psi}) = \sum_{i=1}^{k} \omega_i \left(\widehat{\gamma}(\boldsymbol{h}_i) - \gamma(\boldsymbol{h}_i, \boldsymbol{\psi})\right)^2, \tag{5.2}$$

where ω_i are nonnegative weights specifying the contribution of each pair. The (weighted) nonlinear least squares fitting aims to determine the vector $\widehat{\boldsymbol{\psi}}$ that minimizes the expression (5.2), that is,

$$\widehat{\boldsymbol{\psi}} = \underset{\boldsymbol{\psi} \in \boldsymbol{\Psi}}{\arg\min}\, Q(\boldsymbol{\psi}), \tag{5.3}$$

where $\boldsymbol{\Psi}$ is a set of admissible parameters. So, it is natural to adapt the strategies introduced in geostatistics to this context.

The remainder of the chapter is organized as follows. In Section 5.1, we derive a closed form expression of the spatio-temporal λ-madogram and Pickands dependence function in the framework of spectrally separable space-time max-stable processes. Two semi-parametric estimation schemes for space-time max-stable are described in Section 5.2. Section 5.3 illustrates the performance of our method through a number of simulation studies, where also a comparison with the semi-parametric estimation [23] is performed. In Section 5.4, we apply our inferential procedure to quantify the extremal behavior of radar rainfall data in a region in the State of Florida. Concluding remarks are given in Section 5.5.

5.1 Extensions of spatial extremal summary measures

In the next two propositions, we derive explicit formulas for the λ-madogram and Pickands dependence function in the case of the spectrally separable spatio-temporal model (recall Section 2.3.2), i.e.,

$$X(\boldsymbol{s}, t) = \max\{\delta X(\boldsymbol{s} - \boldsymbol{\tau}, t-1), (1-\delta)H(\boldsymbol{s}, t)\}, \ (\boldsymbol{s}, t) \in \mathcal{S} \times \mathcal{T}. \tag{5.4}$$

Proposition 5.1. *(λ-madogram for space-time max-stable process with spectral separability) In the case of (5.4), for $\boldsymbol{h} \in \mathbb{R}^2$ and $l \in \mathbb{R}$, the spatio-temporal λ-madogram for any $\lambda \in (0,1)$ is given by*

$$\nu_\lambda(\boldsymbol{h}, l) = \frac{(1-\lambda)\{V_{\mathbf{0},\boldsymbol{h}-l\boldsymbol{\tau}}\left(\lambda, (1-\lambda)\delta^{-l}\right)\} + 1 - \delta^l}{(1-\lambda)\{1 + V_{\mathbf{0},\boldsymbol{h}-l\boldsymbol{\tau}}\left(\lambda, (1-\lambda)\delta^{-l}\right)\} + 1 - \delta^l} - c(\lambda), \tag{5.5}$$

where $c(\lambda) = \frac{3}{2(1+\lambda)(2-\lambda)}$.

Proof. We follow the idea of the proof from [81, 60]. In the case of (5.4), the spatio-temporal λ-madogram, stemming from the spatial version (recall Definition 2.14) is defined by

$$\nu_\lambda(\boldsymbol{h}, l) = \frac{1}{2}\mathbb{E}[|F^\lambda(X(\mathbf{0},0)) - F^{1-\lambda}(X(\boldsymbol{h},l))|],$$

where F is the unit Fréchet distribution function. Applying the equation $|a-b| = 2\max\{a,b\} - (a+b)$ with $a = F^\lambda\{X(\mathbf{0},0)\}$ and $b = F^{1-\lambda}\{X(\boldsymbol{h},l)\}$ and recalling that $\mathbb{E}[F^\alpha\{X(\boldsymbol{h},l)\}] = 1/(1+\alpha)$ for any $\alpha > 0$, we have

$$\begin{aligned}\nu_\lambda(\boldsymbol{h}, l) &= \mathbb{E}\left[\max\{F^\lambda\{X(\mathbf{0},0)\}, F^{1-\lambda}\{X(\boldsymbol{h},l)\}\}\right] \\ &\quad - \frac{1}{2}\mathbb{E}\left[F^\lambda\{X(\mathbf{0},0)\} + F^{1-\lambda}\{X(\boldsymbol{h},l)\}\right] \\ &= \mathbb{E}\left[\max\{F^\lambda\{X(\mathbf{0},0)\}, F^{1-\lambda}\{X(\boldsymbol{h},l)\}\}\right] - c(\lambda).\end{aligned}$$

Considering the random variable $W = \max\left\{F^\lambda\{X(\mathbf{0},0)\}, F^{1-\lambda}\{X(\boldsymbol{h},l)\}\right\}$. The probability distribution function G of W satisfies (recall (2.57))

$$\begin{aligned}G(x) &= \mathbb{P}[W \leq x] \\ &= \mathbb{P}\left[X(\mathbf{0},0) \leq -\frac{\lambda}{\log x}, X(\boldsymbol{h},l) \leq -\frac{1-\lambda}{\log x}\right] \\ &= \exp\left\{\log x\left(V_{\mathbf{0},\boldsymbol{h}-l\boldsymbol{\tau}}\left(\lambda,(1-\lambda)\delta^{-l}\right) + \frac{1-\delta^l}{1-\lambda}\right)\right\}.\end{aligned}$$

Differentiating G with respect to x gives the p.d.f. of the random variable W. So, we obtain

$$\mathbb{E}[W] = \int_0^1 \Omega(.)\exp\{\log x\,\Omega(.)\}\,dx = \frac{\Omega(.)}{1+\Omega(.)},$$

where $\Omega(.) = V_{\mathbf{0},\boldsymbol{h}-l\boldsymbol{\tau}}\left(\lambda,(1-\lambda)\delta^{-l}\right) + \frac{1-\delta^l}{1-\lambda}$. This yields the spatio-temporal λ-madogram (5.5). □

Unlike the extremal coefficient θ and thus the F-madogram, this measure provides information on $V_{\mathbf{0},\boldsymbol{h}-l\boldsymbol{\tau}}(x_1,x_2)$ for $x_1 \neq x_2$. Hence, exploring the whole space.

Remark 5.1. *Based on expression (5.5), we easily rediscover the expressions of the following special cases:*

(i) Setting $\lambda = 0.5$, we obtain $\nu_{0.5}(\boldsymbol{h},l) = \frac{2\theta(\boldsymbol{h},l)}{1+2\theta(\boldsymbol{h},l)} - \frac{2}{3}$, where $\theta(\boldsymbol{h},l) = V_{\mathbf{0},\boldsymbol{h}-l\boldsymbol{\tau}}\left(1,\delta^{-l}\right) + 1 - \delta^l$, recall (2.58). Using the one-to-one relationship between the spatio-temporal extremal dependence function and the spatio-temporal F-madogram, that is, $\theta(\boldsymbol{h},l) = 1 + 2\nu_F(\boldsymbol{h},l)/1 - 2\nu_F(\boldsymbol{h},l)$, yields $\nu_{0.5}(\boldsymbol{h},l) = \frac{8\nu_F(\boldsymbol{h},l)}{3\{3+2\nu_F(\boldsymbol{h},l)\}}$. This induces that the spatio-temporal F-madogram is a special case of the λ-madogram when $\lambda = 0.5$.

(ii) Setting the time lag $l = 0$, we rediscover the formula of spatial λ-madogram $\nu_\lambda(\boldsymbol{h})$ as calculated in [81].

Proposition 5.2. ***(Pickands dependence function for space-time max-stable process with spectral separability)*** *In the case of (5.4), for* $\boldsymbol{h} \in \mathbb{R}^2$ *and* $l \in \mathbb{R}$*, the spatio-temporal Pickands dependence function* $A_{\boldsymbol{h},l}$ *for any* $\varphi \in (0,1)$ *is given by*

$$A_{\boldsymbol{h},l}(\varphi) = \varphi(1-\varphi)\left[V_{\mathbf{0},\boldsymbol{h}-l\boldsymbol{\tau}}\left(\varphi, \frac{1-\varphi}{\delta^l}\right) + \frac{1-\delta^l}{1-\varphi}\right], \tag{5.6}$$

where $V_{\mathbf{0},\boldsymbol{h}-l\boldsymbol{\tau}}$ *is the exponent function characterizing the spatial component.*

Proof. The proof can be easily deduced using the spatio-temporal version of the Pickands dependence function A (2.48) and then setting $x_1 = \varphi$ and $x_2 = 1-\varphi$ in the bivariate exponent function of $(X(\mathbf{0},0), X(\boldsymbol{h},l))$ (2.57). □

Remark 5.2. *In the case of (5.4), it is easy to show that the extremal dependence function* $\theta(\boldsymbol{h},l)$ *(2.58) is a special case of the Pickands dependence function* $A_{\boldsymbol{h},l}(\varphi)$ *(5.6) with* $\varphi = 1/2$*, i.e.,*

$$2A_{\boldsymbol{h},l}\left(\frac{1}{2}\right) = \theta(\boldsymbol{h},l) =: V_{\mathbf{0},\boldsymbol{h}-l\boldsymbol{\tau}}\left(1, \delta^{-l}\right) + 1 - \delta^l.$$

5.2 Statistical inference for space-time max-stable processes

In what follows, we shall denote, respectively, by $h = \|\boldsymbol{h}\| =: \|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$, $\boldsymbol{h} \in \mathbb{R}^2$ and $l' = |l| =: |t_1 - t_2|$, $l \in \mathbb{R}$ the Euclidean norm of spatial lag $\boldsymbol{h}$ and the absolute value of temporal lag l.

We now describe two semi-parametric estimation schemes for space-time max-stable processes (recall Section 2.3) based on a closed form expression of the spatio-temporal F-madogram (recall Definition 2.17), which stems from a classical geostatistical tool; the madogram [77]. It has a clear link with extreme value theory throughout the spatio-temporal extremal dependence function $\theta(.)$, i.e.,

$$\nu_F(\boldsymbol{h},l) = \frac{1}{2} - \frac{1}{\theta(\boldsymbol{h},l)+1}. \tag{5.7}$$

In practice, measurements are typically taken at various locations, sometimes on a grid, and at regularly spaced time intervals. In the following, the process $X := \{X(\boldsymbol{s},t) : (\boldsymbol{s},t) \in \mathcal{S}\times\mathcal{T}\}$ is assumed to be a stationary space-time max-stable process. It is observed on locations assumed to lie on a regular 2-dimensional (2D) grid, i.e.,

$$S_n = \left\{\boldsymbol{s}_i : i = 1,\ldots,n^2\right\} = \{(x,y),\ x,y \in \{1,\ldots,n\}\},$$

and at equidistant time moments, given by $\{t_1,\ldots,t_T\} = \{1,\ldots,T\}$. This sampling scheme has been adopted in various studies in the literature, see, e.g., [41, 24, 23]. For statistical inference on the process X, we develop the following two semi-parametric estimation schemes.

5.2.1 Scheme 1

Let $\boldsymbol{\psi} = (\boldsymbol{\psi}^{(s)}, \boldsymbol{\psi}^{(t)})$ denotes the vector gathering the parameters of the process X to be estimated, where $\boldsymbol{\psi}^{(s)}$ and $\boldsymbol{\psi}^{(t)}$ denote, respectively, the vectors gathering the spatial and temporal parameters. In this scheme, we consider how the process evolves at given time of reference (a merely

spatial process), and its evolution over time at a given location (a merely temporal process). So, $\boldsymbol{\psi}^{(s)}$ and $\boldsymbol{\psi}^{(t)}$ can be estimated separately. More precisely, denote by $\mathcal{H} \subset [0,\infty)$ and $\mathcal{K} \subset [0,\infty)$ finite sets of spatial and temporal lags on which the estimation is performed. Let the set $\mathcal{B}_h$ summarizes all pairs of $\mathcal{S}_n$ which give rise to the same spatial lag $h \in \mathcal{H}$, i.e.,

$$\mathcal{B}_h = \{(\ell, p) \in \{1,\ldots,n^2\}^2 : \|\boldsymbol{s}_\ell - \boldsymbol{s}_p\| = \|\boldsymbol{h}\| = h\}.$$

The inferential methodology is summarized in the following steps:

(i) As a first step, we estimate the purely spatial/temporal F-madogram nonparametrically by the empirical version. Denote by $\widehat{\nu}_F^{(t)}(\boldsymbol{h})$, $\|\boldsymbol{h}\| \in \mathcal{H}$ $\left(\text{respectively, } \widehat{\nu}_F^{(s)}(l'),\ l' \in \mathcal{K}\right)$ the nonparametric estimate of the purely spatial (respectively, temporal) F-madogram. As is standard in geostatistics, we compute $\widehat{\nu}_F^{(t)}(\boldsymbol{h})$ from the empirical spatio-temporal F-madogram $\widehat{\nu}_F(\boldsymbol{h}, l)$ at spatio-temporal distances $(\boldsymbol{h}, 0)$, that is for all $\{t_1,\ldots,t_T\}$,

$$\widehat{\nu}_F^{(t)}(\boldsymbol{h}) = \widehat{\nu}_F(\boldsymbol{h},0) = \frac{1}{2|\mathcal{B}_h|} \sum_{\substack{p=1}}^{n^2} \sum_{\substack{\ell=1 \\ \|\boldsymbol{s}_\ell - \boldsymbol{s}_p\| = \|\boldsymbol{h}\| = h}}^{n^2} |F\{X(\boldsymbol{s}_\ell, t)\} - F\{X(\boldsymbol{s}_p, t)\}|,\ h \in \mathcal{H},$$

where $|.|$ denotes the cardinality of the set B_h and F is the standard Fréchet probability distribution function. Let us remark that, a similar estimator in the framework of λ-madogram has been adopted by [81] in an analysis of Bourgogne (France) annual maxima of daily rainfall measurements recorded over a period of 51 years (1953-2003) at 146 locations. In this study, the authors focused on a single realization of a random field of maxima computed over 51 years instead of 51 random fields of maxima.

On the other hand, $\widehat{\nu}_F^{(s)}(l')$ is computed from the empirical spatio-temporal F-madogram $\widehat{\nu}_F(\boldsymbol{h}, l')$ at spatio-temporal distances $(\boldsymbol{0}, l')$, that is for all $\boldsymbol{s} \in \mathcal{S}_n$,

$$\widehat{\nu}_F^{(s)}(l') = \widehat{\nu}_F(\boldsymbol{0},l') = \frac{1}{2(T-l')} \sum_{k=1}^{T-l'} |F\{X(\boldsymbol{s}, t_k)\} - F\{X(\boldsymbol{s}, t_{k+l'})\}|,\ l' \in \mathcal{K}.$$

(ii) Then, the overall purely spatial (respectively, temporal) F-madogram estimates $\widehat{\nu}_F(\boldsymbol{h})$ (respectively, $\widehat{\nu}_F(l')$) are computed from the averages over the temporal moments (respectively, the spatial locations). More precisely,

$$\widehat{\nu}_F(\boldsymbol{h}) = \frac{1}{T} \sum_{\substack{k=1 \\ \|\boldsymbol{h}\| = h}}^{T} \widehat{\nu}_F^{(t_k)}(\boldsymbol{h}),\ h \in \mathcal{H}. \tag{5.8}$$

$$\widehat{\nu}_F(l') = \frac{1}{n^2} \sum_{\ell=1}^{n^2} \widehat{\nu}_F^{(\boldsymbol{s}_\ell)}(l'),\ l' \in \mathcal{K}. \tag{5.9}$$

(iii) Finally, a nonlinear least squares procedure is applied to estimate the parameters of interest.

$$\widehat{\boldsymbol{\psi}}^{(s)} = \underset{\boldsymbol{\psi}^{(s)} \in \boldsymbol{\Psi}^{(s)}}{\arg\min} \sum_{\|\boldsymbol{h}\| = h \in \mathcal{H}} \omega^{\boldsymbol{h}} \left(\widehat{\nu}_F(\boldsymbol{h}) - \nu_F^{(\boldsymbol{s})}(\boldsymbol{h}, \boldsymbol{\psi}^{(s)})\right)^2,\ h \in \mathcal{H}, \tag{5.10}$$

$$\widehat{\boldsymbol{\psi}}^{(t)} = \underset{\boldsymbol{\psi}^{(t)} \in \boldsymbol{\Psi}^{(t)}}{\arg\min} \sum_{l' \in \mathcal{K}} \omega^{l'} \left(\widehat{\nu}_F(l') - \nu_F^{(t)}(l', \boldsymbol{\psi}^{(t)}) \right)^2, \; l' \in \mathcal{K}, \tag{5.11}$$

where $\nu_F^{(s)}(\boldsymbol{h}, \boldsymbol{\psi}^{(s)}) = \nu_F(\boldsymbol{h}, 0, \boldsymbol{\psi}^{(s)})$ and $\nu_F^{(t)}(l', \boldsymbol{\psi}^{(t)}) = \nu_F(\mathbf{0}, l', \boldsymbol{\psi}^{(t)})$ denote, respectively, the spatial and temporal model-based F-madogram counterparts. $\omega^{\boldsymbol{h}} \geq 0$ and $\omega^{l'} \geq 0$ denote, respectively, the spatial and temporal weights. Since it is anticipated that the spatio-temporal pairs which are far away in space or in time, have only little influence on the dependence parameters to be estimated, a simple choice for these weights is $\omega^{\boldsymbol{h}} = \mathbb{1}_{\{\|\boldsymbol{h}\| \leq r\}}$, $\omega^{l'} = \mathbb{1}_{\{l' \leq q\}}$, where $\mathbb{1}(.)$ denotes the indicator function and (r, q) to be specified, recall Section 3.1.

Note that the setup of the inferential methodology in Scheme 1 is close to the one proposed in [23], in which the spatio-temporal extremogram was adopted. The latter is defined for a stationary space-time process $X := \{X(\boldsymbol{s}, t) : (\boldsymbol{s}, t) \in \mathcal{S} \times \mathcal{T}\}$ by

$$\rho_{\mathscr{A}_1, \mathscr{A}_2}(\boldsymbol{h}, l) = \lim_{x \to \infty} \frac{\mathbb{P}\left\{x^{-1} X(\boldsymbol{s}, t) \in \mathscr{A}_1, x^{-1} X(\boldsymbol{s} + \boldsymbol{h}, t + l) \in \mathscr{A}_2\right\}}{\mathbb{P}\{x^{-1} X(\boldsymbol{s}, t) \in \mathscr{A}_1\}}, \; \boldsymbol{h} \in \mathbb{R}^d, \; l \in \mathbb{R}. \tag{5.12}$$

[23] considered the special case when the Borel sets $\mathscr{A}_1 = \mathscr{A}_2 = (1, \infty)$, where $\rho_{(1,\infty),(1,\infty)}(\boldsymbol{h}, l) = \chi(\boldsymbol{h}, l)$. The two cases $\chi(\boldsymbol{h}, l) = 0$ and $\chi(\boldsymbol{h}, l) = 1$ correspond to the boundary cases of asymptotic independence and complete dependence.

5.2.2 Scheme 2

We now generalize Scheme 1 in order to estimate temporal and spatial parameters simultaneously. Thus, we consider how the process X evolves in both space and time. In the classical geostatistics, for a stationary spatio-temporal process $\{X(\boldsymbol{s}, t) : (\boldsymbol{s}, t) \in \mathcal{S} \times \mathcal{T}\}$, the spatio-temporal empirical classical semivariogram is defined by

$$\widehat{\gamma}(\boldsymbol{h}, l) = \frac{1}{2|\mathcal{B}_{(\boldsymbol{h}, l)}|} \sum_{\mathcal{B}_{(\boldsymbol{h}, l)}} \left(X(\boldsymbol{s}_i, t_i,) - X(\boldsymbol{s}_j, t_j) \right)^2,$$

where $\mathcal{B}_{(\boldsymbol{h}, l)} = \left\{ (\boldsymbol{s}_i, t_i)(\boldsymbol{s}_j, t_j) : \boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h} \text{ and } t_i - t_j = l \right\}$, see, e.g., [78, 80]. By adapting this estimator to our framework, we consider the following estimation procedure:

(i) First, the spatio-temporal F-madogram is estimated nonparametrically by its empirical version. Assume the set $\mathcal{B}_{(h,l')}$ summarizes all pairs of $\mathcal{S}_n$ which give rise to the same spatial lag $h \in \mathcal{H} \subset [0, \infty)$ and the same temporal lag $l' \in \mathcal{K} \subset [0, \infty)$. In other words, combining the spatial and the temporal lags from Scheme 1, i.e.,

$$\mathcal{B}_{(h,l')} = \left\{ \left((\boldsymbol{s}_i, t_i), (\boldsymbol{s}_j, t_j) \right) : \|\boldsymbol{s}_i - \boldsymbol{s}_j\| = h, |t_i - t_j| = l' \right\}.$$

We estimate $\nu_F(\boldsymbol{h}, l')$ by

$$\widehat{\nu}_F(\boldsymbol{h}, l') = \frac{1}{2|\mathcal{B}_{(h,l')}|} \sum_{\mathcal{B}_{(h,l')}} |F\{X(\boldsymbol{s}_i, t_i) - F\{X(\boldsymbol{s}_j, t_j)\}|, \; (h, l') \in \mathcal{H} \times \mathcal{K}, \tag{5.13}$$

where $|.|$ denotes the cardinality of the set $\mathcal{B}_{(h,l')}$.

(ii) Then, we apply a nonlinear least squares fitting to obtain the estimates of the process parameters; $\boldsymbol{\psi}$, i.e.,

$$\widehat{\boldsymbol{\psi}} = \underset{\boldsymbol{\psi}\in\boldsymbol{\Psi}}{\arg\min} \sum_{l'\in\mathcal{K}} \sum_{\substack{h\in\mathcal{H} \\ \|\boldsymbol{h}\|=h}} \omega^{\boldsymbol{h},l'} \left(\widehat{\nu}_F(\boldsymbol{h},l') - \nu_F(\boldsymbol{h},l',\boldsymbol{\psi})\right)^2, \ (h,l') \in \mathcal{H}\times\mathcal{K}, \tag{5.14}$$

where $\omega^{\boldsymbol{h},l'} \geq 0$ denotes the spatio-temporal weights and $\nu_F(\boldsymbol{h},l',\boldsymbol{\psi})$ is the model-based spatio-temporal F-madogram counterpart.

The idea underlying the construction of Scheme 2 is that when modeling and predicting a given phenomenon, significant benefits may be obtained by considering how it evolves in both space and time rather than only considering its spatial distribution at a given time of reference (a merely spatial process), or its evolution over time at a given location (a merely temporal process), such as those described in Scheme 1.

Lastly, the establishment of the asymptotic properties of the resulting pairwise dependence estimates is deferred to future work. The derived asymptotic properties of the "unbinned" empirical λ-madogram in the spatial context, see [81] (Proposition 3 and 4) might provide a starting point. Nevertheless, this setting is more specialized. In the real data example of that study, a binned version of the empirical λ-madogram is adopted and deriving the convergence of this estimator as the cardinality of the distance class (i.e., $\mathcal{B}_h$) increases is still challenging. Therefore, for this moment, we will provide some numerical indications for the asymptotic properties of our pairwise dependence estimates.

Example 5.1. ***(Computing spatio-temporal distances and the number of spatio-temporal pairs of points for such distances, see also [80] for more details)*** *Assume that we observe a set of spatio-temporal data at three moments of time t_1, t_2 and t_3, on a regular 4×4, 1 km spaced grid, see Figure 5.1. Assuming an isotropic, stationary random field. The set of possible spatial lags between the pairs of the locations considered in the grid is $\left\{0, 1, \sqrt{2}, 2, \sqrt{5}, \sqrt{8}, 3, \sqrt{10}, \sqrt{13}, \sqrt{18}\right\}$ km, whereas the temporal lags are:* $0, 1, 2$. *Combining the spatial and temporal lags, we obtain* 10×3 *spatio-temporal distances. For instance, at the spatio-temporal distance* $(0,1)$, *there are* $16\times2 = 32$ *spatio-temporal pairs of points, that is,*

$$\{(\boldsymbol{s}_1,t_2),(\boldsymbol{s}_1,t_1)\},\{(\boldsymbol{s}_2,t_2),(\boldsymbol{s}_2,t_1)\},\ldots,\{(\boldsymbol{s}_{16},t_2),(\boldsymbol{s}_{16},t_1)\}$$

$$\{(\boldsymbol{s}_1,t_3),(\boldsymbol{s}_1,t_2)\},\{(\boldsymbol{s}_2,t_3),(\boldsymbol{s}_2,t_2)\},\ldots,\{(\boldsymbol{s}_{16},t_3),(\boldsymbol{s}_{16},t_2)\}.$$

Analogously, there are $16\times1 = 16$ *spatio-temporal pairs of points at distance* $(0,2)$. *It is also easy to check that there are* $24\times3 = 72$ *spatio-temporal pairs of points at distance* $(1,0)$, $16\times3 = 48$ *at distance* $(2,0)$, *and so forth. On the other hand, we need to be more careful when identifying the spatio-temporal pairs of points that are separated by a specific distance if both spatial and time lags are non zero. For example, we can count 16 pairs of points separated by a spatio-temporal distance* $(3,2)$, *i.e.,*

$$\{(\boldsymbol{s}_1,t_3),(\boldsymbol{s}_4,t_1)\},\{(\boldsymbol{s}_5,t_3),(\boldsymbol{s}_8,t_1)\},\{(\boldsymbol{s}_9,t_3),(\boldsymbol{s}_{12},t_1)\},\{(\boldsymbol{s}_{13},t_3),(\boldsymbol{s}_{16},t_1)\}$$

$$\{(\boldsymbol{s}_4,t_3),(\boldsymbol{s}_1,t_1)\},\{(\boldsymbol{s}_8,t_3),(\boldsymbol{s}_5,t_1)\},\{(\boldsymbol{s}_{12},t_3),(\boldsymbol{s}_9,t_1)\},\{(\boldsymbol{s}_{16},t_3),(\boldsymbol{s}_{13},t_1)\}$$
$$\{(\boldsymbol{s}_1,t_3),(\boldsymbol{s}_{13},t_1)\},\{(\boldsymbol{s}_2,t_3),(\boldsymbol{s}_{14},t_1)\},\{(\boldsymbol{s}_3,t_3),(\boldsymbol{s}_{15},t_1)\},\{(\boldsymbol{s}_4,t_3),(\boldsymbol{s}_{16},t_1)\}$$
$$\{(\boldsymbol{s}_{13},t_3),(\boldsymbol{s}_1,t_1)\},\{(\boldsymbol{s}_{14},t_3),(\boldsymbol{s}_2,t_1)\},\{(\boldsymbol{s}_{15},t_3),(\boldsymbol{s}_3,t_1)\},\{(\boldsymbol{s}_{16},t_3),(\boldsymbol{s}_4,t_1)\}.$$

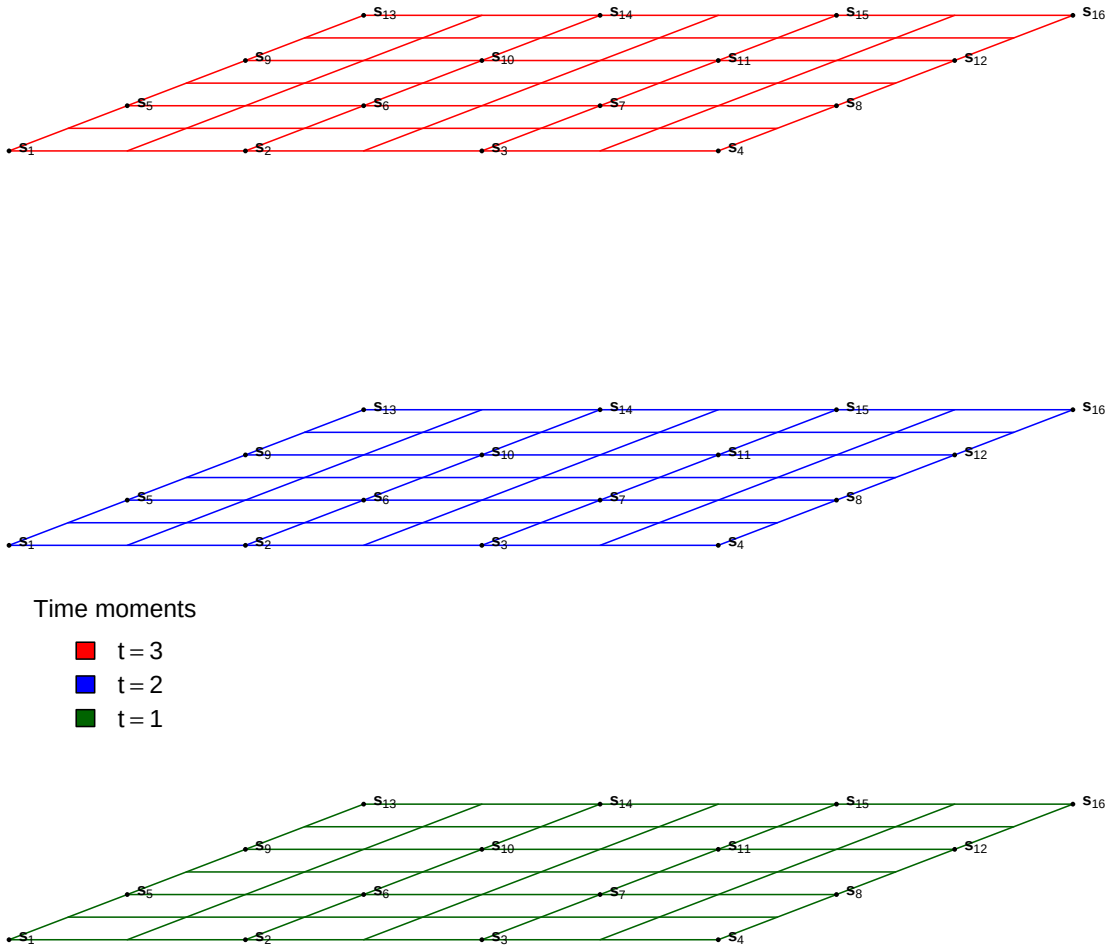


Figure 5.1: Regular 4×4, 1 km spaced grid.

5.2.3 Illustration examples

In order to illustrate how the proposed estimation schemes perform, we consider the following two examples, which we will revisit in Section 5.3.

Example 5.2. ***(Estimation of isotropic space-time max-stable BR)*** *Let us consider the space-time max-stable BR process in (2.49) with bivariate c.d.f. (2.50), where the dependence structure*

is given by the following stationary isotropic FBM spatio-temporal semivariogram (recall Section 2.1.2)

$$\gamma(\boldsymbol{h}, l) := \gamma(h, l') = 2\phi_s h^{\kappa_s} + 2\phi_t l'^{\kappa_t}, \tag{5.15}$$

where the scalar distance $h = \|\boldsymbol{h}\| = \|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$, $l' = |l| = |t_1 - t_2|$, $\phi_s, \phi_t > 0$ determine spatial and temporal scale parameters and $\kappa_s, \kappa_t \in (0, 2]$ relate to the smoothness of the underlying Gaussian process in space and time. The associated spatio-temporal F-madogram with this process is

$$\nu_F(h, l') = \frac{1}{2} - \frac{1}{2\Phi\left(\sqrt{\phi_s h^{\kappa_s} + \phi_t l'^{\kappa_t}}\right) + 1}, \tag{5.16}$$

where $\theta(h, l') = 2\Phi\left(\sqrt{\phi_s h^{\kappa_s} + \phi_t l'^{\kappa_t}}\right)$ is the associated spatio-temporal extremal dependence function. Figure 5.2 visualizes a 3D representation of the spatio-temporal FBM semivariogram in (5.15) and the associated dependence summary measures: the spatio-temporal extremal dependence function $\theta : \mathbb{R}^2 \times \mathbb{R}^+ \mapsto [1, 2]$ and the spatio-temporal F-madogram $\nu_F : \mathbb{R}^2 \times \mathbb{R}^+ \mapsto [0, 1/6]$. Complete dependence (respectively, complete independence) is achieved at lower boundaries (respectively, upper boundaries). Moreover, Figure 5.3 displays the theoretical behaviors of the purely spatial FBM semivariogram $\gamma^{(s)}(h, \kappa_s)$ and the related purely spatial F-madogram $\nu_F^{(s)}(h, \kappa_s)$. Obviously, depending on the value of the smoothness parameter κ_s, these measures exhibit a large variety of dependence behaviors.

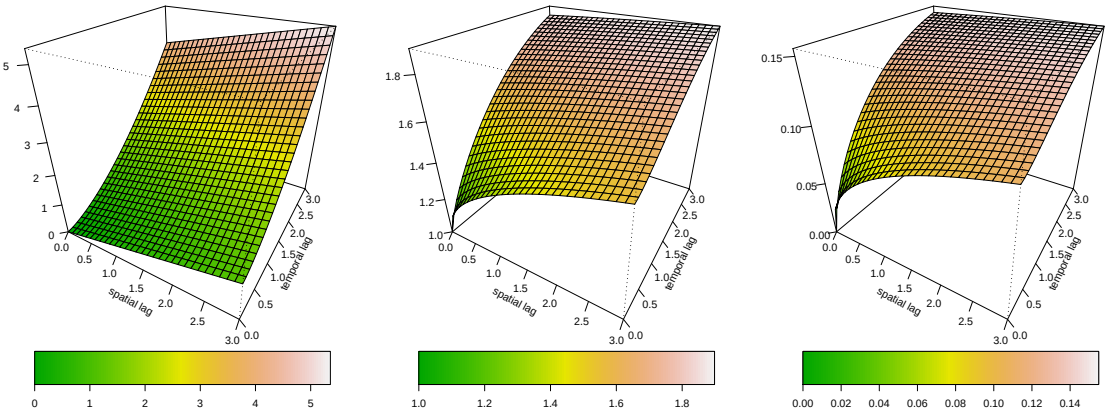


Figure 5.2: Spatio-temporal FBM semivariogram $\gamma(h, l') = 0.8h^{1.5} + 0.4l'$ (left panel). The associated spatio-temporal extremal dependence function (middle panel). The associated spatio-temporal F-madogram (right panel).

With this construction, based on Scheme 1, the nonlinear least squares optimization problems in (5.10) and (5.11) can be expressed as

$$\begin{pmatrix} \widehat{\kappa}_s \\ \widehat{\phi}_s \end{pmatrix} = \underset{\substack{\phi_s > 0 \\ \kappa_s \in (0,2]}}{arg\,min} \sum_{h \in \mathcal{H}} \omega^h \left(\widehat{\nu}_F(h) - \left\{ \frac{1}{2} - \frac{1}{2\Phi\left(\sqrt{\phi_s h^{\kappa_s}}\right) + 1} \right\} \right)^2, \; h \in \mathcal{H}, \tag{5.17}$$

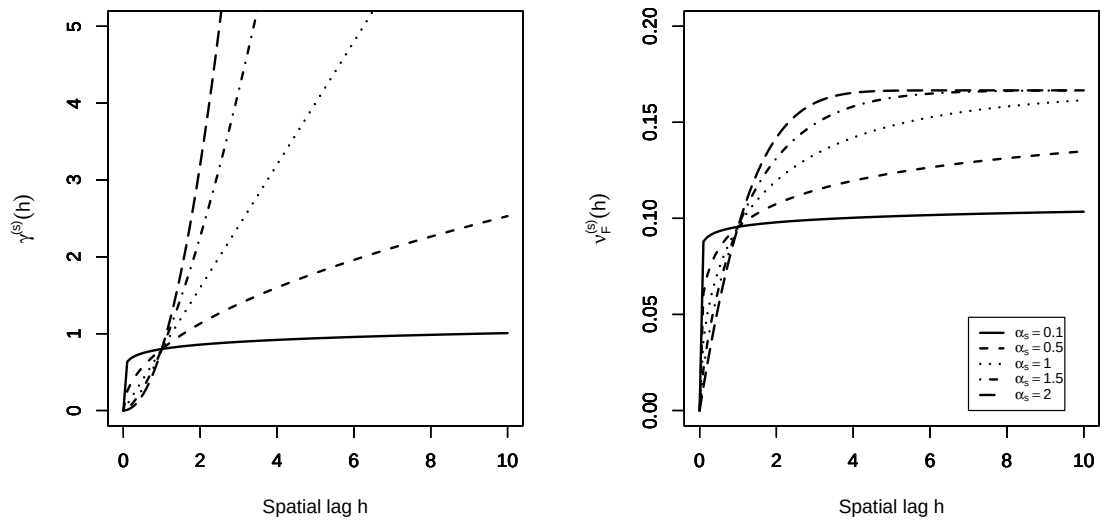


Figure 5.3: The FBM semivariogram $\gamma^{(s)}(h,\kappa_s) = 0.8h^{\kappa_s}$ (left panel) and the related spatial F-madogram $\nu_F^{(s)}(h,\kappa_s) = 0.5 - \left\{2\Phi\left(\sqrt{0.4h^{\kappa_s}}\right)+1\right\}^{-1}$ (right panel) plotted as functions of space lag h, with different smoothness parameter $\kappa_s \in \{0.1, 0.5, 1, 1.5, 2\}$.

$$\begin{pmatrix}\widehat{\kappa}_t\\ \widehat{\phi}_t\end{pmatrix} = \underset{\substack{\phi_t>0\\ \kappa_t\in(0,2]}}{arg\ min} \sum_{l'\in\mathcal{K}} \omega^{l'}\left(\widehat{\nu}_F(l') - \left\{\frac{1}{2} - \frac{1}{2\Phi\left(\sqrt{\phi_t l'^{\kappa_t}}\right)+1}\right\}\right)^2,\ l'\in\mathcal{K}. \tag{5.18}$$

Lastly, with $(h,l') \in \mathcal{H}\times\mathcal{K}$ and on the basis of Scheme 2, the nonlinear least squares estimation problem in (5.14) has the form

$$\begin{pmatrix}\widehat{\kappa}_s\\ \widehat{\phi}_s\\ \widehat{\kappa}_t\\ \widehat{\phi}_t\end{pmatrix} = \underset{\substack{\phi_s,\phi_t>0\\ \kappa_s,\kappa_t\in(0,2]}}{arg\ min} \sum_{l'\in\mathcal{K}}\sum_{h\in\mathcal{H}} \omega^{h,l'}\left(\widehat{\nu}_F(h,l') - \left\{\frac{1}{2} - \frac{1}{2\Phi\left(\sqrt{\phi_s h^{\kappa_s} + \phi_t l'^{\kappa_t}}\right)+1}\right\}\right)^2. \tag{5.19}$$

Example 5.3. ***(Estimation of spectrally separable space-time max-stable Smith process)*** *We now describe the way to fit the spectrally separable space-time max-stable Smith process. Indeed, the estimation procedure can be simplified since the purely spatial parameters can be estimated independently of the purely temporal parameters. Formally, we consider the process in (5.4), where the innovation process H is derived from independent replications of a spatial Smith process with covariance matrix*

$$\boldsymbol{\Sigma} = \begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{pmatrix}. \tag{5.20}$$

We donte by $\boldsymbol{\psi}$ the vector gathering the parameters to be estimated, i.e., $\boldsymbol{\psi} = (\sigma_{11},\sigma_{12},\sigma_{22},\boldsymbol{\tau}^t,\delta)^t$. It is possible to separate the estimation. Firstly, the estimation of the spatial parameters $\boldsymbol{\psi}^{(s)} = (\sigma_{11},\sigma_{12},\sigma_{22})^t$ is carried out. Secondly, once $\boldsymbol{\psi}^{(s)}$ is known, it is held fixed and we estimate the temporal parameters $\boldsymbol{\psi}^{(t)} = (\boldsymbol{\tau}^t,\delta)^t = (\tau_1,\tau_2,\delta)^t$. Subsequently, under Scheme 1,

the nonlinear least squares optimization problems in (5.10) and (5.11) can be expressed as

$$\begin{pmatrix}\widehat{\sigma}_{11}\\ \widehat{\sigma}_{12}\\ \widehat{\sigma}_{22}\end{pmatrix} = \underset{\substack{\sigma_{11},\sigma_{22}>0\\ \sigma_{12}\in\mathbb{R}}}{arg\,min} \sum_{\substack{h\in\mathcal{H}\\ \|\boldsymbol{h}\|=h}} \omega^{\boldsymbol{h}} \left(\widehat{\nu}_F(\boldsymbol{h}) - \left\{ \frac{1}{2} - \frac{1}{2\Phi\left(\sqrt{\boldsymbol{h}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{h}}/2\right)+1} \right\} \right)^2, \; h\in\mathcal{H}, \tag{5.21}$$

$$\begin{pmatrix}\widehat{\delta}\\ \widehat{\tau}_1\\ \widehat{\tau}_2\end{pmatrix} = \underset{\substack{a\in(0,1)\\ \tau_1,\tau_2\in\mathbb{R}}}{arg\,min} \sum_{l'\in\mathcal{K}} \omega^{l'} \left(\widehat{\nu}_F(l') - \left\{ \frac{1}{2} - \frac{1}{\theta(l')+1} \right\} \right)^2, \; l'\in\mathcal{K}, \tag{5.22}$$

where $\theta(l') := \theta(\mathbf{0},l') = \Phi\left(\dfrac{b^*(l')}{2} + \dfrac{1}{b^*(l')}\log\left(\delta^{-l'}\right)\right) + \delta^l\Phi\left(\dfrac{b^*(l')}{2} + \dfrac{1}{b^*(l')}\log\left(\delta^{l'}\right)\right) + 1 - \delta^{l'}$

with $b^*(l') = \sqrt{(\mathbf{0}-l'\boldsymbol{\tau})^t\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{0}-l'\boldsymbol{\tau})}$.

In order to figure out the role of the temporal parameter δ *for this process. For a fixed site* $\boldsymbol{s}\in\mathcal{S}$*, Figure 5.4 displays the temporal extremal function* $\theta(l')$ *and the associated temporal F-madogram* $\nu_F^{(t)}$ *for* $\delta\in\{0.1,0.3,0.5,0.7,0.9\}$*. We set* $\boldsymbol{\Sigma} = 10\,Id_2$ *and* $\boldsymbol{\tau} = (1,1)^t$ *(translation to the top right). Clearly, as the value of* δ *increases, the independece (i.e.,* $\theta(l')\to 2$*) occurs at larger time lags* l'*.*

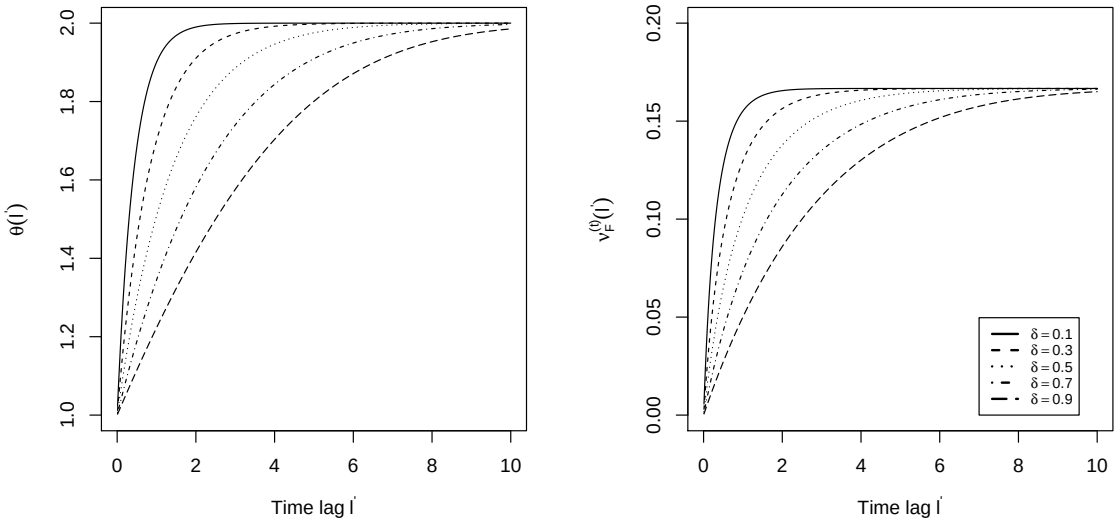


Figure 5.4: $\theta(l')$ and the associated $\nu_F^{(t)}(l')$ plotted as functions of time lag l' for $\delta\in\{0.1,0.3,0.5,0.7,0.9\}$ based on the process (5.4), where H is a sequence of i.i.d. spatial Smith processes with covariance matrix $\boldsymbol{\Sigma} = 10\,\mathrm{Id}_2$.

Lastly, based on Scheme 2, the nonlinear least squares estimator $\widehat{\boldsymbol{\psi}} = \left(\widehat{\sigma}_{11},\widehat{\sigma}_{12},\widehat{\sigma}_{22},\widehat{\tau}_1,\widehat{\tau}_2,\widehat{\delta}\right)^t$ *is given by*

$$\widehat{\boldsymbol{\psi}} = \underset{\boldsymbol{\psi}\in\boldsymbol{\Psi}}{arg\,min} \sum_{l'\in\mathcal{K}} \sum_{\substack{h\in\mathcal{H}\\ \|\boldsymbol{h}\|=h}} \omega^{\boldsymbol{h},l'} \left(\widehat{\nu}_F(\boldsymbol{h},l') - \left\{ \frac{1}{2} - \frac{1}{\theta(\boldsymbol{h},l')+1} \right\} \right)^2, \; (h,l')\in\mathcal{H}\times\mathcal{K}, \tag{5.23}$$

where $\theta(\boldsymbol{h},l') = \Phi\left(\frac{b(\boldsymbol{h},l')}{2} + \frac{1}{b(\boldsymbol{h},l')}\log\left(\delta^{-l'}\right)\right) + \delta^{l'}\Phi\left(\frac{b(\boldsymbol{h},l')}{2} + \frac{1}{b(\boldsymbol{h},l')}\log\left(\delta^{l'}\right)\right) + 1 - \delta^{l'}$
with $b(\boldsymbol{h},l') = \sqrt{(\boldsymbol{h}-l'\boldsymbol{\tau})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{h}-l'\boldsymbol{\tau})}$.

5.3 Simulation study

Throughout this section, we investigate the performance of the semi-parametric estimation procedures introduced in Section 5.2 with three simulation studies.

5.3.1 Simulation study 1: Fitting space-time max-stable BR process

In this study, we adopt the same experiment plan that has been proposed in [23] (Section 5), in order to make the results obtained there comparable with the results here.

Setup for a simulation study

We simualte the space-time BR process with spectral representation (2.49) and dependence function γ modeled as in (5.15). Namely,

$$\gamma(\boldsymbol{h},l) = 0.8h^{3/2} + 0.4l'. \tag{5.24}$$

The simulations have been carried out using the function *RFsimulate* of the R package RandomFields [91] and based on the exact method proposed by [52]. The space-time observation area is assumed to be on a $n \times n$ spatial grid and the time moments are equidistantly, i.e.,

$$\mathscr{A} = \{(x,y) : x,y \in \{1,\ldots,n\}\} \times \{1,\ldots,T\}.$$

Figure 5.5 visualizes a realization simulated from space-time BR process with a spatio-temporal FBM semivariogram model (5.24) at six consecutive time points. As in [23], we choose the sets $\mathcal{H} = \{1, \sqrt{2}, 2, \sqrt{5}, \sqrt{8}, 3, \sqrt{10}, \sqrt{13}, 4, \sqrt{17}\}$ and $\mathcal{K} = \{1,\ldots,10\}$, where permutation tests (which we describe at the end of Section 5.4) show that these lags are enough to capture the relevant extremal dependence structure. The adoption of larger lags may introduce a bias in the estimation, see [23, 41]. Figure 5.6 depicts the locations with influence on the estimation. Equal weights are assumed. We repeat this experiment 100 times to obtain summary plots of the resulting estimates and to compute performance metrics: the mean estimate, RMSE, and MAE.

Estimation using Scheme 1

Simulation of space-time max-stable BR processes based on the exact method proposed in [52] can be time-consuming. Hence, for the sake of time-saving and due to the fact that the estimation of the purely spatial (respectively, purely temporal) parameters depends on a large number of spatial observations (respectively, a large number of observed time instants), we examine the performance of the purely spatial (respectively, purely temporal) estimates using two different space-time observation areas, i.e.,

- $\mathscr{A}_1 = \{(x,y) : x,y \in \{1,\ldots,50\}\} \times \{1,\ldots,10\}$.

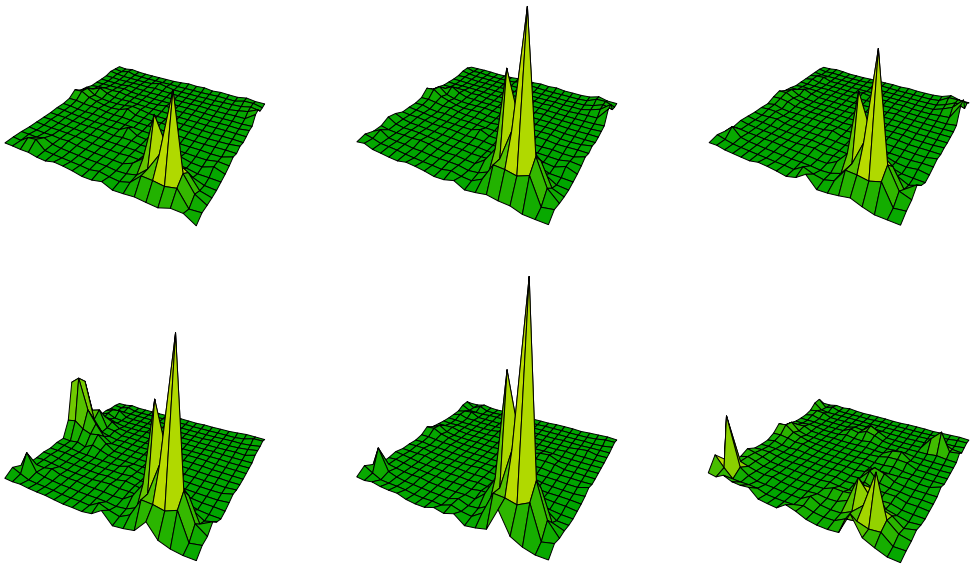


Figure 5.5: Simulation from a space-time max-stable BR process with spatio-temporal FBM semivariogram $\gamma(\boldsymbol{h}, l) = 0.8h^{1.5} + 0.4l'$ at six consecutive time points (from left to right and top to bottom).

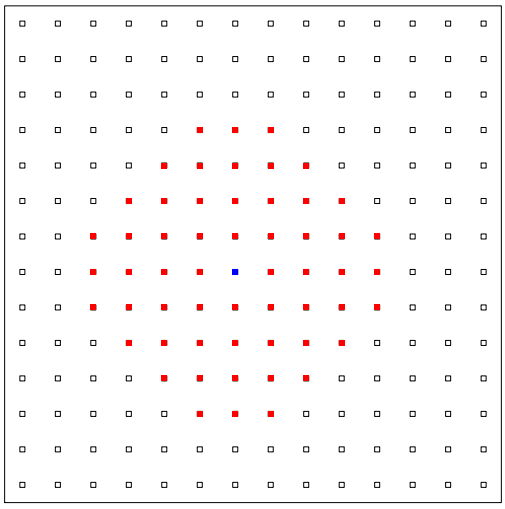


Figure 5.6: A regular 14×14 spatial grid. The distances between the peripheral locations (shown by red square-symbols) and the central one (shown by blue square symbol) belong to the set $\mathcal{H}$.

- $\mathscr{A}_2 = \{(x, y) : x, y \in \{1, \ldots, 5\}\} \times \{1, \ldots, 300\}$.

We assess the quality of the fit between the theoretical values of spatial/temporal F-madograms and their estimates. Figure 5.7 compares empirical estimates of purely spatial/temporal F-madograms with their asymptotic counterparts. Overall, both the purely spatial/temporal empirical versions are consistent with their asymptotic counterparts, with a relatively higher variability for the temporal estimates. This is probably due to the fairly low number of time instants (300) used for the estimation of the purely temporal parameters compared to the number of spatial locations (2500) used for the estimation of the purely spatial parameters.

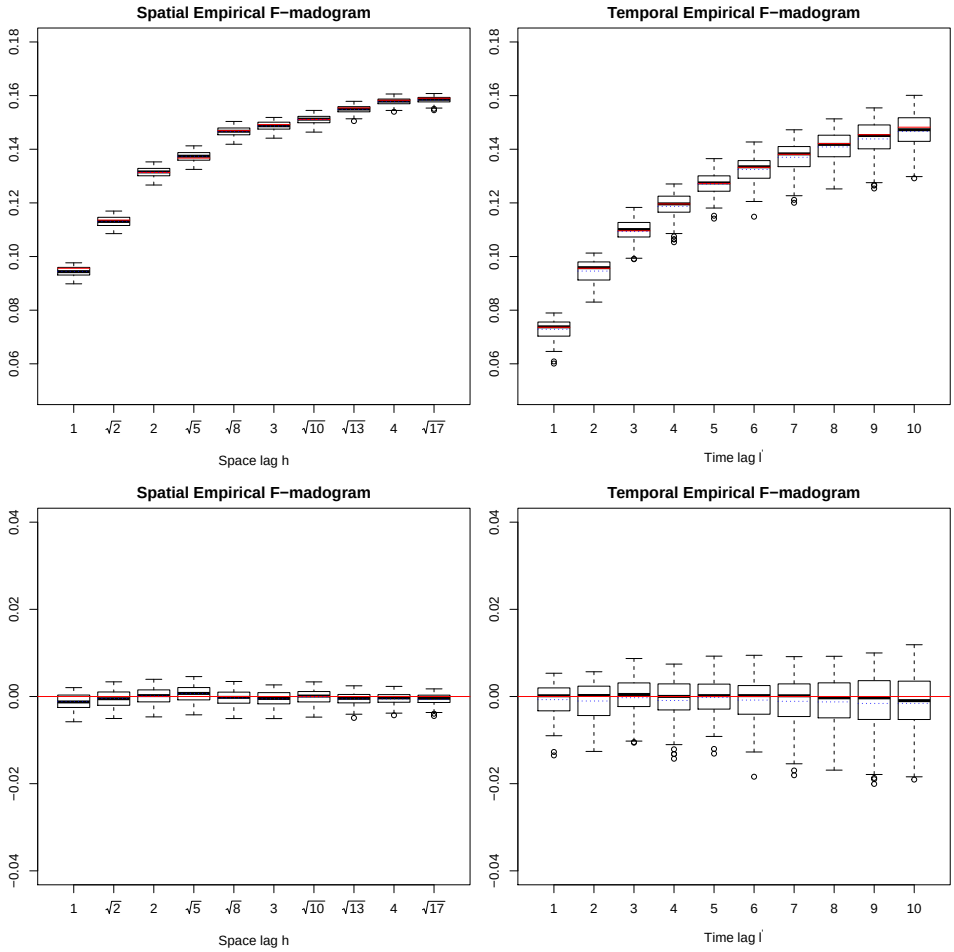


Figure 5.7: Scheme 1—boxplots of purely spatial/temporal empirical F-madograms estimates at lags $(h, l') \in \mathcal{H} \times \mathcal{K}$ for 100 simulated BR processes (2.49) with FBM spatio-temporal semivariogram (5.24) (top row). The middle blue dotted/red solid lines show the overall mean of the estimates/true values. Boxplots of the corresponding estimation errors (bottom row).

Next, we present results for the semi-parametric estimation with Scheme1. Figure 5.8 displays the resulting estimates of the purely spatial parameters (ϕ_s, α_s) and the purely temporal parameters (ϕ_t, α_t). Generally, the estimation procedure appears to work well. Moreover, we observe that the estimation of the purely spatial parameters is more accurate (the RMSE and

MAE are lower), see Table 5.1. Again this probably stems from the large number of spatial locations used in the estimation which is (≈ 8.3) times higher than the time points.

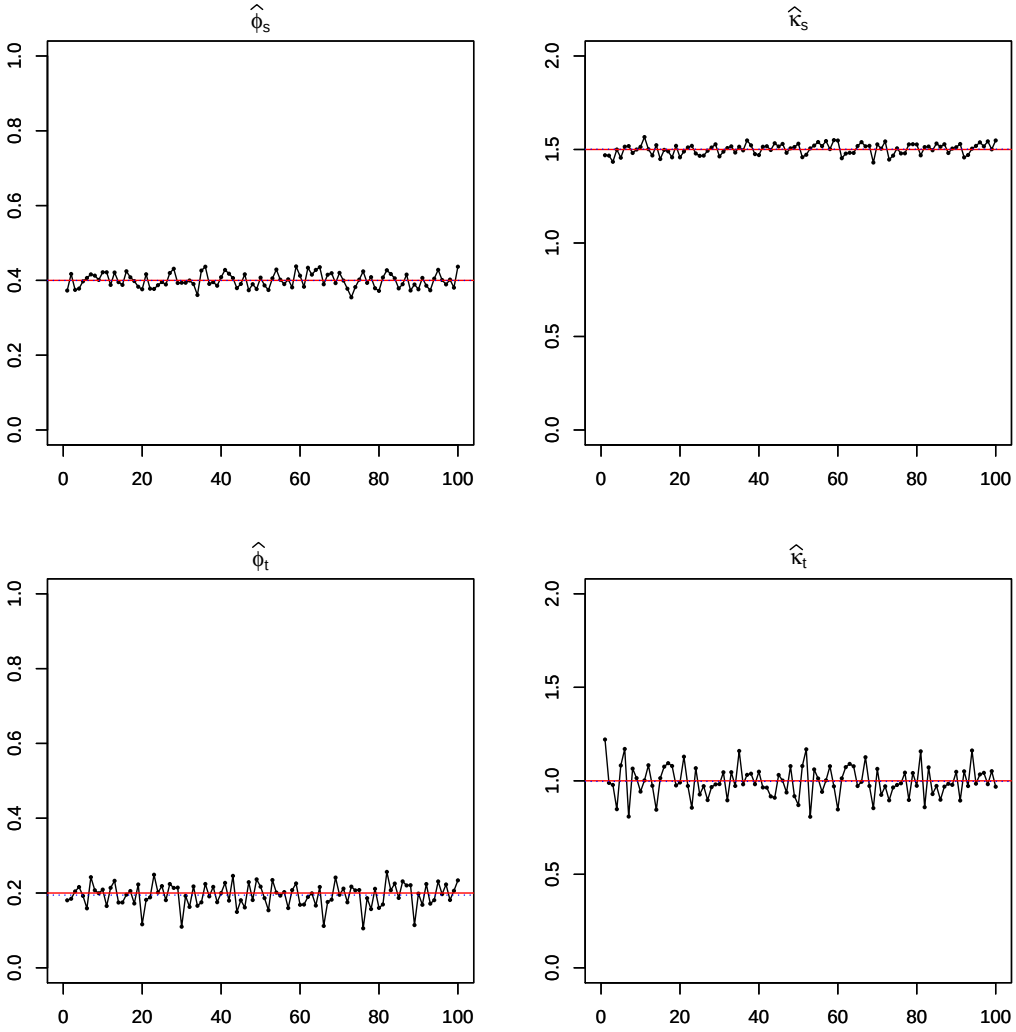


Figure 5.8: Scheme 1—semi-parametric estimates of $\widehat{\boldsymbol{\psi}} = \{\widehat{\phi}_s, \widehat{\kappa}_s, \widehat{\phi}_t, \widehat{\kappa}_t\}$ for 100 simulated BR processes defined by (2.49) with FBM spatio-temporal semivariogram (5.24). The middle blue dotted/red solid lines show the overall mean of the estimates/true values.

As the last step in this study, we compare the statistical efficiency of our method and the one proposed in [23]. Table 5.1 reports the performance metrics for both methods. Although in that study, the authors used a larger grid size ($n = 70$) to estimate the purely spatial parameters, clearly, the F-madogram semi-parametric estimation outperforms their approach which based on the extremogram as an inferential tool (their semi-parametric estimates show a larger bias than ours; the RMSE and MAE are higher). This is probably due to the fact that the estimates obtained in [23] are sensitive to the choice of the threshold used for computing (possibly bias corrected) empirical estimates of the extremogram.

Estimation using Scheme 2

Based on Scheme 2, we estimate the parameters of the space-time max-stable BR process with a similar simulation setting which is previously described in Section 5.3.1. We consider the space-time observation area where the spatial locations consisted of a 20×20 grid and equidistantly time points, $\{1, \ldots, 200\}$. Figure 5.9 displays the boxplots of the empirical spatio-temporal F-madogram estimates at a subset of spatio-temporal lags; $\{(1, l'), l' \in \mathcal{K}\}$, and the corresponding errors boxplots. Whereas, Figure 5.10 compares the empirical spatio-temporal F-madogram estimates $\widehat{\nu}_F(h, l')$ with their model-based counterparts $\nu_F(h, l')$ over the spatio-temporal lags $(h, l') \in \mathcal{H} \times \mathcal{K}$. There is a good agreement overall. These diagnostic plots provide a satisfactory representation of the empirical spatio-temporal F-madogram estimates. Generally, the results lend support to the agreement between the empirical spatio-temporal F-madogram estimates and model-based counterparts, especially once sampling variability is taken into account.

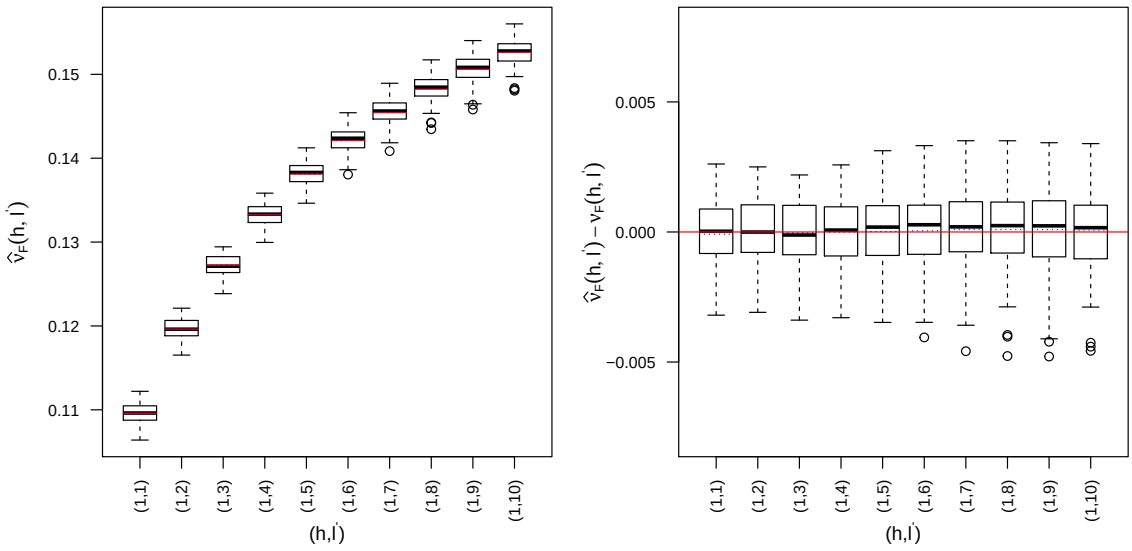


Figure 5.9: Scheme 2—boxplots of empirical spatio-temporal F-madogram estimates at spatio-temporal lags $\{(1, l'), l' \in \mathcal{K}\}$ for 100 simulated BR processes defined by (2.49) with FBM spatio-temporal semivariogram (5.24) (left panel). The middle blue dotted/red solid lines show the overall mean of the estimates/model-based counterparts. The associated boxplots of the errors $\widehat{\nu}_F(1, l') - \nu_F(1, l')$, $l' \in \mathcal{K}$ (right panel).

Figure 5.11 shows the estimation performance of the estimated parameters. Overall, the parameters are well estimated. Moreover, we observe that the estimation of the scale parameters $\{\phi_s, \phi_t\}$ is more accurate than the smoothness parameters $\{\kappa_s, \kappa_t\}$ (the RMSE and MAE are lower), see Table 5.1.

To sum up, for both schemes, Table 5.1 reports the mean estimate, RMSE, and MAE of the estimated parameters $\widehat{\boldsymbol{\psi}} = \{\widehat{\phi}_s, \widehat{\kappa}_s, \widehat{\phi}_t, \widehat{\kappa}_t\}$. Let us remark that the comparison between the resulting parameter estimates from the two estimation schemes is not completely straightforward because we consider non-unified space-time observation areas due to the above-mentioned computational

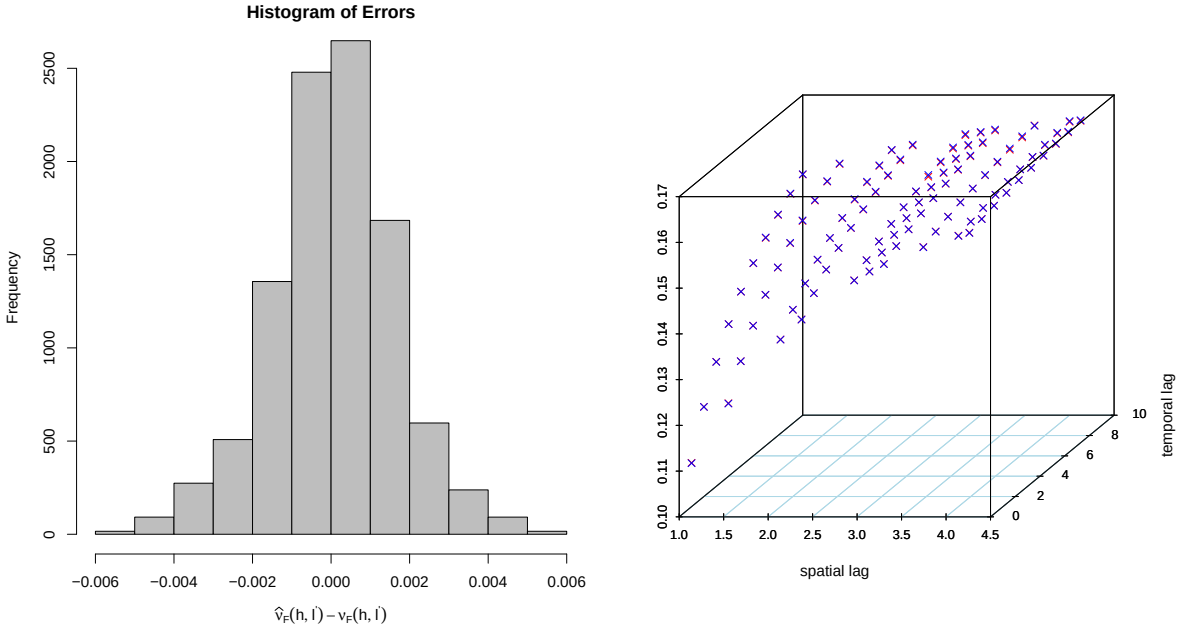


Figure 5.10: Scheme 2—diagnostic plots of the empirical spatio-temporal F-madogram estimates for 100 simulated BR processes defined by (2.49) with FBM spatio-temporal semivariogram (5.24). Histogram of the errors, $\widehat{\nu}_F(h,l') - \nu_F(h,l')$, $(h,l') \in \mathcal{H} \times \mathcal{K}$ (left panel). Blue/red cross-symbols show the overall mean of the empirical spatio-temporal F-madogram estimates/model-based counterparts (right panel).

reasons. However, with the above sampling schemes, we observe that the estimation of the purely spatial parameters is more accurate when using Scheme 1 (the RMSE and MAE are lower). On the other hand, we notice a slight outperformance for Scheme 2 in estimating purely temporal parameters. Finally, the QQ-plots against a normal distribution in Figure 5.12 provide an indication for asymptotic normality of the resulting estimates.

Table 5.1: Performance of the estimation for 100 simulated BR processes defined by (2.49) with FBM spatio-temporal semivariogram (5.24). The mean estimate, RMSE, and MAE of the estimated parameters.

	Scheme 1			Scheme 1, [23]			Scheme 2		
True	Mean estimate	RMSE	MAE	Mean estimate	RMSE	MAE	Mean estimate	RMSE	MAE
Purely Spatial									
$\phi_s = 0.4$	0.3998	0.0191	0.0162	0.4033	0.0678	0.0559	0.4093	0.0389	0.0307
$\kappa_s = 1.5$	1.5019	0.0289	0.0243	1.4984	0.0521	0.0400	1.4921	0.1399	0.1083
Purely temporal									
$\phi_t = 0.2$	0.1944	0.0314	0.0246	0.2249	0.0649	0.0526	0.1909	0.0251	0.0201
$\kappa_t = 1$	0.9969	0.0831	0.0657	0.9563	0.0939	0.0767	1.0278	0.0785	0.0619

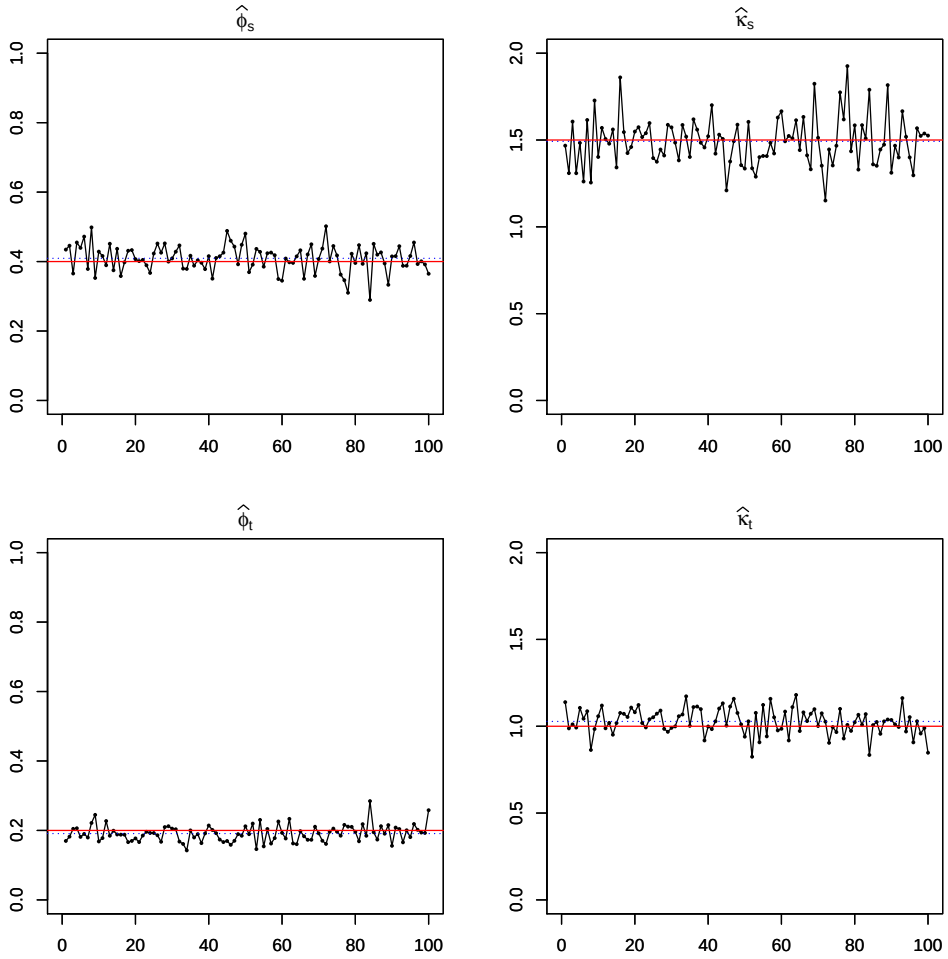


Figure 5.11: Scheme 2—semi-parametric estimates of $\widehat{\boldsymbol{\psi}} = \{\widehat{\phi}_s, \widehat{\kappa}_s, \widehat{\phi}_t, \widehat{\kappa}_t\}$ for 100 simulated BR processes defined by (2.49) with FBM spatio-temporal semivariogram (5.24). The middle blue dotted/red solid lines show overall mean of the estimates/true values.

5.3.2 Simulation study 2: Fitting spectrally separable spatio-temporal Smith process

Setup for a simulation study

We simulate data from the spatio-temporal Smith process considered in Example 5.3, with parameter vector $\boldsymbol{\psi} = (1, 0, 1, 1, 1, 0.7)^t$. As a reasonable compromise between accuracy and computation time, the locations are assumed to lie on a regular 2D grid of size $n = 20$. The time points are equidistant, given by the set $\{1, \ldots, 200\}$. The simulations have been carried out using R SpatialExtremes package with *rmaxstab* function, see [86]. The spatial lags set $\mathcal{H}$ and temporal lags set $\mathcal{K}$ are fixed as before, recall Section 5.3.1. Equal weights are assumed. We repeat this experiment 100 times.

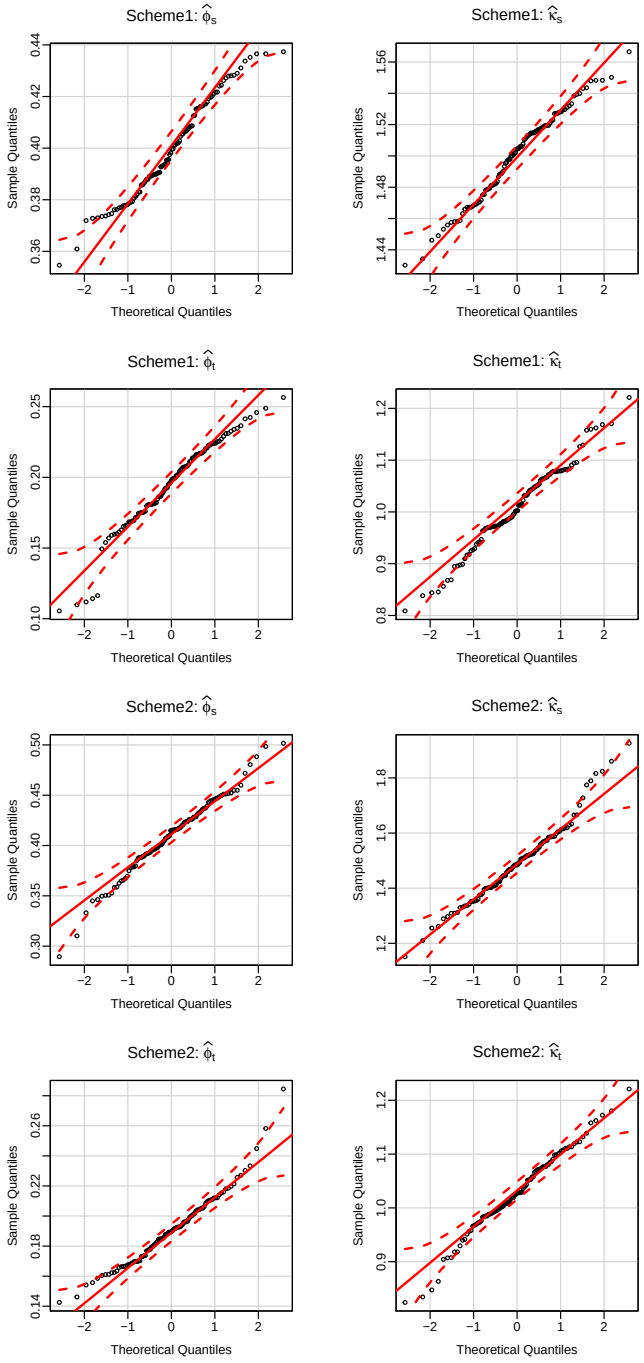


Figure 5.12: QQ-plots of the estimates resulting from both estimation schemes for 100 simulated BR processes defined by (2.49) with the FBM spatio-temporal semivariogram (5.24) against the normal distribution. Scheme 1—purely spatial parameters (top row) and purely temporal parameters (second row). Scheme 2—purely spatial parameters (third row) and purely temporal parameters (bottom row). Dashed red lines correspond to 95% confidence intervals.

Results for the two estimation schemes

The top row of Figure 5.13 displays the density of the errors between the empirical estimates of the purely spatial/temporal F-madograms and their model-based counterparts, whereas the bottom row displays the density of the errors between empirical spatio-temporal F-madogram estimates and model-based counterparts. Generally, all of the empirical versions are congruous with their asymptotic counterparts. Clearly, the density of the errors is close to a centered Gaussian distribution.

Figure 5.14 displays boxplots the errors of the resulting estimates from both schemes: $(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi})$. The top row displays the estimation errors of purely spatial parameters $(\sigma_{11}, \sigma_{12}, \sigma_{22})$ and purely temporal parameters (τ_1, τ_2, δ) resulting from Scheme 1, whereas the bottom row displays the estimation errors resulting form Scheme 2. Overall, the inference procedures perform well. Altogether, we observe that the estimates are close to the true values.

To sum up, for both schemes, Table 5.2 reports the mean estimate, RMSE, and MAE of the estimated parameters $\widehat{\boldsymbol{\psi}} = \{\widehat{\sigma}_{11}, \widehat{\sigma}_{12}, \widehat{\sigma}_{22}, \widehat{\tau}_1, \widehat{\tau}_2, \widehat{\delta}\}$. Contrary to Scheme 2, we observe that the estimation of purely spatial parameters $\boldsymbol{\Sigma}$ is more precise than the estimation of purely temporal parameters ($\boldsymbol{\tau}$ and δ) when using Scheme 1 (RMSE and MAE are lower). This probably can be justified by the fact that in Scheme 1 the number of spatial locations used is higher than time moments. Additionally, there is probably an impact of the estimated covariance matrix $\widehat{\boldsymbol{\Sigma}}$ on the estimation efficiency of the purely temporal parameters, whereas, the purely temporal parameters are estimated independently of purely spatial parameters when using Scheme 2. Moreover, we notice that the estimation of purely spatial parameters is less precise when using Scheme 2 (RMSE and MAE are higher). This is probably owing to the fact that in Scheme 2 the number of pairs used is higher than in Scheme 1, leading more variability. Whereas, both schemes seem to have the same performance order in estimating purely temporal parameters.

We also show QQ-plots against a normal distribution for all parameters in Figure 5.15. For both schemes, it seems that the semi-parametric estimates are approximately normally distributed.

Table 5.2: Performance of the estimation for 100 simulated spectrally separable space-time max-stable Smith processes considered in Example 5.3, with parameter $\boldsymbol{\psi} = (1, 0, 1, 1, 1, 0.7)^t$. The mean estimate, RMSE, and MAE of the estimated parameters.

	Scheme 1			Scheme 2		
True	Mean estimate	RMSE	MAE	Mean estimate	RMSE	MAE
Purely Spatial						
$\sigma_{11} = 1$	0.9973	0.0331	0.0259	0.9929	0.0888	0.0727
$\sigma_{12} = 0$	0.0081	0.0470	0.0369	−0.0357	0.0770	0.0609
$\sigma_{22} = 1$	0.9848	0.0440	0.0346	1.0295	0.0805	0.0647
Purely temporal						
$\tau_1 = 1$	1.0021	0.0549	0.0426	1.0261	0.0747	0.0591
$\tau_2 = 1$	1.0107	0.0646	0.0505	0.9962	0.0620	0.0516
$\delta = 0.7$	0.7012	0.0595	0.0482	0.6939	0.0510	0.0400

Finally, let us remark that a simulation study has been carried out in [54], where only

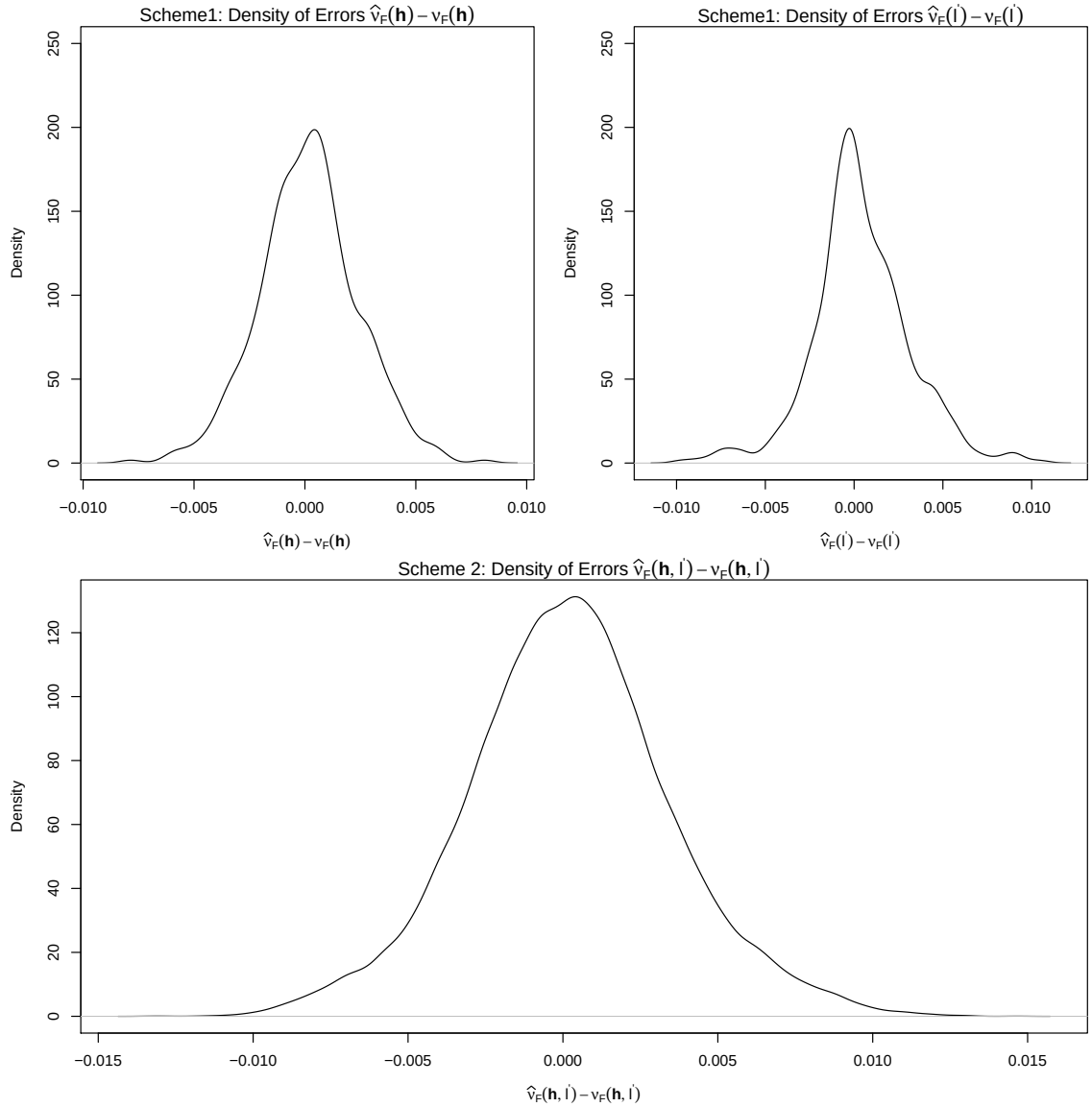


Figure 5.13: Density of the errors between the empirical versions of the F-madogram estimates and their model-based counterparts for 100 simulated spectrally separable space-time max-stable Smith processes with parameter $\boldsymbol{\psi} = (1, 0, 1, 1, 1, 0.7)^t$. Scheme 1 (Top row): $\widehat{\nu}_F(\boldsymbol{h}) - \nu_F(\boldsymbol{h})$, $\|\boldsymbol{h}\| \in \mathcal{H}$ (left panel) $\widehat{\nu}_F(l') - \nu_F(l')$, $l' \in \mathcal{K}$ (right panel). Scheme 2 (Bottom row): $\widehat{\nu}_F(\boldsymbol{h}, l') - \nu_F(\boldsymbol{h}, l')$, at spatio-temporal lags $(\|\boldsymbol{h}\|, l') \in \mathcal{H} \times \mathcal{K}$.

the spectrally separable spatio-temporal Smith process has been fitted with parameters $\boldsymbol{\psi} = (1, 0, 1, -1, -1, 0.7)^t$. Irregularly spaced locations have been considered. Two estimation approaches based on pairwise likelihood have been adopted: (i) a two-step approach, first estimating the purely spatial parameters. Once these parameters are known, they are held fixed and the purely temporal parameters are estimated. (ii) a one-step approach, all parameters are estimated simultaneously. More precisely, assume that we observe the space-time max-stable process X at D sites $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_D$ and T dates $t_1, \ldots, t_T$. Let f denotes the p.d.f. of the process X, then the

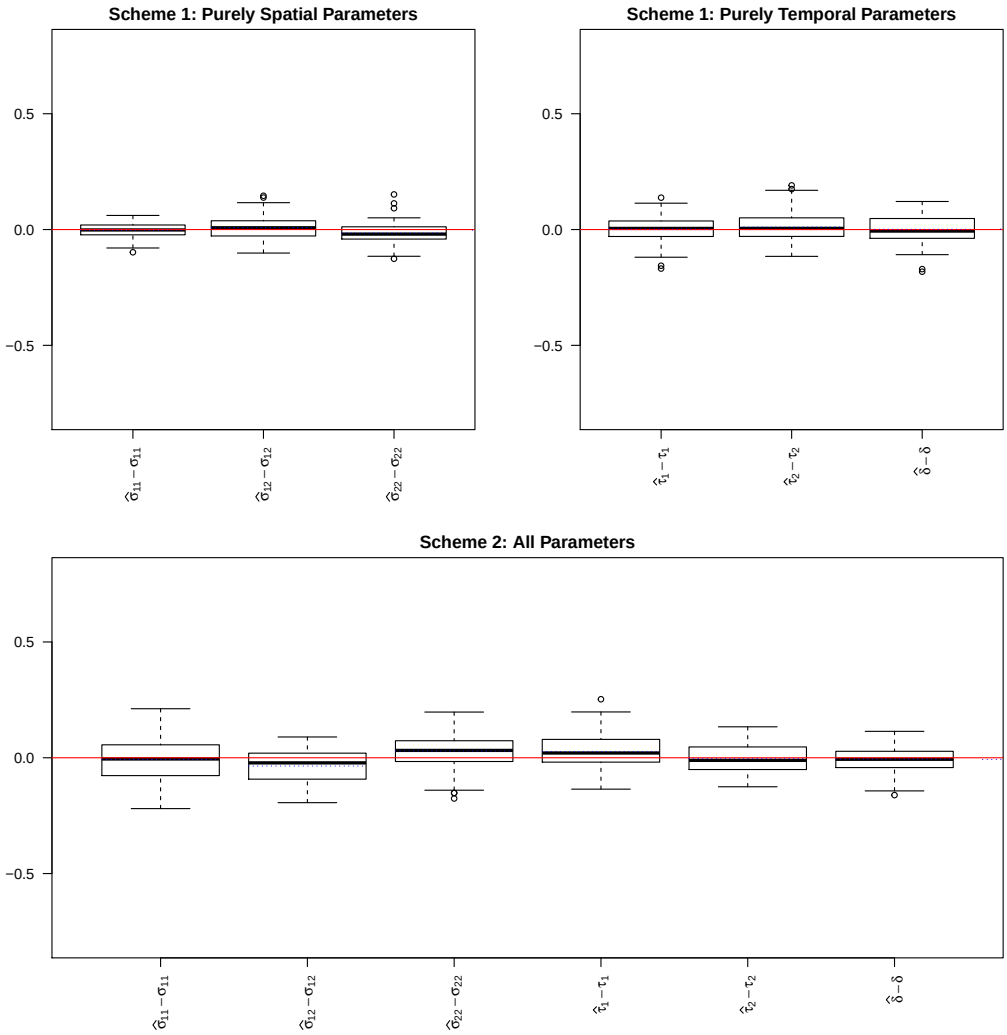


Figure 5.14: Boxplots of the errors $\widehat{\boldsymbol{\psi}}-\boldsymbol{\psi}$ resulting from both estimation schemes for 100 simulated spectrally separable space-time max-stable Smith processes with parameter $\boldsymbol{\psi} = (1,0,1,1,1,0.7)^t$. Scheme 1 (Top row): purely spatial parameters (left panel) and purely temporal parameters (right panel). Scheme 2 (Bottom row): all parameters. The middle blue dotted/red solid lines show the overall mean of errors estimates/zero value.

spatio-temporal pairwise log-likelihood is defined by (see, [41])

$$\ell_p(\boldsymbol{\psi}) = \sum_{k=1}^{T-1}\sum_{l=k+1}^{T}\sum_{i=1}^{D-1}\sum_{j=i+1}^{D} \omega_{ij}\,\omega_{kl}\log f_{(s_i,t_k)(s_j,t_l)}(x_{ik},x_{jl};\boldsymbol{\psi}),$$

where ω_{ij}, ω_{lk} are nonnegative spatial and temporal weights, respectively, and x_{ik} denotes the observation of X at site i and date k. The maximum pairwise likelihood estimator is given by $\widehat{\boldsymbol{\psi}} = \text{argmax}\ \ell_p(\boldsymbol{\psi})$. The obtained results have shown that, the estimation of purely spatial parameters is more accurate with a two-step approach.

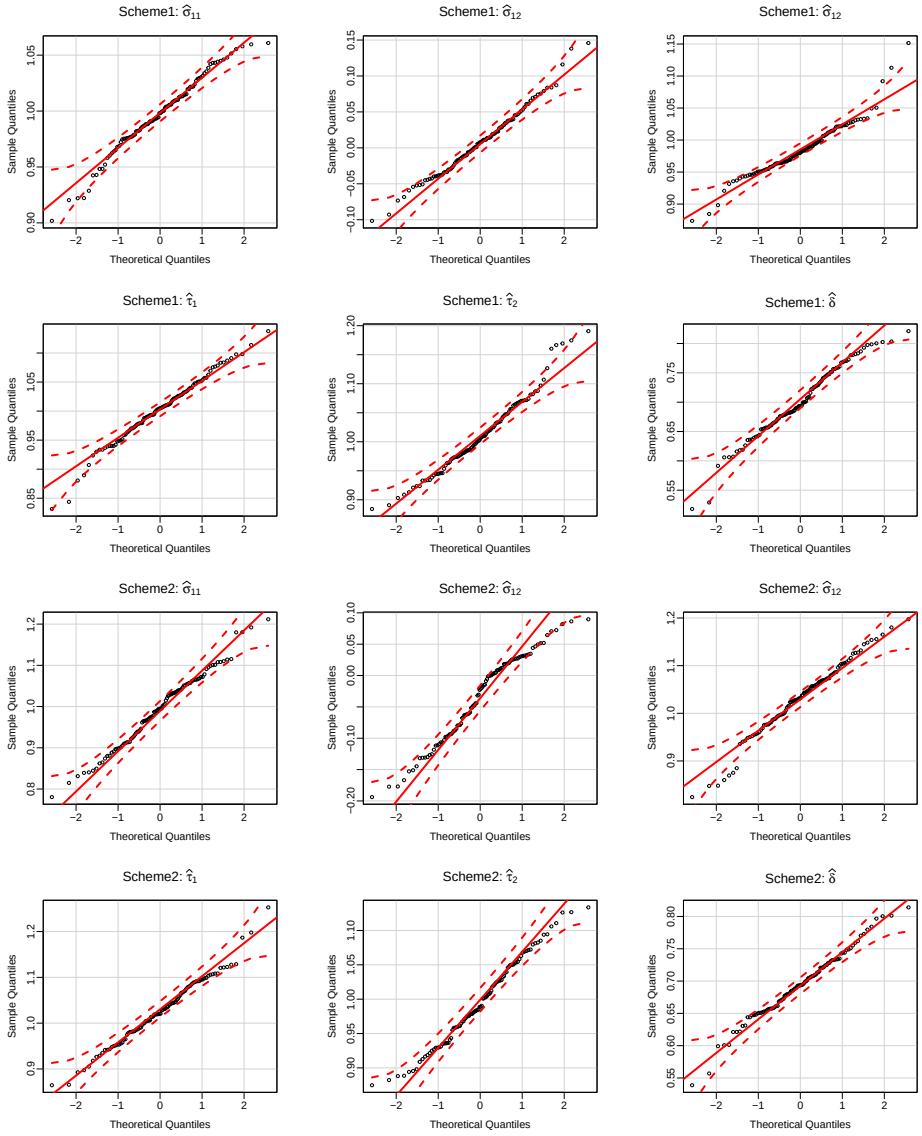


Figure 5.15: QQ-plots of the estimates from both estimation schemes for 100 simulated spectrally separable space-time max-stable Smith processes with parameter $\boldsymbol{\psi} = (1, 0, 1, 1, 1, 0.7)^t$ against the normal distribution. Scheme 1: purely spatial parameters (top row) and purely temporal parameters (second row). Scheme 2: purely spatial parameters (third row) and purely temporal parameters (bottom row). Dashed red lines correspond to 95% confidence intervals.

5.3.3 Simulation study 3: Fitting spectrally separable spatio-temporal Schlather process

Finally, we perform a third simulation study to fit spectrally separable space-time max-stable Schlather process. The innovation process H is derived from independent replications of a spatial Schlather process (recall 2.13) with correlation function of powered exponential type defined, for all $\|\boldsymbol{h}\| \geq 0$, by $\rho(\boldsymbol{h}) = \exp[-(\|\boldsymbol{h}\|/\phi)^{\kappa}]$, $\phi > 0$ and $0 < \kappa < 2$, where ϕ and κ denote, respectively, the range and the smoothing parameters. We denote by $\boldsymbol{\psi} = (\phi, \kappa, \tau_1, \tau_2, \delta)^t$

the vector gathering the model parameters. We take $\phi = 3$, $\kappa = 3/2$, $\boldsymbol{\tau} = (1,0)^t$ and $\delta = 0.3$. As previously, we consider the same simulation setup used in Section 5.3.2. The results are summarized in Figure 5.16 and Table 5.3. Generally, we obtain equally satisfying results.

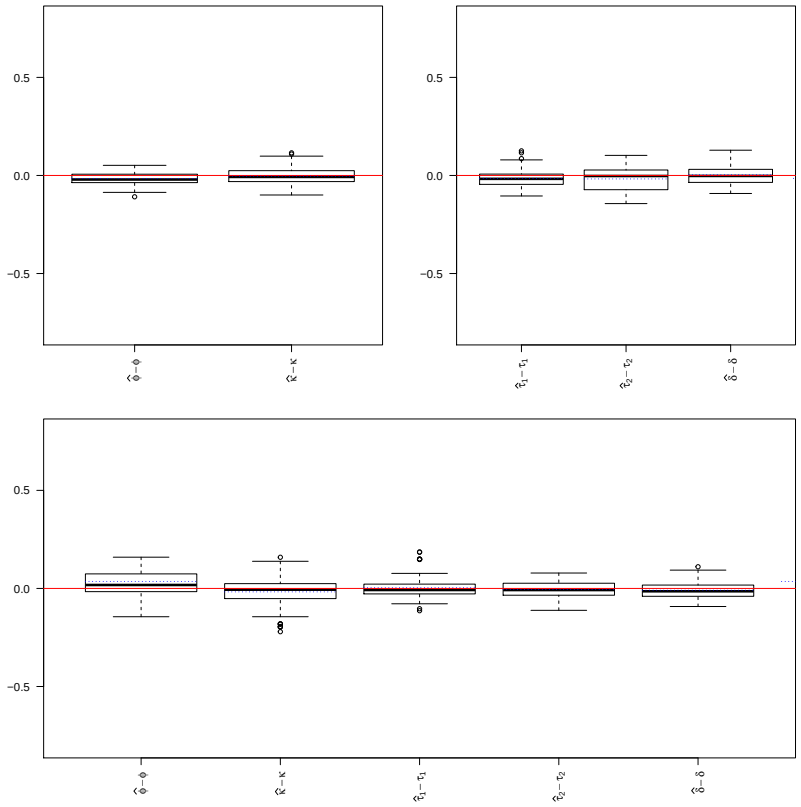


Figure 5.16: Boxplots of errors $\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}$ from both estimation schemes for 100 simulated spectrally separable STMS Schlather processes with parameter $\boldsymbol{\psi} = (2, 1.5, 1, 0, 0.3)^t$. Scheme 1 (Top row): purely spatial parameters (left panel) and purely temporal parameters (right panel). Scheme 2 (Bottom row): all parameters. The middle blue dotted/red solid lines show the overall mean of errors estimates/zero value.

Table 5.3: Performance of the estimation for 100 simulated spectrally separable STMS Schlather processes, with parameter $\boldsymbol{\psi} = (2, 1.5, 1, 0, 0.3)^t$. The mean estimate, RMSE, and MAE of the estimated parameters.

	Scheme 1			Scheme 2		
True	Mean estimate	RMSE	MAE	Mean estimate	RMSE	MAE
Purely Spatial						
$\phi = 2$	1.9841	0.0368	0.0309	2.0357	0.0812	0.0599
$\kappa = 1.5$	1.4967	0.0407	0.0327	1.4814	0.0771	0.0557
Purely temporal						
$\tau_1 = 1$	0.9852	0.0442	0.0346	1.0036	0.0556	0.0393
$\tau_2 = 0$	−0.0177	0.0636	0.0512	−0.0053	0.0427	0.0353
$\delta = 0.3$	0.3031	0.0473	0.0383	0.2913	0.0393	0.0318

5.4 Real data analysis

In this section, we aim to quantify the extremal behavior of radar rainfall data in a region located in the State of Florida. Our approach is to fit the data by different space-time max-stable classes based on a space-time block maxima design using the proposed semi-parametric estimation procedure.

5.4.1 Description of the dataset

The dataset analyzed in this section is composed of radar rainfall values (in inches) measured on a square of 140 × 140 km region containing 4900 grid locations in the State of Florida. The database consists of radar hourly rainfall values measured on a regular grid with squared cells of size 2 km covering a region of 70 × 70 cells in the State of Florida. A map of the study area is shown in Figure 5.17. We only consider the wet season (June-September) over the years 2007−2012. The data were collected by the Southwest Florida Water Management District (SWFWMD) and freely available on ftp://ftp.swfwmd.state.fl.us/pub/radar_rainfall. Moreover, the dataset is available in the Supplementary Material: http://math.univ-lyon1.fr/homes-www/abuawwad/Florida_RadarRainfall/.

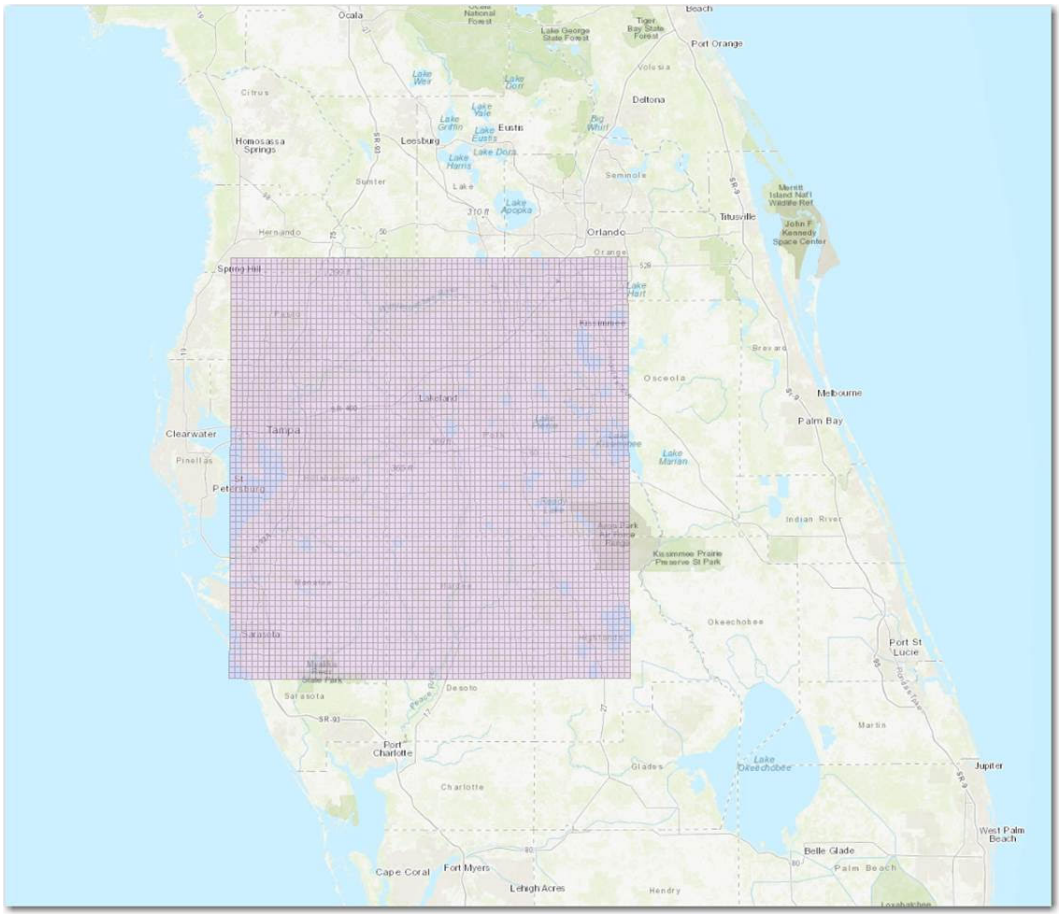


Figure 5.17: Radar rainfall observation area in the State of Florida. Source: Southwest Florida Water Management District (SWFWMD).

5.4.2 Data fitting

We perform a block maxima design in space and time as follows: we take block maxima over 24 consecutive hours and over 10 km × 10 km areas (the daily maxima over 25 grid locations), resulting in 14×14 grid in space for all 6×122 days of the wet seasons. So, this gives a time series of dimension 14×14 and of length 732. For the sake of notational simplicity, we denote the set of resulting grid locations by $\mathbb{S} = \{(x, y) : x, y \in \{1, \ldots, 14\}\}$ and the space-time realizations by $\{X(\boldsymbol{s}, t), \boldsymbol{s} \in \mathbb{S},\ t \in \{t_1, \ldots, t_{732}\}\}$. This setup has been also considered in [23, 41] for analyzing radar rainfall measurements in a region from the State of Florida over the years 1999−2004, where only space-time max-stable BR process has been fitted to the data by a semi-parametric approach in [23] and a pairwise likelihood approach in [41]. Let us remark that both regions here and in the above-mentioned two studies are located in the central portion of Florida District, which would probably have the best square area of coverage. Having larger grid size will lead to some cells missing in the southwestern 'corner' due to the coastline. Figure 5.19 shows the obtained time series for daily maxima observations at four grid locations. One location lies on the boundary of the study region and three are located in the interior, see Figure 5.18.

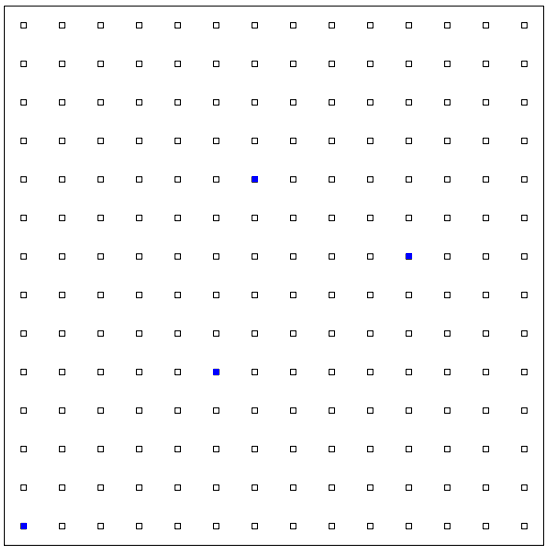


Figure 5.18: Blue square-symbols show the reference locations for the time series plotted in Figure 5.19.

As has already been mentioned in Chapter 3, when we are interested in modeling the joint occurrence of extremes over a region, then the dependence structure of a multivariate variable has to be explicitly stated. The usual modeling strategy consists of two steps: firstly, estimating the marginal distribution. Secondly, characterizing the dependence via a model issued by the multivariate extreme value theory, see, e.g., [14, 84]. For marginal modeling, we explain the procedure as follows:

(i) We transform the data to stationarity by removing possible seasonal effects using a simple moving average with a period of 122 days (the number of days in the wet season considered

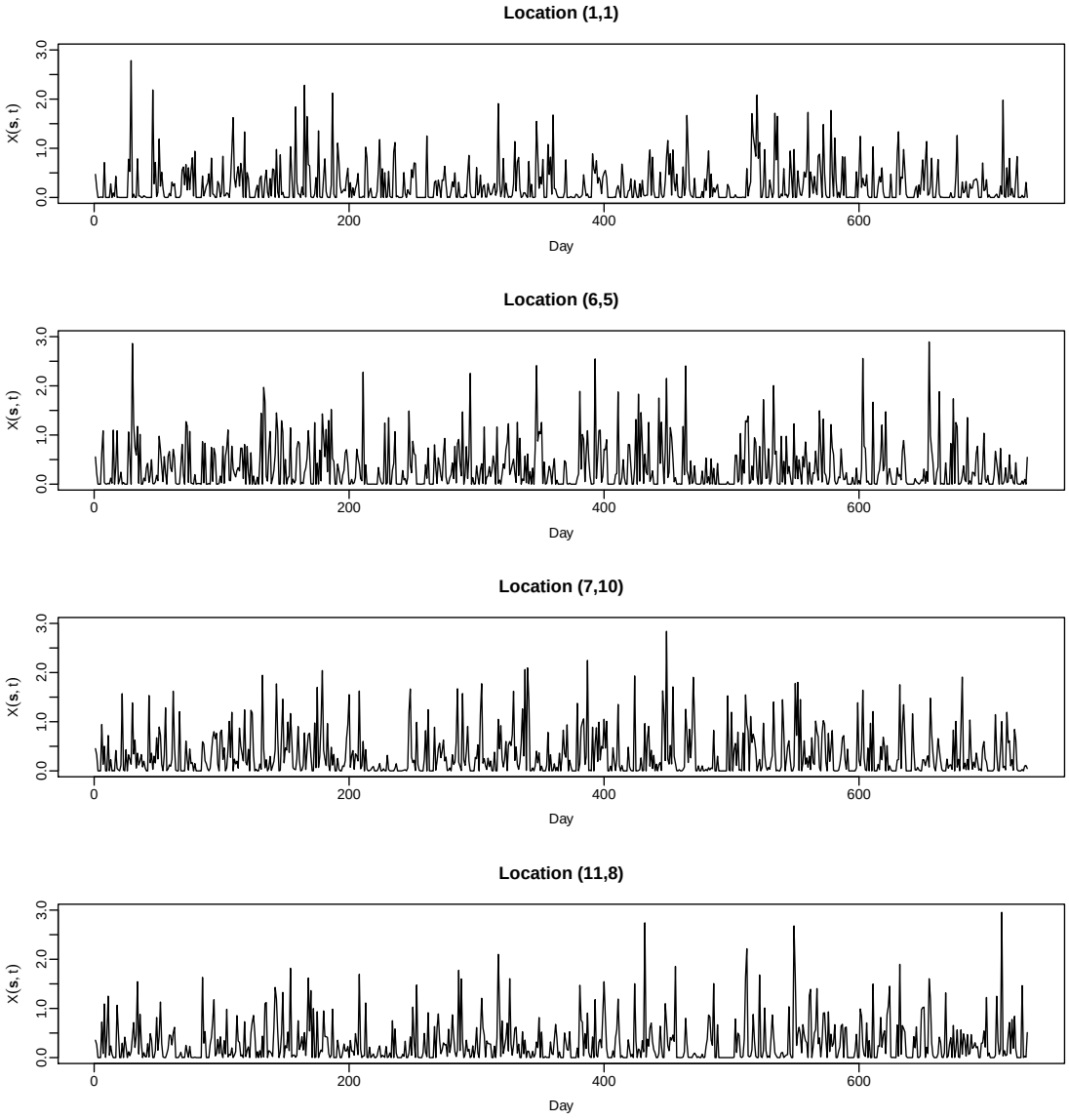


Figure 5.19: Plots of daily maximal rainfall in inches for four grid locations with simplified coordinates: (1,1), (6,5), (7,10) and (11,8).

in one particular year). More precisely, for each fixed location $\boldsymbol{s} \in \mathbb{S}$, we deseasonalize the time series $\{X(\boldsymbol{s},t), t \in \{t_1,\ldots,t_{732}\}\}$ by computing

$$\tilde{X}(\boldsymbol{s},t_{i+122(j-1)}) = X(\boldsymbol{s},t_{i+122(j-1)}) - \frac{1}{6}\sum_{j=1}^{6} X(\boldsymbol{s},t_{i+122(j-1)}), \text{ for } i = 1,\ldots,122. \quad (5.25)$$

(ii) For each fixed location $\boldsymbol{s} \in \mathbb{S}$, the deseasonalized observations are fitted to the generalized extreme value distribution,

$$\mathrm{GEV}_{\mu(\boldsymbol{s}),\sigma(\boldsymbol{s}),\xi(\boldsymbol{s})}(x) = \exp\left\{-\left[1+\xi(\boldsymbol{s})\left(\frac{x-\mu(\boldsymbol{s})}{\sigma(\boldsymbol{s})}\right)\right]^{-1/\xi(\boldsymbol{s})}\right\}, \quad (5.26)$$

for some location $\mu(\boldsymbol{s}) \in \mathbb{R}$, scale $\sigma(\boldsymbol{s}) > 0$, and shape $\xi(\boldsymbol{s}) \in \mathbb{R}$. The estimated shape parameters $\xi(\boldsymbol{s})$ are sufficiently close to zero with confidence interval containing zero, see Figure 5.20. This suggests a Gumbel distribution (GEV with $\xi = 0$) as appropriate model. Therefore, we fit directly a Gumbel distribution $\mathrm{GEV}_{\mu(\boldsymbol{s}),\sigma(\boldsymbol{s}),0}(x) = \left\{\exp\left[-\exp\left(-\frac{x-\mu(\boldsymbol{s})}{\sigma(\boldsymbol{s})}\right)\right]\right\}$. For each spatial location, we assess the goodness of the marginal fits by QQ-plots of deseasonalized rainfall series versus the fitted Gumbel distribution. The results at four spatial locations $(1,1),(6,5),(7,10)$ and $(11,8)$ are summarized in Figure 5.21. All plots provide a plausible fit.

(iii) The deseasonalized observations may be transformed either to standard Gumbel or standard Fréchet margins. More precisely, let $\widehat{\mu}(\boldsymbol{s})$, $\widehat{\sigma}(\boldsymbol{s})$ are the parameter estimates obtained from (ii), then we may use:

(a) $\tilde{\tilde{X}}(\boldsymbol{s},t) = \frac{\tilde{X}(\boldsymbol{s},t)-\widehat{\mu}(\boldsymbol{s})}{\widehat{\sigma}(\boldsymbol{s})}$, $t \in \{1,\ldots,732\}$ to transform the deseasonalized observations to standard Gumbel margins;

(b) $\tilde{\tilde{X}}(\boldsymbol{s},t) = -\frac{1}{\log\{\mathrm{GEV}_{\widehat{\mu}(\boldsymbol{s}),\widehat{\sigma}(\boldsymbol{s}),0}(\tilde{X}(\boldsymbol{s},t))\}}$, $t \in \{1,\ldots,732\}$ to transform the deseasonalized observations to standard Fréchet margins. This transformation is called the probability integral transformation. In this study, we adopt this case.

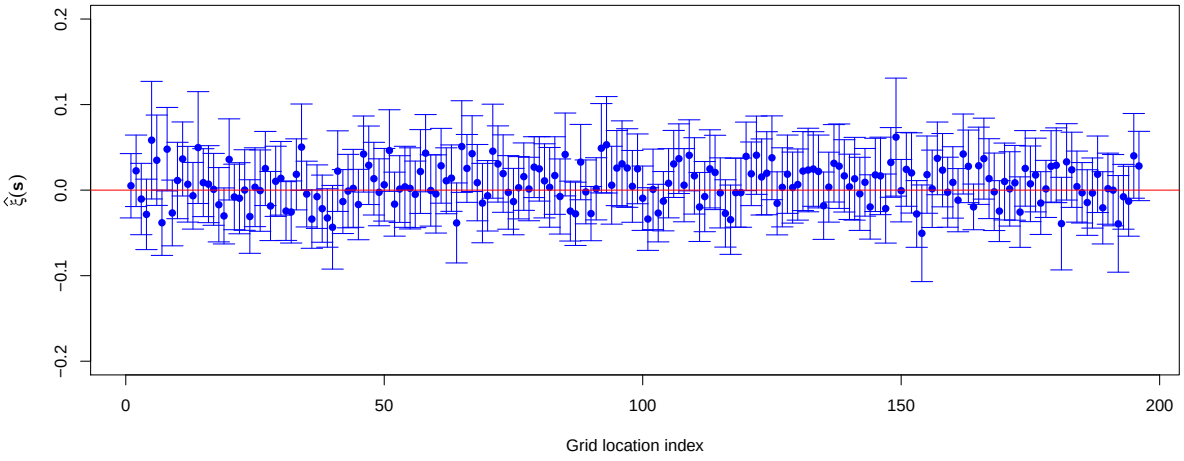


Figure 5.20: Estimated GEV shape parameter $\widehat{\xi}(\boldsymbol{s})$ at all grid locations with 95% confidence intervals.

In [23, 41], the authors assume that the observations $\tilde{\tilde{X}}(\boldsymbol{s},t)$ are realizations from the space-time max-stable BR process. The contribution of the present section is to broaden the dependence structure by considering the spectrally separable space-time max-stable processes, that allow interactions between spatial and temporal components.

In the sequel, we estimate the extremal dependence structure for the daily maxima of rainfall measurements. Based on our findings in the simulation studies, we notice that Scheme 1 outperforms Scheme 2 generally. So, one may first estimate the extremal dependence parameters using Scheme 2. Afterward, re-estimating the parameters using Scheme 1, where the estimates

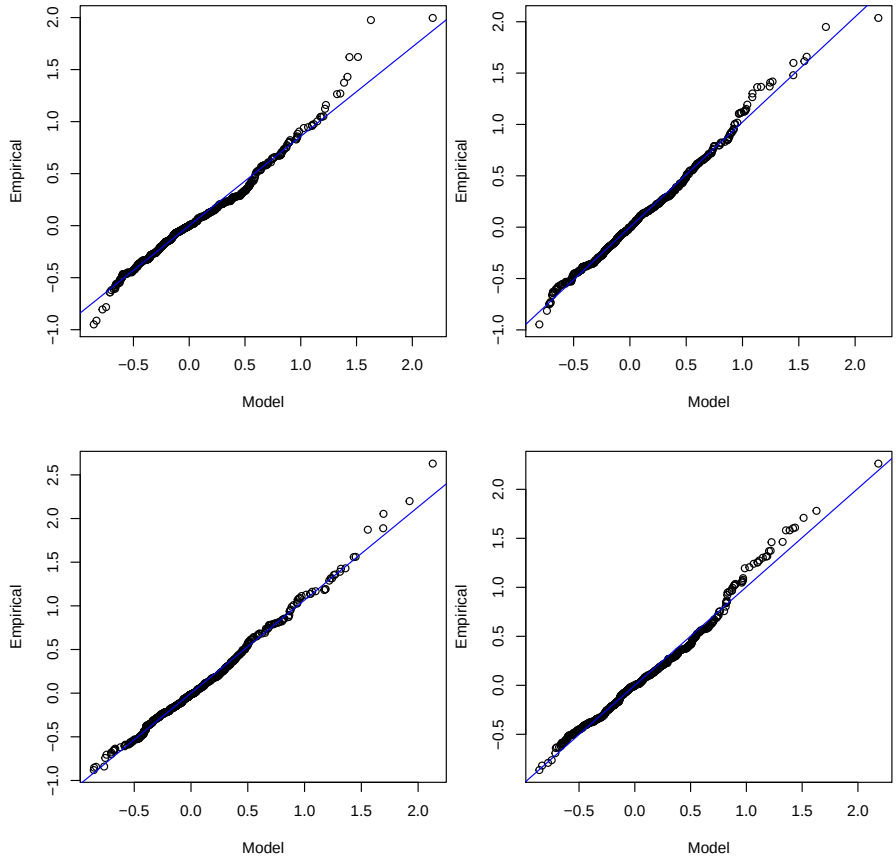


Figure 5.21: QQ-plots of deseasonalized rain series versus the fitted Gumbel distribution (GEV with $\widehat{\mu}(\boldsymbol{s})$, $\widehat{\sigma}(\boldsymbol{s})$ and 0) on the basis of the time series corresponding to the four grid locations shown in Figure 5.19.

resulting from Scheme 2 serve as starting values for the optimization routine used in Scheme 1. To that aim, we consider the following five spatio-temporal max-stable models:

(i) Class A: consists of two non-spectrally separable models A_1 and A_2.

- A_1: a space-time max-stable BR model (2.49), with dependence function $\gamma(\boldsymbol{h}, l) = 2\phi_s \|\boldsymbol{h}\|^{\kappa_s} + 2\phi_t l'^{\kappa_t}$, where $l' = |l|$, recall Example 5.2.
- A_2: a space-time max-stable Schlather model (2.13). The space-time correlation function is chosen to be separable such that $\rho(\boldsymbol{h}, l) = \exp\{-[(\|\boldsymbol{h}\|/\phi_s)^{\kappa_s} + (l'/\phi_t)^{\kappa_t}]\}$, where the range parameters $\phi_t, \phi_s > 0$ and the smoothing parameters $0 < \kappa_t, \kappa_s < 2$.

(ii) Class B: consists of spectrally separable models B_1, B_2, and B_3.

- B_1: a spectrally separable space-time max-stable model (5.4), where the innovation process H is derived from independent replications of a spatial BR process with semivariogram $\gamma(\boldsymbol{h}) = (\|\boldsymbol{h}\|/\phi)^{\kappa}$, for some range parameter $\phi > 0$ and smoothness parameter $\kappa \in (0, 2]$. Obviously, models A_1 and B_1 are equivalent when $l' = 0$.

- B_2: a spectrally separable space-time max-stable model (5.4), where the innovation process H is derived from independent replications of a spatial Smith process with covariance matrix $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$, recall Example 5.3.
- B_3: a spectrally separable space-time max-stable model (5.4), where the innovation process H is derived from independent replications of a spatial extremal-t process (2.16) with degrees of freedom $\nu \geq 1$ and correlation function of powered exponential type, defined for all $\|\boldsymbol{h}\| \geq 0$, by $\rho(\boldsymbol{h}) = \exp[-(\|\boldsymbol{h}\|/\phi)^{\kappa}]$, $\phi > 0$ and $0 < \kappa < 2$, where ϕ and κ denote, respectively, the range and the smoothing parameters.

To select the best-fitting model, we use the Akaike Information Criterion (AIC) which was first developed by [6] under the framework of maximum likelihood estimation. The AIC is one of the most widely used methods for selecting a best-fitting model from several competing models given a particular dataset. A concise formulation of the AIC under the framework of least squares estimation has been derived by [13]. The AIC under Scheme 1 is defined as

$$\mathrm{AIC}_{\mathrm{NLS}} = |\mathcal{H}| \log\left(\frac{\mathcal{L}(\widehat{\boldsymbol{\psi}}^{(s)})}{|\mathcal{H}|}\right) + 2(k_s + 1) + |\mathcal{K}| \log\left(\frac{\mathcal{L}(\widehat{\boldsymbol{\psi}}^{(t)})}{|\mathcal{K}|}\right) + 2(k_t + 1), \tag{5.27}$$

where $\mathcal{L}(\widehat{\boldsymbol{\psi}}^{(s)})$ and $\mathcal{L}(\widehat{\boldsymbol{\psi}}^{(t)})$ are the estimated objective functions in space and time with $\omega^{\boldsymbol{h}} = \omega^{l'} = 1$, i.e.,

$$\mathcal{L}(\widehat{\boldsymbol{\psi}}^{(s)}) = \sum_{\|\boldsymbol{h}\| = h \in \mathcal{H}} \left(\widehat{\nu}_F(\boldsymbol{h}) - \nu_F^{(\boldsymbol{s})}(\boldsymbol{h}, \widehat{\boldsymbol{\psi}}^{(s)})\right)^2, \ h \in \mathcal{H},$$

$$\mathcal{L}(\widehat{\boldsymbol{\psi}}^{(t)}) = \sum_{l' \in \mathcal{K}} \left(\widehat{\nu}_F(l') - \nu_F^{(t)}(l', \widehat{\boldsymbol{\psi}}^{(t)})\right)^2, \ l' \in \mathcal{K}.$$

where $|\mathcal{A}|$ denotes the cardinality of the set $\mathcal{A}$, and k_s and k_t are, respectively, the total number of purely spatial and purely temporal parameters in the underlying model. If $|\mathcal{H}|/k_s + 1 < 40$ and $|\mathcal{K}|/k_t + 1 < 40$, it is suggested to use an adjusted "corrected" version of $\mathrm{AIC}_{\mathrm{NLS}}$ (5.27), see [13], i.e.,

$$\mathrm{AIC}_{\mathrm{NLS}_c} = \mathrm{AIC}_{\mathrm{NLS}} + \frac{2(k_s + 1)(k_s + 2)}{|\mathcal{H}| - k_s} + \frac{2(k_t + 1)(k_t + 2)}{|\mathcal{K}| - k_t}. \tag{5.28}$$

Our results are summarized in Table 5.4. Model A_1 has the lowest $\mathrm{AIC}_{\mathrm{NLS}_c}$ value and therefore would be considered as the best candidate for this dataset, closely followed by model B_3. Obviously, the temporal estimates ($\widehat{\phi}_t$ and $\widehat{\kappa}_t$) in the best-fitting model A_1 indicate that there is a weak temporal extremal dependence. Recall that the purely temporal F-madogram for this model is given by

$$\nu_F^{(t)}(l') = 0.5 - \left\{2\Phi\left(\sqrt{\phi_t l'^{\kappa_t}}\right) + 1\right\}^{-1}, \ l' > 0.$$

Accordingly, $\nu_F^{(t)}(l')$ is close to $1/6$ for large values of ϕ_t, indicating asymptotic independence. On the other hand, $\nu_F^{(t)}(l')$ is approximately constant when κ_t is small, indicating that the extremal dependence is the same for all l'. So, both large ϕ_t and small κ_t lead to temporal asymptotic independence.

For comparison, we present the semi-parametric estimates obtained by [23]; $\widehat{\phi}_s = 0.3611, \widehat{\kappa}_s = 0.9876, \widehat{\phi}_t = 2.3650$ and $\widehat{\kappa}_t = 0.0818$. On the other hand, the pairwise likelihood estimates obtained by [41] are $\widehat{\phi}_s = 0.3485, \widehat{\kappa}_s = 0.8858, \widehat{\phi}_t = 2.4190$ and $\widehat{\kappa}_t = 0.1973$. Obviously, these estimates are close to our estimates, except the temporal smoothness estimate $\widehat{\kappa}_t$ which is relatively large.

Figure 5.22 shows the empirical values of $\nu_F(h),\ h \in \mathcal{H}$ and $\nu_F(l'),\ l' \in \mathcal{K}$, and their model-based counterparts from the three best-fitting models according to the AIC_{NLS}. It seems that the three models give a quite reasonable fit with a little outperformance for model A_1. So, considering these plots and the AIC_{NLS} values there is overall evidence in favor of model A_1.

Table 5.4: Summary of the fitted models based on the block maxima design from the radar rainfall measurements in a region in the State of Florida.

Model	Purely spatial parameters	Purely temporal parameters	$\text{AIC}_{\text{NLS}_c}$
A_1	$\widehat{\phi}_s = 0.4109,\ \widehat{\kappa}_s = 0.9527$	$\widehat{\phi}_t = 2.1686,\ \widehat{\kappa}_t = 0.5410$	**−64.3921**
A_2	$\widehat{\phi}_s = 2.6023,\ \widehat{\kappa}_s = 1.2600$	$\widehat{\phi}_t = 2.1902,\ \widehat{\kappa}_t = 0.3464$	−43.0257
B_1	$\widehat{\phi} = 1.2289,\ \widehat{\kappa} = 0.9527$	$\widehat{\boldsymbol{\tau}} = (-0.2990, 0.1661)^t,$ $\widehat{\delta} = 0.5821$	**−58.7364**
B_2	$\widehat{\sigma}_{11} = 3.7253,\ \widehat{\sigma}_{12} = -0.4181,$ $\widehat{\sigma}_{22} = 4.2100$	$\widehat{\boldsymbol{\tau}} = (0.5379, -0.1452)^t,$ $\widehat{\delta} = 0.1830$	−21.4420
B_3	$\widehat{\phi} = 5.9293,\ \widehat{\kappa} = 1.2491,$ $\widehat{\nu} = 6.0820$	$\widehat{\boldsymbol{\tau}} = (1.4074, 0.8505)^t,$ $\widehat{\delta} = 0.5317$	**−59.7906**

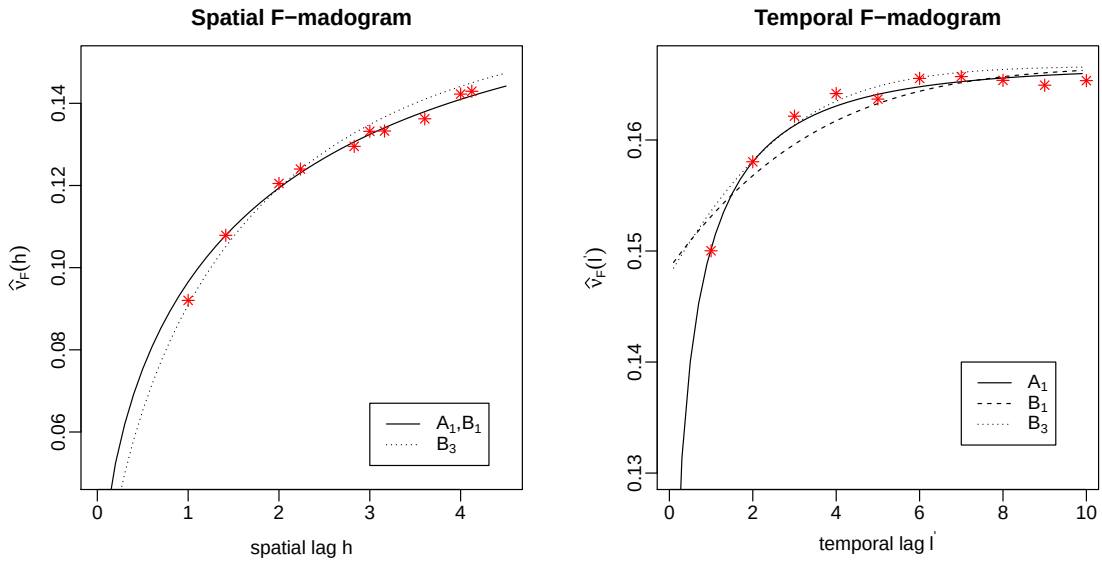


Figure 5.22: Red star-symbols show the empirical values of $\nu_F(h)$ and $\nu_F(l')$ used for estimation. The curves show the fitted $\nu_F(h)$ and $\nu_F(l')$ from the three best-fitting models (A_1, B_1 and B_3).

Lastly, permutation tests (or sometimes called randomization tests) can be useful to determine the range of clear dependence. So, in order to examine whether the extremal dependence

in space and time is significant, we perform a permutation test. We randomly permute the space-time data and compute the empirical spatial/temporal F-madograms. More precisely, to check how the extremal dependence lasts in space, for each fixed time point $t \in \{t_1, \ldots, t_{732}\}$ we permute the spatial locations. Afterward, the spatial F-madogram is computed and the procedure is repeated 1000 times. From the obtained spatial F-madogram sample, we compute 97.5% and 2.5% empirical quantiles which form a 95% confidence region for spatial extremal independence. On the other hand, to test the presence of temporal extremal independence, the analog procedure is done for the temporal F-madogram. In particular, for each fixed location $\boldsymbol{s} \in \mathbb{S} = \{(x, y) : x, y \in \{1, \ldots, 14\}\}$ we sample without replacement from the corresponding time series and compute the empirical temporal F-madogram. Our findings are shown in Figure 5.23 together with the fitted values of spatial/temporal F-madograms derived from the best-fitting models A_1. Inspecting these plots, it appears that the spatial extremal dependence vanishes for spatial lags larger than four (the fitted values for the spatial F-madogram lies within the obtained independence confidence region), whereas the temporal extremal dependence vanishes for time lags larger than three. The same conclusions are obtained in [23], where the permutation tests have been carried out based on the extremogram.

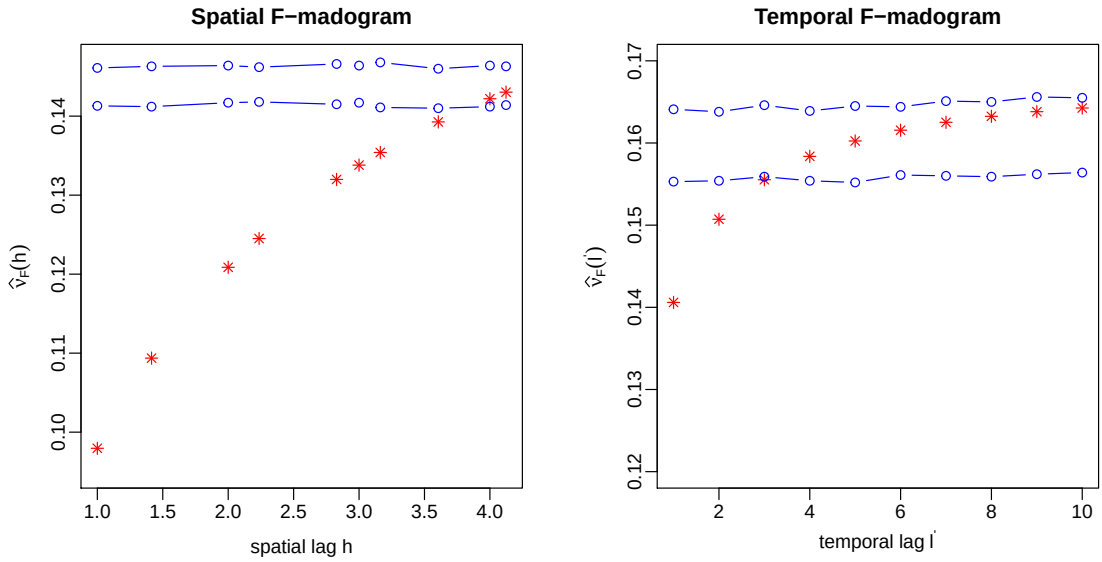


Figure 5.23: Permutation test for extremal independence in space (left panel) and time (right panel). Upper/lower blue lines show 97.5% and 2.5% quantiles of empirical F-madograms for 1000 spatial (right) and temporal (left) permuations of the space-time observations. Red star-symbols show the fitted values of $\nu_F(h)$, $h \in \mathcal{H}$ and $\nu_F(l')$, $l' \in \mathcal{K}$ derived from the best-fitting model A_1.

5.5 Conclusions

In summary, we proposed two novel and flexible semi-parametric estimation schemes for space-time max-stable processes based on the spatio-temporal F-madogram, $\nu_F(\boldsymbol{h}, l)$. Working with

the madogram has a few advantages. In addition to its simple definition and the computational facility, it has a clear link with extreme value theory throughout the extremal dependence function. The new estimation procedure may be employed as an alternative or a prerequisite to the widely used pairwise likelihood; the semi-parametric estimates could serve as starting values for the optimization routine used to maximize the pairwise log-likelihood function to decrease the computational time and also improve the statistical efficiency, see [26].

A simulation study has shown a visual overview of the quality of the proposed inference procedure. Overall, the procedure performs well. Moreover, our estimation methodology outperforms the semi-parametric estimation procedure suggested by [23] which was based on the dependence measure extremogram. The introduced method is applied to radar rainfall measurements in a region in the State of Florida (Section 5.4) in order to quantify the extremal properties of the space-time observations. Finally, our attention is concentrated on fitting space-time max-stable processes based on gridded datasets. In the future, we plan to generalize our method in order to fit space-time max-stable processes with extensions to irregularly spaced locations that may have a fundamental interest in practice.

Chapter 6

Conclusions and outlook

6.1 Concluding remarks

In summary, we have considered hypothesis testing for the mixing parameter a of a spatial max-mixture process using two pairwise likelihood-based statistics: Z_a and LR_a. Pairwise marginal likelihood appears to be an attractive tool for modeling complex data and has received increasing attention in handling high dimensional datasets when the joint distribution is computationally difficult to evaluate, or intractable due to the complex structure of dependence. Afterward, we have suggested a madogram-based selection criterion for the mixing parameter a of a spatial max-mixture process. Finally, we have also explored a madogram-based inference procedure for some spatio-temporal max-stable processes.

Hypothesis testing presents some difficulties: Wald-type tests lack invariance under reparameterizations, score-type tests suffer from numerical instability, whereas the asymptotic null distribution of the composite likelihood ratio statistic depends both on the statistical model and on the dimension of the parameter of interest. It involves a linear combination of independent chi-squared variables with coefficients given by the eigenvalues of a matrix related to Godambe information matrix. Many different adjustments of the composite likelihood ratio statistic have been proposed to recover the usual chi-squared distribution which depends only on the dimension of the parameter of interest. A regularity condition underlying these adjustments is that the parameter of interest is interior to its parameter space. However, this is usually not satisfied where the parameter of interest lies on the boundary. Accordingly, testing whether the data are asymptotically dependent or independent, i.e., $H_0 : a = a_0,\ a_0 \in \{1, 0\}$ are much more tricky to deal with because standard asymptotic theory does not apply. For this reason, we applied the statistics when the model is near to asymptotic dependence (i.e., we tested $H_0 : a = 0.99$) or asymptotic independence (i.e., we tested $H_0 : a = 0.01$). In a simulation study, the two statistics Z_a and LR_a showed a reasonable performance. They can control the type I error rate α. The LR_a test seems to be more powerful than the Z_a one when the true mixing parameter $a \in (0, 1)$, while the contrary seems to hold for $a \in \{0, 1\}$. Furthermore, the power to reject asymptotic dependence, i.e., $H_0 : a = 1$ (respectively, asymptotic independence, i.e., $H_0 : a = 0$) improves as the tested value $a_0 \to 0$ (respectively, $a_0 \to 1$). We applied our testing framework in an analysis of exceedances over a large threshold of daily rainfall data from the East of Australia,

where pairwise likelihood is adopted for estimating the dependence parameters. We showed that our testing procedure could be an efficient model validation tool on a. In addition, our testing framework could be generalized to test for all model parameters. Hence, providing a tool for model checking in the validation stage.

Furthermore, we have derived a closed form expression of the F^{λ}-madogram $\nu_F^{\lambda}(\boldsymbol{h})$ in the framework of spatial max-mixture models. It can detect more than one dependence structure in the model: asymptotic dependence and independence. Indeed, the F^{λ}-madogram presents the advantage of having both extremal dependence functions of the max-mixture model components by its formula, i.e., $\theta_X(\boldsymbol{h})$ and $\theta_Y(\boldsymbol{h})$. Afterward, motivated by defects in existing model-based inference approaches for spatial extremes, we have developed an estimation procedure for max-mixture model parameters based on $\nu_F^{\lambda}(\boldsymbol{h})$. It consists of using a moment-based approach to estimate the mixture parameter a and the bivariate dependence summaries for the two processes (max-stable, inverted max-stable). The estimation algorithm proceeds by estimating those bivariate dependence summaries for fixed values of the mixture parameter a, and then the best value of a is selected by profiling the values of a goodness-of-fit criterion with respect to a. Unlike the estimation techniques used so far in this context, an appealing feature of this approach is that the estimation of max-mixture model parameters can be performed without specifying a parametric distribution family prior to fitting the model. In practice, the main advantage of this strategy is to reduce the effect resulting from inaccurate choices of stochastic processes used for describing the joint tail distribution. Such inaccurate choices may lead to severe mis-estimation of probabilities associated with simultaneous extreme events. However, the proposed procedure does not allow for estimating the entire process: only the mixing parameter and the bivariate summaries of the two components in the mixture are estimated. So, it can be seen as an exploratory tool to assess (in a preliminary step) what kind of parametric max-mixture model would be appropriate to fit the data. Also, to perform model checking at the validation step. In a simulation study, the proposed procedure performs well, even when we have considered the boundary values for a. It improves the empirical dependence summaries for the class of max-mixture models. In an analysis of monthly maxima of Australian daily rainfall data, we implemented the F^{λ}-madogram estimation procedure for diagnostic and confirmatory purposes.

Finally, we have also proposed a semi-parametric estimation procedure for space-time max-stable processes based on the spatio-temporal F-madogram, $\nu_F(\boldsymbol{h}, l)$. In a simulation study, the proposed procedure results in quite reliable estimates. It outperforms the semi-parametric procedure suggested by [23] in which the extremogram was employed for inferential purposes. Then, we applied our proposed inference framework to analyze radar daily rainfall maxima measurements in a region of Florida. The proposed procedure could be used as an alternative or a prerequisite to likelihood-based methods, where the semi-parametric estimates could serve as starting values for the optimization routine used to maximize the pairwise log-likelihood function in order to reduce the computational time and also improves the statistical efficiency for estimators.

6.2 Future research

To conclude the thesis, we mention some remaining issues which are subject to future research.

- In Chapter 4, besides the proposed estimation approach for bivariate dependence summaries of the two processes, i.e., $\theta_X(\boldsymbol{h})$ and $\theta_Y(\boldsymbol{h})$, let us remark that a maximum likelihood-based method could also be implemented as an alternative parametric approach. Indeed, using the idea of profiling goodness-of-fit criterion with respect to the value of the mixture parameter a could also be used to devise a composite likelihood procedure for estimating the mixture parameter a and the bivariate dependence summaries, similar to the moment-based procedure presented. More precisely, for a fixed value of the mixture parameter a, the functions $\theta_X(\boldsymbol{h})$ and $\theta_Y(\boldsymbol{h})$ can be estimated by maximizing a composite likelihood based on setting $z = z_1 = z_2$ in (2.35).

- Anisotropic situations may have a fundamental interest in practice. Especially, environmental phenomena often have strong anisotropic spatial variation when the regions monitored are very large, see [8]. Our suggestion is to introduce anisotropic models. There are two types of anisotropy: geometric or elliptical, and zonal. An easy way to introduce spatial anisotropy to a model is given by geometric anisotropy, see [53]. For example, spatial geometrically anisotropic models may be obtained by replacing the Euclidean distance $\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$ with the Mahalanobis distance, i.e.,

$$h^2_{\mathcal{M}} = (\boldsymbol{s}_1 - \boldsymbol{s}_2)^t \begin{pmatrix} \cos(\varrho) & -\sin(\varrho) \\ \sin(\varrho) & \cos(\varrho) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & b^2 \end{pmatrix} \begin{pmatrix} \cos(\varrho) & -\sin(\varrho) \\ \sin(\varrho) & \cos(\varrho) \end{pmatrix}^t (\boldsymbol{s}_1 - \boldsymbol{s}_2), \tag{6.1}$$

  with $b > 0$ and $\varrho \in [-\pi, \pi]$. The parameter b reflects the degree of anisotropy, as it corresponds to the ratio of the principal axes of dependence contours, whereas ϱ is the angle with respect to the west-east direction. When $b = 1$, the model is isotropic, see [18, 66].

- Equally weighted inference approaches have been widely implemented. However, using non-constant weights seems appealing for at least two reasons. First from a computational point of view, for example discarding distant pairs, the central processing unit (CPU) load for the evaluation might be smaller and the fitting procedure would be less time-consuming. On the other hand, as neighboring pairs are expected to be strongly dependent, thus providing valuable information for the estimation of dependence parameters, this may improve the statistical efficiency. Therefore, it could be interesting to investigate the gain in statistical efficiency of estimators as well as computational efficiency by adopting different weighting strategies. For the semi-parametric estimation of space-time max-stable processes, since the number of spatial and temporal lags are limited, we could consider weights such that locations and time points which are further apart from each

other have less influence on the estimation, i.e.,

$$\begin{aligned}
\omega^{\boldsymbol{h}} &= \exp\{-c_1\|\boldsymbol{h}\|\} \text{ or } \exp\left\{-c_1\|\boldsymbol{h}\|^2\right\} \text{ or } \|\boldsymbol{h}\|^{-c_1},\\
\omega^{l'} &= \exp\{-c_2 l'\} \text{ or } \exp\left\{-c_2 l'^2\right\} \text{ or } l'^{-c_2},\\
\omega^{\boldsymbol{h},l'} &= \exp\{-c\,(\|\boldsymbol{h}\|+l')\} \text{ or } \exp\left\{-c\left(\|\boldsymbol{h}\|^2+l'^2\right)\right\} \text{ or } (\|\boldsymbol{h}\|+l')^{-c},
\end{aligned}$$

where $c_1,\ c_2,\ c>0,\ \|\boldsymbol{h}\|\in\mathcal{H}$ and $l'\in\mathcal{K}$.

- Owing to our motivating radar rainfall dataset, we focused on fitting space-time max-stable processes based on gridded datasets using the spatio-temporal F-madogram. We plan to generalize our method in order to fit space-time max-stable processes with extensions to irregularly spaced locations (typically, a more realistic scenario). For non-gridded sampling locations, the estimators need to be modified to account for the fact that very few or no pairs of locations will be separated by a specific spatial lag, $\boldsymbol{h}$. One solution to this challenge is to specify a distance tolerance (ε) such that lags having length $\|\boldsymbol{h}\|+\varepsilon$ are included in estimating the F-madogram at lag $\boldsymbol{h}$. We could also use the λ-madogram to estimate the spatio-temporal extremal dependence function $V_{\boldsymbol{h},l}$. Furthermore, it could be interesting to extend the proposed F^{λ}-madogram estimation approach in Chapter 4 to fit spatio-temporal max-stable processes with unknown extremal dependence class. Hence, providing better model selection at an exploratory stage and better model checking at the validation stage.

- Asymptotic properties of our pairwise dependence estimators were analyzed numerically. However, still, much more theoretical research has to be undertaken to establish the asymptotic properties of these estimators. Some helpful ideas can be found in [81, 23].

- Finally, an important modeling issue is that asymptotic independence cannot be captured by the spatio-temporal max-stable processes. To handle this issue, one possibility is to extend the spatial max-mixture model to the spatio-temporal setting. However, this will lead to a highly parameterized spatio-temporal model. Accordingly, novel classes of spatial extremal models described by a small number of parameters and making a smooth transition between the two dependence paradigms have been proposed recently in the literature. For instance,

  1. let $\{W(\boldsymbol{s})\}_{s\in\mathcal{S}}$ be a stationary spatial process with standard Pareto margins, which satisfies for any $x\geq 1$

  $$\mathbb{P}\{W(\boldsymbol{s}_j)>x\}=x^{-1},$$

  $$\mathbb{P}\{W(\boldsymbol{s}_j)>x, W(\boldsymbol{s}_k)>x\}=\mathcal{L}_W(x)x^{-1/\eta_W(\boldsymbol{h})},\ j\neq k,$$

  where $\mathcal{L}_W(.):(0,\infty)\to(0,\infty)$ is slowly varying at infinity and the parameter $\eta_W(\boldsymbol{h})$ with $\boldsymbol{h}=\boldsymbol{s}_j-\boldsymbol{s}_k$ is the coefficient of tail dependence (recall Section 2.2.2). It summarizes the joint tail decay of the process $W(\boldsymbol{s})$. With $W(\boldsymbol{s})$ as described, let

R be a standard Pareto random variable independent of W. A spatial dependence model is defined by [67] through the random field constructed as

$$X(\boldsymbol{s}) = R^{\alpha} W(\boldsymbol{s})^{1-\alpha},\ \alpha \in [0,1].$$

According to this construction: when $\alpha > 0.5$ then R^{α} is heavier-tailed than $W^{1-\alpha}$ and this induces asymptotic dependence; the converse is true when $\alpha < 0.5$, and this induces asymptotic independence.

2. let $\{W(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ be a standard Gaussian process with correlation function $\rho(\boldsymbol{s}_1, \boldsymbol{s}_2)$ and $R \sim F(r)$ is a positive random variable. A Gaussian scale mixture (GSM) process (i.e., a Gaussian process with random variance) is defined as follows:

$$X(\boldsymbol{s}) = RW(\boldsymbol{s}),\ \boldsymbol{s} \in \mathcal{S}.$$

Several new models have been proposed in [66] based on GSM models that provide a transition from one asymptotic dependence regime to the other. For example, R can be assumed to have a two-parameter Weibull-type distribution, for $r \geq 1$

$$F(r) = \begin{cases} 1 - \exp\left\{-\gamma(r^{\beta} - 1)/\beta)^{\alpha}\right\}, & \beta > 0, \\ 1 - r^{-\gamma}, & \beta = 0. \end{cases}$$

Therefore, GSM models can add modeling flexibility to spatial extreme analysis. They have non-Gaussian marginals and can exhibit asymptotic dependence unlike Gaussian processes, which are asymptotically independent except in the case of perfect dependence.

Hence, it could be interesting to adapt these approaches to the spatio-temporal setting.

Bibliography

[1] Abrahamsen, P. (1997). A review of Gaussian random fields and correlation functions, Technical Report 917, Norwegian Computing Center, Blindern, Norway.

[2] Abu-Awwad, A.; Maume-Deschamps, V. and Ribereau, P. Censored pairwise likelihood-based tests for mixture parameter of spatial max-mixture models, *accepted for publication to Revista de Investigacion Operacional*.

[3] Abu-Awwad, A.; Maume-Deschamps, V. and Ribereau, P. Fitting spatial max-mixture processes with unknown extremal dependence class: an exploratory analysis tool, *published online to TEST journal*.

[4] Abu-Awwad, A.; Maume-Deschamps, V. and Ribereau, P. Semi-parametric estimation for space-time max-stable processes: F-madogram-based estimation approach, *submitted for publication*.

[5] Ahmed, M.; Maume-Deschamps, V.; Ribereau, P. and Vial, C. A semi-parametric estimation for max-mixture spatial processes, *submitted for publication*.

[6] Akaike, H. (1974). A new look at the statistical model identification, *IEEE transactions on automatic control*, 19(6): 716–723.

[7] Antoniadis, A.; Berruyer, J. and René, C. (1992). *Régression non linéaire et applications*, Economica, Paris.

[8] Arbia, G. and Lafratta, G. (2002). Anisotropic spatial sampling designs for urban pollution, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 51(2):223–234.

[9] Bacro, J-N.; Bel, L. and Lantuéjoul, C. (2010). Testing the independence of maxima: from bivariate vectors to spatial extreme fields, *Extremes*, 13(2):155–175.

[10] Bacro, J-N. and Gaetan, C. (2014). Estimation of spatial max-stable models using threshold exceedances, *Statistics and Computing*, 24(4):651–662.

[11] Bacro, J-N.; Gaetan, C. and Toulemonde, G. (2016). A flexible dependence model for spatial extremes, *Journal of Statistical Planning and Inference*, 172:36–52.

[12] Bacro, J-N. and Toulemonde, G. (2013). Measuring and modelling multivariate and spatial dependence of extremes, *Journal de la Société Française de Statistique*, 154(2):139–155.

[13] Banks, H.T. and Joyner, M.L. (2017). AIC under the framework of least squares estimation, *Applied Mathematics Letters*, 74:33–45.

[14] Beirlant, J.; Goegebeur, Y.; Segers, J. and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*, John Wiley, NewYork.

[15] Bel, L.; Bacro, J-N. and Lantuéjoul, C. (2008). Assessing extremal dependence of environmental spatial fields, *Environmetrics*, 19(2):163–182.

[16] Beniston, M.; Stephenson, D.B.; Christensen, O.B.; Ferro, C.A.; Frei, C.; Goyette, S.; Halsnaes, K.; Holt, T.; Jylhä, K.; Koffi, B. and Palutikof, J. (2007). Future extreme events in European climate: an exploration of regional climate model projections, *Climatic change*, 81(1):71–95.

[17] Besag J. (1974). Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society. Series B (Methodological)*, 192–236.

[18] Blanchet, J. and Davison, A.C. (2011). Spatial modeling of extreme snow depth, *The Annals of Applied Statistics*, 1699–1725.

[19] Bevilacqua, M.; Gaetan, C.; Mateu, J. and Porcu, E. (2012). Estimating space and space-time covariance functions for large data sets: a weighted composite likelihood approach, *Journal of the American Statistical Association*, 107(497), 268–280.

[20] Brown, B.M. and Resnick, S.I. (1977). Extreme values of independent stochastic processes, *Journal of Applied Probability*, 14:732–739.

[21] Brown, P.E.; Diggle, P.J.; Lord, M.E. and Young, P.C. (2001). Space-Time Calibration of Radar Rainfall Data, *Journal of the Royal Statistical Society: Series C (AppliedStatistics)*, 50(2):221–241.

[22] Brown, P.E.; Karesen, K.F.; Roberts, G.O. and Tonellato, S. (2000). Blur-Generated Non-Separable Space-Time Models, *Journal of the Royal Statistical Society: Series B (StatisticalMethodology)*. 62(4):847–860.

[23] Buhl, S.; Davis, R.A.; Klüppelberg, C. and Steinkohl, C. Semiparametric estimation for isotropic max-stable space-time processes, *submitted for publication*.

[24] Buhl, S. and Klüppelberg, C. (2016). Anisotropic Brown-Resnick space-time processes: estimation and model assessment, *Extremes*. 19(4):627–660.

[25] Capéraà, P.; Fougères, A.L. and Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas, *Biometrika*, 84:567–577.

[26] Castruccio, S.; Huser, R. and Genton, M.G. (2016). High-order composite likelihood inference for max-stable distributions and processes, *Journal of Computational and Graphical Statistics*, 25:1212–1229.

[27] Cattelan, M. and Sartori, N. (2016). Empirical and simulated adjustments of composite likelihood ratio statistics, *Journal of Statistical Computation and Simulation*, 86(5):1056–1067.

[28] Chandler, R.E. and Bate, S. (2007). Inference for clustered data using the independence loglikelihood, *Biometrika*, 167–183.

[29] Chen, Y. and Liang, K.Y. (2010). On the asymptotic behaviour of the pseudolikelihood ratio test statistic with boundary problems, *Biometrika*, 97(3):603–620.

[30] Chen, Y.; Ning, J.; Ning, Y.; Liang, K.Y. and Bandeen-Roche, K. (2017). On pseudolikelihood inference for semiparametric models with boundary problems, *Biometrika*, 104(1):165–179.

[31] Cho, Y.B.; Davis, R.A. and Ghosh, S. (2016). Asymptotic properties of the empirical spatial extremogram, *Scandinavian Journal of Statistics*, 43(3):757–773.

[32] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer, London.

[33] Coles, S.; Heffernan, J. and Tawn, J. (1999). Dependence measures for extreme value analyses, *Extremes*, 2(4):339–365.

[34] Coles, S. and Tawn, J. (1991). Modelling extreme multivariate events, *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(2):377–392.

[35] Cooley, D.; Naveau, P. and Poncet, P. (2006). Variograms for spatial max-stable random fields, *Dependence in probability and statistics*, 373–390.

[36] Cressie, N. (1993). *Statistics for spatial data*, New York: Wiley.

[37] Cressie, N. and Wikle, C. K. (2015). *Statistics for spatio-temporal data*, John Wiley and Sons.

[38] Davies, R.B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, 64(2):247–254.

[39] Davies, R.B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternatives, *Biometrika*, 33–43.

[40] Davis, R.A.; Klüppelberg, C. and Steinkohl, C. (2013). Max-stable processes for modeling extremes observed in space and time, *Journal of the Korean Statistical Society*, 42(3):399–414.

[41] Davis, R.A.; Klüppelberg, C. and Steinkohl, C. (2013). Statistical inference for max-stable processes in space and time, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(5):791–819.

[42] Davis, R.A. and Mikosch, T. (2009). The extremogram: A correlogram for extreme events, *Bernoulli*, 15(4):977–1009.

[43] Davis, R.A. and Resnick, S. I. (1989). Basic properties and prediction of max-ARMA processes, *Advances in applied probability*, 21(4):781–803.

[44] Davison, A.C. and Gholamrezaee M.M. (2012). Geostatistics of extremes, *Proc. R. Soc. A*, 468:581–608.

[45] Davison, A.C.; Huser, R. and Thibaud, E. (2013). Geostatistics of dependent and asymptotically independent extremes, *Mathematical Geosciences*, 45(5):511–529.

[46] Davison, A.C.; Padoan, S.A. and Ribatet, M. (2012). Statistical modeling of spatial extremes, *Statistical science*, 161–186.

[47] Davison, A.C. and Smith, R.L. (1990). Models for exceedances over high thresholds, *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3):393–425.

[48] De Cesare, L.; Myers, D. E. and Posa, D. (2001). Product-sum covariance for space-time modeling: an environmental application, *Environmetrics*, 12(1) 11–23.

[49] De Haan, L. (1984). A spectral representation for max-stable processes, *The annals of probability*, 1194–1204.

[50] De Haan, L. and Pereira, A. (2006). *Extreme Value Theory: An Introduction*, Springer, NewYork.

[51] De Haan, L. and Pereira, T.T. (2006). Spatial extremes: Models for the stationary case, *The annals of statistics*, 146–168.

[52] Dombry, C.; Engelke, S. and Oesting, M. (2016). Exact simulation of max-stable processes, *Biometrika*, 103(2):303–317.

[53] Ecker, M.D. and Gelfand, A.E. (2003). Spatial modeling and prediction under stationary non-geometric range anisotropy, *Environmental and Ecological Statistics*, 10(2):165–178.

[54] Embrechts, P.; Koch, E. and Robert, C. (2016). Space–time max-stable models with spectral separability, *Advances in Applied Probability*, 48:77–97.

[55] Fougères A.-L. (2004). Multivariate Extremes. In *Extreme Values in Finance, Telecommunications, and the Environment*, Ed. B. Finkenstädt and H. Rootzèn, 373–388. Chapman Hall/CRC, New York.

[56] Gaetan, C. and Guyon, X. (2010). *Spatial statistics and modeling*, Springer, New York.

[57] Gelfand, A.E. and Schliep, E.M. (2016). Spatial statistics and Gaussian processes: a beautiful marriage, *Spatial Statistics*, 18(A):86–104.

[58] Genton, M.G.; Ma, Y. and Sang, H. (2011). On the likelihood function of Gaussian max-stable processes, *Biometrika*, 98:481–488.

[59] Gneiting, T. (2002). Nonseparable, stationary covariance functions for space–time data, *Journal of the American Statistical Association*, 97(458):590–600.

[60] Guillou, A.; Naveau, P. and Schorgen, A. (2014). Madogram and asymptotic independence among maxima, *REVSTAT-Statistical Journal*, 12(2).

[61] Heagerty, P.J., and Lele, S.R. (1998). A composite likelihood approach to binary spatial data, *Journal of the American Statistical Association*, 93(443):1099–1111.

[62] Heffernan, J.E. (2000). A directory of coefficients of tail dependence, *Extremes*, 3(3):279–290.

[63] Huser, R. and Davison, A.C. (2013). Composite likelihood estimation for the Brown-Resnick process, *Biometrika*, 100(2):511–518.

[64] Huser, R. and Davison, A.C. (2014). Space-time modelling of extreme events, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2):439–461.

[65] Huser, R.; Davison, A.C. and Genton, M.G. (2016). Likelihood estimators for multivariate extremes, *Extremes*, 19(1):79–103.

[66] Huser, R.; Opitz, T. and Thibaud, E. (2017). Bridging asymptotic independence and dependence in spatial extremes using gaussian scale mixtures, *Spatial Statistics*, 21:166–186.

[67] Huser, R. and Wadsworth, J.L. (2018). Modeling spatial processes with unknown extremal dependence class, *Journal of the American Statistical Association*, 1–11.

[68] Kabluchko, Z.; Schlather, M. (2010). Ergodic properties of max-infinitely divisible processes, *Stochastic Processes and their Applications*, 120(3):281–295.

[69] Kabluchko, Z.; Schlather, M. and De Haan, L. (2009). Stationary max-stable fields associated to negative definite functions, *The Annals of Probability*, 2042–2065.

[70] Kent, J.T. (1982). Robust properties of likelihood ratio tests, *Biometrika*, 69(1):19–27.

[71] Ledford, A.W. and Tawn, J.A. (1996). Statistics for near independence in multivariate extreme values, *Biometrika*, 83(1):169–187.

[72] Ledford, A.W. and Tawn, J.A. (1997). Modelling dependence within joint tail regions, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(2):475–499.

[73] Lindsay, B.G. (1988). Composite likelihood methods, *Contemporary mathematics*, 80(1):221–239.

[74] Marcon, G.; Padoan, S. A.; Naveau, P.; Muliere, P. and Segers, J. (2017). Multivariate nonparametric estimation of the Pickands dependence function using Bernstein polynomials, *Journal of Statistical Planning and Inference*, 183:1–17.

[75] Matérn, B. (2013). *Spatial variation*, Vol. 36. Springer, New York.

[76] Matheron, G. (1962). *Traité de géostatistique appliquée*, Éditions Technip, Paris.

[77] Matheron, G. (1987). Suffit-il, pour une covariance, d'être de type positif, *Sciences de la Terre, série informatique géologique*, 26:51–66.

[78] Matheron, G. (1989). *Estimating and choosing*, Springer, New York.

[79] Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*, Springer, New York.

[80] Montero, J.M. and Mateu, J. (2015). *Spatial and spatio-temporal geostatistical modeling and kriging*, John Wiley & Sons.

[81] Naveau, P.; Guillou, A.; Cooley, D. and Diebolt, J. (2009). Modelling pairwise dependence of maxima in space, *Biometrika*, 96:1–17.

[82] Opitz, T. (2013). Extremal t processes: Elliptical domain of attraction and a spectral representation, *Journal of Multivariate Analysis*, 122:409–413.

[83] Pace, L.; Salvan, A. and Sartori, N. (2011). Adjusting composite likelihood ratio statistics, *Statistica Sinica*, 129–148.

[84] Padoan, S.A.; Ribatet, M. and Sisson, S.A. (2010). Likelihood-based inference for max-stable processes, *Journal of the American Statistical Association*, 105(489):263–277.

[85] Pickands, J. (1981). Multivariate extreme value distributions (with discussion), In *Bulletin de l'Institut International de Statistique*, 49(2):859–878, 894–902.

[86] Ribatet, M. (2015). SpatialExtremes: R Package, version 2.0–2, Available at http://spatialextremes.r-forge.rproject.org/.

[87] Ribatet, M. and Sedki, M. (2013). Extreme value copulas and max-stable processes, *Journal de la Société Française de Statistique*, 154:138–150.

[88] Rotnitzky, A. and Jewell, N.P. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data, *Biometrika*, 485–497.

[89] Sang, H., and Genton, M.G.(2014). Tapered composite likelihood for spatial max-stable models, *Spatial Statistics*, 8:86–103.

[90] Schlather, M. (2002). Models for stationary max-stable random fields, *Extremes*, 5(1):33–44.

[91] Schlather, M. (2001). RandomFields, contributed package on random field simulation for R, Available at *https://CRAN.R-project.org/package=RandomFields*.

[92] Schlather, M. and Tawn, J.A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference, *Biometrika*, 90(1):139–156.

[93] Sibuya, M. (1960). Bivariate extreme statistics, *Annals of the Institute of Statistical Mathematics*, 11(2):195–210.

[94] Smith, R.L. (1990). Max-stable processes and spatial extremes, *Unpublished manuscript 205*.

[95] Stein, M. L. (2012). *Interpolation of spatial data: some theory for kriging*, Springer, New York.

[96] Thibaud, E.; Mutzner, R. and Davison, A.C. (2013). Threshold modeling of extreme spatial rainfall, *Water resources research*, 49(8):4633–4644.

[97] Thibaud, E. and Opitz, T. (2015). Efficient inference and simulation for elliptical Pareto processes, *Biometrika*, 102(4):855–870.

[98] Varin, C. (2008). On composite marginal likelihoods, *AStA Advances in Statistical Analysis*, 92(1):1–28.

[99] Varin, C.; Reid, N. and Firth, D. (2011). An overview of composite likelihood methods, *Statistica Sinica*, 5–42.

[100] Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection, *Biometrika*, 519–528.

[101] Wadsworth, J.L. and Tawn, J.A. (2012). Dependence modelling for spatial extremes, *Biometrika*, 99(2):253–272.

[102] Wadsworth, J.L. and Tawn, J.A. (2013). Efficient inference for spatial extreme value processes associated to log-Gaussian random functions, *Biometrika*, 101(1):1–15.

[103] Yaglom, A.M. (1987). *Correlation Theory of Stationary and Related Random Functions*, Springer, New York.

Sur l'inférence statistique pour des processus spatiaux et spatio-temporels extrêmes

Résumé. Cette thèse traite de l'inférence statistique pour les événements extrêmes dans le cadre spatial et spatio-temporel. Dans la partie spatiale, nous proposons deux tests sur le paramètre de mélange a d'un processus spatial max-mélange : le test statistique Z_a et le rapport de vraisemblance par paire LR_a. Une étude de simulations permete d'avoir une vision globale de leurs performances. Nous appliquons ces tests dans le cadre d'une analyse d'excès au delà d'un grand seuil pour des données de précipitations dans l'Est de l'Australie. Nous proposons aussi une nouvelle procédure d'estimation, basée sur l'utilisation de la méthode des moindres carrés sur le F^λ-madogramme, pour ajuster des processus spatiaux max-mélanges lorsqu'on ne connait pas la classe de dépendance extrêmal. La nouveauté de cette procédure est qu'elle permet de faire de l'inférence sans spécifier au préalable la famille de distributions, laissant ainsi parle les données et guider l'estimation. Nous prouvons la convergence des estimateurs obtenus. Une indication sur la normalité asymptotique est donnée numériquement. Une étude sur simulation montrent que la méthode proposée améliore les coefficients empiriques pour la classe de modèles max-mélange. Nous implémentons notre procédure d'estimations sur des données de maximas mensuels de précipitations en Australie dans un but exploratoire et confirmatoire. Dans la partie spatio-temporelle, nous proposons une méthode d'estimation semi-paramétrique pour les processus max-stables spatio-temporels en nous basant sur une expression explicite du F-madogramme spatio-temporel. En particulier, pour des observations sur grille régulière, nous estimons le F-madogramme spatio-temporel par sa version empirique et nous appliquons une procédure basée sur les moments pour obtenir les estimations des paramètres d'intérêt. Nous illustrons les performances de cette procédure par une étude sur simulations. Ensuite, nous appliquons cette méthode pour quantifier le comportement extrêmal de maximum de données radar de précipitations dans l'Etat de Floride. Cette méthode peut être une alternative ou une première étape pour la vraisemblance composite.

Mots-clés : Dépendance/Indépendance asymptotique, vraisemblance composite, événement extrême, F^λ-madogramme, processus max-stable, processus max-mélange, précipitations, estimation semi-paramétrique, processus max-stable spatio-temporel.

On statistical inference for spatial and spatio-temporal extreme processes

Abstract. This thesis deals with the statistical inference of extreme events in both spatial and spatio-temporal settings. In the spatial part, we consider hypothesis testing for the mixture parameter of a spatial max-mixture model using two classical statistics: the Z-test statistic and the pairwise likelihood ratio statistic. A simulation study gives a visual overview of their performances. Afterward, we apply this testing framework in an analysis of exceedances over a large threshold of daily rainfall data from the East of Australia. Using nonlinear least squares fit based on the F^λ-madogram, we also propose a novel estimation procedure to fit spatial max-mixture processes with unknown extremal dependence class. The novelty of this procedure is to let the data speak for themselves. We establish the consistency of the estimators. An indication for asymptotic normality is given numerically. A simulation study shows that this procedure improves empirical coefficients for this class of models. In an analysis of monthly maxima of Australian daily rainfall data, we implement this procedure for diagnostic and confirmatory purposes. In the spatio-temporal part, based on the spatio-temporal F-madogram, we suggest a semi-parametric estimation approach for space-time max-stable processes. For regular grid observations, the F-madogram is estimated nonparametrically by its empirical version and a moment-based procedure is applied to obtain the estimates of interest. The performance of the method is investigated by various simulation studies. Lastly, we apply this method to quantify the extremal behavior of radar daily rainfall maxima data from a region in Florida. This approach could serve as an alternative or a prerequisite to pairwise likelihood estimation.

Keywords: Asymptotic dependence/independence, composite likelihood, extreme event, F^λ-madogram, max-stable process, max-mixture process, rainfall data, semi-parametric estimation, space-time max-stable process.

ED 512
INFOMATHS
Ecole doctorale

UNIVERSITÉ DE LYON