Spatiotemporal modeling of extreme events and point patterns

Thomas Opitz

Spatiotemporal modeling of extreme events and point patterns

Document scientifique pour l'habilitation à diriger des recherches en sciences

Thomas Opitz
Biostatistique et Processus Spatiaux, INRAE, 84914, Avignon, France

Section CNU 2600 – Mathématiques appliquées et applications des mathématiques

Aix-Marseille Université – Faculté des Sciences

June 24, 2021

Thanks

This manuscript distills the results of my research projects that I have undertaken with many inspiring colleagues at BioSP, at other INRAE units, at other institutions in France, and abroad. There have been many visits, conversations and discussions, and some of the colleagues have become friends over the years. I would like to thank all them wholeheartedly.

I am also very grateful to the members of my habilitation committee for the considerable time they have invested in assessing my work of research, and in particular also for the many relevant and interesting questions and comments they have brought up in our discussion.

Finally, I would like to say very big thank you to my family and friends who have always supported me, and most of all to Caroline and my son Paul.

Habilitation committee

The committee was composed as follows:

- Véronique Maume-Deschamps, Professeure des Universités, Université Claude Bernard Lyon 1 (Referee)
- Philippe Naveau, Directeur de recherche, CNRS (Referee)
- Jonathan Tawn, Distinguished Professor, Lancaster University (Referee)
- Clément Dombry, Professeur des Universités, Université de Franche-Comté (Examinator)
- Nicolas Eckert, Directeur de Recherche, INRAE (Examinator)
- Frédéric Richard, Professeur des Universités, Aix-Marseille Université (Examinator)
- Thordis L. Thorarinsdottir, Chief Research Scientist, Norwegian Computing Center (Examinator)

Scientific preamble

My personal perception of probability and statistics throughout my work is strongly driven by the practical utility that one may derive from it. This requires to address a readership that is as wide as possible. Therefore I may sometimes give preference to a rather intuitive presentation of problems, techniques and results, while I intend to avoid jargon that I deem as overly technical and therefore not useful. Albert Einstein's famous words

Everything Should Be Made as Simple as Possible, But Not Simpler

should always serve as guidance when thinking about how science should be conducted and communicated.

I seek to develop new theory and methodology motivated by the relevant practical issues raised by my colleagues in the applied sciences, where the discussion of methods and results should be done in a simple – yet clear and consistent – language to be fruitful for both sides.

While the main purpose of this manuscript is to provide a summary of my research activities over the past years, I have also written many parts of it to serve as a general introduction to numerous aspects of state-of-the-art spatiotemporal modeling of extreme and rare events.

Contents

Résumé long en français

Mes travaux de recherche contribuent à la méthodologie de modélisation et d'estimation en statistique spatiale et spatio-temporelle. Dans ce cadre probabiliste, je développe des applications pour répondre à des défis en sciences environnementales, climatiques, écologiques et agro-épidémiologiques, dans le but d'inférer les facteurs explicatifs des phénomènes observés, et de fournir des prédictions, souvent sous forme d'une cartographie spatiale ou spatio-temporelle. Je me concentre sur des phénomènes rares et extrêmes, typiquement caractérisés par une probabilité d'occurrence relativement faible ou des magnitudes relativement extrêmes. De tels événements entrainent souvent des impacts importants sur les systèmes climatiques, écologiques et anthropiques. En raison de cette focale, mes approches se situent dans la plupart des cas au-delà de la géostatique classique centrée sur les processus gaussiens, et elles possèdent une assise solide dans deux domaines développés par des communautés mathématiques relativement indépendantes, à savoir la géométrie stochastique et la théorie des valeurs extrêmes. La géométrie stochastique fournit les modèles de processus ponctuels, bien adaptés pour modéliser les positions et temps d'occurrence d'événements d'intérêt. Dans le contexte des risques environnementaux, elle fournit ainsi une perpective de modélisation à une échelle souvent relativement macroscopique en opérant avec des données discrètes de semis de points. En revanche, la théorie des valeurs extrêmes traite des processus à variables continues et met à disposition des modèles à fondement asymptotique pour capter les caractéristiques extrémales des lois marginales et de la dépendance. Elle permet de modéliser la variabilité du phénomène pendant des épisodes extrêmes, à une échelle spatiale et temporelle relativement fine, souvent en travaillant avec un continuum pour le support spatial. Selon les problématiques abordées, mes approches méthodologiques puisent dans l'une ou l'autre de ces champs disciplinaires théoriques avec des chevauchements occasionnels mais de plus en plus fréquents dans mes travaux de recherche les plus récents. Les prolongements envisagés dans mes recherches actuelles ont visée à rapprocher ces deux domaines pour faciliter une modélisation intégrée des risques spatio-temporels dans les disciplines scientifiques susmentionnées en alliant les échelles globale et locale grâce à une boîte à outils basée sur un langage commun et sur une plus forte intégration des méthodes de l'apprentissage artificiel.

En analyse des valeurs extrêmes, la partie théorique de mes travaux concerne la caractérisation de la dépendance extrémale. Les processus limites en théorie asymptotique, et un grand éventail d'autres modèles *sous-asymptotiques* plus souples, se construisent en multipliant un processus stochastique à structure relativement simple, comme par exemple un processus gaussien, avec une variable aléatoire d'échelle, et on peut alors parler d'une *décomposition profile-échelle* par rapport à ces deux composantes. Sur ce fond, mes développements méthodologiques proposent des approches d'estimation et de simulation pour des nouveaux modèles spatiaux et spatio-temporels, pour lesquels je cherche à caractériser des indices de résumé du comportement extrémal. Dans les approches de modélisation de la dépendance extrémale, une grande attention est prêtée à la distinction entre les deux régimes très différents de dépendance asymptotique et d'indépendance asymptotique. Avec le premier, caractéristique des modèles asymptotiques tels que les processus max-stables et les processus de Pareto généralisés, la force de dépendance est stable à travers toutes les magnitudes extrêmes, et par conséquent l'étendue spatiale et la durée temporelle des épisodes extrêmes ne dépendent pas de la magnitude d'un épisode extrême. Toutefois, ce cadre se révèle être trop restrictif pour la plupart des processus envi-

ronnementaux observés, dont les caractéristiques empiriques pointent vers le deuxième régime, l'indépendance asymptotique. La découverte empirique de ce comportement de dépendance est à l'origine des modèles “sous-asymptotiques", qui gardent une forte motivation asymptotique mais sont capables de prendre en compte de façon flexible des taux de décroissance relativement rapide dans les probabilités de dépassements conjoints de plusieurs composantes au dessus de seuils de plus en plus élevés. Les classes de nouveaux modèles sous-asymptotiques de dépendance extrémale issues de mes travaux se basent principalement sur des constructions faisant apparaître une décomposition profile-échelle. Des caractérisations théoriques très générales sont proposées pour décrire la structure de dépendance résultant de cette approche dans les lois bivariées. L'inférence statistique est mise au point pour des processus construits comme des mélanges d'échelle de champs gaussiens dans le cadre de la modélisation des dépassements de seuil spatialisés, et pour des processus de type max-infiniment divisible pour les observations de données de maxima ; ces nouveaux modèles ne manifestent pas la stabilité asymptotique des modèles max-stables classiques. Outre ces propositions pour une meilleure prise en compte de la dépendance extrémale, des nouveaux modèles de lois marginales ont été développés, d'une part en adoptant une optique sous-asymptotique pour construire des lois de probabilité proches des lois limites de Pareto généralisées pour les dépassements de seuils, mais plus souples, d'autre part en intégrant des covariables et des effets aléatoires dans des modèles de queue de distribution structurés en trois composantes : un modèle de régression pour la distribution complète des observations, permettant de définir un seuil nonstationnaire élevé, puis un modèle de régression logistique pour la probabilité de dépassement de ce seuil, et enfin un modèle de régression avec une loi de réponse de type Pareto généralisée pour les excès au-dessus de ce seuil.

La deuxième grande partie de mes travaux concerne la construction et l'estimation de nouveaux modélès bayésiens hiérarchiques, basés sur des processus gaussiens latents, pour les semis de points spatio-temporels de données d'occurrence de risques environnementaux, tels que les feux de forêts et les mouvements de terrain. Ces modèles de Cox log-gaussiens intégrènt l'influence linéaire ou nonlinéaire d'une multitude de variables explicatives, et des effets aléatoires spatiaux ou spatio-temporels sont déployés pour tenir compte de la variabilité d'intensité d'occurrence non expliquée par les variables observées seules. Un premier volet de ces travaux concerne les incendies de forêts en France, dont les positions et les temps des départ représentent un semis de points spatio-temporel ; la surface parcourue et brûlée par les feux peut être attachée à ces points comme une marque numérique. Les modèles bayésiens hiérarchiques développés pour ces données intègrent une multitude de covariables liées au couvert et à l'utlisation des sols ("Land Cover – Land Use"), aux infrastructures et aux conditions météorologiques. Ces covariables nécessitent souvent des pré-traitements importants pour obtenir une meilleure capacité prédictive dans le contexte de prédiction des incendies de forêt, et aussi pour ramener ces données de type multi-source et multi-échelle à une échelle spatiale et temporelle commune. Outre l'inférence sur ces facteurs de risque, qui sont susceptibles de manifester une influence fortement non linéaire sur le risque d'occurrence des feux de forêt, il faut tenir compte d'autres sources de variabilité spatio-temporelle méconnues en rajoutant des effets aléatoires spatio-temporels. Cette approche rend possible une cartographie spatio-temporelle réaliste de l'intensité d'occurrence des feux de forêt, et l'attribution de la variabilité spatio-temporelle observée aux facteurs de risque devient statistiquement fiable. Finalement, la surface brûlée est modélisée à l'aide d'une composante du modèle prêtant une grande attention aux plus grands feux de forêt ayant une influence dominante sur les surfaces brûlées cumulées en raison des

queues de distribution lourdes, typiques pour les surfaces brûlées. Un deuxième volet traite des mouvements de terrain en proposant des modélisations innovantes basées sur les processus ponctuels, en particulier sur les processus de Cox log-gaussiens. Les mouvements de terrain ont souvent lieu après l'occurrence d'un événement déclencheur, tel qu'un événement de précipitations intenses ou un séisme, dont la variation d'intensité spatiale et temporelle est observée seulement partiellement, ou même pas du tout dans la plupart des cas historiques. En identifiant un mouvement de terrain avec un point en espace planaire, typiquement défini comme la position ayant l'altitude la plus élevée dans la cicatrice infligée au paysage par le mouvement de terrain, la modélisation spatiale est abordée en incorporant diverses variables géomorphologiques dans le modèle, ainsi que des effets aléatoires pour représenter l'influence spatialement variable de l'événement déclencheur non observé. L'analyse des résultats à visée prédictive se concentre sur une bonne prise en compte de l'effet aléatoire, et sur l'influence nonlinéaire de variables auxiliaires, comme par exemple la pente du terrain, et de leurs interactions potentielles avec l'effet aléatoire. Un travail novateur étend ces approches à un cadre spatio-temporel pour analyser un inventaire multi-événementiel de plusieurs milliers de mouvements de terrain, observés dans une région d'Italie pendant une période d'observation de près d'un siècle. Cette modélisation a permis de réveler et interpréter des patrons d'occurrence spatiaux et temporels des mouvements de terrain.

Afin d'estimer les nouveaux modèles développés dans les approches décrites précédemment, je fais appel à des méthodes d'estimation fréquentistes et bayésiennes. Pour maîtriser le grand nombre d'observations dans ces applications spatio-temporelles, il est souvent nécessaire de développer et déployer des techniques adaptées à l'utilisation de modèles de covariance en grande dimension. En inférence bayésienne hiérarchique, mes contributions consistent en des extensions de l'approche récente de l'Integrated Nested Laplace Approximation (INLA), qui est souvent combinée à l'utilisation de modèles gaussiens markoviens spatiaux ou spatio-temporels, obtenus grâce à l'approche des équations aux dérivées partielles stochastiques ("SPDE approach"). Ce cadre de modélisation a généré une véritable révolution en analyse bayésienne spatiale en facilitant l'implémentation de modèles à processus gaussiens latents très sophistiqués. La complexité des composantes de ces modèles peut être contrôlée d'une façon très intuitive à l'aide des lois a priori de type "Penalized Complexity", que je déploie systématiquement. Mes travaux à fondement bayésien se focalisent sur des problématiques liées aux modèles de régression pour les dépassements de seuil, et aux processus de Cox log-gaussiens dans un contexte spatio-temporel. En revanche, l'estimation fréquentiste est préférée pour la plupart des modèles de dépendance extrémale, caractérisés par des structures non-gaussiennes, qui nécessitent souvent des approches d'intégration numérique pour gérer des variables latentes non gaussiennes ou pour tenir compte d'un mécanisme de censure des observations en dessous d'un seuil (c'est-à-dire en dehors de la queue de distribution). L'adaptation des méthodes de vraisemblance par paires représente souvent une solution robuste permettant le passage à l'échelle avec des grands jeux de données, tout en garantissant des propriétés des estimateurs semblables au maximum de vraisemblance classique, telles que la consistance et la normalité asymptotique.

Une priorité dans les prolongements méthodologiques que je prévois développer dans les prochaines années concerne une meilleure modélisation des risques multiples, surtout des risques extrêmes. Il s'agit de risques pour lesquels plusieurs aléas ou enjeux interagissent, ce qui nécessite de faire appel à une modélisation stochastique conjointe de plusieurs processus impliqués, souvent définis à différentes échelles spatiales et temporelles. Par exemple, les incendies de forêt se

produisent de préférence en situation de conjonction de l'accumulation d'un fort déficit hydrique avec des températures élevées, générant un environnement très sec en présence d'un combustible de végétation relativement sèche. La cascade des risques liés aux incendies de forêt peut encore être prolongée vers les glissements de terrain, la pollution de l'air, et des rétroactions avec le réchauffement climatique. Les processus faisant partie d'un tel système à composantes multiples sont le plus souvent encore modélisés de façon séparée, et souvent par des chercheurs travaillant dans des disciplines distinctes. Un autre aspect de modélisation étroitement lié à ce défi, auquel je souhaite m'attaquer, est une meilleure gestion de données volumineuses multi-source et multi-échelle dans les modèles et algorithmes d'estimation statistiques. Un autre objectif concerne une meilleure intégration entre les méthodes d'apprentissage artificiel, souvent à visée purement prédictive et non inférentielle, et les méthodes d'estimation de modèles stochastiques permettant une prédiction probabiliste et l'inférence des facteurs de risque. Dans les domaines d'application de mes approches, je souhaite apporter mon expertise pour la solution de problèmes de modélisation spatio-temporelle en (agro-)écologie et épidémologie. En particulier, je contribuerai à une meilleure exploitation des données opportunistes, pour lesquelles il faut corriger les biais d'observation en cas de protocoles d'observation peu documentés et d'observations entachées d'effets d'échantillonnage préférentiel.

List of notations and abbreviations

Abbreviations

- BHM – Bayesian hierarchical model(ing)
- EVT – extreme-value theory
- GAM – generalized additive model(ing)
- GEV – Generalized Extreme-Value (Distribution)
- GPD – Generalized Pareto Distribution
- i.i.d. – independent and identically distributed
- INLA – Integrated Nested Laplace Approximation
- LGCP – log-Gaussian Cox process
- LGM – latent Gaussian model
- MEVT – multivariate extreme-value theory
- POT – peaks-over-threshold
- PPP – Poisson point process
- SPDE – stochastic partial differential equation

Functions

- a_+ – $\max(a, 0)$ for a value $a \in \mathbb{R}^d$, using the componentwise maximum if $d > 1$
- $f(x) = o(g(x))$ – $f(x)/g(x) \to 0$ for functions f, g (with $g(x) \neq 0$ for $x > x_0$), as $x \to \infty$
- $f \sim g$ – $f(x)/g(x) \to 1$ for functions f, g (with $g(x) \neq 0$ as $x > x_0$), for $x \to \infty$
- $\mathbb{I}_A$ – the indicator function of the event A, *i.e.*, $\mathbb{I}_A(x) = 1$ if $x \in A$, and $\mathbb{I}_A(x) = 0$ otherwise

Random variables and probability distributions

- F_X^{-1} – (generalized) inverse of the distribution function F_X of a random vector X
- $F_X \star F_Y$ – convolution of $X \sim F_X$ and $Y \sim F_Y$, *i.e.*, the distribution of $X + Y$
- $X_1 \perp X_2$ – random variables/processes X_1 and X_2 are stochastically independent
- $\mathrm{P}(A)$ – the probability of event A

Specific probability distributions

- B$in(n, p)$ – binomial distribution with $n \in \mathbb{N}$ trials and success probability $p \in [0, 1]$
- GPD(ξ, σ) – generalized Pareto distribution with shape ξ and scale $\sigma > 0$
- Pareto(α, β) – Pareto distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$

Other notation

- $\boldsymbol{x}_{-i}$ – a vector $\boldsymbol{x}$ with its ith component removed
- A^{c} – the complement set of a set $A \subset \mathbb{R}^D$

Chapter 1

Introduction

1.1 Short biography

Born January 20th, 1983, in Nuremberg, Germany, I effected my undergraduate studies in applied mathematics ("Econo-Mathematics") at Ulm University, Germany. I decided to conclude my studies with a Master Thesis under the supervision of Volker Schmidt on the topic of 3D-modeling and inference with Gibbs point processes. While this study program puts relatively strong focus on applications in finance and actuarial sciences, I developed strong interest for a research career in environmental and life sciences, a fact that motivated me to enrol for the Master of Science in Biostatistics at Montpellier University, France. After successful graduation, I opted for a PhD project at IMAG, the mathematical institute of Montpellier University, with focus on new models and inference tools in spatial extreme-value analysis, under the supervision of Jean-Noël Bacro and Pierre Ribereau.

After defending my PhD in October 2013, I undertook the adventure of a post-doc project at LIRMM, the Computer Science institute of Montpellier University, in collaboration with Sandra Bringay and other members of her team. My work was part of a larger project geared towards the development of statistics and data mining techniques for extracting clinically relevant information from unstructured textual data retrieved from internet forums visited by breast cancer patients. During the 10 months spent on this new research field, I have become familiar with many data extraction and analysis techniques for high-dimensional data such as unstructured texts, and my work has led to a number of conference papers and journal publications. Some of these papers are co-authored with Mike Donald-Tzapi, a PhD student who took over the torch when I left the project for joining INRA in 2014. Note that INRA is now called INRAE after the 2020 merger with Irstea, another French public research institute with numerous sites implanted in the French territory.

Since September 2014, I am a research associate and part of the BioSP unit of INRAE, located at INRAE's Avignon site, and whose research focuses on various modeling approaches for spatially indexed processes. My recruitment aimed at maintaining the lab's strong expertise in modeling and inference approaches related to stochastic geometry and to modeling of complex data, and at reinforcing BioSP's links to academic research in France and worldwide. Using the solid knowledge that I had already acquired on statistical modeling at the intersection of geostatistics and extreme-value theory, I have since then substantially expanded my research

interests towards spatiotemporal modeling of various types of extreme and rare events – using a blend of frequentist and Bayesian inference techniques – to the service of fields such as climate science, environmental risk analysis and mapping, and agroecology. A summary of selected research results, and an outlook to future developments, is presented in this manuscript.

In 2020, I have won the Early Career Award of INRAE (Lauréat "Espoir Scientifique").

1.2 Challenges in modern spatiotemporal statistics

Environmental, climatological, ecological and epidemiological risks pose major challenges to societies and ecosystems, and these may be further exacerbated in the current era of global changes and ecological transitions. Events related to such risks show variation over space and dynamics through time. Authorities, scientists and other stakeholders are in need of tools to better understand and quantify relevant risk factors, to predict occurrences, and to provide a range of realistic future scenarios, with the aim to inform policy for reducing disaster risk. Such tools can be developed and trained using increasingly abundant and "big" datasets.

Events occurring close in space and/or time often tend to be similar, as Tobler (1970) has enacted into his first law of geography:

Everything is related to everything else, but near things are more related than distant things.

Therefore, the modeling of spatial-temporal stochastic co-occurrence patterns is crucial for understanding and predicting such processes, and for drawing sound statistical inferences. To succeed with such challenging data analyses, often combining data from multiple sources and observed at different spatial and temporal scales, the development of new methodology at the interface of the two following, up to now relatively disconnected research fields is a promising approach: *extreme-value theory*, which provides a framework for predicting probabilities of events with very extreme magnitudes, and *stochastic geometry*, useful for studying geometric patterns in occurrence locations/times. Constructing easily interpretable and deployable predictive models is challenging for high-dimensional data with many predictor variables, and it excludes the use of "black-box" Machine Learning approaches.

In contrast to purely spatial modeling, spatio-temporel modeling adds the difficulty of appropriately capturing long-term trends and short-term variability along the time arrow. Temporal dynamics and causal effects have a forward direction along the time axis, and the role of the time dimension is fundamentally different from that of planar or 3D geographical space. The study of consequences of global and local changes, such as climate change, changes in land cover and land use, or ecological invasions, necessitates designing models that are often required to be fundamentally nonstationary.

1.2.1 Practical goals of space-time modeling

A first typical goal is to design explanatory models that allow us to reveal and quantify the influence of observed predictor variables – called *covariates* in the following – on one or several target variables – called *responses* in the following. For the sake of simplicity, we use the term *covariates* for both numeric and categorical explanatory variables. This goal requires the development of regression approaches to formally decide if covariate effects are significant, and

to show how they contribute to variability of the response variable (linearly or nonlinearly). In particular, one must take into account the dependence of observations of the response over space and time to avoid wrong inferences based on the erroneous assumption of conditional independence of the response given the observed, deterministic covariates. More specifically, the detection and quantification of nonstationarities with respect to the time covariate may be the prime inferential goal (*e.g.*, in climate change studies). While nonstationarities are most commonly modeled in the marginal distributions, it may also be of interest do detect time-variant dependence structures, *e.g.*, for inference on the evolution of the size of *hot spots* of climatic variables.

A second goal, intertwined with the first, is to provide prediction of partially observed phenomena over space and time, and in particular forecasting, that is, prediction forward in time for the near future. For the communication of results to scientists and other stakeholders, we have to provide decision tools, for which spatiotemporal mapping of predictions and uncertainties is particularly important.

1.2.2 Spatiotemporal statistical modeling in the era of big data

The fast development and deployment of remote or embedded sensing technologies (satellites, drones, GPS-based mobile sensors, Lidar units) has led to an explosion of spatiotemporal datasets in recent years. For the task of *turning raw data into understanding*, as formulated by Wickham and Grolemund (2017), statistical learning techniques have become indispensible to extract knowledge from a heap of data. Major research efforts have gone into spatial and spatiotemporal statistics leading to appropriate extensions of available models, inference techniques and mapping tools to cope with very large samples, *i.e.*, to have *scalable* methods. In contrast to the panoply of methods developed in the fields of artificial intelligence and Machine Learning, the focus of the spatial statistics community remains on sound inferences, often model-based, such that uncertainties stemming from the data-generating process and from the estimation approach can be properly identified and assessed (Dunson, 2018).

In this age marked by global changes and ecological transitions, we often seek to better understand what has physically existed at a given time in the past and may continue to exist (*e.g.*, geological features, past weather and climate), but we also aim to predict or to scenarize what will happen in the nearest future (*forecasting*), or at medium-to-long-term temporal horizons, by using *probabilistic predictions*. The development of generative stochastic models, for instance stochastic generators of extreme weather events, is necessary to produce and communicate such scenarios, to explain how likely they are, and to feed impact models with realistic scenarios of external forcings.

A recurring theme in my research is to jointly characterize spatial-temporal trends and dependencies for complex phenomena. Achieving this goal often requires combining different data sources with data arising at various spatial and temporal scales.

Large multi-source and multi-scale space-time datasets

The surge of new tools and techniques for data acquisition and storage in recent decades has led to an exponential growth in available data with a large diversity of georeferenced information, data quality and accessibility.

As of today, my work focuses strongly on studying the dynamics in environmental and climatological risks and in ecological processes, which are often strongly driven by a combination of properties related to land cover/land use (available for various grid sizes based on sampling campaigns at irregular time steps) and climate and weather (available as weather station data, gridded reanalysis data, for various periods and spatial resolutions). Models and estimation methods must be flexible enough to cope with such heterogeneous, multi-scale data. Moreover, we often aim for the detection of multiple relevant spatial and temporal scales in the response variable(s). The scales that are important for the understanding and prediction of the physical processes may be different from the scales at which data have been observed and collected. This situation often gives rise to an intricate structure of combinations of observed and latent scales of space and time between predictor variables and the response. An important goal is to reveal the relevant latent (*i.e.*, unobserved) scales of processes through appropriate inference approaches, often involving hierarchical formulations of statistical models.

Data quality is another potential issue. The sampling design may not follow a strict protocol to ensure that it is uncorrelated with the measured variables, such that models have to be able to take into account varying observation efforts and preferential sampling, for instance with respect to spatial location or extreme values. Such issues are exacerbated in many citizen science datasets, whose statistical analysis has found of major interest in recent years in disciplines such as ecology.

Manipulating high-dimensional covariance matrices

For appropriate representations of spatiotemporal correlation, we are bound to manipulate large covariance matrices and infer their structure from large datasets. This leads to the *big n problem* where the standard algorithms for covariance matrices (solving linear systems, inversion, calculation of the determinant) of size $n \times n$ possess complexity involving $\mathcal{O}(n^3)$ basic arithmetic operations, if a generic covariance structure is assumed. Therefore, the classical geostatistical setting, where common covariance functions (*e.g.*, Matérn, powered-exponential) and related covariance matrices are manipulated directly, is often too unwieldy in practice, and usually comes with prohitively high computational cost for full likelihood inference. However, a relatively recent shift towards direct manipulation of sparse approximations to inverse covariance matrices (*i.e.*, of Gauss–Markov processes, see Rue and Held, 2005) provides an elegant solution thanks to theoretical links between Gaussian processes and stochastic partial differential equations (SPDEs) exploited by the seminal work of Lindgren et al. (2011).

Disentangling nonstationarities in marginal distributions and dependence

The correct specification of marginal distributions and dependence structures, and their interplay in the case of nonstationary processes, is a challenging task in practice. While marginal distributions are intended to capture long-term behavior and variations therein, dependence structures should characterize the short-range and short-term stochastic variability of the process in space and time. In cases where observations tend to have similar values (*e.g.*, in the case of a variable observed at different locations at a given day of the year), this may be due to a marginal effect (*e.g.*, the mean parameter of the marginal distributions is very high for all locations during the current season), or due to a dependence effect (*e.g.*, observations at

the observed locations are always similar for any given day). When distances between the locations are very small, one may hypothesize a pure dependence effect, while for locations that are mutually separated very far one may rather hypothesize a marginal effect. For illustration, consider the case of climatic variables: marginal distributions at a given point in space and time are related to the notion of *climate*, while variability arising locally in space and time (*i.e.*, space-time dependence) is related to the notion of *weather*.

Appropriate identifcation of marginal and dependence effects in the presence of nonstationarities may be facilitated by the availability of space-time datasets with large temporal depth. Nevertheless, even in this setting the correct attribution of variability in the process arising at intermediate spatial and temporal scales to either trends in marginal distributions or long-range dependence patterns may be intricate, especially when we concentrate on rare events.

1.2.3 Modeling rare events with high impact

In practice, primary interest of statistical modeling should be geared towards those realizations of physical processes that actually "matter", such as those that represent important risks and are typically associated with relatively extreme values of some observed variables. Often, the observations of such high-impact events represent only a small fraction of the data sample. Therefore, classical statistical theory where key tools are based on central limit theorems characterizing "average" behavior of phenomena is not appropriate. Especially purely Gaussian models do not provide an appropriate conceptual framework and tend to show bad performance in many cases. Depending on the nature of the phenomenon and how it was observed, two alternative frameworks come to the rescue: extreme-value theory (Coles, 2001; Beirlant et al., 2004) in the case of continuous variables; and point process theory, or stochastic geometry (Chiu et al., 2013) in the broader sense, in the case of discretely observed rare events.

Extreme values

The statistical estimation of a model using all observations from a data sample supplies little control over the goodness-of-fit in the extreme regions of the distribution, because the latter contribute only a small fraction to the sample. If we use standard statistical distributions and methods, then the influence of the extreme observations is rather small on the estimation of model parameters. Certain distributions may be endowed with shape parameters that provide some flexibility in the tails, but usually there are no parameters to allow for precise and separate control of behavior in tail regions. Fortunately, extreme-value theory provides limit theory and statistical tools tailored to the analysis of extreme events. Two standard approaches – *block maxima* and *peaks-over-threshold* – have been established for inferring extremal behavior. Without loss of generality, we assume that the extreme events of interest are located in the upper tail of the distribution. Lower-tail extremes may also be of interest (*e.g.*, small temperatures for analyzing cold spells), and it would be straightforward to adapt statements and tools for the upper-tail by switching tails, *e.g.*, by replacing a random variable X with its negation $-X$. In the *block maxima approach*, we consider the sample of maxima over blocks of data of the same size (*e.g.*, yearly maxima for each weather station in the case of climatological extremes). In the *peaks-over-threshold approach* (POT), we fix a high threshold and then study the exceedance probability and the size of the excess above the threshold. Due to its easier interpretation with

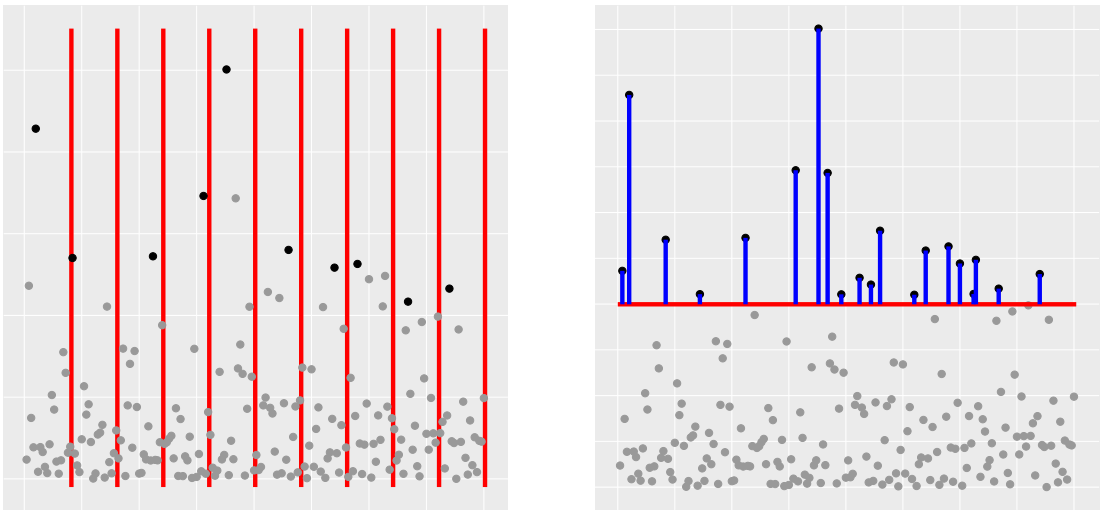


Figure 1.1: Illustration of two important modeling approaches in univariate EVT. Left display: the block maximum approach, with black dots showing maxima over blocks delimited by vertical red segments. Right display: the peaks-over-threshold approach, with blue segments showing positive excesses above the threshold (horizontal red line).

respect to the original events and its higher flexibility through the choice of the threshold level (which may be nonstationary), recent approaches often resort to the POT approach, and the majority of approaches presented in my work follow this paradigm. The two approaches are illustrated in Figure 1.1.

In POT-based inference, one usually has to decide about the approach used to handle the values in the bulk of the distribution, *i.e.*, the values below the threshold. One possibility is to fully remove bulk data from the sample used for estimating the exceedance model. Then, it is useful to provide an estimation for the mixture probability between the bulk component (for which we may not specify a model) and the exceedance component. Another possibility is to specify a model that applies to the full distribution (*i.e.*, the bulk region and the exceedances) in principle, but to *censor* non-exceeding observations during the inference procedure such that they only contribute the binary information about exceedance of the threshold but not the actual value below the threshold. This approach to calibrating models to data ensures that the fit is good for the region of threshold exceedances, while less importance is attached to the fit for values below the threshold. Such models may be preferred if it is desired that realizations of the model include bulk values with smooth behavior around the threshold. In practice, another benefit is that such models can be expected to provide a useful representation of quantiles below the threshold but close to it.

In situations where extremes may occur in several components of a system (*e.g.*, for different variables, or for the same variable observed at different locations in space), it is important to appropriately take into account the co-occurrence patterns of extremes for providing inferences and predictions about joint occurrences of extreme values. In multivariate and spatial POT-based modeling, the task of choosing a multivariate or spatial threshold is usually difficult. There is no unique way of ordering observation vectors according to their magnitude, such that an appropriate criterion has to be defined *ad hoc* (*e.g.*, exceedances of the maximum, the

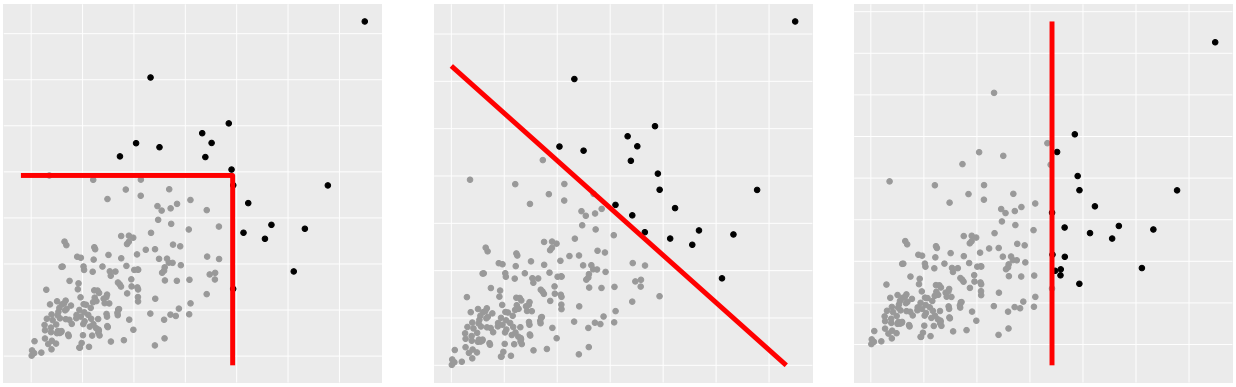


Figure 1.2: Illustration of different types of threshold exceedances in the bivariate setting. The threshold is shown in red. From left to right: exceedances in at least one of the two components; exceedances of the sum of the two components; exceedances in a fixed component.

minimum or the sum of the vectors' components). Examples of different types of bivariate threshold exceedances are illustrated in Figure 1.2. The threshold criterion may depend on the practical aim of the analysis but also on considerations regarding the tractability of estimation and prediction methods. It is a major concern of my work to develop inference procedures for extreme events characterized by the threshold exceedance of some practically useful aggregated value of multiple variables, *i.e.*, characterized by an appropriate exceedance region representing a relatively small subset of the full support of the stochastic process of interest.

Point patterns

In many applications, primary interest lies in modeling the occurrence intensity of the events, *i.e.*, the focus is on counting the number of times that an event of interest happens in a given area and a given period of time. Examples of such data are the positions in space and time where we observe the ignition of a wildfire, the occurrence of a landslide, or the presence of a plant or animal species. Such data then correspond to *point patterns*, for which stochastic point process models can be used.

With point patterns, additional categorical or numerical information may be available for each of the observed events, and such information can be represented mathematically through marks of the points of the point pattern, leading to the notion of a *marked point process*. For example, the burnt area of a wildfire is a numerical mark associated to the point representing its ignition location and time.

It is worth pointing out that there exists a natural and well-studied link between the theory of extreme values and point patterns (*e.g.*, Resnick, 1987). This link originates from considering the pattern of points where a variable exceeds a high threshold; *i.e.*, the location and time of a point identifies a local "hotspot" of the process under study. This approach allows adopting a relatively macroscopic view on extreme event episodes in a continuous spatiotemporal stochastic process, where information about such episodes is only considered through their occurrence location and time.

Particularities in modeling processes of rare events

The use of probabilistic frameworks for modeling and prediction allows taking into account stochastic behavior of processes, and estimation uncertainty. Specifically, the rare character of extreme events calls for appropriate theoretical and statistical tools to assess uncertainties. In general, rareness may refer to rarely observed magnitudes (*e.g.*, windstorms, heat waves, extreme precipitation), or to occurrence numbers that are small when we compare them to to all the possible configurations of space, time and covariates where no event occurred (*e.g.*, wildfires, landslides, temperature hotspots). In particular, the use of the Bayesian paradigm is beneficial for accurate assessment of uncertainties through exploration of posterior distributions, and this framework further makes it possible to incorporate prior information, such as expert knowledge, into the modeling approach to obtain more reliable and stable results.

Even if Gaussian processes are widely used in spatiotemporal analysis thanks to their theoretical foundation in central limit theory and advantageous numerical properties, they cannot be directly used for rare event data whose characteristics require the use of more appropriate non Gaussian models, *e.g.*, of max-stable models, or of generalized Pareto processes for modeling extreme episodes of a spatiotemporal process. While estimation of Gaussian dependence models can be achieved through numerically convenient and well-mastered techniques of weighted-least squares, the construction and deployment of appropriate non Gaussian models typically requires more sophisticated estimation and optimization routines, and their numerical complexity may not scale well with the size of datasets.

1.2.4 Fusing Bayesian and frequentist inference

In the history of statistics and today's statistical practice, a distinction between two major inference frameworks has been carved out: *frequentist* and *Bayesian* statistics. Frequentist statistical approaches put focus on the proportions or frequencies of occurrence of observation values in the sample of observations. Estimated parameters and models should attribute high probabilities to the observed sample. Parameters are usually considered as fixed but unknown values, and hypothesis to be tested are considered as being fundamentally true or false. In contrast, the Bayesian view of inference considers parameters as random variables, and focuses on assigning probabilities to hypotheses. Bayesian modeling aims to treat all sources of uncertainty as unknown quantities through the use of random variables. A prior probability is defined for hypotheses to be investigated, and the prior probability is then updated in light of available data to obtain posterior probabilities, which are then used to draw statistical inferences.

While rather strong opinions may be carried by communities promoting the advantages of one view over the other (see, *e.g.*, the content and discussion of the paper by Gelman, 2008), I promote the idea of a pragmatic consideration of the benefits and drawbacks of the two approaches depending on the problem to be solved, which leads to a blend of frequentist and Bayesian inference techniques with the aim to bring together the best of both worlds.

In the frequentist approach, it may be easier to deploy estimation techniques that scale well with increasingly large sample sizes, *i.e.*, that are suitable for very high-dimensional problems with millions of observations. Parameter inference using composite likelihoods (Lindsay, 1988; Varin et al., 2011) is an important example where the use of prior distributions in the Bayesian sense may be awkward in practice since direct interpretation of results is difficult, and adjust-

ments to obtain results allowing for usual interpretation based on classical full likelihoods are not possible or unwieldy in practice.

Bayesian inference is particularly relevant when prediction uncertainties are high. Due to generally small occurrence numbers of rare events, the assessment of natural and statistical uncertainties through appropriate stochastic modeling approaches is crucial, and the propagation of such uncertainties across different model components must be handled with care. Expert knowledge can be elicited for the construction of prior distributions in the often data-poor situation of rare and extreme events. State-of-the-art Bayesian implementations enable statistical inference for very complex predictive models with many unknown parameters (up to tens of thousands). As such, Bayesian techniques are particularly useful for latent variable models, *e.g.*, latent Gauss–Markov models. Often, the size of the latent random parameter vector is relatively high but different from the even larger size of the observation vector, such that dimension reduction can be achieved in modeling large datasets.

1.3 Main tools in my approaches

This section shortly summarizes the main theoretical and inferential tools in my work.

As a side note, in order to disseminate fundamentals of spatiotemporal analysis, I have strongly contributed to a workshop tutorial and a publication (RESSTE Network, 2017) with colleagues from INRAE's RESSTE research network. It covers the full workflow from data import to model validation and geostatistical spatiotemporal prediction (*kriging*) in the setting of frequentist estimation of stationary Gaussian or log-Gaussian models.

1.3.1 Extreme-value theory

Gaussian models are ubiquitous in classical statistics thanks to the central limit theorems for samples of random variables and stochastic processes, but they are not well adapted to modeling rare and extreme events. In contrast, extreme-value theory (de Haan and Ferreira, 2007) provides a set of asymptotic results (*e.g.*, maximum domain of attraction conditions), limit models (*e.g.*, max-stable processes, generalized Pareto processes), and summary statistics (*e.g.*, tail correlation functions, extremal coefficients) for continuous-valued processes. These tools are useful for the analysis of extreme events of stochastic processes for which observations over a sufficiently long period of time are available, such that a large number of independent observations, or near-independent observations owing to mixing properties, can be used to statistically describe the behavior of extreme quantiles.

1.3.2 Point processes and stochastic geometry

If the focus is on counting events and studying geometric patterns, with data available as georeferenced observations of discrete variables and structures, then stochastic geometry provides appropriate exploratory tools (*e.g.*, Ripley's K-function, g-function) and model classes (*e.g.*, Poisson processes, Cox processes, Gibbs processes); see Chiu et al. (2013). For the sake of spatial and spatiotemporal modeling of occurrences of risks, the class of log-Gaussian Cox processes (LGCP Møller et al., 1998) allows for flexible inclusion of covariate information and

random effects in the intensity function of the point process, and inference tools for latent Gaussian modeling such as the integrated nested Laplace approximation (INLA, Rue et al., 2009) can be deployed. In applications, high-risk situations can be attributed to specific risk factors, geographic regions or time periods, and mapping of point process intensities and uncertainties can be provided as a decision support for stakeholders.

1.3.3 Generalized additive modeling

Generalized additive models (Hastie and Tibshirani, 1990; Fahrmeir and Tutz, 2013) are flexible regressions models where response variables X_i, $i = 1, \ldots, n$ are linked to a set of predictors $\boldsymbol{z}_i = (z_{i,1}, \ldots, z_{i,m})$ through a strictly monotonically increasing link function g according to the following equation:

$$g(\mathbb{E}X_i) = f_1(\boldsymbol{z}_i) + \ldots f_K(\boldsymbol{z}_i), \quad i = 1, \ldots, n,$$

with functions f_k that may depend on a single predictor $z_{i,j}$, or on two or more of the predictors in the case where f_k is an interaction effect. Usually a parametric or semi-parametric specification is given for the functions f_k, $k = 1, \ldots, K$, and parameters have to be estimated. Trends in marginal distributions of observations along the time axis, and in fact all such relatively smooth trends due to time, season or other covariates, can be captured in a nonlinear way by using GAMs, which usually scale well for small to moderately large datasets with hundreds to millions of observations.

In frequentist inference, appropriate smoothness of the estimated effects is achieved through smoothness penalties on the shape of the functions f_k, $k = 1, \ldots, K$. Optimal smoothness parameters can be determined through cross-validation approaches. In the Bayesian framework, a prior distribution can be specified for the parameters defining the effects f_k, and then the component f_k can be interpreted as a random effect. Smoothness of estimated effects is achieved through appropriate specification of multivariate prior distributions for the parameters involved in the construction of f_k. A benefit of the Bayesian approach is that smoothness parameters (*i.e.*, hyperparameters of the model) can be estimated simultaneously with the other parameters, and information about posterior uncertainty in the smoothness is available.

1.3.4 Composite likelihood estimation

When strong model misspecification can be excluded, classical full likelihood inference can be considered as the commendable standard approach when the likelihood function is numerically tractable. In cases with many observation locations and times, full likelihoods may be numerically intractable even if they can be written in closed analytical form. Possible causes are matrix operations that become intractable, or a combinatorial explosion of the number of terms to be computed for the likelihood. The latter problem frequently arises with models for extreme-value analysis, even for low-dimensional settings with more than two variables. To remedy such issues, composite likelihood estimation (Lindsay, 1988; Varin et al., 2011) has become a useful alternative ensuring consistent estimation of parameters with asymptotic normality of estimators. Its idea consists in constructing a likelihood from blocks of data while assuming independence between blocks. The most popular choice is the pairwise likelihood, where blocks consist of two observations each. If the parameters to be estimated can be identified from the information

available in the chosen blocks, then composite likelihood estimation usually comes at a very moderate loss of efficiency as compared to the full likelihood.

Often, especially in the spatial and spatiotemporal setting, estimation is carried out in a two-step approach where one estimates marginal distributions in the first step, and then the dependence structure in the second step (Joe, 2014). Marginal parameters estimated in the first step using an independence likelihood (*i.e.*, composite likelihood of univariate distributions), or alternatively empirical marginal distributions, may have to be "plugged in" during the second step. This approach has the benefit of estimating a smaller number of parameters during each step as compared to joint estimation of all parameters. Moreover, it is possible to use different estimation approaches in the two steps. Asymptotic theory still works with plug-in approaches thanks to variants of the continuous mapping theorem and results for rank-based (*i.e.*, nonparametric and distribution-free) transformations (Joe, 2014).

1.3.5 Hierarchical modeling and INLA

Complex observation processes (*e.g.*, multi-source, multi-scale, multi-type, preferential sampling) often require a hierarchy of model layers to appropriately include and link the different inputs. Inference is usually sought for unobserved processes, whose structure has to be appropriately specified at latent layers. Therefore, I adopt the paradigm of Bayesian hierarchical modeling (BHM) in many approaches. With the lower-case notations typically applied in this setting for the vectors of data ($\boldsymbol{y}$), latent variables ($\boldsymbol{x}$) and hyperparameters ($\boldsymbol{\theta}$), the three layer-structure of hierarchical models can be summarized as follows:

Stage I:	**data**	$\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}$	–	likelihood of observations,
Stage II:	**process**	$\boldsymbol{x} \mid \boldsymbol{\theta}$	–	latent stochastic components
Stage III:	**hyperparameters**	$\boldsymbol{\theta}$	–	hyperpriors.

The latent process of Stage II encapsulates parameters of the likelihood of the observed data in Stage I, while hyperparameters in Stage III consist of parameters governing the prior distributions of process parameters, and global parameters of the likelihood. In this hierarchy, Stage I could also be considered as the *observation process*, while Stage II corresponds to a relatively smooth process that is not directly observed and typically represents an important component of the actual physical process that we want to estimate and predict. Often, the observations are considered as conditionally independent given the latent process and the hyperparameters. In particular, this assumption may considerably simplify and speed up estimation algorithms.

For the purpose of confronting the prior model with observed data to obtain posterior inferences and predictions, the integrated nested Laplace approximation (INLA Rue et al., 2009; Lindgren and Rue, 2015; Rue et al., 2017; Opitz, 2017) provides relatively fast and accurate analytical approximations in Bayesian hierarchical models with latent Gaussian processes, and typically requires only a fraction of the time of simulation-based Markov Chain Monte Carlo (MCMC) estimation. INLA sidesteps issues with the construction of appropriate proposal distributions and slow mixing of Markov chains in simulation-based techniques, while providing highly accurate results. The method astutely combines Laplace approximations (Tierney and Kadane, 1986), used to compute expectations with respect to high-dimensional multivariate Gauss–Markov random vectors with up to hundreds of thousand of components, with efficient

numerical integration schemes for integration with respect to a relatively small number of hyper-parameters (*e.g.*, controling the variance and dependence of Gaussian components). Therefore, it bypasses issues that may arise with simulation-based MCMC inference, where the design of stable algorithms for fast exploration of the posterior distribution may be hampered by intricate dependencies between the components of the model (*e.g.*, Rue and Held, 2005). INLA is implemented in the `INLA` package (Lindgren and Rue, 2015) of the `R` statistical software, and over the last decade it has been widely adopted for Bayesian additive regression modeling of spatial and spatiotemporal data (Opitz, 2017; Krainski et al., 2018). The Bayesian framework of INLA allows for joint estimation and uncertainty assessment of latent components, hyper-parameters and predictions. Recently, the speed and stability of INLA with high-dimensional latent Gaussian structures were further leveraged through its integration with the sparse matrix computation library PARDISO (van Niekerk et al., 2019). Technical details on how INLA works are presented in the Appendix §A.1.

The choice of prior distributions is a crucial step in Bayesian modeling, even more if only a small sample of rare events is available. While a number of general classes of prior distributions based on theoretical concepts of informativeness have been proposed in the Bayesian paradigm (*e.g.*, objective priors, Jeffrey's priors), it is notoriously difficult to specify appropriately informative and easily interpretable prior distributions in the high-dimensional setting of space-time modeling, where models are often constructed by combining predictor components of various types. In my work, I often make use of the recent approach of *Penalized Complexity priors* (*PC priors*, Simpson et al., 2017) where the distance of the prior of a model (or of a model component) with respect to a simpler baseline specification (the *reference*) is penalized at a constant rate. This mechanism is implemented through an exponential prior on this distance, such that the specific parametrization of the model (*e.g.*, the standard deviation, the variance, or the precision, which are equivalent up to strictly monotonic transformations) is without importance for the resulting prior, and the modeler has to choose only the penalty rate. These priors take the geometry induced by the choice of model parametrization into account, therefore avoiding intricate interpretation problems that may arise otherwise. They are designed to allow for shrinkage towards a simpler reference model in virtue of the principle of parsimony. They also provide an objective (*i.e.*, automatic) way of choosing the prior distribution family, while keeping some degree of subjectivity in selecting the penalization rate parameter. An example of a baseline for the variance of a Gaussian random variable is the value 0, *i.e.*, the absence of any stochasticity. Even better, the penalty rate can usually be set indirectly by prior specifications based on fixing a reference value u of the parameter of interest and the probability $\mathrm{P}(\text{parameter} > u)$, with "$>$" replaced by "$<$" in some cases, depending on the role of the parameter, such as $\mathrm{P}(\text{standard deviation} > 1) = 0.5$.

1.3.6 Gauss–Markov structures and the SPDE approach

The characterization of spatiotemporal stochastic variability usually involves the specification of a spatiotemporal covariance function and the manipulation of high-dimensional variance-covariance matrices, whose number of rows/columns typically corresponds to the number of observations and can reach several millions. While it is possible to break down such high-dimensional matrices into smaller pieces during estimation (*e.g.*, by using composite likelihoods) or prediction (*e.g.*, by using local kriging environments), it is often desirable to work with the

full matrix for optimal inference based on the full likelihood, especially in Bayesian hierarchical modeling (BHM). BHM provides a remedy by resorting to space-time Gaussian processes at the process layer, recall §1.3.5, where the dimension of the latent process can be chosen to be substantially lower than the number of observations. Moreover, such latent processes can be specified through their *precision matrix* (*i.e.*, the inverse variance-covariance matrix), which provides the benefit of working with sparse matrices (*i.e.*, matrices with mostly 0 entries) when using Gauss–Markov dependence structures (Rue and Held, 2005). The Stochastic Partial Differential Equation (SPDE) approach of Lindgren et al. (2011) has been a huge step forward for high-dimensional geostatistical modeling thanks to providing Gauss–Markov representations with closed-form precision matrices that accurately approximate the flexible and widely used Matérn covariance function. This fundamental result provides spatial Gauss–Markov structures, and these can be readily extended to include temporal dependence without losing sparsity of the precision matrix by specifying temporal autoregressive structures (Cameletti et al., 2013; Opitz, 2017). Technical background on the SPDE approach is given in §A.2.1.

1.4 Summary of main results

This section gives an elementary summary of main results in my work from the perspective of their impact on practical spatial-temporal modeling and risk assessment of rare events. Detailed technical background will be exposed in the following chapters.

1.4.1 General methodological developments and theory

Bayesian spatial-temporal inference and mapping of risks

An important part of my applied work aims to successfully adapt the INLA method to cope with moderately many predictor variables (say, up to 100, *e.g.*, in Lombardo et al., 2018; Opitz et al., 2020b; Pimont et al., 2021; Opitz et al., 2020a), often with nonlinear influence on response variables, and to capture temporal dynamics through models that have been purely spatial before in the existing literature (*e.g.*, for landslides in Lombardo et al., 2020). The development of INLA-based methodology for spatiotemporal modeling of the occurrence intensity of environmental risks in the point process setting to predict the expected number of events is central in my work, with important applications to wildfires and landslides. High-dimensional models and optimization problems arise in this setting since one has to include a very large number of control locations without observed events into the data vector. Typically, some of the relevant predictor components have not been measured directly (*e.g.*, the precipitation intensity triggering landslides, or human activity at the origin of wildfires) and must be represented in the model through space-time-indexed parameter surfaces (random effects) to be estimated. Standard algorithms often turn out to be numerically prohibitive for estimation in this setting, and numerical tricks are necessary to propose techniques for data dimension reduction (*e.g.*, subsampling) while keeping the loss of data information small.

To cope with the large variety of possible models that become available by combining several stochastic components into sophisticated models, their complexity is controlled through appropriately chosen penalized complexity priors, and model selection is performed by comparing models based on multiple criteria, including predictive checks. A detailed discussion is provided

in Opitz et al. (2020a) for a spatial application to landslide modeling, and a short exposition will also be given in Chapter §7. Model selection is often achieved through stratified cross-validation schemes; that is, in an objective, data-driven way by repeatedly removing part of the observations during the estimation step to numerically assess the forecasts through prediction scores that quantify how close a model can predict unknown quantities of interest, and in particular those values that have not (yet) been observed.

From theory to practice of flexible dependence modeling for space-time extremes

A collection of my results on the theory and implementation of realistic extreme-value models, with the most foundational output in Huser et al. (2017); Engelke et al. (2019); Mhalla et al. (2019); Bacro et al. (2019); Huser et al. (2021), allows addressing the question of how the spatial extent and the duration of extreme event episodes depend on the event magnitude. Spatiotemporal persistence of such events is a major factor for economic, sanitary and ecological catastrophes (*e.g.*, heatwaves, cold spells, air pollution episodes, floods, droughts).

If we suppose that the spatial extent of increasingly extreme events remains comparable in size, then we observe a certain stability of co-occurrence patterns of high values when moving towards higher and higher event magnitudes. This case is known as *asymptotic dependence* and corresponds to the classical limit models obtained by multivariate and spatial extreme-value theory. This situation implies that very high aggregated observation values may arise in case of extreme episodes, for instance very extreme cumulated precipitation amounts over a catchment when assessing flood risk. However, empirical findings and the physics of meteorological and environmental processes often contradict such behavior.

The lack of appropriate statistical extreme-value tools for handling decreasing dependence strength with increasing event magnitudes in spatial processes has been a strong motivation to construct novel theoretically motivated but practically more flexible models. These models accommodate *asymptotic independence* with decreasing spatial-temporal scales of extreme clusters for increasingly high magnitudes. At first glance, the ubiquitous Gaussian processes – the "Swiss knife" of spatial statistics – are promising candidates in this context thanks to their asymptotic independence, but they usually lack flexibility for realistic modeling when the application focus is on extreme events, *i.e.*, on the tail of probability distributions. I have therefore strongly contributed to promote a more general modeling framework using the notions of *scale* and *profile* with intuitive interpretation. Starting from a relatively simple and easily tractable baseline process (the profile), we modify it through a random variable (the scale) to jointly drive the extremes of all events, which may occur in several variables, over space and/or time. This yields high flexibility in comparison to the baseline model. The scaling variable acts as a common "shock" impacting all components, and it naturally encodes the strength of concurrence of extreme events in several components or locations/times. The use of Gaussian baseline models allows us to assess how strongly the extremal behavior deviates from Gaussian-based approaches used in "classical" geostatistics to capture the behavior of "ordinary" events.

The applications of such novel models confirm that atmospheric variables (*e.g.*, wind speeds, precipitation, air pollution) are most often asymptotically independent. The illustrating Figure 1.3 uses simulated data and shows cases where the same extreme quantile is observed in the central pixel of the three images, the same Gaussian correlation function is used, but different random scale variables are chosen, such that the peak value and the spatial extent of extreme

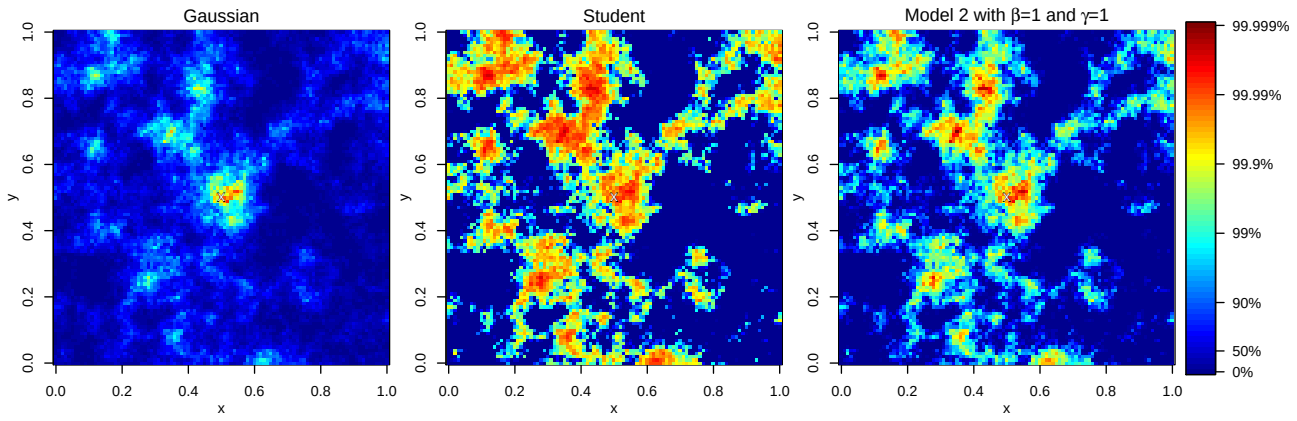


Figure 1.3: Illustration of various strengths of spatial extremal dependence based on the same correlation model for the Gaussian profile process, the same conditioning quantile in the central pixel of the space, but different random scale variables.

clusters arising farther away from the central pixel are fundamentally different between the three models.

Moreover, asymptotic theory can be combined with flexible Bayesian modeling as mentioned in the previous §1.4.1, especially through the use of INLA. This allows for a realistic representation of the variability and dynamics of extreme events through space and time by fitting models with many parameters to observed extreme episodes of variables such as precipitation or temperatures. Specifically, the paper of Opitz et al. (2018) extends INLA and related frameworks to estimate high quantiles and co-occurrence probabilities varying nonlinearly with space, time and other predictor variables. By implementing extreme-value models for threshold exceedances with INLA, reliable estimation and uncertainty assessment for extreme event probabilities become possible while making allowance for the small sample size in the extreme-value context. Moreover, our recent preprint Simpson et al. (2020) applies INLA-based modeling of space-time extremal dependencies to the flexible conditional extremes framework.

I promote hierarchical modeling of extremes since it allows using asymptotically justified probability distributions for extreme values, and we can embed latent (*i.e.*, unobserved) processes capturing the influence of external predictors, space and time on the occurrence probabilities and magnitudes. Estimating relatively smooth latent components and their interaction allows understanding the genesis of extreme states of the process under study, and then predicting them.

1.4.2 Spotlight applications

Spatial and spatiotemporel modeling of landslide occurrences

The concept of occurrence intensity, until then unknown in the geomorphological community, has been popularized for landslides in several of my papers, whose purpose is to propose INLA-based mapping of landslide intensities and uncertainties using point process models (Lombardo et al., 2018, 2019, 2020; Opitz et al., 2020a). Estimated models can be used to generate maps highlighting the spatial or spatiotemporal distribution of risk components for different spatial

scales. Using a century-spanning landslide inventory for a region in Italy, we have invented models that allow modeling and understanding long-term temporal dynamics in spatial landslide distributions (Lombardo et al., 2020).

Spatiotemporal modeling of wildfire activity in France

Models similar to those for landslides, but with larger event numbers and stronger focus on spatiotemporal modeling, are built for wildfire occurrences in Mediterranean France in the papers Gabriel et al. (2017); Fargeon et al. (2018); Opitz et al. (2020b); Pimont et al. (2021). The models, constructed at daily resolution and incorporating weather conditions and land-use land-cover covariates, allow for identifying the influence of various risk factors and for mapping fire occurrence risk. Moreover, by combining a model component for occurrences (ignition locations and times) with a model component for fire sizes, it is possible to reliably predict wildfire counts and aggregated burnt areas at various spatial and temporal scales (Pimont et al., 2021). State-of-the-art approaches to modeling wildfire risk strongly rely on the commonly used Fire Weather Index (FWI, van Wagner, 1977), but our models indicate a transfer function from FWI to wildfire activity that is strongly nonlinear for the study region. This finding highlights weaknesses of this index for the study region, and it points out the need to develop more appropriate fire danger indices.

Space-time dynamics of extreme Cévenol precipitation episodes

Extreme precipitation events in the French Mediterranean area, known as Cévenol or Mediterranean episodes, occur in autumn when relatively hot and humid air enters onshore from the Mediterranean and hits mountain ranges that have already cooled down. Important flooding events with many casualties have been the consequence. By building on new extreme-value models for space-time episodes in Bacro et al. (2019), we develop an innovative, hierarchically structured model for asymptotically independent threshold exceedances. It enables physical interpretation with respect to the spatial extent, temporal duration and velocity of precipitation episodes by using a construction based on geometric objects moving through space and time. The shape of such slanted space-time cylinders is then estimated to determine interacting spatiotemporal scales. Precisely, we apply spatiotemporal kernels to a gamma process to generate the latent layer of the model that jointly controls exceedance probabilities and excesses, the latter following a generalized Pareto distribution. The implementation of fast frequentist estimation techniques (composite likelihood) has enabled estimation for massive "high-frequency" datasets of hourly precipitation data observed over 20 years.

Simulation of unprecedented extreme heatwaves in France

Motivated by the exceptional French summer heatwave in 2019, the work in Opitz et al. (2021) capitalizes on the theoretical framework of scale-profile decompositions of Pareto limit process in the POT framework to build a spatial scenario generator for extreme events of yet unseen magnitude, featuring new yet realistic spatial patterns. Minimal modeling assumptions, guided by asymptotic theory, are embedded into powerful nonparametric resampling techniques (*e.g.*, Direct Sampling, Mariethoz and Caers, 2014) to generate extreme heatwave scenarios for France based on a gridded reanalysis dataset of maximum daily temperatures. The illustration in

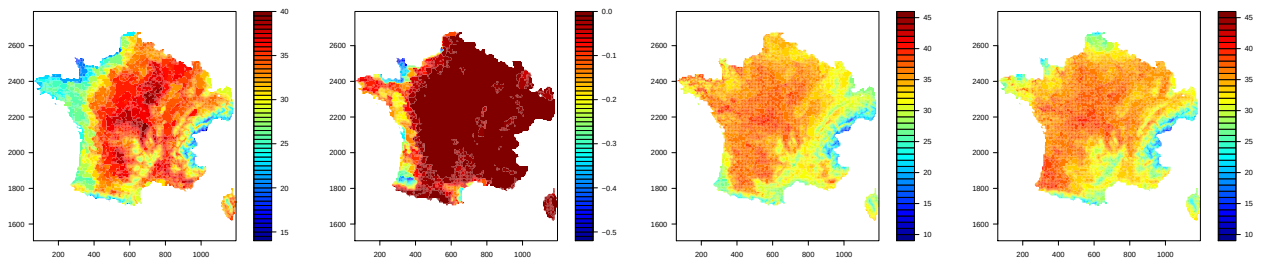


Figure 1.4: Illustration of semi-parametric resampling of extreme spatial temperature episodes in France. From left to right: most extreme event on original scale; same event on a scale highlighting the observed value relative to the local climatology; two artificial scenarios obtained through resampling.

Figure 1.4 shows an observed extreme episode and simulations obtained through resampling. A similar resampling approach is developed for space-time extreme precipitation episodes in the French Mediterranean region in Palacios-Rodriguez et al. (2020).

Predicting extreme space-time hotspots in sea surface temperatures

Large georeferenced datasets, often based on remote sensing, have become abundant in many domains. However, gaps may arise when sensors are defective or cannot provide useful data, for instance, due to cloud occusion with satellite-based instruments. Then, direct calculation of data summaries, such as the minimum value within a extreme space-time hotspot, is not possible. By adapting the INLA framework for extremes to a big space-time dataset of Red Sea surface temperatures, our method in Castro-Camilo et al. (2020) succeeds in filling such gaps and provides probabilistic predictions of hotspots. Developed in the context of the data challenge of the 2019 Extreme-Value Analysis conference (Zagreb, Croatie), our method achieves superior prediction performance for summaries of extreme space-time clusters in comparison to all other competing teams. Figure 1.5 illustrates the gap filling approach for a specific observation day.

1.5 Structure of the remainder of this manuscript

Chapter 2 provides background on point process analysis and extreme-value theory. Flexible modeling of univariate extremes is treated in Chapter 3, where flexibility stems either from incorporating covariates in generalized additive models, or is obtained in a univariate distributional sense by defining extensions of the generalized Pareto distribution. Results for the theory and practice of modeling bivariate extremes are exposed in Chapter 4. The Chapters 3 and 4 focus on theory and statistical tools for univariate and bivariate extremes; then, the modeling, simulation and prediction of extremes in the spatial and spatiotemporal setting is the topic of Chapter 5 in the asymptotic model setting, of Chapter 6 for the subasymptotic extensions, including approaches residing more specifically in the hierarchical modeling framework. General methodology for Bayesian modeling of point patterns using INLA is presented in Chapter 7, while Chapter 8 focuses on landslides and Chapter 9 on wildfires. Next, Chapter 10 recaps my

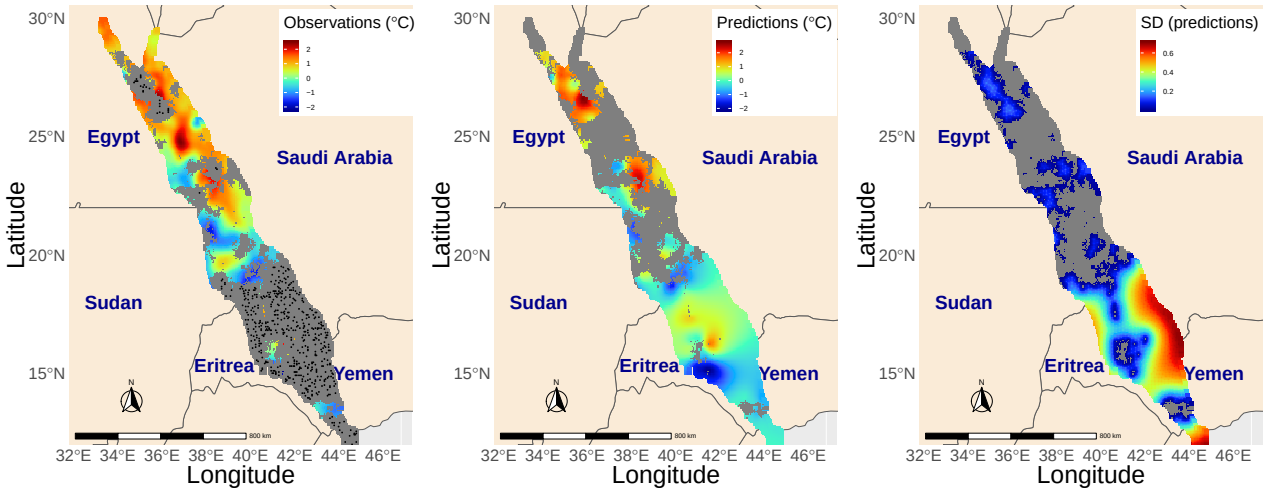


Figure 1.5: Illustration of gap filling in Red Sea surface temperatures. Data with gaps (left); model-based prediction of gaps (middle); associated prediction uncertainty (right).

scientific environment (projects, students, collaborating teams) and also mentions other research work that is not presented in detail in this manuscript. Finally, Chapter 11 summarizes important future developments that I envisage to extend the results presented for this habilitation. The appendix chapters A and B provide some technical background on specific aspects related to the INLA-SPDE approach and to extreme-value analysis, respectively.

Chapter 2

General background on extreme-value analysis and point processes

This chapter reviews theoretical and statistical background for point processes in the first section, and in the subsequent sections for univariate, bivariate and spatial extremes, as far as it directly concerns the results presented afterwards. In the context of extreme-value modeling, shortcomings of already existing approaches are discussed; they have been inspiration and motivation for tackling unresolved problems and developing the approaches presented in this habilitation. Some connections between extreme-value theory and point processes are pointed out. Textbooks providing a more extensive introduction into these fields are Cox and Isham (1980); Illian et al. (2008); Moller and Waagepetersen (2003); Chiu et al. (2013) for point processes and Resnick (1987); Embrechts et al. (1997); Coles (2001); Beirlant et al. (2004); de Haan and Ferreira (2007) for extreme-value analysis, among others.

2.1 Point processes for modeling rare event occurrences

Stochastic geometry (Illian et al., 2008; Chiu et al., 2013) is a branch of mathematics studying random geometric patterns, in particular the properties of collections of random geometric objects. A central concern is the study of point processes, *i.e.*, of random point patterns whose realizations represent locally finite sets of points in some space, typically $\mathbb{R}^d$ for $d \in \{1, 2, 3\}$ in practice. The term *point process* emphasizes that we consider such objects as stochastic processes, while the term *point pattern* puts more focus on the observed collection of points and its geometric patterns. The distinction of these two terms blurries in practice and may also strongly depend on the background of the modeler. We here use the abbreviation PP to refer to both entities, *point processes* and *point patterns*, as far as no annoying ambiguity arises. Many theoretical developments found their inspiration in important applications, such as telecommunication networks, stereology, forestry, species distribution modeling and occurrences of natural disasters.

Practical interest of studying random point patterns is usually direct towards the spatial variation of its intensity (*i.e.*, the expected number of points per spatial unit), and the interactions arising among points, such as clustering or inhibition. The case of clustered point patterns may arise if points occur preferentially in specific environmental conditions, such that

intensity variations are driven by the underlying spatial distribution of environmental factors. Examples of such factors are climatic conditions and land cover types (*e.g.*, type of vegetation) when modeling species distributions or wildfire occurrences. Model classes used to capture this type of interaction are known as *empirical* or *correlative* point process models. The variables describing relevant environmental conditions may be available as auxiliary data (*covariates*). However, often some or all of them have not been observed, are not available, or it is not known which variables might help to explain the variations in the intensity function. In these cases, we can add random components to the intensity function to represent the *a priori* unspecified environmental conditions. If points occur independently from each other given the known and unknown environmental conditions, then the class of Cox processes, and in particular the subclass of log-Gaussian Cox processes (LGCPs, Møller et al., 1998), provide an appropriate modeling framework for correlative modeling with random effects.

Other interaction types, of mechanistic nature, arise when points interact directly. Examples of such mechanisms are the competition or interaction among individuals whose position is represented by the points, or causal relationships in the spatiotemporal dynamics of point occurrence, where earlier points may spawn or inhibit the occurrence of later points. Such patterns can be captured through the class of Gibbs point processes, for which the intensity variation and interactions between points are encoded into an energy functional, which is additively composed of potential functions for singletons, pairs, triplets, and so on, of points. The energy function weights the point pattern configurations according to the "energy" required to generate them, which translates into a probability distribution over all PPs. A popular subclass of Gibbs processes are pairwise Markov processes, in which interactions are restricted to pairs of points and are usually defined locally in space; *e.g.*, two points interact if they are closer than a certain fixed distance.

Finally, points of a PP may be further characterized by additional numerical or type information. This information can be associated to points as *marks*. Marks could represent the category of a point if there are several types of points in a multi-type PP, or the magnitude of an event if points represent rare event occurrences.

Since the Poisson distribution $\text{Pois}(\lambda)$ is the limit of the binomial distribution with $np_n \to \infty$ as the number of trials n goes to infinity and the success probability p_n goes to 0, the Poisson distribution is the natural model for discrete rare events. More generally, Poisson point processes arise as the limit of appropriately defined binomial processes based on a grid discretization of $\mathbb{R}^d$ with mesh size going to 0 (see, *e.g.*, Chiu et al., 2013). The main works in this manuscript are based on LGCPs, *i.e.*, Poisson processes with Gaussian random effects in the log-intensity. They are detailed in Chapters 7 (general LGCP modeling), 8 (landslide modeling) and 9 (wildfire modeling), where correlative modeling approaches are developed. Some mention will also be made of mechanistic interactions and how to model them, especially in the outlook on ongoing work and research perspectives in Chapters 10 and 11.

2.1.1 Intensity function and Poisson point processes

We suppose that a point process is observed over a bounded domain $\mathcal{X} \subset \mathbb{R}^d$, also called the *observation window*. We write $\boldsymbol{X} = \{x_1, \dots, x_n\}$ to refer to the set of points of the pattern. We use notation $\boldsymbol{X}_B$ for $\boldsymbol{X} \cap B$, and $N(B) = |\boldsymbol{X}_B|$ for the number of points, where B is a Borel set. Therefore, N designates the random counting measure defined over the Borel sets of $\mathcal{X}$,

such that $N(B) \in \mathcal{N}_0$ are random variables corresponding to the number of points Borel sets $B \subset \mathcal{X}$. In practice, the observation of a PP is usually available for bounded $\mathcal{X}$, but considering unbounded domains such as $\mathcal{X} = \mathbb{R}^d$ may be relevant from a conceptual or theoretical stance. In particular, the observed point pattern may be considered as the realization of a point process defined over a domain that is larger than the observation window, which has consequences for practical modeling if observed points within $\mathcal{X}$ can directly interact with points outside $\mathcal{X}$.

The intensity function is fundamental in point pattern analysis, similar to mean surfaces in classical geostatistics. It indicates the expected number of events per unit of space at any position $x \in \mathcal{X}$; as such, it characterizes the *first-order* behavior of the point process. We note the intensity function

$$\lambda(x), \quad x \in \mathcal{X}. \tag{2.1}$$

A point x may refer to position in geographic space ($\mathcal{X} \subset \mathbb{R}^2$), or to a combination of time and geographic position in the observation window $\mathcal{X} = \{(s,t) \in \mathcal{S} \times \mathcal{T}\}$ with spatial and temporal observation windows $\mathcal{S}$ and $\mathcal{T}$, respectively.

A fundamental theoretical and practical model in stochastic geometry, and more generally in the field of stochastic processes, is given by the class of *Poisson point processes*, or *Poisson processes* in short. We point out that Poisson PPs are also used in the constructive representations of important limit processes in extreme-value theory: max-stable processes, and the more general class of max-id processes introduced in Chapter 6.

In Poisson PPs, the points arise independently according to the deterministic intensity function $\lambda(x)$, $x \in \mathcal{X}$, defined in (2.1). Therefore, in Borel sets $B \subset \mathcal{X}$ we observe Poisson-distributed counts

$$N(B) \sim \text{Pois}\left(\int_B \lambda(x)\,\mathrm{d}x\right). \tag{2.2}$$

The probability density of a point pattern $\boldsymbol{X} = \{x_1, \ldots, x_n\}$ observed in the window $\mathcal{X}$ is given as

$$f(\boldsymbol{X}) = \exp\left(-\int_{\mathcal{X}} \lambda(x)\,\mathrm{d}x\right) \prod_{i=1}^{n} \lambda(x_i), \tag{2.3}$$

where we use the convention $\prod_{i=1}^{0} \lambda(x_i) = 1$ in case of an empty pattern.

2.1.2 Log-Gaussian Cox processes

In the works presented in this habilitation, heavy use is made of the class of log-Gaussian Cox process (LGCP) models, wich are part of the larger class of Cox processes. Such processes include random components in the intensity function, and conditional to the intensity function a Poisson process is obtained. Suppose that $\boldsymbol{\Lambda} = \{\Lambda(x) \geq 0,\ x \in \mathcal{X}\}$ is a nonnegative stochastic process. The point process (or counting measure) N is a Cox process with intensity process $\boldsymbol{\Lambda}$ if, by analogy with (2.2), we have

$$N(B) \mid \boldsymbol{\Lambda} \sim \text{Pois}\left(\int_B \Lambda(x)\,\mathrm{d}x\right)$$

for all Borel sets B. In LGCPs, $\log \Lambda(s)$ is given as a Gaussian process. Usually there is no closed-form expression of the density

$$\mathbb{E}_{\boldsymbol{\Lambda}} \exp\left(-\int_{\mathcal{X}} \Lambda(x)\,\mathrm{d}x\right) \prod_{i=1}^{n} \Lambda(x_i),$$

such that for likelihood-based statistical inference we have to resort to techniques that allow handling the latent Gaussian process $\log \Lambda(x)$, such as INLA.

2.1.3 Marked point processes

Each point may carry additional information, different from its coordinates in the window $\mathcal{X}$. We can represent this information numerically and attach it as a *mark* to the point, defining a *marked point process*. For example, we can denote the points of a spatiotemporal marked point pattern $\boldsymbol{X} = \{x_1, \ldots, x_n\}$ as $x_i = (s_i, t_i, m_i)$, where s refers to spatial location, t to time, and m to the mark, which is element of a mark space $\mathcal{M}$. Marks may either be categorical, *i.e.*, $\mathcal{M} = \{1, \ldots, k_{\mathcal{M}}\}$, or they could be numeric, *e.g.*, $\mathcal{M} = \mathbb{R}_+$ in the case of wildfire ignition points marked by the burnt surface. Usually, marks are used to encode information that only exists if a point exists (*e.g.*, size or magnitude of the event occurring at the point, traits of the individual represented by the point), but not information that exists everywhere in space (*i.e.*, covariates, such as land cover, land use, climate and weather).

2.1.4 Estimation of the point process intensity

In the case of a stationary intensity function over a bounded observation window $\mathcal{X}$, the intensity parameter λ can be estimated empirically as the number of points per spatial unit, *i.e.*, $\hat{\lambda} = N(\mathcal{X})/|\mathcal{X}|$. In the absence of covariate information to be included into the model, nonparametric estimation of nonstationary point process intensities has been studied intensively and is part of the larger field of nonparametric statistics using kernel-based estimators. An intensity $\hat{\lambda}(x)$ can be estimated through an appropriate choice of kernel function $k(x_1, x_2)$ in $\hat{\lambda}(x) = \sum_{i=1}^{n} k(x, x_i)$. We do not further detail this approach here.

Instead, we suppose that we aim to estimate the parameter vector in a parametric model λ_θ making use of the information contained in covariates $z_j(x)$, $j = 1, \ldots, J$. For instance, θ may represent the vector of covariate coefficients in the log-linear intensity model

$$\lambda(x) = \beta_0 + \sum_{j=1}^{J} \beta_j z_j(x), \quad \boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_m)^T. \tag{2.4}$$

The estimation of the parameters $\boldsymbol{\beta}$ in (2.4) can be done in a similar way for Poisson processes and for more general processes featuring interaction of points (Waagepetersen and Guan, 2009). The estimation is formulated as a regression problem (Baddeley et al., 2010). Some extra care is needed for assessing the estimation uncertainty if the point pattern does not stem from a Poisson point process. A main difficulty in likelihood-based estimation of the parameters $\boldsymbol{\beta}$ in the Poisson PP model (2.3) is the computation of the integral $\int_{\mathcal{X}} \lambda(x)\,\mathrm{d}x$. Different approximation strategies have been developed to bypass this problem. A simple and robust approach is to discretize the

observation window, typically using a fine grid, and to assume that the intensity function $\lambda(x)$ is (approximately) constant within each grid cell. Then, the number of events in a cell C_k of the discretization of $\mathcal{X}$ has Poisson distribution, $N(C_k) \sim \text{Pois}(|C_k|)$, and the count variables $N(C_k)$ are mutually independent between different cells. Therefore, the estimation problem can be reformulated as a logistic regression equation with Poisson responses:

$$N_k \sim \text{Pois}(|C_k|), \quad \mathbb{E}N_k = \lambda_k, \quad \log(\lambda_k) = \beta_0 + \sum_{j=1}^{J} \beta_j z_{jk}. \tag{2.5}$$

One further has to decide which covariate values z_{jk} should be attributed to a cell C_k. For very small cells with little intra-cell variabillity of covariates, one may simple take the covariate value observed at a representative point within the cell, *e.g.*, at its centerpoint. When covariates can vary more strongly within individual cells, various approaches are possible, such as using an average value taken over all of the covariate values observed in the cell.

Other approaches for numerically approximating the integral $\int_{\mathcal{X}} \lambda(x)\,\mathrm{d}x$ consist in using appropriately weighted sums $\sum_k \omega_k \lambda(\tilde{x}_k)$ with discretization points $\tilde{x}_k$ and weights $\omega_k > 0$. A widely used technique of this type is known as the *Berman–Turner device* (Berman and Turner, 1992; Baddeley and Turner, 2000), for which the discretization points consist of the observed points x_i, $i = 1, \ldots, n$, augmented with a set of other points used to provide a relatively homogeneous discretization of space, *e.g.*, by using a regular grid. This discretization scheme allows writing the likelihood of the parameters $\boldsymbol{\beta}$ in the log-linear representation (2.4) as the likelihood of a logistic regression equation, such that standard implementations of logistic regression can be used. It becomes feasible to estimate parameter vectors $\boldsymbol{\beta}$ with a moderately large number J of components.

2.1.5 Challenges with the spatiotemporal modeling of point patterns

With spatiotemporal point patterns, a very large number of discretization points k (up to several millions) may arise in (2.5). Moreover, many sources of spatiotemporal variability in the point process intensity may not be appropriately captured through a log-linear influence of covariates, or appropriate covariates may not be available. Then, nonlinear and random effects should be included into the linear predictor in (2.5). My work on spatial and spatiotemporal modeling of landslides and wildfires, detailed in Chapters 8 and 9, tackles these challenges in the Bayesian framework of INLA, with the SPDE approach used for modeling spatially indexed random effects. The generic modeling framework, especially the use of subsampling techniques to cope with the large observation vectors, is exposed in Chapter 7.

In some applications (*e.g.*, wildfires) it is important to differentiate between moderate and extreme events, which can be done based on the magnitude of the event attached as a mark to the point. Then, the model developed for the marks should pay special attention to the modeling of extreme magnitudes. An example of such modeling is presented in Chapter 9 through our *Firelihood* model for wildfires.

2.2 Univariate extremes

We shortly recall asymptotic results and formulas of limit distributions in univariate extreme-value theory (EVT) of continuous variables, which are useful for statistical approaches. For statistical modeling, we focus on two approaches. First, in the *block maximum approach*, we divide the original data sample into blocks of same size and then extract the maximum from each block to obtain a sample of block maxima. Second, in the peaks-over-threshold approach, we fix a high threshold and then model the positive excesses above this threshold.

2.2.1 The trinity of univariate extreme-value limits

Univariate EVT is usually presented from the starting point of sample maxima and their extreme-value limit distributions resulting from appropriate rescaling, where the term *rescaling* refers to linear transformations, *i.e.*, location-scale transformations. Limit theory for sample maxima was the historical inception of EVT (Fréchet, 1927; Gnedenko, 1943). A more general treatment is possible through three equivalent limits arising for sample maxima, for threshold exceedances, and for the point process of sample points; on may call them the *trinity of univariate extreme-value limits* to appreciate their ubiquity in EVT. We shortly review these three approaches, their interconnections, and how they can be used for statistical inference.

Consider an i.i.d. sample $X_1, \ldots, X_n$ with $X_1 \sim F_X$, and denote by $M_n = \max_{i=1,\ldots,n} X_i$ the sample maximum. The fundamental *maximum domain of attraction* (MDA) condition is given by the following convergence property: if deterministic sequences a_n and $b_n > 0$ exist such that

$$F_X^n(a_n + b_n z) \to G(z), \quad n \to \infty, \quad z \in \mathbb{R}, \tag{2.6}$$

with a nondegenerate limit distribution G, then G is a *generalized extreme-value distribution* (GEVD). Equivalently, $(M_n - a_n)/b_n \to Z \sim G$ for $n \to \infty$. GEVDs constitute a class of three-parameter distributions defined as

$$\mathrm{GEV}(z; \xi, \sigma, \mu) = \exp(-T(z)), \quad T(z) = \begin{cases} \left(1 + \xi \frac{z-\mu}{\sigma}\right)_+^{-1/\xi}, & \xi \neq 0, \\ \exp\left(-\frac{z-\mu}{\sigma}\right), & \xi = 0, \end{cases}$$

with shape ξ, location μ and scale $\sigma > 0$. The support of the GEVD is parameter-dependent and given as $(-\infty, \mu - \sigma/\xi]$ if $\xi < 0$, $\mathbb{R}$ if $\xi = 0$, and $[\mu - \sigma/\xi, \infty)$ if $\xi > 0$. The shape parameter ξ is the main driver of the tail decay rate, with fast polynomial decay to a finite upper endpoint for $\xi < 0$, exponential decay for $\xi = 0$, and power-law decay for $\xi > 0$; a huge literature treats its estimation (de Haan and Ferreira, 2007).

The MDA condition can be reformulated equivalently as follows, which leads over to the *peaks-over-threshold* (POT) limit formulation. Consider a random variable $X \overset{d}{=} X_1$ with essential supremum $x^\star = \sup\{x : F(x) < 1\}$. Then, there exists a function $b(t) > 0$ such that

$$\frac{\mathrm{P}(X \geq t + y/b(t))}{\mathrm{P}(X \geq t)} \to (1 + \xi y)_+^{-1/\xi}, \quad t \to x^\star, \quad y \geq 0, \tag{2.7}$$

i.e., the distribution of $X \geq t + y/b(t) \mid X \geq t$ converges, and we obtain a limit distribution for the excesses above the threshold. Therefore, the asymptotic behavior of threshold exceedances

is closely related to the MDA condition, and the existence of the limits in the two frameworks is equivalent. As a consequence of the limit in (2.7), the *generalized Pareto distribution* (GPD) arises asymptotically as the distribution of rescaled threshold exceedances $(X - t) \mid X \geq t$, and its distribution function is given by

$$\mathrm{GPD}(y; \sigma_{\mathrm{GPD}}, \xi) = \begin{cases} 1 - (1 + \xi y/\sigma_{\mathrm{GPD}})_+^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp(-y/\sigma_{\mathrm{GPD}}), & \xi = 0, \end{cases}$$

with parameters $\sigma_{\mathrm{GPD}} > 0$ and $\xi \in \mathbb{R}$, where the tail index is the same as for the GEVD in the MDA limit, while the scale parameter σ_{GPD} can be expressed as a function of the threshold and the GEVD scale and location parameters.

Finally, we recall the *point process limit*, equivalent to the MDA and POT conditions, which provides a direct transition between the two above limits. Consider the point process of rescaled points, $\{(X_i - a_n)/b_n,\ i = 1, \ldots, n\}$, for $n = 1, 2, \ldots$. Then, (2.6) is equivalent to the weak convergence towards a Poisson point process

$$\{(X_i - a_n)/b_n,\ i = 1, \ldots, n\} \to \mathrm{PPP}(\kappa), \quad n \to \infty,$$

where the Poisson intensity measure is defined as $\kappa[z, \infty) = T(z)$ for values z in the support of the corresponding GEVD. To grasp the relationship to the two other limits, suppose that z is the maximum point of the limit process $\mathrm{PPP}(\kappa)$. The Poisson process probability of observing no point larger than z is $\exp(-\kappa[z, \infty)) = \exp(-T(z))$, which is the expression of the GEVD and establishes the link to limits for maxima. On the other hand, we can show straightforwardly that the distribution of the excess $y = z - u > 0$ of point z above a threshold u corresponds to the GPD as defined above.

2.2.2 Statistical considerations for univariate tail modeling

In contrast to the block maximum approach, POT techniques using threshold exceedances operate with the original event data and and allow detailed modeling of trends, seasonality and extremal clustering characteristics stemming from short-term dependence. Moreover, they also give more flexibility for balancing bias and variance through an appropriate choice of the threshold. In practice, the choice of a good threshold u should reflect the transition around which the asymptotic regime takes place for the tail approximation through the GPD to be valid. This implies a bias-variance trade-off, as a high threshold u leads to a good approximation (*i.e.*, low bias) but yields a small number of exceedances (*i.e.*, high variance), and vice versa for a low threshold. Experience shows that automatic threshold selection procedures are not always reliable. It is often difficult to find a good, natural and interpretable threshold, and parameter estimates are often sensitive to this choice (Scarrott and MacDonald, 2012). In principle, the block maximum approach would allow adapting the block size to handle the bias-variance trade-off, but often there is a natural choice of block size (*e.g.*, yearly maxima) according to cyclic behavior, such as seasons, which leads to less flexibility of this approach in comparison to the POT method.

A first systematic methodological treatment of the POT approach goes back to Davison and Smith (1990), who advocate to capture systematic variation in extreme events by including fixed covariate effects into the GPD parameters; parameters may be estimated by maximum

likelihood or the method of moments. Useful diagnostics for careful bias-variance assessment to choose a suitable threshold u above which observations are deemed to be extreme are discussed in Davison and Smith (1990), Northrop and Jonathan (2011) and Scarrott and MacDonald (2012).

2.2.3 A wider terminology of tail classes

There is no widely recognized standard for ordering univariate tail decay rates from the slowest to the fastest, although a broad characterization is given by the three domains of attraction of the maximum according to the value of the tail index: Fréchet ($\xi > 0$), Gumbel ($\xi = 0$) and reverse Weibull ($\xi < 0$). The Gumbel limit ($\xi = 0$) attracts distributions with highly diverse tail behavior such as finite upper bounds or heavy tails, and the classification according to the three domains of attraction may be too coarse in many situations. In addition to the maximum domains of attraction, we can refer to various other commonly used classes of upper tail behavior of a random variable $X \sim F$, for which we recall definitions for later reference. Relationships among such tail classes, and the membership of well-known parametric univariate distributions in such tail classes, are summarized in the Appendix B.3.

Definition 1 (Light-, heavy- and superheavy-tailed distributions)**.** *The distribution F is* heavy-tailed *if* $\exp(\lambda x)\overline{F}(x) \to \infty$ *as* $x \to \infty$*, for any* $\lambda > 0$*. Further, F is* superheavy-tailed *if* $F(\exp(\cdot))$ *is heavy-tailed. If F is not heavy-tailed, it is* light-tailed*.*

Definition 2 (Regularly varying functions and distributions (RV_α^0 and $\mathrm{RV}_\alpha^\infty$))**.** *A measurable function g is* regularly varying at infinity or at zero with index $\alpha \in \mathbb{R}$ *if* $g(tx)/g(t) \to x^\alpha$ *as* $t \to \infty$ *or* $t \to 0$*, respectively, for any* $x > 0$*. We write* $g \in RV_\alpha^\infty$ *or* $g \in RV_\alpha^0$*, respectively. If* $\alpha = 0$*, then g is said to be* slowly varying*. A probability distribution F with upper endpoint* $x^\star = \infty$ *is called* regularly varying with index $\alpha \geq 0$ *if* $\overline{F} \in RV_{-\alpha}^\infty$ *with the survivor distribution* $\overline{F} = 1 - F$ *of F. If* $x^\star < \infty$*, then F is* regularly varying at $x^\star$ with index α *if* $\overline{F}(x^\star - \cdot) \in RV_\alpha^0$*.*

A distribution F is in the Fréchet domain of attraction if and only if it is regularly varying at ∞ with $\alpha = 1/\xi > 0$, *i.e.*, $\overline{F} \in \mathrm{RV}_{-1/\xi}^\infty$.

Definition 3 (Exponential-tailed distributions (ET_α, $\mathrm{ET}_{\alpha,\beta}$))**.** *The distribution F with upper endpoint* $x^\star = \infty$ *is* exponential-tailed with rate $\alpha \geq 0$ *if for any* $x > 0$*,* $\overline{F}(t+x)/\overline{F}(t) \to \exp(-\alpha x)$*,* $t \to \infty$*. More specifically, if* $\alpha > 0$ *and* $\overline{F}(x) = r(x)\exp(-\alpha x)$*,* $r \in RV_\beta^\infty$*, we write* $F \in \mathrm{ET}_{\alpha,\beta}$*.*

By definition, $F \in \mathrm{ET}_\alpha$ with $\alpha \geq 0$ if and only if $\overline{F}(\log(\cdot)) \in \mathrm{RV}_{-\alpha}^\infty$. The class $\mathrm{ET}_{\alpha,\beta}$ with $\beta > -1$ is referred to as *gamma-tailed distributions*. Another important subclass of ET_α are the convolution-equivalent distributions.

Definition 4 (Convolution-equivalent distributions (CE_α))**.** *The distribution F is* convolution equivalent with index $\alpha \geq 0$ *if* $F \in \mathrm{ET}_\alpha$ *and* $\overline{F \star F}(x)/\overline{F}(x) \to 2\int_{-\infty}^{\infty} \exp(\alpha x)F(\mathrm{d}x) < \infty$ *when* $x \to \infty$*. We write* $F \in \mathrm{CE}_\alpha$*. We refer to the class* CE_0 *with* $\overline{F \star F}(x)/\overline{F}(x) \to 2$ *as* subexponential distributions*.*

Subexponential distributions are a large class of heavy-tailed distributions characterized by the following property, the principle of the *single big jump*: given i.i.d. random variables $X_j \sim F$, $j = 1, \ldots, d$, we have

$$\frac{\mathrm{P}\left(\sum_{j=1}^{d} X_j > x\right)}{\mathrm{P}\left(\max_{j=1}^{d} X_j > x\right)} \to 1, \quad x \to \infty.$$

In practice, all of the commonly used heavy tailed distributions belong to the class of subexponential distributions, CE_0.

Definition 5 (Weibull- and log-Weibull tailed distributions (WT_β, LWT_β))**.** *The distribution F is* Weibull-tailed with index $\beta > 0$ *if there exist $\alpha > 0$, $\gamma \in \mathbb{R}$, and $r \in RV_\gamma^\infty$ such that $\overline{F}(x) \sim r(x)\exp(-\alpha x^\beta)$. F is* log-Weibull-tailed *with index $\beta > 0$ if $F(\exp(\cdot)) \in \mathrm{WT}_\beta$.*

2.3 Dependent extremes

This section provides a quick overview over extremal dependence concepts for random vectors with a focus asymptotic representations used in the following chapters.

2.3.1 Dependence summaries of bivariate extremes

A random vector (X_1, X_2) with $X_j \sim F_{X_j}$, $j = 1, 2$, is said to display *asymptotic dependence* if the limit

$$\chi = \lim_{q \to 1} \mathrm{P}\left(X_1 \geq F_{X_1}^{-1}(q), X_2 \geq F_{X_2}^{-1}(q)\right)/(1-q) \tag{2.8}$$

exists and is positive; a limit of 0 defines *asymptotic independence*. If $X_1 \overset{d}{=} X_2$, then $\chi = \lim_{x \to x^\star} \mathrm{P}(X_2 \geq x \mid X_1 \geq x)$ with $x^\star$ the finite or infinite upper endpoint of the marginal distribution. The parameter $\chi \in [0, 1]$ is termed the (upper) *tail dependence coefficient* or the (upper) *tail correlation*, and a value $\chi > 0$ summarizes the strength of the dependence within the class of asymptotically dependent variables. *Asymptotic dependence* ($\chi > 0$) means that a certain joint tail stability prevails where joint exceedance behavior is independent of the threshold level.

Under asymptotic independence ($\chi = 0$)) where the probability of joint exceedances conditional to a marginal exceedance vanishes when moving to the most extreme quantile levels, a more useful summary is the rate at which the convergence to zero in Equation (2.8) occurs. A widely satisfied assumption (Ledford and Tawn, 1997) is

$$\mathrm{P}\{X_1 \geq F_{X_1}^{-1}(q), X_2 \geq F_{X_2}^{-1}(q)\} = \ell_{\mathrm{SV0}}(1-q)(1-q)^{1/\eta_X}, \qquad \eta \in [0, 1], \tag{2.9}$$

where $\ell_{\mathrm{SV0}} : [0, 1] \to \mathbb{R}_+$ is a function slowly varying at zero, *i.e.*, $\lim_{s \to 0} \ell_{\mathrm{SV0}}(sx)/\ell_{\mathrm{SV0}}(s) = 1$, $x > 0$. The parameter η is termed the *residual tail dependence coefficient* or the *coefficient of tail dependence*. Positive and negative extremal association are indicated respectively by $\eta \in (1/2, 1]$ and $\eta \in [0, 1/2)$, whilst asymptotically dependent variables have $\eta = 1$ and $\chi = \lim_{q \to 1} \ell_{\mathrm{SV0}}(1-q)$. The case of asymptotic independence with $\chi = 0$ and $\eta = 1$ is possible when $\ell_{\mathrm{SV0}}(1-q) \to 0$ as $q \to 1$. A value of $\eta = 0$ means that the left-hand side of (2.9) decays faster

than any power of $1-q$, whilst if the left-hand side is exactly zero for some $q<1$, we say that η is not defined.

Sometimes, the alternative index

$$\overline{\chi} = 2\eta - 1 \in [-1, 1] \tag{2.10}$$

is used, which was introduced by Coles et al. (1999). For bivariate Gaussian vectors, $\overline{\chi}$ is equal to the linear correlation coefficient.

2.3.2 Max-stable limits for componentwise maxima

Classical multivariate extreme-value theory (MEVT) provides support for the use of max-stable models for componentwise block maxima in random vectors, because they are the only possible limits of linearly rescaled componentwise maxima. This property further extends to stochastic processes (de Haan and Ferreira, 2007, Chapter 9). Max-stable processes provide a natural modeling framework for asymptotically dependent extremes. We first state the multivariate limit theory for maxima and threshold exceedances, and we then lead over to the spatial case in §2.4, and more generally to the case of stochastic processes.

For a sequence of i.i.d. random vectors $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d})^T \sim F_{\boldsymbol{X}}$, $i = 1, 2, \ldots$, the componentwise maximum

$$\boldsymbol{M}_n = (M_{n,1}, \ldots, M_{n,d})^T = \left(\max_{i=1}^{n} X_{i,1}, \ldots, \max_{i=1}^{n} X_{i,d} \right)^T \tag{2.11}$$

has joint distribution function $F_{\boldsymbol{X}}^n$. If vector sequences $\boldsymbol{a}_n = (a_{n,1}, \ldots, a_{n,d})^T$ and $\boldsymbol{b}_n = (b_{n,1}, \ldots, b_{n,d})^T > \boldsymbol{0}$ exist such that

$$F_{\boldsymbol{X}}^n(\boldsymbol{a}_n + \boldsymbol{b}_n \boldsymbol{z}) \to G, \quad n \to \infty, \tag{2.12}$$

with a nondegenerate joint limit distribution G, then G is called a *multivariate extreme-value distribution* (MEVD). Equivalently, $(\boldsymbol{M}_n - \boldsymbol{a}_n)/\boldsymbol{b}_n \to \boldsymbol{Z} \sim G$ for $n \to \infty$. Given the convergence (2.12), we say that F is in the *multivariate domain of attraction* of G. The class of limit distributions G coincides with the class of *max-stable distributions*, which are characterized by the existence of deterministic normalizing sequences $\boldsymbol{\alpha}_n = (\alpha_{n,1}, \ldots, \alpha_{n,d})^T$ and $\boldsymbol{\beta}_n = (\beta_{n,1}, \ldots, \beta_{n,d})^T > 0$ such that

$$G^n(\boldsymbol{\alpha}_n + \boldsymbol{\beta}_n \boldsymbol{z}) = G(\boldsymbol{z}), \quad \boldsymbol{z} \in \mathbb{R}^d, \quad n \in \mathbb{N};$$

i.e., max-stable distributions satisfy equality with the limit in (2.12) when used for $F_{\boldsymbol{X}}$. The univariate marginal distributions in MEVDs are of GEVD type, and the class of GEVDs defines the class of univariate max-stable distributions.

To focus solely on the extremal dependence structure, it is useful to abstract away from the marginal distributions F_j of X_j by transforming them to a standardized marginal distribution $F^\star$. Here we transform a random variable $X \sim F$, supposed to be continuous, towards a marginal distribution of unit Fréchet or standard Pareto type as follows:

$$X^F = -\frac{1}{\log F(X)}, \quad X^P = \frac{1}{1 - F(X)}, \tag{2.13}$$

with superscripts F and P indicating the Fréchet and Pareto case, respectively. When the specific choice of one or the other does not make a difference, we simple use the $\star$-notation for both and write $X^\star \sim F^\star$. Both of these transformations ensure $X^\star \geq 0$ and $\mathrm{P}(X^\star > x) \sim 1/x$ for $x \to \infty$. The subsequent limit results hold in both cases, and they even hold in the more general case of any probability integral transform that leads to a standardized distribution with these two properties. As for notation, given a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)^T$, we write $F^\star_{\boldsymbol{X}}$ for the distribution of the random vector $\boldsymbol{X}^\star = (X_1^\star, \ldots, X_d^\star)^T$ with normalized margins:

$$\boldsymbol{X}^\star \sim F^\star_{\boldsymbol{X}}, \quad F^\star_{\boldsymbol{X}}(\boldsymbol{x}) = F^\star_{\boldsymbol{X}}(x_1, \ldots, x_d) = F_{\boldsymbol{X}}\left(F_1^{-1}(F^\star(x_1)), \ldots, F_d^{-1}(F^\star(x_d))\right),$$

which is similar to the idea of copula modeling (Joe, 2014), where the normalized marginal distributions are assumed to be uniform. Benefits of standardizing to marginal Pareto distributions are summarized in Klüppelberg and Resnick (2008), where the authors introduce the concept of *Pareto copula*.

The multivariate extreme value limit (2.12) is equivalent to the following two conditions:

1. the univariate domain of attraction condition holds for X_j, $j = 1, \ldots, d$, such that

   $$F_j^n(a_{n,j} + b_{n,j} z) \to G_j(z), \quad n \to \infty,$$

   with a GEVD G_j;

2. the multivariate domain of attraction condition holds for standardized data $\boldsymbol{X}^\star$, such that

   $$(F^\star_{\boldsymbol{X}})^n(n\boldsymbol{z}) \to G^\star(\boldsymbol{z}), \quad n \to \infty;$$

   equivalently, we observe the convergence in distribution

   $$\frac{\boldsymbol{M}_n^\star}{n} = \left(\max_{i=1}^n \frac{X_{i,1}^\star}{n}, \ldots, \max_{i=1}^n \frac{X_{i,d}^\star}{n}\right)^T \to \boldsymbol{Z}^\star \sim G^\star, \quad n \to \infty.$$

Then, the joint distribution $G^\star$ has unit Fréchet marginal distributions $G_j^\star(z) = \exp(-1/z)$ for $z > 0$, and it is max-stable satisfying $(G^\star)^n(n\boldsymbol{z}) = G^\star(\boldsymbol{z})$ for any $n \in \mathbb{N}$. We say that $G^\star$ is *simple max-stable*.

An important and useful implication of this result is that the convergence of marginal distributions and convergence of the dependence structure (*i.e.*, of the marginally standardized multivariate distribution) can be considered separately, which allows for important simplifications in statistical modeling.

2.3.3 Representations of multivariate asymptotic dependence

A simple max-stable distribution $G^\star$ has representation

$$G^\star(\boldsymbol{z}) = \exp(-V^\star(\boldsymbol{z})), \quad \boldsymbol{z} \geq \boldsymbol{0},$$

with the *exponent function* $V^\star$, which is homogeneous of order -1, *i.e.*, $tV^\star(t\boldsymbol{z}) = V^\star(\boldsymbol{z})$ for $t > 0$. It satisfies $V^\star(\boldsymbol{z}) = \infty$ if $\min_{j=1}^d z_j = 0$. More generally, and by analogy with

the univariate case, we can uniquely define an *exponent measure* $\kappa^\star$ on $[0,\infty]^d$ such that $V^\star$ corresponds to the "survival measure" of $\kappa^\star$:

$$V^\star(\boldsymbol{z}) = \kappa^\star([\mathbf{0},\boldsymbol{z}]^C), \quad \boldsymbol{z} \in [\mathbf{0},\boldsymbol{\infty}).$$

The measure $\kappa^\star$ can be evaluated for any Borel set B in $\mathbb{R}_+^d$ and is (-1)-homogeneous: $t\kappa(tB) = \kappa(B)$ for $t > 0$.

To lead over to threshold-based representations, we can reformulate the domain of attraction condition using *multivariate regular variation* (MRV). A multivariate distribution $F_{\boldsymbol{X}}^\star$ is attracted by $G^\star$ with exponent function $V^\star$ if and only if it satisifies the following MRV condition:

$$t\,\mathrm{P}(\boldsymbol{X}^\star \not\leq t\boldsymbol{x}) = t\,(1 - F_{\boldsymbol{X}}^\star(t\boldsymbol{x})) \to V^\star(\boldsymbol{x}), \quad t \to \infty, \quad \boldsymbol{x} > \mathbf{0}. \tag{2.14}$$

Combined with the (-1)-homogeneity of $V^\star$, MRV suggests using the following asymptotic approximation in practice:

$$\mathrm{P}(\boldsymbol{X}^\star \not\leq \boldsymbol{x}) = \overline{F}_{\boldsymbol{X}}^\star(\boldsymbol{x}) = 1 - F_{\boldsymbol{X}}^\star(\boldsymbol{x}) \approx V^\star(\boldsymbol{x}) \quad \text{if} \quad \min_{j=1}^{d} x_j \gg 0.$$

The MRV condition, and the peaks-over-threshold stability that it represents, can be reformulated more generally. For this, we consider a homogeneous function ℓ continuous in $\mathbf{0}$, where homogeneity means that $\ell(t\boldsymbol{x}) = t\ell(\boldsymbol{x})$ for any $t > 0$, $\boldsymbol{x} \geq \mathbf{0}$. We call this function the *cost functional*, or also the *risk functional* or *aggregation functional*. Examples for ℓ are norms, the minimum, or an order statistics. Then, the MRV condition implies

$$t\,\mathrm{P}\left(\frac{\boldsymbol{X}^\star}{\ell(\boldsymbol{X}^\star)} \in A,\ \ell(\boldsymbol{X}^\star) > tr\right) \to S_\ell(A) \times \frac{\theta_\ell}{r}, \quad t \to \infty, \tag{2.15}$$

with the so-called *angular distribution* $S_\ell(\cdot)$ (also called *angular measure*, *spectral distribution*, *profile distribution*) defined on the unit sphere $\mathbb{S}_\ell$ with respect to ℓ, and the *extremal coefficient* $\theta_\ell \geq 0$ associated to ℓ. Here, we define $S_\ell(\cdot)$ as a probability distribution with overall mass 1, but sometimes one rather considers the measure $\theta_\ell S_\ell(\cdot)$ with overall mass θ_ℓ. The angular distribution indicates along which directions $\boldsymbol{W} = \boldsymbol{X}^\star/\ell(\boldsymbol{X}^\star)$ the extreme values in $\boldsymbol{X}^\star$ tend to concentrate when the magnitude $R = \ell(\boldsymbol{X}^\star)$ increases. The classical extremal coefficient $\theta_d = V^\star(1,\ldots,1)$ (Schlather and Tawn, 2003) arises for the maximum norm $\ell(\boldsymbol{x}) = \|\boldsymbol{x}\|_\infty$. To further extend this framework, we can allow choosing two different risk functionals for defining the direction and the magnitude.

In practice, we choose a aggregation functional ℓ and fix a high threshold r, and then we consider events $\boldsymbol{X}$ with $\ell(\boldsymbol{X}^\star) \geq r$ as being extreme. This procedure is based on the following convergence in distribution resulting from (2.15):

$$\boldsymbol{t}^{-1}\boldsymbol{X}^\star \mid \ell(\boldsymbol{X}^\star) \geq t \to \boldsymbol{Y}^\star \sim H_\ell^\star, \quad t \to \infty \tag{2.16}$$

where we call $H_\ell^\star$ an *ℓ-Pareto distribution* (Dombry and Ribatet, 2015). By construction, we have

$$H_\ell^\star(\cdot) = \frac{\kappa_\ell(\cdot)}{\kappa_\ell(\{\boldsymbol{y} \in \mathbb{R}_+^d \mid \ell(\boldsymbol{y}) \geq 1\})}.$$

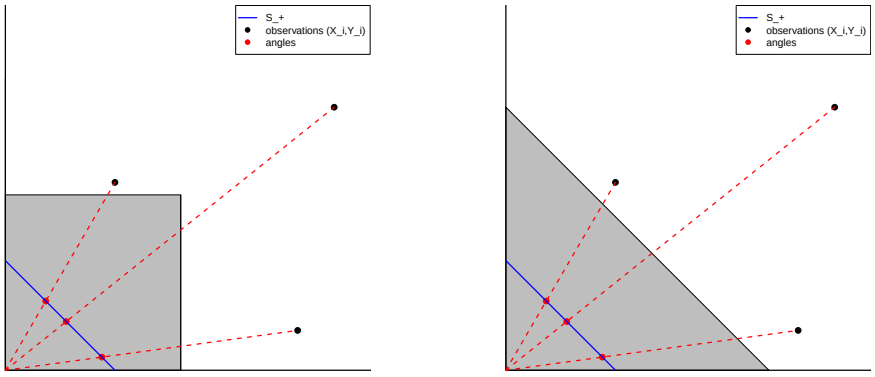


Figure 2.1: llustration of empirical spectral distributions for aggregation functional ℓ given as the maximum norm (left) and the L^1-norm (right). Observations outside the gray boxes are considered as multivariate threshold exceedances, and their projection on the unit sphere with respect to the norm defines observations of the spectral distribution S_ℓ.

The case $\theta_\ell S_\ell \equiv 0$ is possible, where we set $\theta_\ell = 0$ while S_ℓ and $H_\ell^\star$ are not defined. For instance, this situation arises in the case of asymptotic independence with $\ell = \min$. We further point out that the choice of $\ell(\boldsymbol{x}) = x_{j_0}$, *i.e.*, of using the value of a fixed component j_0 as aggregation functional, refers to a special case of the conditional extremes approach (Heffernan and Tawn, 2004).

An illustration of empirical angular distributions using the maximum norm or the L^1-norm for ℓ to define bivariate exceedances is given in Figure 2.1. Finally, we study the example of using the (generalized) maximum norm for the aggregation functional in more detail. Suppose that we consider an event $\boldsymbol{X}$ as extreme if we observe a threshold exceedance $X_j > u_j$ in at least one of the components $j = 1, \ldots, d$. Then, standardized marginal thresholds are $u_j^\star = 1/(1 - F_j(u_j))$, $j = 1, \ldots, d$. Therefore, we use the generalized maximum norm defined by

$$\ell(\boldsymbol{x}^\star) = \max_{j=1}^{d} x_j^\star / u_j^\star,$$

and we fix the threshold of the risk functional ℓ to $r = 1$ to identify extreme multivariate events with $\ell(X^\star) \geq r$. The extremal coefficient associated to this norm is given as $\theta_\ell = V^\star(u_1^\star, \ldots, u_d^\star)$. Moreover, the ℓ-Pareto distributions $H_\ell^\star$ arising for exceedances with respect to this norm, and taking into account the marginal transformations according to the tail parameters ξ_j, $\sigma_j > 0$ and μ_j for $j = 1, \ldots, d$, are known as multivariate generalized Pareto distributions (Rootzén and Tajvidi, 2006; Rootzén et al., 2018; Rootzén et al., 2018).

2.4 Spatial extremes

Based on the univariate and multivariate limit theory exposed in the preceding sections, there has been increasing development of theoretical and statistical methodology for modeling spatial extremes over the last 20 years, motivated by numerous applications in climatology and environmental sciences. Given a stochastic process $\{X(s),\, s \in \mathbb{R}^D\}$, written $X(s)$ in short, we say that

it is in the maximum domain of attraction of a max-stable process $\{Z(s),\, s \in \mathbb{R}^D\}$, written $Z(s)$ in short, if the max-stable convergence (2.12) holds for all finite-dimensional distributions. To provide mathematical formulas, we first fix some notations: for a collection of distinct locations $\boldsymbol{s} = (s_1, \ldots, s_m)$, we write $X(\boldsymbol{s}) = (X(s_1), \ldots, X(s_m))^T \sim F_{\boldsymbol{s}}$. Then, the convergence condition requires that appropriate normalizing deterministic vector sequences $\boldsymbol{a}_n(\boldsymbol{s})$ and $\boldsymbol{b}_n(\boldsymbol{s}) > \boldsymbol{0}$ and max-stable limit distributions $G_{\boldsymbol{s}}(\boldsymbol{z})$ exist for any choice of $\boldsymbol{s}$, *i.e.*,

$$F_{\boldsymbol{s}}(\boldsymbol{a}_n(\boldsymbol{s}) + \boldsymbol{b}_n(\boldsymbol{s})\boldsymbol{z}) \to G_{\boldsymbol{s}}(\boldsymbol{z}), \quad n \to \infty, \quad \boldsymbol{z} \in \mathbb{R}^m.$$

Therefore, many results and modeling mechanisms obtained within multivariate extreme-value theory and statistics can be adapted to the setting of stochastic processes by focusing on their bivariate or higher-dimensional multivariate marginal distributions.

A constructive representation of max-stable processes, known as the *spectral representation* (De Haan, 1984; Schlather, 2002), is often useful in spatial modeling. It states that any simple max-stable process $Z^\star(s)$ (*i.e.*, having unit Fréchet margins) can be constructively represented as

$$Z^\star(s) = \max_{i=1,2,\ldots} R_i W_i(s), \qquad s \in \mathcal{S} \subset \mathbb{R}^d, \tag{2.17}$$

where $\{R_i\}$ are the points of a Poisson process on $\mathbb{R}_+$ with intensity measure $r^{-2}\,\mathrm{d}r$, and $W_i(s)$ are independent copies of a random process $W(s)$ satisfying $\mathrm{E}[\max\{W(s), 0\}] = 1$ and being independent of $\{R_i\}$. The commonly used spatial models for max-stable processes are constructed from specific choices of $W(s)$ in (2.17). Based on a centered Gaussian process $\tilde{W}(s)$, constructions with $W(s) = (\tilde{W}(s))_+^\nu$ lead to extremal-t processes (Opitz, 2013) with degrees of freedom parameter $\nu > 0$, whereas constructions with $W(s) = \exp(\tilde{W}(s) - \mathrm{Var}(\tilde{W}(s))/2)$ lead to max-stable processes of Brown–Resnick type (Kabluchko et al., 2009). In all cases, appropriate multiplicative scaling constants have to be applied to the spectral process $W(s)$ to ensure the unit moment condition.

2.4.1 Peaks-over-threshold limits for stochastic processes

The generalization of POT limits to the setting of stochastic processes is more intricate since the exceedance regions in multivariate space $\mathbb{R}^d$ used in conjunction with the POT-convergence (2.15) must be replaced by function spaces that are usually infinite-dimensional (Ferreira and De Haan, 2014). Most results in the literature have been derived for the space of continuous functions over a nonempty compact domain $\mathcal{K} \subset \mathbb{R}^D$, denoted by $C(\mathcal{K})$ and endowed with the supremum norm $\|f\|_\infty = \sup_{s \in \mathcal{K}} |f(s)|$. The restriction of this function space to non-negative functions is denoted by $C_+(\mathcal{K})$.

As before, we can work with homogeneous aggregation functionals ℓ. A functional $\ell : C_+(\mathcal{K}) \to [0, \infty)$ is called *risk functional* (or *cost functional*, *aggregation functional*) if it is continuous and homogeneous, *i.e.*, $\ell(tf) = t\,\ell(f)$ for $t \geq 0$. Functional limits equivalent to the multivariate limits (2.15) and (2.16) arise (Dombry and Ribatet, 2015). Limits arising in (2.16) are known as *ℓ-Pareto processes* (Ferreira and De Haan, 2014; Dombry and Ribatet, 2015; Thibaud and Opitz, 2015; Palacios-Rodriguez et al., 2020), and these limit processes can be transformed towards more general asymptotic univariate marginal distributions based on the parameters $\xi(s)$, $\sigma(s) > 0$, $\mu(s)$ for shape, scale and location, respectively, to obtain the class

of *generalized ℓ-Pareto processes*. We first provide a constructive definition of these two model classes.

Definition 6 (Standard ℓ-Pareto process)**.** *Suppose that $Y^\star = \{Y^\star(s),\, s \in \mathcal{K}\}$ is a stochastic process in $C_+(\mathcal{K})$. We call $Y^\star$ a* standard ℓ-Pareto process *if it can be represented as*

$$Y^\star(s) \overset{d}{=} RW(s), \quad s \in K, \tag{2.18}$$

where

1. *W is a stochastic process in $C_+(\mathcal{K})$ satisfying $\ell(W) = 1$;*
2. *R has a standard Pareto distribution;*
3. *W and R are stochastically independent.*

The process W is called *spectral process* or *profile process*; is is defined on the unit sphere in $C_+(\mathcal{K})$ with respect to ℓ. This definition can be shown to be equivalent to the POT stability property: for any $u \geq 1$, the distribution of the renormalized exceedances $\{u^{-1}Y^\star | \ell(Y^\star) \geq u\}$ is equal to the distribution of $Y^\star$. By construction, we get $W \overset{d}{=} Y^\star/\ell(Y^\star)$ and $R \overset{d}{=} \ell(Y^\star)$. We now define generalized ℓ-Pareto process to provide flexibility in the marginal distributions according to the location-scale-shape parametrization of univariate EVT.

Definition 7 (Generalized ℓ-Pareto process)**.** *Given an ℓ-Pareto process $Y^\star$ and continuous real functions $\xi(s)$, $\sigma(s) > 0$ and $\mu(s)$ over $\mathcal{K}$, a* generalized ℓ-Pareto process *is any process constructed as*

$$Y(s) \overset{d}{=} \begin{cases} \mu(s) + \sigma(s)(Y^\star(s)^{\xi(s)} - 1)/\xi(s), & \xi(s) \neq 0, \\ \mu(s) + \sigma(s) \log Y^\star(s), & \xi(s) = 0, \end{cases} \quad s \in \mathcal{K}. \tag{2.19}$$

In the above Definitions 6 and 7, we can think of s as a general index that may refer to locations in time, space or space-time. By analogy with the multivariate setting of random vectors, different choices of ℓ may be used to answer different questions. If $\ell(f) = \max_{j=1}^d f(s_j)/u_j$ for certain points $s_j \in \mathcal{K}$, $j = 1, \ldots, d$, we focus on processes with at least one exceedance of the marginal thresholds $u_j > 0$. By contrast, $\ell(f) = \min_{j=1}^d f(s_j)/u_j$ requires exceedances at each of the d points. The original definition of a Pareto process (Ferreira and De Haan, 2014) uses $\ell(f) = \sup_{s\in\mathcal{K}} f(s)$, but conditioning on a different $\ell(f)$ rather than $\sup_{s\in\mathcal{K}} f(s)$ is desirable in applications where data are only observed at a finite number of irregularly spaced points.

2.4.2 The trinity of limits of stochastic processes

By analogy with the trinity of univariate EVT limits outlined in §2.2.1, we recall the different forms of convergence of extremes of continuous processes in terms of block maxima, threshold exceedances and point processes, as done in our paper Thibaud and Opitz (2015). Throughout, we will use the symbol $\Longrightarrow$ to indicate weak convergence of random elements from the univariate, multivariate or functional domain. For i.i.d. copies $X_1, X_2, \ldots$ of a stochastic process $X = \{X(s),\, s \in \mathcal{K}\}$ with continuous sample paths, we say that X is in the maximum domain of attraction of a max-stable process $Z = \{Z(s),\, s \in \mathcal{K}\}$ if there exist sequences of deterministic

continuous normalizing functions $a_n(s)$ and $b_n(s) > 0$ and such that (de Haan and Ferreira, 2007, Ch. 9)

$$\left\{\max_{i=1}^{n} b_n(s)^{-1}\{X_i(s) - a_n(s)\}\right\} \Longrightarrow \{Z(s)\}, \quad n \to \infty, \quad s \in \mathcal{K}, \tag{2.20}$$

in $C(\mathcal{K})$, with the limit process Z having nondegenerate univiariate GEVDs $F_{Z(s)}$. The distribution of the marginally transformed process $Z^\star$ with unit Fréchet margins, *i.e.*, of the simple max-stable process, is fully characterized by the exponent measure Λ on $C(\mathcal{K}) \setminus \{0\}$ through the relation (Giné et al., 1990)

$$\Lambda\left[\bigcup_{j=1,\ldots,d}\left\{f \in C(\mathcal{K}) : \sup_{s\in\mathcal{K}_j} f(s) \geq z_j\right\}\right] = -\log \mathrm{pr}\left\{\sup_{s\in\mathcal{K}_1} Z^\star(s) \leq z_1, \ldots, \sup_{s\in\mathcal{K}_d} Z^\star(s) \leq z_d\right\} \tag{2.21}$$

for any collection of nonempty compact sets $K_j \subset K$ and $z_j > 0$ for $j = 1, \ldots, d$. The measure Λ is uniquely defined if we impose the constraint $\Lambda\{C(\mathcal{K}) \setminus C_+(\mathcal{K})\} = 0$. We obtain the unique version as $\Lambda_+(B) = \Lambda[\{f \in C(\mathcal{K}) : f_+ \in B\}]$, for measurable $B \subset C_+(\mathcal{K}) \setminus \{0\}$, with $f_+(s) = \max\{f(s), 0\}$. When the sets $\mathcal{K}_j = \{s_j\}$ are singletons, expression (2.21) corresponds to the multivariate exponent function $V^\star$.

As before, max-stability of $Z^\star$ implies that the measure Λ_+ is homogeneous of order -1, *i.e.*, $\Lambda_+(tB) = t^{-1}\Lambda_+(B)$, $t > 0$. We now consider the limit behavior according to the scale-profile decomposition of ℓ-Pareto processes given in Definition 6. For a risk functional ℓ and $f \in C_+(\mathcal{K})$ with $\ell(f) > 0$, consider the pseudo-polar coordinates (r, f_0) with $r = \ell(f)$ and $f_0 = f/\ell(f) \in \mathbb{S}_\ell = \{f \in C_+(\mathcal{K} \mid \ell(f) = 1\}$. If $\theta_\ell(\mathcal{K}) = \Lambda_+[\{\ell(f) \geq 1\}] > 0$, arguments similar to the multivariate pseudo-polar representation (2.15) imply the factorization

$$\Lambda_+(\mathrm{d}f) = \theta_\ell(K)\, r^{-2}\mathrm{d}r S_\ell(\mathrm{d}f_0), \quad r > 0, \tag{2.22}$$

with S_ℓ the *ℓ-spectral distribution* on $\mathbb{S}_\ell$.

To establish the trinity of functional EVT limits, we consider the following additional assumptions:
$(A1)$ Marginal convergence: sequences of normalizing continuous functions and $a_n(s)$ $b_n(s) > 0$ exist such that, for each $s \in \mathcal{K}$,

$$\max_{i=1,\ldots,n} b_n(s)^{-1}\{X_i(s) - a_n(s)\} \Longrightarrow Z(s), \quad n \to \infty, \quad s \in \mathcal{K}, \tag{2.23}$$

with a nondegenerate limit $Z(s)$. The convergence (2.23) is uniform in $s \in \mathcal{K}$.
$(A2.i)$ Normalized max-stable convergence:

$$\left\{\max_{i=1,\ldots,n} n^{-1} X_i^\star(s)\right\} \Longrightarrow \{Z^\star(s)\}, \quad n \to \infty, \tag{2.24}$$

where the max-stable limit process $Z^\star(s)$ is characterized by its exponent measure Λ_+.
$(A2.ii)$ Standard point process convergence:

$$\left\{n^{-1} X_i^\star(s),\, i = 1, \ldots, n\right\} \Longrightarrow \mathcal{P}, \quad n \to \infty, \tag{2.25}$$

where $\mathcal{P} = \{P_i(s), i = 1, 2, \ldots\}$ is a Poisson process with intensity measure Λ_+.
$(A2.iii)$ Convergence of standard sup-exceedances:

$$\left\{ n^{-1} X^\star(s) : \sup_{s \in \mathcal{K}} X^\star(s) > n \right\} \Longrightarrow \{Y^\star_{\text{sup}}(s)\}, \quad n \to \infty,$$

where $Y^\star_{\text{sup}}$ is a standard sup-Pareto process with sup-spectral distribution S_{sup} associated to the exponent measure Λ_+ of $Z^\star$ in (2.24) through (2.22). The convergences in assumptions $(A2.i)$, $(A2.ii)$ and $(A2.iii)$ are assumed to hold in $C_+(\mathcal{K}) \setminus \{0\}$ or some suitably defined closure thereof.

Then, from de Haan and Ferreira (2007), assumptions $(A1)$ and $(A2.i)$ are together equivalent to the max-stable convergence in (2.20). From Ferreira and De Haan (2014), assumptions $(A2.i)$, $(A2.ii)$ and $(A2.iii)$ are equivalent. By assuming $(A2.i)$, $(A2.ii)$ or $(A2.iii)$, we get the POT convergence

$$\left\{ n^{-1} X^\star(s) : \ell(X^\star) > n \right\} \Longrightarrow \{Y^\star_\ell(s)\}, \quad n \to \infty, \tag{2.26}$$

from Dombry and Ribatet (2015, Theorem 3), where $Y^\star_\ell$ is a standard ℓ-Pareto process characterized by its ℓ-spectral distribution S_ℓ according to the exponent measure Λ_+, see (2.22).

The convergence (2.26) builds the link between max-stable limits for maxima and ℓ-Pareto limits for threshold exceedances, thus establishing a basis for threshold-based inference as follows. From its constructive definition, the distribution of a standard ℓ-Pareto process is $r^{-2}\mathrm{d}rS_\ell(\mathrm{d}f_0)$, which is also equal to $\Lambda_+(\mathrm{d}f)/\theta_\ell(\mathcal{K})$ from (2.22) with $f = rf_0$. Hence the convergence in (2.26) conveys that

$$\mathrm{P}\left\{ X^\star \in B \mid \ell(X^\star) > r \right\} \approx r\Lambda_+(B)/\theta_\ell(\mathcal{K}), \quad r \to \infty,$$

for $B \subset \{f \in C_+(\mathcal{K}) : \ell(f) > r\}$. From de Haan and Ferreira (2007, Theorem 9.3.1), any of the assumptions $(A2.i)$, $(A2.ii)$ and $(A2.iii)$ imply that $\mathrm{P}\{\ell(X^\star) \geq r\} \sim \theta_\ell(\mathcal{K})/r$ as $r \to \infty$. Thus, the convergence of ℓ-exceedances gives $\mathrm{P}(X^\star \in B) \approx \Lambda_+(B)$ for extreme events B defined in terms of ℓ-exceedances.

Assumption $(A1)$ implies the convergence of marginal pointwise maxima to GEVDs, such that equivalent univariate POT limits exist. From a Pareto process perspective, it is convenient to fix a high threshold function $u(s)$ and to assume that the marginal univariate distributions satisfy

$$\mathrm{P}(X(s) > x) = \left[1 + \xi(s)(x - \mu^{\text{marg}}(s))/\sigma^{\text{marg}}(s)\right]_+^{-1/\xi(s)}, \quad x > u(s), \tag{2.27}$$

which establishes the link to the univariate tail probabilities of the generalized Pareto process through the usual location-scale-shape parameters from univariate EVT. Due to various standardization constants and the conditional process in (2.26), there are slight differences in the location and scale parameters of marginal distributions in (2.27) and of the generalized Pareto process in (7); see the details from our work in Palacios-Rodriguez et al. (2020).

2.5 Statistical considerations and challenges for spatiotemporal extremes

In this section, we shortly summarize statistical approaches for dependent extremes. In particular, we highlight challenges that arise for spatiotemporal statistical modeling of extremes.

At an exploratory stage, empirical versions of the bivariate summaries in §2.3.1 can be explored. They can help choose appropriate models and inference tools. Part of my work presented in Chapter 4 develops flexible modeling and inference techniques for such bivariate summaries, and of more general related summaries, by constructing generalized additive models for the situation where the strength of extremal dependence may depend on covariates in a nonlinear way.

2.5.1 Modeling with asymptotic models

Inference for max-stable models is often tricky, especially if the number of components in the observation vectors is much larger than 2, for instance in spatial and spatiotemporal modeling. Full likelihoods can only be calculated in small dimensions, which motivated the use of alternative, less efficient inference techniques, such as composite likelihoods (Padoan et al., 2010). While componentwise block maxima are relatively easy to extract from data, a general drawback with such maxima data is that the vectors of maxima may contain components from different extreme events, such that interpretation of results in terms of the original extreme event episodes can be awkward.

The above limitations of max-stable models have led to a stronger focus on POT-based methods in recent years. The choice of the threshold level in POT modeling usually requires some thought and exploratory analysis. Moreover, in the multivariate and functional setting, there is no unique ordering of values and therefore no unique way of defining the magnitude of an extreme event and of setting a threshold. In practice, the choice of a norm, or more generally of an aggregation functional ℓ, as presented in the previous section, provides satisfactory flexibility and can be adapted to the specific scientific questions to be answered. In inferential procedures, the choice of the threshold further determines how non-extreme observations are removed from the sample or how their values are censored. Censoring mechanisms may lead to more or less computationally demanding inference procedures, where the effort to calculate multivariate censoring probabilities is often the dominating component of numerical cost. Threshold choice in POT approaches has been discussed extensively in the literature (Scarrott and MacDonald, 2012), while the validity of max-stability with respect to the choice of block length in block maxima techniques is usually not questioned. However, the distribution of observed block maxima may still be far from the asymptotic regime, especially in the case of asymptotic independence, which merits stronger attention and the development of more flexible alternative model classes.

The framework of generalized ℓ-Pareto processes with an aggregation functional ℓ for modeling threshold exceedances has become increasingly popular because it circumvents many of the computational bottlenecks of max-stable processes. Some typical difficulties arise in practice. It is often desirable to work with an ℓ-function that depends on the values of all locations in the spatial domain (*e.g.*, the maximum, minimum or average of all values in space), but observations are often available only at a finite number of irregularly spaced locations. Therefore, the ℓ-function cannot be evaluated based on the observations, and a different ℓ-function (and therefore a different criterion for characterizing extreme events) must be used on the observed data. My contributions to this modeling approach using ℓ-Pareto processes are twofold: first, the statistical inference for the generalized Pareto process associated to the flexible and well-known class of extremal-t processes (Thibaud and Opitz, 2015); second, the development of techniques for semi-parametric resampling of extreme events, even of yet unobserved magnitude, under min-

imal and nonparametric assumptions on the dependence structure (Palacios-Rodriguez et al., 2020; Opitz et al., 2021). In earlier work (during my PhD thesis), I developed pairwise statistical inference for spatial extremes based on the spectral measures S for different aggregation functionals ℓ in bivariate distribution (Opitz et al., 2015).

2.5.2 Limitations of asymptotic models

The strong asymptotic justification of max-stable and generalized Pareto models can be both a blessing and a curse. Max-stability is an expression of theoretical elegance and simplicity, and it provides a robust modeling framework when few extreme data are available. However, this strong assumption may be far from satisfied at subasymptotic levels arising with finite samples. An instructive example is asymptotic independence, where the limiting max-stable distribution is the product of independent margins and cannot capture the potentially strong dependence that remains at extreme subasymptotic levels. Very extreme joint risks tend to be strongly overestimated by asymptotic models if the data exhibit decreasing dependence strength at more extreme levels. Because asymptotic models such as generalized ℓ-Pareto processes are characterized by POT-stability, they may be too rigid in practice. The absence of such stability implies that the spatial or spatiotemporal extent of extreme episodes depends on the overall event magnitude, *e.g.*, expressed through the value of the aggregation functional ℓ, and more realistic models should make allowance for magnitude-dependent formulations of extremal dependence. It is often ambiguous whether data should be modeled using an asymptotically dependent or asymptotically independent distribution, and most families of distributions and processes only exhibit one type of dependence. This implies that asymptotic models may be too restrictive in practice, such that more flexible subasymptotic models can be useful in practice.

2.5.3 Towards increased flexibility through subasymptotic modeling

Because of practical limitations of asymptotic models such as max-stable or Pareto processes, especially their restriction to asymptotic dependence, it is natural to seek subasymptotic models for spatial extremes, which combine tail flexibility with computational tractability and have known tail characteristics. In the case of asymptotic independence, Gaussian dependence models might be reasonable and computationally convenient but they may lack flexibility in the joint tails. Alternatively, Wadsworth and Tawn (2012) proposed inverted max-stable models, but they are as difficult to fit as max-stable models. In the case of asymptotic dependence, subasymptotic models were also developed (*e.g.*, Wadsworth and Tawn, 2012; Krupskii et al., 2018) with the aim of being more flexible than max-stable processes and making inference easier. A number of other subasymptotic models have been proposed in recent years, and some of them able to accommodate both asymptotic dependence and asymptotic independence have been developed in the work presented in this manuscript, see Chapters 4 and 6.

Achieving a gain in flexibility for extremal dependence modeling, combined with flexible representations of marginal distributions, has therefore become one of the main purposes of my work. It has led me to propose more flexible model classes, and to develop related asymptotic theory and statistical methods. One can mention the bivariate random scale constructions in §4.2 and GAM-based techniques for estimating extremal dependence conditional to covariates in §4.1, the Gaussian scale mixture processes in §6.1, max-infinitely divisible processes extending the

class of max-stable processes in §6.2, and hierarchical constructions in §6.3. While approaches to spatial modeling of extremes (*i.e.*, without considering temporal dependence in extremes) have received a lot of attention in the 21st century, their spatiotemporal modeling is yet in its infancy. The models and methods presented in this manuscript contribute to fill this blank space and aim to help lay the foundation for off-the-shelf implementations of space-time extreme value analysis.

Chapter 3

Flexible modeling of univariate extremes

Throughout this chapter on new approaches to statistical modeling of univariate extremes, we focus on the tail behavior of data using the setting of threshold exceedances, and we do not tackle explicit modeling of extremal dependence. Univariate EVT suggests using the asymptotically arising generalized Pareto distribution (GPD), as recalled in Equation (2.8). We here present two conceptually very different solutions to improve flexibility by relaxing the default assumption that exceedance data correspond to an i.i.d. sample of the GPD. In § 3.1, we explain how the generalized additive modeling (GAM) framework can be used to model threshold exceedances and their occurrence conditionally to observed covariates or to random effects. When random effects are included, then Bayesian estimation using INLA is developed, and the penalized complexity (PC) prior of the tail index is derived. In § 3.2, a distributional extension of the GPD is developed by proposing a class of models that can be represented as ratios of two random variables, while keeping the GPD as a special case. This approach may provide better goodness-of-fit in *subasymptotic modeling* when convergence to the asymptotic GPD is not yet attained in data, especially in heavy-tailed data, and it may allow using a larger fraction of the data for the tail model. Moreover, the ratio representation is a good entry point to develop latent process modeling of spatial and spatiotemporal extremes, as shown later in § 6.3.

3.1 Three-stage GAMs for threshold exceedances

3.1.1 Three-stage structure

We here present a general modeling strategy that has been proposed and applied in Opitz et al. (2018); Castro-Camilo et al. (2020). In order to construct a model based on threshold exceedances that allows us to fully characterize nonstationary univariate tail behavior, we use the following three components:

1. a threshold; it may have to be chosen nonstationary, *i.e.*, conditionally to predictor variables, to ensure that it is not too low (where the asymptotic regime would not be valid) and not too high (where too few or no exceedances would occur);

2. the threshold exceedance probability above the threshold in component 1, which may also be nonstationary and depend on predictor variables;

3. a probability distribution for the positive threshold excesses above the threshold in component 1, which may also be nonstationary.

Our general strategy for modeling trends in high quantiles can be decomposed into three stages corresponding to the above components, each consisting of a suitable univariate response distribution combined with a generalized additive regression equation capturing systematic variation with covariates or random effects. An early GAM framework for maxima data using the generalized extreme-value distribution (GEVD) was proposed by Chavez-Demoulin and Davison (2005) where smoothing splines are incorporated into the parameters, and a POT-based approach with similarities to the three-stage procedure outlined below was proposed by Northrop and Jonathan (2011) and Youngman (2019).

We use the following three-stage model. After fitting a full distribution in Stage 1 to all data, a high probability p is fixed to determine a nonstationary threshold for Stages 2 and 3 by setting it to the p-quantile of the distribution fitted in Stage 1. The vector of one or several covariates is denoted by $\boldsymbol{z}$, and it may include components that refer to spatial location or time.

1. **Stage 1 (Full data distribution):** This stage is "context-dependent", in the sense that one chooses a combination of response distribution and link function that is deemed appropriate for the data (bulk and extremes), and is amenable to GAM inference (*e.g.*, Gaussian, Gamma...). A high nonstationary threshold $u(\boldsymbol{z})$ is then chosen as the p-quantile of the fitted response distribution.
   An alternative procedure for Stage 1 would be to directly use quantile regression at the level p without explicitly specifying a response distribution, but it may come along with high estimation uncertainty if p is relatively high, especially with complex predictor structures.

2. **Stage 2 (Binary logistic regression):** Using the threshold $u(\boldsymbol{z})$ obtained in Stage 1, exceedance indicators are modeled as Bernoulli random variables, *i.e.*,

$$1(Y(\boldsymbol{z}) > u(\boldsymbol{z})\} \sim \text{Ber}\{p_u(\boldsymbol{z})\}. \tag{3.1}$$

   In case of multiple observations for the same predictor configuration $\boldsymbol{z}$, the Bernoulli distribution can be replaced with the corresponding binomial distribution.

3. **Stage 3 (GPD regression):** Using the threshold $u(\boldsymbol{z})$ obtained in Stage 1, positive threshold exceedances $Y_u^+(\boldsymbol{z}) = Y(\boldsymbol{z}) - u(\boldsymbol{z}) \mid Y(\boldsymbol{z}) > u(\boldsymbol{z})$ are extracted from the sample and are modelled through the GPD, whose parameters $\xi(\boldsymbol{z})$ and $\sigma_{\text{GPD}}(\boldsymbol{z})$ may depend on the covariate vector $\boldsymbol{z}$.

For reasons of modeling and computational complexity, we use the working assumption that the data are conditionally independent with respect to the trend surfaces. Dependence that arises between observations conditionally to observed covariates can be captured to some extent by including appropriately defined random effects at the layer of the linear predictors. For similar reasons, the tail index ξ is often chosen to be stationary in practice; in this case, it is estimated but it does not depend on covariates or random effects.

If Stage 1 involves a response distribution for which a nonstationary scaling parameter $\mu(\boldsymbol{z})$ is estimated, it may be sensible to include it as a deterministic offset in the GPD scale, *i.e.*, to include $\log(\mu(\boldsymbol{z}))$ in the additive predictor of $\log(\sigma(\boldsymbol{z}))$. Since the nonstationary patterns

in the bulk and the tail of the distribution could be quite similar to each other, this allows borrowing strength from the bulk for more accurate tail estimation. Especially in a Bayesian framework, the estimation may be simplified since the Gaussian prior distributions of additive components in the GPD predictor can be chosen to be relatively smooth and informative, *i.e.*, strongly concentrated around zero. This idea of propagating the scale information from the bulk model to the tail model echos the fundamental question if extreme-value models should only serve to correct deficiencies of models fitted to the full distribution. That is, should the extreme-value model be fitted to some variant of normalized residuals of bulk models? Or is the full distribution not of any interest, in which case relatively sophisticated models are needed to capture only the tail of the distribution in its full complexity?

3.1.2 Bayesian inference of space-time trends with INLA

Fully Bayesian modeling approaches for spatial and/or temporal extremes often rely on latent processes embedded into the GPD parameters to capture trends and dependence. In particular, Gaussian processes can be used to capture spatial dependence and covariate-driven trends in data such as precipitation; *e.g.*, Cooley et al. (2007) take advantage of simulation-based Markov chain Monte-Carlo (MCMC) methods for the estimation of posterior distributions. Here, we adopt a similar model structure based on using Gaussian random effects in the predictors of the three-stage specification in §3.1. We develop efficient statistical inference geared towards estimating space-time random effects by exploiting the speed and accuracy of INLA to approximate posterior distributions. A good choice of prior distributions enables appropriate smoothing of predicted quantiles, which is especially important when predicting rare events. In this setting, we can predict extreme conditional quantiles by combining the posterior mean predictions from the binomial and GPD stages of the three-stage model.

Gaussian process priors can be deployed to capture systematic temporal and spatial trends through semi-parametrically specified random effects in the three stages. We point out that Gaussian dependence is usually not well adapted to extremes since it is often not strong enough in the tails, but the use of latent Gaussian models, *i.e.*, of Gaussian random effect components, still makes sense for capturing nonlinear trends of space, season/time and other covariates.

The dependence between estimators of the two GPD parameters (scale and shape) in Stage 3 may cause unstable inferences with techniques such as INLA or MCMC. Therefore, to avoid confounding problems, we propose to reparametrize the GPD using the tail index ξ and a specific q-quantile κ_q for some fixed probability of interest $q \in (0,1)$, *i.e.*,

$$\mathrm{GPD}(y;\kappa_q,\xi) = \begin{cases} 1-\left[1+\{(1-q)^{-\xi}-1\}y/\kappa_q\right]_+^{-1/\xi}, & \xi \neq 0, \\ 1-(1-q)^{y/\kappa_q}, & \xi = 0, \end{cases} \quad y>0; \tag{3.2}$$

The overall α-quantile $y_\alpha(\boldsymbol{z})$ of the data, for $1-p_u(\boldsymbol{z}) < \alpha < 1$, is then obtained as

$$\begin{aligned} y_\alpha(\boldsymbol{z}) &= u(\boldsymbol{z}) + \mathrm{GPD}^{-1}\{1-(1-\alpha)/p_u(\boldsymbol{z});\kappa_q(\boldsymbol{z}),\xi(\boldsymbol{z})\} \\ &= \begin{cases} u(\boldsymbol{z}) + \kappa_q(\boldsymbol{z})\left[\{(1-\alpha)/p_u(\boldsymbol{z})\}^{-\xi(\boldsymbol{z})}-1\right]/\left\{(1-q)^{-\xi(\boldsymbol{z})}-1\right\}, & \xi(\boldsymbol{z}) \neq 0, \\ u(\boldsymbol{z}) + \kappa_q(\boldsymbol{z})\log\{(1-\alpha)/p_u(\boldsymbol{z})\}/\log(1-q), & \xi(\boldsymbol{z}) = 0, \end{cases}, \end{aligned}$$

where GPD^{-1} denotes the quantile function of the GPD in (3.2).

The additive predictor for INLA or for a frequentist GAM-implementation can be specified for $\log \kappa_q$, and choices such as the median (*i.e.*, setting $q = 0.5$) may be expedient in practice. Moreover, for Bayesian modeling we restrict attention to $\xi \geq 0$ and thus exclude the very light-tailed case $\xi < 0$ with finite and potentially nonstationary upper endpoint of the support of the distribution, which may pose problems for likelihood-based modeling, and in particular for accurate Laplace approximations with INLA.

Penalized complexity prior of the tail index

We derive the penalized complexity (PC) prior distribution (see §1.3.5) for the tail index in the GPD, which provides a principled prior choice (Simpson et al., 2017) for this crucial model parameter. We penalize the distance to a baseline model possessing light exponentially-decaying tails, *i.e.*, by using $\xi = 0$ as baseline. Moreover, when $\xi \geq 1$ the tail is so heavy that the mean does not even exist, and when $\xi \geq 1/2$, the variance is infinite. Because of this, too large values of ξ are unrealistic for many data types. It is therefore important to choose a suitable prior distribution for ξ that giv priority to light and moderately heavy tails while properly downweighting unrealistically heavy tails. Exponential tails usually make sense for environmental data, and the heavier the tail, the stronger the penalty.

Let $f_\xi(y)$ denote the GP density and $f_{\xi_0}(y)$ denote the exponential density. We obtain the following formula for the Kullback-Leibler divergence with respect to the baseline model:

$$\mathrm{KLD}(f_\xi || f_{\xi_0}) = \frac{\xi^2}{1-\xi}, \qquad 0 \leq \xi < 1. \tag{3.3}$$

To define a PC prior $\pi(\xi)$ for the tail index ξ, we assume that f_ξ is penalized at constant rate in terms of its "distance" $d(f_\xi, f_{\xi_0}) = \sqrt{2\,\mathrm{KLD}(f_\xi || f_{\xi_0})}$ to the reference f_{ξ_0}, therefore involving an exponential prior distribution for $d(f_\xi, f_{\xi_0})$. Because the KLD (3.3) converges to infinity as $\xi \to 1$, such a prior will put zero mass on $\xi \geq 1$, hence preventing infinite-mean models to occur. We here propose two possible prior choices, which are based (i) directly on Equation (3.3), or (ii) on an approximation of (3.3) as $\xi \to 0$:

(i) exact PC prior:

$$\begin{aligned} \pi(\xi) &= \lambda \exp(-\lambda d(f_\xi, f_{\xi_0})) \left| \frac{\partial d(f_\xi, f_{\xi_0})}{\partial \xi} \right| = \sqrt{2}\lambda \exp\left\{ -\sqrt{2}\lambda \frac{\xi}{(1-\xi)^{1/2}} \right\} \left\{ \frac{1-\xi/2}{(1-\xi)^{3/2}} \right\} \\ &= \tilde{\lambda} \exp\left\{ -\tilde{\lambda} \frac{\xi}{(1-\xi)^{1/2}} \right\} \left\{ \frac{1-\xi/2}{(1-\xi)^{3/2}} \right\}, \quad 0 \leq \xi < 1, \end{aligned} \tag{3.4}$$

where the penalization rate parameter is $\lambda = \tilde{\lambda}/\sqrt{2} > 0$;

(ii) approximate PC prior obtained by replacing (3.3) with the first-order approximation ξ^2 near 0, leading to an exponential distribution with rate $\tilde{\lambda} = \sqrt{2}\lambda > 0$:

$$\pi(\xi) = \sqrt{2}\lambda \exp(-\sqrt{2}\lambda\xi) = \tilde{\lambda} \exp(-\tilde{\lambda}\xi), \quad \xi \geq 0; \tag{3.5}$$

see Figure 3.1 for an illustration, which shows a good approximation of the exact PC prior (i) by (ii), especially close to $\xi = 0$.

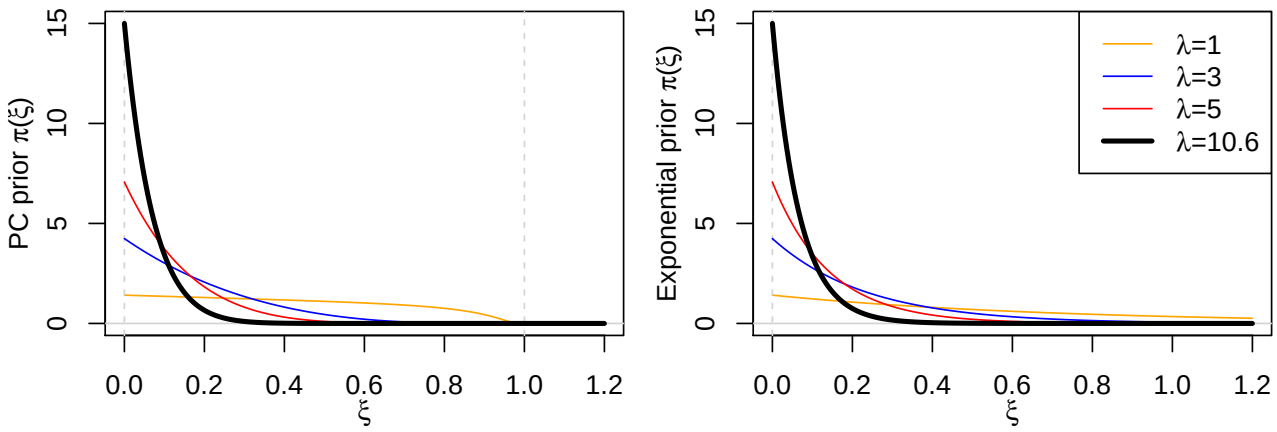


Figure 3.1: PC priors for the GPD tail index ξ using the exact formula (left) or an approximation (right).

Application to predicting extreme spatiotemporal precipitation quantiles

The following work was motivated by the Data Challenge of the 2017 conference on Extreme-Value Analysis (EVA2017, Delft, Netherlands), with the goal of predicting monthly quantiles at the 99.8% level in daily precipitation data for observed and unobserved locations. Data have been measured at 40 stations during the period 1972–2016 and have been divided in a training set (1972–1995) and a validation set (1996–2016), with only some of the stations active during all of the training period.

In Stage 1 of the model, we fitted a gamma regression to the positive precipitation intensities, with the goal to set a high space-time threshold $u(s,t)$ chosen as the quantile for a probability p_+. In the regression equations of all three stages, we separately included an intercept term, a spatial random effect with Matérn covariance, and a weekly-indexed seasonal random effect with cyclic second-order random walk prior. The computational efficiency of INLA allowed us to conduct an extensive cross-validation study for selecting certain model parameters that are crucial for prediction (precisions of latent Gaussian effects, spatial range of Matérn effect, exceedance probability p_+ to set the threshold in Stage 1). The model is misspecified here because it does not account for spatiotemporal dependence within precipitation episodes, such that we cannot expect a good automatic fit of smoothing parameters through posterior distributions. Instead, we used the cross-validated prediction score of the data challenge (quantile score) to compare a grid of values for the smoothing parameters, and we choose the best performing configuration. We devise a stratified cross-validation scheme by holding out complete data either for specific years or for specific stations to appropriately cope with the twin goals of temporal and spatial prediction, respectively.

Figure 3.2 illustrates the resulting INLA-based posterior means and credible intervals of the spatial and seasonal effects for the three stages. A moderately high threshold of $p_+ = 0.92$ is identified for the tail model, and in particular the re-estimated exceedance probability model in Stage 2 helps identify nonstationarity that is different between the tail and the bulk of the distribution. The tail index is estimated at 0.34 with 95% credible interval of $(0.31, 0.38)$,

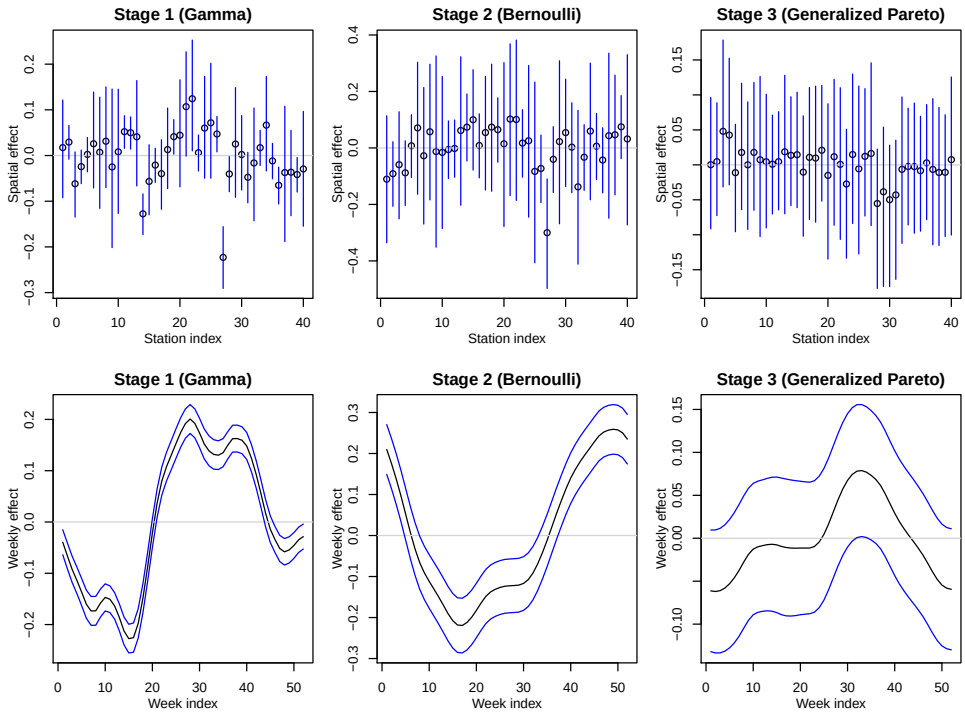


Figure 3.2: INLA fit of random effects in three-stage GAM for Dutch precipitation extremes. Top row: spatial random effect. Bottom row: seasonal random effect.

implying moderately heavy tails, substantially heavier than the gamma tails in Stage 1.

3.1.3 Application to spatiotemporal gap filling for inference on extreme hotspots

In the paper Castro-Camilo et al. (2020), already highlighted in §1.4.2, we describe an approach implemented for the data challenge of the Extreme-Value Analysis conference 2019 in Zagreb. The dataset consists of anomalies of gridded daily Red Sea surface temperatures, where some space-time-contiguous areas of the dataset have been masked artificially. The goal is to predict a summary variables of space-time clusters, defined as the minimum over space-time cylinders, with a focus on extreme episodes. For that purpose, we develop a method for probabilistic prediction of extreme value hotspots in a spatiotemporal framework, tailored to big datasets containing important gaps. In this setting, direct calculation of summaries from data, such as the minimum over a space-time domain, is not possible due to missing data.

To obtain predictive distributions for such cluster summaries, we propose a two-step approach. First, we model marginal distributions with a focus on accurate modeling of the upper tail by using the three-stage GAM outlined above. Then, after transforming the data to a standard Gaussian scale based on the marginal model, we estimate a Gaussian space-time dependence model defined locally in the time domain for the space-time subregions where we want to predict. As an alternative to GAMs for marginal modeling, we also studied the approach of pooling data locally in space using nearest-neighbor techniques and then estimating distribution

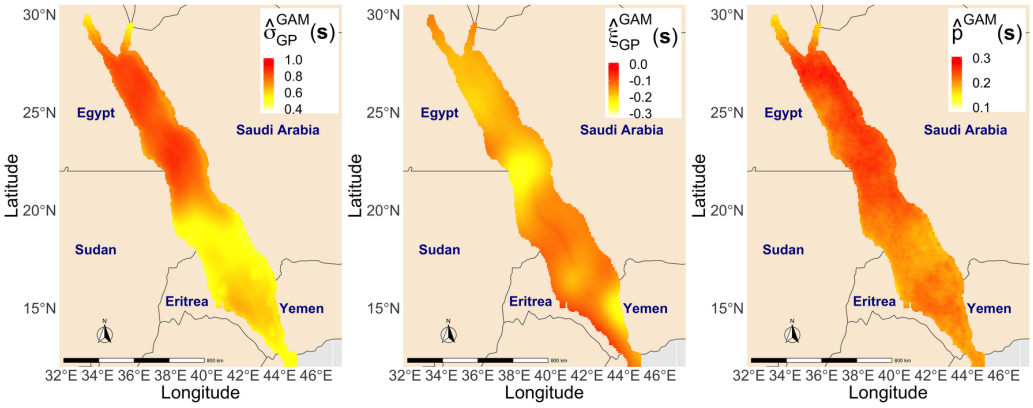


Figure 3.3: GAM-based spatially varying tail parameter estimates for Red Sea surface temperature anomalies based on the three-stage approach. Left display: GPD scale parameter. Middle display: GPD shape parameter. Right display: exceedance probability.

parameters according to the three stages separately for each pooled dataset. To cope with the high space-time resolution of data consisting of 11315 days and 16703 spatial pixels, the local Gaussian models use the Markov representation of the Matérn correlation function based on the SPDE approach. They are fitted in a Bayesian framework through INLA. Finally, we generate posterior samples of the INLA model, backtransform them to the original marginal scale of the data, and then use them to provide statistical inferences through Monte-Carlo estimation, such as predictive distributions of cluster summaries.

For illustration, Figure 3.3 shows the GAM-based estimates of the GPD scale, the GPD shape and the exceedance probability, where the threshold is fixed as a high quantile of a Gaussian GAM for the mean of the Red Sea temperature anomalies.

3.1.4 Possible extensions of multi-stage GAMs for extremes

In future work, it would be interesting to replace the generalized Pareto distribution by more flexible sub-asymptotic response distributions (*e.g.*, Papastathopoulos and Tawn, 2013; Naveau et al., 2016, or the class of models described in the following §3.2) amenable to GAM-based inference. More flexible models could potentially be applied at much lower thresholds, thus avoiding that the non-continuous behavior around the threshold arising at relatively high quantiles. The GAMs, or the mixed GAMs when using INLA, are not well-suited to properly capture space-time dependence. As long as the primary objective is to estimate high *marginal* quantiles, the exact characterization of the dependence structure is a secondary issue, or even perhaps a nuisance in practice. Two-step approaches, such as the one adopted in Castro-Camilo et al. (2020), can be a useful practical solution to jointly model complex marginal behavior and spatiotemporal dependence, especially when prediction over space and time is the goal. Moreover, when the GAM does not appropriately capture the spatiotemporal dependence patterns, then some care is required with respect to assessment of the estimation uncertainty. More accurate standard errors can be obtained by using block bootstrap techniques with re-estimation of parameters for each bootstrap sample (Carlstein, 1986).

3.2 Ratio constructions extending the GPD

In Yadav et al. (2020), we develop new flexible univariate models for light-tailed and heavy-tailed data, which extend a hierarchical representation of the GPD limit for threshold exceedances. These models can accommodate departure from asymptotic threshold stability in finite samples while keeping the asymptotic GPD as a special case, and they can capture the tails and the bulk jointly without losing much flexibility. For the Bayesian setting, we further design PC priors to shrink the model towards a simpler reference given by the GPD.

It is often difficult to find a good, natural and interpretable threshold, and parameter estimates are often sensitive to this choice (Scarrott and MacDonald, 2012). This has motivated the development of *subasymptotic* models for univariate extremes, which are more flexible than the asymptotic GPD at finite levels while keeping a GPD-like behavior in the tail; see, *e.g.*, Frigessi et al. (2003); Carreau and Bengio (2009); Papastathopoulos and Tawn (2013); Naveau et al. (2016), among others. Ideally, subasymptotic models allows us to describe also the distribution of moderate and low values through an appropriate parametrization, which partly or fully separates control over bulk and tail properties.

In our approach, we extend the characterization of the GPD as an exponential mixture with rate parameter following a gamma distribution. Let $\text{Gamma}(\beta, \alpha)$ denote the gamma distribution with shape $\beta > 0$ and rate $\alpha > 0$, then

$$\left.\begin{array}{rl} Y \mid \Lambda & \sim \text{Exp}(\Lambda) \\ \Lambda & \sim \Gamma(\alpha, \beta) \end{array}\right\} \Rightarrow Y \sim \text{GPD}(\alpha/\beta, 1/\beta); \tag{3.6}$$

see Reiss and Thomas (2007). In other words, exponentially-decaying tails become heavier (and of power-law type) by making their rate parameter Λ random. We propose new tail models constructed as in (3.6), but we modify the top and/or lower levels of the hierarchy in order to gain in flexibility, while keeping the GP distribution with $\xi \geq 0$ as a special case. For the Bayesian setting, we derive the PC priors with respect to the GPD baseline model to allow for control over the departure from the asymptotic GPD. We mainly focus on a parsimonious extension of (3.6), which is based on a gamma distribution in both levels of the hierarchy. It can potentially generate a wide variety of new models with light and heavy tails and various behaviors in the bulk. While this section focuses on the univariate construction of this model, its distributional properties and the PC prior, we describe a statistical application in the context of Bayesian hierarchical modeling in § 6.3.

3.2.1 Univariate tail properties in rate mixture constructions

We can construct new rate mixture models for observations Y following a probability distribution F as follows. We consider a family of distributions $F(\,\cdot\,;\lambda)$ with rate parameter λ and having support on $[0, \infty)$, and an independent latent random variable $\Lambda \geq 0$, such that $Y \mid \Lambda \sim F(\cdot;\Lambda)$. Equivalently, we have the following ratio representation, which is useful for simulation and inference:

$$Y \mid \Lambda \overset{d}{=} \frac{\tilde{Y}}{\Lambda}, \quad \text{with} \quad \Lambda \geq 0 \perp \tilde{Y} \geq 0, \quad \tilde{Y} \sim F(\,\cdot\,;1), \tag{3.7}$$

The unconditional upper tail behavior of Y is determined by the interplay between the upper tail of $\tilde{Y}$ and the lower tail of Λ, *i.e.*, the upper tail of $1/\Lambda$. We shortly discuss two particularly

interesting scenarios; recall §2.2.3 for the discussion of various tail classes.

In the first scenario, we assume that $1/\Lambda$ in (3.7) has power-law tail decay, *i.e.*, its distribution is regularly varying with index $-a < 0$. If the distribution $F_Y(\cdot; 1)$ in (3.7) has a lighter upper tail than that of $1/\Lambda$, such that $\mathbb{E}(\tilde{Y}^{a+\varepsilon}) < \infty$ for some $\varepsilon > 0$ with $\tilde{Y} \sim F_Y(\,\cdot\,; 1)$, then Breiman's Lemma 1 (Breiman, 1965) implies that

$$1 - F(y) = \Pr(Y > y) \sim \mathbb{E}(\tilde{Y}^a)\, \Pr(1/\Lambda > y), \quad y \to \infty. \tag{3.8}$$

The heavier-tailed random factor $1/\Lambda$ in (3.7) dominates the tail behavior of Y, while the lighter-tailed random factor $\tilde{Y}$ contributes to extreme survival probabilities only through a scaling factor.

In the second scenario, we assume that both $\tilde{Y}$ and $1/\Lambda$ in (3.7) have tails of Weibull type with Weibull indices $\tilde{\eta}$ and η_Λ, respectively. Then, the variable Y constructed as in (3.7) also has a tail of Weibull type. In particular, its Weibull index is $\eta_Y = (\tilde{\eta}\eta_\Lambda)/(\tilde{\eta} + \eta_\Lambda) < \min(\tilde{\eta}, \eta_\Lambda)$, such that the tail of Y always has a slower decay rate than that of each of the two random factors $\tilde{Y}$ and $1/\Lambda$; see Arendarczyk and Debicki (2011).

3.2.2 Gamma-gamma model

Specifying the gamma distribution for $F(\cdot; 1)$ in (3.6) yields the hierarchical gamma-gamma model, which may be written as

$$Y \mid \Lambda \sim \Gamma(\beta_1, \Lambda), \quad \Lambda \sim \Gamma(\beta_2, \alpha), \quad \alpha, \beta_1, \beta_2 > 0. \tag{3.9}$$

The model (3.6) simplifies to the GPD obtained in (3.6) when $\beta_1 = 1$. The distribution of Y corresponds to a rescaled F_{ν_1,ν_2} distribution with degrees of freedom $\nu_1 = 2\beta_1$ and $\nu_2 = 2\beta_2$, and scaling factor $\alpha\beta_1/\beta_2$, such that $Y \overset{D}{=} (\alpha\beta_1/\beta_2)Z$, with $Z \sim F_{2\beta_1,2\beta_2}$. The gamma-gamma model has a heavy power-law tail with tail index $\xi = 1/\beta_2$.

3.2.3 Model extension with Weibull-type tail behavior

For data with a light upper tail and tail index equal to zero, we propose a model extension that keeps the heavy-tailed GPD on the boundary of the parameter space as follows:

$$\begin{aligned} Y^{1/k} \mid \Lambda &\sim \Gamma(\beta_1, \Lambda), \quad k, \beta_1 > 0, \\ \Lambda &\sim \mathrm{GIG}(\alpha/2, b, \beta_2), \quad (\alpha, b, \beta_2) \in D_{\mathrm{GIG}}, \end{aligned} \tag{3.10}$$

where the latent rate parameter Λ is assumed to follow the generalized inverse Gaussian (GIG) distribution with parameters $\alpha/2$, b and β_2, and where D_{GIG} denotes its parameter space. For the $\mathrm{GIG}(a, b, \beta)$-distribution, it is given by $a, b \geq 0$ and as follows for β: $\beta \in \mathbb{R}$ if $a, b > 0$; $\beta > 0$ if $b = 0, a > 0$; $\beta < 0$ if $a = 0, b \geq 0$. The GIG distribution has an exponentially decaying tail (i.e., Weibull-type tail with Weibull index one).

This model generalizes the gamma-gamma construction, which is on the boundary of the parameter space with $b = 0$, $k = 1$ and $\beta_2 > 0$. It captures a wide range of tail behaviors, from very light tails to relatively heavy tails. Specifically, when $b > 0$, we deduce that Y has Weibull index $\eta_Y = (1/k^2)/(2/k) = 1/(2k) > 0$, such that the model can capture Weibull tails with any Weibull index.

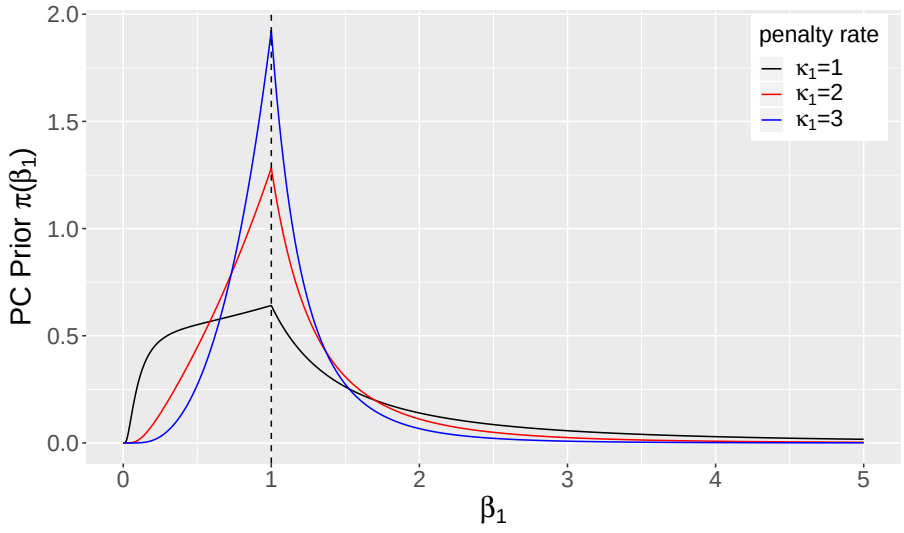


Figure 3.4: PC prior density of $\beta_1 > 0$ in the gamma-gamma ratio model for penalty rates $\kappa_1 = 1, 2, 3$ (blue, red and black curves, respectively).

3.2.4 PC priors of the shape parameters in the gamma-gamma model

We consider the GPD as the baseline distribution for setting PC priors in the gamma-gamma model. The PC priors derived for the tail index parameter ξ of the GPD in §3.1.2 can be straightforwardly adapted to provide a PC prior for $\beta_2 = 1/\xi$. However, special care is required for the shape parameter β_1 of the gamma distribution in the numerator, which represents the "distance" to the GP sub-model with $\beta_1 = 1$. Writing $l(\beta_1) = \sqrt{2\text{KLD}\left\{\gamma(\cdot;\lambda,\beta_1)\|\gamma(\cdot;\lambda,1)\right\}}$ for the Kullback-Leibler divergence of the gamma densities in the baseline and the general model, we calculate the following PC prior for $\beta_1 > 0$:

$$\pi(\beta_1) = \frac{\kappa}{2}\exp\{-\kappa\sqrt{2(\beta_1-1)\psi(\beta_1) - 2\log\{\Gamma(\beta_1)\}}\}\left|\frac{(\beta_1-1)\psi'(\beta_1)}{\sqrt{2(\beta_1-1)\psi(\beta_1) - 2\log\{\Gamma(\beta_1)\}}}\right|$$

with $\kappa > 0$ the a predetermined penalty rate. An illustration is given in Figure 3.4, where a clear peak at the parameter value of the baseline GPD model becomes visible.

Chapter 4

Flexible modeling of bivariate extremes

While Chapter 1 focused on univariate tails, we now turn to the modeling of the dependence of extremes, viewed principally from the most fundamental configuration – a bivariate random vector – in this Chapter 4. This chapter maintains some analogy with the preceding chapter. At the beginning in §4.1, we develop generalized additive models (GAMs) to flexibly capture nonlinear variation in extremal dependence conditional to covariates for bivariate and multivariate random vectors. The second §4.2 provides a catalog of very general and unifying results characterizing extremal dependence in bivariate random scale constructions, which are ubiquitous in the literature on flexible multivariate and spatial tail modeling, including representations of multivariate POT limit distributions. In this chapter, we work with standard representations of extremal dependence based on normalized marginal distributions as introduced in §2.3, such that extremal dependence can be interpreted by ignoring marginal behavior; see also the extremal dependence summaries in §2.3.1.

4.1 GAMs for bivariate extremal dependence

The probability and structure of co-occurrences of extreme values in multivariate data may critically depend on auxiliary information provided by covariates. We develop a flexible GAM framework based on high threshold exceedances for estimating covariate-dependent joint tail characteristics for both regimes of asymptotic dependence and asymptotic independence. The framework is based on suitably defined marginal pretransformations and projections of the random vector along the directions of the unit simplex, which lead to convenient univariate representations of multivariate exceedances based on the exponential distribution. This framework allows combining GAM modeling both for the univariate margins and for bivariate dependence.

In §4.1.1, we recall asymptotic joint tail representations based on exponential marginal distributions, and we derive how exponential distributions arise for appropriately defined univariate projections of the random vector in §4.1.2.

4.1.1 Representations of multivariate extremal dependence

For studying upper tail dependence, normalized marginal distributions are useful; recall (2.13) where we define superscripts X^P and X^F to refer to marginal probability integral transform that

establish standard Pareto or unit Fréchet distributions, respectively, in the case of continuous marginal distributions. Exponential marginal distributions are particularly convenient here since they are available in all standard implementations of generalized linear or additive models, and of survival models where part of their distribution may be censored. We therefore work with the following marginal transformation to a standard exponential scale, which is based on a log-transformation of the standard Pareto representation:

$$X^E = \log X^P.$$

If we start from the unit Fréchet representation, we can use an inverse transformation to obtain an *inverted standard exponential scale*:

$$X^{E\downarrow} = 1/X^F,$$

i.e., we switch the upper and the lower tail of the distribution to obtain standard exponential margins.

Consider a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$, whose joint extremal behavior is characterized by the exponent function $V^\star(\boldsymbol{z})$ as defined in §2.3. We here use a transformation $L(\boldsymbol{z}) = V^\star(1/\boldsymbol{z})$ of the exponent function with inverted arguments, known as *stable tail dependence function*. The multivariate regular variation limit (2.14) of classical multivariate limit theory implies

$$1 - \Pr(\boldsymbol{X}^{E\downarrow} \geq \boldsymbol{x}) \sim 1 - \exp(-L(\boldsymbol{x})) = 1 - G^\star(1/\boldsymbol{x}), \quad \max_{j=1}^{d} x_j \to 0, \tag{4.1}$$

where G is a max-stable distribution with unit Fréchet marginal distribution. The stable tail dependence function can be expressed through *Pickands' dependence function* A as

$$L(\boldsymbol{x}) = (x_1 + \ldots + x_d)\, A\left(\frac{x_d}{x_1 + \ldots x_d}, \ldots, \frac{x_1}{x_1 + \ldots + x_d}\right) \tag{4.2}$$

where A is defined for any *direction* (or *weight vector*) $\boldsymbol{\omega}$ on the unit simplex $\mathbb{S}_d$:

$$\boldsymbol{\omega} \in \mathbb{S}_d = \left\{ (\omega_1, \ldots, \omega_d) \geq \mathbf{0} \mid \sum_{j=1}^{d} \omega_j = 1 \right\}. \tag{4.3}$$

Without loss of generality, we can drop $\omega_d = 1 - \sum_{j=1}^{d-1} \omega_j$ and write $A(\omega_1, \ldots, w_{d-1})$, such that A is a univariate function when $d = 2$. The function A must satisfy certain shape constraints to be valid, among them $\max_{j=1}^{d} \omega_j \leq A(\boldsymbol{\omega}) \leq 1$, where the lower bound is attained for perfect asymptotic dependence. The above equations establish a useful *directional representation* of multivariate tails in the case of asymptotic dependence where A has a non-trivial form.

With asymptotic independence, $L(\boldsymbol{x}) = x_1 + \ldots + x_d$ has "trivial" form without parameters. For the case of asymptotic independence, we propose to exploit the following more general multivariate tail representation (Wadsworth et al., 2013):

$$\Pr\left(X_1^E > \omega_1 x, \ldots, X_d^E > \omega_d x\right) \sim \ell(\exp(x); \boldsymbol{\omega}) \exp(-\lambda(\boldsymbol{\omega})\, x), \quad x \to \infty, \tag{4.4}$$

with the *angular dependence function* $\lambda(\boldsymbol{\omega})$ and a slowly varying function $\ell(x; \boldsymbol{\omega})$. This representation generalizes the bivariate representation in (2.9), which considers only the direction along the diagonal, and where the bivariate coefficient of tail dependence is given by

$\eta_2 = 1/(2\lambda(1/2, 1/2))$. The following forms arise for the angular dependence function. In the case of asymptotic dependence, we have $\lambda(\boldsymbol{\omega}) = \max_{j=1}^{d} \omega_j$. With asymptotic independence, we obtain $\lambda(\boldsymbol{\omega}) > \max_{j=1}^{d} \omega_j$. Classical mutual independence of the components of $\boldsymbol{X}$ leads to $\lambda(\boldsymbol{\omega}) = 1$. If there were some form of "negative dependence" in the tail of $\boldsymbol{X}$ (*e.g.*, negative correlation), it is possible to obtain $\lambda(\boldsymbol{\omega}) > 1$; we do not consider this case in the following.

We concentrate our efforts on the construction of estimators of $A(\boldsymbol{\omega})$ and $\lambda(\boldsymbol{\omega})$ for fixed $\boldsymbol{\omega}$ but given additional auxiliary predictor information.

4.1.2 Univariate projections with exponential limits

Given the weight vector $\mathbf{0} \leq \boldsymbol{\omega} = (\omega_1, \ldots, \omega_d) \in \mathbb{S}_d$ and a vector $\boldsymbol{x}$, we define two types of univariate projections of $\boldsymbol{x}$ along $\boldsymbol{\omega}$. The *max-projection of* $\boldsymbol{x}$ along $\boldsymbol{\omega}$ is given as

$$\max\nolimits_{\boldsymbol{\omega}}(\boldsymbol{x}) = \max_{j=1}^{d} \omega_j x_j, \quad \boldsymbol{\omega} \in \mathbb{S}_d,$$

while the *min-projection of* $\boldsymbol{x}$ along $\boldsymbol{\omega}$ is given as

$$\min\nolimits_{\boldsymbol{\omega}}(\boldsymbol{x}) = \min_{j=1}^{d} x_j/\omega_j, \quad \boldsymbol{\omega} \in \mathbb{S}_d.$$

These projections are constructed to give little influence on the projection to components x_j with small weights ω_j. The two projections are linked by the following inversion transformation:

$$\max\nolimits_{\boldsymbol{\omega}}(\boldsymbol{x}) = 1/\min\nolimits_{\boldsymbol{\omega}}(1/\boldsymbol{x}).$$

Based on the tail representation (4.1) of $\boldsymbol{X}$ with Pickands' dependence function $A(\boldsymbol{\omega})$, which is useful mainly for asymptotic dependence, we obtain that the lower tail of $M_{\boldsymbol{\omega}}^{\min\downarrow} = \min_{\boldsymbol{\omega}}(X^{E\downarrow})$ is exponential with rate $A(\boldsymbol{\omega})$. In practice, we can model the lower tail through a right-censored exponential distribution. On the other hand, if $\boldsymbol{X}$ allows for the more general tail representation (4.4) under asymptotic independence with angular dependence function $\lambda(\boldsymbol{\omega})$, then exceedances of the min-projection $M_{\boldsymbol{\omega}}^{\min} = \min_{\boldsymbol{\omega}}(\boldsymbol{X}^E)$ satisfy

$$\Pr\left(M_{\boldsymbol{\omega}}^{\min} > x + u \mid M_{\boldsymbol{\omega}}^{\min} > u\right) \to \exp\left\{-x\lambda(\boldsymbol{\omega})\right\}, \quad u \to \infty, \quad x > 0.$$

In practice, such exceedances can be modeled with an exponential distribution with rate $\lambda(\boldsymbol{\omega})$. Figure 4.1 gives an empirical illustration of the two above projections based on a bivariate sample of the logistic extreme-value distribution.

4.1.3 GAM-based estimation

The exponential distribution and the right-censored exponential distribution, both of which arise asymptotically for the directional projections in §4.1.2, are standard response distributions in regression modeling. Here, generalized linear models may not give realistically flexible specifications of covariate influence on extremal dependence, such that we consider generalized additive modeling.

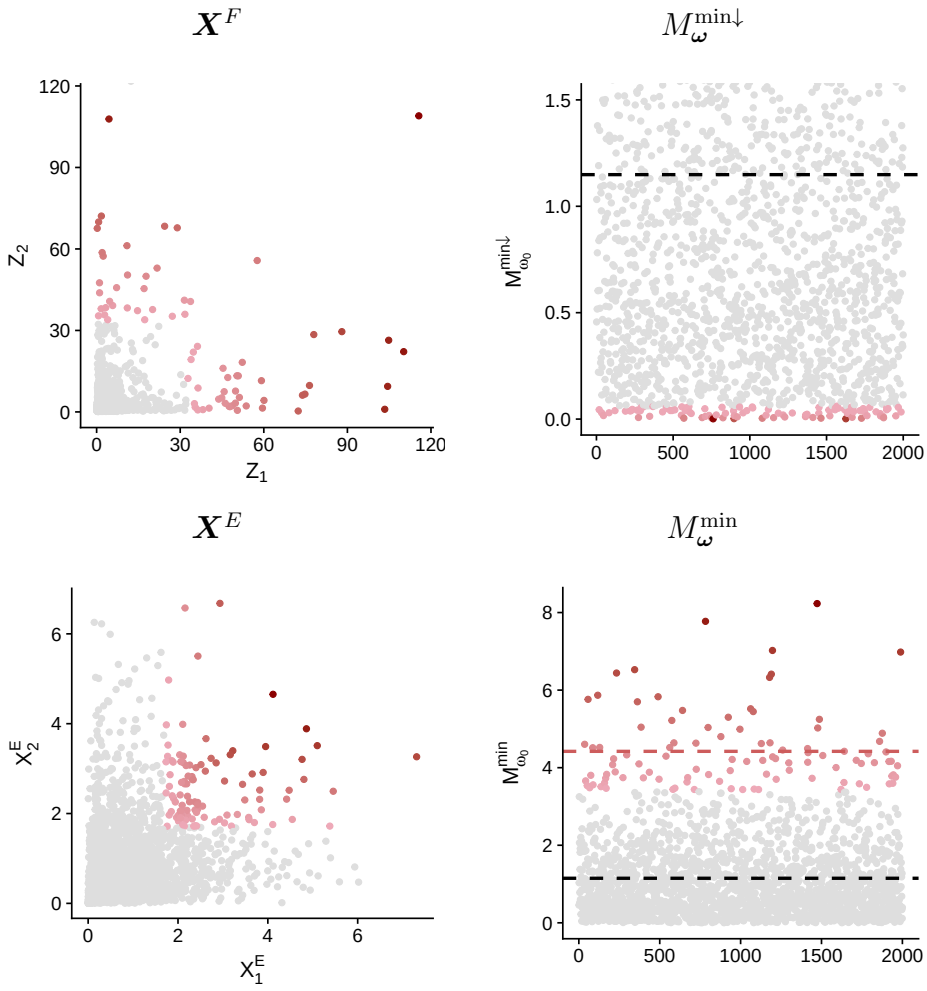


Figure 4.1: Illustration of univariate directional projections of multivariate extremes based on a sample of size 2000 from the bivariate logistic extreme-value distribution. Left column: normalized bivariate data sample. Right column: Projections along $\boldsymbol{\omega} = (1/2, 1/2)^T$. Black dashed lines in the displays on the right indicate the expected values of projections; the red dashed line in the lower right display indicates the expected value of exceedances.

Given observation vectors $\boldsymbol{X}_i$, $i = 1, \ldots, n$, each with a covariate vector $\boldsymbol{z}_i$, we first transform margins to the exponential or inverted exponential scale. The original marginal distributions can also be modeled through GAMs by using the available covariates; recall §3.1.

After transformation to the marginal exponential scale in the case of asymptotic dependence, the observations $M_{\boldsymbol{\omega},i}^{\min\downarrow} = \min_{\boldsymbol{\omega}}(\boldsymbol{X}_i^{E\downarrow})$ are assumed to be exponential, and they are right-censored above a low threshold u_{lower}. In the case of asymptotic independence, we extract the positive threshold excesses $\left(M_{\boldsymbol{\omega},i}^{\min} - u_{\text{upper}}\right) \mid M_{\boldsymbol{\omega},i}^{\min} > u_{\text{upper}}$ of $M_{\boldsymbol{\omega},i}^{\min} = \min_{\boldsymbol{\omega}}(\boldsymbol{X}_i^{E})$ above a high threshold u_{upper}.

For estimating extremal dependence conditional to covariates, we then estimate $\boldsymbol{z} \to A(\boldsymbol{\omega}; \boldsymbol{z})$ and $\boldsymbol{z} \to \lambda(\boldsymbol{\omega}; \boldsymbol{z})$ for fixed $\boldsymbol{\omega}$. In both cases, relevant values of the dependence function are contained in the interval $[\max_{j=1}^{d} \omega_j, 1]$, such that we propose to use the link function

$$h_{\boldsymbol{\omega}}(x) = \log \frac{x - \max_{j=1}^{d} \omega_j}{1 - x}, \quad x \in \left(\max_{j=1}^{d} \omega_j, 1 \right),$$

to link the exponential rate parameter to the GAM predictor. Specifically for $d = 2$, fixing $\boldsymbol{\omega} = (0.5, 0.5)$ allows estimating the extremal dependence summaries $\chi(\boldsymbol{z})$ and $\eta(\boldsymbol{z})$. In practice, it may be useful to obtain estimates for both dependence regimes, and to investigate the results to draw conclusions about the membership in one of the regimes. In some cases, the covariate value $\boldsymbol{z}$ may govern the dependence regime.

For estimation, one could opt for the classical GAM framework using penalized likelihood methods. Alternatively, Bayesian methods could be implemented, such as INLA, which already provides support for censored exponential distributions in the context of survival analysis. We recall the standard GAM approach. We aim to estimate the following semi-parametric specification

$$f(\boldsymbol{\omega}; \mathbf{z}) = h^{-1}\left\{ \mathbf{z}^T \boldsymbol{\beta} + \sum_{k=1}^{K} h_k(t_k; \boldsymbol{z}) \right\}$$

with f being one of $\{A, \lambda\}$, link function h, and semi-parametric effects h_k, such as spline functions described through B-splines and their coefficients to be estimated. The we denote by $\boldsymbol{\lambda}$ the column vector that gathers all parameters to be estimated in the model. The dimension (*e.g.*, number of spline knots) and smoothness of semi-parametric functions h_k can be determined based on the *penalized log-likelihood* with *penalty parameters* $\gamma_k \geq 0$, $k = 1, \ldots, K$, given as

$$\ell(\lambda, \boldsymbol{\gamma}) = \ell(\lambda) - \frac{1}{2} \sum_{k=1}^{K} \gamma_k \int_{\mathbb{H}_k} h_k''(t_k)^2 \, \mathrm{d}t_k,$$

Coefficients in the semi-parametric functions can be optimized given fixed penalty parameters, while optimal penalty parameters can be found through techniques such as cross-validation. This avoids overfitting and allows finding the optimal smoothness of functions h_k. In Bayesian estimation, this regularization mechanism is obtained through the choice of prior variance and dependence of coefficients to be estimated.

A simulation study in Mhalla et al. (2019) confirms good estimation behavior of classical GAM estimation for the the directional dependence functions in practice, especially when the sample size is relatively large.

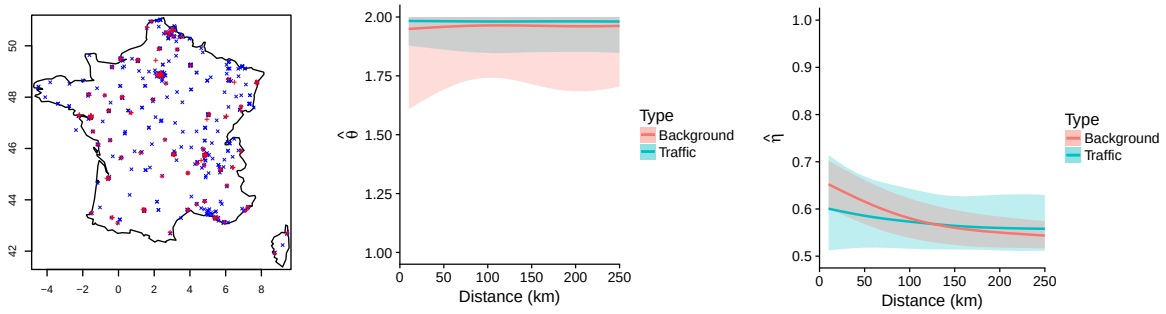


Figure 4.2: Nonparametric extremal dependence modeling of NO_2 air pollution in France. Left: measurement stations with road traffic stations in red. Middle: Effect of spatial distance assuming asymptotic dependence. Right: effect of spatial distance assuming asymptotic independence.

4.1.4 Application to NO_2 air pollution in France

We illustrate the usefulness of our modeling framework on a large dataset of nitrogen dioxide (NO_2) measurements recorded at 569 monitoring stations in France between 1999 and 2012. We use the GAM framework for modeling marginal distributions and tail dependence in large concentrations observed at pairs of stations. We tackle the central question of how extremal dependence evolves with spatial distance and over years.

Our results imply asymptotic independence of data observed at different stations, and we find that the estimated coefficients of tail dependence decrease as a function of spatial distance and show distinct patterns for different years and for different types of stations (traffic vs. background).

Due to the very large data set, we focus modeling on monthly maxima extracted from the original data, and we estimate covariate-driven dependence coefficients that are useful under asymptotic dependence or asymptotic independence. Univariate distributions were fitted using GAMs for marginal GEVDs to normalize margins to the exponential scales. Then, the GAMs for extremal dependence summaries were constructed as follows by using nonlinear functions of spatial distance (dist) between locations and of year of observation (t), and we consider two types of measurement stations (road traffic stations vs. background stations):

$$\begin{aligned} h\{\theta(t, \text{dist}, \text{type})/2\} =& \theta_0 + f_1^\theta(t) + f_2^\theta(\text{dist}) + f^\theta(t, \text{dist}), \\ h\left[\{2\eta(t, \text{dist}, \text{type})\}^{-1}\right] =& \eta_0 + f_1^\eta(t) + f_2^\eta(\text{dist}) + f^\eta(t, \text{dist}), \end{aligned}$$

with the classical bivariate extremal coefficient $\theta = 2 - \chi$ (in the case of asymptotic dependence) and the coefficient of tail dependence η in the case of asymptotic independence.

The network of monitoring stations and some estimation results with respect to spatial distance are illustrated in Figure 4.2. The results clearly point towards asymptotic independence in data, and distinct extremal dependence behavior was observed for different years, and for stations near road traffic as compared to background stations; see Mhalla et al. (2019) for details.

4.1.5 Possible extensions of GAM-based extremal dependence

The method developed above was centered on estimating response surfaces for fixed direction $\boldsymbol{\omega}$. In contrast, to estimate the full dependence functions A or λ for a fixed covariate level $\boldsymbol{z}$, it is necessary to satisfy the shape constraints on these functions (*e.g.*, convexity). For this purpose, constrained optimization tools can be made available (*e.g.*, related to the *cobs* technique – COnstrained B-Splines Nonparametric Regression Quantiles). Moreover, in some cases visual inspection of estimation results may allow us to decide on the asymptotic dependence regime with high certainty, as in the case of the above air pollution data, but we deem important that formal statistical decision tools are developed to distinguish between asymptotic dependence and asymptotic independence in data conditional to auxiliary data. In a Bayesian setting, where dependence characteristics for asymptotic dependence and asymptotic independence could be estimated jointly, the design of prior distributions could be geared towards a clear preference for one of the two regimes in the posterior distributions. More generally, in cases where a mix of the two regimes is possible depending on the predictor information, *e.g.*, due to different weather patterns, we could seek to identify a partition of the covariate space into two subspaces corresponding to the two regimes, respectively.

4.2 Extremal dependence in random factor constructions

The preceding section exploited "model-free" asymptotic representations to develop flexible formulations of extremal dependence conditional to auxiliary information. In this section, we turn away from covariate-based approaches and take a more theoretical perspective by characterizing extremal dependence properties of large classes of bivariate constructions, most of which are non-asymptotic distributions and can be used to model the full bivariate distribution. The question of how to construct flexible and interpretable dependence models accounting for a variety of bivariate extremal behavior with respect to summaries such as χ and $\overline{\chi}$ is recurrent in the literature, and no general consensus and unified treatment has emerged so far. In Engelke et al. (2019), we decide to consider the extremal dependence properties of vectors with a random scale construction $(X_1, X_2) = R(W_1, W_2)$, motivated by the ubiquity of such constructions in the literature and their high flexibility. In most of the spatial and spatiotemporal dependence models considered in the literature and particularly in my work, the bivariate distributions can be cast into such random scale representations, such that this section provides a very general theoretical framework and a catalog of useful constructions for modeling extremal dependence. At the end of this section, it is illustrated how our results can be used to propose new models that encompass both dependence classes of asymptotic dependence and independence.

A rich variety of bivariate dependence models have a pseudo-polar representation

$$(X_1, X_2) = R(W_1, W_2), \qquad R > 0 \text{ independent of } (W_1, W_2) \in \mathcal{W} \subseteq \mathbb{R}^2, \tag{4.5}$$

where we term R the *radial variable*, assumed to have a non-degenerate distribution, and (W_1, W_2) the *angular variables*. Indeed, many well-known copula families, including the elliptical, Archimedean, Liouville and multivariate Pareto families have such a representation. We here focus on the upper tail dependence of such constructions. In particular, we examine whether a given (X_1, X_2) displays asymptotic dependence or asymptotic independence, and the strength of dependence within these classes.

Using the common summary parameters χ and $\eta = 2(\overline{\chi} - 1)$, we obtain that broad factors affecting the results are the heaviness of the tails of R and (W_1, W_2), the shape of the support of (W_1, W_2), and dependence between (W_1, W_2). When R is distinctly lighter-tailed than (W_1, W_2), the extremal dependence of (X_1, X_2) is typically the same as that of (W_1, W_2), whereas similar or heavier tails for R compared to (W_1, W_2) typically result in increased extremal dependence. Similar tail heavinesses in R and (W_1, W_2) represent the most interesting and technical cases, and we find both asymptotic independence and dependence of (X_1, X_2) possible in such cases when (W_1, W_2) exhibit asymptotic independence. The bivariate case often directly extends to higher-dimensional vectors and spatial processes, where the dependence is mainly analyzed in terms of summaries of bivariate sub-vectors.

A broad split in representations of type (4.5) is the dimension of $\mathcal{W}$, the support of (W_1, W_2). The most common case in the literature is that $\mathcal{W}$ is a one-dimensional subset of $\mathbb{R}^2$, such as the unit sphere defined by some norm or other homogeneous function. Examples include the Mahalanobis norm (elliptical distributions), L_1 norm (Archimedean and Liouville distributions) or L_∞ norm (multivariate Pareto distributions). On top of the support $\mathcal{W}$, to obtain distributions within a particular family, R or (W_1, W_2) may be specified to have a certain distribution. Where $\mathcal{W}$ is two-dimensional, it may sometimes be reduced to the one-dimensional case by redefining R, such as in the Gaussian scale mixtures of Huser et al. (2017). Where $\mathcal{W}$ is two-dimensional, the possible constructions stemming from (4.5) form an especially large class, since (W_1, W_2) can itself have any copula. In this case, we focus on how the multiplication by R changes the extremal dependence of (W_1, W_2), summarized by the coefficients (χ_W, η_W), to obtain the extremal dependence of the modified vector (X_1, X_2) in terms of its coefficients (χ_X, η_X). The marginal distributions of (W_1, W_2) and R play a crucial role, since, intuitively, the heavier the tail of R the more additional dependence is introduced in the vector (X_1, X_2).

As we are focused on the upper tail of (X_1, X_2), we henceforth assume $(W_1, W_2) \in \mathbb{R}^2_+$; by the invariance of copulas to monotonic marginal transformations, this also covers random location constructions of the form $(Y_1, Y_2) = S + (V_1, V_2)$, $S \in \mathbb{R}$, $(V_1, V_2) \in \mathcal{V} \subseteq \mathbb{R}^2$, with independence between S and (V_1, V_2). For simplicity of presentation, we will often make the restriction that W_1 and W_2 have the same distribution.

Various authors have focused on extremal dependence arising from certain types of polar representation, but from the perspective of conditional extremes proposed by (Heffernan and Tawn, 2004). This is different to our focus; here we examine the extremal dependence as both variables grow at the same rate. Next, some of the most interesting and useful results of Engelke et al. (2019) are summarized, and a literature review framing a large number of existing examples in terms of our general results is given afterwards. For the following results, recall the list of tail classes and associated notations given for reference in §2.2.3. If Q represents a bivariate random vector $Q = (Q_1, Q_2)$, we denote the minimum of its margins by $Q_\wedge = Q_1 \wedge Q_2$.

4.2.1 Constrained angular variables

We first focus on the case where $\mathcal{W}$ is defined by a norm ν; specifically let $\mathcal{W} = \{(w_1, w_2) \in \mathbb{R}^2_+ : \nu(w_1, w_2) = 1\}$. Other types of constrained spaces may sometimes be of interest, but norm spheres are a common restriction, and this focus allows greater generality in other aspects. We examine the extremal dependence based on the heaviness of the tail of R.

To begin, the case where R belongs to the Fréchet MDA is the least delicate: as long as

R has a much heavier tail than each of (W_1, W_2), results do not strongly depend on other considerations. No equality in distribution is assumed between W_1, W_2 in this case.

Proposition 1 (R in Fréchet MDA)**.** *Let* $\overline{F}_R \in RV^{\infty}_{-\alpha}$, $\alpha \geq 0$, $P(W_1 > 0) = P(W_2 > 0) = 1$, *and* $E(W_j^{\alpha+\varepsilon}) < \infty$, $j = 1, 2$, *for some* $\varepsilon > 0$. *Then* $\eta_X = 1$, *and*

$$\chi_X = E\left[\min\left\{W_1^{\alpha}/E(W_1^{\alpha}), W_2^{\alpha}/E(W_2^{\alpha})\right\}\right]. \tag{4.6}$$

The condition $\mathrm{E}(W_j^{\alpha+\varepsilon}) < \infty$ is guaranteed when $\mathcal{W}$ is the unit sphere of a norm ν. Proposition 1 notably also covers the general case where $(W_1, W_2) \in \mathbb{R}^2_+$. Small modifications to Proposition 1 yield the bivariate exponent function of X, given by

$$V_X(x_1, x_2) = \mathrm{E}\left[\max\left\{W_1^{\alpha}/(\mathrm{E}(W_1^{\alpha})x_1), W_2^{\alpha}/(\mathrm{E}(W_2^{\alpha})x_2)\right\}\right]. \tag{4.7}$$

Next, we suppose that R is in the Gumbel MDA, with upper endpoint $r^\star \in (0, \infty]$, *i.e.*,

$$\lim_{t \to r^\star} \overline{F}_R(t + r/b(t))/\overline{F}_R(t) = e^{-r},$$

where $b(t)$ is termed the auxiliary function. Such distributions can be expressed as

$$\overline{F}_R(r) = c(r) \exp\left\{-\int_z^r b(t)\mathrm{d}t\right\}, \tag{4.8}$$

where $z < r < r^\star$, $c(r) \to c > 0$ as $r \to r^\star$, and the function $a = 1/b$ is absolutely continuous with density a' satisfying $\lim_{t \to r^\star} a'(t) = 0$ (e.g. Embrechts et al., 1997, Chapter 3.3). If $r^\star = \infty$, we also have that for any $\lambda > 1, \rho \in \mathbb{R}$,

$$\lim_{r \to \infty} (rb(r))^{\rho} \overline{F}_R(\lambda r)/\overline{F}_R(r) = 0. \tag{4.9}$$

Before stating results for the Gumbel case of R, we first have to introduce some technical notation. Suppose that $W_1 \overset{d}{=} W_2 \overset{d}{=} W \in [0, 1]$ and $\nu(W_1, W_2) = 1$. To this end, we assume that ν is a symmetric norm, *i.e.*, $\nu(x, y) = \nu(y, x)$, and scaled to satisfy $\nu(x, y) \geq \max(x, y)$, such that the unit sphere of ν is contained in that of the maximum, with $\nu(b, 1 - b) = b$ for some $b \geq 1/2$. Let $\tau(z) = z/\nu(z, 1 - z) = 1/\nu(1, 1/z - 1)$. The random variable $Z = W_1/(W_1 + W_2) \in [0, 1]$ satisfies

$$(W_1, W_2) = (Z, 1 - Z)/\nu(Z, 1 - Z) = (\tau(Z), \tau(1 - Z)). \tag{4.10}$$

Define $I_\nu = [b_1, b_2] \subseteq [1/2, 1]$ as the interval such that $\tau(z) = 1$ for all $z \in I_\nu$, and $\tau(z) < 1$ for $z \notin I_\nu$, and write $\tau(z) = \tau_1(z)$ for $z \in [0, b_1]$, $\tau(z) = 1$ for $z \in [b_1, b_2]$, and $\tau(z) = \tau_2(z)$ for $z \in [b_2, 1]$, with τ_1 strictly increasing and τ_2 strictly decreasing. Figure 4.3 illustrates τ for a particular ν.

We assume further that

(Z1): Z has a Lebesgue density, f_Z, positive everywhere on $(0, 1)$, and that its survival function is regularly varying at 1, with $\overline{F}_Z(1 - \cdot) \in \mathrm{RV}^0_{\alpha_Z}$, $\alpha_Z > 0$,

and make the following mild regularity assumptions on the norm, ν, or equivalently τ:

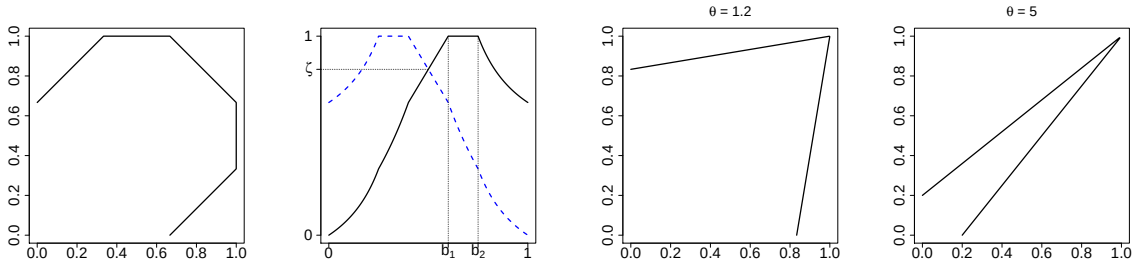


Figure 4.3: Illustration of unit spheres of norms. Left: the unit sphere for a particular norm ν; centre-left: illustration of $\tau(z)$ (solid line) and $\tau(1-z)$ (dashed line) for the same ν. Centre-right and right: illustration of the unit sphere of $\nu(x,y) = \theta \max(x,y) + (1-\theta)\min(x,y)$ for two different values of θ.

(N1): The function τ is twice (piecewise) continuously differentiable except for finitely many points, at which we only require existence of left and right derivatives of first and second order.

(N2): τ is regularly varying as it approaches 1 from either side, i.e., $1-\tau_1(b_1-\cdot) \in \mathrm{RV}^0_{1/\gamma_1}$, and, if $b_2 < 1$, $1-\tau_2(b_2+\cdot) \in \mathrm{RV}^0_{1/\gamma_2}$, $\gamma_j \in (0,1]$, $j=1,2$. We label $\gamma = \min(\gamma_1,\gamma_2)$ with $\gamma = \gamma_1$ if $b_2 = 1$.

In practice, (N1) and (N2) are satisfied by a wide variety of commonly used norms, and the upper limit of $\gamma_1, \gamma_2 \leq 1$ in (N2) is a consequence of convexity of the norm ν. Note that $\tau(z) \lesseqgtr \tau(1-z) \Leftrightarrow z \lesseqgtr 1-z$, so that

$$\min(\tau(z), \tau(1-z)) = \begin{cases} \tau(z), & z \in [0, 1/2], \\ \tau(1-z), & z \in [1/2, 1]. \end{cases}$$

Finally, denote $\zeta = \tau(1/2) \in [1/2, 1]$, so that $W_\wedge = \min(\tau(Z), \tau(1-Z)) \in [0, \zeta]$.

Proposition 2 (R in Gumbel MDA)**.** *Assume $\overline{F}_R$ satisfies (4.8) and a number of other mild assumptions, see Engelke et al. (2019). Then:*

1. *If $\zeta < 1$, then $\chi_X = 0$ and $\eta_X = \lim_{x \to r^\star} \log \overline{F}_R(x) / \log \overline{F}_R(x/\zeta)$, which is defined only for $r^\star = \infty$.*

2. *If $\zeta = 1$, then $\eta_X = 1$. Further, $b_1 = 1/2$ and*

$$\chi_X = \begin{cases} 0 & \text{if } b_2 > 1/2, \text{ i.e., } P(W=1) > 0, \\ \dfrac{2\tau_2'(1/2_+)}{\tau_2'(1/2_+) - \tau_1'(1/2_-)} & \text{otherwise.} \end{cases}$$

We observe that asymptotic independence arises for $\zeta < 1$, with the residual tail dependence coefficient determined by the properties of $\overline{F}_R$. The following corollary covers an important subclass of distributions in the Gumbel MDA.

Corollary 1. *If $\zeta < 1$ and $-\log \overline{F}_R \in RV^{\infty}_{\delta}$, $\delta \geq 0$, then $\eta_X = \zeta^{\delta}$.*

Specifically if $\delta = 0$, as in the case of log-normal R, then $\eta_X = 1$.

Finally, we suppose that $R > 0$ is in the negative Weibull MDA with upper endpoint $r^{\star} > 0$, *i.e.*,

$$\overline{F}_R(r^{\star} - s) = \ell(s)s^{\alpha_R}, \qquad \ell \in \mathrm{RV}^0_0,\ \alpha_R > 0;$$

equivalently, we have that $\overline{F}_R(r^{\star} - \cdot) \in \mathrm{RV}^0_{\alpha_R}$.

Proposition 3 (R in negative Weibull MDA)**.** *Assume $\overline{F}_R(r^{\star} - \cdot) \in RV^0_{\alpha_R}$ and that (N1), (N2) and (Z1) hold. Then:*

1. *If $\zeta < 1$, then $\chi_X = 0$ and η_X is not defined.*
2. *If $\zeta = 1$, then $b_1 = 1/2$ and*

$$\chi_X = \begin{cases} 0 & \text{if } P(W = 1) > 0, \\ \dfrac{2\tau_2'(1/2_+)}{\tau_2'(1/2_+) - \tau_1'(1/2_-)} & \text{otherwise,} \end{cases} \qquad \eta_X = \begin{cases} \frac{\alpha_R}{1+\alpha_R} & \text{if } P(W = 1) > 0, \\ 1 & \text{otherwise.} \end{cases}$$

4.2.2 Unconstrained angular variables

We now treat the case where the support $\mathcal{W}$ is two-dimensional. To avoid additional complications, we assume throughout this subsection that W_1 and W_2 share the common marginal distribution F_W. We also generally assume that the tail dependence coefficient χ_W and the residual tail dependence coefficient η_W of (W_1, W_2) exist. For two-dimensional $\mathcal{W}$, the variety of marginal and dependence behaviors possible for (W_1, W_2) means that systematic characterization according only to the MDA of R is more difficult. In fact, we need to consider different tail decays of both the radial variable R and the angular variable W since the combination of the two is crucial to classify the extremal dependence of $(X_1, X_2) = R(W_1, W_2)$. We focus on some interesting sub-classes that still incorporate a wide variety of structures and cover most of the parametric univariate distributions available for R and W. We structure results according to the tail heaviness assumed for R, W, or both of them. In decreasing order we consider distributions with superheavy tails, regularly varying distributions, distributions of log-Weibull and Weibull type in the following exposition. Table 4.1 summarizes the general results developed in the following, and Table 4.2 contains the extremal dependence coefficients for all combinations of tail decays of R and W for the specific, yet interesting example where W_1 and W_2 are independent.

Radius R	additional assumptions	χ_X	η_X
Superheavy tails			
a) $\overline{F}_W(x)/\overline{F}_R(x) \to c$	Prop. 4	$\frac{1+c\chi_W}{1+c}$	1
b) $\overline{F}_R = o(\overline{F}_W)$	$\chi_W > 0$	χ_W	1
	$\chi_W = 0$, $\overline{F}_R(x) \le C\overline{F}_{W_\wedge}(x)$	0	η_W
	$\chi_W = 0$, $\overline{F}_{W_\wedge} = o(\overline{F}_R)$	0	(2b)
$\mathrm{RV}^\infty_{-\alpha_R}$			
a) $\mathrm{E}(W^{\alpha_R+\varepsilon}) < \infty$	$\mathrm{P}(W > 0) = 1$	(4.6)	1
b) $\overline{F}_W \in \mathrm{RV}^\infty_{-\alpha_W}$			
(i) $\alpha_R > \alpha_W$		χ_W	(4.12)
(ii) $\alpha_R = \alpha_W$	Prop. 6	Prop. 6	1
$\mathrm{LWT}_{\beta_R>1}$	$F_W, F_{W_\wedge} \in \mathrm{LWT}_{\beta_R}$	χ_W	(4.14)
WT_{β_R}	$F_W \in \mathrm{WT}_{\beta_W}, F_{W_\wedge} \in \mathrm{WT}_{\beta_{W_\wedge}}$	Prop. 8	Prop. 8

Table 4.1: Tail dependence summaries χ_X and η_X for $(X_1, X_2) = R(W_1, W_2)$ with different tail decay rates of the radial variable $R > 0$ and unconstrained variables $W_1 \overset{d}{=} W_2$.

Angle W Radius R	Super-heavy	Reg. varying	log-Weibull ($\beta_W > 1$)	Weibull	Neg. Weibull
Super-heavy	$\chi_X = (1+c)^{-1}$ η_X: Prop. 4(2)	$\chi_X = 1$	$\chi_X = 1$	$\chi_X = 1$	$\chi_X = 1$
Reg. varying	*	$\alpha_R < \alpha_W : \chi_X = (4.6) > 0$ $\alpha_R = \alpha_W$: Prop. 6 $\alpha_W < \alpha_R < 2\alpha_W$: $\eta_X = \alpha_W/\alpha_R$ $\alpha_R > 2\alpha_W$: $\eta_W = 1/2$	$\chi_X = (4.6) > 0$	$\chi_X = (4.6) > 0$	$\chi_X = (4.6) > 0$
log-Weibull ($\beta_R > 1$)	*	*	$\beta_R = \beta_W : \eta_X = (4.14)$	unknown	$\chi_X = 0$ $\eta_X = 1$
Weibull	*	*	unknown	$\eta_X = 2^{-\beta_R/(\beta_R+\beta_W)}$	$\chi_X = 0$ $\eta_X = 1$
Neg. Weibull	*	*	*	*	$\eta_X = \dfrac{\alpha_W + \alpha_R}{2\alpha_W + \alpha_R}$

Table 4.2: The values of χ_X and η_X for $(X_1, X_2) = R(W_1, W_2)$ with $W_1, W_2 \overset{d}{=} W$ independent, with different tail decay rates of the radial and angular variables. The *'s indicate that multiplication with R does not change the tail dependence of (W_1, W_2), *i.e.*, $\chi_X = \chi_W = 0$ and $\eta_X = \eta_W = 1/2$. The combinations of Weibull and log-Weibull tails remain open problems.

Superheavy-tailed variables

Suppose that R or W is superheavy-tailed, *i.e.*, $\log R$ or $\log W$ is heavy-tailed. This case naturally arises when considering random location constructions $\log R + (\log W_1, \log W_2)$; we thus further assume $W > 0$ so that $\log W_j$, $j = 1, 2$, are well defined.

Proposition 4 (Superheavy-tailed variables)**.**

1. *If* $F_{\log R} \in \mathrm{CE}_0$ *and* $\overline{F}_W(x)/\overline{F}_R(x) \to c \geq 0$ *as* $x \to \infty$*, then* $\eta_X = 1$ *and*
$$\chi_X = (1 + c\,\chi_W)/(1 + c) > 0. \tag{4.11}$$
2. *If* $F_{\log W} \in \mathrm{CE}_0$ *and* $\overline{F}_R = o(\overline{F}_W)$*, then* $\chi_X = \chi_W$*. If further* $F_{\log R} \in \mathrm{CE}_0$ *and*
   (a) $F_{\log W_\wedge} \in \mathrm{CE}_0$ *with* $\overline{F}_R(x)/\overline{F}_{W_\wedge}(x) \leq C$ *for a constant* $C > 0$ *as* $x \to \infty$*, then* $\eta_X = \eta_W$*;*
   (b) $\overline{F}_{W_\wedge} = o(\overline{F}_R)$*, then, provided the limit exists,*
$$\eta_X = \lim_{x\to\infty} \log \overline{F}_W(x)/\log \overline{F}_R(x).$$

Regularly varying variables

In this section we consider the case where R, W or both of them are regularly varying. When R is regularly varying with index $\alpha_R > 0$ and $\mathrm{E}(W^{\alpha_R+\varepsilon}) < \infty$ for some $\varepsilon > 0$, then the tail dependence coefficient χ_X is as described in Proposition 1. We first consider the case where W is regularly varying with index $\alpha_W > 0$ and R is lighter tailed, *i.e.*, either also regularly varying with $\alpha_R > \alpha_W$ or even lighter-tailed such as distributions in the Gumbel or negative Weibull domain of attraction. Second, we study the case where both R and W are regularly varying with the same index $\alpha_W = \alpha_R$, which turns out to be particularly involved, and which requires additional assumptions.

Proposition 5 (W regularly varying with R lighter)**.** *Let* $\overline{F}_W \in RV^\infty_{-\alpha_W}$*,* $\alpha_W \geq 0$*, and suppose that either* $\overline{F}_R \in RV^\infty_{-\alpha_R}$ *with* $\alpha_R > \alpha_W$*, or* R *is in the Gumbel or negative Weibull domain of attraction; denote the latter case by* $\alpha_R = \infty$*. Then* $\chi_X = \chi_W$ *and*
$$\eta_X = \begin{cases} \alpha_W/\alpha_R, & \text{if } \alpha_R < \alpha_W/\eta_W, \eta_W = 0 \text{ or } \eta_W \text{ not defined,} \\ \eta_W, & \text{if } \alpha_R > \alpha_W/\eta_W \text{ or } \alpha_R = \infty. \end{cases} \tag{4.12}$$

The case where R and W are regularly varying with the same index $\alpha > 0$ leads to various scenarios for the extremal dependence in (X_1, X_2).

Proposition 6 (Regularly varying R and W with the same index)**.** *Let* $\overline{F}_R, \overline{F}_W \in RV^\infty_{-\alpha}$ *with* $\alpha > 0$*. Then* $\eta_X = 1$*, and we have the following:*

1. *If* $F_{\log R} \in \mathrm{CE}_\alpha$*, and if* $\overline{F}_W(x)/\overline{F}_R(x) \to c \geq 0$ *as* $x \to \infty$*, then*
$$\chi_X = \frac{E(W^\alpha_\wedge) + c\,\chi_W\,E(R^\alpha)}{E(W^\alpha) + c\,E(R^\alpha)}.$$

2. *If* $F_{\log W} \in \mathrm{CE}_\alpha$ *and* $\overline{F}_R = o(\overline{F}_W)$, *then* $\chi_X = \chi_W$.

3. *Let* $F_{\log R} \in \mathrm{ET}_{\alpha,\beta_R}$ *with* $\beta_R \geq -1$ *and* $E(R^\alpha) = \infty$ *if* $\beta_R = -1$, *and let* $F_{\log W} \in \mathrm{ET}_{\alpha,\beta_W}$.

   (a) *If* $\chi_W > 0$ *and if either* $\beta_W > -1$ *or* $\beta_W = -1$ *and* $E(W^\alpha) = \infty$, *then* $\chi_X = \chi_W$.

   (b) *If* $\chi_W \geq 0$ *and if either* $\beta_W < -1$ *or* $\beta_W = -1 < \beta_R$ *and* $E(W^\alpha) < \infty$, *then* $\chi_X = E(W_\wedge^\alpha)/E(W^\alpha)$.

   (c) *If* $\beta_R > -1$, $\beta_W > -1$ *and* $E(W_\wedge^{\alpha+\varepsilon}) < \infty$ *for some* $\varepsilon > 0$, *then* $\chi_X = 0$.

We remark that the tail of R is not dominated by that of W in Proposition 6(1), while it is dominated in Proposition 6(2). Proposition 6(3) shifts focus to relatively heavy tails in R with $\mathrm{E}(R^\alpha) = \infty$, such as the gamma tails of $\mathrm{ET}_{\alpha,\beta}$ with $\beta > -1$.

Log-Weibull-type variables

In this and the following subsubsection we concentrate on radial and angular variables in the Gumbel domain of attraction. Due to the large variety of distributions in this domain we consider subsets that include the most commonly used distribution families. We first study the case where both R and W are log-Weibull-tailed; equivalently, $\log R$ and $\log W$ are Weibull-tailed. We recall that a random variable Y is log-Weibull-tailed if

$$\overline{F}_Y(y) = \ell(\log y)(\log y)^\gamma \exp\left(-\alpha(\log y)^\beta\right), \qquad \ell \in \mathrm{RV}_0^\infty, \gamma \in \mathbb{R}, \alpha, \beta > 0, \tag{4.13}$$

and we write $F_Y \in \mathrm{LWT}_\beta$. In the following, we denote the β-parameters of R and W by β_R and β_W, respectively. The superheavy-tailed case, $\beta_R < 1$ or $\beta_W < 1$, is already covered by §4.2.2, and the case of regularly varying tails with $\beta_R = 1$ or $\beta_W = 1$ is treated in §4.2.2. We therefore study the remaining case $\beta_R > 1$ and $\beta_W > 1$, which encompasses important distributions such as the log-Gaussian. It is more intuitive to consider the random location construction $\log R + \log(W_1, W_2)$, where we can apply convolution-based results. When independent heavy-tailed summands are involved in the convolution, typically only one of the values of summands has a dominant contribution to a high values of the sum, resulting in relatively simple formulas; see §4.2.2. On the contrary, in the light-tailed set-up all summands may contribute significantly when high values arise in the sum, rendering the tail analysis more intricate. Only relatively few general results on convolutions with tails lighter than exponential are available in the literature.

We consider the set-up where the components R, W and $W_\wedge$ are log-Weibull-tailed with the same coefficient $\beta > 1$ and a simplified form of the slowly varying function ℓ by assuming that it is asymptotically constant, *i.e.*, $\ell(x) \sim c > 0$.

Proposition 7 (Light-tailed random location with $F_R \in \mathrm{LWT}_\beta$, $\beta > 1$)**.** *Suppose that* $F_R, F_W, F_{W_\wedge} \in \mathrm{LWT}_\beta$ *with possibly different parameters* α, γ *indexed by the corresponding* R, W *and* $W_\wedge$, *but where* $\beta = \beta_R = \beta_W = \beta_{W_\wedge} > 1$. *Assume that the slowly varying functions* ℓ *behave asymptotically like positive constants.*

1. *If* $\chi_W > 0$, *then* $\chi_X = \chi_W > 0$.

2. *If* $\chi_W = 0$, *then* $\chi_X = 0$ *and*

$$\eta_X = \eta_W \times \left(\frac{\alpha_{W_\wedge}^{1/(\beta-1)} + \alpha_R^{1/(\beta-1)}}{\alpha_W^{1/(\beta-1)} + \alpha_R^{1/(\beta-1)}} \right)^{\beta-1}, \tag{4.14}$$

where $\eta_W = \alpha_W/\alpha_{W_\wedge}$, *and* $\eta_X = \eta_W$ *if* $\alpha_W = \alpha_{W_\wedge}$.

Example 1 (Gaussian factor model)**.** Suppose that $\log R$ is univariate standard Gaussian and that $\log(W_1, W_2)$ is bivariate standard Gaussian, independent of R and with Gaussian correlation $\rho_W \in (-1, 1]$. Then we have log-Weibull tails with parameters $\beta_R = \beta_W = \beta_{W_\wedge} = 2$, $\alpha_R = \alpha_W = 1/2$ and $\alpha_{W_\wedge} = 1/(1+\rho_W)$ (see Example 2). Applying (4.14) gives $\eta_X = \eta_W \times (3+\rho_W)/(2(1+\rho_W)) = (3+\rho_W)/4$.

Weibull-type variables

We now consider the case where R and W follow a Weibull-type distribution, a rich class in the Gumbel MDA. Recall that a variable Y is of Weibull-type, $F_Y \in \mathrm{WT}_\beta$, if

$$\overline{F}_Y(y) = \ell(y) y^\gamma \exp\left(-\alpha y^\beta\right), \qquad \ell \in \mathrm{RV}_0^\infty, \gamma \in \mathbb{R}, \alpha, \beta > 0. \tag{4.15}$$

Well-known examples of Weibull-tailed distributions are the Gaussian with $\beta = 2$, the gamma with $\beta = 1$ or, more generally, the Weibull where β is called the Weibull index.

For developing useful results, we further assume that, in addition to R and W, $W_\wedge$ also has a Weibull-type tail. As previously, we index the corresponding ℓ functions and the parameters α, γ in (4.15) by the variable name.

Proposition 8 (Weibull-type variables)**.** *Suppose that* $F_R \in \mathrm{WT}_{\beta_R}$, $F_W \in \mathrm{WT}_{\beta_W}$ *and* $F_{W_\wedge} \in \mathrm{WT}_{\beta_{W_\wedge}}$. *We have the following hierarchy of dependence structures:*

1. *If* $\beta_{W_\wedge} = \beta_W$, $\alpha_{W_\wedge} = \alpha_W$, $\gamma_{W_\wedge} = \gamma_W$, *then* $\chi_X = \chi_W = \lim_{x\to\infty} \ell_{W_\wedge}(x)/\ell_W(x)$, *if the limit exists, and* $\eta_X = \eta_W = 1$.

2. *If* $\beta_{W_\wedge} = \beta_W$, $\alpha_{W_\wedge} = \alpha_W$, $\gamma_{W_\wedge} < \gamma_W$, *then* $\chi_X = 0$ *and* $\eta_X = \eta_W = 1$.

3. *If* $\beta_{W_\wedge} = \beta_W$, $\alpha_{W_\wedge} > \alpha_W$, *then* $\chi_X = 0$ *and*

$$\eta_X = \eta_W^{\beta_R/(\beta_R+\beta_W)} = (\alpha_W/\alpha_{W_\wedge})^{\beta_R/(\beta_R+\beta_W)}.$$

4. *If* $\beta_{W_\wedge} > \beta_W$, *then* $\chi_X = 0$ *and* $\eta_X = \eta_W = 0$.

In all of the cases encompassed by Proposition 8, (X_1, X_2) and (W_1, W_2) have the same tail dependence coefficient χ, which can be positive only in case 1. In all other cases the variables are asymptotically independent, and only in case 3 the residual tail dependence coefficient η changes under the multiplication of the radial variable R. Since $\beta_R/(\beta_R+\beta_W) \in (0,1)$, this always leads to an increase in dependence, that is, $\eta_X > \eta_W$.

Example 2 (Gaussian scale mixtures)**.** To illustrate the most interesting case 3 in Proposition 8 we consider (W_1, W_2) following a bivariate Gaussian distribution with standardized margins and correlation ρ_W. We have that

$$\overline{F}_W(x) \sim r_W(x)\exp(-x^2/2), \qquad \overline{F}_{W_\wedge}(x) \sim r_{W_\wedge}(x)\exp\{-x^2/(1+\rho_W)\},$$

where the tail distribution of the minimum follows from bounds on the multivariate Mills ratio (*e.g.*, Hashorva and Hüsler, 2003), and r_W and $r_{W_\wedge}$ are regularly varying functions. Therefore, $\eta_W = (1+\rho_W)/2$, and

$$\eta_X = \eta_W^{\beta_R/(\beta_R+2)} = \{(1+\rho_W)/2\}^{\beta_R/(\beta_R+2)}.$$

Literature review and examples

Here we detail how existing distribution families fit into the framework of the above results.

Elliptical copulas Let Σ be a positive-definite covariance matrix with Cholesky decomposition $\Sigma = AA^T$, and (U_1, U_2) be uniformly distributed on the L_2 sphere $\{(w_1, w_2) : (w_1, w_2)(w_1, w_2)^T = 1\}$. Then $(X_1, X_2) = RA(U_1, U_2)^T$ has an elliptical distribution for any $R > 0$, called the *generator*. Therefore $(W_1, W_2)^T = A(U_1, U_2)^T$ lies on the Mahalanobis sphere $\mathcal{W} = \{(w_1, w_2) : (w_1, w_2)\Sigma^{-1}(w_1, w_2)^T = 1\}$, and the extremal dependence in the upper right orthant is unchanged by taking $(W_j)_+ = \max(W_j, 0)$. It is well known that (X_1, X_2) is asymptotically dependent if and only if R is in the Fréchet MDA; then, the tail dependence coefficient χ_X is given by (4.6), with W_j replaced by $(W_j)_+$. For R in the Gumbel MDA, the scaling condition on ν such that $\tau(w) \in [0, 1]$ yields Σ with diagonal elements 1, off-diagonal elements $\rho \in (-1, 1)$, and residual tail dependence coefficient is given by Proposition 2(1) with $\zeta = \tau(1/2) = \{(1+\rho)/2\}^{1/2}$. In particular, the Gaussian distribution arises when $\overline{F}_R(r) = e^{-r^2/2}$, so that by Corollary 1, $\eta_X = \zeta^2 = (1+\rho)/2$.

Archimedean and Liouville copulas Archimedean copulas, and the larger class of Liouville copulas, arise as the survival copula when (W_1, W_2) is uniformly (respectively Dirichlet) distributed on the positive part of the L_1 sphere $\mathcal{W} = \{(w_1, w_2) \in [0, 1]^2 : w_1 + w_2 = 1\}$, and $R > 0$. That is, $(X_1, X_2) = R(W_1, W_2)$ has an *inverted* Archimedean or Liouville copula, whilst the Archimedean or Liouville copula itself is that of $(t(X_1), t(X_2))$, for a monotonic decreasing transformation t. By taking $t(x) = 1/x$, we obtain $1/(X_1, X_2) = (\tilde{X}_1, \tilde{X}_2) = \tilde{R}(\tilde{W}_1, \tilde{W}_2)$. The dependence properties of Liouville copulas are studied in Belzile and Nešlehová (2017). For (X_1, X_2), their Theorem 1 states that R in the Fréchet MDA leads to asymptotic dependence, whilst the Gumbel and negative Weibull MDAs lead to asymptotic independence. The exponent function given in their Theorem 1 matches Equation (4.7). In their Theorem 2, Belzile and Nešlehová (2017) consider the extremal dependence properties of $1/(X_1, X_2) = \tilde{R}(\tilde{W}_1, \tilde{W}_2)$, *i.e.*, the Liouville copula itself. Since the reciprocal of Dirichlet random variables have regularly varying tails, this links with Proposition 5 which states that asymptotic independence arises if $(\tilde{W}_1, \tilde{W}_2)$ themselves are asymptotically independent and heavier-tailed than R. Proposition 6(3c) is relevant if $\tilde{R}$ and $\tilde{W}$ are regularly varying with the same index.

Multivariate (ρ-)Pareto copulas Let $\rho : (0,\infty)^2 \to (0,\infty)$ be a positive homogeneous function. Multivariate ρ-Pareto copulas arise when $\overline{F}_R(r) = r^{-1}$, *i.e.*, standard Pareto, and the random vector (W_1, W_2) is concentrated on $\mathcal{W} = \{(w_1, w_2) \in \mathbb{R}_+^2 : \rho(w_1, w_2) = 1\}$ with marginals satisfying $\mathrm{E}(W) < \infty$ (Dombry and Ribatet, 2015). The case of $\rho(x_1, x_2) = \max(x_1, x_2)$ leads to the multivariate Pareto copula associated to multivariate generalized Pareto distributions. Such copulas are asymptotically dependent (except for degenerate distributions) with χ_X given by (4.6). Although we have focused on norms and ρ need not be convex, there is nothing in Proposition 1 requiring this.

Model of Wadsworth et al. (2017) They consider the copula induced by taking R to be generalized Pareto, $\overline{F}_R(r) = (1+\lambda r)_+^{-1/\lambda}$, and $\mathcal{W} = \{(w_1, w_2) \in [0,1]^2 : \|(w_1, w_2)\|_* = 1\}$ where $\|\cdot\|_*$ is a symmetric norm subject to certain restrictions. These restrictions mean that $\lambda \leq 0$ corresponds to asymptotic independence; the residual tail dependence coefficient η_X is as given in Proposition 2 for $\lambda = 0$ with $\zeta = \tau(1/2) = \|(1,1)\|_*^{-1}$, and Proposition 3 for $\lambda < 0$. When R is in the Fréchet MDA ($\lambda > 0$) then asymptotic dependence holds with χ_X given by (4.6).

Model of Krupskii et al. (2018) They consider location mixtures of Gaussian distributions, corresponding to scale mixtures of log-Gaussian distributions. Asymptotic dependence occurs when the location variable is of exponential type, *i.e.*, the scale is of Pareto type; the given χ_X can then be obtained via (4.6). When the location is Weibull-tailed but with shape in $(0,1)$, the scale is superheavy-tailed, with $\overline{F}_R \in \mathrm{RV}_0^\infty$, and perfect extremal dependence ($\chi_X = 1$) arises. When the random location is Weibull-tailed with shape in $(1,\infty)$, then the random scale R is in the Gumbel MDA and asymptotic independence arises. If $\overline{F}_{\log R} \in \mathrm{WT}_2$ has the same Weibull coefficient 2 as the standard Gaussian $\log W$ and as $\log W_\wedge$ (provided that $\rho = \mathrm{cor}(\log W_1, \log W_2) \in (-1, 1]$), then we can apply Proposition 7 to calculate the value of η_X given as

$$\eta_X = \eta_W \frac{\alpha_{W_\wedge} + \alpha_R}{\alpha_W + \alpha_R} = \frac{1+\rho}{2} \frac{(1+\rho)^{-1} + \alpha_R}{1/2 + \alpha_R} = \frac{1 + (1+\rho)\alpha_R}{1 + 2\alpha_R},$$

which extends the results of Krupskii et al. (2018). Specifically, with standard Gaussian $\log R$ we get $\eta_X = (3+\rho)/4$, see Example 1.

Model of Huser and Wadsworth (2019) They consider scale mixtures of asymptotically independent vectors where both R and W have Pareto margins with different shape parameters. Asymptotic dependence arises when R is heavier tailed; χ_X is then given by (4.6), whilst asymptotic independence arises when W is heavier tailed and η_X is given by (4.12). When R and W have the same shape parameter, their assumption $\eta_W < 1$ implies that $\mathrm{E}(W_\wedge^{\alpha+\varepsilon}) < \infty$ for some $\varepsilon > 0$, giving asymptotic independence by Proposition 6(3c).

A new model class

We here exemplify a new construction using unconstrained (W_1, W_2) that has the desirable property of smoothly interpolating between asymptotic dependence and asymptotic independence,

whilst yielding non-trivial structures within each class. By smoothness, we mean that the transition between classes occurs at an interior point, θ_0, of the parameter space Θ, and, assuming increasing dependence with θ, $\lim_{\theta\to\theta_{0+}} \chi_X = 0$, $\lim_{\theta\to\theta_{0-}} \eta_X = 1$.

The following proposition provides a general principle for constructing new dependence models permitting both asymptotic dependence and asymptotic independence.

Proposition 9. *Let R be in the MDA of a generalized extreme value distribution with shape parameter $\xi \in \mathbb{R}$, and let (W_1, W_2) with $W_1 \overset{d}{=} W_2 \overset{d}{=} W$ have $\chi_W = 0$, well-defined $\eta_W \in (0,1)$, and $\overline{F}_W(w^\star - \cdot) \in RV^0_{\alpha_W}$, $\alpha_W > 0$. Then*

1. *For $\xi > 0$, $\chi_X = E(W_\wedge^{1/\xi})/E(W^{1/\xi})$, $\eta_X = 1$,*
2. *For $\xi = 0$, $\chi_X = 0$, $\eta_X = 1$,*
3. *For $\xi < 0$, $\chi_X = 0$, $\eta_X = (1 - \xi\alpha_W)/(1 - \xi\alpha_W/\eta_W)$.*

The model construction opportunities from Proposition 9 are quite varied; specifically taking $\overline{F}_R$ that permits all three tail behaviors produces a flexible range of models spanning the two dependence classes.

4.2.3 Discussion of results

In §4.2.1, where (W_1, W_2) is constrained to the sphere of some norm, classical results on multivariate Pareto copulas are recovered for regularly varying R, whereas new structures are obtained for distributions of R with light tails or finite upper endpoint. In particular, for the Gumbel MDA we get a large variety of behaviors for asymptotically independent (X_1, X_2) that strongly depend on the auxiliary function b of R and the shape of the ν-sphere.

For unconstrained distributions of both R and W, §4.2.2 formalizes the general intuition that heavier tails of R introduce more additional dependence in (X_1, X_2). The results summarized in Table 4.2 for the special case of the independence model allow for several conclusions. The most interesting (and involved) situations figure along the main diagonal where R and W have similar tail behavior. Above this main diagonal, R is so heavy that it mostly dominates the extremal dependence in (X_1, X_2). On the other hand, below the diagonal, R is too light tailed, relatively to W, to have an impact on the tail dependence coefficients χ_X and η_X. Similar observations hold true for the more general case of arbitrary dependence in (W_1, W_2) summarized in Table 4.1.

Multivariate analogs of the upper and residual tail dependence coefficients are obtained by considering the d-variate survival function $\mathrm{P}(X_1 \geq x_1, \ldots, X_d \geq x_d)$ recalled in Equations (2.8) and (2.9). For random scale constructions in d dimensions, the results from §4.2.2 are all directly applicable if the W_j components have common margins, since similarly to the bivariate case, we only need to consider the two variables $X_\wedge = R\min(W_1, \ldots, W_d)$ and $X_j = RW_j$. An assumption of common margins is more realistic in spatial models, where dependence is often analyzed in terms of bivariate margins anyway.

Although our focus was on dependence, knowledge on how the marginal scales of R and W and the dependence properties of (W_1, W_2) influence the dependence of (X_1, X_2) makes it easier to construct models (X_1, X_2) that naturally accommodate both marginal distributions and

dependence of multivariate data. Such modeling avoids what may be construed as the artificial separation of modeling of margins and dependence known as copula modeling. For example, in factor constructions based on independent random variables, such as the ones with independent W_1 and W_2 discussed throughout, our results give guidance on the relative tail heaviness of R with respect to (W_1, W_2) necessary to transition from asymptotic independence to asymptotic dependence in (X_1, X_2), and both heavy- or light-tailed marginal distributions are possible by considering the distribution of either (X_1, X_2) or $\log(X_1, X_2)$ as a model for data.

Chapter 5

Spatial-temporal extreme-value modeling based on asymptotically stable dependence

In this chapter, *asymptotic models* are understood as constructions that correspond to limit distributions in extreme-value theory of stochastic processes. The two main representatives are max-stable processes, which arise asymptotically for rescaled componentwise maxima taken over an increasing number of independently replicated processes, and ℓ-Pareto processes, which arise asymptotically for the normalized process conditional to the exceedance of an aggregation functional ℓ over an increasingly large threshold tending towards the upper endpoint of its distributions. We have presented these two model in classes in §2.4. While first approaches to statistical inference for max-stable processes go back to Smith (1990) and have been followed by a large number of publications, the spatial and spatiotemporal modeling using ℓ-Pareto processes has been developed more recently following the seminal theoretical work of Ferreira and De Haan (2014) and Dombry and Ribatet (2015). The first paper having proposed parametric statistical inference in a wide class of models, known as elliptical Pareto processes, will be discussed in §5.1 (Thibaud and Opitz, 2015).

The customary parametric models for spatial and spatiotemporal extremes usually involve Gaussian processes in their construction and may be seen as too unwieldy and inflexible in practice for extremes of complex processes such as temperature or precipitation. In many applications, methods are needed to generate an essentially infinite number of new simulations of extreme spatial and spatiotemporal episodes associated with magnitudes beyond the range of observed values. These simulations can be fed to impact models, *e.g.*, to assess and project potential risks associated with extreme climatic events in hydrological, agricultural or public health processes. Therefore, when gridded datasets are available, it is useful to develop resampling algorithms that use only minimal asymptotic assumptions from EVT of stochastic processes but abstain from more restrictive parametric assumptions. An approach to semi-parametric resampling of spatial extreme episodes was proposed in Opitz et al. (2021), where we combine a lifting step based on the threshold stability of ℓ-Pareto processes with generic nonparametric resampling techniques such as *Direct Sampling*; see §5.2 for details. Moreover, extensions to the spatiotemporal setting were untertaken in Palacios-Rodriguez et al. (2020); we discuss the choice of spatiotemporal aggregation functionals ℓ, and operational solutions to take into account the intermittence in precipitation data (*i.e.*, the singularity at the lower endpoint 0 corresponding to absence of precipitation) are proposed and validated.

5.1 Statistical inference and simulation for elliptical Pareto processes

ℓ-Pareto processes have been established as the natural limits for extreme events defined in terms of exceedances of a risk functional; see §2.4. In Thibaud and Opitz (2015), we provide statistical inference methods based on a tractable yet flexible extremal dependence model. We introduce the class of elliptical ℓ-Pareto processes, which arise as the limit of threshold exceedances of certain elliptical processes and are characterized by a correlation function and a shape parameter. Elliptical processes are characterized by their finite-dimensional distributions that all possess elliptically contoured densities; see §B.1. If the index of components is not finite (*e.g.*, with processes defined over $\mathbb{R}^D$), then such processes can always be constructed as variance mixtures of Gaussian processes.

The *extremal-t process* represents the max-stable counterpart to elliptical ℓ-Pareto processes (Opitz, 2013) and can be seen as a generalization of two models that have proven flexible for modelling extremal dependence in environmental data based on asymptotic theory (Davison et al., 2012, 2019): the extremal Gaussian model (Schlather, 2002), and the Brown–Resnick model (Kabluchko et al., 2009). The extremal-t process arises for asymptotically dependent elliptical processes. The elliptical ℓ-Pareto processes form the corresponding class of limiting process for threshold exceedances of such elliptical processes. We propose an efficient inference approach for parametric models based on a full likelihood with partial censoring. The resulting novel inferential procedures are more efficient than composite likelihood methods, and we assess efficiency gains over a pairwise likelihood in a simulation study in Thibaud and Opitz (2015). In addition, we propose a new approach to exact simulation from extremal-t and elliptical Pareto processes, and we show how conditional simulations can be obtained very easily for the latter.

5.1.1 Likelihood inference for ℓ-Pareto processes

We now work in the setting of the generalized ℓ-Pareto limit processes as obtained in Equation (2.26) and use the assumptions preceding this equation. We first provide the general construction of ℓ-Pareto likelihoods. Consider a collection $\boldsymbol{s} = (s_1, \ldots, s_d)$ of d sites in the compact domain $\mathcal{K} \subset \mathbb{R}^D$. We assume that the finite-dimensional observation vector $X(\boldsymbol{s}) = \{X(s_1), \ldots, X(s_d)\}$ is embedded in a process $\{X(s)\}_{s\in\mathcal{K}}$ that satisfies assumption $(A1)$ and the equivalent assumptions $(A2.i)$, $(A2.ii)$ and $(A2.iii)$. The finite-dimensional marginal measures of the exponent measure Λ relative to d sites $\boldsymbol{s} = (s_1, \ldots, s_d)$ are written as $\Lambda_{\boldsymbol{s}}$, *i.e.*, $\Lambda_{\boldsymbol{s}}(\times_{j=1,\ldots,d}[a_j, b_j]) = \Lambda[\bigcap_{j=1,\ldots,d}\{f \in C(\mathcal{K}) : f(s_j) \in [a_j, b_j]\}]$ for $0 < a_j < b_j$, $j = 1, \ldots, d$. In particular, $V(z_1, \ldots, z_d) = \Lambda_{\boldsymbol{s}}\left\{(\times_{j=1}^{d}[-\infty, z_j])^C\right\}$. Further, we consider a sample $X_1(\boldsymbol{s}), \ldots, X_n(\boldsymbol{s})$ of independent replicates of $X(\boldsymbol{s})$. The finite-dimensional vectors of the marginally normalized process $X^\star$ (recall Equation (2.13)) are denoted by $X^\star(\boldsymbol{s}) = (X^\star(s_1), \ldots, X^\star(s_d))$. It is possible to estimate marginal parameters separately before estimating the dependence from the normalized vector $X^\star(\boldsymbol{s})$, which avoids the challenging maximization of a likelihood of parameters related to both the dependence and marginal distributions. Here we describe the second step, which aims at estimating the measure Λ_+ based on ℓ-exceedances of $X^\star(\boldsymbol{s})$ with a suitably chosen risk functional ℓ. Different choices of ℓ yield different approaches to inference, and it is crucial that $\ell(X^\star)$ can be determined from $X^\star(\boldsymbol{s})$; *i.e.*, we need

$\ell(X^\star) = \ell(X^\star(\boldsymbol{s}))$.

The condition $\ell(X^\star(\boldsymbol{s})) \geq 1$ is used to select the exceedance observation vectors to which the ℓ-Pareto process is fitted. For a standard ℓ-Pareto process $Y_\ell^\star$, the probability density of the vector $Y^\star(\boldsymbol{s}) = \{Y_\ell^\star(s_1), \ldots, Y_\ell^\star(s_d)\}$ on $\{\boldsymbol{y} \in \mathbb{R}_+^d \setminus \{\boldsymbol{0}\} : \ell(\boldsymbol{y}) \geq 1\}$ is $\lambda_{+,\boldsymbol{s}}(\boldsymbol{y})/\kappa_\ell(\mathcal{K})$, where $\lambda_{+,\boldsymbol{s}}$ is the density of $\Lambda_{+,\boldsymbol{s}}$, the finite-dimensional marginal measure of Λ_+ relative to the sites $\boldsymbol{s}$. When $\Lambda_{+,\boldsymbol{s}}$ is absolutely continuous with respect to Lebesgue measure, $\lambda_{+,\boldsymbol{s}}$ is the full derivative $-V_{1:d}(\boldsymbol{y})$, where V is the exponent function. Otherwise, when $\Lambda_{+,\boldsymbol{s}}$ puts positive mass to lower-dimensional subspaces of $\mathbb{R}_+^d$, we get slightly different expressions for $\lambda_{+,\boldsymbol{s}}$ on those subspaces (Coles and Tawn, 1991).

Based on the sample of ℓ-exceedances $X_k^\star(\boldsymbol{s})$, $k = 1, \ldots, N_u$, satisfying $\ell(X_k^\star(\boldsymbol{s})) > 1$, the ℓ-Pareto process has full likelihood

$$\prod_{k=1}^{N_u} \frac{\lambda_{+,\boldsymbol{s}}(X_k^\star(\boldsymbol{s}))}{\kappa_\ell(\mathcal{K})}. \tag{5.1}$$

When the constant $\kappa_\ell(\mathcal{K})$ cannot be calculated explicitly, Monte-Carlo approximations are required to evaluate the likelihood function (5.1). For a choice of ℓ that is both tractable and useful in practice, we here focus on $\ell(f) = \max_{j=1}^d f(s_j)/u_j$ with a high multivariate threshold $\boldsymbol{u} = (u_1, \ldots, u_d) > 0$, leading to an exceedance for at least one of the sites in $\boldsymbol{s}$ when $\ell(f) \geq 1$. Then, $\kappa_\ell(\mathcal{K}) = V(\boldsymbol{u})$, and we get (Ferreira and De Haan, 2014)

$$\mathrm{P}(Y^\star(\boldsymbol{s}) \leq \boldsymbol{y}) = \frac{V\{\min(\boldsymbol{y}, \boldsymbol{u})\} - V(\boldsymbol{y})}{V(\boldsymbol{u})}, \quad \boldsymbol{y} \not\leq \boldsymbol{u},$$

which is the multivariate Pareto distribution defined by Rootzén and Tajvidi (2006). Specifying $\kappa_\ell(\mathcal{K}) = V(\boldsymbol{u})$ in (5.1) yields the corresponding likelihood

$$\tilde{L}_1(\psi) = \prod_{k=1}^{N_u} \frac{\lambda_{+,\boldsymbol{s}}(X_k^\star(\boldsymbol{s}))}{V(\boldsymbol{u})},$$

with parameter vector ψ. Still, inference based on $\tilde{L}_1$ might be inefficient in practice for two reasons. First, using the full information from an observation $X_k^\star(\boldsymbol{s})$ with $\ell(X_k^\star(\boldsymbol{s})) > 1$ might be inefficient since the asymptotic distribution might model the non-exceeding components badly and thus induce bias in the estimators. Second, positive mass on the boundary of $\mathbb{R}_+^d \setminus \{\boldsymbol{0}\}$ creates a discontinuity due to the weak convergence of the data process (without positive mass on the boundary) to the ℓ-Pareto process in (2.26), as is the case for the extremal elliptical model. The margins of $X_k^\star(\boldsymbol{s})$ are chosen to be strictly positive, which is incoherent with the possible mass on the axis for $\Lambda_{+,\boldsymbol{s}}$. In some applications, such mass may have a physical interpretation (*e.g.*, absence of precipitation corresponding to an observed value of 0). However, the normalized marginal distributions must then allow for this singularity at 0, which raises many questions on how to implement the marginal normalization in practice; see Palacios-Rodriguez et al. (2020) for a discussion and an example of a distribution normalized in this way. To overcome the above two issues, we propose the use of a censoring scheme. We consider the censored observations $X_k^c(\boldsymbol{s}) = \max(X_k^\star(\boldsymbol{s}), \boldsymbol{u})$, where the maximum is taken componentwise. The corresponding likelihood is

$$\tilde{L}_2(\psi) = \prod_{k=1}^{N_u} \frac{-V_{I_k}(X_k^c(\boldsymbol{s}))}{V(\boldsymbol{u})},$$

where V_{I_k} denotes the partial derivative of V with respect to the indices $I_k \subset \{1, \ldots, d\}$ associated to the components that exceed their corresponding marginal thresholds. If both n and N_u are observed, we further propose to incorporate the information provided by the binomial variable $n - N_u$, representing the number of fully censored observations, into the modified likelihoods

$$L_m = \{1 - V(\boldsymbol{u})\}^{n-N_u} V(u)^{N_u} \times \tilde{L}_m, \quad m = 1, 2.$$

The threshold vector $\boldsymbol{u}$ must be high enough to yield $V(\boldsymbol{u}) \leq 1$.

Full likelihood inference based on L_1 or L_2 is possible if we know $\lambda_{+,\boldsymbol{s}}$, the function V and its partial derivatives. We derive these expressions for extremal elliptical processes in §5.1.2. Finally, we point out that the likelihood L_2 is closely related to a censored likelihood approach proposed for inference with the Brown–Resnick model in Wadsworth and Tawn (2014).

For modeling in practice, we propose to use flexible parametric families of elliptical Pareto processes, described in the following §, to define the measure Λ_+.

5.1.2 The elliptical Pareto model

The construction and properties of multivariate distributions with elliptically contoured densities, called *elliptical distributions* in short, are recalled in §B.1. We refer to a random vector with elliptical distribution as an *elliptical random vector*. An elliptical random vector can be written as

$$X = RAU + \mu, \tag{5.2}$$

with R a nonnegative random variable, A a $d \times d$ deterministic nonsingular matrix defining the dispersion matrix $\Sigma = AA^T$, U a random vector independent of R and distributed uniformly on the Euclidean unit sphere $\{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x}^T\boldsymbol{x} = 1\}$ and $\boldsymbol{\mu} \in \mathbb{R}^d$ a deterministic shift vector. The restriction to nonsingular square matrices A excludes some special, degenerate cases of minor practical importance. As an extension of (5.2), a random process X is called elliptical if all its finite-dimensional distributions are elliptical with dispersion matrices Σ defined through a correlation function. The max-stable limits in (2.20) for elliptical processes are either processes with independent univariate marginal distributions in the case of asymptotic independence, as for instance the limits of Gaussian processes, or are extremal-t processes in all other cases (Opitz, 2013). In terms of standard Fréchet margins, extremal-t processes can be constructively represented through the spectral construction (2.17) as

$$Z^\star(s) = m_\alpha \max_{i=1,2,\ldots} W_i(s)_+^\alpha / Q_i, \quad m_\alpha = \pi^{1/2} 2^{1-\alpha/2} \Gamma\{(\alpha+1)/2\}^{-1}, \tag{5.3}$$

where $0 < Q_1 < Q_2 < \cdots$ are the points of a unit-rate Poisson process on the positive half-line, and $W_i = \{W_i(s)\}$ are independent replicates of a standard Gaussian process with continuous sample paths and correlation function ρ (Opitz, 2013). By interpreting the processes W_i as independent marks of the points of the Poisson process $\{Q_i\}$, we see that the point process

$$\{P_i,\ i = 1, 2, \ldots\} = \{m_\alpha (W_i)_+^\alpha / Q_i,\ i = 1, 2, \ldots\} \tag{5.4}$$

is Poisson with intensity measure Λ_+. From Poisson process theory (see Chapter 2 in Daley and Vere-Jones, 2003), the points P_i with $\ell(P_i) \geq 1$ are independent and have distribution $\Lambda_+(\mathrm{d}f)/\kappa_\ell(\mathcal{K})$; they are realizations of the corresponding ℓ-Pareto process, see §2.4. We coin

the term *elliptical ℓ-Pareto process* since the tails of its finite-dimensional distributions correspond to elliptical distributions with a Pareto-distributed radial variable R in (5.2). The finite-dimensional dependence structure associated to d sites $\boldsymbol{s} = (s_1, \ldots, s_d)$ is characterized by the exponent function (Nikoloulopoulos et al., 2009; Opitz, 2013)

$$\begin{aligned} V(\boldsymbol{z}) &= -\log \mathrm{P}\{Z^\star(s_1) \leq z_1, \ldots, Z^\star(s_d) \leq z_d\} \\ &= \sum_{j=1}^{d} z_j^{-1} t_{\alpha+1}\left\{(z_{-j}/z_j)^{1/\alpha}; \Sigma_{-j,j}, (\alpha+1)^{-1}\left(\Sigma_{-j,-j} - \Sigma_{-j,j}\Sigma'_{-j,j}\right)\right\}, \end{aligned} \tag{5.5}$$

with the correlation matrix $\Sigma = \{\rho(s_{j_1}, s_{j_2})\}_{1\leq j_1, j_2 \leq d}$ related to the correlation function ρ, and with $t_\alpha(\cdot; \mu, \Sigma)$ the cumulative distribution function of a multivariate t distribution with α degrees of freedom and parameters $\boldsymbol{\mu}$ (median vector) and Σ (scatter matrix). Dependence structures of Brown–Resnick type can be interpreted as a special case of extremal-t dependence, arising asymptotically when $\alpha \to \infty$. The truncation of W_i at zero in (5.3) implies that the measure Λ_+ of an extremal-t process has positive mass on the set $\{f \in C_+(\mathcal{K}) \setminus \{0\} : \min_{s\in\mathcal{K}} f(s) = 0\}$, which is not the case for Brown–Resnick processes.

Densities and partial derivatives of the exponent function

Ribatet (2013) gives the density $\lambda_{\boldsymbol{s}}$ of $\Lambda_{\boldsymbol{s}}$,

$$\begin{aligned} \lambda_{\boldsymbol{s}}(\boldsymbol{y}) = \alpha^{1-d}\pi^{(1-d)/2}|\Sigma_{\boldsymbol{s}}|^{-1/2}\Gamma\{(\alpha+1)/2\}^{-1}\Gamma\{(\alpha+d)/2\} \\ \times \prod_{j=1}^{d} |y_j|^{1/\alpha-1}\{T_{1/\alpha}(\boldsymbol{y})^T\Sigma_{\boldsymbol{s}}^{-1}T_{1/\alpha}(\boldsymbol{y})\}^{-(\alpha+d)/2}, \quad \boldsymbol{y} \in \mathbb{R}^d, \end{aligned}$$

where $\Sigma_{\boldsymbol{s}} = \{\rho(s_{j_1}, s_{j_2})\}_{1\leq j_1, j_2 \leq d}$ denotes the finite-dimensional correlation matrix according to the correlation function ρ of the extremal-t dependence structure relative to the sites $\boldsymbol{s} = (s_1, \ldots, s_d)$, and $T_{1/\alpha}$ is a transformation defined as $T_a(x) = \mathrm{sign}(x)|x|^a$ for $a > 0$. The density $\lambda_{+,\boldsymbol{s}}$ of $\Lambda_{+,\boldsymbol{s}}$ on $(0,\infty)^d$ is equal to $\lambda_{\boldsymbol{s}}$. The partial derivatives V_{I_k} of the exponent function V are calculated by integrating $\lambda_{\boldsymbol{s}}$ with respect to the components in the set complementary to I_k. The integration is carried out using conditional intensities.

Given a collection $\boldsymbol{s}_0 = (s_{0,1}, \ldots, s_{0,d'})$ of d' conditioning locations with values $\boldsymbol{y}_0$, the conditional intensity $\lambda_{\boldsymbol{s}|\boldsymbol{s}_0,\boldsymbol{y}_0}(\boldsymbol{y}) = \lambda_{(\boldsymbol{s},\boldsymbol{s}_0)}(\boldsymbol{y}, \boldsymbol{y}_0)/\lambda_{\boldsymbol{s}_0}(\boldsymbol{y}_0)$ equals (Ribatet, 2013)

$$\begin{aligned} \lambda_{\boldsymbol{s}|\boldsymbol{s}_0,\boldsymbol{y}_0}(y) = \alpha^{-d}\pi^{-d/2}(d+\alpha)^{-d/2}|\tilde{\Sigma}|^{-1/2}\Gamma\{(\alpha+d')/2\}^{-1}\Gamma\{(\alpha+d+d')/2\} \\ \times \prod_{j=1}^{d} |y_j|^{1/\alpha-1}\left[1 + \frac{\{T_{1/\alpha}(\boldsymbol{y}) - \tilde{\mu}\}'\tilde{\Sigma}^{-1}\{T_{1/\alpha}(\boldsymbol{y}) - \tilde{\mu}\}}{d'+\alpha}\right]^{-(\alpha+d+d')/2}, \end{aligned} \tag{5.6}$$

with

$$\tilde{\mu} = \Sigma_{\boldsymbol{s}:\boldsymbol{s}_0}\Sigma_{\boldsymbol{s}_0}^{-1}T_{1/\alpha}(\boldsymbol{y}_0), \quad \tilde{\Sigma} = \frac{T_{1/\alpha}(\boldsymbol{y}_0)^T\Sigma_{\boldsymbol{s}_0}^{-1}T_{1/\alpha}(\boldsymbol{y}_0)}{d'+\alpha}\left(\Sigma_{\boldsymbol{s}} - \Sigma_{\boldsymbol{s}:\boldsymbol{s}_0}\Sigma_{\boldsymbol{s}_0}^{-1}\Sigma_{\boldsymbol{s}_0:\boldsymbol{s}}\right),$$

where $\Sigma_{\boldsymbol{s}:\boldsymbol{s}_0}$ denotes the matrix of covariances between the random vectors corresponding to the location vectors $\boldsymbol{s}$ and $\boldsymbol{s}_0$. Expression (5.6) is the density of a random vector $T_\alpha(\boldsymbol{X})$, where $\boldsymbol{X}$

follows a d-dimensional elliptical t distribution with $d' + \alpha$ degrees of freedom and parameters $\tilde{\mu}$ and $\tilde{\Sigma}$.

Without loss of generality, we consider the partial derivative $V_{1:d'}(\boldsymbol{y})$ of V with respect to the indices 1 to d' such that $I_k = \{1, ..., d'\}$, obtained by calculating the integral of $\lambda_{\boldsymbol{s}_{(d'+1):d}|\boldsymbol{s}_{1:d'},\boldsymbol{y}_{1:d'}}(\boldsymbol{y}_{(d'+1):d})$ and by multiplying the resulting expression by $\lambda_{\boldsymbol{s}_{1:d'}}(\boldsymbol{y}_{1:d'})$. The required integral of the conditional density is $t_{d'+\alpha}(\boldsymbol{y}_{(d'+1):d}^{1/\alpha} - \tilde{\mu}; \boldsymbol{0}, \tilde{\Sigma})$. We get

$$-V_{1:d'}(\boldsymbol{y}) = t_{d'+\alpha}\left(\boldsymbol{y}_{(d'+1):d}^{1/\alpha} - \tilde{\mu}; \tilde{\Sigma}\right) \alpha^{1-d'} \pi^{(1-d')/2} |\Sigma_{1:d'}|^{-1/2} \Gamma\{(\alpha+1)/2\}^{-1}$$
$$\times \Gamma\{(\alpha + d')/2\} \left(\prod_{j=1}^{d'} |y_j|\right)^{1/\alpha - 1} \left\{(\boldsymbol{y}_{1:d'}^T)^{1/\alpha} \Sigma_{1:d'}^{-1} \boldsymbol{y}_{1:d'}^{1/\alpha}\right\}^{-(\alpha+d')/2}, \tag{5.7}$$

with

$$\tilde{\mu} = \Sigma_{(d'+1):d,1:d'} \Sigma_{1:d'}^{-1} \boldsymbol{y}_{1:d'}^{1/\alpha}$$

and

$$\tilde{\Sigma} = (d' + \alpha)^{-1} (\boldsymbol{y}_{1:d'}^T)^{1/\alpha} \Sigma_{1:d'}^{-1} \boldsymbol{y}_{1:d'}^{1/\alpha} (\Sigma_{(d'+1):d} - \Sigma_{(d'+1):d,1:d'} \Sigma_{1:d'}^{-1} \Sigma_{1:d',(d'+1):d}).$$

There exists positive mass on the boundary $\{\boldsymbol{y} \in \mathbb{R}_+^d \setminus \{\boldsymbol{0}\} : \|\boldsymbol{y}\|_\infty > 0\}$. Equation (5.7) gives the densities $\lambda_{+,\boldsymbol{s}}$ for a point $\boldsymbol{y}$ on the boundary of $\mathbb{R}_+^d \setminus \{0\}$: if $\boldsymbol{y}_{1:d'} > 0$ and $y_{(d'+1):d} = \boldsymbol{0}$, then the density on the corresponding subset of $\mathbb{R}_+^d \setminus \{\boldsymbol{0}\}$ is $-V_{1:d'}(\boldsymbol{y})$, see Coles and Tawn (1991, §3.1).

The results of an extensive simulation study in Thibaud and Opitz (2015) suggest that censored approaches are the best in practice, when the model is misspecified. Moreover, full likelihood inference improves estimation efficiency when the distribution of extremes is close to the limiting model, but a pairwise likelihood approach appears more robust to certain kinds of model misspecification. Thibaud and Opitz (2015) conduct an application to spatial precipitation extremes around the city of Zurich, Switzerland, where the elliptical Pareto models outperform the Brown–Resnick model on the boundary of the their parameter space. Conditional simulations from the selected model are then used to perform spatial mapping of conditional means and uncertainties during an extreme episode using observed measurements at 44 locations.

Exact simulation procedures for the extremal elliptical dependence

Based on the results in Thibaud and Opitz (2015), we now describe exact finite-dimensional simulation procedures for three processes related to elliptical extremes: max-stable extremal-t processes, elliptical Pareto processes and conditional processes when conditioning on fixed values at certain sites. The specificities of the elliptical structure in finite-dimensional distributions allow for fast and exact simulation algorithms. In the literature, there also exist alternative exact simulation algorithms for max-stable process classes including the extremal-t processes, proposed more recently by (Dombry et al., 2016), among others. In the recent preprint of Zhong et al. (2021), we have further extended such approaches to the wider class of max-infinitely divisible processes presented later in §6.2.

Due to the elliptical structure of the points $P_i^{1/\alpha}$ of the point process $\{P_i\}$ in (5.4), an equivalent representation of the finite-dimensional projection of an extremal-t process relative

to d sites $\boldsymbol{s} = (s_1, \ldots, s_d)$ is obtained by setting

$$P_{\boldsymbol{s},i} = \{\mathrm{E}(U_{1,1})_+^\alpha\}^{-1}(A_{\boldsymbol{s}}\boldsymbol{U}_i)_+^\alpha/Q_i, \tag{5.8}$$

with $A_{\boldsymbol{s}}$ the Cholesky root of $\Sigma_{\boldsymbol{s}} = A_{\boldsymbol{s}}A_{\boldsymbol{s}}^T$ and $\boldsymbol{U}_i = (U_{i,1}, \ldots, U_{i,d})'$ independent and identically distributed copies of some vector $\boldsymbol{U}$ uniformly distributed on the Euclidean unit sphere $\mathcal{S}_d$. This representation is a special case of Theorem 3.2 of Opitz (2013) and allows exact simulation of both max-stable and Pareto processes due to the boundedness $\|(A_{\boldsymbol{s}}\boldsymbol{U}_i)_+^\alpha\|_\infty \leq 1$.

In practice, max-stable processes are simulated using only a finite number of $P_{\boldsymbol{s},i}$. When a finite boundary $b < \infty$ exists for the components of $Q_iP_{\boldsymbol{s},i}$ such that $\mathrm{P}(\max_{i=1,2,\ldots} Q_iP_{\boldsymbol{s},i}(s_j) \leq b) = 1$ for $j = 1, ..., d$, exact simulation of $Z_{\boldsymbol{s}}^\star$ can be achieved from a finite number of points $P_{\boldsymbol{s},i}$ (Schlather, 2002, Theorem 4). Since the components of $\{\mathrm{E}(U_{1,1})_+^\alpha\}^{-1}(A_{\boldsymbol{s}}\boldsymbol{U}_i)_+^\alpha$ in (5.8) are always bounded by $b = \{\mathrm{E}(U_{1,1})_+^\alpha\}^{-1}$, exact simulation of extremal-t processes is possible. For $i \geq 1$, $\|P_{\boldsymbol{s},i}\|_\infty \leq b/Q_i$ with a decreasing sequence $\{Q_i\}_{i>0}$. If $\|\max_{i=1,\ldots,\tau_b} P_{\boldsymbol{s},i}\|_\infty \geq b/Q_{\tau_b}$ for some $\tau_b > 1$, then the points $P_{\boldsymbol{s},i}$ for $i > \tau_b$ cannot contribute to the maximum in (5.3) and we have $Z_{\boldsymbol{s}}^\star = \max_{i=1}^{\tau_b} P_{\boldsymbol{s},i}$. Two numerical limitations may restrict the applicability of this simulation approach: first, standard algorithms for determining the Cholesky root $A_{\boldsymbol{s}}$ of $\Sigma_{\boldsymbol{s}}$ require $\mathcal{O}(d^3)$ basic operations; second, b may be large if α or d are large, requiring the simulation of a very large number of points $P_{\boldsymbol{s},i}$. More precisely,

$$b = 2\pi^{1/2}\frac{\Gamma\{(d+\alpha)/2\}}{\Gamma\{(\alpha+1)/2\}\Gamma(d/2)} \approx 2^{1-\alpha/2}\pi^{1/2}\frac{(d+\alpha-2)^{\alpha/2}}{\Gamma\{(\alpha+1)/2\}}, \quad d \to \infty,$$

using Stirling's formula. In certain situations, notably when d indexes a fine spatial grid of points, these limitations are too restrictive. Then the conventional approach for approximate simulation can be used, where we assume that the distribution $W(s)$ has some finite upper endpoint used to stop the iterative simulation, leading to an approximation error in the simulated max-stable process. Since the tails of $W(s_j)_+^\alpha$ become heavier when α increases, the approximation error also increases.

The simulation of the points $P_{\boldsymbol{s},i}$ in (5.8) yields an algorithm for the simulation of ℓ-Pareto processes: the points $P_{\boldsymbol{s},i}$ with $\ell(P_{\boldsymbol{s},i}) \geq 1$ are independent realizations from the standard ℓ-Pareto processes; see Figure 5.1. Moreover, for $u_0 > 0$, the homogeneity of $\Lambda_{+,\boldsymbol{s}}$ implies that the points $u_0^{-1}P_{\boldsymbol{s},i}$ with $\ell(P_{\boldsymbol{s},i}) \geq u_0$ are also realizations from the standard ℓ-standard process, *i.e.*, we can use the POT stability of ℓ-Pareto processes. The existence of the upper bound b allows us to simulate all the points $P_{\boldsymbol{s},i}$ in a set $A = ([0, u]^d)^C \subset \mathbb{R}_+^d$ for $u > 0$. Since the set $\ell(y) \geq u_0$ is a subset of A for suitably chosen u, we can obtain exact simulations from every elliptical ℓ-Pareto process.

Instead of simulating the points of the Poisson process, it is possible to use an acceptance-rejection algorithm to generate realisations of $Y_\ell^\star(\boldsymbol{s})$ without dealing with a random number of realisations. First, we consider the simulation of a standard ℓ-Pareto process with $\ell(f) = \max_{j=1}^d f(s_j)$. We need $\{\mathrm{E}(U_{1,1})_+^\alpha\}^{-1}/Q_i \geq 1$ at a point P_i that satisfies $\ell(P_i) \geq 1$. On $[1, \infty)$, the distribution of a point $\{\mathrm{E}(U_{1,1})_+^\alpha\}^{-1}/Q_i$, conditional to an exceedance of 1, corresponds to a standard Pareto random variable R. Hence, any vector $R(A_{\boldsymbol{s}}\boldsymbol{U})_+^\alpha$ with $\boldsymbol{U}$ independent of R is a standard ℓ-Pareto process if $\ell\{R(A_{\boldsymbol{s}}\boldsymbol{U})_+^\alpha\} \geq 1$. When ℓ is different from the componentwise maximum, we proceed as before and fix $u_0 > 0$ such that $\max_{j=1}^d f(s_j) \geq 1$ whenever $\ell(f) \geq u_0$. Then the vector $u_0^{-1}R(A_{\boldsymbol{s}}\boldsymbol{U})_+^\alpha$, given that $\ell\{R(A_{\boldsymbol{s}}\boldsymbol{U})_+^\alpha\} \geq u_0$, is a realisation of $Y_\ell^\star(\boldsymbol{s})$. We

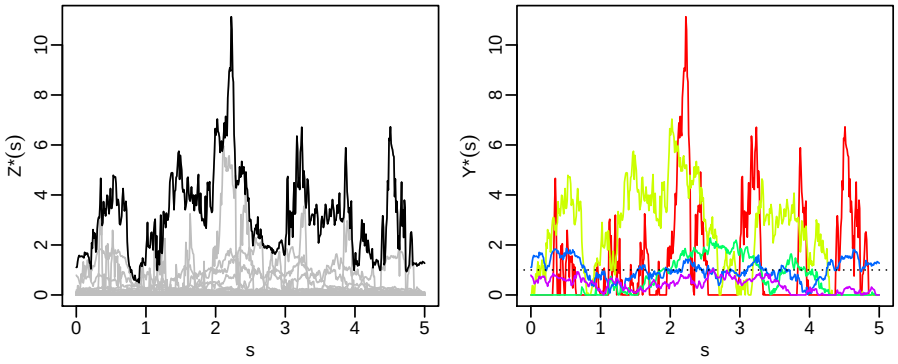


Figure 5.1: Simulation of extremal elliptical processes. Left: simulation (black line) from an extremal-t process with $\alpha = 1$ and $\rho(h) = \exp(-\|h\|)$. The grey lines show the points $P_{\boldsymbol{s},i}$ in the spectral decomposition (5.8). Right: independent simulations from the corresponding elliptical ℓ-Pareto process with $\ell(f) = \sup_{\boldsymbol{s}\in[0,5]} f(s)$ are given by the points $P_{\boldsymbol{s},i}$ with $\ell(P_{\boldsymbol{s},i}) \geq 1$.

can get a sample of $Y_\ell^\star(\boldsymbol{s})$ by repeatedly simulating random vectors $u_0^{-1} R(A_{\boldsymbol{s}}\boldsymbol{U})_+^\alpha$ and retaining only those vectors fulfilling the condition $\ell\{R(A_{\boldsymbol{s}}\boldsymbol{U})_+^\alpha\} \geq u_0$. To minimize the frequency of rejections, u_0 should be chosen as small as possible.

Whereas conditioning Λ on exceedances of $\ell(f)$ over 1 yields the distribution of the ℓ-Pareto process, one might instead be interested in the conditional distribution when values $\boldsymbol{y}_0 > \boldsymbol{0}$ for a collection of sites $\boldsymbol{s}_0 = (s_{0,1}, \ldots, s_{0,d'})$ are fixed. The finite-dimensional conditional distribution for the sites $\boldsymbol{s} = (s_1, \ldots, s_d)$, disjoint with $\boldsymbol{s}_0$, has density (5.6). The conditional process defined on $\mathcal{K} \setminus \{\boldsymbol{s}_0\}$ corresponds to a transformed t process that can easily be simulated.

5.1.3 Discussion of parametric inference with POT-stable processes

The class of ℓ-Pareto models for POT-modeling of stochastic processes sits on a solid theoretical foundation and provides flexibility with respect to the definition of extreme events through risk functionals. Their functional POT-stability may not always be satisfied at observed levels of data, but it provides a useful and elegant modeling framework. Statistical inference based on partial censoring of non-extreme components is efficient but leads to relatively high computational cost even for moderate numbers of components d (*e.g.*, less than 50). In Thibaud and Opitz (2015), we illustrated the estimation approach for extreme precipitations observed at 44 locations around Zurich, Switzerland. For higher-dimensional problems with a larger number of components, one has to resort to alternative estimation approaches such as pairwise likelihoods. Exact simulation and conditional simulation of elliptical ℓ-Pareto processes is straightforward and fast, which stands in contrast to the intricate conditional simulation procedures arising for max-stable processes (Dombry et al., 2013).

5.2 Semi-parametric resampling of spatial extremes

Resampling refers to simulation techniques used to generate new scenarios that realistically reproduce statistical features of observed data, such as trends or spatial dependence patterns.

These scenarios can be fed into process-based impact models (*e.g.*, models for property insurance, rainfall-runoff, agricultural yield, energy production) to study their sensitivity with respect to variations in the input scenarios. Examples of parametric models have been outlined in the previous sections, and they can be used for simulation if they appropriately represent the data-generating process. However, they impose relatively strong assumptions on the structure of the observed data. Moreover, parameter estimation may be very computer-intensive in the case of data with a large number of observations, especially in the spatial setting with many observation locations, such as gridded data. As an alternative, nonparametric resampling methods such as Multiple-Point Geostatistics (Mariethoz and Caers, 2014; Tahmasebi, 2018) have emerged in the geoscience literature for fast simulation using only minimal assumptions on the distribution of data. Nonparametric resampling allows the simulation of new datasets preserving important data features such as spatial patterns from observed datasets. Observations are often called *training data*, while one refers to resampled data as *simulation data*. Training data are usually given on some regular grid spanning over time, space or space-time. In this section, we discuss solutions to appropriately extrapolate statistical features of extreme events beyond the observed range of training data, which poses problems with many standard resampling algorithms when the variables of interest are continuous, *i.e.*, when they are not restricted to a finite number of levels. The use of nonparametric resampling techniques is hampered by their inability to produce simulations with new extreme events beyond the observed range of data values. We show how to combine such techniques with extreme-value theory of stochastic processes, *i.e.*, with asymptotic theory leading to ℓ-Pareto process limits, to extrapolate observed data towards yet unobserved high quantiles.

The asymptotic dependence structure in extreme-value limits is fundamentally nonparametric, even in the bivariate case, which is is contrast to the central limit theory for random vectors and stochastic processes, for which a single parameter – the linear correlation coefficient – fully characterizes the bivariate dependence in the Gaussian limits. Therefore, parametric models for extreme-value dependence may impose unrealistically restrictive assumptions, especially in the context of spatially indexed processes. Moreover, all of the asymptotic spatial extreme-value models that are commonly deployed in practice are based on Gaussian or log-Gaussian processes arising in the spectral processes used to construct the max-stable process in the spectral construction (2.17). Therefore, dependence in these extreme-value models inherits many of the properties of Gaussian dependence. Multiple-point statistics and related nonparametric resampling techniques such as Direct Sampling Mariethoz et al. (2012) aim to better represent complex spatial dependence patterns arising for more than two locations, as compared to the Gaussian models whose correlation functions describe behavior only between two locations.

In the new approach that we propose (Opitz et al., 2021), original data are first enriched with new values in the tail region, and then classical resampling algorithms are applied to enriched data. In a first approach to enrichment that we label "naive resampling", discussed in §5.2.1, we generate an independent sample of the marginal distribution while keeping the rank order of the observed data. We point out inaccuracies of this approach around the most extreme values, and we therefore develop a second approach that works for datasets with many replicates. It is based on the asymptotic representation of extreme events through two stochastically independent components: a magnitude variable, and a profile field describing spatial variation; recall the Definition (6) of ℓ-Pareto processes. To generate enriched data, we fix a target range of return levels of the magnitude variable, and we resample magnitudes constrained to this range.

5.2.1 Naive resampling

With *naive resampling*, simulation must take place over the same domain as the original dataset, and the same grid resolution must be used. Based on a stationarity assumption for marginal distributions, we estimate the marginal distribution using only minimal assumptions by combining a kernel density estimate for the body of the distribution with a simple tail model based on univariate EVT. We first compute the kernel density estimate (kde), and we then fix a high threshold to model the tail using a GPD. The kde is used to estimate the exceedance probability above the threshold, and by assuming continuity of the density around the threshold we obtain the value of the GPD scale parameter from the value of the kde-based density at the threshold. It remains to estimate the tail index, for which a variety of approaches can be implemented; see Opitz et al. (2021) for examples.

The idea for data enrichment is then relatively simple: we generate a new i.i.d. sample of the same size as the training sample according to the estimated marginal distribution, and the indices of the order statistics of the original and new samples remain the same. For instance, the maximum value of the new sample is at the same spatial location as the original maximum value. Then, standard resampling algorithms can be applied to the enriched dataset, in which the dependence structure has been preserved by keeping the ranks of the training data. Alternatively, one could also proceed in a slightly different way after the generation of the new i.i.d. sample: resampling could be applied to the dataset in which the actual values of the variable are replaced by their ranks in the training sample, and then we attribute the order statistics of the newly generated i.i.d. sample to the corresponding ranks in the resampled dataset of ranks.

If the stochastic process from which the training data have been generated is mixing, *i.e.*, if it has long-range independence in the observation window, simulated quantiles will correspond well to original quantiles in the central part of the distribution where data are dense, owing to the law of large numbers. Similar behavior is obtained for replicated data with a sufficiently large number of replicates, for instance independent replicates of a spatial process. Therefore, we can expect that resampling is faithful to statistical features in the training data as far as the central part of the distribution is concerned.

The main issue with naive resampling is the independent sampling of the new values in the tails of the distribution. By consequence, simulated quantiles can differ substantially from quantiles of the training data distribution close to the extremes. In particular, the spacings between order statistics $Y_{(i)}$ and $Y_{(i+1)}$ are relatively large when i is close to 1or to n. As a consequence, too large maxima will arise in the naive resampling approach, and it would be difficult to target specific ranges of return periods for summary statistics that are relatively strongly correlated with the maximum. Another consequence is that naive resampling tends to produce very strong spatial variability in the pixels having a value close to the maximum value $Y_{(n)}$. Therefore, even if a training image shows relatively smooth behavior around the pixel containing the maximum value, naive resampling will tend to produce a relatively rugged surface around the maximum in simulated images. Finally, we underline that the naive resampling procedure depends strongly on the grid resolution. While measures of effective sample size for dependent data may be comparable when studying the same process at different spatial resolutions, they can strongly vary in the case of an i.i.d. sample. The impact of i.i.d. resampling can be assessed using theoretical results such as Renyi's theorem, see Opitz et al. (2021) for details.

5.2.2 Lifting using functional extreme-value theory

For appropriate extrapolation of extreme values while preserving the spatial coherence in training data, we propose to make use of EVT for stochastic processes, where generalized Pareto processes arise in the limit, as recalled in §2.4. We assume that data have been generated by a stochastic process with realizations in the space $\mathcal{C}(\mathcal{K})$ of continuous functions over a compact domain $\mathcal{K}$. Given a homogeneous aggregation functional ℓ and a relatively large threshold u, the POT convergence in (2.26) suggests using the following POT assumption:

$$\{u^{-1}X^\star(s)\} \mid \ell(\boldsymbol{X}^\star) > u \quad \stackrel{d}{=} \quad \{Y_\ell^\star(s)\}, s \in \mathcal{K},$$

where $\boldsymbol{Y}_\ell^\star$ is an ℓ-Pareto process as given in Definition 6. Then, we can exploit the following *scale-profile decomposition* of Pareto processes:

$$Y_\ell^\star(s) = RW(s), \quad R \perp \boldsymbol{W}, \quad R \sim \text{Pareto}(1,1), \quad s \in \mathcal{K},$$

where $\ell(\boldsymbol{W}) = 1$, *i.e.*, the *profile* $\boldsymbol{W}$ is a random element on the functional "unit sphere" with respect to the functional ℓ. The *scale* R is a random variable, independent of $\boldsymbol{W}$.

Given an aggregation functional ℓ, the steps of the resampling procedure including data enrichment with respect to extreme event episodes can be summarized as follows, where the necessity of the post-processing Step 6 will be explained in the following §:

1. Estimate marginal distributions, for instance as described in the previous §5.2.1.
2. Apply the marginal probability integral transform based on the marginal distribution from Step 1 to standardize the marginal distribution of the original data.
3. Extract observed extreme episodes based on exceedances of the aggregation functional ℓ.
4. Decompose extreme episodes into their empirical scale and profile components.
5. Enrich data by lifting extreme episodes using newly sampled scale variables according to a target value or a target range of return levels of the aggregation functional.
6. Post-process non extreme values in enriched data.
7. Perform nonparametric resampling using enriched extremes.
8. Backtransform margins using the marginal distribution from Step 1.

Details of the steps of the above algorithm are discussed in Opitz et al. (2021).

Postprocessing of non extreme values

A practical difficulty with the lifting approach, which also arises with many other methods of multivariate and spatial extreme-value modeling, concerns the treatment of small and moderately large observation values. Asymptotic theory does not uniquely determine how such values should be modified when lifting the extreme sample fraction. Whereas in theory we consider sample sizes that increase to infinity, such that non-extreme values are compressed to the lower

bound of the support of the limit distribution due to the marginal location-scale normalization in (2.12), in practice we have to cope with finite and fixed sample sizes. In this lifting approach, we remedy this issue by proposing a rescaling of non-extreme standardized values based on a minimum entropy principle to avoid unrealistic artefacts in simulations. Given a marginal threshold u_{marg} below which we do not trust the asymptotic model (*e.g.*, the quantile at 80%), we do not directly lift the observations below the threshold during the data enrichment step. Instead, given that the resampled scale variable transforms the threshold u_{marg} to the value $\tilde{u}_{\text{marg}}$ and that the lower bound of the support of the standardized marginal distribution is $x^{\star}_{\text{lower}}$, we linearly rescale the observations below the threshold such that their support covers the full interval $[x^{\star}_{\text{lower}}, \tilde{u}_{\text{marg}}]$; see the details in Opitz et al. (2021).

5.2.3 Spatiotemporal extensions

The lifting procedure is adapted to the simulation of extreme episodes with spatial and temporal extent (*i.e.*, spanning several time steps) in Palacios-Rodriguez et al. (2020). For the case where extreme episodes are allowed to span several time steps with temporal dependence in the simulations, the paper discusses how to extract meaningful extreme episodes as training data. This involves choosing an aggregation functional used to aggregate standardized observations over space and time. This choice further requires setting the length of extreme episodes, for which exploratory statistics on temporal extremal dependence can be studied. In a similar manner, we can restrict the spatial extent of extreme episodes to areas smaller than the full study domain, in case where it is very large. Moreover, we focus on the particularities of precipitation data, for which intermittence (*i.e.*, absence of precipitation) may arise at some locations and time steps during an extreme episode, such that we have to appropriately handle the singular mass at 0 in the univariate marginal distribution of precipitation during the three consecutive tasks of marginal standardization, lifting and marginal backtransformation. Details can be found in Palacios-Rodriguez et al. (2020).

5.2.4 Application examples

We shortly illustrate the practical implementation and results of such resampling procedures for extreme event episodes of two distinct meteorological variables: temperatures in mainland France (Opitz et al., 2021), and precipitation in the Mediterranean region of France (Palacios-Rodriguez et al., 2020).

Heatwaves in mainland France

The temperature application is motivated by the extreme heatwaves observed during the summer of 2019 in France and over large parts of Europe. We use daily gridded data (8 km resolution) from the SAFRAN reanalysis of Météo France for years 2010–2016. We standardize marginal distributions of data to a uniform scale on $(-1, 0)$ by using a pixel-specific transformation combining a kde for the bulk with the GPD for the tail. We put focus on days with high temperatures over relatively large areas of the French territory by using the median as a spatial aggregation functional. Another advantage of the median is that the ranking of extreme episodes

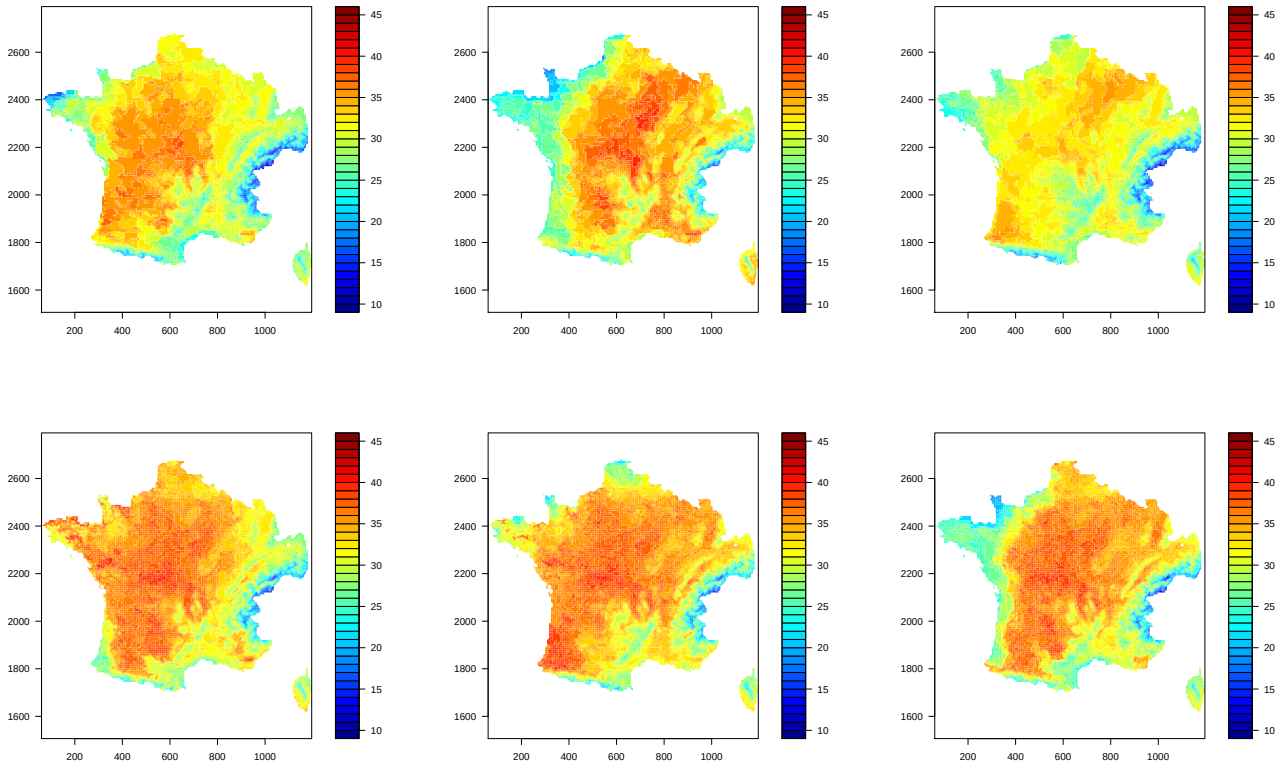


Figure 5.2: Semiparametric resampling of heatwave episodes in France. Top row: examples of selected extreme training data for lifting. Bottom row: examples of simulations obtained through the combination of lifting and nonparametric resampling.

does not depend on the choice of the standardized distribution, which is uniform here to establish a relatively close link to copula modeling.

Then, a day is considered as an extreme episode if the median of standardized observations is in the tail of the distribution of all the observed median values. We consider only data for the summer months from June to September. After applying a declustering step to avoid selecting days from the same heatwave, we keep the 6 most extreme days as training data, and we aim to simulate new spatial extreme episodes. We here apply the lifting step to obtain enriched data composed of 6 episodes lifted to correspond to a return period of approximately 10 years. For this step, we have to estimate the ℓ-extremal coefficient θ_ℓ in $\mathrm{P}(\ell(\boldsymbol{X}) \geq r) \approx -\theta_\ell r$, and we use an estimate $\hat{\theta}_\ell$ obtained by matching empirical and theoretical exceedance probability of the aggregation functional ℓ for a relatively high quantile. Direct Sampling (DS Mariethoz et al., 2010) is then performed on the 6 lifted episodes to be able to produce an essentially infinite number of new scenarios. In DS, we start with an "empty" simulation image and then we iteratively fill the simulation as follows: first, randomly selected an empty pixel in the simulation image to be filled; then, scan for a pixel in the training data possessing a very similar "neighborhood" of already simulated pixels; finally, fill the empty pixel with the selected training data value. As criterion to check if a neighborhood in the training data matches the already simulated parts of the neighborhood of a pixel to fill in the simulation image, we use a distance

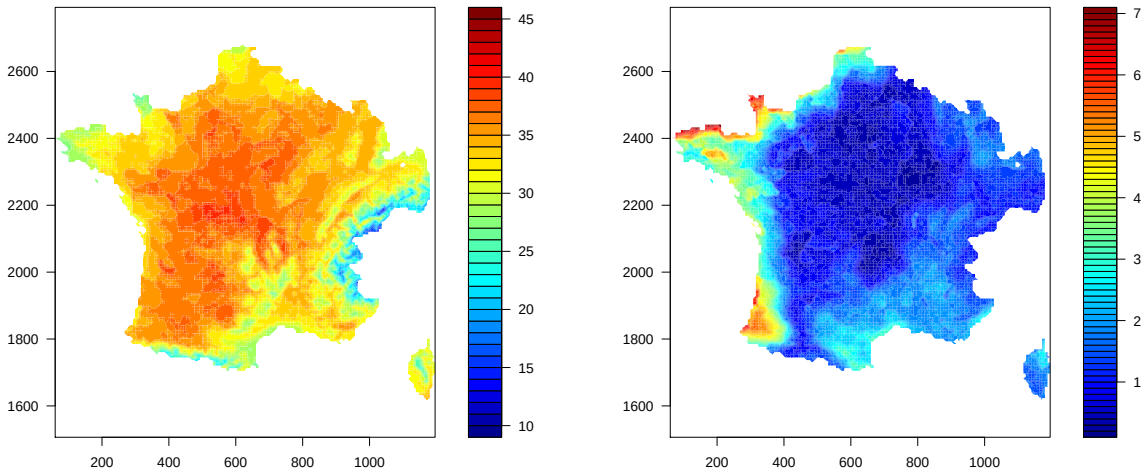


Figure 5.3: Summary statistics based on semiparametric resampling simulations of extreme events. Pixel-wise empirical means (left) and standard deviations (right) calculated over 100 simulations of heatwaves in France with 10-year return period.

criterion weighted at 90% by the nearest-neighbor pattern of data values and at 10% by the spatial coordinates. The weight on spatial coordinates ensures that we we fill a pixel in the simulation image with the value of a training pixel that is located not too far away. This allows us to take into account spatial nonstationarities in the dependence patterns of extreme values. To summarize mean and variance of simulations obtained with this stochastic generator, we have computed pixel-wise empirical mean and variances over 100 final simulations in Figure 5.3; the resulting maps highlight the strong spatial nonstationarities in the distribution of extreme events.

To check if various summary statistics of data are appropriately captured by the simulation approach, we implement a procedure similar to cross-validation, where we use a larger number of the most extreme episodes from the original data. Summary statistics considered here are the minimum, maximum, median, mean, range and interquartile range of the values for all the spatial pixels during an extreme episode. The procedure is based on complete extrapolation since we do not use the 20 most extreme episodes (in terms of the aggregation functional) for generating the simulations but only lower-ranked episodes. By simulating new episodes for return levels above the quantile of ℓ that separates validation and training episodes, we can compare the distribution of simulated summary statistics to the validation sample of the 20 observed summary statistics; see Figure 5.4. While it is difficult to perfectly match all summaries in complete extrapolation, we can conclude that the simulation method provides plausible new scenarios of extreme heatwaves.

Mediterrean precipitation episodes in France

In Palacios-Rodriguez et al. (2020), we consider gridded hourly precipitation reanalysis data for the south of France for the period 1997–2007. This Mediterranean region of France is known

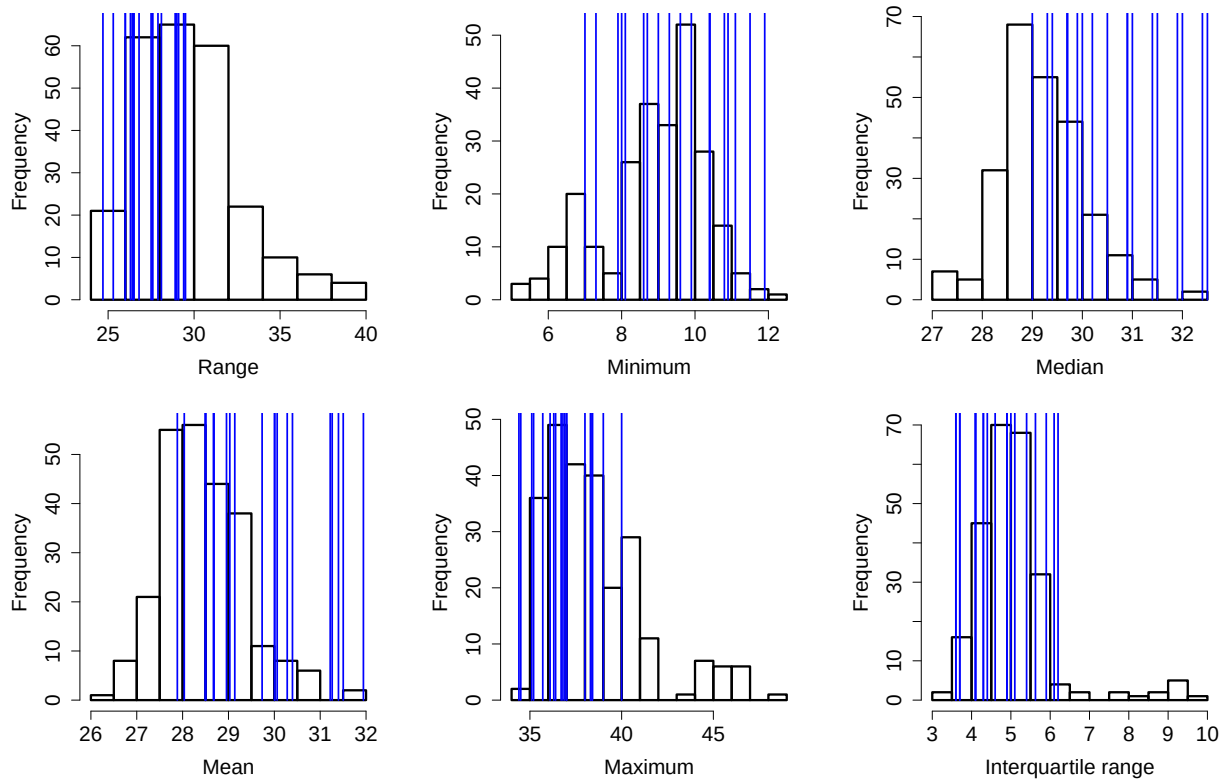


Figure 5.4: Validation study with summary statistics calculated for each extreme episode. Histograms show summary statistics of 250 simulations in a setting of complete extrapolation. Blue lines indicate the summary values of validation data consisting of the 20 most extreme observed episodes. From left to right: range, minimum, median (top row); mean, maximum, interquartile range.

for violent precipitation episodes during autumn, which are triggered when warm and humid air enters the land from the sea and then clashes with colder mountainous regions. While we do not apply nonparametric resampling on lifted extreme space-time episodes in this application, we give a detailed analysis of how the choice of two spatiotemporal different aggregation functionals (maximum, median) can lead to very different results in terms of risk measures calculated for a large ensemble of lifted episodes. Whereas the maximum aggregation puts focus only on the largest value and allows capturing very localized high peaks of precipitation, the median aggregation puts focus on the typical amount of precipitation during an extreme episode and is stronger oriented towards taking into account high values of space-time accumulated precipitation, for instance over a catchment area.

Figure 5.5 highlights differences in two risk measures (quantile, conditional tail expectation) with respect to the type of aggregation functional and the quantile level used for calculating these risk measures. The quantile levels are specific to each episode and correspond to a quantile of the values occurring during the episode; for instance, the quantile level 98% means that the risk measure focuses on the 2% of pixels with the most extreme values. We observe important differences in the values of each of the two risk measures along each of the three dimensions of comparison (quantile level of the risk measure; target return level of the aggregation functional; type of aggregation functional).

5.2.5 Discussion of resampling approaches with extremes

A benefit of naive resampling is that it can be applied to data that do not have any structure of independent replication, such as spatial datasets without temporal replicates. With this approach, EVT is used only for the modeling of univariate marginal distributions. While simulations reproduce the behavior around the global maximum of training data in a rather crude manner, synthetic data examples in Opitz et al. (2021) show that this approach may still be relevant and useful in practice thanks to its simplicity, especially in cases where the strength of extremal dependence in the training data is relatively weak.

The more sophisticated lifting procedure anchored in EVT for dependent extremes exploits POT stability to jointly lift observed extreme-event patterns to more extreme magnitudes. As already outlined earlier, this stability property may not be satisfied by many meteorological and environmental data of interest, where often a weakening of dependence strength is observed with increasing event magnitudes. Nevertheless, the asymptotic stability may still be considered as a useful and easily interpretable working assumption in such cases, although certain results may have to interpreted with care when the aim of simulation is to achieve extrapolation very far beyond the observed quantiles. In such cases, simulated extremal dependencies may be too strong when extrapolating very far into the tail, which typically leads to a conservative assessment of the impact of joint risks.

Validity checks with respect to the assumptions underlying resampling procedures are important but notoriously difficult in the extreme-value setting due to the scarcity of extreme data. The POT property is useful since it allows checking the stability of data summaries across different threshold levels. It allows implementing checks of the soundness of extrapolation above the observed range of values by retaining the largest extreme episodes as a validation set, as done in our example. We suggest a more systematic use of this approach as commendable practice in extreme-value studies.

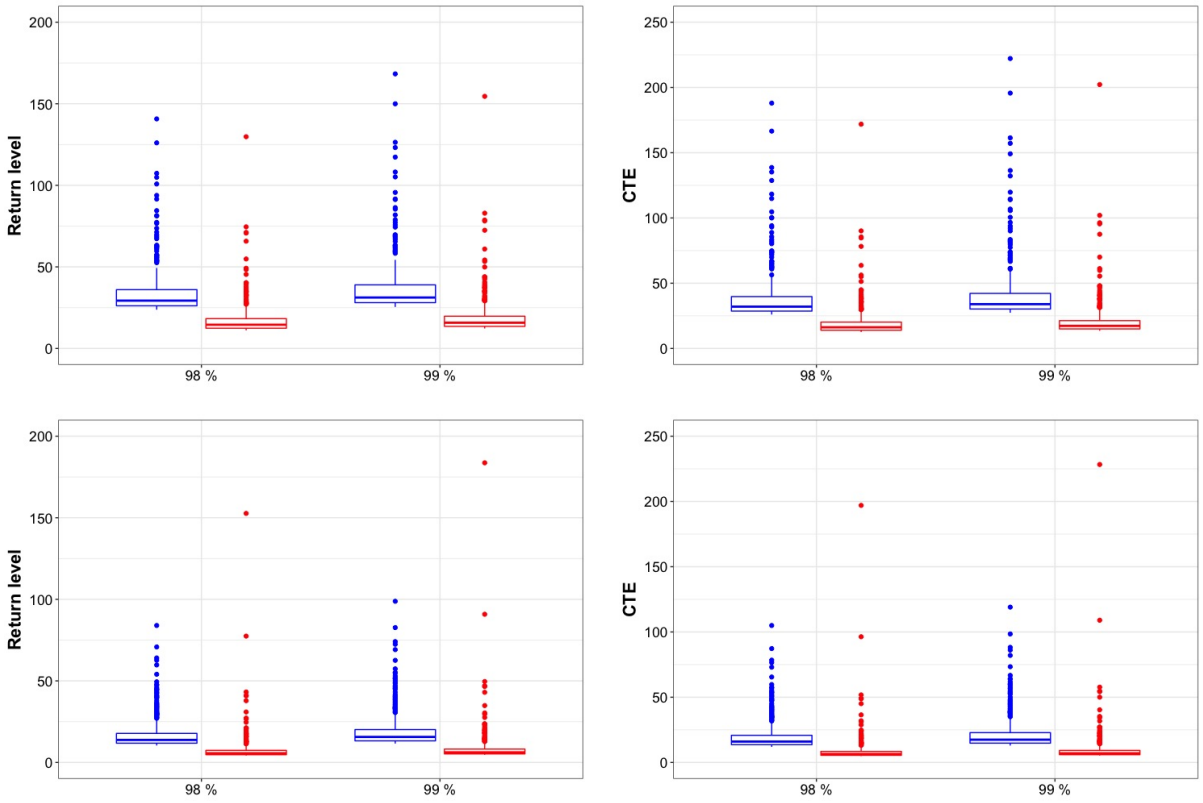


Figure 5.5: Comparison of risk measures for extreme spatiotemporal precipitation episodes in the south of France. Boxplots are based on 500 extreme episodes simulated through lifting of the 50 most extreme observed episodes. First row: spatiotemporal median aggregation functional. Second row: spatiotemporal maximum aggregation functional. Red boxplots: target return level of at least 6 months; blue boxplots: target return level of at least 10 years.

In cases where the assumption of POT-stability is not warranted, an interesting extension of the above approach could be developed based on subasymptotic modeling frameworks, such as the conditional extremes approach (Heffernan and Tawn, 2004; Wadsworth and Tawn, 2019). However, stronger semi-parametric assumptions may be required in this setting.

Finally, the combination of parametric assumptions suggested by EVT with nonparametric resampling techniques can be seen as an approach that paves the way towards more general hybrid parametric-nonparametric resampling techniques.

Chapter 6

Subasymptotic spatial-temporal extreme-value modeling

The previous chapter was centered on modeling and simulation tools for asymptotic models featuring stability properties in marginal distributions and the dependence structure. As already explained before, the dependence stability is often not observed in real datasets, such that co-occurrences of extreme values at higher quantiles become relatively less frequent. For instance, the spatial extent and temporal duration of extreme episodes could become smaller and shorter, respectively, when higher magnitudes are observed. This implies that the cluster size of extremes depends on event magnitude. Bivariate diagnostics and more flexible subasymptotic extremal dependence models for multivariate data have been developed since the mid of the 1990s (Ledford and Tawn, 1996; Coles et al., 1999; Heffernan and Tawn, 2004), see also the contributions discussed in Chapter 4, but the extension to models that systematically address relatively fast joint tail decays in the spatial and spatiotemporal setting is more recent. This chapter summarizes several contributions to this very active area of research. In asymptotic models, the peaks-over-threshold stability can be expressed through a scale-profile decomposition of the ℓ-Pareto limit processes with a Pareto-distributed scale variable, as outlined in §2.4. The construction principle of the subasymptotic models below is based on replacing the Pareto-distribution of the scaling variable by more flexible alternatives. Moreover, we will also discuss more specific subasymptotic models that strongly capitalize on hierarchical constructions.

6.1 Modeling threshold exceedances with Gaussian scale mixture processes

The pseudo-polar representations of multivariate limit distributions (*i.e.*, the scale-profile decomposition of generalized Pareto limits; recall Definition 6) have motivated Wadsworth et al. (2017) to explore how more flexible transitions between dependence classes can be achieved through a common random scaling applied to a random vector on the unit sphere, the latter being defined from a norm on $\mathbb{R}^2$. The work in Engelke et al. (2019), summarized in §4.2, provides an in-depth analysis of bivariate dependence structures arising from such constructions. In this section, we use the pseudo-polar representation of multivariate elliptical distributions, such as the multivariate Gaussian, and we argue that a flexible and natural extension of this approach

to spatial modeling consists in using the wide class of randomly scaled Gaussian processes, also known as Gaussian scale mixtures. This class comprises all infinite-domain processes possessing finite-dimensional nondegenerate elliptical distributions (Huang and Cambanis, 1979). In exceedance-based modeling, the gain in tail flexibility as compared to Gaussian models or asymptotic models for extremes could allow fixing lower thresholds, thus improving statistical efficiency. In Gaussian scale mixtures, the Gaussian correlation structure can be viewed as a mechanism to capture certain properties of the bulk of the distribution, like the range of dependence, while the random scale parameters give separate control over the joint tail decay rates.

A main purpose of this approach is to propose new, parsimonious and flexible subasymptotic dependence models to achieve a smooth transition between asymptotic independence and asymptotic dependence. We pay strong attention to appropriately capturing the tail decay in asymptotically independent scenarios while keeping a highly flexible asymptotically dependent submodel. The asymptotically dependent submodels in the class of Gaussian scale mixtures are closely related to the elliptic Pareto process of Thibaud and Opitz (2015), see §5.1, while the asymptotically independent counterparts described below provide more flexible parametric alternatives. Parameter inference allows the data to provide evidence about the asymptotic dependence class without fixing it a priori.

The results in this section are published in Opitz (2016); Huser et al. (2017).

6.1.1 Gaussian scale mixture processes

To create flexible spatial models, we define a Gaussian scale mixture process (*i.e.*, a Gaussian process with random variance) as follows:

$$X(s) = RW(s), \quad s \in \mathcal{S} \subset \mathbb{R}^D, \tag{6.1}$$

where $W(s)$ is a standard Gaussian process with correlation function $\rho(s_1, s_2)$, and RF is a positive random variable with distribution F and density f, independent of $W(s)$. Conditional on R, the random process $X(s)$ is Gaussian with zero mean and variance R^2. Gaussian processes arise as a special case when $R = r_0$ almost surely for some $r_0 > 0$. We propose to use the copula associated to (6.1) as a model for extremal dependence, *i.e.*, we only use the dependence structure for modeling extremes, but we may modify the marginal distributions by transforming them in a monotonically strictly increasing manner.

Finite-dimensional distributions

When the process (6.1) is observed at d spatial locations $s_1, \ldots, s_d \in \mathcal{S}$, we write $X_j = X(s_j)$ and $W_j = W(s_j)$, $j = 1, \ldots, d$, yielding the random vectors $\boldsymbol{X} = (X_1, \ldots, X_d)^T$ and $\boldsymbol{W} = (W_1, \ldots, W_d)^T$. From (6.1), one has the representation

$$\boldsymbol{X} = R\boldsymbol{W}, \qquad R \sim F \perp \boldsymbol{W} \sim \mathcal{N}_d(0, \boldsymbol{\Sigma}), \tag{6.2}$$

where Σ is a correlation matrix determined by the spatial configuration of sites. The Gaussian scale mixture vectors given in (6.1) correspond to elliptically contoured distributions (Cambanis

et al., 1981); see B.1. By conditioning on R, we deduce that the distribution G and density g of $\boldsymbol{X}$ are

$$G(\boldsymbol{x}) = \int_0^\infty \Phi_d(\boldsymbol{x}/r; \Sigma) f(r) \mathrm{d}r, \qquad g(\boldsymbol{x}) = \int_0^\infty \phi_d(\boldsymbol{x}/r; \Sigma) r^{-d} f(r) \mathrm{d}r, \qquad \boldsymbol{x} \in \mathbb{R}^d, \tag{6.3}$$

where $\Phi_d(\cdot; \Sigma)$ and $\phi_d(\cdot; \Sigma)$, respectively, denote the d-variate Gaussian distribution and density with zero mean and covariance matrix Σ. Some non-trivial choices of the mixing density $f(r)$ lead to a closed-form expression of the density $g(\boldsymbol{x})$, including the Student-t, Laplace and slash models (Kotz et al., 2005). In general, the unidimensional integrals in (6.3) can be accurately approximated using numerical integration. Marginal distributions G_k and their corresponding densities g_k, $k = 1, \ldots, d$, are

$$G_k(x_k) = \int_0^\infty \Phi(x_k/r) f(r) \mathrm{d}r, \qquad g_k(x_k) = \int_0^\infty \phi(x_k/r) r^{-1} f(r) \mathrm{d}r, \qquad x_k \in \mathbb{R}, \tag{6.4}$$

where $\Phi(\cdot) = \Phi_1(\,\cdot\,; 1)$ and $\phi(\cdot) = \phi_1(\,\cdot\,; 1)$ denote the univariate standard Gaussian distribution and density, respectively. The use of a censored likelihood requires the partial derivatives of the distribution G in (6.3), which can be expressed as univariate integrals where the integrand involves conditional Gaussian distribution functions (which can be estimated without bias (Genz and Bretz, 2009)) and probability densities; see Huser et al. (2017) for details.

Conditional distributions and simulation algorithm

We develop an efficient algorithm for conditional simulation of Gaussian scale mixtures, which is crucial for prediction and estimation of complex functionals in spatial and spatiotemporal settings. For $\boldsymbol{X} = (\boldsymbol{X}_1^T, \boldsymbol{X}_2^T)^T = R(\boldsymbol{W}_1^T, \boldsymbol{W}_2^T)^T$ a d-dimensional Gaussian scale mixture partitioned into subvectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ of dimensions $d_1 \in \{1, \ldots, d-1\}$ and $d_2 = d - d_1$, respectively, we first derive the conditional distributions of $\boldsymbol{X}_2$ given $\boldsymbol{X}_1$ and of the latent variable R given $\boldsymbol{X}_1$. Let $\Sigma_{i,j}$, $i, j \in \{1, 2\}$, denote the corresponding blocks of the covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{W} = (\boldsymbol{W}_1^T, \boldsymbol{W}_2^T)^T$ and $\boldsymbol{\Sigma}_{i|j} = \Sigma_{i,i} - \Sigma_{i,j}\Sigma_{j,j}^{-1}\Sigma_{j,i}$.

Proposition 10 (Conditional distributions)**.** *The conditional distribution of $\boldsymbol{X}_2$ given $\boldsymbol{X}_1 = \boldsymbol{x}_1$ is elliptic with density*

$$f_{\boldsymbol{X}_2|\boldsymbol{X}_1=\boldsymbol{x}_1}(\boldsymbol{x}_2) = c_0^{-1} |\boldsymbol{\Sigma}_{2|1}|^{-1/2} h_d \left\{ (\boldsymbol{x}_2 - \boldsymbol{\mu}_{2|1})^T \Sigma_{2|1}^{-1} (\boldsymbol{x}_2 - \boldsymbol{\mu}_{2|1}) + c_1 \right\}, \qquad \boldsymbol{x}_2 \in \mathbb{R}^{d_2}, \tag{6.5}$$

where $\boldsymbol{\mu}_{2|1} = \boldsymbol{\Sigma}_{2,1}\boldsymbol{\Sigma}_{1,1}^{-1}\boldsymbol{x}_1$, $c_0 = A_{d_2} \int_0^\infty h_d(r^2 + c_1) r^{d_2 - 1} \mathrm{d}r$, $c_1 = \boldsymbol{x}_1^T \Sigma_{1,1}^{-1} \boldsymbol{x}_1$ and

$$h_d(t) = A_d^{-1} t^{(1-d)/2} f_{RR_{\boldsymbol{W}}}(\sqrt{r}), \qquad f_{RR_{\boldsymbol{W}}}(r) = \int_0^\infty F(s^{-1}) \left\{ f_{R_{\boldsymbol{W}}}(rs) + f'_{R_{\boldsymbol{W}}}(rs) rs \right\} \mathrm{d}s.$$

It has pseudo-polar representation $\boldsymbol{\mu}_{2|1} + R_{2|1} \Sigma_{2|1}^{1/2} \boldsymbol{U}$ with $\Sigma_{2|1} = \Sigma_{2|1}^{1/2} \Sigma_{2|1}^{T/2}$, $\|\boldsymbol{U}\|_2 = 1$ and radius $R_{2|1}$ whose density is $f_{R_{2|1}}(r) = A_{d_2}\, c_0^{-1} r^{d_2 - 1} h_d(r^2 + c_1)$, $r > 0$. The conditional density of R given $\boldsymbol{X}_1 = \boldsymbol{x}_1 \in \mathbb{R}^{d_1}$, with $1 \leq d_1 \leq d$, is

$$f_{R|\boldsymbol{X}_1=\boldsymbol{x}_1}(r) = r^{-d_1} f(r) \phi_{d_1}(\boldsymbol{x}_1/r; \Sigma_{1;1}) / g(\boldsymbol{x}_1), \qquad r \geq 0. \tag{6.6}$$

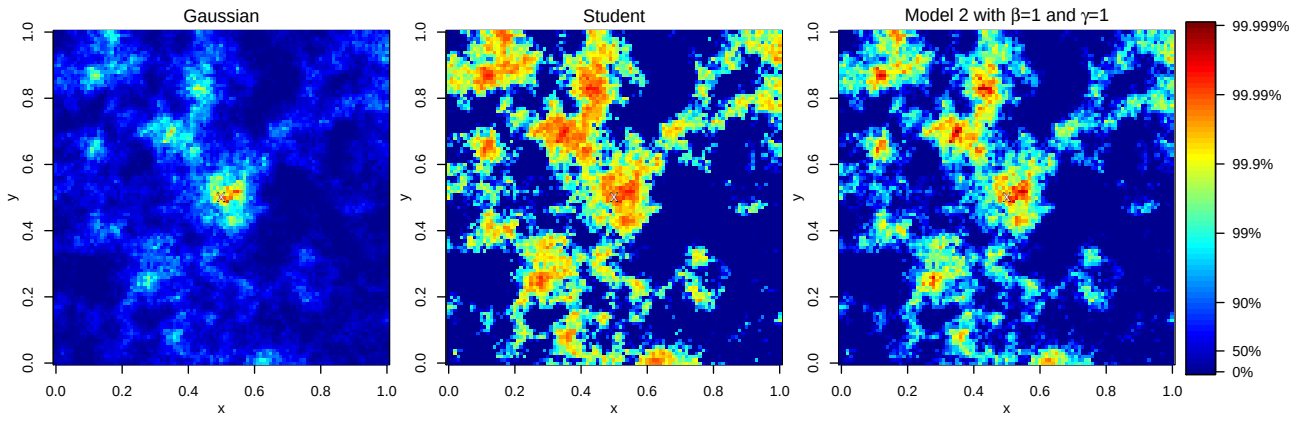


Figure 6.1: Conditional simulations from an asymptotically independent Gaussian process (left), an asymptotically dependent Student-t process with 3 degrees of freedom (middle), and from one of the newly proposed Gaussian scale mixture processes with $\beta = \gamma = 1$ leading to asymptotic independence (right), displayed using exponential marginal distributions. The underlying Gaussian field has exponential correlation function $\rho(s_1, s_2) = \exp(-\|s_1 - s_2\|/0.1)$. Simulation is done conditional to the central grid cell at $(0.5, 0.5)$, whose value is fixed to the 99.99%-quantile. The conditional simulation of the scale variable $\tilde{R}$ is based on a Metropolis–Hastings algorithm. The three conditional simulations of the Gaussian component $\tilde{\boldsymbol{W}}$ use the same random seed.

Based on the above result, simulation of $\boldsymbol{X}_2$ conditional on $\boldsymbol{X}_1 = \boldsymbol{x}_1$ can be done either by directly simulating the elements of the pseudo-polar representation $\boldsymbol{\mu}_{2|1} + R_{2|1}\Sigma_{2|1}^{1/2}\boldsymbol{U}$, or by exploiting the latent Gaussian structure in a two-step procedure. To simulate $R\boldsymbol{W}_2$ conditional on $R\boldsymbol{W}_1 = \boldsymbol{x}_1$, we first generate a realization $\tilde{r}$ of the conditional scale variable $\tilde{R}$ according to its density $f_{R|\boldsymbol{X}_1=\boldsymbol{x}_1}$ in (6.6), and we then sample a realization $\tilde{\boldsymbol{w}}_2$ of $\boldsymbol{W}_2$ conditional on $\tilde{R} = \tilde{r}$ and $\boldsymbol{X}_1 = \boldsymbol{x}_1$, *i.e.*, we sample $\tilde{\boldsymbol{w}}_2$ according to the conditional Gaussian distribution $\boldsymbol{W}_2 \mid \boldsymbol{W}_1 = \boldsymbol{x}_1/\tilde{r}$ with mean $\boldsymbol{\mu}_{2|1}/\tilde{r}$ and covariance matrix $\Sigma_{2|1}$. Then, $\tilde{r}\tilde{\boldsymbol{w}}_2$ is a realization of the conditional vector $\boldsymbol{X}_2$ given $\boldsymbol{X}_1 = \boldsymbol{x}_1$.

For an illustration of conditional simulation, Figure 6.1 displays conditional realizations of three Gaussian scale mixture models, where the impact of choosing different random scale distributions emerges clearly.

6.1.2 Flexible spatial dependence with Gaussian scale mixtures

Bivariate dependence properties of Gaussian scale mixtures can be obtained from the general results for random scale constructions in §4.2; see Example 2. For a regularly varying scale distribution F we get asymptotically dependent extremal-t limit processes (Opitz, 2013). We now outline two models for flexible modeling of threshold exceedances in the spatial setting, which are discussed in detail in Opitz (2016) and Huser et al. (2017), along with several other models. The following more specific result for Weibull-tailed random scale variables, established in Huser et al. (2017), will motivate the so-called *HOT model* and also applies to the *Laplace model*. It states that asymptotic independence arises for Weibull-tailed scale variables, with flexible expressions for the dependence mesure $\overline{\chi}$.

Proposition 11 (Weibull-tailed Gaussian scale mixtures)**.** *Suppose that*

$$P(R \geq r) = 1 - F(r) \sim \alpha r^{\gamma} \exp(-\delta r^{\beta}), \qquad r \to \infty, \tag{6.7}$$

for some constants $\alpha > 0$, $\beta > 0$, $\gamma \in \mathbb{R}$ *and* $\delta > 0$. *Then* $\chi = 0$ *and*

$$\overline{\chi} = 2\left\{(1+\rho)/2\right\}^{\beta/(\beta+2)} - 1.$$

The joint tail can be written as

$$P(F_{X_1} > 1 - 1/x, F_{X_2} > 1 - 1/x) = \mathcal{L}(x) x^{-2/(1+\overline{\chi})}, \quad x \to \infty,$$

where $\mathcal{L}(x) \sim K \log(x)^{(1-1/\eta)\frac{2\gamma+\beta}{2\beta}+1/(2\eta)-1}$ *is a slowly varying function as* $x \to \infty$ *and* K *is a positive constant depending on* α, β, γ *and* δ.

The HOT model

The HOT model, which was named as such in the literature after the authors' initials in Huser et al. (2017), bridges asymptotic dependence and independence. It can generate any value of $\overline{\chi} \in [\rho, 1]$ for fixed Gaussian correlation $\rho < 1$. Its novel two-parameter Weibull-type distribution F with support $[1, \infty)$ contains the Dirac mass at 1 as limiting case, yielding asymptotically independent standard Gaussian processes, and the Pareto distribution as boundary case, yielding asymptotic dependence. The distribution of the scale variable R with parameters $\beta \geq 0$ and $\gamma > 0$ is defined as

$$F(r) = \begin{cases} 1 - \exp\left\{-\gamma(r^{\beta} - 1)/\beta\right\}, & \beta > 0, \\ 1 - r^{-\gamma}, & \beta = 0, \end{cases} \quad r \geq 1. \tag{6.8}$$

The distribution (6.8) forms a continuous parametric family with respect to β since the term $(r^{\beta} - 1)/\beta$ converges to $\log r$ as $\beta \downarrow 0$. It is the result of applying the inverse Box–Cox transform with power β to an exponential variable with rate γ. The type of asymptotic dependence is determined by the value of β. When $\beta > 0$, (6.8) coincides with the tail representation (6.7) with the same tail parameter β, yielding asymptotic independence. When $\beta = 0$ or $\beta \downarrow 0$, the variable R is Pareto distributed with $F(r) = 1 - r^{-\gamma}$, $r \geq 1$, yielding asymptotic dependence. The Dirac mass at 1 is obtained as $\beta \to \infty$ or as $\gamma \to \infty$. The benefit of this model is to provide a smooth transition from asymptotic independence to asymptotic dependence with $\overline{\chi} \uparrow 1$ for $\beta \downarrow 0$ and $\gamma > 0$ and ρ fixed; moreover, it still keeps a smooth transition from asymptotic dependence to asymptotic independence with $\overline{\chi} \downarrow 0$ as $\gamma \downarrow 0$ and $\beta = 0$ and ρ are fixed, leading to a Gaussian limit.

The Laplace model

The Laplace model, proposed prior to the HOT model in Opitz (2016), uses the distribution $F(r) = 1 - \exp(-r^2/2))$, $r > 0$, known as Rayleigh distribution. Therefore, it is asymptotically independent with Weibull index $\beta = 2$. An interesting feature of the corresponding Gaussian scale mixture model is that its upper and lower univariate tails are part of the generalized Pareto family. Indeed, $\mathrm{P}(X(s) > x) = 0.5 \exp(-x)$ for $x > 0$, which defines the tail of a generalized

Pareto distribution with tail index $\xi = 0$. More generally, owing to the sum stability of elliptical distribution families, all nondegenerate weighted sums of components of $\boldsymbol{X}$ again have GPD tails with $\xi = 0$, such that using this model in the extreme-value context is very natural and convenient.

Moreover, it is possible to characterize the joint tail behavior with respect to how extrapolation is done for various aggregations of a random vector with this Gaussian scale mixture structure. The following proposition from Opitz (2016) summarizes the extrapolation of probabilities for exceedance sets, where some extrapolation relationships are exact and not only asymptotic. We focus on four useful choices that are common in the literature. Given a fixed threshold u, a vector $\boldsymbol{x}$ is a *marginal exceedance* in x_1 if $x_1 \geq u$, it is a *sum exceedance* if $\sum_{j=1}^{d} x_j \geq u$, it is a *max exceedance* if $\max_{j=1,\ldots,d} x_j \geq u$, and it is a *min exceedance* if $\min_{j=1,\ldots,d} x_j \geq u$. For the corresponding exceedance sets, we write $A_{x_1}(u) = \{\boldsymbol{x} \mid x_1 \geq u\}$, $A_{\text{sum}}(u) = \{\boldsymbol{x} \mid \sum_j x_j \geq u\}$, $A_{\max}(u) = \{\boldsymbol{x} \mid \max_j x_j \geq u\}$ and $A_{\min}(u) = \{\boldsymbol{x} \mid \min_j x_j \geq u\}$.

Proposition 12 (Extrapolation of exceedance probabilities in the Laplace model)**.** *For $\boldsymbol{X} \sim \mathcal{L}(\Sigma)$ a d-dimensional centered Laplace vector with dispersion matrix Σ, $u > 0$ a threshold and $\boldsymbol{t} = (t, \ldots, t) > 0$ defining a translation $\boldsymbol{X} - \boldsymbol{t}$ of the vector, we have the following properties:*

$$
\begin{aligned}
P(\boldsymbol{X} - \boldsymbol{t} \in A_{\text{sum}}(u)) &= \exp\left(-\frac{d}{\sqrt{\boldsymbol{e}^T \Sigma \boldsymbol{e}}} \times t\right) P(\boldsymbol{X} \in A_{\text{sum}}(u)), \\
P(\boldsymbol{X} - \boldsymbol{t} \in A_{x_1}(u)) &= \exp\left(-t/\sqrt{\sigma_{11}}\right) P(\boldsymbol{X} \in A_{x_1}(u)), \\
P(\boldsymbol{X} - \boldsymbol{t} \in A_{\max}(u)) &\sim \exp\left(-t/\sigma\right) P(\boldsymbol{X} \in A_{\max}(u)), \quad u \to \infty,
\end{aligned}
$$

where $\sigma = \sqrt{\max(\sigma_{11}, \ldots, \sigma_{dd})}$ in the last equation. When $\Sigma = \Sigma^\star$ is a correlation matrix, we further have

$$
P(\boldsymbol{X} - \boldsymbol{t} \in A_{\min}(u)) \sim \exp\left(-\sqrt{\boldsymbol{e}^T (\Sigma^\star)^{-1} \boldsymbol{e}} \times t\right) P(\boldsymbol{X} \in A_{\min}(u)), \quad u \to \infty.
$$

6.1.3 Statistical inference

To estimate the extremal dependence structure from the observed high spatial threshold exceedances, we use a two-step procedure that is often used with likelihood models for dependence in extremes: marginal distributions are estimated nonparametrically based on ranks, and then dependence parameters are estimated using a full likelihood with partial censoring to prevent estimates from being influenced by low and moderate values; see the closely related approach advocated for elliptical Pareto processes in §5.1 and the general estimation principle recalled in §1.3.4. Results from Thibaud and Opitz (2015) and Huser et al. (2017) suggest that it provides a reasonable compromise between bias and variance compared to alternative approaches using threshold exceedances. Even faster inference of tail behavior is possible when censoring is not partial but is only applied in the case where no component exceeds its marginal threshold. In a simulation study in Huser et al. (2017), we have demonstrated good performance of the partial censoring approach for the HOT model.

6.1.4 Applications

Two applications of spatial modeling of extreme wind speeds through the above Gaussian scale mixture models have been implemented in Opitz (2016); Huser et al. (2017). We here describe

only the approach used in Huser et al. (2017) where we analyze hourly wind speed extremes recorded during 2012–2014 in the Pacific Northwest, US, a region with large wind energy resources. Data are available year-round at 20 meteorological towers. To avoid modeling complex spatiotemporal nonstationary patterns, we restrict attention to winter months for the 12 stations located on the East side of the Cascade mountain range; see the upper panel of Figure 6.2. Selected data comprise up to 6504 hourly observations at each site. Wind patterns are mainly characterized by easterly and westerly winds.

We fitted Gaussian scale mixtures of type (6.8), with or without parameter constraints such as $\beta = 0$, and we further fitted a range of natural competitor models. A model comparison based on likelihood-based criteria for goodness-of-fit, and on CRPS for spatial prediction, indicates superior performance of the most complex HOT model considered for this application. The use of conditional simulation of the selected model based on Proposition 10 is illustrated for spatial prediction in the lower panel of Figure 6.2.

6.1.5 Discussion of spatial subasymptotic POT modeling

A number of approaches with mechanisms similar to Gaussian scale mixtures have been proposed in the recent literature to construct subasymptotic spatial peaks-over-threshold models (Huser and Wadsworth, 2019; Castro-Camilo et al., 2020). A strong benefit of Huser and Wadsworth (2019) for statistical inferencer is that the transition between asymptotic dependence and asymptotic independence takes place in the interior of the parameter space; however, this comes at the price of a construction of the dependence model that is more copula-like and less intuitive as compared to Gaussian scale mixtures. Castro-Camilo et al. (2020) implement peaks-over-threshold inference for the so-called *Gaussian factor copulas* of Krupskii et al. (2018), a model that is a variant of Gaussian location mixtures shortly discussed in §4.2; it leads to asymptotic dependence with limits of the Brown–Resnick type (Kabluchko et al., 2009).

6.2 Modeling maxima with max-infinitely divisible processes

Max-stable processes are the natural models for spatial extremes when considering componentwise maxima data, as recalled in §2.4. Max-stable models are appropriate when asymptotic stability properties in the dependence structure are satisfied at the observed levels, and they have been established as useful statistical models in practice (Davison et al., 2012, 2019). As pointed out in §2.5 and the preceding section, a fast growing body of empirical studies of environmental and climatic extremes in the literature has provided evidence that the asymptotic stability arising asymptotically is often violated at finite levels, and that the spatial dependence strength is weakening as events become more extreme (see, e.g., Huser et al., 2017; Tawn et al., 2018; Huser and Wadsworth, 2019; Bacro et al., 2019; Castro-Camilo et al., 2020; Bopp et al., 2020). In particular, under asymptotic independence, maxima become ultimately independent at the highest levels, whichs calls for more specialized models to capture the decay rate towards independence. In this setting, the commonly used max-stable models are not able to capture the rate of joint tail decay, and to estimate joint extremal probabilities beyond observed levels.

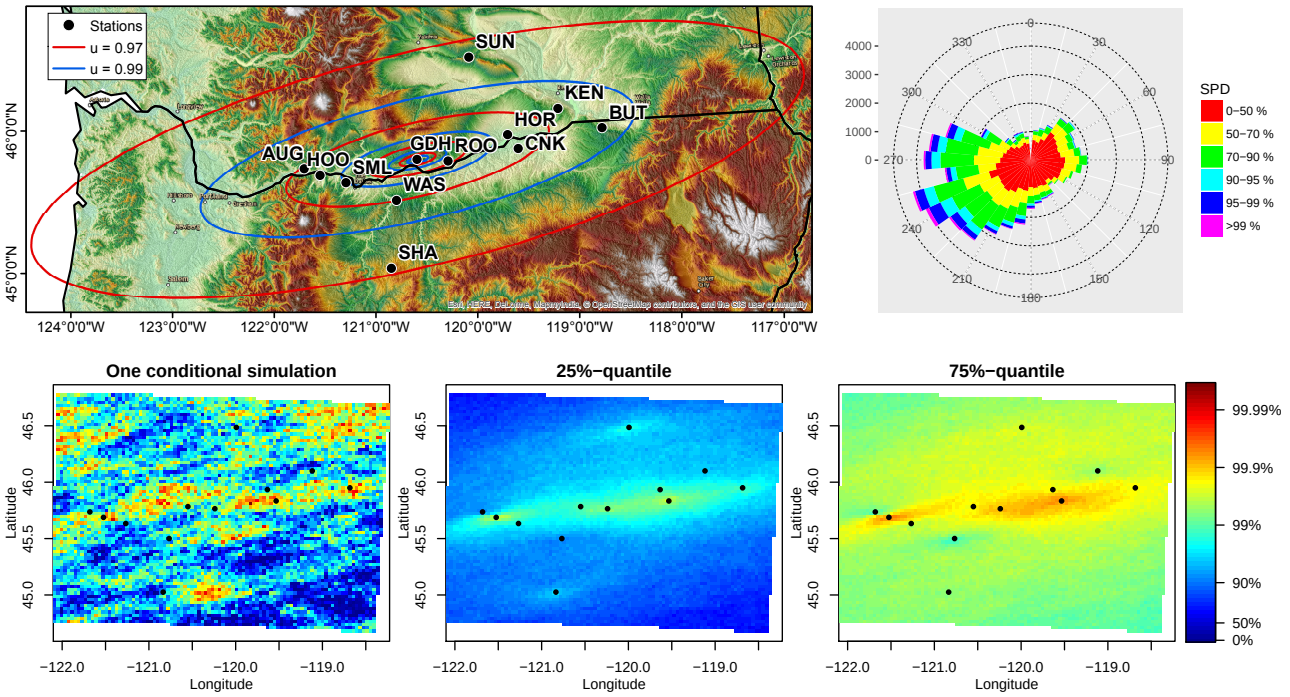


Figure 6.2: Application of spatial peaks-over-threshold modeling to wind speeds with Gaussian scale mixtures. Top left display: Topographic map with meteorological towers and US state boundaries (black). Overlaid ellipses are centered at Goodnow Hills (GDH) and correspond to the isocontours of the fitted subasymptotic dependence coefficients $\chi(u) = 0.5, 0.4, 0.3, 0.2$ (from the center outward) using the thresholds $u = 0.97$ (red) and $u = 0.99$ (blue), for the best geometrically anisotropic model of Huser et al. (2017). Top right display: Wind rose of winter wind speeds for the 12 stations, preliminarily transformed to the uniform scale. The color scale corresponds to different marginal quantile ranges. Bottom left display: Conditional simulation for the best model found in Huser et al. (2017) with conditioning values observed at the twelve stations on February 22, 2012, a day of very strong wind. Bottom middle and right displays: corresponding 25% and 75%-conditional quantiles based on 500 simulations. The color scale in the bottom row indicates quantile levels.

Figure 6.3 empirically illustrates weakening dependence strength at higher levels through level-dependent extremal coefficients estimated for wind speed maxima at 30 weather stations in the Netherlands, using different block sizes. Especially with relatively small blocks (daily, weekly), data show a relatively strong evolution towards larger extremal coefficients (corresponding to weaker extremal dependence) at higher quantiles.

We therefore propose a more flexible modeling framework based on the class of max-infinitely divisible processes, which extend max-stable processes while retaining dependence properties that are natural for maxima. We develop two parametric constructions for max-infinitely divisible models, which relax the max-stability property but remain close to popular max-stable models obtained as special cases. The first model considers maxima over a finite, random number of independent observations, while the second model generalizes the spectral representation (2.17) of max-stable processes. By analogy with max-stable processes, inference can be

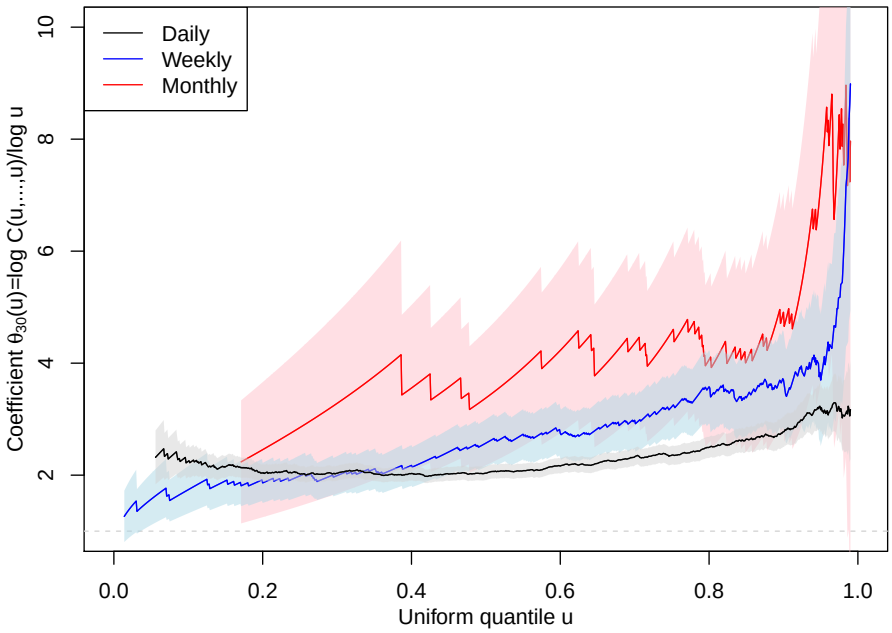


Figure 6.3: Empirical level-dependent extremal coefficients for different block sizes (solid curves), as defined in Equation (6.14), for Netherlands wind speed data, plotted as a function of uniform quantiles u. Shaded areas are 95%-confidence intervals.

performed using a pairwise likelihood.

6.2.1 Definition and Poisson process construction

A multivariate distribution function G on $\mathbb{R}^d$, $d \geq 1$, is max-infinitely divisible (max-id) if and only if G^t (with $G^t(z) = G(z)^t$) is a valid distribution function on $\mathbb{R}^d$ for any $t > 0$. A max-id distribution G describes the componentwise maximum of m independent random variables with distribution $F = G^{1/m}$, for any $m = 1, 2, \ldots$. This property permits to switch from the joint distribution G of the componentwise maximum over a given time unit to alternative time units and in particular to the distribution F of the original events. Unlike the univariate case, multivariate distributions are not always max-id. However, any monotonically increasing marginal transformation of a max-id distribution preserves the max-id property. We will exploit a constructive characterization of max-id distributions based on Poisson processes (Resnick, 1987, Chapter 5). For simplicity, we focus on multivariate distributions, and skip some technical details in the following.

We consider a Poisson point process (PPP) defined on the domain $E = [l_1, \infty] \times \ldots \times [l_d, \infty] \subseteq [-\infty, \infty]^d$ with a locally finite mean measure $\Lambda \geq 0$. Given the Poisson points

$$\{\boldsymbol{X}_i,\, i = 1, \ldots, N\} \sim \mathrm{PPP}(\Lambda), \quad N \in \mathbb{N}_0 \cup \{\infty\}, \tag{6.9}$$

we can define a random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_d)^{\mathrm{T}} \in \mathbb{R}^d$ as the componentwise maximum over the Poisson points $\boldsymbol{X}_i$ and the lower endpoint $\boldsymbol{l}$, *i.e.*,

$$\boldsymbol{Z} = \max\big(\max_{i=1}^{N} \boldsymbol{X}_i, \boldsymbol{l}\big). \tag{6.10}$$

The value $\boldsymbol{Z} = \boldsymbol{l}$ arises when $N = 0$, *i.e.*, when the Poisson process contains no points in $E \setminus \{\boldsymbol{l}\}$. Any max-id distribution can be represented constructively as in (6.10), and this approach can be extended to max-id processes (Giné et al., 1990). The joint distribution function of $\boldsymbol{Z}$ is

$$G(\boldsymbol{z}) = \exp\left\{-\Lambda\left([-\boldsymbol{\infty}, \boldsymbol{z}]^{\mathrm{c}}\right)\right\}, \ \boldsymbol{z} \in E, \qquad G(\boldsymbol{z}) = 0, \ \boldsymbol{z} \in E^{\mathrm{c}}, \tag{6.11}$$

where $\boldsymbol{z} = (z_1, \ldots, z_d)^{\mathrm{T}}$ and $[-\boldsymbol{\infty}, \boldsymbol{z}] = [-\infty, z_1] \times \cdots \times [-\infty, z_d]$. By analogy with max-stable distributions, the measure Λ is called the exponent measure or mean measure of G, and $V(\boldsymbol{z}) = \Lambda([-\boldsymbol{\infty}, \boldsymbol{z}]^{\mathrm{c}})$ is called the exponent function. The Poisson process representation is helpful for intuitive interpretation, modeling and simulation of max-id distributions and processes.

6.2.2 Dependence properties

Max-id random vectors $\boldsymbol{Z}$ are associated(Resnick, 1987, Proposition 5.29), such that a certain form of positive dependence prevails. Extremal dependence is closely related to the tail behavior of the exponent measure Λ since

$$1 - G(\boldsymbol{z}) = 1 - \exp\left\{-\Lambda\left([-\boldsymbol{\infty}, \boldsymbol{z}]^{\mathrm{c}}\right)\right\} \sim \Lambda\left([-\boldsymbol{\infty}, \boldsymbol{z}]^{\mathrm{c}}\right), \quad \min_{j=1,\ldots,d} z_j \to \infty. \tag{6.12}$$

If a max-id distribution G with exponent measure Λ is used to model the componentwise maximum over m independent random vectors with distribution F such that $F^m = G$, then

$$F(\boldsymbol{z}) = G^{1/m}(\boldsymbol{z}) = \exp\left\{-\Lambda\left([-\boldsymbol{\infty}, \boldsymbol{z}]^{\mathrm{c}}\right)/m\right\}, \tag{6.13}$$

which gives the first-order tail approximation $1 - F(\boldsymbol{z}) \approx \Lambda([-\boldsymbol{\infty}, \boldsymbol{z}]^{\mathrm{c}})/m$ when $\boldsymbol{z}$ has large components and/or m is large. therefore, the extremal dependence structures of F, G and Λ are of the same form.

The dependence strength may be summarized by the level-dependent extremal coefficient. For any d-dimension random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_d)^{\mathrm{T}}$ with joint distribution F, assumed to be continuous for simplicity and possessing marginal distributions $F_1, \ldots, F_d$, we define the level-dependent coefficient $\theta_d(u)$, for probability level $u \in (0, 1)$, as

$$\theta_d(u) = \frac{\log\left[F\left\{F_1^{-1}(u), \ldots, F_d^{-1}(u)\right\}\right]}{\log(u)}, \quad u \in (0, 1). \tag{6.14}$$

For max-id distributions, $\theta_d(u) = -\Lambda^{\star}\left([\boldsymbol{0}, (u, \ldots, u)]^{\mathrm{c}}\right)/\log u$. This coefficient can be interpreted as the equivalent number of independent variables amongst the d variables at the probability level $u \in (0, 1)$. Furthermore, in the bivariate case we obtain the link to the tail correlation measure $\chi(u) =\sim 2 - \theta_2(u)$ as $u \to 1$.

6.2.3 Construction principles

Useful max-id models can be built either by (i) directly specifying the exponent measure Λ in (6.9), or (ii) defining the points $\boldsymbol{X}_i$ constructively in the representation (6.10), or (iii) exploiting the fact that max-id distributions arise as limits of F_m^m as $m \to \infty$ where the distributions F_m are not necessarily identical. We propose two new general construction principles: we can follow (i) by defining a finite exponent measure Λ, or we follow (ii) and directly define the points $\boldsymbol{X}_i$ in (6.10), generalizing the spectral representation of max-stable processes in (2.17) with infinite exponent measure Λ.

Models with finite exponent measure Λ

Using a finite exponent measure $\Lambda = cH$ parametrized by an arbitrary probability distribution H on E and a constant $c > 0$, the resulting max-id vector $\boldsymbol{Z}$ has joint distribution

$$G_{c,H}(\boldsymbol{z}) = \exp[-c\{1 - H(\boldsymbol{z})\}], \ \boldsymbol{z} \in E, \qquad G(\boldsymbol{z}) = 0, \ \boldsymbol{z} \in E^{\mathrm{c}}, \tag{6.15}$$

and can be interpreted as the componentwise maximum over a finite number N of independent events, where N follows the Poisson distribution with mean c. As Λ is finite and the event $\{N = 0\}$ has probability $\exp(-c) > 0$, this yields positive mass at the lower boundary $\boldsymbol{l}$. In practice, this singularity is rather a nuisance than a relevant model feature, and we may restrict c to the range $[c_0, \infty)$ with a relatively large value of $c_0 > 0$, to ensure that $\exp(-c) \approx 0$. Once a parametric model for H has been chosen, the additional parameter c refines the tail behavior of G as compared to that of H and adds flexibility. Consider the distribution $F = G_{c,H}^{1/m}$ of the original observations, for some fixed $m > 0$. Using (6.13),

$$1 - F(\boldsymbol{z}) = 1 - G_{c,H}^{1/m}(\boldsymbol{z}) = 1 - \exp[-(c/m)\{1 - H(\boldsymbol{z})\}] \sim (c/m)\{1 - H(\boldsymbol{z})\}, \tag{6.16}$$

as $m \to \infty$ and/or $\min_{j=1}^{d} z_j \to \infty$, so that the constant c controls the tail weight of F with respect to that of H. The finite measure model $G_{c,H}$ in (6.15) therefore interpolates between the tail behavior of H (for $c = 1$) and that of the max-stable limit of H for $c \to \infty$, which is a useful feature when modeling sub-asymptotic block maxima. In the spatial context, this construction generalizes to max-id processes constructed as the pointwise maximum $Z(s) = \max\{X_1(s), \ldots, X_N(s), l(s)\}$, where $X_1(s), \ldots, X_N(s)$ are independent realizations of $X(s)$ conditionally on N, and $l(s)$ is their lower bound function. Relevant choices for the process $X(s)$ include Gaussian processes or Student-t processes, for which efficient implementations of routines to compute multivariate distribution functions exist, or more general elliptical processes. When $X(s)$ is Gaussian, then $Z(s)$ is asymptotically independent, and when $X(s)$ is Student-t with $\alpha > 0$ degrees of freedom, then $Z(s)$ is asymptotically dependent with the max-stable extremal-t limit process (Opitz, 2013).

Generalized spectral construction

To prevent the singularity at the lower endpoint $\boldsymbol{l}$ in max-id models with finite exponent measures, we mimick the spectral representation of max-stable processes in (2.17), but we use a more flexible Poisson point process intensity for the random scales $\{R_i\} > 0$. This allows proposing parametric models that can smoothly bridge asymptotic dependence and independence through a mechanism very similar to the Gaussian scale mixtures discussed in §6.1.1, and with the same joint dependence properties. Let $W_i(s)$ be independent copies of a random process $W(s)$ with $0 < \mathrm{E}[\max\{W(s), 0\}] < \infty$, independent of $\{R_i\}$. We construct the max-id process as

$$Z(s) = \max_{i=1,2,\ldots} R_i W_i(s), \qquad s \in \mathcal{S} \subset \mathbb{R}^D, \qquad 0 < \{R_i\} \sim \mathrm{PPP}(\kappa_{\boldsymbol{\gamma}}), \tag{6.17}$$

where the mean measure $\kappa_{\boldsymbol{\gamma}}$, parametrized by the vector $\boldsymbol{\gamma} \in \Gamma \subset \mathbb{R}^q$, is such that $\kappa_{\boldsymbol{\gamma}}([0, \infty)) = \infty$ but $\kappa_{\boldsymbol{\gamma}}([r, \infty)) < \infty$ for any $r > 0$. We specify $\kappa_{\boldsymbol{\gamma}}$ in order to recover max-stable models as a special case. As negative values of $W_i(s)$ do not contribute to the maximum $Z(s)$ in (6.17) we

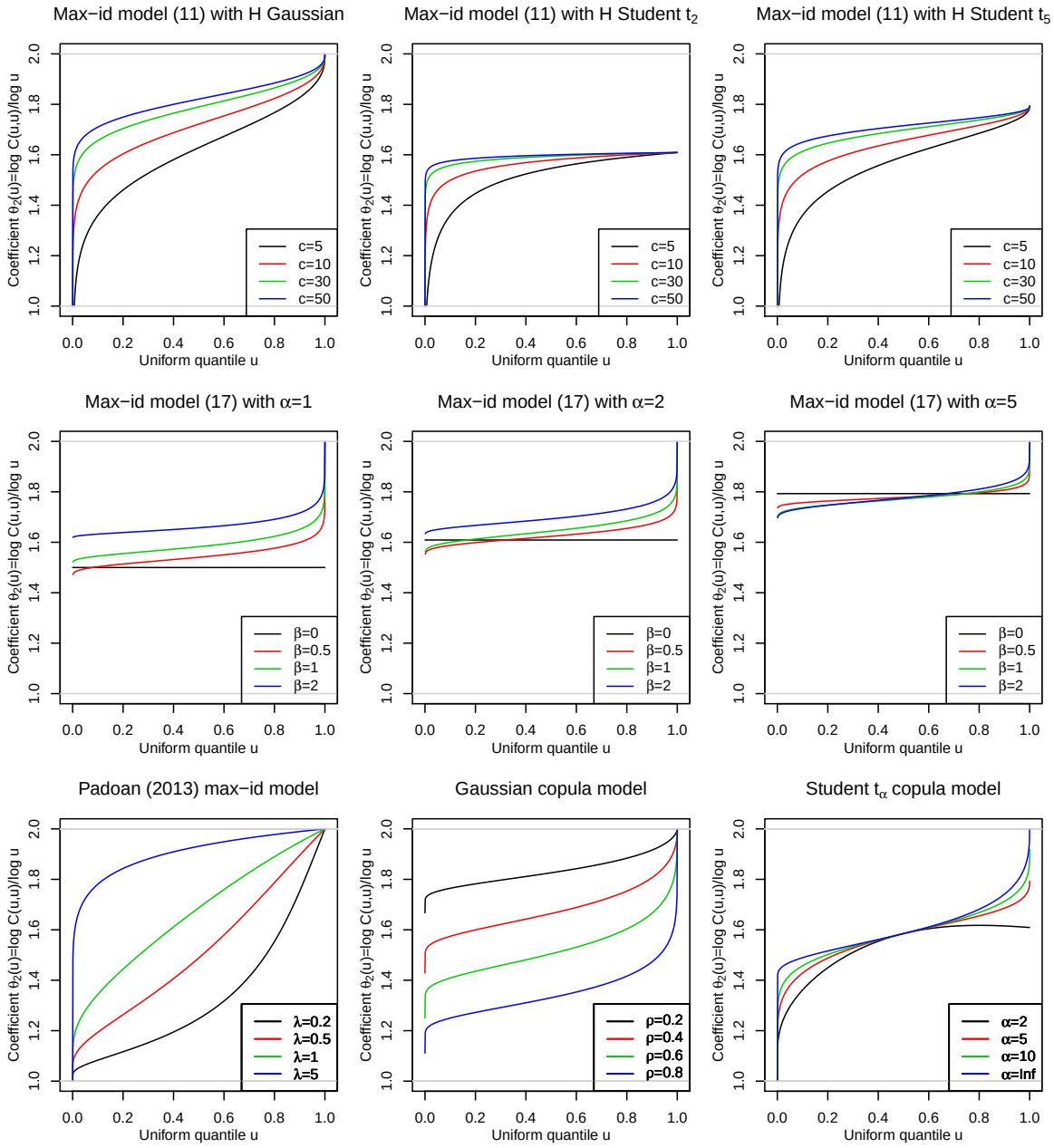


Figure 6.4: Bivariate level-dependent extremal coefficients for a selection of models. Top row: finite exponent measure model (6.15). Middle row: infinite exponent measure model (6.20). Bottom row: model of Padoan (2013) for different variogram values λ (left), Gaussian copula model for different correlation values (middle), Student t copula model for different values of the degrees of freedom parameter (right). The models in the top and middle rows and the Student t copula models are based on an underlying standard Gaussian vector $(W(s_1), W(s_2))^T$ with correlation $\rho(h) = 0.5$.

may replace $W_i(s)$ by $\max\{W_i(s), 0\}$ and set $l(s) = 0$. The finite-dimensional exponent measure Λ in (6.11) is given by

$$\Lambda\left([\mathbf{0}, \boldsymbol{z}]^{\mathrm{c}}\right) = \int_0^\infty \{1 - F_W(\boldsymbol{z}/r)\}\, \kappa_{\boldsymbol{\gamma}}(\mathrm{d}r) < \infty, \qquad \boldsymbol{z} \in (0, \infty)^d, \tag{6.18}$$

where F_W denotes the distribution of the process $W(s)$ observed at any finite collection of $d \geq 1$ sites $s_1, \ldots, s_d \in \mathcal{S}$. Since it must be Radon on $E \setminus \{\boldsymbol{l}\}$, a constraint to be verified is that values $\Lambda\left([\mathbf{0}, \boldsymbol{z}]^{\mathrm{c}}\right)$ are finite for any $\boldsymbol{z} \in (0, \infty)^d$. An intuitive interpretation of (6.17) is to view the max-id process $Z(s)$ as the pointwise maximum of an infinite number of independent "storms" $R_i W_i(s)$ characterized by their amplitude R_i and their spatial extent $W_i(s)$. In practice, we can take $W_i(s)$ as Gaussian processes, and then the scaling variables R_i adjust this baseline model for accommodating more specific joint tail decay rates; recall the results shown in §4.2.2.

The power-law tail of the measure $\kappa([r, \infty]) = 1/r$, $r > 0$, in the max-stable construction (2.17) yields asymptotic dependence. To extend this to asymptotic independence, we propose lighter-tailed models with a Pareto tail on the boundary of the parameter space to keep the max-stable spectral representation as a special case. We say that a measure κ is Weibull-tailed if $\kappa([r, \infty)) \sim c r^\tau \exp(-\alpha r^\beta)$ as $r \to \infty$ for some constants $c > 0$, $\alpha > 0$, $\beta > 0$ and $\tau \in \mathbb{R}$, where we refer to β as the Weibull coefficient of κ. We propose the following two models for the measure $\kappa_{\boldsymbol{\gamma}}$ in (6.17):

$$\kappa_{\boldsymbol{\gamma}}^{[1]}([r, \infty)) = r^{-(1-\alpha)} \exp\{-\alpha(r^\beta - 1)/\beta\}, \qquad r > 0, \boldsymbol{\gamma} = (\alpha, \beta)^{\mathrm{T}} \in [0, 1) \times [0, \infty), \tag{6.19}$$

$$\kappa_{\boldsymbol{\gamma}}^{[2]}([r, \infty)) = r^{-\beta} \exp\{-\alpha(r^\beta - 1)/\beta\}, \qquad r > 0, \boldsymbol{\gamma} = (\alpha, \beta)^{\mathrm{T}} \in (0, \infty) \times [0, \infty). \tag{6.20}$$

For $\beta = 0$, we interpret $\kappa_{\boldsymbol{\gamma}}^{[1]}$ and $\kappa_{\boldsymbol{\gamma}}^{[2]}$ as the limits as $\beta \downarrow 0$, giving $\kappa_{\boldsymbol{\gamma}}^{[1]}([r, \infty)) = r^{-1}$ and $\kappa_{\boldsymbol{\gamma}}^{[2]}([r, \infty)) = r^{-\alpha}$, $r > 0$. With standard Gaussian $W(s)$, the max-stable extremal-t process with $\alpha > 0$ degrees of freedom arises from $\kappa_{\boldsymbol{\gamma}}^{[2]}$ when $\beta = 0$. In the non max-stable cases, the tail decay of $\kappa_{\boldsymbol{\gamma}}^{[k]}$, $k = 1, 2$, is of Weibull type and yields asymptotic independence with Gaussian $W(s)$. With Gaussian correlation function $\rho(h)$, the coefficient of tail dependence between two sites s_1, s_2 at distance $h = \|s_1 - s_2\|$ is $\eta(h) = [\{1 + \rho(h)\}/2]^{\beta/(\beta+2)}$, such that the parameter β plays a crucial role for the joint tail decay rate, while the parameter α also impacts the dependence structure but to a milder degree.

An important distinction between the max-stable and max-id constructions is that the assumption of independence between R_i and $W_i(s)$ is not essential in the max-id case. In the preprint Zhong et al. (2020), we extend the above model by letting the correlation function $\rho(\cdot)$ of the standard Gaussian process W_i depend on R_i, such that $\rho(s_1, s_2; R_i)$ may decrease as R_i increases, and the spatial dependence strength weakens when the overall event magnitude represented by the points $\{R_i\}$ gets larger. For stationary and isotropic $\rho(\cdot)$, we could consider the exponential correlation function

$$\rho(s_1, s_2; R_i) = \exp\{-\|s_1 - s_2\|(1 + R_i)^\nu/\lambda\}, \tag{6.21}$$

for some baseline range parameter $\lambda > 0$, and "modulation" parameter $\nu \in \mathbb{R}$. When $\nu > 0$, the spatial range parameter $\lambda(1 + R_i)^{-\nu}$ gets smaller as R_i increases, and vice versa when $\nu < 0$. Introducing dependence between R_i and W_i adds considerable flexibility to the model

and improves its ability to appropriately capture the dependence of moderately extreme events. Moreover, over large study areas or long periods of time, the strength of extremal dependence, and therefore the spatial extent of clusters of extreme values, may vary. We therefore may extend t exponential correlation model by incorporating spatiotemporal covariates. We index the correlation function of the process W_i by time t as $\rho_t(s_1, s_2; R_i)$ to emphasize that it may vary over time. Building upon propositions for nonstationary Gaussian correlation structures from the literature, a flexible nonstationary correlation function on $\mathbb{R}^2$ is given by

$$\rho_t(s_1, s_2; R_i) = |\Omega_{s_1,t}(R_i)|^{1/4} |\Omega_{s_2,t}(R_i)|^{1/4} \left| \frac{\Omega_{s_1,t}(R_i) + \Omega_{s_2,t}(R_i)}{2} \right|^{-1/2} C\left(Q^{1/2}_{s_1;s_2,t}(R_i)\right),$$

where $\Omega_{s,t}(R_i)$ is a 2-by-2 covariance matrix that may depend on spatial location s, time t and the Poisson points $\{R_i\}$, where $C(h)$ is a stationary isotropic correlation function with unit range, e.g., $C(h) = \exp(-h)$, $h \geq 0$, and $Q_{s_1;s_2,t}(R_i)$ is the quadratic form

$$Q_{s_1;s_2,t}(R_i) = (s_1 - s_2)^T \left(\frac{\Omega_{s_1,t}(R_i) + \Omega_{s_2,t}(R_i)}{2} \right)^{-1} (s_1 - s_2).$$

Covariates, such as time and altitude as used in our temperature data application in Zhong et al. (2020), can be linked to the matrix $\Omega_{s,t}(R_i)$. More precisely, we propose the following general model for the covariance matrix $\Omega_{s,t}(R_i)$:

$$\Omega_{s,t}(R_i) = \lambda^2_{s,t}(1+R_i)^{-2\nu} A(\theta), \quad A(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & a \end{bmatrix} \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}^T, \quad (6.22)$$

where $\lambda_{s,t} > 0$ is a baseline range parameter, $\nu \in \mathbb{R}$, $a > 0, \theta \in [0, \pi/2]$ define a geometric anisotropy. Covariates may be included in $\lambda_{s,t}$, for example by specifying $\lambda_{s,t} = \exp(\lambda_0 + \lambda_1 \times \text{alt}_s + \lambda_2 \times t)$, where $\lambda_0, \lambda_1, \lambda_2 \in \mathbb{R}$ are range parameters corresponding to the intercept, the effect of altitude, and the effect of time, respectively, on the spatial dependence range.

To illustrate the flexibility of some of the max-id models discussed above, Figure 6.4 displays bivariate level-dependent extremal coefficient $\theta_2(u)$ for a variety of models.

Using the spectral construction (6.17), simulation mechanisms for max-id models are similar to those for max-stable models. The Gaussian-based models can be simulated exactly by exploiting multivariate elliptical representations, by analogy with the exact simulation procedures discussed for the extremal-t models in §5.1.2.

6.2.4 Inference

By analogy with the special case of max-stable likelihoods, there is a combinatorial explosion of the number of terms to be computed for the max-id likelihoods in dimension larger than two, such that full likelihood inference is not feasible. We propose using two-step pairwise likelihood inference §1.3.4. In the first step, we fit GEV marginal distributions using some standard method, location-wise maximum likelihood fits, or GAM models to capture spatial trends in the GEV parameters. For the second step, data are transformed to a uniform marginal scale (*i.e.*, we use a copula approach), and then we estimate the max-id dependence parameters with a pairwise likelihood.

6.2.5 Applications

In Huser et al. (2021), we illustrate the benefits of our new modeling framework on Dutch wind gust maxima calculated over different time units. Results strongly suggest that our proposed models outperform other natural models, such as the Student-t copula process and its max-stable limit, even for large block sizes. In the preprint Zhong et al. (2020), we tackle the nonstationary modeling of annual temperature maxima for a century-spanning dataset of 44 European weather stations with a focus on temporal nonstationarities due to climate change. Here, we shortly present the approach and results for the Dutch wind data.

We consider extremes in daily wind gusts from the Netherlands at 30 monitoring stations from 1999 to 2008, and we focus on months October–March, which experience the strongest wind gusts. To study wind gust extremes on various time scales, we compute daily, weekly, monthly and yearly block maxima.

We proceed by a two-step approach for estimating marginal distributions and dependence parameters. We model marginal distributions separately at each location, but jointly across time scales to borrow strength across time series when few observations are available. Specifically, let $\tilde{z}_{ij;k}$ denote the i-th observation at the j-th monitoring station for the k-th time scale. We assume that the daily maxima, $\tilde{z}_{i_1 j;1}$, follow a generalized extreme-value (GEV) distribution $\tilde{G}_{j;1}(z)$ with location, scale and shape parameters $\mu_j \in \mathbb{R}$, $\sigma_j > 0$ and $\xi_j \in \mathbb{R}$, respectively, and that maxima for larger time scales, $\tilde{z}_{i_k j;k}$ $(k = 2, 3, 4)$, are also GEV-distributed according to

$$\tilde{G}_{j;k}(z) = \tilde{G}_{j;1}(z)^{b_k \theta_j} = \exp\left\{ -\left(1 + \xi_j \frac{z - [\mu_j - \sigma_j \{1 - (b_k \theta_j)^{\xi_j}\}/\xi_j]}{\sigma_j (b_k \theta_j)^{\xi_j}} \right)_+^{-1/\xi_j} \right\}, \qquad k = 2, 3, 4,$$

where $a_+ = \max(a, 0)$, $b_2 = 7$, $b_3 = 30$ and $b_4 = 182$ are (approximate) block sizes for weekly, monthly and yearly data, respectively, and $\theta_j \in (0, 1]$ is the *extremal index* specific to each station, representing the proportion of independent extremes within each block. This univariate model utilizes the common summary of temporal extremal dependence without the need to specify a full multivariate distribution for all the daily observations in a block. For each site j, we then maximize a composite likelihood constructed by multiplying the univariate likelihood contributions of all maximum values. The estimated shape parameters are all negative, suggesting short bounded tails, and the extremal index roughly lies in the interval $[0.5, 0.6]$, revealing some mild extremal dependence in the daily time series. For modeling dependence, we standardize the data to the Unif$(0, 1)$-scale and treat the transformed margins as perfectly uniform in the pairwise likelihood approach used to estimate the dependence parameters in the second step.

The special dependence structure of componentwise maxima suggests that these data might be well described over space by a max-id process since the pointwise confidence bounds of the level-dependent extremal coefficients in Figure 6.3 have no overlap for the lowest and highest quantiles of weekly maxima.

We estimate a selection of max-id models based on the above construction principles and using a powered exponential correlation function with range and smoothness parameter for the Gaussian processes involved in the constructions; see Huser et al. (2021) for details. Throughout all of the fitted models, estimated range parameters $\hat{\lambda}$ are large and suggest that spatial dependence is quite strong. In contrast, estimated smoothness parameters show that there is small-scale variability. A comparison of the max-stable models with the max-id extensions reveals that max-stability may be realistic assumption for yearly maxima, while for weekly and

monthly maxima the estimates of $\hat{\beta}$ in non-max-stable models are always significantly different from zero at the 95% confidence level. This implies that these block maxima tend to be closer to a max-stable process as the block size increases, while β provides extra flexibility at subasymptotic regimes associated with small block sizes. In comparison with the infinite exponent measure models, the max-id models with finite exponent measure, based on the Gaussian or Student-t processes in their construction, are also very competitive. In particular, the Student-t max-id model shows the best performance overall, as measured in terms of the Composite Likelihood Information Criterion (CLIC). The parameter c, corresponding to the expected number of independent replicates used to compute maxima, is always estimated to lie above 7.9, such that the singular mass arising in the density of these models is negligible. Overall, our novel max-id models outperform their max-stable extremal-t counterparts for small and moderate block sizes, as well as classical copula models from geostatistics for large block sizes.

6.3 Hierarchical subasymptotic POT models with dependence

Hierarchical constructions of statistical models involve several layers of model components and have found widespread use in spatial and spatiotemporal statistical modeling; recall §1.3.5. The top layer usually refers to the *data*. In this layer, the probability distribution of the observed data is defined, and the aspects related to the sampling design, the observation protocol, the nature of recorded variables and potential observation errors must be taken into account properly. The second layer typically refers to a latent (*i.e.*, unobserved) *process*, which represents the process that one seeks to reveal and study. It could be given as a spatial-temporal parameter surface, often driven by environmental conditions, and it is typically smoother than the top layer of data, especially in the case of discretely observed data, such as event counts or occurrence positions and times of events. In the third and deepest layer (from the perspective of the observer), *hyperparameters* are defined, *i.e.*, parameters that control the behavior of the latent process (variance, dependence, smoothness...), and sometimes also parameters related to the shape of the univariate distribution of observations conditional to the latent process. For reasons of parsimonious and identifiable model structures, but also for the sake of computational benefits, the observations in the data layer are usually assumed to be conditionally independent with respect to the latent process and the hyperparameters.

In this section, several novel constructions and methods for hierarchical modeling of threshold exceedances in continuous variables are discussed. The conditional independence assumption is very useful in the context of threshold exceedances, since exceedances then arise independently in the observed data conditional to the latent process. Therefore, likelihood expressions taking into account the censoring below the threshold correspond to products of censored univariate distributions, which pose less numerical challenges than the joint censoring of several dependent variables.

Hierarchical modeling is often associated with approximation techniques applied in a Bayesian framework, since in many cases the "unconditional" likelihoods (*i.e.*, after integrating out the latent variables) are not available in simple form. Alternatively, if we keep the latent variables as parameters of the model, then a very large number of parameters may have to be estimated. An example falling into this category is the INLA-based model with spatiotemporal Gaussian

random effects embedded into three-stage POT approach using GAM equations for predicting high precipitation quantiles in Opitz et al. (2018), as outlined in §3.1.2. In this model, first a nonstationary marginal model is estimated for the full distribution and used to fix a high threshold corresponding to a high predicted quantile of the fitted model. In the second stage, a nonparametric logistic regression model is used to accurately (re)estimate the exceedance probability above this threshold. Finally, a nonparametric GPD model can be fitted to the excesses above the threshold in the third stage. To account for unobserved or unavailable predictor variables, and to capture stochastic behavior not explained by the response distributions in the three stages, random effects can be incorporated into the linear predictors in the GAM formulas.

In this section, we consider hierarchical models that put stronger focus on appropriately capturing the extremal dependence structure than the above INLA-based model. We discuss a spatial model with a latent copula structure, based on the flexible univariate gamma-gamma construction presented in §3.2.2, where we adapt simulation-based Markov Chain Monte Carlo estimation to handle latent variables. As a second contribution, we highlight another spatiotemporal hierarchical model involving the GPD, for which the "unconditional" marginal and dependence parameters can be estimated in a frequentist setting by using a pairwise likelihood approach. This model has a strong physical interpretation thanks to spatiotemporal kernel smoothing of a gamma noise process for generating dependence at the latent process layer. Moreover, the latent process is shared between the two components for exceedance probabilities and excesses, so that positive correlations between exceedance probabilities and excess sizes can be captured by the model. A further benefit of this model is that its unconditional univariate distributions exactly correspond to the GPD limit from EVT, while in most other hierarchical constructions only the data distribution conditional to the latent process possesses such a strong asymptotic grounding.

6.3.1 Bayesian spatial modeling of extreme event episodes with flexible ratio constructions

We now discuss the approach taken in Yadav et al. (2020). It combines a flexible, subasymptotic univariate model given by the ratio of two gamma variables with a spatial dependence model embedded within one of the gamma-parameters. It permits fully Bayesian inference, and can naturally incorporate covariate information. The number of latent variables is large in this model, with one latent variable for each observation. Nonetheless, it can be fitted in fairly high dimensions using MCMC by exploiting the Metropolis-adjusted Langevin algorithm (MALA), which guarantees fast convergence of Markov chains with efficient block proposals for the latent variables. We also develop an adaptive scheme to calibrate the MALA tuning parameters.

Recall from §3.2.2 that the gamma-gamma model, and similar ratio constructions, can be expressed as

$$Y \mid \Lambda \overset{d}{=} \frac{\tilde{Y}}{\Lambda}, \quad \text{with} \quad \Lambda \geq 0 \perp \tilde{Y} \geq 0, \quad \tilde{Y} \sim F_Y(\,\cdot\,; \Lambda = 1),$$

where in the gamma-gamma case $F_Y(\,\cdot\,; \Lambda = 1)$ corresponds to a $\Gamma(\beta_2, 1)$-distribution, and $\Lambda \sim \Gamma(\beta_2, \alpha)$. Our general spatial hierarchical construction for observations Y_j, $j = 1, \ldots, d$ and

latent variables Λ_j is as follows:

$$\begin{aligned} Y_j \mid \boldsymbol{\Lambda}, \boldsymbol{\Theta}_{\boldsymbol{Y}} &\overset{\text{ind}}{\sim} F_Y(\,\cdot\,; \Lambda_j, \boldsymbol{\Theta}_{\boldsymbol{Y}}), \qquad j = 1, \ldots, d, \\ \boldsymbol{\Lambda} \mid \boldsymbol{\Theta}_{\boldsymbol{\Lambda}} &\sim C_{\boldsymbol{\Lambda}} \left\{ F_{\Lambda_1}(\cdot; \boldsymbol{\Theta}_{\boldsymbol{\Lambda}}^{mar}), \ldots, F_{\Lambda_d}(\cdot; \boldsymbol{\Theta}_{\boldsymbol{\Lambda}}^{mar}); \boldsymbol{\Theta}_{\boldsymbol{\Lambda}}^{dep} \right\}, \\ \boldsymbol{\Theta} &\sim \pi(\boldsymbol{\Theta}), \end{aligned} \tag{6.23}$$

where $C_{\boldsymbol{\Lambda}}$ refers to the spatial copula of $\boldsymbol{\Lambda}$, $F_{\Lambda_j}(\cdot; \boldsymbol{\Theta}_{\boldsymbol{\Lambda}}^{mar})$ denotes the marginal distribution of Λ_j, $j = 1, \ldots, d$, and $\pi(\boldsymbol{\Theta})$ is the prior distribution of the parameter vector $\boldsymbol{\Theta}$.

The joint distribution of $\boldsymbol{Y}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{\Theta}$ can be decomposed into conditional distributions as $\pi(\boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{\Theta}_{\boldsymbol{Y}}, \boldsymbol{\Theta}_{\boldsymbol{\Lambda}}) = \pi(\boldsymbol{Y} \mid \boldsymbol{\Lambda}, \boldsymbol{\Theta}_{\boldsymbol{Y}})\ \pi(\boldsymbol{\Lambda} \mid \boldsymbol{\Theta}_{\boldsymbol{\Lambda}})\ \pi(\boldsymbol{\Theta})$, where $\pi(\cdot)$ denotes a generic (conditional) distribution. The joint posterior distribution $\pi\left(\boldsymbol{\Lambda}, \boldsymbol{\Theta} \mid \boldsymbol{Y}\right)$ of latent variables $\boldsymbol{\Lambda}$ and hyperparameters $\boldsymbol{\Theta}$ is then proportional to $\pi(\boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{\Theta})$, and the posterior distribution of hyperparameters $\boldsymbol{\Theta}$ is obtained by integrating out the latent parameters $\boldsymbol{\Lambda}$, *i.e.*,

$$\pi\left(\boldsymbol{\Theta} \mid \boldsymbol{Y}\right) = \int \pi\left(\boldsymbol{\Lambda}, \boldsymbol{\Theta} \mid \boldsymbol{Y}\right)\ \mathrm{d}\boldsymbol{\Lambda}. \tag{6.24}$$

The dimension of the integration domain in (6.24) can be very high if there are a lot of latent variables. We solve this issue by implementing an MCMC algorithm, in which the latent variables $\boldsymbol{\Lambda}$ are imputed and updated at each iteration.

Joint upper tail behavior

The joint upper tail behavior of the hierarchical model is driven by the interplay of the joint lower-tail behavior of the latent process $\boldsymbol{\Lambda}$ and the upper-tail behavior of the conditionally independent random variables at the observation layer of the model. We here provide more details for the case where $1/\Lambda_j$ has univariate regularly varying distribution with positive tail index ξ, and where $\tilde{Y}$ is lighter-tailed such that $\mathbb{E}(\tilde{Y}^{1/\xi+\varepsilon}) < \infty$ for some $\varepsilon > 0$, which includes the gamma-gamma model. If, in addition, the multivariate distribution $F_{1/\boldsymbol{\Lambda}}$ of $1/\boldsymbol{\Lambda}$ is regularly varying at infinity as defined in (2.14), we have

$$\frac{1 - F_{1/\boldsymbol{\Lambda}}(t\boldsymbol{y})}{1 - F_{1/\boldsymbol{\Lambda}}(t\boldsymbol{1})} \to V_{1/\boldsymbol{\Lambda}}(\boldsymbol{y}), \quad t \to \infty, \quad \boldsymbol{y} > \boldsymbol{0},$$

where $\boldsymbol{1} = (1, \ldots, 1)^T \in \mathbb{R}^d$ and $V_{1/\boldsymbol{\Lambda}}(\boldsymbol{y})$ is some positive limit function. Theorem 3 of Fougeres and Mercadier (2012) then implies multivariate regular variation of $F_{\boldsymbol{Y}}$, *i.e.*,

$$\frac{1 - F_{\boldsymbol{Y}}(t\boldsymbol{y})}{1 - F_{\boldsymbol{Y}}(t\boldsymbol{1})} \to V_{\boldsymbol{Y}}(\boldsymbol{y}) = \int_0^\infty \ldots \int_0^\infty V_{1/\boldsymbol{\Lambda}}(\boldsymbol{y}/\boldsymbol{x}) \prod_{j=1}^{d} F_Y(\mathrm{d}x_j; \Lambda = 1), \quad t \to \infty, \quad \boldsymbol{y} > \boldsymbol{0}.$$

The functions $V_{1/\boldsymbol{\Lambda}}$ and $V_{\boldsymbol{Y}}$ represent exponent measures of multivariate max-stable distributions (not necessarily simple as defined in §2.3.2), and here they are homogeneous of order $-1/\xi$, *i.e.*, $V_{1/\boldsymbol{\Lambda}}(t\boldsymbol{y}) = t^{-1/\xi} V_{1/\boldsymbol{\Lambda}}(\boldsymbol{y})$ and $V_{\boldsymbol{Y}}(t\boldsymbol{y}) = t^{-1/\xi} V_{\boldsymbol{Y}}(\boldsymbol{y})$ for positive values of t and $\boldsymbol{y}$. This result fully characterizes the extremal dependence structure of the process $\boldsymbol{Y}$ resulting from the construction (6.23) in the heavy-tailed case. In practice, we consider either a Gaussian copula in $\boldsymbol{\Lambda}$, or an

elliptical Student-t copula, *i.e.*, the dependence structure of the Gaussian scale mixtures with Student-t marginal distributions. The Gaussian copula leads to asymptotic independence in the limit, while the Student-t copula generates asymptotic dependence.

For modeling replicated observations of spatial extreme event episodes, we suppose that observed data $Y_i(s_j)$, $i = 1, \ldots, n$, $j = 1, \ldots, d$, are composed of n independent time replicates of the d components of a random vector $\boldsymbol{Y} = (Y(s_1), \ldots, Y(s_d))^T$ indexed by locations $s_1, \ldots, s_d$ in $\mathbb{R}^D$, typically $D = 2$.

Censored likelihood with data augmentation for latent variables

Based on the case of the Gaussian copula in $\boldsymbol{\Lambda}$, we illustrate the likelihood formula of data censored below a relatively high threshold and augmented with the latent variables. We write $y_{ij} = Y_i(\boldsymbol{s}_j), \lambda_{ij} = \Lambda_i(s_j)$, $i = 1, \ldots, n$, $j = 1, \ldots, d$, and use the symbols ϕ and ϕ_ρ for the univariate and multivariate Gaussian densities corresponding to Φ and Φ_ρ, respectively. Given a data vector $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{id})^T$ and a fixed threshold vector $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{id})^T \in [0, \infty)^d$, we introduce the exceedance indicator vector $\boldsymbol{e}_i = (e_{i1}, \ldots, e_{id})^T$ with $e_{ij} = 1$ if $y_{ij} \geq u_{ij}$ and $e_{ij} = 0$ otherwise. If $u_{ij} = 0$, no censoring is applied to the value y_{ij}, on the other hand, if $u_{ij} = \infty$ then the observation y_{ij} is treated as fully censored. This may be used to handle missing data and prediction at unobserved locations. In the augmented censored likelihood contribution of $\boldsymbol{y}_i$, we consider both $\boldsymbol{\Theta}$ and $\boldsymbol{\lambda}_i = (\lambda_{i1}, \ldots, \lambda_{id})^T$ as parameters. The density of observations (y_{ij}, e_{ij}) conditional on λ_{ij} is $f_c(y_{ij}, e_{ij}; \lambda_{ij}, \beta_1) = \Gamma(u_{ij}; \beta_1, \lambda_{ij})$ if $e_{ij} = 0$ and $f_c(y_{ij}, e_{ij}; \lambda_{ij}, \beta_1) = \gamma(y_{ij}; \beta_1, \lambda_{ij})$ if $e_{ij} = 1$, where $\gamma(\,\cdot\,; \beta_1, \lambda_{ij})$ is the gamma density with rate λ_{ij} and shape β_1. The augmented censored likelihood contribution for the data vector $(\boldsymbol{y}_i^T, \boldsymbol{e}_i^T)^T$ is thus

$$
\begin{aligned}
L(\boldsymbol{\Theta}, \boldsymbol{\lambda}_i; \boldsymbol{y}_i, \boldsymbol{e}_i) = & \prod_{j=1}^{d} f_c(y_{ij}, e_{ij}; \lambda_{ij}, \beta_1) \\
& \times \phi_\rho[\Phi^{-1}\{\Gamma(\lambda_{i1}; \alpha, \beta_2)\}, \ldots, \Phi^{-1}\{\Gamma(\lambda_{id}; \alpha, \beta_2)\}] \times \prod_{j=1}^{d} \frac{\gamma(\lambda_{ij}; \alpha, \beta_2)}{\phi[\Phi^{-1}\{\Gamma(\lambda_{ij}; \alpha, \beta_2)\}]},
\end{aligned}
$$

where the first line refers to the observation model and the second line to the latent model. The overall augmented censored likelihood is

$$
L_n(\boldsymbol{\Theta}, \boldsymbol{\lambda}; \boldsymbol{y}, \boldsymbol{e}) = \prod_{i=1}^{n} L(\boldsymbol{\Theta}, \boldsymbol{\lambda}_i; \boldsymbol{y}_i, \boldsymbol{e}_i),
$$

where $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^T, \ldots, \boldsymbol{\lambda}_n^T)^T$, $\boldsymbol{y} = (\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_n^T)^T$, and $\boldsymbol{e} = (\boldsymbol{e}_1^T, \ldots, \boldsymbol{e}_n^T)^T$. Thanks to data augmentation and to the conditional independence assumption, only univariate censoring is required, which facilitates computations.

Markov chain Monte Carlo inference

We use Markov chain Monte Carlo (MCMC) sampling to generate a representative posterior sample of the hyperparameter vector $\boldsymbol{\Theta}$ and the latent variables $\boldsymbol{\Lambda}$ involved in the hierarchical model (6.23), conditional on observed data. For prior distributions of hyperparameters, we use

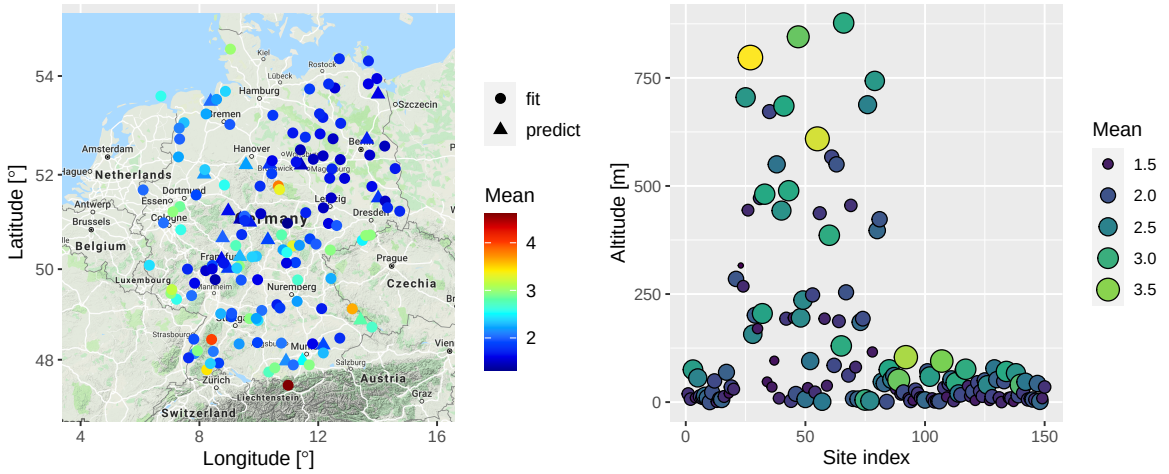


Figure 6.5: Daily precipitation data for Germany. Left: Mean precipitation in mm at study sites calculated over the days of selected extreme events. Right: Mean precipitation in mm plotted with respect to the stations' altitude in m.

appropriately defined PC priors; see their derivation for β_1 and β_2 in §3.2.2. To handle the high dimensionality of the vector of latent variables, we generate MCMC block proposals that ensure a relatively fast exploration of the high-dimensional parameter space of $\mathbf{\Lambda}$. Specifically, we propose using the Metropolis-adjusted Langevin algorithm (MALA), which exploits the gradient of the log-posterior density evaluated at the current parameter configuration to design an efficient multivariate Gaussian proposal density. Details are given in Yadav et al. (2020), where a simulation study shows that our algorithm works correctly and provides accurate inferences.

Application to precipitation extremes in Germany

We apply our hierarchical models to daily precipitation data in Germany for a set of $d = 150$ locations with some missing data for the study period from 2009 to 2018. To avoid modeling complex seasonal nonstationarities, we consider only the observations for the months of September to December, resulting in $n = 1220$ temporal replicates. The location-specific mean precipitation intensities reported in Figure 6.5 show a tendency towards higher values in regions with higher altitudes. The precipitation intensities are zero or very small for most of the days in the observation period, and we first extract extreme events (*i.e.*, specific days) used to fit our spatial hierarchical model. We identify extreme events as threshold exceedances of the average daily precipitation amount taken over all study locations. Using the time series of the binary occurrence indicators of such exceedances, we capture temporal dependence by modeling this time series through a logistic regression with a random effect defined as a first-order autoregressive Gaussian process; see Yadav et al. (2020) for the specification of this model, its inference using INLA and the results. Then, in a second step, we fit the marginally censored gamma-gamma model with latent Gaussian copula (and some other models) to the time series of the selected extreme events. Since the model is subasymptotic and quite flexible, we have explored the use of moderately high thresholds corresponding to marginal empirical quantiles at levels 85%, 90% and 95%. A by-product of censoring low values is that the tricky explicit treatment

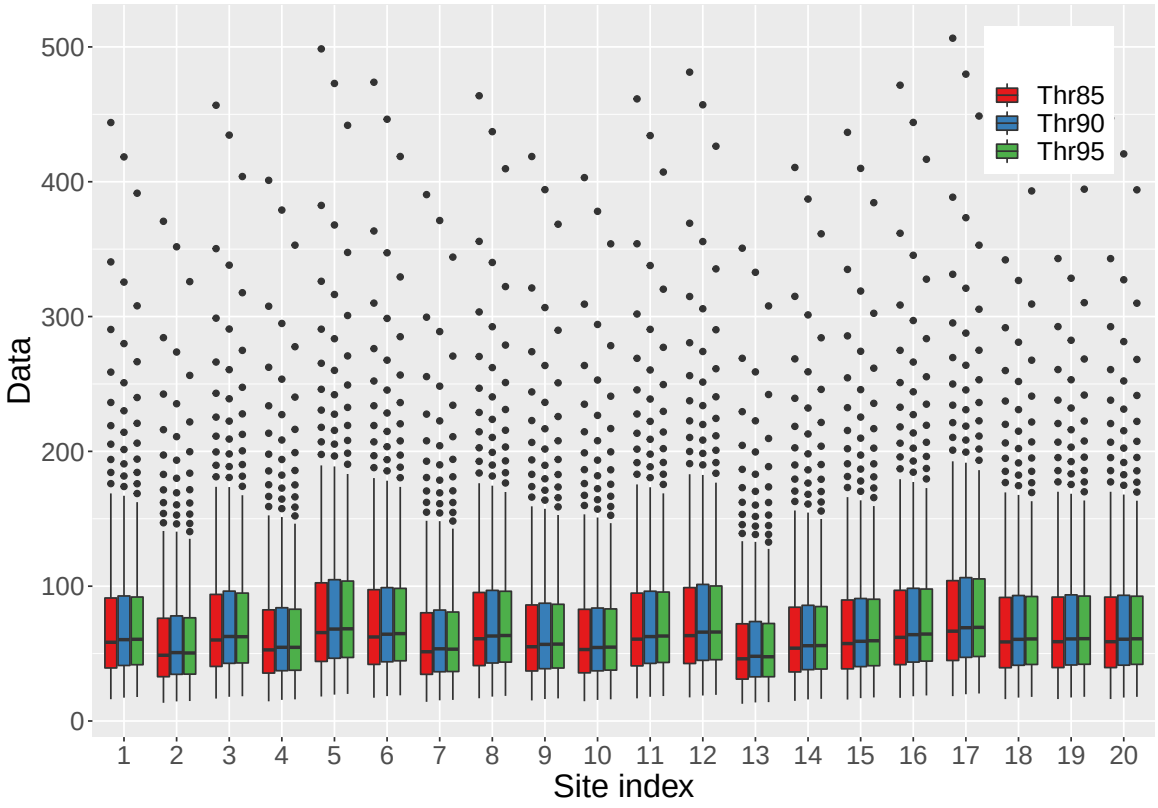


Figure 6.6: Boxplots of posterior predictive samples at 20 hold-out locations for three different marginal thresholds.

of observations of value zero for precipitation intensities is avoided.

The estimated gamma-gamma model shows a satisfactory goodness-of-fit and spatial predictive performance as measured by CRPS and tail-weighted CRPS, and it performs better than a number of simpler models for the precipitation data. Relatively little difference is found between results for different marginal threshold levels, which confirms the flexibility of the model at subasymptotic levels. For illustration, Figure 6.6 shows boxplots of MCMC-based posterior predictive samples at 20 locations held out during the estimation.

6.3.2 Spatiotemporal Gamma-Pareto models

With the exception of the conditional extremes models of Simpson and Wadsworth (2020), the space-time models for extremes available in the current literature typically capture asymptotic dependence or exact asymptotic independence at small distances, while they are not suitable for dealing with residual dependence in asymptotic independence. In Bacro et al. (2019), we propose a novel approach to space-time modeling of asymptotically independent data, where we avoid the tendency of asymptotically-stable models to potentially strongly overestimate joint extreme risks. The model provides a hierarchical formulation for modeling spatiotemporal exceedances over high thresholds. It is defined over a continuous space-time domain and allows for a physical interpretation of extreme events spreading over space and time. Strong motivation comes from

the time series models of Bortot and Gaetan (2014) by developing a generalization of their latent temporal process. Alternatively, our latent process may be viewed as a space-time version of the temporal trawl processes introduced by Barndorff-Nielsen et al. (2014). Our approach is based on representing a generalized Pareto distribution as a gamma mixture of an exponential distribution, as already shown in Equation (3.6). This representation enables us to keep easily tractable marginal distributions which remain coherent with univariate EVT. For the latent process with gamma marginal distributions, we use a kernel convolution of a space-time gamma random process (Wolpert and Ickstadt, 1998) based on influence zones defined as cylinders with an ellipsoidal basis to generate anisotropic spatiotemporal dependence in exceedances. Bivariate densities are available in closed form for this model, and we propose efficient statistical inference based on a pairwise composite likelihood approach, which scales well for relatively large datasets such as the hourly precipitations in the French Mediterranean area studied in the data application.

Hierarchical formulation

We consider a stationary space-time random field $\boldsymbol{Z} = \{Z(x),\, x \in \mathcal{X}\}$ with spatiotemporal index $x = (s,t) \in \mathcal{X} = \mathbb{R}^2 \times \mathbb{R}^+$, such that s indicates spatial location and t time. Without loss of generality (since we can apply marginal transformations to the model), we assume that the margins $Z(x)$ belong to the Fréchet domain of attraction with positive shape parameter ξ. To infer the tail behavior of $\boldsymbol{Z}$, we focus on values exceeding a fixed high threshold u, and we consider the exceedances over u,

$$Y(x) = (Z(x) - u) \times \mathbb{I}_{(u,\infty)}(Z(x)).$$

We now formulate a two-stage model that induces spatiotemporal dependence arising in both the exceedance indicators $\mathbb{I}_{(u,\infty)}(Z(x))$ and the positive excesses $Z(x) - u > 0$ by integrating space-time dependence into a latent gamma component. A key feature of our model is that it naturally links the exceedance probability to the size of the excess, and therefore it provides a joint space-time structure of the zero part and the positive part in the zero-inflated distribution of $Y(x)$. A quite natural assumption is that larger exceedance probabilities may come along with higher exceedances, and the model allows capturing such positive correlation.

In the first stage (observation layer) of the model, we condition on a latent space-time random field $\{\Lambda(x)\}$ with marginal distributions $\Lambda(x) \sim \text{Gamma}(\alpha, \beta)$ with shape $\alpha > 0$ and rate $\beta > 0$. We assume that

$$\begin{aligned} Y(x) \mid (\Lambda(x), Y(x) > 0) &\sim \text{Exp}\left(\Lambda(x)\right), \\ \text{P}(Y(x) > 0 \mid \Lambda(x)) &= e^{-\kappa\Lambda(x)}, \end{aligned}$$

where $\kappa > 0$ is a parameter controling the rate of exceedances of the threshold. The resulting marginal distribution of $Y(x)$ conditionally on $Z(x) > u$ corresponds to the GPD, and the unconditional marginal distribution function of $Y(x)$ is

$$F(y;\sigma,\xi) = \begin{cases} p & \text{if } y = 0, \\ p + (1-p)\text{GPD}(y;\xi,\sigma) & \text{if } y > 0, \end{cases}$$

with shape parameter $\xi = 1/\alpha$, scale parameter $\sigma = (\kappa + \beta)/\alpha$, and with $1 - p$ the probability of an exceedance over u, *i.e.*, $\mathrm{P}(Z(x) > u) = \mathrm{P}(Y(x) > 0) = 1 - p$. The probability of exceeding u,

$$\mathrm{P}(Z(x) > u) = \mathrm{E}\left(\mathrm{P}(Y(x) > 0 | \Lambda(x))\right) = \mathrm{E}\left(e^{-\kappa\Lambda(x)}\right) = \left(\frac{\beta}{\kappa + \beta}\right)^{\alpha}, \tag{6.25}$$

depends on κ, and it is the Laplace transform of $\Lambda(x)$ evaluated at κ. The constraint $\xi > 0$ is not restrictive for dealing with the French precipitation data, which are known to be heavy-tailed. For general modeling purposes, we can relax this assumption by considering a marginal transformation within the class of GPDs for threshold exceedances.

Latent spatiotemporal gamma process

In the latent second stage of the model, spatiotemporal dependence is introduced by means of a space-time stationary random field $\{\Lambda(x),\, x \in \mathcal{X}\}$ with Gamma(α, β) marginal distributions. In principle, we could use an arbitrarily wide range of models with any kind of space-time dependence structure, for instance by marginally transforming a space-time Gaussian random field using the copula idea (Joe, 2014); see the preceding section presenting the work in Yadav et al. (2020) for a related approach. However, we here aim to propose a construction where gamma marginal distributions arise naturally without applying rather artificial marginal transformations. Inspired by the gamma process convolutions of Wolpert and Ickstadt (1998), we develop a space-time gamma convolution process with gamma marginal distributions. The kernel shape in our construction allows for a straightforward interpretation of the dependence structure, and it offers a physical interpretation of real phenomena such as mass and participle transport. We fix $\mathcal{X} = \mathbb{R}^D$ and consider $A \in \mathcal{B}_b(\mathcal{X})$, a subset of $\mathcal{X}$ belonging to the σ-field $\mathcal{B}_b(\mathcal{X})$ restricted to bounded sets of $\mathcal{X}$. A gamma random field $\Gamma(dx)$ is a non negative random measure defined on $\mathcal{X}$, characterized by a base measure $\alpha(dx)$ and an inverse scale parameter β and obeying the following two conditions:

1. $\Gamma(A) = \int_A \Gamma(dx) \sim \text{Gamma}(\alpha(A), \beta)$, with $\alpha(A) = \int_A \alpha(dx)$;
2. for any $A_1, A_2 \in \mathcal{B}_b(\mathcal{X})$ such that $A_1 \cap A_2 = \emptyset$, $\Gamma(A_1)$ and $\Gamma(A_2)$ are independent random variables.

Likelihood-related calculations for our model rely on the Laplace exponent of the random measure given by

$$\mathcal{L}(\phi) = -\log \mathrm{E}\left(\exp\left\{-\int \phi(x)\Gamma(dx)\right\}\right) = \int_{\mathcal{X}} \log\left\{1 + \frac{\phi(x)}{\beta}\right\}\alpha(dx)$$

where ϕ may be any positive measurable function; in our case, it will represent the kernel function. With $\phi(x) = v\mathbb{I}_A(x)$, we get

$$\mathcal{L}(\phi) = -\log \mathrm{E}\left(\exp\{-v\Gamma(A)\}\right) = \int_A \log\left\{1 + \frac{v}{\beta}\right\}\alpha(dx) = \alpha(A)\log\left\{1 + \frac{v}{\beta}\right\},$$

i.e.,

$$\mathrm{E}\left(\exp\{-v\Gamma(A)\}\right) = \left(\frac{\beta}{v + \beta}\right)^{\alpha(A)}.$$

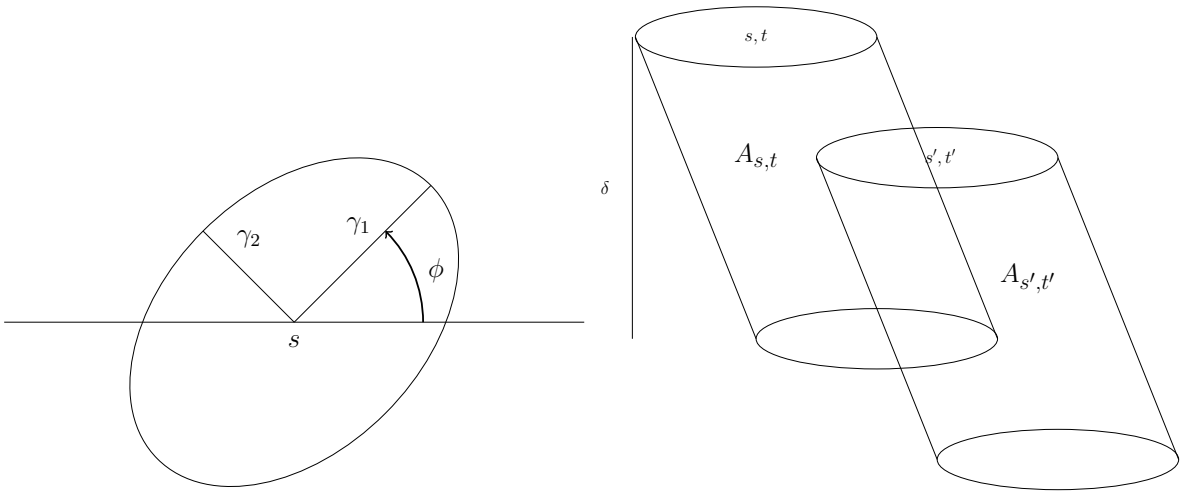


Figure 6.7: Illustration of indicator kernels for space-time Gamma-Pareto processes. Left display: elliptical basis in space. Right display: intersection of two space-time cylinders with the same elliptical basis.

For bivariate analyses we choose $\phi(x) = v_1 \mathbb{I}_{A_1}(x) + v_2 \mathbb{I}_{A_2}(x)$, which yields

$$\mathrm{E}(\exp\{-v_1\Gamma(A_1) - v_2\Gamma(A_2)\}) = \left(1 + \frac{v_1}{\beta}\right)^{-\alpha(A_1 \setminus A_2)} \left(1 + \frac{v_1 + v_2}{\beta}\right)^{-\alpha(A_1 \cap A_2)} \left(1 + \frac{v_2}{\beta}\right)^{-\alpha(A_2 \setminus A_1)}.$$

We propose to model $\{\Lambda(x),\, x \in \mathcal{X}\}$ as a convolution using a 3D indicator kernel $K(x, x')$ (*i.e.*, an indicator function) with an indicator set of finite volume used to convolve the gamma random field $\Gamma(dx)$ (Wolpert and Ickstadt, 1998), *i.e.*, $\Lambda(x) = \int K(x, x')\Gamma(dx')$. The shape of the kernel can be very general (though non indicator kernels usually do not lead to gamma marginal distributions), and particular choices may lead to nonstationary random fields, or to stationary random fields with given dependence properties such as full symmetry, separability or independence beyond some spatial distance or temporal lag. In order to limit model complexity and computational burden to a reasonable amount, we propose using the indicator kernel $K(x, x') = \mathbb{I}_A(x - x')$, for $A \in \mathcal{B}_b(\mathcal{X})$, where A is given as a slated elliptical cylinder, defining a D-dimensional set A_x that moves through $\mathcal{X}$ according to some velocity vector. More precisely, let $E(s, \gamma_1, \gamma_2, \phi)$ be an ellipse centered at $s = (s_1, s_2) \in \mathbb{R}^2$ (see the left display of Figure 6.7), with its axes rotated counterclockwise by the angle ϕ with respect to the coordinate axes, and with the semi-axes' lengths in the rotated coordinate system denoted by γ_1 and γ_2, respectively. A physical interpretation of this construction is that the ellipse describes the spatial influence zone of a storm centered at s. For the temporal dynamics, we assume that the ellipses (storms) $E(s, \gamma_1, \gamma_2, \phi)$ move through space with a velocity $\omega = (\omega_1, \omega_2) \in \mathbb{R}^2$ for a duration of $\delta > 0$. The volume of the intersection of two slated elliptical cylinders (see the right display of Figure 6.7) is given by

$$V(s, t, s', t') = (\delta - |t - t'|)_+ \times \nu_2(E(s, \gamma_1, \gamma_2, \phi) \cap E(\tilde{s}, \gamma_1, \gamma_2, \phi))$$

where $\tilde{s} = (\tilde{s_1}, \tilde{s_2})$ with $\tilde{s_i} = s'_i - |t' - t| \times \omega_i$, $i = 1, 2$.

For two fixed locations, the strength of dependence in the random field $\Lambda(x)$ is an increasing monotone function of the intersection volume; other choices of A are possible, provided that we are able to calculate efficiently the volume of the intersection.

We consider the measure

$$\alpha(B) = \alpha\nu_D(B)/\nu_D(A), \quad B \in \mathcal{B}_b(\mathcal{X}), \tag{6.26}$$

where $\nu_D(\cdot)$ is the Lebesgue measure on $\mathbb{R}^D$. It follows that $\Lambda(x) \sim \text{Gamma}(\alpha, \beta)$. The univariate Laplace transform of $\Lambda(x)$ is

$$LP_x^{(1)}(v) := \mathrm{E}\left(e^{-v\Lambda(x)}\right) = \left(\frac{\beta}{v+\beta}\right)^{\alpha},$$

and the bivariate Laplace transform of $\Lambda(x)$ and $\Lambda(x')$ is

$$LP_{x,x'}^{(2)}(v_1, v_2) = \left(\frac{\beta}{v_1+\beta}\right)^{\alpha(A_x \backslash A_{x'})} \left(\frac{\beta}{v_1+v_2+\beta}\right)^{\alpha(A_x \cap A_{x'})} \left(\frac{\beta}{v_2+\beta}\right)^{\alpha(A_{x'} \backslash A_x)}.$$

The intersecting volume $|A_{x_1} \cap A_{x_2}|$ tends to 0 if $\|x_2 - x_1\| \to \infty$, which establishes the property of α-mixing over space and time for the processes $\Lambda(x)$ and $Y(x)$.

Joint tail behavior of Gamma-Pareto processes

Thanks to closed-form expressions of bivariate Laplace transforms, the values of the bivariate distribution function of $Z(x)$ are also available in closed form. It is easy to show that for any $(x, x') \in \mathcal{X}^2$, $x \neq x'$, and for values v exceeding a threshold $u \geq 0$, we get

$$\begin{aligned}\Pr(Z(x) > v, Z(x') > v) =& LP_{x,x'}^{(2)}(v-u+\kappa, v-u+\kappa) \\ =& \left(1 + \frac{v-u+\kappa}{\beta}\right)^{-\alpha(A_x \backslash A_{x'})} \left(1 + \frac{2v-2u+2\kappa}{\beta}\right)^{-\alpha(A_x \cap A_{x'})} \\ & \times \left(1 + \frac{v-u+\kappa}{\beta}\right)^{-\alpha(A_{x'} \backslash A_x)}.\end{aligned}$$

To simplify notations, we set $c_0 = \alpha(A_x)$, $c_1 = \alpha(A_x \backslash A_{x'})$, $c_2 = \alpha(A_x \cap A_{x'})$, $c_3 = \alpha(A_{x'} \backslash A_x)$, such that $c_1 = c_3 = c_0 - c_2 \geq 0$ and $c_1 + 2c_2 + c_3 = 2c_0$. For $c_2 = 0$ characterizing disjoint indicator sets A_x and $A_{x'}$, it is clear that $Z(x)$ and $Z(x')$ are independent.

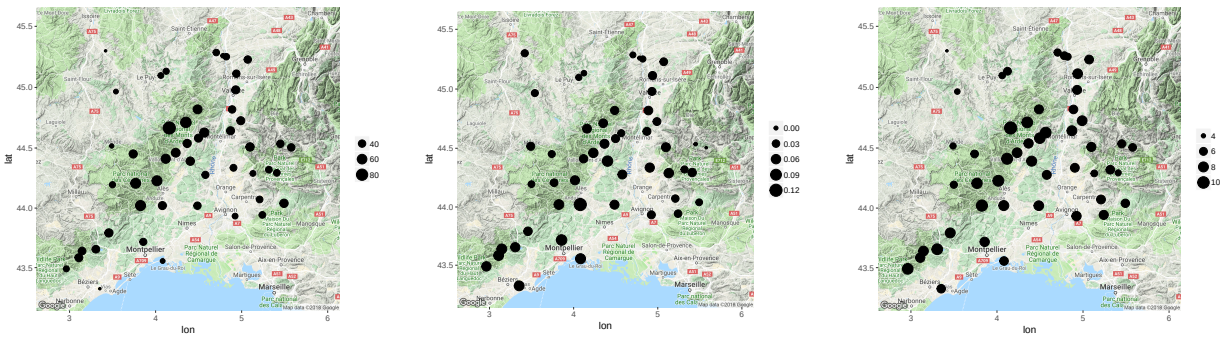
The process $Z(x)$ is asymptotically independent. If we assume $u = 0$ without loss of generality and $x \neq x'$, then the tail correlation coefficient can be calculated as follows:

$$\chi_{x,x'}(v) = \frac{\mathrm{P}(Z(x) > v, Z(x') > v)}{\mathrm{P}(Z(x') > v)} \sim 2^{-c_2} \left(\frac{v}{\beta}\right)^{c_2 - c_0}, \quad v \to \infty.$$

Since $c_2 < c_0$, we obtain $\chi_{x,x'} = 0$ in the limit as $v \to \infty$. To characterize the faster joint tail decay, we can calculate

$$\overline{\chi}_{x,x'} = \lim_{v \to \infty} \overline{\chi}_{xx'}(v) = \frac{2c_0}{c_1 + c_2 + c_3} - 1 = \frac{c_2}{2c_0 - c_2},$$

which describes the ratio between the intersecting volume of A_x and $A_{x'}$ and the volume of the union of these two sets.



u ξ σ

Figure 6.8: Precipitation data for southern France. Dots correspond to the stations used for fitting; their diameter is proportional to station-wise empirical 99% quantiles $u(s)$ (left display) and estimates of the GPD parameters $\xi(s)$ (middle display) and $\sigma(s)$ (right display).

Pairwise likelihood inference

Based on the above formula for the bivariate survivor function of $Z(x)$, it is straightforward to derive the expression of the pairwise likelihood (PL) of the model. We use temporal and spatial cutoff distances beyond which pairs are not included into the PL. This keeps the number of terms in the PL tractable, and it also avoids that its value is dominated by a large number of intermediate-range distances where spatial and temporal dependence has almost completely vanished. Due to the mixing properties of the Gamma-Pareto field, we obtain asymptotic normality of the PL estimator, and standard errors and information criteria such as the composite likelihood information criterion (CLIC) can be obtained through block subsampling techniques (Carlstein, 1986). A simulation study in Bacro et al. (2019) shows that PL estimation is efficient even in high-dimensional settings and provides unbiased parameter estimates for the complex hierarchical space-time models that we propose.

Application to Mediterranean precipitation episodes

We apply the hierarchical model to precipitation extremes observed over a study region in the south of France. Extreme rainfall events usually occur during fall season in this area. They are mainly due to southern winds driving warm and moist air from the Mediterranean sea towards the relatively cold mountainous areas, leading to a situation which often provokes severe thunderstorms. The dataset for September to November months of the study period spanning the years 1993–2010 consists of observations for 50 weather stations over 54542 hours. The spatial design of the stations is illustrated in Figure 6.8, where marginal tail behavior is highlighted.

Figure 6.9 displays empirical estimates of extremal dependence measures χ and $\overline{\chi}$ for different threshold levels and strongly supports the assumption of asymptotic independence at all positive distances and at all positive temporal lags. In particular, the relatively stable behavior of the $\overline{\chi}$-estimates for different threshold levels indicates the presence of residual tail dependence that

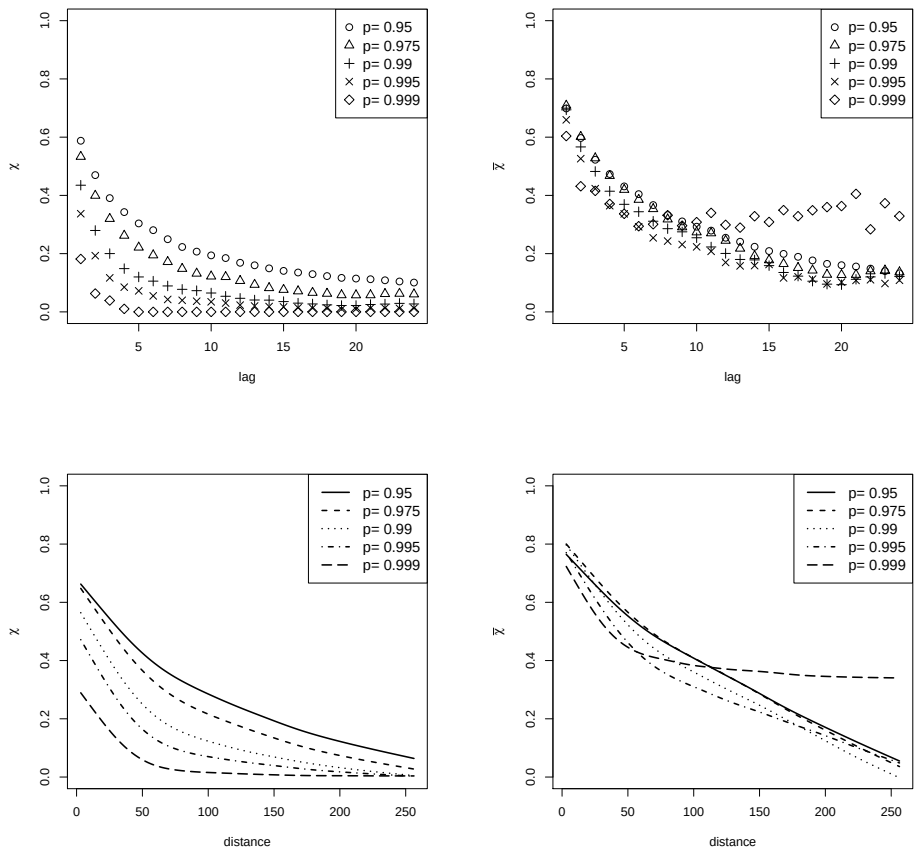


Figure 6.9: Empirical estimates of extremal dependence measures χ (left column) and $\overline{\chi}$ (right column) for the precipitation data in southern France with respect to temporal lags (top row) and spatial distances (bottom row).

vanishes only asymptotically.

Joint estimation of nonstationary margins and dependence would be highly intricate here. Therefore, we adopt a two-step procedure. We first fit a GPD separately to each station with thresholds chosen as the empirical 99.5% quantile. Next, we use the estimated parameters $\hat{\xi}$ and $\hat{\sigma}$ to transform the raw exceedances $Y(x)$ observed at station s to a marginally normalized scale. Then, we fit our hierarchical models to the censored pretransformed data $\tilde{Y}(x)$ by numerically maximizing the pairwise likelihood. We consider two settings for the hierarchical model, with and without velocity, and we further compare these two models to three variants of a censored Gaussian space-time copula model pertaining to the class of asymptotic independent processes. Estimation results, especially CLICs, show a preference for our hierarchical models with the best value for model the most complex model. For visual interpretation, we refer to Figure 6.10, which shows a simulation of G1 where a slight movement of rainfall along the west-east direction becomes apparent. More details about the application can be found in Bacro et al. (2019).

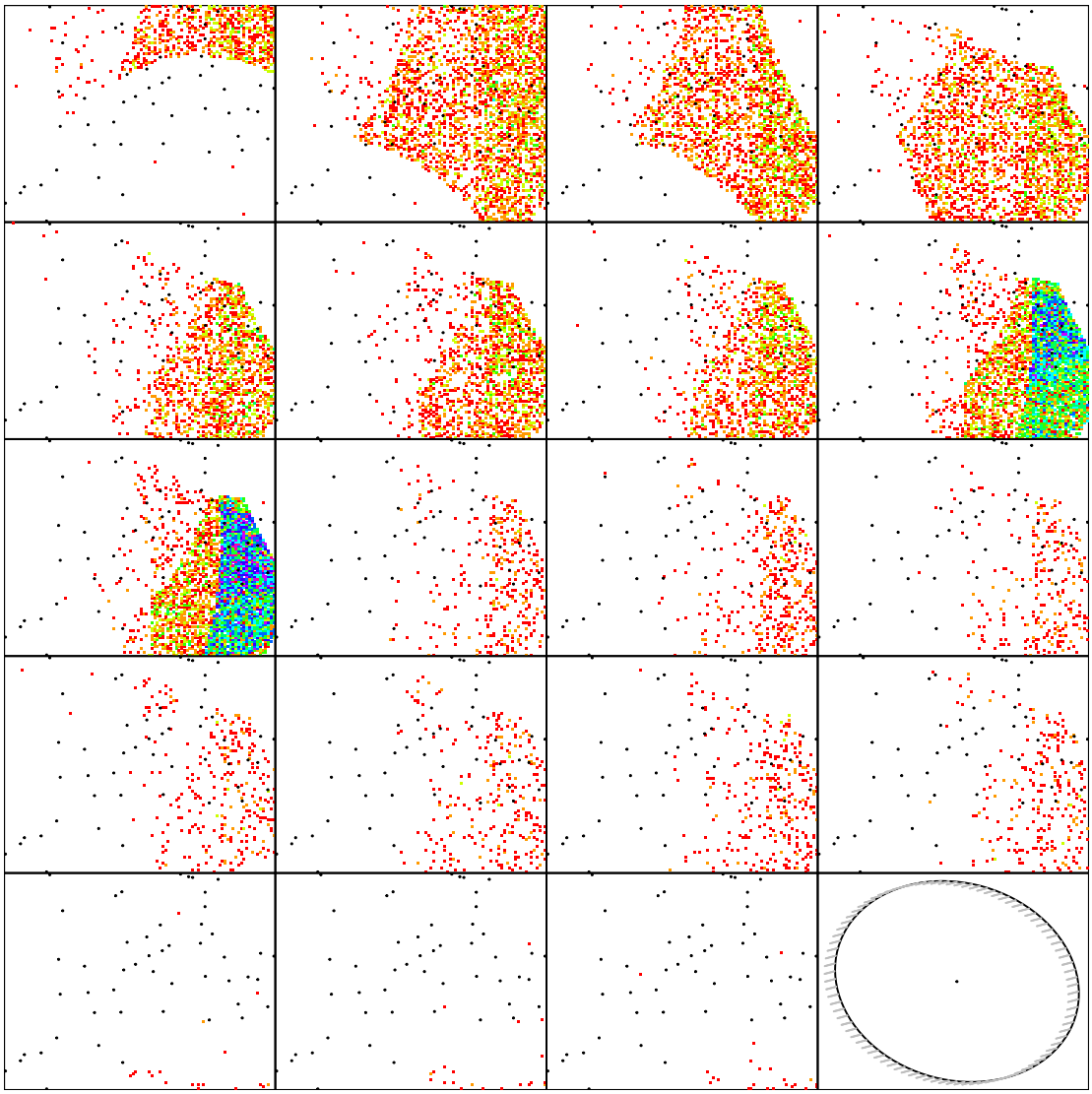


Figure 6.10: A simulation example showing exceedances above the 0.95-quantile for the best Gamma-Pareto model for French precipitation data. Dots correspond to the stations used for fitting. The evolution over time during 19 hours is presented row-wise starting from the top left. The bottom right display illustrates the estimated ellipse basis of space-time kernel sets, centred at the barycenter of the locations, and the movement induced by the velocity vector.

6.4 Discussion of subasymptotic modeling of extremes

Subasymptotic models for spatial and spatiotemporal extremes do not directly arise as extreme-value limits, such as max-stable or Pareto processes. They provide higher flexibility for capturing joint tail behavior at finite levels, especially when the strength of extremal dependence decreases at higher quantile levels (*e.g.*, smaller spatial area of excursion sets above increasingly high thresholds, shorter duration of more extreme episodes). Aside from high flexibility, useful subasymptotic models should preserve a strong asymptotic motivation, for instance by keeping customary asymptotic models as special boundary cases, or by providing formulas for extremal dependence measures that are simple to compute and to interpret. Since a much larger range of potentially useful model classes is available for subasymptotic modeling thanks to dropping the asymptotic stability properties, model selection and validation has to be carried out very carefully. For that purpose, we define novel tools such as the level-dependent extremal coefficient for max-id processes, and in several publications (*e.g.*, Huser et al., 2017; Zhong et al., 2020) we propose the use of predictive scores for comparing models, such as variants of CRPS for spatial prediction. Efficient inference of parameters is another important step towards statistical practice with subasymptotic models. We therefore adapt estimation tools of classical limit models (pairwise likelihoods, partially censored likelihoods) to the new model classes, and we show through simulation studies that estimation and parameter inference work well.

Because of the use of random scaling variables that do not vary over space, spatial Gaussian scale mixture processes and max-id processes are not mixing, *i.e.*, independence cannot be attained between locations that are far separated. When we seek to model extreme events of moderate spatial extent over very large study areas, these models may therefore not be appropriate. Extensions towards models with similar construction mechanisms but allowing for dependence decay towards independence for increasingly large spatial distances would therefore be welcome and are part of ongoing work. Spatiotemporal extensions of random scaling techniques for modeling extreme event episodes spanning over several discrete time steps, or arising in continuous time, are another active field of research. In this context, it would also help to provide mechanisms allowing the temporal dependence to decay towards independence for times that are separated by an increasingly large time lag.

The hierarchical spatiotemporal Gamma-Pareto models present an interesting option for asymptotically independent data. Mixing in space and time arises naturally, and pairwise likelihood estimation is fast due to to simple analytical expressions for bivariate distribution functions. Some aspects still require further developments for such models, such as spatiotemporal prediction, *e.g.*, through simulation conditional to observed values.

Another very flexible class of subasymptotic models for multivariate, spatial and spatiotemporal data has been developed through the conditional extremes approach (Heffernan and Tawn, 2004; Wadsworth and Tawn, 2019; Simpson and Wadsworth, 2020), where one models the process conditional to an exceedance of a high fixed threshold at a given location. By using a marginal scale with exponential tails, the conditional process can be flexibly modeled through a nonstationary Gaussian random field. In this setting, our preprint Simpson et al. (2020) tackles spatial and spatiotemporal modeling and proposes particularly flexible, semi-parametric model specifications for the conditional mean, and we embed the model into the INLA-SPDE framework to achieve efficient Bayesian inference for datasets with a large number of several thousands of observation locations.

Chapter 7

Spatiotemporal Bayesian modeling of occurrences of environmental risks

This chapter provides a brief general introduction to flexible Bayesian modeling of intensity functions of spatiotemporal point patterns. We aim to answer the question of *how many* events will happen and *where* they will happen. By using marked point processes, that is, by additionally modeling information about the type or magnitude of events, we can further answer the question of *how severe* the events will be, or of *which type* they will be. The purpose of constructing and estimating models is to explain the factors contributing to event occurrence and magnitude, and to provide prediction over space and time for hazard and risk mapping. Specific approaches and results for applications to landslides and wildfires are presented in the following Chapters 8 and 9, respectively.

7.1 General framework: Log-Gaussian Cox processes

Log-Gaussian Cox processes (LGCPs, Møller et al., 1998) can be viewed as Poisson point processes with Gaussian random effects in their intensity function. As such, they are natural models to capture clusters of occurrences of points that are not explainable through the deterministic components (*e.g.*, fixed covariate effects) in the intensity function. The following exposition focuses on models of LGCP type. Conditional to the random intensity function, LGCPs become nonstationary Poisson processes, for which a wide variety of estimation techniques are available, including standard likelihood estimation. A difficulty to be handled is that some discretization of continuous space and time is necessary to estimate an integral of the intensity function over the observation window. Moreover, the inclusion of the Gaussian random effects requires estimation techniques that allow handling latent Gaussian variables, and we here focus on INLA (Illian et al., 2012). Spatially indexed random effects can be defined through the SPDE approach.

7.1.1 Susceptibility vs. occurrence intensity

The statistical modeling of occurrences of environmental risks in the geosciences, such as landslide processes, has spawned a rich literature. In most of these approaches, the probability of occurrence of such events is represented only through the notion of susceptibility when con-

ducting estimation and hazard mapping. This means that one models the probability of presence/absence of events in the areal units in a fixed partition of space. For instance, one may consider a fine-meshed regular grid of pixels covering geographic space, or one may consider more problem-specific partitions of space, such as slope units as used in Lombardo et al. (2020) for landslide occurrences, or administrative units. In contrast, the intensity function of a point process is defined over continuous space, and it is of prime interest in the modeling of point processes. While the approach of modeling susceptibilities gives access to a wide range of classification algorithms from statistics and machine learning communities to estimate models, we identify two major drawbacks from focusing attention solely on the probability of presence or absence in a given areal unit. First, a probability of occurrence is associated to a fixed areal unit, such that we cannot easily provide occurrence probabilities for other areal units than those fixed in the estimated model. In particular, upscaling of probabilities from smaller to larger areal units is awkward. By contrast, an intensity function allows calculating intensities for arbitrary spatial units, and probabilities for the occurrence of at least one landslide event are readily derived. Second, the notion of susceptibility is binary since it considers only presence or absence, but it does not provide any information on the number of events that may arise in an area. Again, the intensity framework is more general since it allows calculating expected event counts for any spatial support, and the probability distribution of the number of events can be calculated analytically or obtained through Monte-Carlo simulation of the estimated point process models. Our seminal paper Lombardo et al. (2018) has introduced the point process concept to the geomorphological community of landslide modelers, and in several follow-up publications and ongoing work we capitalize on this approach by providing relevant modeling extensions.

It is important to note that many point process models can be accurately approximated and estimated through variants of logistic regression, *i.e.*, through a representation with presence/absence values. Therefore, classification methods can still be applied analogously for point process models. By consequence, we recommend to systematically use a point process framework where possible, since all the features and methods of susceptibility modeling are still available, but it comes with important additional conceptual and practical benefits.

7.1.2 Random effects for capturing unavailable environmental predictors

Let us consider a Poisson process over an observation window $\mathcal{X}$ with intensity function $\lambda(x)$, $x \in \mathcal{X} \subset \mathbb{R}^D$, as already introduced in §2.1. By including random effects in the intensity functions, we obtain a Cox process. As outlined above, the class of log-Gaussian Cox processes (LGCPs, Møller et al., 1998) has proven to be particulary flexible and useful for modeling complex spatial and spatiotemporal point patterns. With Cox processes, we use upper-case notation $\Lambda(x)$ to emphasize the randomness in the intensity function.

Effects of observed covariates $z_j(x)$, $j = 1, \ldots, J$, can be incorporated into the intensity function through a log-linear specification, *i.e.*, as *fixed effects*. The log-intensities of flexible LGCP models with fixed and random effects can be structured according to the following formula defining a generalized additive mixed model (GAMM) with log-link function:

$$\log \Lambda(x) = \beta_0 + \sum_{j=1}^{J} \beta_j z_j(x) + \sum_{\ell=1}^{L} W_\ell(x), \quad x \in \mathcal{X}. \tag{7.1}$$

The Gaussian random effects $W_\ell(x)$ may directly depend on location x, for instance in the case of a spatial random effect ($D = 2$). Spatial random effects may be defined through the SPDE approach, or as spatial conditionally autoregressive models based on areal units with an adjacency graph (Besag, 1974). In other cases, random effects may depend only indirectly on location x through a covariate $z_j(x)$ observed at x; for instance, $W_\ell(z_j(x))$ can be used to capture the nonlinear influence of the covariate z_j. The role of random effects is to capture sources of variation of the intensity function that cannot be explained through the deterministic and log-linear influence of the fixed effects $\beta_j z_j(x)$ of the observed covariates z_j, $j = 1, \ldots, J$.

When adopting Bayesian estimation techniques as described below, we systematically scale continuous covariates to have empirical mean 0 and empirical variance 1. This facilitates defining priors for fixed effect coefficients with similar influence of different covariates, and it allows for a more straightforward interpretation and comparison of the absolute values of estimated fixed effect coefficients.

7.1.3 Event magnitudes as marks in log-Gaussian Cox processes

Additional information may be available to characterize events and can be modeled as marks of the points. Events may be of different type (*e.g.*, shallow and deep landslides, or small and large wildfires), in which case the mark is categorical and indicates the type, leading to a multi-type point process. Marks may also be represented on a continuous numerical scale; a typical example is event magnitude (*e.g.*, burnt area of wildfires, or size of landslide scars). An alternative solution to including this additional information as marks consists of extending the observation space $\mathcal{X}$ of points, such that it also contains the dimensions along which the marks are defined. For instance, instead of using real-valued marks, points of the point pattern could be defined as the vector combining the point location and the mark, such that $\mathcal{X} = \mathbb{R}^{D+1}$.

In many applications, it is challenging to define a model for both point occurrences and marks that appropriately captures the potential interactions between these two components. In the simplest case, mark values arise independently from the point process intensity, and we are in the situation of *geostatistical marking*. Then, the estimation of the process describing the mark distribution can be done separately from the estimation of the point process model without marks.

7.2 General estimation strategy using INLA-SPDE

Estimation methods for point processes are a lively area of research. For spatial and spatiotemporal LGCPs whose log-linear predictor in (7.1) comprises fixed effects and a spatial or spatiotemporal Gaussian random field, relatively fast and robust methods are available to estimate fixed effects coefficients and the parameters of the Gaussian covariance function. These approaches do not explicitly handle the latent Gaussian variables but rather focus on second-order properties of the point process (*i.e.*, the characterization of the interaction in point pairs) that are available in closed form. For estimating fixed effect coefficients, it would be possible to adapt the approach to intensity estimation outlined in §2.1.4. To estimate also the hyperparameters of the Gaussian process (*e.g.*, variance, range), we can mention moment-based methods (Illian et al., 2008; Waagepetersen and Guan, 2009) or composite likelihood-based approach

(Guan, 2006). However, these approaches do not provide the surface of the latent process conditional to the observed point pattern, *i.e.*, , they do not provide an estimation $\hat{\Lambda}(x)$ of $\Lambda(x)$. Though, estimation of the field Λ, together with the associated uncertainties, is required for intensity mapping, which is often a crucial goal in applications to environmental risks, such as landslides or wildfires. Therefore, we here focus on likelihood-based techniques that allow for inferences on all model components, especially on the latent field $\Lambda(x)$. More precisely, we adopt a fully Bayesian approach using INLA-SPDE, where we can set prior distributions for the hyperparameters controling the latent Gaussian predictor components.

Recall that the probability density function of an observed finite point pattern $\boldsymbol{X} = (x_1, \ldots, x_N)^T$ (without considering marks), composed of a random but finite number $N \geq 0$ of points x_i in the observation window $\mathcal{X}$, corresponds to the expectation

$$f_{\text{LGCP}}(\boldsymbol{X}) = \mathbb{E}_\Lambda \left[\exp\left(- \int_{\mathcal{X}} \Lambda(x)\, \mathrm{d}x \right) \prod_{i=1}^{N} \Lambda(x_i) \right],$$

using the convention that $\prod_{i=1}^{N} \Lambda(x_i) = 1$ if $N = 0$. Closed-form expressions of this expectation are not available in general, but Bayesian inference techniques, such as INLA, have been developed to approximate it numerically. In the Bayesian framework, the Gaussian processes arising in the log-Gaussian intensity function $\Lambda(x)$ can be viewed as prior distributions for deterministic components of the intensity function of a Poisson process.

Two major challenges arise for likelihood-based inference in LGCPs: (i) intensity functions $\Lambda(x)$ are conceptually defined over continuous space; (ii) the Gaussian random effects lead to an intractable likelihood with no general closed-form expression. As to (i), different approximation strategies allow numerical computation of the integral $\int_{\mathcal{X}} \Lambda(x)\, \mathrm{d}x$ conditional on Λ. A standard approach is to discretize the observation window and to assume that the intensity function does not vary within the cells C_k of the resulting partition of the observation window, where $\bigcup_k^K C_k = \mathcal{X}$ and $C_{k_1} \cap C_{k_2} = \emptyset$ if $k_1 \neq k_2$. Then, conditional on Λ, the number of points observed in a cell C_k, $k = 1, \ldots, K$, is Poisson distributed with $N(C_k) \sim \text{Pois}(|C_k| \times \Lambda_k)$, where $|C_k|$ is the Lebesgue volume, and $N(C_k)$ are mutually independent. Therefore, estimating the LGCP corresponds to performing a (mixed) Poisson regression with the canonical log-link:

$$N_k \mid \Lambda_k \overset{\text{ind}}{\sim} \text{Pois}(|C_k| \times \Lambda_k), \quad \log(\Lambda_k) = \mu_k, \quad k = 1, \ldots, K, \tag{7.2}$$

where the *linear predictor* μ_k is additively composed of fixed and random effects, as described above. The multiplicative constants $|C_k| > 0$ appear as an offset $\log(|C_k|)$ in the intercept of the linear predictor of the Poisson regression. For space-varying random effects, we use its value at s_k, defined as the center of the grid cell C_k, to compute μ_k.

Other approaches for numerically approximating the integral $\int_{\mathcal{X}} \Lambda(x)\, \mathrm{d}x$ in (7.2) exist. Typically, they use appropriately weighted sums $\sum_k \omega_k \lambda(\tilde{x}_k)$ with discretization points $\tilde{x}_k$ and weights $\omega_k > 0$, which lead to variants of Poisson and logistic regression. In the *Berman–Turner device* (Berman and Turner, 1992), discretization points consist of observed points x_i, $i = 1, \ldots, N$, augmented with *dummy points*. Both point sets together are used to define a partition of the observation window, typically given by the Voronoi tessellation of the observation window using the set of observed points and dummy points as Voronoi cell centers. Weights then correspond to the hypervolume of the Voronoi cells, and this approach allows writing the likelihood

of the log-linear representation (7.1) as a variant of the likelihood of Poisson regression with log-link. Alternatively, if the observed counts N_k in (7.2) are binary, *i.e.*, $N_k \in \{0, 1\}$, then $\mathrm{P}(N_k = 1) = 1 - \exp(-\exp(-(\log(|C_k|) + \mu_k))$. This means that the Poisson regression model with log-link is equivalent to a regression model with Bernoulli response distribution and complementary log-log link function, which is available in standard implementations of GAMs and GAMMs, especially in `R-INLA`. Baddeley et al. (2010) provide an insightful comparative discussion of various representations.

If there are marks, then we can define a GAMM with Gaussian random effects and an appropriate response distribution for the marks, and estimate it also with INLA. Moreover, some of the random effects may be shared between the mark model and the occurrence model, such that stochastic interactions between the occurrence of events and the mark-generating mechanism can be captured. Sharing of random effects is described in Krainski et al. (2018), for instance.

This general framework of regression models with random effects could also be used to estimate point process models with Gibbs-like mechanistic interactions, where random effects may arise in the trend component and in the interaction coefficients. However, since full likelihoods involve an intractable normalizing constant in Gibbs models, we can only work with variants of pseudo-likelihoods. In the setting of Bayesian inference with pseudo-likelihoods, consistency results and the Bernstein von-Mises theorem will ensure convergence of posterior means to the true parameters in appropriately defined asymptotic settings (*e.g.*, Soubeyrand and Haon-Lasportes, 2015); however, the uncertainties conveyed by posterior distributions, such as those obtained from INLA, are wrong.

7.3 Data aggregation and subsampling schemes

Spatiotemporal hierarchical modeling is notoriously computer-intensive due to the large datasets and the numerical challenges with spatiotemporal covariances. With the implementation of INLA in `R-INLA`, up to several hundreds of thousands of observations can be handled, and the recent integration of the PARDISO sparse matrix library within `R-INLA` has further improved numerical stability and speed (van Niekerk et al., 2019). However, stable inferences with INLA may require compromises with respect to the complexity of the latent model and the number of observations, where both aspects jointly determine the size and sparsity structure of the Gaussian precision matrices, and therefore the computation times, memory requirements and the well-conditioned numerical behavior of operations such as solving linear systems and inverting matrices. Usually even stronger restrictions arise with alternative methods such as Markov Chain Monte Carlo (MCMC) to achieve approximation quality comparable to INLA (*e.g.*, Taylor and Diggle, 2014; van Niekerk et al., 2019). Krainski et al. (2018, Section 8.4) develop some strategies for LGCPs by aggregating the events to relatively large mapping units and by lowering the spatial-temporal resolution of Gaussian random effects to decrease model complexity and computing times. However, this solution would cause a deterioration of results especially for small-scale structures (*e.g.*, spatiotemporal prediction for small mapping units and time intervals), and it is problematic with respect to covariates, which would also have to be aggregated to larger spatial-temporal units, thus impeding predictions and interpretation at small scales.

Another useful solution to cope with large space-time datasets consists in subsampling techniques, where the model is estimated using an appropriately reweighted subsample of data points, while keeping the loss of information as small as possible. Subsampling techniques for point processes have been proposed (Baddeley and Turner, 2000; Rathbun et al., 2007; Tokdar and Kass, 2010; Baddeley et al., 2010; Rathbun, 2013; Baddeley et al., 2014). As outlined above, the estimation of LGCP models can be numerically represented as a regression problem with Gaussian random effects. Let us consider the approach of using event counts on a partition of the observation window, where we use the Poisson response distribution with log-link. If the grid is very dense, such that event counts are (almost) all binary (*i.e.*, 0 or 1), then logistic regression would be an interesting alternative; see the discussion in (Baddeley et al., 2010; Lombardo et al., 2018). Since maximum likelihood is equivalent to maximizing the empirical expectation of the log-density of observations, a subsampling scheme is appropriate if it ensures a faithful approximation of this expectation. Subsampling schemes in likelihood-based estimation can be interpreted as importance sampling (Tokdar and Kass, 2010): the original sample with constant observation weight 1 is replaced by a subsample with many observation weights typically larger than 1, and potentially non-constant, to compute the empirical expectation. The history of estimation techniques using weighted subsampling goes back to Horvitz and Thompson (1952).

Based on the representation (7.2) of a LGCP model conditional to Λ, we consider the Poisson intensities Λ_k, $k = 1, \ldots, K$, as the parameters to be estimated, and we aim to select a subsample N_{k_j} with weights ω_j, $j = 1, \ldots, J$, such that the subsample likelihood function is as close as possible to the full data likelihood. To make `R-INLA`-based estimation feasible (van Niekerk et al., 2019), we can devise subsampling schemes that strongly reduce the number of observations in a stratified way (with respect to predictor subspaces) to keep the loss of information and its impact on posterior inferences small. We can use the rule of thumb of Baddeley et al. (2014, 2015) as a guideline: it recommends that the subsample of observations 0 should be larger than the sample of occurrence points by a factor of at least 4, but using higher factors is advised if estimation remains numerically feasible. We do not subsample observations $N_k > 0$ since they are rare and the goal is to appropriately characterize their occurrence; *i.e.*, they are kept with weight 1. For the other observations $N_k = 0$, subsampling can be closely linked to Poisson additivity. Indeed, the weighted likelihood contribution $\exp(-\Lambda_k)^{\omega_k} = \exp(-\omega_k \Lambda_k)$ of the observation $N_k = 0$ with weight $\omega_k \in \mathbb{N}$ corresponds to the likelihood of the weighted sum $\sum_{\ell=1}^{\omega_k} 0 = 0$ of ω_k observations with count 0, and the size of the initial sample with ω_k observations 0 is divided by the factor ω_k. The set of predictors arising in the regression equation (7.1) may be different between all k in a LGCP model (*e.g.*, if there is there is a spatial random effect that varies continuously in space), but one can use the working assumption that this set is often very similar for cells C_k that are close in space (and also close in time, if there is a temporal dimension). A stratified subsampling scheme should result in a known positive sampling probability $p_k > 0$ for each observation N_k in (7.2). Then, the weights that we use for the likelihood of the subsampled observations are given by $\omega_k = 1/p_k$ for selected indices k.

So far, no standard solutions exist for designing efficient subsampling strategies. In particular, future research should investigate into stratification techniques that ensure appropriate subsampling of high-risk subdomains of the predictor space, *i.e.*, of subdomains that may be relatively small in terms of the number of mapping units but that tend to be associated to high occurrence intensities of points.

7.4 Model selection and validation

Model selection and validation of point processes, and particularly of LGCPs, is a wide area with a multitude of approaches developed in the extant literature. In general, model comparison for very sophisticated models, often estimated through approximate estimation procedures such as INLA, should not be based on a single criterion. Different models of various complexity may provide insights into complementary aspects of the same applied problem.

7.4.1 Goodness-of-fit using likelihood-based information criteria

We can compare models through classical likelihood-based information criteria adapted to the Bayesian context: the Deviance Information Criterion (DIC), and the Watanabe–Akaike Information Criterion (WAIC). These goodness-of-fit criteria take the effective dimension of the posterior latent Gaussian model into account, thus penalizing model complexity. Their close relationship to the predictive performance measured through leave-one-out cross-validation has been established, and WAIC is known to better take the stochasticity of the posterior predictive distributions into account (Gelman et al., 2014). With INLA, these quantities are calculated through sensible approximation techniques (Rue et al., 2009)

7.4.2 Assessing posterior predictive distributions

Posterior predictive distributions may be of interest for a variety of quantities, for instance for various spatiotemporal aggregations of event numbers or of event magnitudes. For aggregating event magnitudes in the case of a marked point process with marks representing event magnitude, we can add up the magnitudes for the points observed within the sets forming a partition of the observation window into relevant mapping units.

To focus on criteria evaluating the predictive performance of models, we can implement a cross-validation scheme. Specifically, with an estimation approach involving a division of the observation window into small mapping units, we can randomly partition the set of mapping units into k folds (*e.g.*, $k = 10$), each containing approximately the same number of mapping units (or approximately the same aggregated hypervolume of the mapping units in each fold). We can then calculate predictive scores for various choices of mapping units. If we want to assess predictive performance over mapping units larger than those used for estimation, we can aggregate observed and predicted counts from the smaller mapping units of the model. Predictive scores can assess information about either the predicted counts $\hat{\lambda}_k = \mathbb{E}(\Lambda_k \mid \boldsymbol{y})$, the predicted probabilities of occurrences $\hat{p}_k = 1 - \exp(-\hat{\lambda}_k)$, or the full posterior predictive distribution of $\Lambda_k \mid \boldsymbol{y}$, where $\boldsymbol{y}$ denotes the vector of observed counts used in the GAMM to estimate the LGCP.

An alternative approach for constructing the hold-out sets for cross-validated predictive diagnostics, studied by Leininger et al. (2017), consists of constructing hold-out sets by removing points at random from the point pattern; this is known as *thinning* (Chiu et al., 2013). In general, we prefer the more challenging task of predicting entire spatially-contiguous areas where all data within some mapping units have been removed, which is also more suited to assessing mapping-unit-based hazard predictions.

In yet another approach for very large point patterns, we could also divide the observation window into two sub-windows, one relatively large, used for training (*i.e.*, for estimating the model), and the other kept for validation.

Posterior predictive distributions of various quantities of interest can be estimated by generating a large number of posterior samples of counts (and marks if part of the model) for each cross-validation fit. While INLA does not directly provide posterior samples because of its use of deterministic and not simulation-based approximations, these can be generated conveniently from the estimated posterior model (Rue et al., 2017). Using `R-INLA`'s internal, discrete approximations for posterior distributions of hyperparameters and latent Gaussian fields, the simulation algorithm first generates a realization of the hyperparameter vector; next, conditional on these hyperparameters, a latent Gaussian field is sampled according to the posterior precision matrix computed through Laplace approximation; finally, counts (and marks) are simulated from the mapping-unit-based Poisson distributions (and mark distributions) with intensities defined according to the simulated latent Gaussian fields. Cross-validation results using simulations of the posterior predictive distributions should be based on a relatively large number of samples of the full posterior model (*e.g.*, at least 1000) to avoid too noisy estimation of prediction scores.

7.4.3 Prediction scores for susceptibility

If we consider only presence or absence of events, we can use scores that allow assessing binary classification methods. In particular, the area-under-the-curve (AUC, Fawcett, 2006) measures prediction quality for the presence or absence of points. AUC only considers the relative ordering of predicted probabilities, or of any monotonically strictly increasing transformation of such probabilities, but it does not allow checking for systematic biases. It summarizes the structure of true/false positives negatives for probability thresholds between 0 and 1, and it provides a value in $[0, 1]$, with 1 indicating perfect classification while 0.5 indicates random prediction. However, when we assess predictions for relatively large mapping units where counts larger than one are frequent, the presence/absence distinction may be too coarse, and predictive scores for counts may be more appropriate and informative.

In cases where a model also predicts magnitudes given as point marks, AUC can be used to assess predicted exceedance probabilities over severity thresholds. Alternatively, we can also consider the Brier score for binary outcomes (Brier, 1950). It corresponds to the Mean Squared Error, and is known to be a proper scoring rule in the sense of Gneiting and Raftery (2007).

Due to the complexity of LGCP models with various terms in the linear predictor (7.1), it is recommendable practice to assess not only a single predictive score for comparing and ranking models but rather a basket of several scores measuring different aspects of predictive performance.

7.4.4 Prediction scores for count-valued predictions

For assessing predictions of counts to rank models according to predictive performance, we can calculate the residual sum of squared errors (RSS), the residual sum of absolute errors (RSA), and a χ^2-statistics corresponding to a "studentized" version of RSS, where these three scores use predicted and observed counts. We can further calculate the continuous ranked probability score (CRPS, Gneiting and Katzfuss, 2014) using the predictive distribution functions and observed

counts. RSS and CRPS are proper scoring rules (Gneiting and Raftery, 2007). The formulas for mapping-unit-based RSS, RSA and χ^2 are as follows:

$$\text{RSS} = \sum_{k=1}^{N} (y_k - \hat{\lambda}_k)^2, \quad \text{RSA} = \sum_{k=1}^{N} |y_k - \hat{\lambda}_k|, \quad \chi^2 = \sum_{k=1}^{N} \frac{(y_k - \hat{\lambda}_k)^2}{\hat{\lambda}_k},$$

where $\hat{\lambda}_k = \int_0^\infty \pi_k(\log(\lambda) \mid \boldsymbol{y})\, \mathrm{d}\lambda$ with $\pi_k(\cdot \mid \boldsymbol{y})$ the posterior density of the linear predictor $\log(\Lambda_k)$. The general CRPS formula for a single observation y_{obs} and a corresponding (posterior) predictive distribution $\hat{F}(y)$ from a model may be expressed as $\int_{-\infty}^{\infty} \{\hat{F}(y) - \mathbb{I}(y \geq y_{\text{obs}})\}^2\, \mathrm{d}y$. For the mapping-unit-based CRPS score, we add up the CRPS values over all mapping units and therefore use

$$\text{CRPS} = \sum_{k=1}^{N} \sum_{y=0}^{\infty} \sum_{\ell=0}^{y} \left[\int_0^\infty \exp(-\lambda) \frac{\lambda^\ell}{\ell!} \pi_k\{\log(\lambda) \mid \boldsymbol{y}\} \frac{1}{\lambda}\, \mathrm{d}\lambda - \mathbb{I}(y \geq y_k) \right]^2.$$

For criteria involving larger mapping units different from those used for estimation, the original mapping-unit-based observed counts y_k and intensities λ_k must be aggregated over the predictive mapping units. This requires resorting to the joint posterior distributions of all Λ_k corresponding to the mapping units that we want to aggregate. Since such CRPS formulas are difficult to calculate analytically, we use posterior sampling as implemented in `R-INLA`, and we compute a Monte-Carlo approximation of CRPS values based on a large number of posterior samples.

These three measures rely on counts: while RSS and RSA focus on point predictions defined through the posterior mean of intensities at pixel level, CRPS also accounts for the uncertainty of the predictive distributions and yields good scores for models that provide predictions that are both calibrated (*i.e.*, correct on average) and sharp (*i.e.*, having little prediction uncertainty).

7.5 Mapping

Producing spatial maps of risk components, or time-indexed spatial maps or dynamic maps in the case of spatiotemporal resolution of models, is usually an important goal of modeling the occurrences of environmental risks. Such maps can be produced based on predicted intensities $\hat{\lambda}(x)$, and spatial variability in prediction uncertainties can be highlighted by mapping uncertainty measures associated with $\hat{\lambda}(x)$, such as posterior standard deviations or lengths of credible intervals. More specifically, we can also map posterior means and uncertainties of spatially indexed random effects in the linear predictor (7.1). In particular, we could highlight areas where the random effect is significant, *e.g.*, where credible intervals of $W_k(x)$ do not overlap 0, which puts focus on areas where the other predictor components alone (such as fixed effects) do not succeed in accurately predicting the occurrence intensities.

It is further possible to aggregate predictions of quantities such as counts or magnitudes to various mapping units of interest, and to produce the corresponding maps where the mapped values are constant within each mapping unit.

Chapter 8

Bayesian spatiotemporal modeling and prediction of landslides

Landsliding processes represent a major category among environmental risks. Landslides are nearly ubiquitous phenomena worldwide and pose severe threats to people, properties and the environment in many areas, often causing large numbers of human fatalities and huge economic costs. Investigators have for long attempted to estimate landslide hazard in an effort to determine where, when (or how frequently), and how large (or how destructive) landslides are expected to be in an area. This information may prove useful to design landslide mitigation strategies, and to reduce landslide risk and societal and economic losses. Typical triggers for such movements of land masses are heavy precipitation and earthquakes, and short-to-medium term processes related to vegetation cover and soil erosion may favor the likelihood of landslides. In this chapter, we outline a new modeling framework for landslides due to weather triggers based on log-Gaussian Cox processes (LGCPs). We draw advantage from Bayesian hierarchical modeling using INLA to construct and estimate sophisticated models that enable us to capture the complex influence of numerous observed predictor variables related to geomorphological features and their interactions, and to include spatial or spatiotemporal random effects to represent the contribution of other unobserved environmental factors, especially of the triggering event which is usually not observed at appropriately fine resolution over space and time.

In the literature of geomorphology, most of the attempts at predicting the occurrence of populations of landslides by adopting statistical approaches are based on the empirical observation that landslides occur as a result of multiple, interacting, conditioning and triggering factors. Based on this observation, and under the assumption that at the spatial and temporal scales of our investigation individual landslides are georeferenced points in the landscape, we build our modeling effort on point processes, specifically on a LGCP, recall Chapter 7. Individual landslides in an area form a point pattern generated by an unknown intensity function. The modeling framework has two stochastic components: (i) a Poisson component, which models the observed random landslide count in each terrain subdivision according to a given landslide "intensity", *i.e.*, the expected number of landslides per terrain subdivision; and (ii) a Gaussian component, used to account for the spatial or spatiotemporal distribution of the local observed or unobserved environmental conditions that influence landslide occurrence. The landslide intensity may be transformed into a corresponding landslide susceptibility.

Our seminal paper Lombardo et al. (2018) has introduced the point process concept to the

geomorphological community of landslide modelers, and in follow-up publications (Lombardo et al., 2019, 2020; Opitz et al., 2020a) and ongoing work we capitalize on this approach by providing relevant modeling extensions. In several of these papers, we analyze a collection of 4874 separate debris-flow landslides triggered by a major rainfall discharge in an area of around 100 km^2 on the island of Sicily (Southern Italy), on October 1, 2009. In the recent preprint Opitz et al. (2020a), we propose and compare the most sophisticated of our spatial models, including models from previous papers but also novel models that can account for interaction between the latent spatial random effect and a covariate (*e.g.*, the slope steepness) through the idea of space-varying regression. In this preprint, we also roll out a detailed discussion of how model selection can be achieved with very complex models. Finally, in Lombardo et al. (2020), we also consider the time domain and propose a space-time modeling approach to study the question of stochastic temporal dependence in landslide inventories, which had so far not yet been examined in the geomorphological literature. Based on a landslide occurrence dataset spanning almost a century for a study region in Italy, we give a detailed discussion of various aspects that distinguish the spatiotemporal from the purely spatial case.

8.1 The intensity concept for modeling landsliding risk

The modeling of landslide processes has spawned a rich literature of statistical approaches, see the review of Reichenbach et al. (2018). However, for representing the probability of occurrence of such events only the notion of susceptibility was used for estimation and mapping purposes. This means that one models the probability of presence/absence for the areal units in a fixed partition of space. For instance, one may consider a fine-meshed regular grid of pixels covering geographic space, or one may consider more problem-specific partitions of space, such as slope units (SUs) or administrative units. In contrast to the concept of an intensity function defined over continuous space, as usually employed in point process modeling, there are two drawbacks from focusing attention solely on the probability of presence or absence in a given areal unit. First, a probability of occurrence is associated to a fixed areal unit, such that we cannot easily provide occurrence probabilities for other areal units than those fixed in the model used for estimation. In particular, upscaling of probabilities from smaller to larger areal units is awkward. By contrast, an intensity function allows calculating intensities for arbitrary spatial units, and probabilities for the occurrence of at least one landslide event are readily derived. Second, the notion of susceptibility may be overly simplistic because of its binary nature: it considers only presence or absence, but it does not provide any information on the number of events that may arise in an area, which is especially problematic when relatively large mapping units are used, where many landslides may occur in a single unit. Again, the intensity framework is more general since it allows calculating expected event counts for any spatial support, and the probability distribution of the number of events can be calculated analytically or obtained through Monte-Carlo simulation of estimated point process models.

8.2 Spatial scales and covariates

Geomorphologically relevant covariates are given at high raster resolution. The list of available covariates typically includes variables such as Elevation, Aspect (*i.e.*, the slope angle in $[0, 2\pi)$

describing the exposition of the area with respect to the North), Slope Steepness, Planar Curvature (measured perpendicular to the steepest slope angle and characterizing the convergence and divergence of flow across the surface), Profile Curvature (indicating the direction of maximum slope), the Distance to the closest tectonic fault line, Normalized Difference Vegetation Index (NDVI, measuring the "greenness" of a landscape and serving as a proxy for vegetation), Landform (with around 10 categories), Lithology (*i.e.*, soil type with more than 20 categories, where we summarize rare soil types into a single class "other"), and Land Use (with around 10 categories). In Lombardo et al. (2018, 2019); Opitz et al. (2020a), the original high-resolution raster images of these covariates are aggregated to a 15m $\times$ 15m grid, which yields a representation of the study area through $449,038$ pixels. Then, the LGCP models are discretized through a Poisson regression formulation with respect to this grid, as described through Equation (7.2).

We additionally exploit mapping units at an intermediate resolution (between small pixels and the full study area) known as *slope units* (SUs). The SU partition defines a physically-motivated and moderately-sized spatial discretization, which is useful for capturing latent random effects such as the influence of the spatially-varying precipitation event. SUs can be viewed as relatively homogeneous mapping units with respect to geomorphological and geophysical features relevant to landslide activations. For instance, the study area in Sicily, Italy, is composed of 3484 SUs; see Figure 8.1. Some landslides triggered within the same SU may be due to a joint triggering mechanism, which could lead to some residual stochastic dependence in the landslide occurrence process (conditional to the geomorphological structure and the precipitation trigger), but such spatial dependence is likely to be very weak when considering events arising in separate SUs. In other words, landslide data in different SUs can be safely assumed to be conditionally independent given fixed and random effects, such that the LGCP framework is appropriate.

8.3 Modeling nonlinear responses and unobserved triggers

We first summarize the most relevant spatial model structures and then present a selection of important results.

8.3.1 Model structure

We define a log-linear model for the point process intensity of landslides where we may include fixed effects of observed covariates and several types of random effects. Since usually the observations of the landslide trigger (*e.g.*, precipitation) are not available at satisfactorily high spatial resolution, we include a latent spatial random effect in the linear predictor of the point process intensity to capture the influence of such unobserved and spatially varying effects. Moreover, some covariates are known to have a strongly nonlinear influence on the log-intensity of the point process. Therefore, in some cases it is important to allow for nonlinear covariate effects through the use of spline-like random effect terms. In practice, a simple and robust solution consists of using piecewise constant splines with a first-order random walk prior, as implemented for the models in this chapter. PC priors can be specified for the random walk precision parameter in order to estimate spline curves with appropriate posterior smoothness. More generally, we systematically use PC priors for all estimated hyperparameters in LGCP models for landslides.

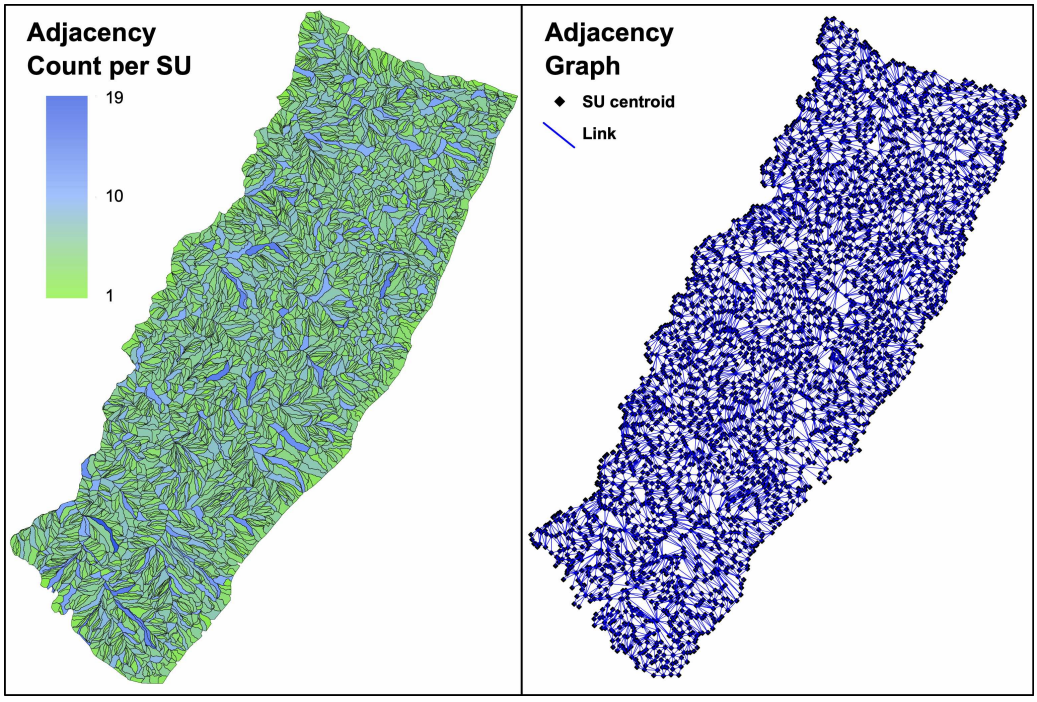


Figure 8.1: Illustration of the slope units (SUs) in the Sicily study area in Italy, used to define latent Gaussian random effects with conditionally autoregressive spatial structure. Left display: map indicating the number of adjacent SUs. Right display: adjacency graph of SUs.

In Opitz et al. (2020a), we define a baseline model including fixed effects, a spline effect for the Aspect covariate, and a spatial random effect, as follows:

$$\log \Lambda_0(s) = \sum_{j=1}^{J_{\text{cont}}} \beta_j^{\text{cont}} z_j^{\text{cont}}(s) + \sum_{j=1}^{3} \sum_{\ell=1}^{\ell_j} \beta_{j,\ell}^{\text{cat}} z_{j,\ell}^{\text{cat}}(s) + x_{\text{Aspect}}^{\text{CRW1}}(s) + x_{\text{SU}}^{\text{CAR}}(s). \tag{8.1}$$

The precisions in the centered Gaussian priors of fixed effect coefficients β_j^{cont} estimated for continuous variables are fixed to obtain moderately informative priors, without any hyperparameters to be estimated. By contrast, we use PC priors for the precision parameters of coefficients β_j^{cat} of factor effects (*i.e.*, of categorical effects), where a sum-to-zero constraint is imposed for each of the factors (Lithology, Landform, Land Use in our case) to ensure identifiability. A separate precision parameter τ is used for each factor with the *a priori* specification that $\text{P}(\sqrt{1/\tau} > 1) = 0.01)$. In a similar way, we specify the hyperpriors for the marginal precisions of the cyclic random walk prior used for the Aspect covariate, cyclic over a subdivision of $[0, 2\pi)$ into 16 categories, and of the conditionally autoregressive (CAR) spatial effect (Besag, 1974) defined at the level of SUs based on their adjacency structure, again with a sum-to-zero-constraint for identifiability reasons.

We explore several extensions of the baseline regression equation (8.1). We can include additional i.i.d. effects specified either at the 15 m pixel level or the SU level, or additional nonlinear spline effects for other covariates, such as the Slope Steepness (in which case we remove its fixed effect), or an additional space-varying regression (SVR) term for covariates such as the Slope Steepness. In our most complex model, we include both a nonlinear spline

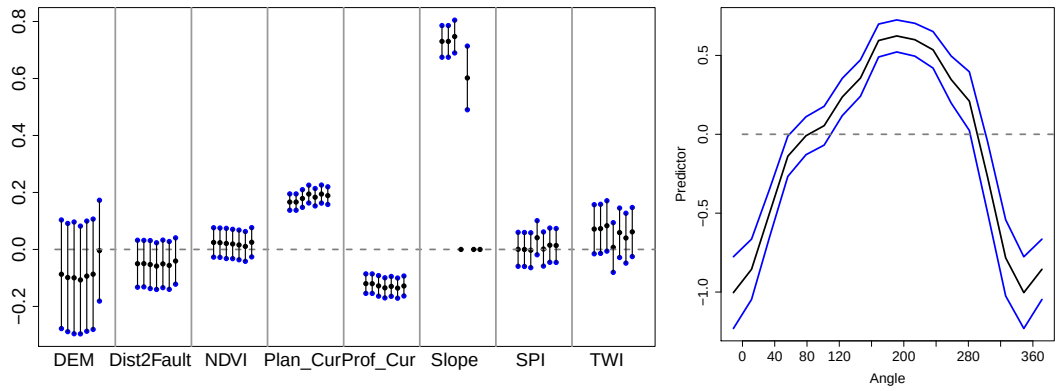


Figure 8.2: Posterior estimates (mean, 95% credible bounds) of covariate effects on spatial landslide intensity in the Sicily area. Left: fixed effect coefficients for the different models in Opitz et al. (2020a). Right: Nonlinear cyclic effect of Aspect in the baseline model (8.1).

effect $x^{\text{RW1}}_{\text{Slope}}$ and a SVR coefficient $x^{\text{CAR}}_{\text{Slope}}$ with spatial CAR prior for the Slope Steepness, leading to the following regression formula:

$$\log \Lambda_{\text{ext-1}}(s) = \log \Lambda_{0,-\text{Slope}}(s) + x^{\text{RW1}}_{\text{Slope}}(s) + \text{Slope}(s) \times x^{\text{CAR}}_{\text{Slope}}(s), \tag{8.2}$$

where $\Lambda_{0,-\text{Slope}}(s)$ is the equation of the baseline model after removal of the fixed effect of Slope Steepness.

Finally, we construct a model similar to the preceding one but which links the latent spatial effect $x^{\text{CAR}}_{\text{SU}}(s)$ and the SVR component. If the latent spatial effect acts as a proxy for the precipitation trigger, then its low values indicate a weak or absent trigger effect, and then the Slope Steepness value could become irrelevant since no landslides occur, whatever the geomorphological conditions. In this case, the SVR component can locally neutralize the globally estimated Slope Steepness effect. We here consider the following model:

$$\log \Lambda_{\text{ext-2}}(s) = \log \Lambda_{0,-\text{Slope}}(s) + x^{\text{RW1}}_{\text{Slope}}(s) + \beta \times \text{Slope}(s) \times x^{\text{CAR}}_{\text{SU}}(s), \tag{8.3}$$

with the interaction coefficient $\beta \in \mathbb{R}$ to be estimated. Unlike the most complex model (8.2) (where complexity is understood in terms of the number of latent Gaussian variables), this model features only one single CAR effect, $x^{\text{CAR}}_{\text{SU}}(s)$, instead of the two a priori independent effects, $x^{\text{CAR}}_{\text{SU}}(s)$ and $x^{\text{CAR}}_{\text{Slope}}(s)$, such that we consider it as a parsimonious variant of space-varying regression.

8.3.2 Summary of spatial modeling results in the Sicily area

Covariates included as fixed effects show relatively stable estimates of their posterior coefficients when we compare models with differently specified random effects, except for cases where the random effect is used to capture a covariate's effect partially or fully, as it is the case for Slope Steepness in models (8.2) and (8.2); see the left display in Figure 8.2. The right display in Figure 8.2 exemplifies the estimation of a nonlinear effect using the example of Aspect, for which a cyclic random walk prior was used to make angles 0 and 2π identical.

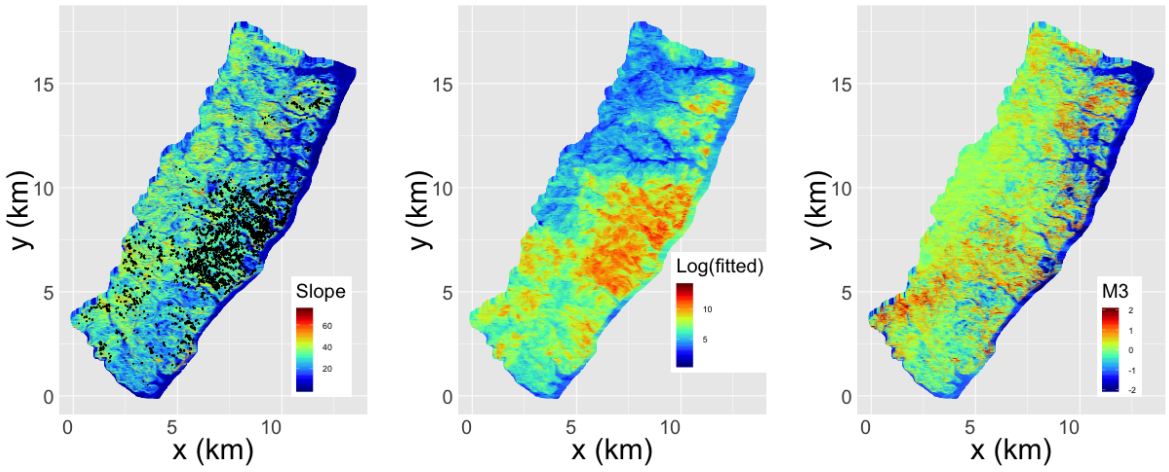


Figure 8.3: Mapping of components of predicted landslide intensities for the Sicily study area. Left: Slope Steepness covariate. Middle: Predicted landslide log-intensity of the baseline model (8.1). Right: Combined effect of Slope Steepness components in the space-varying regression model (8.2).

In Figure 8.3, certain components of the predicted landslide intensities are shown: a map of the Slope Steepness – a key covariate – is plotted on the left display; the spatial prediction obtained from the baseline model (8.1) is given in the middle display; the combined effect of the predictor components related to Slope Steepness in the most complex model (8.2) with SVR is shown in the right display. Both the middle and the right display reveal the spatial distribution of the strength of the precipitation trigger, which entered the study area from the West (around the middle of the vertical axis), and then crossed the study area while losing intensity and exiting through the southwestern boundary.

Finally, Figure 8.4 shows why the parsimonious formulation of the SVR component with respect to Slope Steepness in model (8.3) is a sensible choice. The patterns of the slope contribution to the linear predictor are very similar between the model with and without SVR when we plot this contribution as a function of the value of the global spatial random effect $x_{\text{SU}}^{\text{CAR}}(s)$ and the Slope Steepness value. This holds especially in the region where the small gray dots are concentrated, *i.e.*, where we have observations of landslide counts. These results corroborate the following interpretation of the parsimonious model: the interaction between the latent spatial effect and the Slope Steepness allows us to establish a direct link between the trigger intensity and the Slope Steepness effect, and the latter effect is mostly irrelevant in regions where the trigger was absent or only weak.

We further compare models based on information criteria such as WAIC, and by using a 10-fold cross-validation scheme where SUs are held out entirely. Cross-validation allows us to check for overfitting with very complex models and to assess the models' performance in out-of-sample prediction.

For predictive performance, we compare a number of useful scores to contrast observations and predictions either at the pixel level or at the SU level; recall §7.4.2 for a summary of such scores. Some scores use only presence/absence-information (*e.g.*, AUC), while others allow using the actual count value (*e.g.*, CRPS). We summarize main findings from Opitz et al. (2020a). The

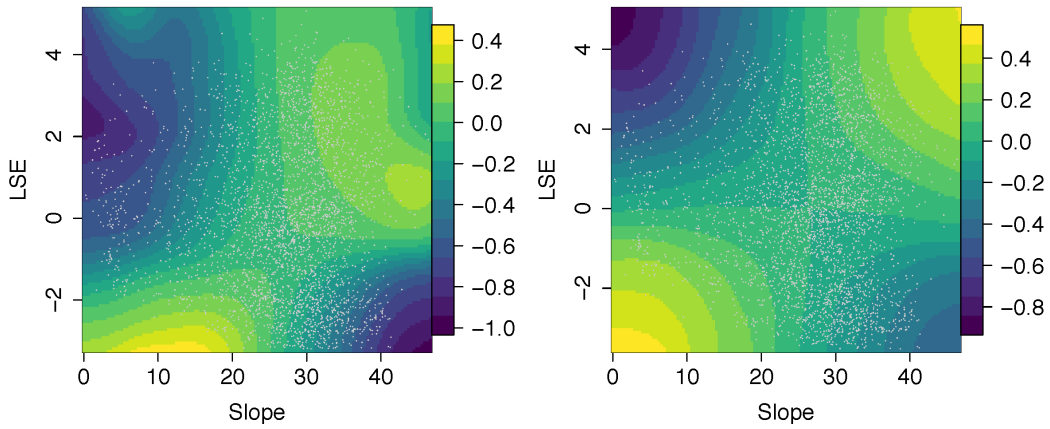


Figure 8.4: Illustration of the Slope Steepness effect in models with space-varying regression. The overall Slope effect is shown through the color map in a coordinate system given by the Slope value and the posterior mean of the latent spatial effect. Light grey dots indicate the position of the observed SUs in this coordinate system. Left display: most complex model (8.2) with an independently specified global spatial random effect. Right display: parsimonious model (8.3) where the SVR-coefficient corresponds to the rescaled global spatial random effect.

results show that the inclusion of a nonlinear effect of the Slope Steepness, implemented through a piecewise constant spline curve with random walk prior, is important for good prediction. We point out that the baseline model is already relatively competitive and shows stable and quite good performance throughout, and it does not suffer from several relatively bad count-based scores arising for some of the more complex models. Overall, the ranking of models based on their predictive performance looks quite different from the one based on WAIC (although scores and WAIC values are often close between different models). A possible reason is that very high stochasticity and complexity of prior models may lead to more unstable, noisy predictions.

In general, we recommend a careful inspection of the fitted models based on several criteria, for goodness-of-fit and for out-of-sample prediction. Models such as (8.2) and (8.3), which possess extra flexibility thanks to an SVR component, are relatively competitive overall for prediction. Despite their increased computational complexity, such models are useful by offering insights into the "physical" interaction of Slope Steepness (or of other covariates) with the unobserved precipitation trigger, to which we have alluded in the discussion of Figure 8.4. In terms of WAIC, the inclusion of an additional SVR component for the Slope Steepness (but without using a spline curve effect for this covariate) provides the most substantial improvement over the baseline model. The relatively complex models (8.2) and (8.3) also lead to substantial improvement in WAIC, while no relevant improvement in goodness-of-fit could be achieved by adding spatially independent random effects at the pixel or SU level.

8.4 Modeling spatiotemporal dependence over several decades

Some available landslide datasets comprise inventories for several triggers or several time periods. So far, the existing landslide literature had tackled statistical modeling of such data from a purely spatial perspective, *i.e.*, by assuming that subsequent time periods or time steps are

linked solely through covariate coefficients that are constant over time. In Lombardo et al. (2020), we go beyond this restriction and propose spatiotemporal modeling approaches with a focus on revealing temporal dependence patterns that cannot be explained through available geomorphological predictor variables alone. To improve over the existing techniques, we therefore construct innovative LGCP models that consider latent random effects of (i) spatial, (ii) temporal, and (iii) spatiotemporal type among adjacent terrain mapping units, among the same mapping units but subsequent in time, or using both conditions together, respectively.

We use a data inventory covering the period from before 1941 to 2014 for the Collazzone area in in Central Italy, for which most of the relevant geomorphological covariates are also available as spatial rasters. We subdivide the 79 km^2 area into 889 slope units (SUs). In this work, we do not retain the very high pixel resolution for defining models, mostly for computational reasons. Instead, we directly use observations given at SU resolution by appropriately aggregating landslide counts and covariate information to the SU scale. For continuous covariates, both the SU-wise empirical average and the empirical standard deviation of raster-based covariate values are used as predictors. The aggregation through the standard deviation provides information about the homogeneity of covariate values with a SU. We further conduct appropriate preprocessing steps with the original landslide data, consisting of spatial inventories for 19 periods, to aggregate them into 6 relatively homogeneous time intervals, which are approximately of the same length except for the first interval spanning a longer time. Figure 8.5 illustrates the original spatiotemporal distribution of the landslide dataset and its aggregation towards 6 time intervals. Figure 8.6 gives an illustration of the SUs, and in particular of their adjacency structure used for defining the conditionally autoregressive spatial random effects in the LGCP models.

8.4.1 Model structure

The model structure in space corresponds to the spatial baseline model (8.1). In all of the considered spatiotemporal models, we define moderately informative PC priors for hyperparameters. As in the spatial models, all hyperparameters are estimated, except for the precisions of the centered Gaussian priors of fixed effects of continuous covariates.

We consider a simple baseline model where the spatial covariates enter only as fixed effects. In this model, we allow for time-interval-specific regression constants,

$$\log\left(\Lambda_j^{\text{base}}(s_i)\right) = \log(|A_i|) + \beta_0 + \beta_j + \sum_{k=1}^{8} \beta_{k,1}^{\text{morph}} z_k^{\text{mean}}(s_i) + \sum_{k=1}^{8} \beta_{k,2}^{\text{morph}} z_k^{\text{sd}}(s_i) + \sum_{k=1}^{13} \beta_k^{\text{them}} z_k^{\text{prop}}(s_i), \tag{8.4}$$

where $j = 1, \ldots, 6$ indexes the time intervals and each s_i, $i = 1, \ldots, 889$, corresponds to a different SU with area $|A_i| > 0$. Overall, the model comprises 35 covariate coefficients β to be estimated. In Equation (8.4), fixed effects are separated according to the SU-wise means and standard deviations of continuous morphometric variables with superscript "morph", and according to the 13 categories of thematic properties with superscript "them", expressed through SU-wise proportions. We impose a sum-to-zero constraint on the 6 time-interval-specific intercepts β_j to ensure identifiability of the global intercept β_0, and the common precision parameter of these 6 intercepts is estimated. This time-varying intercept can be considered as a relatively simple temporal random effect.

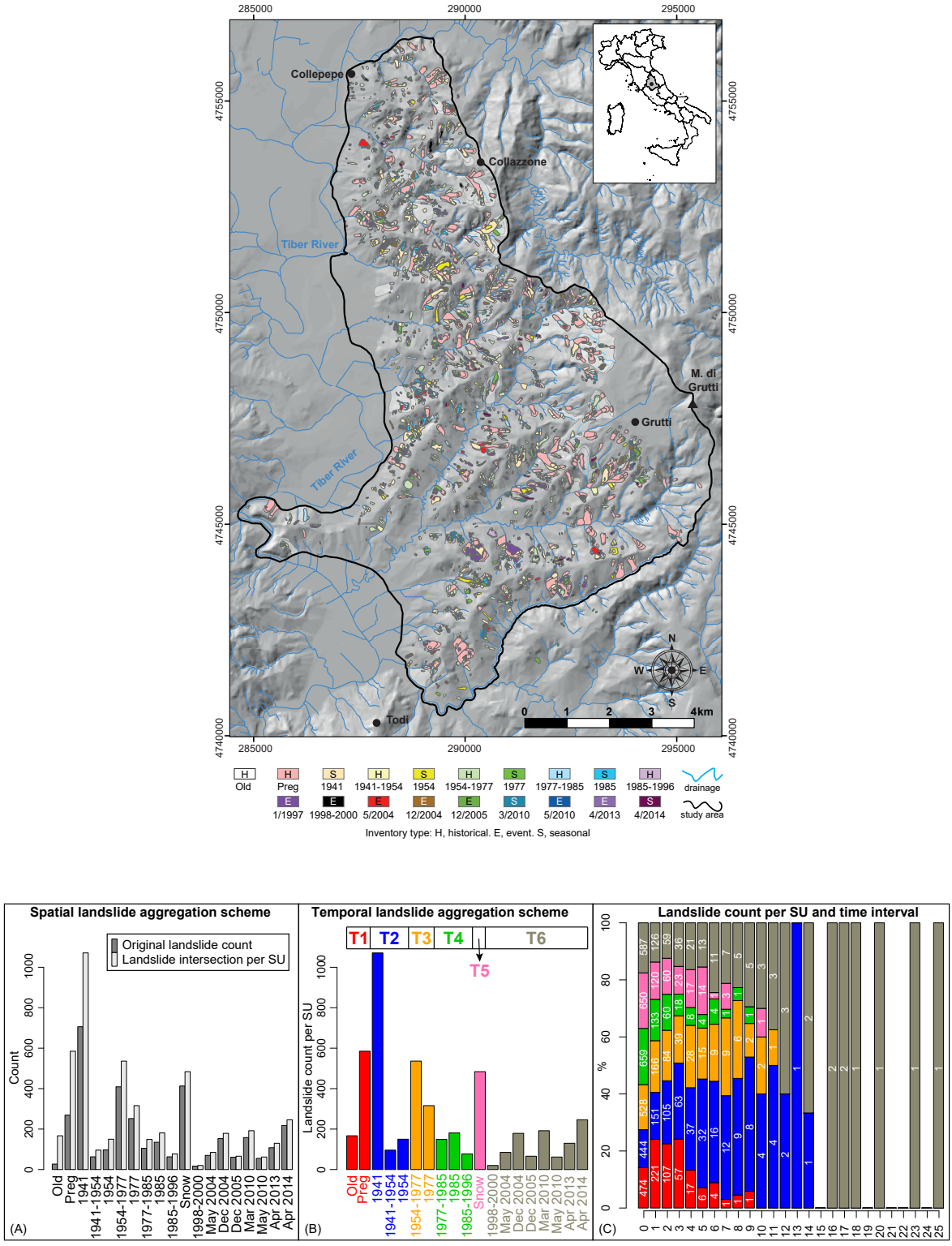


Figure 8.5: Illustration of the spatiotemporal landslide inventory of the Collazzone area, Central Italy. The upper display shows the spatial distribution of landslide scars according to the 19 original time intervals. The lower display puts focus on the temporal structure and the temporal aggregation into 6 relatively homogeneous time intervals.

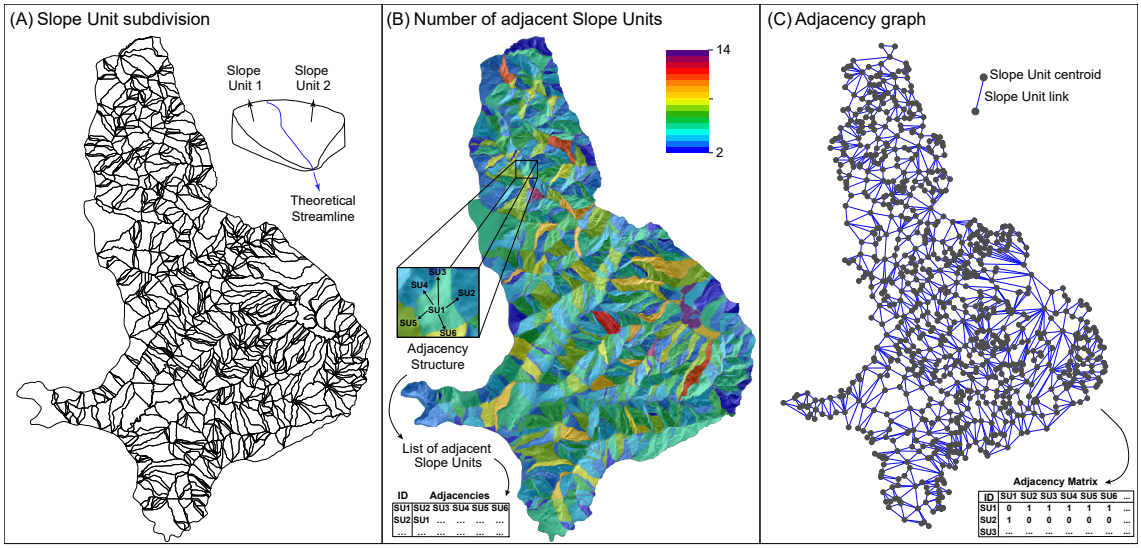


Figure 8.6: Illustration of SUs in the spatiotemporal landslide inventory of the Collazzone area, Central Italy, and of their adjacency structure used for defining conditionally autoregressive spatial random effects.

To extend the baseline model in (8.4), we add spatial random effects according to three different specifications. Our assumption prior to estimating models is that the spatial random effect should differ between SUs and between different time intervals, but that it may be similar for observations of landslide counts arising close in space or in time. In a first extension, we use a Gaussian random effect with conditionally autoregressive (CAR) spatial dependence and 6 independent replicates in time. The role of the CAR effect is to explain variations in the landslide intensity that cannot be explained by the observed covariates, and the full model can be written as follows:

$$\log\left(\Lambda_j^{\mathrm{S}}(s_i)\right) = \log\left(\Lambda_j^{\mathrm{base}}(s_i)\right) + W_j^{\mathrm{S}}(s_i), \quad i = 1, \ldots, 889, \; j = 1, \ldots, 6. \tag{8.5}$$

In a second extension, featuring a structure similar to (8.5) at first sight, we now assume prior temporal dependence of SU-based random effects within the same SU, while disregarding any direct spatial relationship between adjacent SUs:

$$\log\left(\Lambda_j^{\mathrm{T}}(s_i)\right) = \log\left(\Lambda_j^{\mathrm{base}}(s_i)\right) + W_i^{\mathrm{T}}(t_j), \quad i = 1, \ldots, 889, \; j = 1, \ldots, 6, \tag{8.6}$$

using first-order auto-regressive (AR1) Gaussian time series $W_i^{\mathrm{T}}(t_j)$ within SUs, whose prior distributions are mutually independent between SUs. Finally, we combine the two preceding extensions with either spatial or temporal dependence into a model that features both dependence mechanisms in its spatiotemporal Gaussian random effect:

$$\log\left(\Lambda_j^{\mathrm{ST}}(s_i)\right) = \log\left(\Lambda_j^{\mathrm{base}}(s_i)\right) + W^{\mathrm{ST}}(s_i, t_j), \quad i = 1, \ldots, 889, \; j = 1, \ldots, 6, \tag{8.7}$$

where the Gaussian process $W^{\mathrm{ST}}(s_i, t_j)$ combines a spatial CAR structure with SU-wise temporal AR1 structure. This model is able to capture spatial and temporal dependence in a parsimonious way by estimating the spatial precision parameter of the CAR model and the temporal

correlation parameter, constraint to the interval $(-1, 1)$, for the temporal AR1 structure. This is the most complex spatiotemporel model that we consider here. It can learn about spatial clustering and temporal persistence in the structure of the landslide-triggering mechanism.

8.4.2 Summary of spatiotemporal modeling results

Similar to the model comparison tools developed for spatial models, see §8.3, we compare models using visual displays but also numerical scores based on presence-absence or on counts of landslides in SUs for each of the 6 time intervals. For checking out-of-sample prediction performance, we consider two cross-validation schemes: *Space-CV*, which is 10-fold and stratified in space by holding out the same SUs over the 6 time intervals; *Time-CV*, which is 6-fold since we hold out full time intervals, one at a time. Overall, we identify the model (8.7) with both spatial and temporal dependence in its spatiotemporal random effect as the best model with good relative and absolute performance in both Space-CV and Time-CV. This model provides the additional convenience of interpreting spatial and temporal dependence parameters. We find substantial temporal correlation with a significant AR1 correlation parameter. Figure 8.7 summarizes results for out-of-sample validation using either Space-CV or Time-CV.

Based on time-interval-specific SU-based posterior intensity estimates of the landslide point process, we further propose a novel classification approach to identify spatiotemporal regions with similar posterior susceptibility and intensity behavior; see Figure 8.8 for the resulting maps. We distinguish between *clearly stable* and *clearly unstable* SUs, and we further consider two intermediate classes of *rather stable* or *rather unstable* SUs; see Lombardo et al. (2020) for details.

Finally, as an outlook towards joint modeling of occurrence locations and sizes of landslides using a marked LGCP model, we can study the relationship of landslide counts (observed or predicted) to observed aggregated landslide areas in the spatiotemporal Collazzone dataset; see Figure 8.9. Interestingly, the plots highlight a strong dependence with an almost linear correlation pattern between intensities and aggregated landslide areas, such that the intensity of the point process model seems to provide a good proxy for the aggregated magnitudes of landslides, up to a scaling factor. If we consider the area of a landslide as a mark of the landslide point, then this exploratory finding hints at marks whose univariate distribution could be independent and identical across space and time.

8.5 Outlook towards modeling extensions

Based on the foregoing work presented in this chapter, we have identified a number of important extensions that we have started to implement in ongoing work.

A first extension towards multi-type point processes will allow for modeling of several types of landslides (*e.g.*, shallow vs. deep, small vs. large) in a single model. Predictor components, such as covariates or random effects, can have the same or different influence in the linear predictors of the intensity functions associated to the different landslide types. Random effects capturing triggers and other unobserved environmental conditions may be shared between the intensity functions of several landslide types. Joint modeling of several landslide types may therefore provide insights into the similarities and differences among the generating mechanisms

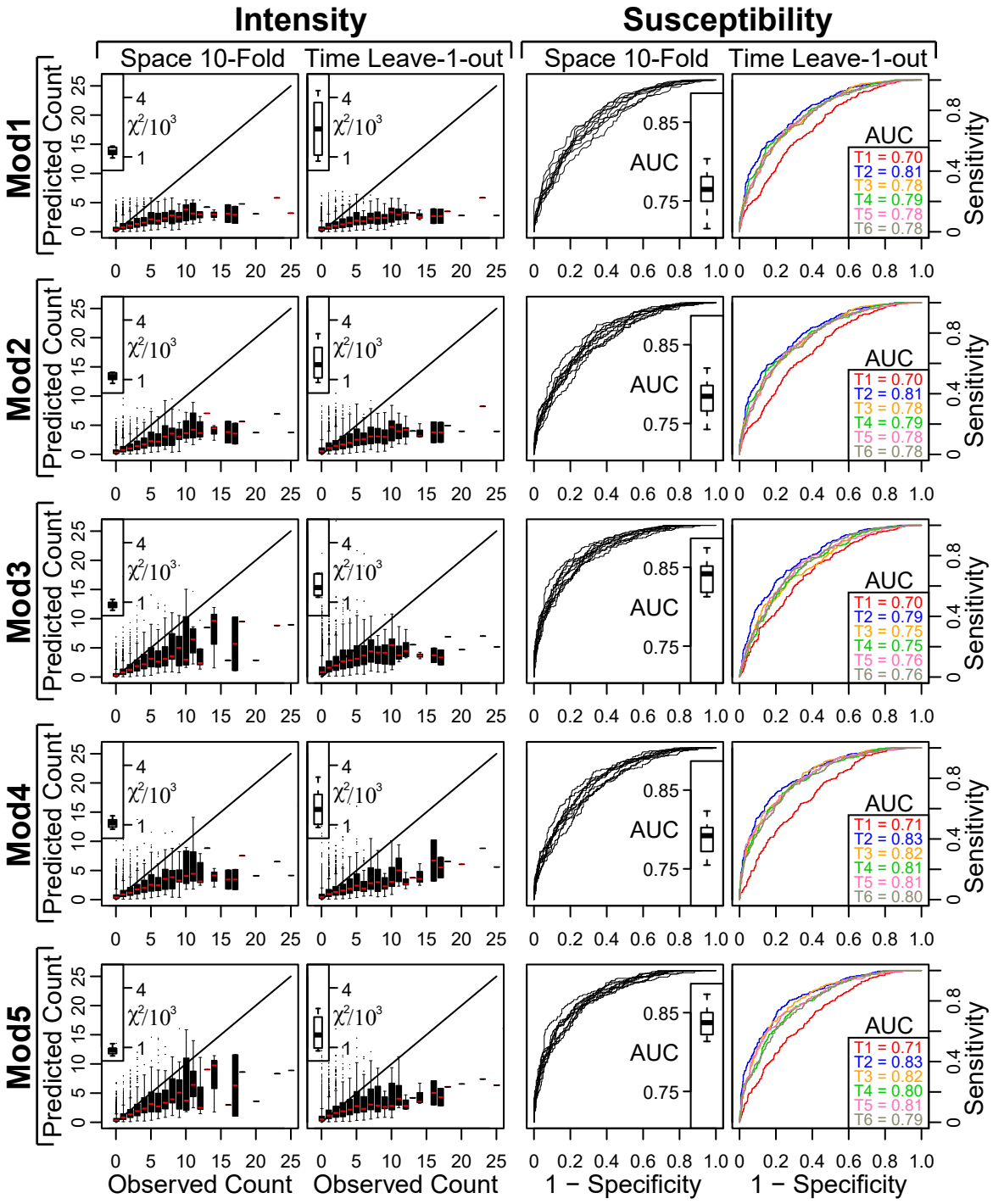


Figure 8.7: Out-of-sample validation of the models fitted to the spatiotemporal landslide inventory in the Collazzone study area in Italy, using Space-CV and Time-CV.

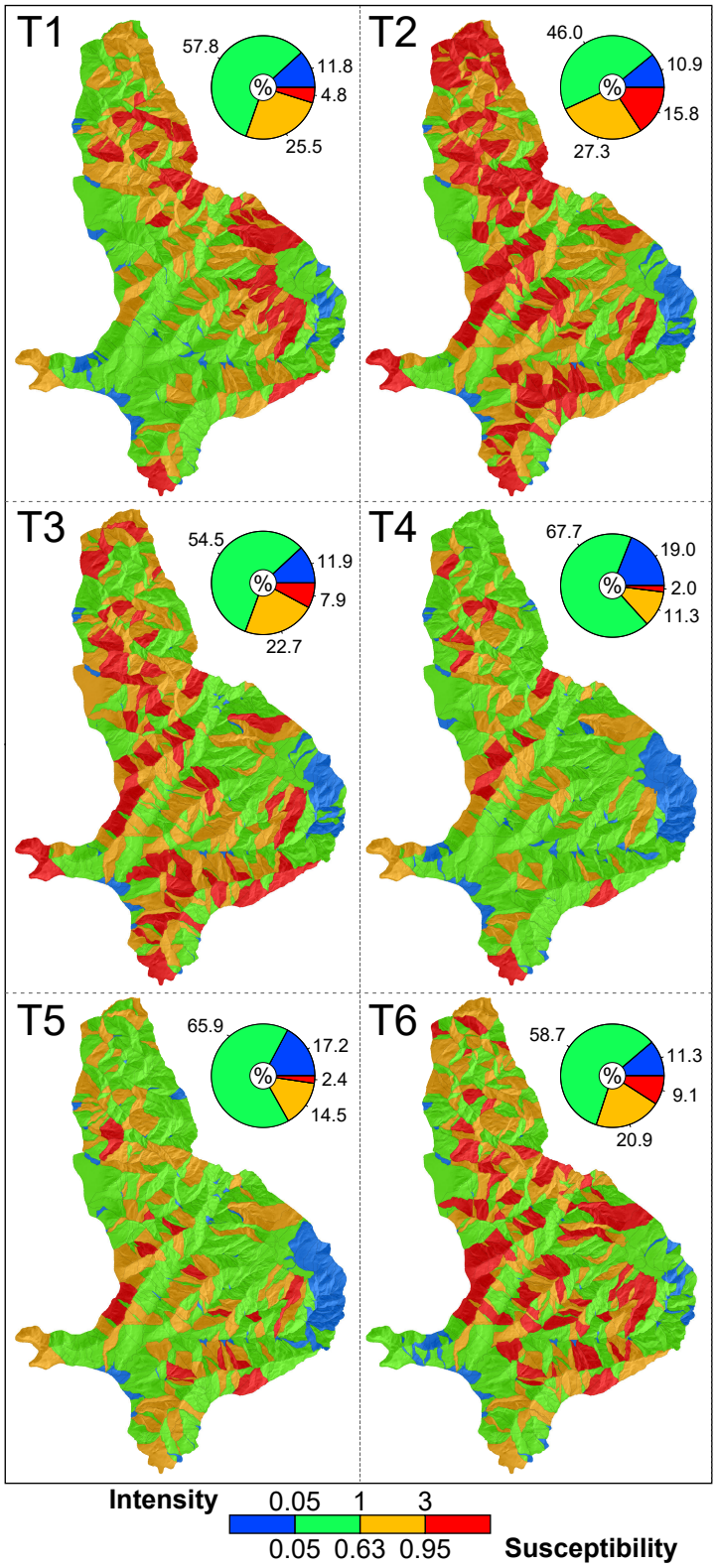


Figure 8.8: Classification of SUs and time intervals according to the posterior intensities of the spatiotemporal LGCP model (8.7). For each time interval, pie charts show the percentage of SUs falling into one of the four considered classes.

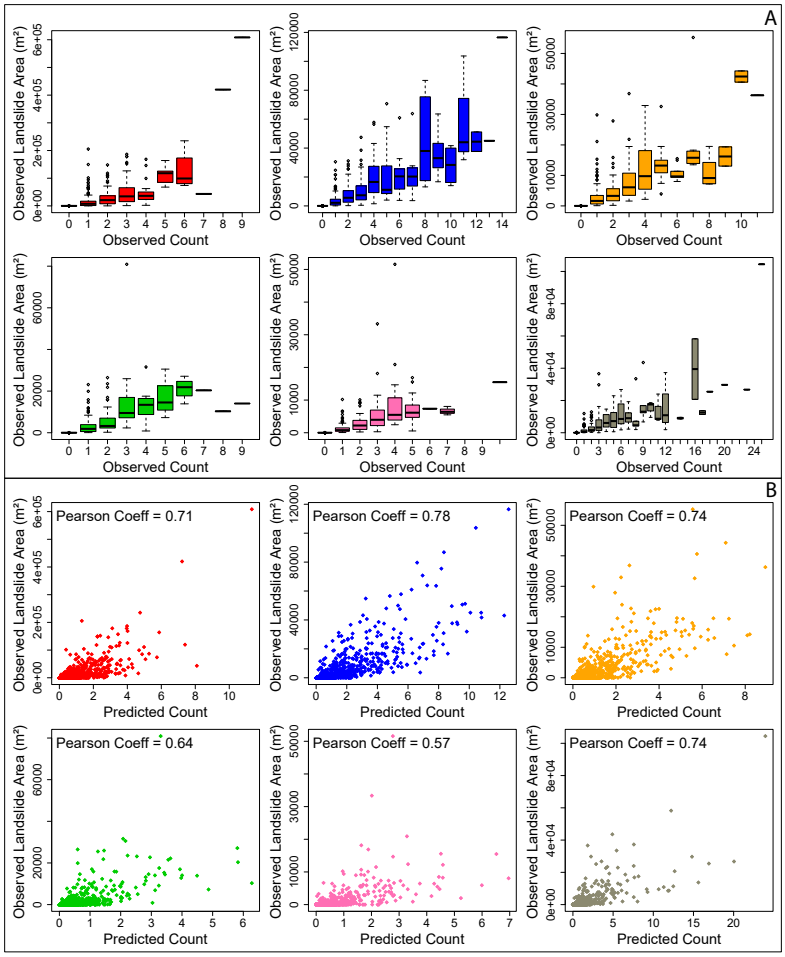


Figure 8.9: Relationship of landslide counts (observed in upper two rows, predicted using model (8.7) in lower two rows) to observed aggregated landslide areas.

of different landslide types. Moreover, the estimation of components that are shared between different types (where possible and appropriate) is a means to reduce estimation uncertainties. A promising approach consists of using a multi-type Poisson process with inhomogeneous intensity functions $\lambda_i(s)$, $i = 1, \ldots, m$ for the m different types as starting point, such that the process obtained by superposing all point patterns is still of Poisson type with intensity function $\lambda(s) = \sum_{i=1}^{m} \lambda_i(s)$, and marks follow a multinomial distribution whose probability parameters are given by $\lambda_i(s)/\lambda(s)$, $i = 1, \ldots, m$. Through the inclusion of random effects, which may be correlated or shared between point types, we obtain a flexible modeling framework of multivariate log-Gaussian Cox processes, and statistical inference can be performed in a Bayesian framework using the same tools as in the single-type case, *i.e.*, we can use INLA and further combine it with the SPDE approach, see Chapter 7. `R-INLA` provides functionality to implement such models through the use of several response vectors corresponding to multiple regression equations.

With the purpose of modeling the magnitude of landslides, the aforementioned approach of defining a multi-type point process according to different magnitude categories is a practical solution, but another avenue would consist of using point processes with a continuous mark space to exactly model the observed magnitudes. This would establish a predictive modeling approach that simultaneously answers three questions: *where* will landslides occur? *how many* will occur? *how large* will they be? As pointed out in the summary of spatiotemporal results in §8.4.2, a preliminary analysis revealed a quite strong, seemingly linear correlation between the observed/predicted counts for SUs and the aggregated landslide areas. This could hint at the mark distribution being independent from the local occurrence intensity, but a deeper analysis and proper modeling tools are required to examine this relationship, which might show different behavior on other landslide inventories. By analogy with the multi-type case, a marked Poisson process could provide an appropriate framework, and the use of random effects leads to a marked LGCP. Its intensity function $\Lambda(s)$ and the distribution of a mark attached to an event s could be positively or negatively correlated, *i.e.*, a larger number of events may locally come along with relatively larger or smaller event magnitudes, respectively. The choice of an appropriate probability distribution for marks is also very important, and again the mechanisms for estimating multiple regressions in INLA with the possibility of sharing random effects between different responses would allow for fully Bayesian inference.

Finally, the joint stochastic modeling of observed triggering weather events and landsliding events could be tackled in a spatiotemporal setting at a relatively coarse spatial scale. For regions where relatively complete landslide inventories spanning several decades have been constructed, we could additionally collect historical weather data from stations within the region or nearby. Then, a joint spatiotemporal model could be developed for extreme precipitation episodes and landsliding activity, with the aim to reveal the interactions between the two processes. This idea represents an example of joint stochastic modeling of multiple risks (or of a risk and its driver(s)) and their interactions, which is one of the major research perspectives to my work that will be further discussed in the scientific outlook in Chapter 11.

Chapter 9

Bayesian spatiotemporal modeling and prediction of wildfires

Wildfires are uncontrolled fires of vegetation. The annual burnt area worldwide is on average approximately 3.5 million km^2 (Giglio et al., 2013), but the characteristics of fire activity in terms of numbers, size, intensity and severity vary substantially over space and time, depending on many factors related to weather, climate, vegetation, orography, as well as local and regional human influences. Wildfires represent major environmental and ecological risks worldwide and show strong interactions and feedbacks with many other climatic, environmental and ecological processes. They provoke many human casualties and substantial economic costs, and they can trigger extreme air pollution episodes and lead to the important losses of biomass and biodiversity. Climate change may further exacerbate their frequency and extent. Moreover, wildfires contribute an important fraction of global greenhouse gas emissions, such that they can further accelerate climate change and drive problematic feedback loops in the Earth system when climate extremes further increase wildfire activity.

9.1 Challenges in wildfire modeling

In the study of wildfire activity, explanatory and predictive modeling approaches are both important. Explanatory models allow us to attribute components of wildfire activity to risk factors, such as land cover and land use types and their spatial structure, meteorological conditions and socio-economic variables. This informs wildfire management about action levers to reduce risks. On the other hand, predictive models with spatiotemporal resolution aim to produce precise risk maps and forecasts. In particular, this may require incorporating random effects in models to capture spatiotemporally structured variability of wildfire risk that cannot be explained by the available predictor information from one of the aforementioned risk factor classes.

In France, most wildfires are caused by human activity (accidents or arson). Natural causes such as lightning are much less frequent and amount to around 10% of wildfire occurrences in Europe. A major difficulty in accurate modeling is to assess and quantify the influence of human activity and its spatiotemporal dynamics. Usually only relatively static proxy variables are available for this risk factor (*e.g.*, spatially resolved indices characterizing wildland-to-urban interface, touristic zones, and their evolution over longer time periods such as years or decades).

Dynamic data of mobile phone activity could help gain better understanding and predictions of human behavior related to wildfires, but data collection and exploitation raise strong concerns about privacy. Another important variable is fuel and soil moisture, for which only relatively rough spatiotemporal estimations can be derived from the observation of meteorological and surface variables at irregulary spaced weather stations, from satellite-based remote sensing, and from climate reanalyses that assimilate observation data into climate models. Finally, wildfire prevention policies and measures may not be well documented, or applied heterogeneously, in space and time. Overall, it is therefore highly challenging to disentangle how the space-time dynamics of wildfire activity across years are driven by climate cycles, global warming and changes in preventive measures, human activity and fuel structure.

The study of wildfire activity has led to a large body of statistical and machine learning literature (see the reviews of Preisler et al., 2004; Xi et al., 2019; Pereira and Turkman, 2019) and to the rise of a number of specialized scientific journals, with the goal to provide methods for identifying risk factors, producing risk maps and proposing wildfire management policies to reduce wildfire activity. Most statistical studies focus on modeling either occurrence numbers or sizes, the latter usually represented by the burnt areas of spatially and temporally contiguous wildfire events. In occurrence modeling, the spatial or spatiotemporal pattern of ignition points (or other representative points of separate wildfire events) can be analyzed with point process tools. Often, data are available as presences/absences or counts over dense spatial or spatiotemporal grids; or data have been transformed to such representations in order to facilitate modeling and to harmonize different spatial-temporal scales of wildfire and auxiliary data (*e.g.*, weather conditions, land cover, land use).

A first component of wildfire activity that we aim to model is the occurrence intensity, especially the location, time and numbers of wildfire ignitions (*where? when? how many?*). This can be achieved through point process models. If presence/absence in relatively small grid cells is considered, then binary classification models can also be used. Another key characteristic of wildfire impacts is burnt area. It can provide a direct approximation of the loss of biomass and emission of greenhouse gases, and it allows for interpretation with respect to loss of ecosystem services. Distributions of wildfire sizes usually show quite heavy tails, such that a small number of wildfires have a very dominant contribution to aggregated burnt areas. Moreover, spatial and temporal predictability of the occurrences and sizes of wildfires may be low, especially in the case of large wildfires. Therefore, while extreme values in wildfire sizes represent only a small fraction of the sample, it is very important to carefully address their distributional properties and to identify the presence of spatial and temporal structures in their occurrence patterns as far as possible. A large number of univariate probability distributions has been explored for wildfire sizes, but no consensus has arisen so far on which parametric distribution family provides the best fit (Pereira and Turkman, 2019). Specifically, distributions suggested by extreme-value theory, such as the generalized Pareto distribution (GPD, §2.2.1) arising for threshold exceedances, have been studied (*e.g.*, approaches by De Zea Bermudez et al., 2009; Turkman et al., 2010; Amaral-Turkman et al., 2011; Pereira and Turkman, 2019), and we will use them to model exceedances above a range of increasingly high severity thresholds in Pimont et al. (2021).

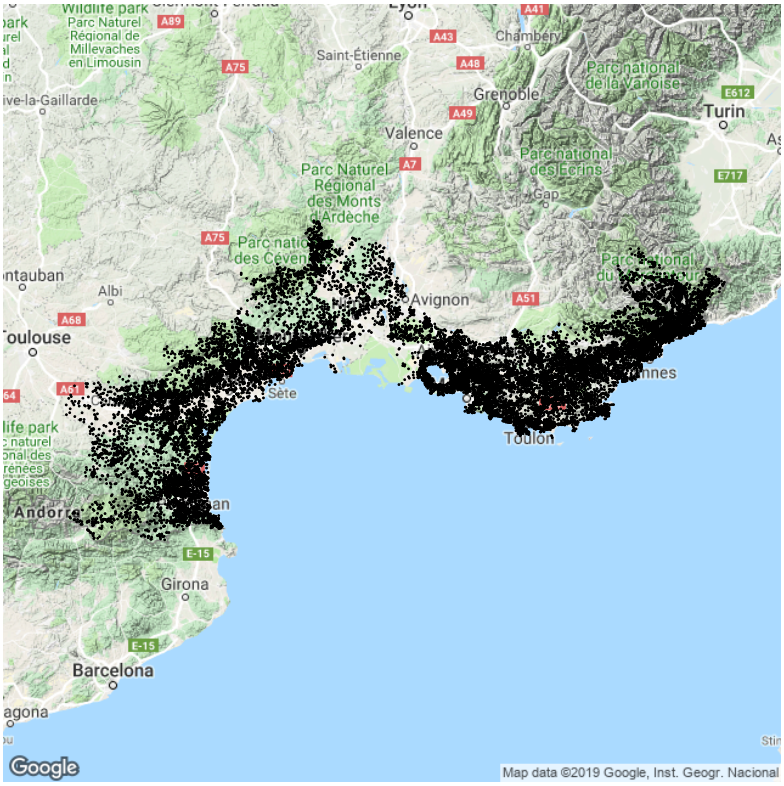


Figure 9.1: Prométhée wildfire dataset (without the island of Corsica here). The map shows the locations of the 23,309 wildfires observed during the 1995–2018 period in one of the 9,562 DFCI grid cells in the Mediterranean basin in Southern France.

9.1.1 Multi-scale multi-source datasets

Another issue in data analysis and predictive modeling is the multi-source multi-scale structure of relevant variables and data. Daily wildfire occurrences in France are available from the Prométhée database since the 1970s; they are georeferenced using the so-called DFCI grid at 2 km resolution. A map showing spatial position of more than 20,000 wildfire occurrences recorded in the DFCI grid during the period 1995–2018 is presented in Figure 9.1. Weather variables are provided by Météo France, *e.g.*, as observations of irregularly-spaced weather stations, or as SAFRAN reanalysis data at 8 km grid resolution. Land cover and land use (LCLU) variables, especially forest cover, can be obtained from the CORINE land cover database and from databases of the national IGN service. Some databases are available in vector format as shapefiles, while others are of raster type. Some temporal dynamics in LCLU variables can be exploited since databases have been updated at several occasions. Besides multi-scale structures in the datasets, we also expect that different spatial and temporal scales, not necessary congruent with the data scales, are of interest for wildfire modeling and prediction.

9.1.2 Weather predictors

Wildfire activity shows a strong response to weather conditions, where especially dry and hot conditions are very fire-prone. In Gabriel et al. (2017), we have done some exploratory analysis

with respect to weather variables for the Bouches-du-Rhône administrative region in Southern France. We aimed to understand the temporal scales of weather factors, here given by precipitation, temperature and wind speeds. Based on the observation series for a weather station in the center of this study domain, Figure 9.2 illustrates these three variables by differentiating their values for days with and without wildfire occurrences. Low precipitation and high temperatures appear clearly as factors favoring wildfire occurrences, while the role of wind speeds is less clear.

Instead of using raw weather variables, we can also work with one of the fire weather indices defined in the literature. These indices aggregate weather variables (and sometimes other inputs such as LULC variables) into a single structure variable, whose high values ideally represent weather conditions that are particularly prone to wildfires. In the papers Fargeon et al. (2018) and Pimont et al. (2021) discussed below, we focus on the widely used Fire Weather Index (FWI van Wagner, 1977), originally established for forest regions in Canada. The FWI first aggregates weather conditions into fuel moisture indices for vegetation and top soil layers, which are then further aggregated into indices for buildup (related to wildfire occurrence) and spread (related to wildfire sizes). The final FWI value is obtained by combining buildup and spread indices. Other wildfire indices exist (*e.g.*, the McArthur Forest Fire Danger Index developed in Australia, or the Fosberg Fire Weather Index developed in the USA), but we focus on the Canadian FWI in this chapter.

9.1.3 Land-use and land-cover predictors

Wildfire activity is also very strongly conditioned by how land is covered and used. First, the availability of fuel is related to the vegetation cover (forest, type of tree species, other natural or cultivated vegetation) and its fuel moisture state due to weather conditions. In dry conditions, some tree species are known to ignite more easily and then allow for faster spreading of wildfires (*e.g.*, coniferous forests, by contrast with deciduous forests). In contrast, areas covered by water (lakes, rivers), and more generally very humid zones, do not suffer from wildfire activity. In France, most wildfire ignitions are caused by human activity, such that the presence of buildings and population density, transport networks and touristic infrastructures play an important role for the spatiotemporal variation of wildfire activity. Especially the wildland-to-urban interface, corresponding to transition zones between densely populated urban areas and thick forest cover, is known to represent relatively high wildfire risk. Figure 9.3 illustrates four variables related to land use and land cover (LULC).

9.1.4 Building novel tools for the analysis of wildfire activity

The goal of my work on wildfire activity is to provide novel exploratory tools and Bayesian hierarchical models to improve the exploration and modeling of the complex multi-scale space-time structures arising for wildfire occurrences. Since the largest wildfires account for most of the total aggregated burnt area of all wildfires, particular attention is required to accurately model the tail of the wildfire size distribution, especially of exceedances above high thresholds.

Some novel exploratory tools, as well as modeling approaches based on INLA and the SPDE approach, are proposed in our early paper Gabriel et al. (2017) for the spatiotemporal analysis of wildfire occurrences using data from the Bouches-du-Rhône area in Southern France. We explore the role assigned to temporal scales of the weather variables given as prepitation, temperature

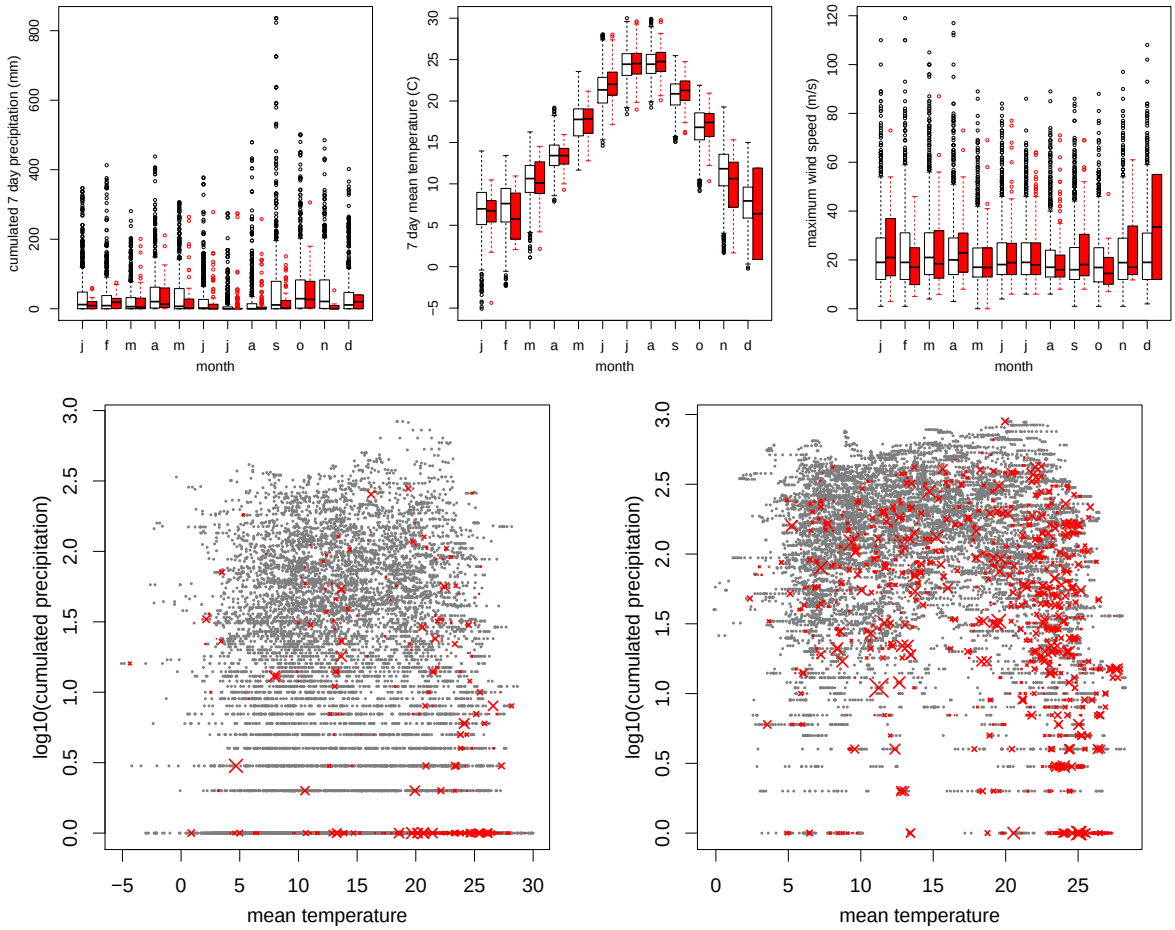


Figure 9.2: Exploratory plots for wildfire-weather relationships for the Bouches-du-Rhône département. Top row: Monthly boxplots of weather data, where red boxplots represent the subsample of days with observed wildfires; the left display shows 7 day cumulated precipitation preceding the reported day; the middle display shows 7 day average temperature preceding the reported day; the right display shows wind speed for the reported day. Bottom row: scatterplots of weather data for all days of the study period: the left display shows average 7 day temperature preceding the reported day and base-10 logarithm of (1+cumulated precipitation); the right display shows the same averages and aggregations but over 28 days preceding the reported day; in both display in the bottom line, the days with wildfire occurrences are shown in red, and the size of the $\times$-symbol is proportional to the wind speed observed at the day.

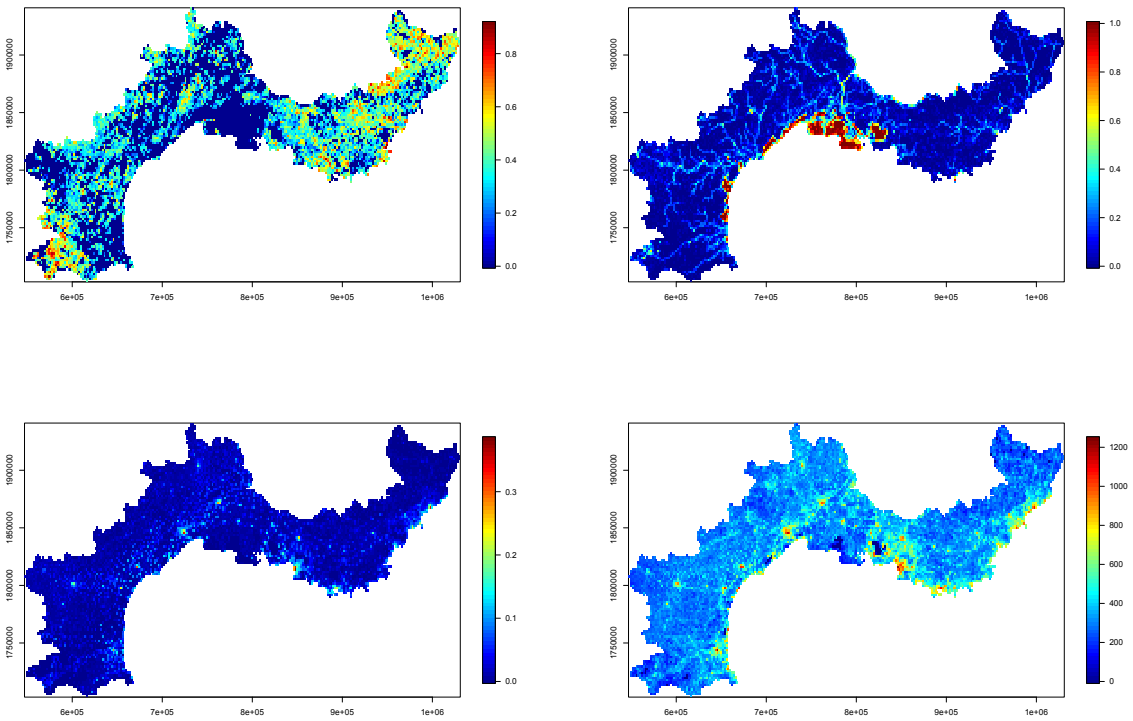


Figure 9.3: Examples of LULC covariates for the Prométhée area. Top left: coverage of coniferous trees; top right: water coverage; bottom left: building coverage; bottom right: total length of roads (in kilometers).

and wind speeds in the point process intensity of wildfire ignition points. Moreover, using spatiotemporal inhomogeneous K-functions (Illian et al., 2008) implemented in the `stpp` package of `R`, we study how differently specifiied spatiotemporal model structures allow us to capture spatiotemporal clustering of wildfire ignition points, and more specifically if space and time are separable in the clusters not explained by the spatiotemporal intensity functions of the models. The models were estimated at annual temporal resolution and have been obtained either through kernel intensity estimation or as variants of models including weather and LULC covariates, as well as different specifications of random effects (spatial; spatial, and temporal random walk; temporally replicated spatial random effect) in the intensity function. Some of the exploratory analyses implemented in Gabriel et al. (2017) would merit to be deployed at larger spatial scales over the full Prométhée area in Mediterranean France.

In our paper Opitz et al. (2020b), detailed in §9.2, we develop spatiotemporal point process modeling (without marks) for the whole Prométhée area using INLA and the SPDE approach to study the influence of a wide variety of LULC-related pedictor variables, some of them designed specifically to provide information about the wildland-to-urban interface. The model has resolution at monthly scale and uses the 2 km pixel resolution of the Prométhée wildfire database. Random effects were used to explain residual seasonal effects and the spatiotemporal variation at annual scale not explained by the observed covariates (LULC, precipitation, temperature).

In §9.3, we present results from the work in Pimont et al. (2021), focused on spatiotemporal joint modeling of occurrences and sizes, especially of large wildfires. In this approach, we use the FWI to capture weather influence, and we explore through nonlinear spline effects how well it captures variations in the occurrence and size responses at daily scale. All components of the model are estimated using the INLA-SPDE approach, except for the exceedance distribution corresponding to the category of largest wildfires, where we use a frequentist maximum likelihood estimator of the parameters of a GPD model with negative shape (not available with INLA). We baptize this model *Firelihood*, and it is a solid basis for work in progress on several modeling extensions to further improve its predictive performance and its interpretability.

9.2 Spatiotemporal occurrence modeling with focus on land cover and land use

For this work (Opitz et al., 2020b), we have considered the spatiotemporal point pattern (without marks) of wildfires in the Prométhée study region (without the island of Corsica) for the years 1995–2018, and we work with a monthly resolution for time and with the 2 km DFCI grid for space. We have put particular attention to constructing a relatively large number of potentially useful covariates related to LULC.

9.2.1 Land-use land-cover covariates

We first unify datasets with different spatial resolution towards the DFCI grid by constructing summary covariates of the land cover variables which are available at very high spatial resolution from the IGN service We here do not consider temporal changes in LULC variables, which may be relatively small during the study period. First, we calculate proportions of forest cover type and buildings, and lengths of road types, for an intermediate regular grid at 200 m resolution.

Then, for each DFCI cell at 2 km resolution and each covariate, we compute the mean and standard deviation of the 100 subgrid values. This two-step approach keeps valuable information concerning the fine-scale structure of covariate values when we aggregate them to the DFCI grid. The mean summary shows the overall trend inside each DFCI cell, and the standard deviation measures the variability around this trend and informs about small-scale heterogeneity. Moreover, we generate synthetic covariates as interactions of several of the original covariates in order to highlight the interface of forest areas to human activity. Wildland-to-urban interfaces are heterogeneous areas where the two main prerequisites of wildfires, vegetation and human activity, are expected to strongly coincide with wildfire outbreaks. We compute two additional covariates to take into account the interface between open forests and urbanized areas, and between open forests and paths. For this, we multiply the percentage of open forests by the percentage of buildings or by the total length of paths inside each DFCI cell, respectively.

For weather influence, we here work with covariates obtained through the spatiotemporal kriging of observations of precipitation and temperature.

9.2.2 Model structure

The spatiotemporal log-intensity function of the log-Gaussian Cox process in Opitz et al. (2020b) has the following additive structure, which includes 30 covariates z_j^{land} related to LULC and 3 covariates $\hat{z}_j^{\text{clim}}$ related to weather conditions:

$$\log \Lambda(s,t) = \beta_0 + \beta^{\text{time}}\tilde{t} + \sum_{j=1}^{30} \beta_j^{\text{land}} z_j^{\text{land}}(s) + \sum_{j=1}^{3} \beta_j^{\text{clim}} \hat{z}_j^{\text{clim}}(s,t) \tag{9.1}$$

$$+ f(\text{month}(t)) + W(s,a(t)); \tag{9.2}$$

the first line in the above formula shows fixed effects with coefficients β_j^{type} to estimate, and the second line shows random effects. Based on a transformation $\tilde{t} = \frac{t-t_{\min}}{t_{\max}-t_{\min}}$ of the observation period to the interval $[0,1]$, a linear trend is estimated through the fixed effect coefficient β^{time}. A Gaussian space-time random effect $W(s,a(t))$ is defined at the level of the years $a(t)$ associated with the observation month t. A seasonal effect $f(\text{month}(t))$ is defined at monthly resolution with 12 levels. The hat-notation $\hat{z}_j^{\text{clim}}$ indicates that climate covariates have been estimated beforehand through kriging of monthly observation series recorded at 17 weather stations. We represent the Gaussian fields $W(s,a(t))$ by using the SPDE approach (see §A.2.1 for background), such that we work with a Gauss–Markov approximation of the Matérn covariance function, see Figure 9.4 for the spatial triangulation mesh that we use. For different spatiotemporal structures in $W(s,a(t))$, we considered four choices, and we here report only the finally selected (and most sophisticated) prior model of time-stationary autoregression defined as follows:

$$W(s,a(t)) = \rho W(s,a(t)-1) + \sqrt{1-\rho^2}\varepsilon_{a(t)}(s), \quad \rho \in (-1,1), \tag{9.3}$$

where $\varepsilon_{a(t)}(s)$ are the spatial Matérn-SPDE innovation fields, independent between years.

Estimation is carried out using INLA.

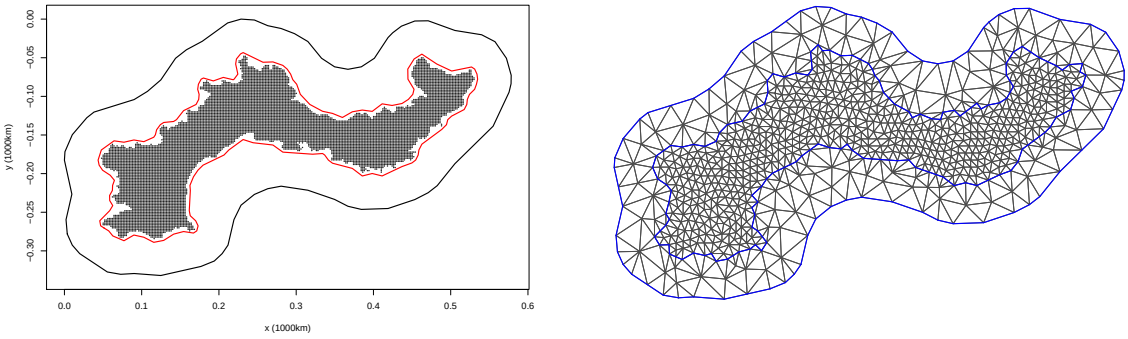


Figure 9.4: Spatial triangulation mesh for the spatiotemporal residual effect of the model of Opitz et al. (2020b). Left: DFCI grid cells and internal and external boundaries for the mesh construction (1,000 km unit). Right: triangulation mesh structure for the INLA-SPDE approach.

9.2.3 Main results

The posterior distributions of hyperparameters of the spatiotemporal residual effect $W(s, a(t))$ indicate relatively strong year-to-year dependence (with posterior mean of autocorrelation at 0.89) and relatively small spatial range (with posterior mean at 20 km). The posterior mean of 1.4 of the standard deviation of $W(s, a(t))$ implies that a substantial part of the spatiotemporal variability in occurrence intensities of wildfires cannot be explained by the LULC and climate covariates incorporated into this model.

Regarding estimated fixed effects, we first highlight a significant decreasing linear time tend over the study period, which corresponds to an almost 40% drop in the point process intensity when comparing the beginning and the end of the 24-year study period. In the context of a warming climate with a tendency towards higher positive temperature anomalies and more arid climate in the Mediterranean basin, which makes weather conditions more favorable to wildfires in principle, this opposite effect may be interpreted as a consequence of improved preventive measures.

The weather and LULC covariates expressed through the mean over the DFCI cell, which contribute to increase the probability of occurrence of forest fires, are the following: temperature anomaly, road length, proportion of conifer trees, slope, and areas of regulated touristic zones. On the other hand, high values of the following factors significantly decrease this probability: altitude, precipitation anomaly, water coverage (*e.g.*, lakes, rivers), building coverage, length of secondary roads, and length of paths. For the French Mediterranean area, our results therefore precisely quantify the influence of factors that are known to favor or limit the occurrence of a forest fire. We believe that altitude, with the strongest coefficient in absolute value, already summarizes a lot of information in the study area. Indeed, low altitude levels can be found near the coastline where the climate is the most Mediterranean, hot and dry, with highly flammable plant species (conifers) and strong human presence (buildings, roads, tourism). By contrast, at higher altitudes temperatures are lower, precipitation is higher, vegetation is less present and less favorable to ignition, and human activity is lower or more strongly supervised in the case of tourism. The total length of roads can be seen as a proxy for human presence, while

conifers are strongly present in the study region and represent a highly flammable tree species. The slope provides complementary information with respect to altitude; for instance, along the Mediterranean coast we find areas with low elevation but with steep slopes (creeks, valleys) and high touristic activity. Slope is also known to be a factor of faster propagation of forest fires, therefore it seems logical to find its coefficient to be significantly positive. The significance of the regulated tourist zone covariate might be explained by the fact that, despite the efforts of conservation and prevention of the forest areas, touristic pressure is so high that it increases the risk of wildfire occurrences. Subregions of the study domain with a very high proportion of buildings (urban areas) or water naturally have a much lower level of forest fire exposure. According to our model, the presence of many secondary roads and paths tends to limit the occurrence of a forest fire. This characteristic contrasts with an opposite effect observed for the total length of all roads, highlighting the major impact of primary roads.

Both interface effects (forest-to-building, forest-to-paths) are significant. The forest-building covariate increases the fire occurrence intensity, while the forest-to-paths factor leads to a decrease. Forest-to-building interfaces concentrate the main cause of wildfire outbreaks: human activity in a forest environment. The risk reduction owing to the forest-to-paths factor is more difficult to explain, but may be due to the fact that the presence of small dirt roads and paths does not necessarily coincide with fire-hazard-prone human activities.

Among the class of standard deviation covariates derived from LULC data, a group of variables significantly increases the probability of fire occurrence: secondary road length, forest cover, path length, building cover, shrubland, conifers and moorland; two other variables are associated to a significant decrease of wildfire numbers: road length, altitude. Overall, these effects are in line with our general understanding: locally heterogeneous environments, where human activity often coincides with presence of combustible material, favor the outbreak of wildfires. The effects of the different road types are not always easy to disentangle, but our model shows that taking into account averages and variances of road lengths in DFCI grid cells for different road types (primary, secondary, paths) considerably improves the goodness-of-fit.

In summary, unusually high temperatures and low precipitation result in strongly increased fire occurrence risk, and we observe a strong effect of LULC, especially of areas where human activity (agriculture, recreation, tourism) takes place in the presence of forest cover. In particular, DFCI cells with high average forest cover alone were not found to be exposed more strongly to fire risk in a significant way; rather, the presence of buildings and forest together, the dominance of coniferous trees, and a fragmented forest cover have been identified as important factors (among others) contributing to increased fire occurrences.

In Figure 9.5, the estimated residual seasonal effect at monthly resolution is shown. It is bimodal, with a large peak in summer but also a second lower peak in spring. For both peaks, we conjecture that the state of the combustible material cannot be characterized completely by the temperature and precipitation covariates. In particular, the spring peak may be due to relatively dry vegetation at the end of winter; moreover, agricultural activity could contribute to accidental wildfire ignitions in this season.

Finally, Figure 9.6 shows maps of posterior predictions of expected wildfire numbers for the summer months of June to September in 2017. Clearly, very high occurrence numbers arise in July and especially in August, and spatial nonstationarities are strong. Predicted intensities are particularly low in the blueish area around the coordinate $x = 0.8$, which corresponds to marshland, *i.e.*, to very humid land cover.

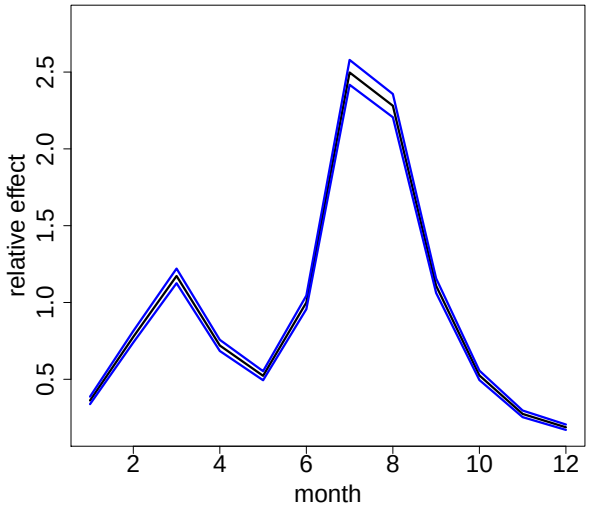


Figure 9.5: Estimated seasonal effect in the final model of Opitz et al. (2020b). The black curve shows the posterior mean of the odds ratio with the month of June as reference (*i.e.*, with its value scaled to 1), with 95% credible intervals (blue curves).

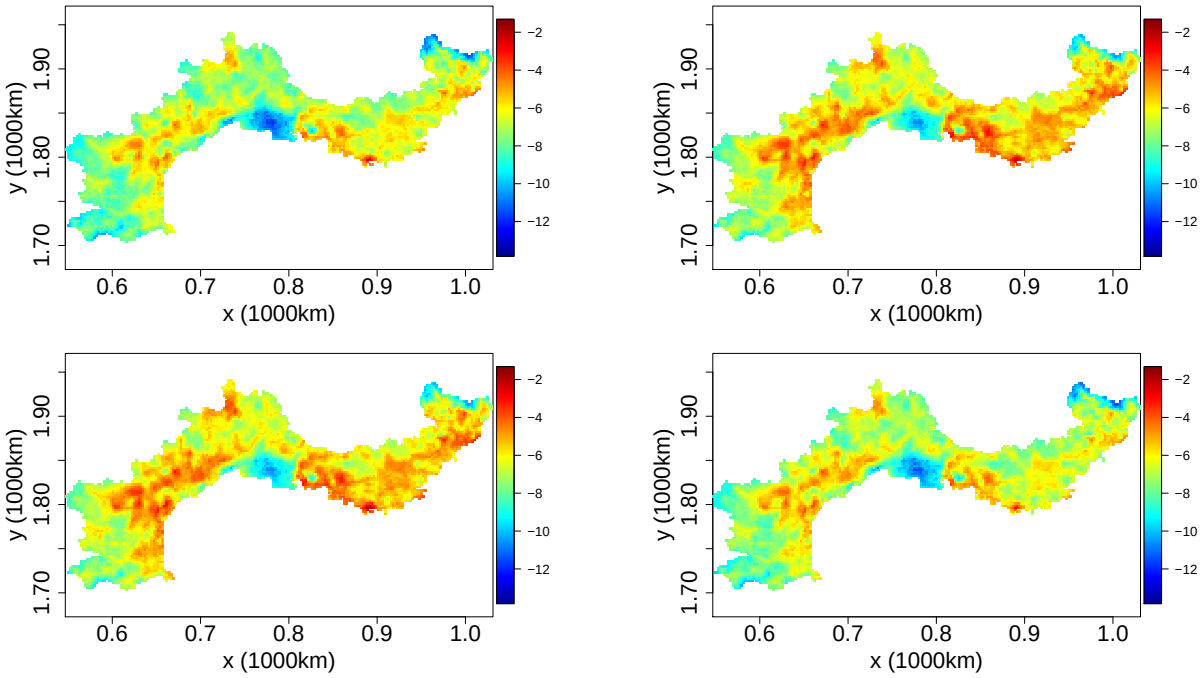


Figure 9.6: Estimated log-intensity functions $\log(\Lambda(s,t))$ for the months of June to September in 2017 based on the final model of Opitz et al. (2020b), with unit km^{-2}. Top row: June, July. Bottom row: August, September.

9.3 *Firelihood*: Joint modeling of occurrences and sizes with focus on weather influence

In the work described in Fargeon et al. (2018); Pimont et al. (2021), we propose joint modeling of wildfire occurrences and sizes. This project was developed independently from Opitz et al. (2020b), which explains some differences in modeling choices. In contrast to the relatively strong focus on new statistical methodology for large space-time point patterns in Opitz et al. (2021), the approaches promoted in Fargeon et al. (2018); Pimont et al. (2021), and in ongoing work on extensions, are more strongly motivated by practical wildfire management considerations. The notion of wildfire occurrences here consists of considering only *escaped* wildfires defined by a burnt area equal to or exceeding 1 ha, *i.e.*, wildfires that have not been extinguished at an early stage; we discard smaller wildfires.

Firelihood is designed as a Bayesian hierarchical regression model with several types of response variables to represent daily fire activity. It corresponds to a marked point process in which individual fires are the points (occurrence component), and the fire sizes (*i.e.*, burnt areas) are the marks (size component). Using a grid approximation of space and a daily resolution for time, a space-time Poisson model with log link function and random effects for the occurrence component is adjusted to gridded fire counts using INLA combined with the SPDE approach to define Gaussian process priors for spatial random effects. To appropriate characterize the wildfire sizes, and especially extreme wildfires, the size model is based on a mixture distribution, where the mixture components have support over disjoint intervals used to subdivide $[1, \infty)$. Mixture components correspond to truncated Pareto and generalized Pareto distributions and are also estimated with INLA except for the most extreme interval. The Fire Weather Index (FWI) and Forest Area (FA) are used as explanatory variables, We focus on the role of the weather drivers in this model, which does not explicitly use LULC predictors. Seasonal and spatial random effects, as well as a post-2003 effect, are included to improve the consistency of the relationship between climate and fire occurrence.

For spatial discretization we use the regular 8 km grid given by the SAFRAN reanalysis of Météo France. We consider this voxel size (8×8 km^2 $\times$ 1 day) as a good approximation for the "true" Poisson distribution of occurrence numbers resulting from aggregating intra-pixel variation, since estimated pixel-per-day probabilities remain small. Another criterion for the choice of pixel resolution was that, to appropriately capture spatial variation, the pixel extent should be by a multiple smaller than the range of spatial correlation estimated in the spatial random effect. Our model fits show that this range is approximately 30 km, which is indeed substantially larger than pixel size. Moreover, most fires observed in the study area spread for less than a day and are smaller than 1000 ha, which is much smaller than the area of a pixel. Therefore, it is also appropriate to stick to the above voxel scale for fire size modeling, even if a very small number of fires spread over more than one voxel. Finally, it is not necessary to subsample the voxel-based wildfire counts in the implementation (see §7.3 for subsampling techniques), such that the chosen voxel dimensions are consistent with the computational and memory costs of INLA.

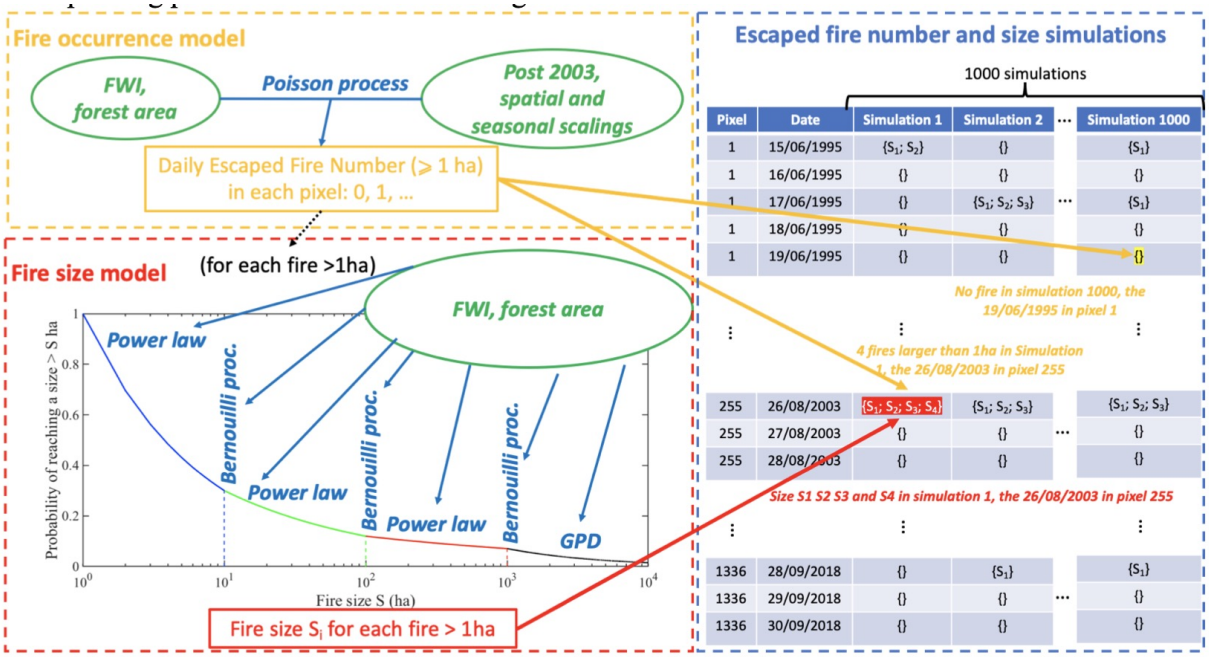


Figure 9.7: Schematic representation of the *Firelihood* model developed in Pimont et al. (2021). The representation of the fire size model in the lower left display suggests a density that is continuous at the severity thresholds, but the model that we have implemented does not impose this continuity constraint.

9.3.1 Structure of *Firelihood* regression models

The general mechanism of the *Firelihood*-model is outlined in the schematic representation in Figure 9.7. A system of regression models is combined to jointly model occurrence numbers, and exceedance probabilities and excess sizes of burnt areas above the four severity thresholds of $1, 10, 100, 1000$ ha. By construction of the dataset used for the model, the exceedance probability of 1 ha is 1, since we consider only escaped fires. Model components are trained with data from 1995–2014 (calibration sample), while the years 2015–2018 (validation sample) are withheld for the evaluation of the predictive performance. The dataset for fire occurrence contains fire counts for approximately 4.44 million pixel-days, the large majority of them being zero, whereas the dataset of fire sizes contains 7193 observations.

We denote by N_{it} the number of wildfire occurrences on day $t \in \{1, \ldots, n\}$ and in the 8km $\times$ 8km pixel $i \in \{1, \ldots, d\}$ with centroid s_i, and by $\mathcal{A}_{i,t} \subset \mathcal{D}$ the space-time cell with volume $|\mathcal{A}_{i,t}| = 64$ (ignoring units). The year corresponding to a given day t is written $a(t) \in \{1995, \ldots, 2018\}$, and the week number within the year as week$(t) \in \{1, \ldots, 52\}$. If at least one fire is observed ($N_{it} > 0$), we denote by $\boldsymbol{Y}_{it} = (Y_{it,1}, \ldots, Y_{it,N_{it}}) \in \mathbb{R}_+^{N_{it}}$ the vector of the corresponding quantitative marks, here restricted to values $[1, \infty)$ since we consider only escaped fires. We write $\boldsymbol{R}_{it}^{(\ell)} = (R_{it,1}^{(\ell)}, \ldots, R_{it,N_{it}}^{(\ell)}) \in \{0,1\}^{N_{it}}$, $\ell = 1, \ldots, 4$, for the corresponding vectors of binary exceedance indicators $R_{it,k}^{(\ell)} = \mathbb{I}(Y_{it,k} > u_\ell)$, $k = 1, \ldots, N_{it}$, with thresholds $u_1 = 1$, $u_2 = 10$, $u_3 = 100$ and $u_4 = 1000$, each in hectare (ha) units. Finally, we write $z_{\mathrm{FA}}(s,t)$ and $z_{\mathrm{FWI}}(s,t)$ for the voxel-based observations of FA and FWI, respectively.

In the next two subsections, we detail the most complex model that we fit, which we consider as the best model based on various model selection and validation steps. To verify the added value of this full model and to avoid overfitting, we also systematically estimate simpler submodels, in which some of the predictor components are removed, and we then conduct model selection and comparison using WAIC (if estimation is done with INLA) or AIC (if estimation is frequentist), as well as visual and predictive criteria. Specifically, we obtain posterior predictive distributions for burnt areas, and in particular for burnt areas aggregated over various spatiotemporal supports. This is done by generating approximate simulations from the posterior model using `R-INLA`, and by appropriately combining the simulated variables to obtain representative samples of the quantity to predict.

Occurrence component

In the most complex model, we use the following general structure of the linear predictor of the occurrence process. A Poisson response distribution is combined with Gaussian random effects in the log-intensity. Within each voxel $\mathcal{A}_{i,t}$, we assume that the Poisson intensity does not vary. This leads to the following formula of the linear predictor:

$$\begin{aligned}\mu_{i,t}^{\text{COX}} &= \log\left(\int_{\mathcal{A}_{i,t}} \lambda(s,t)\mathrm{d}(s,t)\right)\\ &= \log\lambda(s_i,t) + \log|\mathcal{A}_{i,t}|\\ &= \beta_0^{\text{COX}} + \beta_{>2003}\,\mathbb{I}(a(t)>2003) +\\ &\quad + g_{\text{FA}}^{\text{COX}}(z_{\text{FA}}(s_i,t)) + g_{\text{FWI}}^{\text{COX}}(z_{\text{FWI}}(s_i,t)) + g_{\text{season}}^{\text{COX}}(\text{week}(t))\\ &\quad + W_{\text{2D-SPDE}}(s_i) + W_{\text{iid}}(s_i).\end{aligned}$$

The coefficient $\beta_{>2003}$ allows capturing differences in post-2003 wildfire occurrence intensities. The three nonlinear effects $g_{\text{COMP}}^{\text{COX}}$ with $\text{COMP} \in \{\text{FA}, \text{FWI}, \text{season}\}$ have Gaussian process priors of first-order random walk type (RW1, *i.e.*, we model these effects as step functions), with a relatively large number of 30, 18 and 23 segments, respectively, to allow for near-continuous behavior. The RW1 prior model is known to be stable for estimation with INLA (using a sum-to-zero constraint to ensure identifiability), and it has only a single hyperparameter to be estimated. Finally, to capture variation in the occurrence intensity that cannot be explained by the other predictor variables, we have two components indexed by space and endowed with Gaussian process priors. One is of SPDE type with approximate Matérn covariance ($W_{\text{2D-SPDE}}$) to capture residual variation in the occurrence intensity with spatial dependence, and the second one is of i.i.d. type (W_{iid}) to capture spatially unstructured variation, *i.e.*, the prior model corresponds to Gaussian variables that are independent between spatial pixels. All hyperparameters are estimated from data, and we use PC priors.

Size component

For the fire size component, we develop a piecewise model of fire size distribution, that is, a mixture distribution with disjoint supports (*i.e.*, size classes) of the mixture components; recall Figure 9.7. This avoids making restrictive parametric assumptions on the general shape of the

probability density of the distribution in all size classes together. In Pimont et al. (2021), we conduct preliminary analyses of the response of fire size distributions for different FWI levels using mean excess plots (Hall and Wellner, 2020) of the log-transformed sizes of escaped fires, which leads us to fix the segment boundaries to u_ℓ, $\ell = 1, \ldots, 5$, with $u_5 = \infty$. The number of observed exceedances over increasingly high thresholds suggests a slow power-law-like tail decay for all thresholds except for the highest one, for which exceedance numbers seem to decrease much faster as in the power-law setting, consistent with findings in Cui and Perera (2008). We therefore assume that

$$Y_{it,k}/u_\ell \mid (Y_{it,k} \geq u_\ell) \sim \text{Pareto}(\alpha_\ell, 1), \quad \ell = 1, 2, 3,$$

and $\log \alpha_\ell$ is allowed to vary nonlinearly with FA and FWI. We implement estimation of this structure by using the survival model framework of INLA, where only wildfires satisfying $Y_{it,k} \geq u_\ell$ are used to estimate α_ℓ, and we further censor the wildfire sizes above $u_{\ell+1}$. The link function, and the prior structure of the linear predictor, are given as follows:

$$\log \alpha_\ell = g_{\text{FA}}^{\text{SIZE},\ell}(z_{\text{FA}}(s_i, t)) + g_{\text{FWI}}^{\text{SIZE},\ell}(z_{\text{FWI}}(s_i, t)), \quad \ell = 1, 2, 3.$$

For the category with largest fire sizes (exceeding $u_4 = 1000$ ha and containing 33 fires for the training period), we take into account physical considerations and deploy a GPD, which allows for a finite upper endpoint if its shape parameter is negative. Indeed, an upper bound for forest fires — maybe a very large one — must necessarily exist. We fit the GPD to $\log_{10}(Y_{it,k}/1000)$ given that $Y_{it,k} \geq 1000$. This model with the possibility for a negative shape parameter is not available within INLA due to the finite and parameter-dependent upper endpoint of its support. Instead, we estimate the GPD parameters using frequentist maximum likelihood. Owing to the small sample size, we choose a more parsimonious parametrization of covariate influence using only linear coefficients:

$$\begin{aligned} \log_{10}(Y_{it,k}/1000) \mid (Y_{it,k} \geq 1000) &\sim \text{GPD}(\xi_{it}, \sigma_{it}), \\ \log \sigma_{it} &= \sigma_0 + \sigma_{\text{FA}} \times z_{\text{FA}}(s_i, t) + \sigma_{\text{FWI}} \times z_{\text{FWI}}(s_i, t), \\ \log \xi_{it} &= \xi_0 + \xi_{\text{FA}} \times z_{\text{FA}}(s_i, t) + \xi_{\text{FWI}} \times z_{\text{FWI}}(s_i, t). \end{aligned}$$

In principle, we could simply estimate the exceedance probabilities above $u_{\ell+1}$ derived from the power law with parameter α_ℓ for $\ell = 1, 2, 3$, but we cannot expect a good estimation because of the relatively small sample fraction of fire sizes in the segment above $u_{\ell+1}$. Therefore, we use INLA to separately model and estimate these exceedance probabilities, noted $p_{it}^{u_\ell} \in (0, 1)$, above thresholds u_ℓ, $\ell = 2, 3, 4$, based on logistic regressions using the indicator variables of threshold exceedances, $R_{it,k}^{(\ell)}$, as data for the response variable:

$$\log \frac{p_{it}^{u_\ell}}{1 - p_{it}^{u_\ell}} = \beta_0^{\text{EXC},\ell} + g_{\text{FWI}}^{\text{EXC},\ell}(z_{\text{FWI}}(s_i, t)) + g_{\text{FA}}^{\text{EXC},\ell}(z_{\text{FA}}(s_i, t)), \quad \ell = 2, 3, 4,$$

again with Gaussian first-order random walk priors in $g_{\text{FWI}}^{\text{EXC},\ell}$ and $g_{\text{FA}}^{\text{EXC},\ell}$.

9.3.2 Estimated effects and their interpretation

Several of the random effect components estimated for the most complex *Firelihood* model are shown in Figure 9.8 with pointwise 95% credible intervals. In the occurrence component

describing the expected numbers of wildfires, we find a posterior multiplicative effect of FWI that is close to linear for small to moderately large FWI values, but the effect is dampened and essentially flat for higher FWI values. This can be seen as evidence that FWI is not very useful as an index for the occurrence of the most extreme fires in the Mediterranean region, since the proportionality relationship is broken for the most extreme values of FWI, such that FWI would tend to indicate a too high wildfire occurrence risk during high-FWI regimes.

The multiplicative effect of forest area (FA) is also estimated to be strongly nonlinear. A steep increase of relative wildfire occurrence intensity from 0% up to 20% of FA in a pixel is followed by a dampening with a high plateau region between 30% to 60% approximately, before the effect slighlty declines for pixels with very dense forest cover. Dense forest cover often comes along with less human activity, such that less fire ignitions ensue.

The seasonal trend defined at weekly resolution also highlights strong variability with a difference of up to 60% in occurrence numbers at different times of the year. Highest relatively risk arises at the end of August; we conjecture that the FWI does not appropriately capture the fuel moisture conditions over the full wildfire season.

The posterior multiplicative spatial effect highlights that there remains strong spatial variability between pixels that cannot be explained by the other predictors. Spatially aggregated structures arise, where spatial clusters with relatively low or relatively high wildfire occurrence intensities spanning over several adjacent pixels become visible.

Finally, the displays on the right-hand side of Figure 9.8 illustrate estimated exceedance probabilities above a selection of thresholds $(1, 5, 10, 100, 500, 2000$ ha), and how these vary nonlinearly with respect to FA and FWI.

9.3.3 Predictive model comparison

We further conduct a more specific analysis of predicted temporal and spatial patterns, and of how they change if some key components of the linear predictor is removed from the model; that is, if we fit the model without one or several of the random effects in the full model. To realize this analysis, we use 1000 samples of the posterior distributions according to each of the considered models. Figures 9.9 and 9.10 report the results of the comparison of different models and observations for predictions of aggregated occurrence numbers and burnt areas, respectively. The full *Firelihood* model shows generally good correspondence with observations in all displays (yearly trends, seasonal trends, spatial trends) except for the very extreme observation year of 2003, while removing key components from the model (*i.e.*, using only nonlinear FWI, no seasonal effect, no spatial effect, respectively) leads to severe biases in predictions. Predictions of burnt areas are generally more noisy because of the heavy tails in the distribution of burnt areas. In Pimont et al. (2021), we also discuss some diagnostics similar to the above ones but restricted to the year 2003, which stands out for its exceptionally high fire activity. The *Firelihood* model could not fully capture the behavior for this extreme year. The estimation of the numbers of escaped fires was consistent with observations for 2003, but the model underrepresents larger fires and burnt areas. This finding could again be due to limitations of the FWI in rating fire danger during extreme heatwaves, and it highlights the importance of long time series including catastrophic years for accurate fire activity modeling.

Finally, we define several binary representations of predictive behavior, for instance the occurrence of at least one fire in a pixel-day, or the exceedance of a severity threshold given that

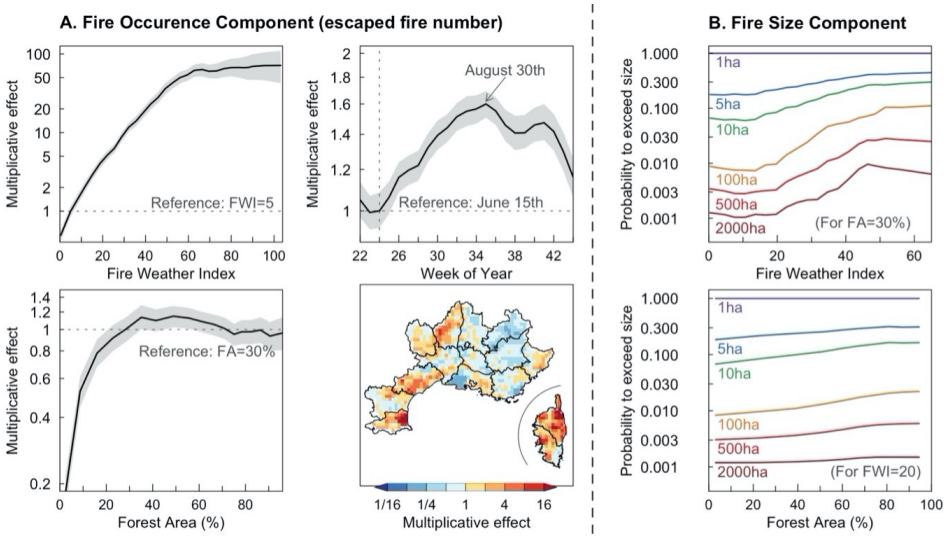


Figure 9.8: Estimated effects of the most complex *Firelihood* model. In the fire occurrence component, the estimated nonlinear curves of FWI, season and forest area are shown with pointwise 95% credible intervals. The effects included in the fire occurrence component are expressed as posterior means of multiplicative factors with respect to some reference value. For the fire size component, we show how posterior exceedance probabilities over a range of severity thresholds vary with FWI and FA (with fixed FA and FWI value, respectively).

a fire occurs. This approach allows computing Area-Under-the-Curve (AUC) scores, which are shown in Figure 9.11. Occurrence of wildfires is generally very well predicted by our full model with the AUC above 0.8 in the validation dataset. By contrast, the prediction of exceedances of severity thresholds of burnt areas is generally more difficult and leads to lower AUC scores between 0.55 and 0.75, depending on the threshold.

From the analyses detailed in Pimont et al. (2021), we conclude that the overall number of escaped fires for the whole study region can be coarsely reproduced at daily scale, and it is accurately predicted on a weekly or longer basis. We also predict the overall weekly number of larger fires (10–100 ha), where the accuracy decays with size since the model uncertainty increases with event rareness. Likewise, more localized predictions of fire numbers or burnt areas, for instance at the level of administrative areas (départements), require a longer temporal aggregation period to maintain model accuracy. In summary, our study sheds new light on the stochastic processes underlying fire hazard, and it our model is a promising tool for predicting and projecting future fire hazard in the context of climate change, see in particular the results presented in Fargeon et al. (2018).

9.4 Outlook towards modeling extensions

The following subsections summarize several possible extensions whose relevancy is grounded on the analysis of still open questions resulting from the work in Gabriel et al. (2017); Fargeon et al. (2018); Opitz et al. (2020b); Pimont et al. (2021). Some of these extensions are already tackled in work in progress.

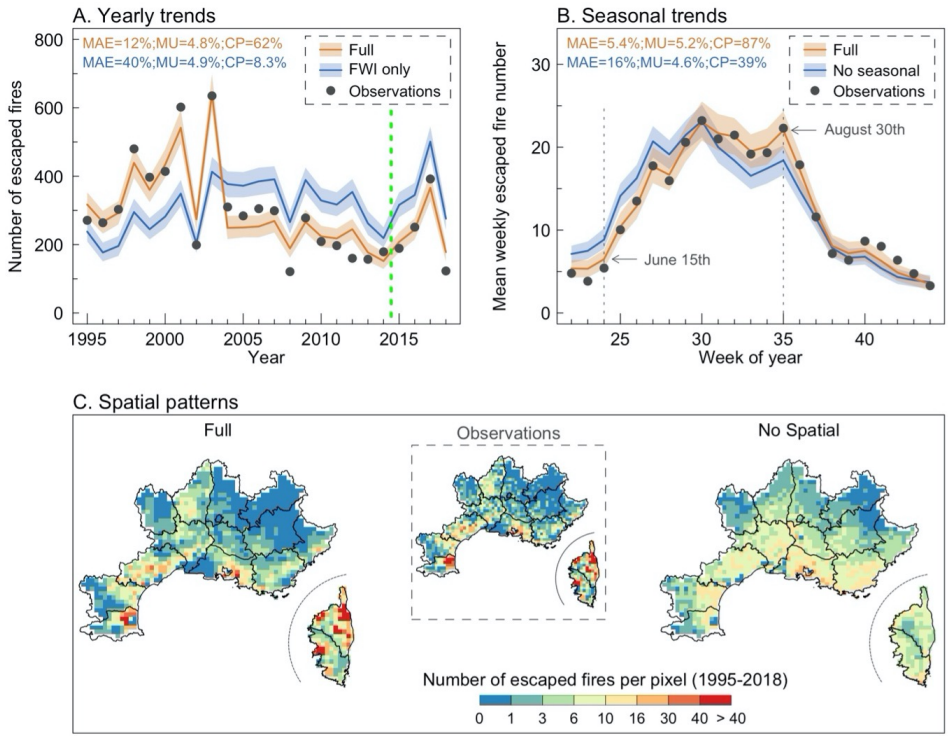


Figure 9.9: *Firelihood*-based predictions of occurrence numbers of escaped fires based on posterior simulations aggregated over time and space. Each display compares predictions of the full model with those from a simpler model where some key random effect components are removed, *i.e.*, , of a model only with FWI effect (upper left display), without seasonal random effect (upper right display), and without spatial effect (lower display).

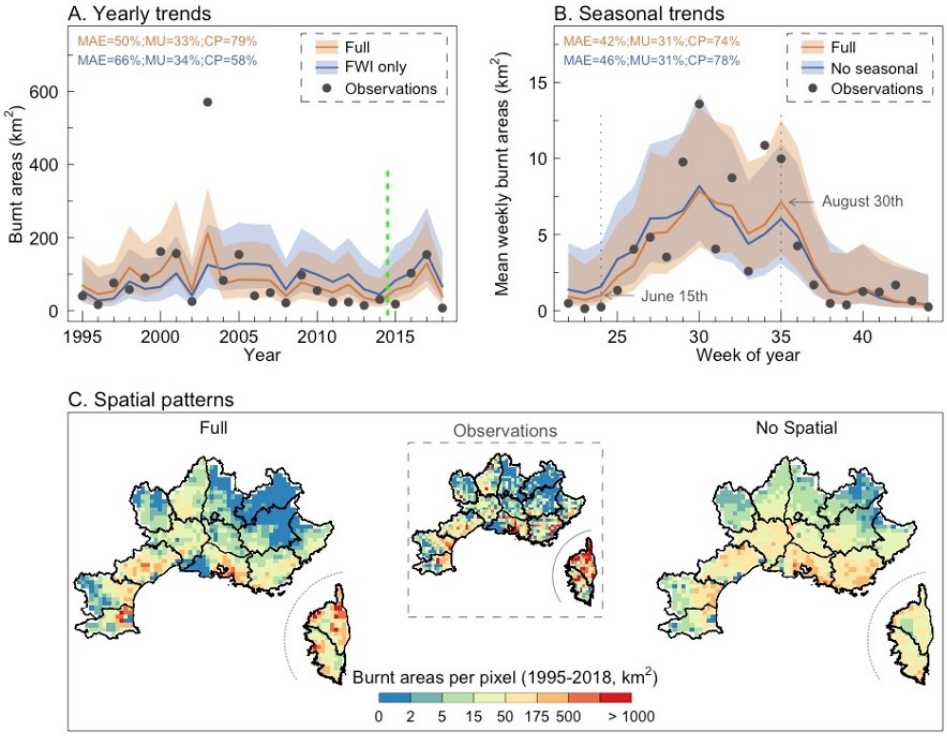


Figure 9.10: *Firelihood*-based predictions of burnt areas aggregated over time and space based on aggregated posterior simulations of the occurrence and the size components. The configuration of the displays is the same as in Figure 9.9.

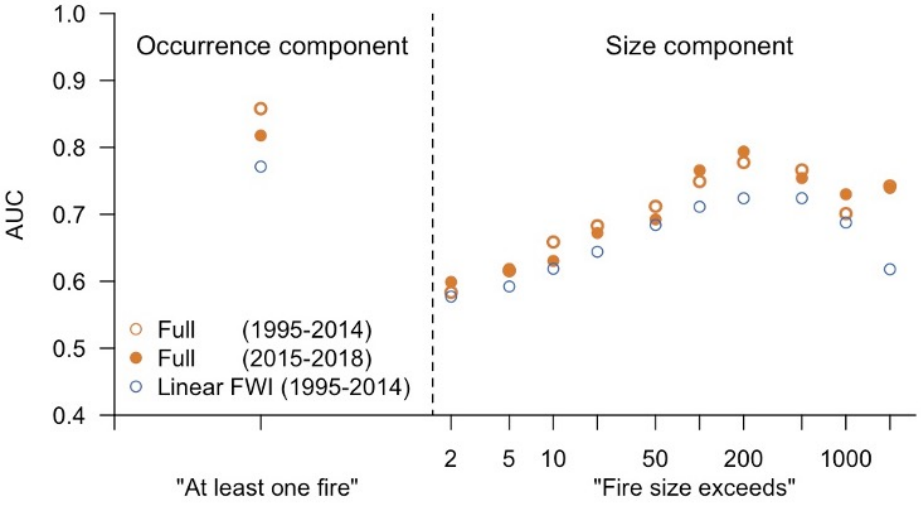


Figure 9.11: AUC scores calculated for the training and validation period for the *Firelihood* model. We compare the Full model (with nonlinear FWI effect) with a simpler model that has linear FWI effect in the linear predictor of the occurrence component.

9.4.1 Small-scale inhibition among wildfire occurrences

While the above exposition has focused on models such as LGCPs that are able to capture spatiotemporal clustering in wildfire occurrence, there may also be notable inhibitive effects among wildfire ignitions at relatively small spatial and temporal scales. In particular, it is natural to assume that no wildfires can take place over the same surface immediately after an outbreak, and ecological knowledge about regeneration cycles of vegetation suggests that this inhibitive pattern between neighboring events can last for several years after a fire. It would be worthwhile to better quantify this claim and provide empirical evidence from data. Hybrid point process models incorporating a mix of aggregative and repulsive structures at different scales are an active area of research (Baddeley et al., 2013) and could be used to incorporate this mechanism into generative models.

In Gabriel et al. (2017), we propose a descriptive measure to quantify such repulsive effects through the *annual normalized empirical intensity ratio index* I_{year}. Values of this novel index below 1 indicate temporal inhibition between fire events at close distances, while values above 1 suggest temporal clustering. It is defined as the ratio of two aggregated counts: in the numerator, the number of events observed in the follow-up years after a wildfire in a reference period (and spatially close to this wildfire), where the reference period consists of the reference year and potentially also some of the preceding years; in the denominator, for the same spatial buffer, the number of wildfire events observed outside the chosen reference and follow-up periods. In addition, we normalize the counts in the numerator and denominator with the global intensities (in the full study region) for the respective periods. Figure 9.12 shows this annual index for various configurations of lengths of reference and follow-up periods, and of wildfire size categories, in a specific administrative area (Bouches-du-Rhône). Especially for larger wildfires, the inhibition effect becomes clearly visible, despite the positional uncertainty about the exact location of wildfires; recall that the observed burnt areas are typically relatively small in comparison to the 2 km grid resolution of the Prométhée database. It would be interesting to further refine this index and its estimation to conduct proper statistical inferences and to apply it at a larger scale in France and in other wildfire-prone areas. A recurrent difficulty is that wildfire ignition locations are often not recorded exactly but only up to a certain grid resolution, *e.g.*, at 2 km resolution in the Prométhée database, such that there is relatively high uncertainty about the exact spatial location of the wildfire scar, especially if most wildfire sizes are much smaller than a grid cell.

9.4.2 Fully Bayesian inference for joint occurrence-size modeling

With respect to the *Firelihood* model proposed in Pimont et al. (2021) for joint modeling of occurrences and sizes in Mediterranean France, we work on several extensions to provide improvement in predictions especially for extreme wildfires, in uncertainty assessment and in interpretation.

First, due to the complexity of *Firelihood*, the occurrence and size model components are estimated separately, such that transfer of information between components, and inference on stochastic interactions between the two components, are not feasible. Moreover, the GPD with negative tail index fitted to the logarithms of the largest burnt areas is estimated in a frequentist setting, since the INLA framework is not appropriate for that purpose. An exploration of alternative formulations of size models that are fully amenable to inference with INLA could

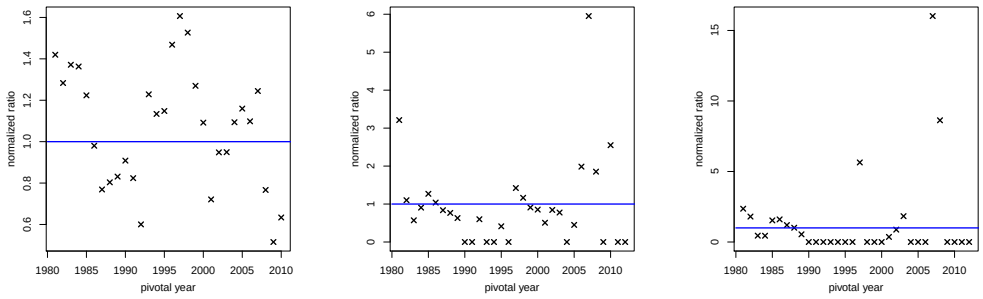


Figure 9.12: Annual normalized empirical intensity ratio index. Left: Results for wildfiires with at least 10 ha of burnt surface and a 5 year follow-up period. Middle: Results for reference wildfires with at least 50 ha of burnt surface and all other events with 20 ha minimum surface, and 3 year follow-up period. Right: Results for reference wildfires with at least 100 ha of burnt surface and other events with 25 ha minimum surface, and a 3 year follow-up period.

lead to conceptually more elegant modeling solution.

Second, temporal stochastic structures in the current *Firelihood* version are relatively simple since they are restricted to spatiotemporal variability of covariates and to a categorical effect with two levels, the latter used to capture a decrease of wildfire numbers after the pivotal year of 2003 marked by exceptionally high wildfire activity (which may have triggered improved wildfire prevention). As a consequence, simulated posterior predictive distributions for some space-time aggregations of wildfire activity do not appropriately capture some very extreme events, specifically the year 2003. A promising lead is to propose more flexible spatiotemporal structures. Moreover, to allow for joint estimation and fully Bayesian inference of all model components in models with more complex temporal structures, suitable data subsampling schemes, similar to the approach implemented in Opitz et al. (2020b), should be developed, and their loss of information must be analyzed carefully. We should take care to achieve subsample sizes that allow running models on standard desktop computers, in contrast to the often highly computer-intensive approaches in the literature, especially plenty of fast access memory; see the discussion in 7.3.

Third, it is important to further conceptualize the combination of spatiotemporal marked point processes with extreme-value theory in the modeling mechanisms combining moderate and extreme wildfires. Extreme wildfires can be defined as exceedances above a high but fixed severity threshold for burnt areas. The point pattern of extreme fires can be viewed as a *thinning* of the full point pattern, and we can conduct threshold selection to identify a suitable threshold for applying the theoretically justified GPD to the excesses of burnt area above the threshold.

Fourth, in order to estimate spatially indexed random effects that affect several model components simultaneously, the idea of *sharing* can be developed: the random effect is estimated within one response variable (*e.g.*, wildfire counts), and we also include it with a scaling coefficient (which may be 0 in the absence of a shared effect) in some other, potentially related response variables (*e.g.*, wildfire size exceedances). The significance of the sharing coefficient can be assessed. Sharing could be used as tool to increase model parsimony and it may offer new scientific insight; see also the preprint (*e.g.*, Opitz et al., 2021). Sharing may be especially relevant to "borrow strength" for estimating complex structures in the linear predictors related

to extreme events, where only a relatively small number of observations is available. By using sharing, we can let the data decide if effects estimated at moderate quantile levels of wildfire sizes carry over to extreme levels. Currently, the size component of *Firelihood* does not have any spatially indexed random effects.

Fifth, we can improve on the nonlinear FWI effect that is global in the current *Firelihood* model. Instead, we may use a more sophisticated model featuring a month-specific nonlinear FWI effect. Indeed, Pimont et al. (2021) identify some shortcomings of the FWI in capturing seasonal variation of wildfire occurrences in France, which could be remediated by estimating a seasonally varying response of wildfire activity to FWI.

A final point of caution concerning spatiotemporal point process models with heavy-tailed marks is model validation and comparison based on posterior predictive distributions. Due to heavy tails and to high uncertainty in the spatiotemporal prediction of individual wildfires, customary validation scores such as mean-squared/absolute errors (MSE, MAE) or CRPS are of only limited informative value. The distribution of the contributions of individual observations to such scores tends to be heavy-tailed, such that adding up or averaging such contributions (as is done in MSE, or when considering average CRPS) will not lead to reliable scores. Methodological developments are still necessary to define appropriate probabilistic prediction scores and techniques, and they could be inspired by the theoretical analyses of Taillardat et al. (2019). In particular, instead of comparing different forecast models based on means of scores of individual observations, one may compare the full empirical distribution of such scores and then use variants of stochastic dominance to rank forecast models.

9.4.3 New indices of wildfire danger

So far, we have considered the widely used FWI as a standard measure of meteorological wildfire danger. However, the shortcomings pointed out for this index in our work suggest that we should seek a more appropriate formula to define meteorogical fire danger in the Mediterranean region based on raw weather variables such as temperature, precipitation, wind speed and humidity. We intend to use statistical learning techniques such as Random Forests or boosted regression to learn about the functional relationship of fire activity components with weather variables and their interactions. This exploratory analysis can help define relatively parsimonious parametric or nonparametric models of fire danger indices, which may include expert knowledge on biophysical mechanisms. Extreme-value theory will be of great use for two reasons: first, we aim to put focus on weather conditions leading to extreme wildfires; second, highly fire-prone vegetation conditions typically correspond to lower-tail extremes of so called live fuel moisture. We further aim to include LULC information in this approach. This work is expected to result in new definitions of wildfire danger indices, and we will use regression models to link them to wildfire activity components while taking into account remaining prediction uncertainties.

9.4.4 Towards operational forecasts of wildfire activity

Currently available spatiotemporal models provide valuable predictions and insights into risk factors and their contribution to components of fire activity, but the use of predictive models for operational purposes requires an even finer spatial and temporal resolution of predictions.

Moreover, specific event types such as multiple wildfire ignitions at several close locations at the same time, and fires that "escaped" at an initial stage by exceeding certain severity thresholds, are particularly problematic for firefighters and merit the development of modeling approaches with even stronger focus on these situations. Multiple fire occurrences within a relatively small space-time buffer are very rare but highly problematic, and by their extreme nature they could be modeled through extreme-value techniques for discrete variables.

Such approaches are part of ongoing work and project proposals, and they will be developed with colleagues from INRAE's URFM lab in Avignon and in collaboration with Météo France, among others.

9.4.5 Decomposition of risk and uncertainty components under climate change

Due to the high complexity and spatiotemporal fragmentation of the processes involved in wildfire activity, and due to the resulting high sophistication of models required to provide realistic predictions, there remain large uncertainties in model outputs, even if we can use relatively large wildfire datasets to calibrate models. Another source of uncertainty related to future projections of wildfire activity is given by different scenarios for climate change and for changes in LULC. With respect to climate change, scenarios such as the Representative Concentration Pathways (RCP) can be injected as new predictors into probabilistic regression models to obtain probabilistic predictions (Fargeon et al., 2018). However, the question of how to apply an appropriate weighting of the results according to different scenarios of severity of climate change (and also different climate models) is still an open question. Moreover, appropriate decomposition of the variability and uncertainty of model outcomes according to different sources of uncertainty is another problem that requires deeper analysis; the following components of uncertainty could be considered: natural uncertainty of the biophysical processes; intrinsic uncertainty of climate change scenarios, and uncertainty in the attribution of climate change effects; uncertainty of model choice; parameter estimation uncertainty.

The development of methods to better identify, separate and quantify these sources of uncertainty from spatiotemporal posterior simulations is therefore another area of research to be investigated.

Chapter 10

Student supervision, teaching and other research output

10.1 Other research output

10.1.1 Text mining

The work during my 10-month post-doc project at the computer science institute LIRMM of Montpellier University has led to the publication of a number of relevant results, and it has laid the foundation for a wide range of follow-up projects (in which I am not involved any more today due to lack of time and focus on core topics of INRAE). During my post-doc, I have used highly noisy and unstructured textual data extracted from online health forums, for which I have developed statistical methods and computer codes to extract structured medical knowledge about the quality of life of breast cancer patients. This involved attribution and statistical analysis of text content to items of quality-of-life questionnaires in Opitz et al. (2014), and non-supervised structure analysis of topics discussed in these forums (Nzali et al., 2017), among other approaches (Tapi Nzali et al., 2015; Nzali et al., 2017; Tapi Nzali et al., 2019).

I have implemented related concepts and models in ongoing joint work with my colleague André Kretzschmar at BioSP, with the goal of automatically analyzing transcribed interviews with beekeepers, in particular with respect to how they cope (from the epidemiological and economical stance) with a parasite (*varroa destructor*) that is strongly harming and decimating bee colonies.

10.1.2 EVT-based analysis of windstorm-based insurance claims

With partners at Allianz France and Lyon's ISFA institute for actuarial sciences, we have used EVT to study the design of highly nonlinear "transfer functions" that explain how the most extreme wind speeds observed during windstorms translate into property damage, and which subregions of the French territory show relatively homogeneous exposure patterns by using ad-hoc designed spatial clustering techniques (Mornet et al., 2015, 2017). Interestingly, the homogeneous wind risk zones that we find through non-supervised clustering reflect orographic structures of the French territory since they grossly correspond to the most important river catchments.

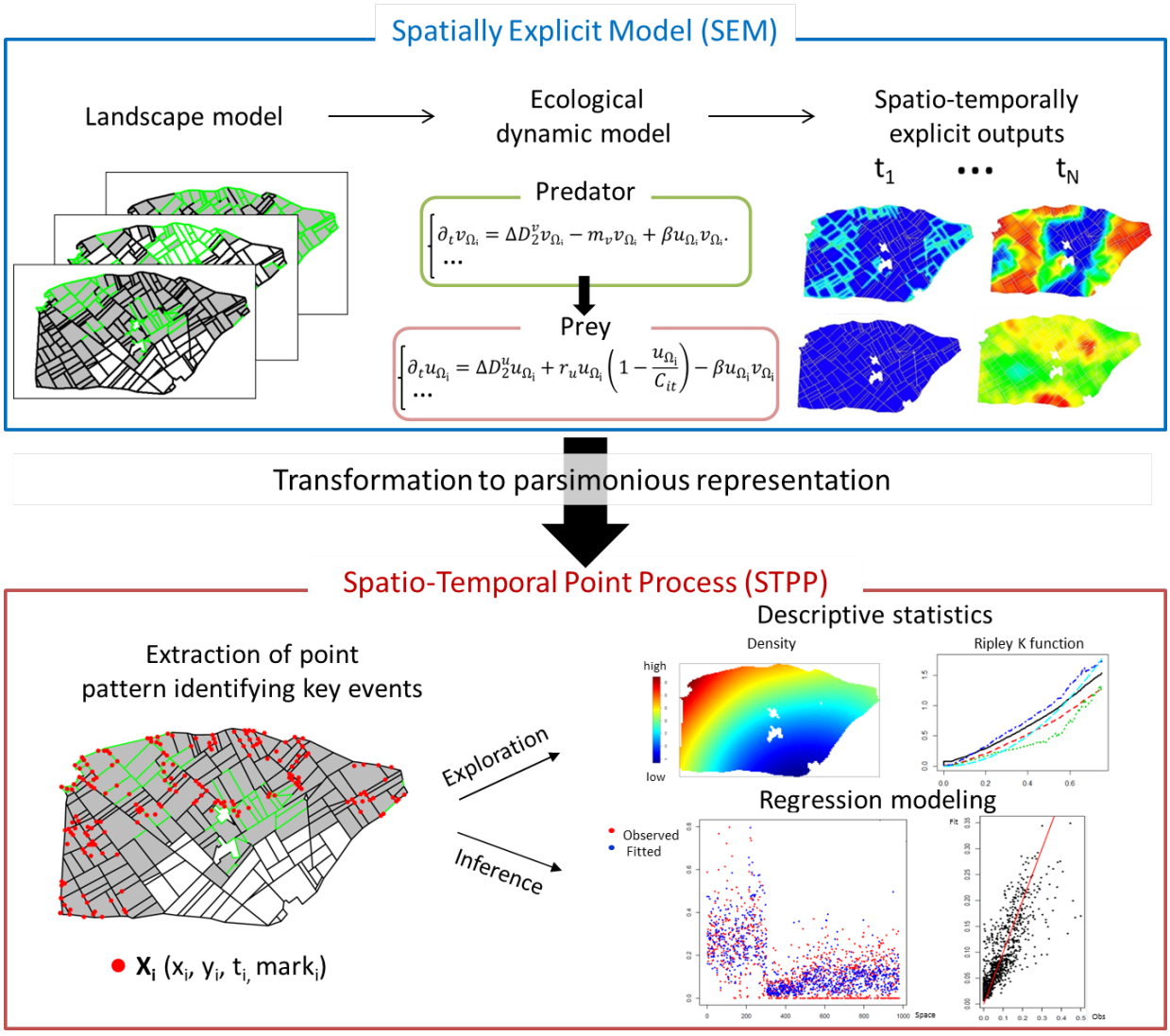


Figure 10.1: Schematic overview of meta-modeling of agroecological events in pest-predator population dynamics through spatiotemporal point processes.

10.1.3 Stochastic modeling of agricultural landscapes and their impact on agroecological processes

This field of research has become very important for me over the past three years thanks to collaborations at BioSP and the PhD project of Patrizia Zamberletti (2018–2021). It contains two major axes, and a schematic summary of important steps is given in Figure 10.1. First, we develop new generative stochastic models for statistical inference and simulation with agricultural landscapes. They are based on a vector-based representation of the spatial landscape support using polygons (2D) and line segments (1D), and the allocation of landscape elements with categories, such as "crop vs. semi-natural habitat" or "hedge vs. no-hedge" is achieved through Markov random field models; see the details in our publication Zamberletti et al. (2021a).

Then, we simulate in cascade a model of these 2D-1D landscapes and a spatially explicit

predator-prey dynamical system based on the numerical solution of systems of partial differential equations (PDEs), and we implement the application of pesticide treatments when a certain threshold value of the pest density is exceeded locally. Based on a large variety of parameter configurations used to control the spatiotemporal dynamics in the landscape-pest-predator system, we study the impacts of landscape structure (composition and configuration) and of species traits on biological control outcomes, such as the number of pesticide treatments and the local peak value of pest density in case of treatments.

Specifically, we develop marked point process models for the joint analysis of locations and magnitudes of pest density peaks after an exceedance of the pest density threshold has occurred, *i.e.*, just before a pesticide treament is applied. Marked PPs are used as meta-models in this approach, and they can be viewed as an upscaling from prey-predator simulations at high spatiotemporal resolution to intermediate scales where agroecological key events can be highlighted and represented parsimoniously through the marked PP. We use generalized additive regression models for the point processes to estimate and infer the drivers of spatiotemporal dynamics in the occurrence and magnitude of key events. For the discretization of the study area (recall the regression equation (7.2)), we here design a problem-specific discretization. We capitalize on the implementation of triangulation mesh generation in the `R-INLA` package of `R` (available for the SPDE models) to obtain mapping units that make allowance for the special role of line segments (*e.g.*, allocated with hedges) in agricultural landscapes; see Figure 10.2. In particular, we distinguish between different types of mapping units: units in a patch center, units connecting exactly two patches, and units connecting more than two patches (blue), where these three categories can be associated with different ecological dynamics. Details of this work are available in the two preprints Zamberletti et al. (2021c,b).

10.1.4 Modeling spatiotemporal trends in soil properties using INLA-SPDE

With colleagues from INRAE's InfoSol unit at INRAE Orléans, who maintain and explore the data and soil samples in the *Conservatoire Européen d'Échantillons de Sol* (*i.e.*, the *European Conservatory for Soil Samples*), we have tackled spatiotemporal soil modeling with focus on space-varying temporal trends. This project has its beginnings in INRAE's RESSTE network for spatiotemporal statistics, and some significant progress was achieved during the PhD project of Bifeng Hu supervised by colleagues at InfoSol. Our collaborative work is still in progress, but important intermediate results have already been collected in Bifeng's PhD manuscript. Our motivation is to develop a novel modeling approach to space-time mapping of soil variables, which allows drawing statistical inferences on spatial-temporal soil dynamics. Soil conditions can vary significantly over space and may show strong temporal dynamics, often triggered or enhanced by anthropogenic activity, *e.g.*, application of fertilizers and pesticides, atmospheric deposition of nitrogen and sulfur, and other environmental factors driven by climate change. Soil contributes to a wide range of ecosystem services since soil fulfills many important functions including biomass production, storage and filtration of water, storage and recycling of nutrients, and habitat for biological activity and carbon storage.

Datasets of soil variables can be massive and result from complex, spatially and temporally heterogeneous sampling procedures. We complete them with covariate data related to soil types,

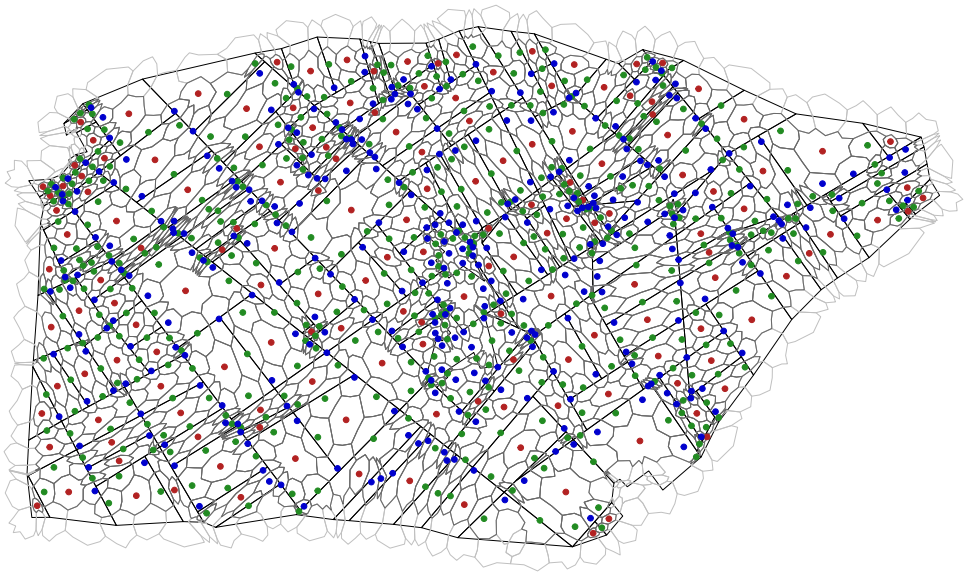


Figure 10.2: Spatial discretisation of the regression models for point process meta-models of agroecological events in pest-predator population dynamics. Complete mesh discretisation (light grey), mesh cells used in the analysis (dark grey), landscape patches (black). Cell centroids of different colour refer to different cell types: cell in patch center (red), cell connecting exactly two patches (green), cell connecting more than two patches (blue).

weather and climate. The sampling process is of opportunistic nature and highly irregular in space and time in the data that we study, and data may be very noisy due to measurement errors. We use a large database containing hundreds of thousands of soil samples of different soil properties for the mainland French territory collected since the 1990s. The data correspond to the collection of the results of soil analyses requested by farmers for their fields. Although we could not detect any strong preferential sampling effects in these data, they can be qualified as opportunistic since there is no strict protocol for setting sampling locations and times, and observations arise only if users (*i.e.*, farmers) make a request.

In the following, I shortly outline major issues that we can solve through the INLA-SPDE approach thanks to the specific structure of the model that we estimate. Statistical inference on temporal trends and their variation through space is paramount for explanatory and predictive modeling of soil variables but has not yet been much explored. In this work, we advocate a model constructed in the INLA-SPDE framework, similar to space-varying regression (*e.g.*, §8.3), to tackle a number of statistical issues that are difficult to solve with state-of-the-art kriging or machine learning approaches in soil science:

1. Data are not available for a factorial design of spatial stations and time steps, such that we may have many missing values in the time series for a fixed location. This renders location-wise estimation (*e.g.*, regression) difficult. Spatial pooling of data is possible, but leads to the question of the choice and weighting of areas to pool, and may lead to spatial oversmoothing of results.

2. Standard regression and machine learning techniques may be used to explain the target variable's variation through available covariates, such as time or geological and climatological features, but the usual assumption of independent residuals leads to wrong inferences. In particular, uncertainty in regression coefficients would be underestimated because spatial dependence implies a smaller effective sample size of data.

3. The estimation of space-varying regression coefficients (*e.g.*, for the time covariate or the intercept) in continuous space is challenging, especially if we wish to conduct a data-driven choice of parameters governing the spatial smoothness of the estimated trend surface.

4. A precise assessment of estimation and prediction uncertainty is challenging in regression models when estimation is performed in several steps (*e.g.*, in the following setting: regression in step 1, followed by spatiotemporal dependence modeling of regression residuals in step 2 for kriging them).

5. Space-time covariance matrices become increasingly intractable with more data points. Even with only several thousands of data points, one may have to resort to dependence structure representations defined locally in space and time (*e.g.*, pairwise likelihood, kriging neighborhoods).

With INLA-SPDE, joint estimation of all model components, including smoothness parameters (Issues 2, 3, 4), becomes feasible for any sampling design (Issue 1) and with many data points (Issue 5), such that estimation uncertainty is propagated through all model components jointly. Through the hierarchical formulation of INLA-based models, we can accommodate measurement errors at the observation layer. The SPDE approach allows using sparse spatiotemporal

precision matrices, tractable in very high dimension (Issue 5) and useful for defining flexible prior structures for space-varying components (Issue 3).

We assume that a soil property S at space-time position (x, t) can be predicted from observations and other covariate information, that is, we use the so-called *scorpan* model

$$S(x,t) = f(\{s_i,\, i = 1, \ldots, n\}, c, o, r, p, a, n \ldots)$$

where $S(x,t)$ is the soil class or attribute to predict, f a function, $\{s_i,\, i = 1, \ldots, n\}$ are the measured soil properties, c refers to the climate, o to organisms, r to relief, p to soil parent material, a to age, and $n = x$ to the spatial position. Among other predictor variables to be included, the time t is of prime importance in our model but is usually not explicitly included in the scorpan-model. Given observations $s_i = s(x_i, t_i)$, our most complex Bayesian model possesses the following prior structure, here written as a regression equation:

$$s(x_i,t_i) = \beta_0 + \sum_{j \in \text{scorpan}} \beta_j^{\text{sc}} z_j^{\text{sc}}(x_i) + f(\text{month}(t_i)) + W_0(x_i) + \tilde{t}_i\, W_1(x_i) + \sum_{\ell=1}^{L} B_\ell(t_i) W_\ell^{\text{res}}(x_i) + \varepsilon(x_i,t_i), \tag{10.1}$$

with the following elements:

- fixed effect coefficients β_j^{sc} with independent Gaussian priors for scorpan-related spatial covariates $z_j^{\text{sc}}(x_i)$;
- normalized time $\tilde{t} = \frac{t - t_{\min}}{t_{\max} - t_{\min}} \in [0,1]$, such that the left bound 0 corresponds to the beginning of the study period $t_{\min}$, and the right bound 1 corresponds to the end of the study period $t_{\max}$;
- a seasonal effect $f(\text{month}(t))$ with first-order random walk prior, defined at monthly resolution;
- a spatial Gaussian process W_0 to define a space-varying intercept;
- a spatial Gaussian process W_1 to define a space-varying slope;
- spatial Gaussian processes $W_\ell^{\text{res}}(s)$, $\ell = 1, \ldots, K$ to capture residual, space-varying nonlinear effects using B-spline functions $B_\ell(t)$ defined over the time dimension;
- mutual independence between all of the above spatial Gaussian processes;
- i.i.d. Gaussian variables $\varepsilon(s,t) \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ to capture measurement errors and variability in the response values not explained by the other effects.

To improve identifiability of the above model components, we can impose sum-to-zero contraints on the Gaussian variables pertaining to certain random effects, such as the monthly random walk, and the spatial Gaussian processes W_0 and W_1. Through the INLA-SPDE approach, we can obtain accurate estimations and uncertainty measures for all of the above components, and for spatiotemporal predictions of $S(x,t)$. Using observations of soil variables such as pH or Phosphorus, we fit this model and compare it to simpler variants, embedded as submodels in the general model by removing some of the additive components.

10.2 Supervision of students

10.2.1 Master students

After my PhD, I have supervised several master students at the M1 and M2 level (note: M2 corresponds to the final year in 5-year bachelor-master studies) working on various subjects:

- Yassine Motie (M2, 2014), topic *Acquisition of Patient Vocabulary From Health Forums*, Master thesis conducted at LIRMM, Montpellier University.
- Khadija Sabir (M2, 2018), topic *Simulation of agricultural landscapes to study population dynamics of pest and auxiliary*, Master thesis conducted at BioSP, defended at SupAgro Montpellier.
- Thomas Minotto (M1, 2018), topic *Calculating tree detectability in forest canopies based on pairwise tree interaction*, master project conducted at BioSP.
- Oussama Ennouri (M2, 2020) topic *Geostatistical kriging of conditional extremes*, one of the final year projects for his studies at MinesParisTech.

10.2.2 PhD projects and post-docs

At BioSP, I am currently co-supervising (with Julien Papaïx, Édith Gabriel) the PhD project (2018–2021) of Patrizia Zamberletti on *Spatiotemporal modeling and simulation of agricultural landscapes, with an application to pest regulation*, see §10.1.3. This PhD project is funded for half by INRAE's (former) MIA and SPE divisions, and for half by the Provence-Alpes Côte d'Azur region.

In the framework of the KAUST Competitive Research Grant (CRG) project *Statistical Estimation and Detection of Extreme Hot Spots, with Environmental and Ecological Applications*, a project during 2018–2021 coordinated by Raphaël Huser, I am co-supervising the PhD projects of three students, Rishikesh Yadav, Peng Zhong, Zhongwei Zhang, at KAUST (King Abdullah University of Science and Technology, Saudi Arabia). With funding from this project, I have also co-supervised a post-doc project of Emma Simpson at Lancaster University, on the topic of flexible high-dimensional spatiotemporal modeling of conditional extremes using the INLA-SPDE framework, see Simpson et al. (2020) for a submitted preprint of our work.

With colleagues at University of Montpellier (Julie Carreau, Gwladys Toulemonde), I have been supervising the 18-month post-doc project of Fátima Palacios Rodriguez on resampling methods for space-time precipitation extremes, with results presented in §5.2. We had obtained funding for this project from Montpellier Université d'Excellence (MUSE), the Labex NUMEV, and Inria.

Moreover, I am closely involved (or have been involved) in a number of PhD projects of young collaborators without being in the position of supervisor, specifically in the PhD projects of Hélène Fargeon (URFM, INRAE Avignon; wildfire modeling), Bifeng Hu (Infosol, INRAE Orléans; soil modeling), Oksana Grente (CEFE-CNRS, OFB; wolf attacks on sheep herds) and Jonathan Koh (Chair of Statistics, EPFL, Lausanne; wildfire modeling).

A recent PhD project funded by the 3IA Côte d'Azur structure (dedicated to fostering artificial intelligence and smart territories) will start in spring of 2021 in co-supervision with

Elena de Bernardino at Nice University. The topic is at the intersection of spatiotemporal statistics, EVT, stochastic geometry and statistical learning, and it will focus on theoretical and statistical tools for excursion sets above high thresholds in stochastic processes. Based on datasets over dense and regular spatiotemporal grids obtained from physical models for weather reanalysis and projection, we will study the distribution of geometric properties of excursion sets over high thresholds, such as the area, perimeter and Euler characteristic. This approach will allow for a relatively sparse representation of extremal dependence in big datasets, and we intend to seek connections to widely used geometric landscape descriptors in landscape ecology; see Chapter 11 for more background.

10.3 My role in research groups and project consortia

INRAE's RESSTE network[1] for spatiotemporal data analysis, launched in 2014 by Denis Allard (BioSP), has become an important forum with more than 50 affiliated members from France and abroad for bringing together researchers, for animating workshops and for exchanging on cutting-edge research geared towards statistical modeling of the interplay of space and time in environmental and climatic processes. I am steering committee member of this network, and I have taken a leading role in organizing and animating several international workshops and tutorials (Avignon, Montpellier, Lancaster, each with 30-50 participants; one of them co-organized with the EcoStat GdR network dealing with Ecological Statistics) on space-time modeling with classical geostatistical methods (see the review paper RESSTE Network (2017) resulting from this work) and with INLA-SPDE techniques.

A major research grant, in which I am involved as co-investigator, is the KAUST Competitive Research Grant entitled *Statistical Estimation and Detection of Extreme Hot Spots, with Environmental and Ecological Applications*, coordinated by Raphaël Huser, from 2018 to 2021. I am one of five international co-investigators, and I co-supervise several PhD projects at KAUST, and a post-doc project of Emma Simpson at Lancaster University, with several visits of these students at BioSP. Moreover, I have significantly contributed to two LEFE projects ("Les enveloppes fluides et l'environnement", funded by INSU for fostering new project-centered collaborations) called "Cerise" (2016–2018) and "Fraise" (2019-2021). Both projects put focus on the topic of space-time modeling of extremes, and I am coordinator of several of the work packages.

Since 2017, I am elected member (reelected in 2020) of the Environment and Statistics group of the French Statistical Society. It fosters discussions on the use of statistics to environmental and ecological problems, and we organize workshops and other activities focusing on the use of statistical methodology in this context. I am webmaster of this group, and I have been principal organizer of an online workshop in November 2020 on environmental risk analysis with more than 100 registered participants.

[1] https://informatique-mia.inrae.fr/reseau-resste/

10.4 Teaching

This section shortly summarizes my activity of teaching and training at undergraduate and graduate level after my PhD defense.

10.4.1 Undergraduate level

During the first semester of the academic year 2016–2017, I have taught an introductory course of descriptive statistics at IUT Avignon.

During a one-week course on Extreme-Value Analysis in 2019 and in 2021, organized for the student exchanges of the European ATHENS network at MinesParisTech, I have given a one-day course with practicals on multivariate extreme-value theory.

I am the coordinator of a Master Course (M2 level, 24 hours) on Spatial Statistics, taught intermittently since 2018–2019 at Aix-Marseille University as part of a Master of Data Science programme (with Denis Allard, Florent Bonneu).

During the academic year 2020–2021, I give a course "Introduction à l'analyse des valeurs extrêmes" (translation: "Introduction to extreme-value analysis"), of 16 hours mixing theory and practicals for implementation with `R`, at École Centrale Marseille.

10.4.2 Graduate level

In the framework of the life-long training of INRA(E) staff, which is also open to external participants, I have teached a four-day course "Processus ponctuels spatio-temporels" (translation: "Space-time point processes") in March 2018 jointly with my colleague Édith Gabriel (BioSP). The workshop took place at the Avignon site of INRAE.

With the colleagues from the steering committee of INRAE's RESSTE network focusing on spatiotemporal statistics, we have elaborated a one-day workshop to teach the full workflow of analyzing and predicting space-time data using geostatistical tools available in the `R` language. We have held this workshop twice: at the Spatial Statistics conference in Lancaster, UK (2017), and at the METMA workshop in Montpellier (2018).

In November 2018, I have given an introduction to the INLA-SPDE method during 6 hours of tutorials and practicals at the SPDE workshop in Avignon, organized with the help of the RESSTE network and the GdR EcoStat, a French research federation for statistical ecology.

In May 2019, I have given a one-day course with practicals to present statistical regression modeling to PhD students of the École Doctorale (PhD school) of Avignon, within a one-week introduction to statistics for PhD students without a strong statistical background.

Chapter 11

Outlook to future research

The research projects that I intend to develop will focus on new methodology for data analysis and statistical inference geared towards spatiotemporal risk modeling. Specific unsolved applied problems, identified in collaboration with scientists in various applied disciplines (climate science, ecology, geomorphology, soil science...), will serve as motivation and for case studies. Some avenues of future research have already been outlined in the preceding chapters presenting results of my finalized work. The following presentation highlights other projects that are already work in progress in some cases, and it outlines more general and fundamental research questions that I aim to tackle in my future work.

11.1 My ecosystem of collaborations

Proposing new research tools and advancing knowledge on agricultural, environmental and ecological questions and problems is a highly collaborative endeavor. Within INRAE, I have established important and stable collaborations. At BioSP, we have a multidisciplinary team assembling experts on spatiotemporal statistics, dynamical systems, ecology and epidemiology. I work with colleagues on methodological advances in spatiotemporal statistics for extremes and point patterns, and also on new tools for the stochastic simulation of agricultural landscapes, and for sensitivity analyses of agroecological processes (*e.g.*, pest-predator systems) with respect to landscape composition and configuration. With colleagues from the URFM unit in Avignon, whose work focuses on Mediterranean forests, I collaborate on wildfire modeling, and with the Sol/Infosol units in Orléans on soil modeling.

Moreover, collaborations with colleagues outside of INRAE have become a cornerstone to propose relatively fundamental theoretical and novel methodological approaches, especially in the context of extreme-value theory, *e.g.*, with collaborators at KAUST, in Lausanne, at the universities of Montpellier, Nice, Lancaster and Venice, and at ITC Twente, among others.

I plan to strengthen these existing collaborations, and amend them with new collaborations, especially within France, and in particular in the Provence-Alpes Côte d'Azur region where the INRAE Avignon site is located.

11.2 Modeling spatiotemporal extremes

In the field of extreme-value theory, I will continue to develop novel theory and methods for spatiotemporal modeling and prediction of extreme episodes.

11.2.1 Learning aggregation functionals that drive extreme impacts

Special care is required when the extreme character of events or of environmental conditions does not become manifest in very extreme observed values of a process, but is rather the result of a specific and rare combination of temporal duration, spatial extent and moderately high observations, *i.e.*, of an extreme value in an appropriately defined aggregation over space and time of certain processes. Inferring this aggregation could be viewed as a problem of high-dimensional (functional) statistical learning. One can mention the example of precipitation deficits, whose accumulation can generate extreme drought conditions that build up slowly, especially when combined with relatively strong positive temperature anomalies. Another relevant case is the combination of the values of environmental and climatic factors occurring during an agricultural cropping season that strongly impact the phenology of crops (*i.e.*, their life cycle) and may provoke extremely low agricultural yields. The modeling of multivariate spatiotemporal dependencies and extreme-value theory will be at the core of the extensions to the state-of-the-art that I intend to develop with my collaborators. The scale-profile decompositions of extreme event episodes discussed in this manuscript are a promising base for developing such approaches.

11.2.2 Nonstationary extremes in space and time

I propose to develop stochastic models to gain precise insights into space-varying behavior of temporal trends in the frequency, magnitude and spatiotemporal extent of extreme episodes, and to allow for the detection of such behavior and its attribution to climate change. Nonstationary extreme-value modeling is yet in its infancy, especially in the space-time context, and a major goal is to propose appropriate extensions to be applied to Earth surface and atmosphere data. Seasonal and multi-annual cycles (e.g., El Niño, La Niña), as well as long-term trends owing to anthropogenic climate change, must be captured appropriately. The theoretically-founded and asymptotically dependent state-of-the-art models in EVT may lack flexibility and may be rather unwieldy for capturing complex nonstationarties in relatively high-dimensional data. In contrast, the conditional extremes framework, originating in the work of Heffernan and Tawn (2004), preserves strong asymptotic motivation but allows for flexible dependence modeling of extremes through marginally transformed Gaussian processes. Its recent extension to spatial and spatiotemporal processes (Wadsworth and Tawn, 2019; Simpson and Wadsworth, 2020), used in combination with generalized additive modeling (GAMs, see Castro-Camilo et al., 2020, for instance), will be a good point of departure for nonstationary modeling in both dependence and marginal distributions, *i.e.*, of weather event structure and climate conditions in the context of climatic applications. The Gaussian dependence in these models can also serve as entrypoint for using probabilistic machine learning models, which are often based on flexible Gaussian structures. The results obtained within the PhD project on the stochastic geometry of extreme excursion sets, starting in spring 2021 and discussed in §10.2.2, will further contribute to this axis of research.

11.2.3 Global extremes

Modeling of large-scale extremal dependence patterns in processes observed over the whole globe have received relatively little attention so far. In climate science, dependence models for global extremes could improve our understanding of large-scale atmospheric processes and climate change, and they could provide new perspectives on the analysis of global circulation patterns and related data, *e.g.*, climate model outputs such as reanalyses or projections. In this setting, models and estimation routines must be adapted to the spherical support of the globe and to a potentially very large number of observation locations and times, especially when gridded data are used. An interesting lead would be to extend the conditional extremes model, formulated and estimated with a Gaussian residual process through the INLA-SPDE approach, and already implemented for standard Euclidean 2D space and relatively small spatial domains (a part of the Red Sea) in Simpson et al. (2020). With the spherical domain of the globe, Matérn-like SPDE models can be constructed as before (Lindgren et al., 2011), and the numerical complexity remains comparable to the setting of a Euclidean space. The implementation of spherical SPDE models is readily available in the `R-INLA` package. Moreover, to take into account nonstationarities in the extremal dependence structure, it would be possible to include a small number of covariates in the range and variance parameter of the Gaussian process used in the conditional extremes model, or to apply a space transformation approach where the spherical support is transformed before deploying the SPDE that will define the Gaussian covariance structure. The inclusion of temporal nonstationarities into the model as mentioned in §11.2.2, and the implementation of statistical inferences for related parameters, could further allow conducting detection and attribution studies with respect to climate change effects.

If modeling the process over the full globe is too intricate due to the complexity of the data-generating process, we could still use the conditional extremes model defined over part of the spherical support to characterize local extreme climate properties and their variability over the globe. In climate science, one usually distinguishes "climate", corresponding to the univariate marginal distributions in the climate processes, from "weather", corresponding to local spatiotemporal dependence in the weather variables. We could therefore consider spatial and spatiotemporal conditional extremes models that are nonstationary with respect to the conditioning location, *i.e.*, we let the parameters of the conditional extremes model vary over the globe and then study the differences in the characteristics of the estimated models. This approach bears some resemblance with local likelihood modeling (Loader, 2006), used in the extreme-value setting with subasymptotic spatial models by Castro-Camilo and Huser (2020). The characteristics of the local conditional extremes model can be thought of as characterizing the local climate, *i.e.*, we would include local dependence properties as part of the local climate. A potential application of this could be a clustering approach based on the local climate characteristics, such that we could achieve a classification of locations (and maybe also seasons) into a moderate number of different climate classes.

11.2.4 Fast robust likelihood-free estimation of dependent extremes

Likelihood-based estimation can be considered as the "gold standard" of statistical inference when likelihood expressions are relatively easy and fast to compute and when we assume that models correctly capture the specifics of the data-generating mechanism. However, in peaks-

over-threshold modeling of dependent spatial or spatiotemporal extremes, the computation of likelihoods is often extremely computer-intensive and time-consuming due to multivariate censoring schemes, and the computational cost may become prohibitively high, especially in cases where we have data for many observation locations, such as gridded data. In this situation, one could resort to estimation approaches that scale better with very large data volumes. An example outside the extreme-value context is the method of regression kriging for Gaussian data in classical geostatistics, which combines (i) a regression model for the trend of the response distribution with (ii) moment-based estimation of the covariance function in the residuals using weighted least-squares techniques.

In the extreme-value context, I propose exploring the following approaches in future work.

Regression kriging for conditional extremes

In the conditional extremes framework for spatial and spatiotemporal data (Wadsworth and Tawn, 2019; Simpson and Wadsworth, 2020; Simpson et al., 2020), the dependence structure can be flexibly and conveniently modeled through a nonstationary Gaussian process with the structure of its mean and covariance function suggested by asymptotic considerations and constraints that ensure a well-defined model. Therefore, it would be interesting to develop and estimate appropriately constrained generalized additive models with covariates in the Gaussian location and scale parameter to estimate the mean and variance function (so-called GAMLSS: *Generalized Additive Models for Location, Scale and Shape*), combined with a weighted-least-squares estimator of the correlation structure. This approach would build a bridge between classical geostatistics and EVT, and it would lead to a relatively robust estimation technique that scales well with the size of datasets. Moreover, the non-Gaussian marginal distributions of tails of the data can also be modeled using approaches of type GAM or GAMLSS, see Mhalla et al. (2019); Castro-Camilo et al. (2020) for examples.

Robust rank-based estimation of marginally transformed Gaussian mixtures

With the Gaussian scale mixture models discussed in §6.1.1, it is possible to estimate correlation matrices by using only ranks. As highlighted in work of Hult and Lindskog (2002), the rank-based correlation coefficient known as Kendall's tau is invariant with respect to the choice of the distribution of the random scale variable, and a simple closed-form bijective transformation to the linear correlation coefficient exists. Therefore, in case of replicated observations of a Gaussian scale mixture process, it is possible to estimate the correlation matrix of the (latent) Gaussian process using only ranks, even if the Gaussian process has a random scale and has been transformed marginally (in a strictly monotonic way). This allows for very flexible modeling of spatial data through Gaussian scale mixtures, especially if these have been transformed marginally towards a marginal distribution that is different from the symmetric marginal distributions of Gaussian scale mixtures possessing heavier-than-Gaussian tails. To model the full distribution of data and not only threshold exceedances, but while still keeping focus on appropriately capturing extremal dependence in the upper tail, one could therefore proceed through the following "robust", likelihood-free approach applied to Gaussian scale mixture models: (i) estimate trends in the median of the process using quantile regression; (ii) estimate the Gaussian correlation coefficients using rank-based estimators of pairwise correlations; (iii) estimate cor-

relation parameters using weighted least squares between empirical and parametric correlation coefficients; (iv) estimate parameters of the distribution of the random scale variable by contrasting empirical and theoretical values of the finite-sample estimates of the χ and $\overline{\chi}$-measures; see §2.3.1. If the data correspond can be represented as a parametric marginal transformation of the Gaussian scale mixture, then the transformation parameters could be estimated by minimizing so-called empirical distribution function statistics, such as variants of the Cramér-von Mises or Anderson–Darling statistics. An approach using the above step (ii) has been proposed for semiparametric modeling of the extremal dependence in multivariate data in Klüppelberg et al. (2008), and consistency of estimators has been established. A combination of steps (ii) and (iii) is proposed in Klüppelberg and Kuhn (2009), again with theoretical results on asymptotic estimator properties. These results from the literature can be used to study the theoretical properties of estimators based on the steps described above. For more general subasymptotic models, such as location-scale mixtures of Gaussian processes, the minimization of some contrast between empirical and theoretical values of the finite-sample estimates of the χ and $\overline{\chi}$-measures could also be explored; asymptotic results for the estimators could be obtained based on recent work of Lalancette et al. (2020).

Scoring rules as loss functions beyond the likelihood

In a more general POT setting beyond the ideas discussed for conditional extremes in §11.2.4, we can replace the loss function, corresponding to the Kullback-Leibler loss when we maximize the likelihood, with a more general loss. A promising lead is the use of *proper scoring rules* (as defined byGneiting and Raftery (2007)) in the loss function. An example is the use of the gradient score in (de Fondeville and Davison, 2018), which bypasses the heavy computation of multivariate integrals in multivariate likelihoods implementing censoring of components falling below a marginal threshold. Such approaches could be generalized to subasymptotic models, *e.g.*, Gaussian mixture processes, and the construction of more general scoring rules for this purpose could be tackled. Alternatively, scoring rules such as the continuous ranked probability score (CRPS), used to assess spatial or spatiotemporal univariate predictions (*e.g.*, at a location whose data were not used during estimation), could be computed based on cross-validation and then minimized to obtain parameter estimates.

One may object that such approaches lead to a conceptual problem with model validation since scoring rules were designed to validate predictions after parameter estimation, but here we would already use them to estimate parameters. However, with relatively large datasets, a separation of datasets into training data (where scores are used for estimation) and validation data (where scores are used for validation) could be envisaged and would present a solution to this problem.

11.2.5 Assessing probabilistic predictions of heavy-tail phenomena

The validation of predictions of extremes of distributions with very heavy tails is notoriously difficult. Most standard validation metrics and scores are not appropriate since they may have infinite mean or variance. For instance, the variance of a distribution does not exist if its tail index is larger than one half, and the mean does not exist if its tail index is larger than one. Even if these moments exist, the behavior of prediction scores may not be stable and robust

enough to warrant their use in practice. An example is the validation of predictions of burnt areas of wildfires, or of burnt areas aggregated for various spatiotemporal mapping units, as discussed in Chapter 9. Wildfires are known to be quite heavy-tailed in most wildfire-prone areas of the globe (Pereira and Turkman, 2019), often with tail indices estimated above 0.5 or even above 1. Some of the strong variability of extreme event sizes can be captured through the inclusion of covariate information and spatial and temporal random effects, but even then the tails conditional to such information remain relatively heavy.

So far, there is no good solution to validate predictions in this context. The additive contributions of predictions for individual observations to measures such as mean squared error, mean absolute error or CRPS inherit the heavy-tail behavior of the observations. Similar issues arise with landslides (Lombardo et al., 2020) and insurance claims after windstorms (Mornet et al., 2015, 2017). The validation problem could be transformed into a problem for ordered categorical predictions if we work with exceedance probabilities over a number of increasingly high severity thresholds, but this is rather an ad-hoc approach that requires application-specific constructions of the score, and it also neglects tail behavior when extrapolating beyond the highest threshold. A different solution for comparing the performance of several forecasters could consist in developing criteria for comparing the distribution of the contributions of the forecasters' predictions of individual observations to the overall score. For example, with CRPS one usually reports the average of CRPS values for the observations to predict. Taillardat et al. (2019) demonstrate that this is not satisfactory for validating extreme-event predictions. Instead, the distributions of CRPS values of different forecasters should be compared more generally by directing stronger attention to the tail behavior of the CRPS distribution.

In practice, I suggest to explore the use of stochastic dominance properties between different empirical CRPS distributions, which could allow concluding on the superior performance of one forecast over another if a clear dominance relation arises. More generally, more thought should be invested to find good solutions for communicating heavy-tailed forecast distributions (*e.g.*, posterior predictive distributions in Bayesian models) in a concise and visually appealing fashion to stakeholders.

11.3 Stochastic geometry of extremes in large gridded data

Stochastic geometry (Chiu et al., 2013) is a branch of probability theory and statistics that is concerned with the study of geometric patterns and shapes, especially of point patterns. The characterization of excursion sets of stochastic processes in Euclidean spaces has also given rise to a large body of literature (see the monographs of Adler, 1981, 2010). When data are available on relatively dense spatial grids over relatively large spatial supports (*e.g.*, climate model output, satellite observations), then new insights could be gained by studying extreme excursion sets in these data, or by considering the locations of local extreme hotspots as the points of a point pattern.

11.3.1 The stochastic geometry of excursion sets

Summary statistics of excursion sets of spatial or spatiotemporal extreme-event episodes can serve as a parsimonious and easily interpretable representation of spatiotemporal extreme-value

behavior when focusing on large-scale properties rather than on small-scale variations like those captured by Gaussian-based random field constructions defined over continuous space. Specifically, the probability distributions of the extent, the perimeter and the number of connected components and holes of excursion sets are of interest. They allow for useful interpretations; for instance, in the ecological context the perimeter of an excursion set for a high threshold can be viewed as measuring the size of the interface between regions where components of an ecosystem are under high stress (excursion) or suffer from only moderate/low stress (no excursion).

While existing theory (see Adler, 1981, 2010) mostly focuses on Gaussian random fields and single replicates of a spatial process, we can extend and transpose those results to the setting of spatiotemporal EVT by working with Gaussian mixture representations and their extreme-value limits (extremal-t dependence, Brown–Resnick-type dependence). Specifically, results on the moments (expectation, variance...) of the three above-mentioned characteristics of excursion sets are relevant and can be derived from the mixture representation; this is work in progress. Based on such developments, new empirical and parametric estimators of characteristics of extreme excursion sets can be developed, and novel summaries for extreme excursions may arise from these developments.

In the spatial analyses used in landscape ecology (Naveh and Lieberman, 2013), geometric summaries play an important role; see also our work on stochastic agricultural landscape simulators in Zamberletti et al. (2021a). Therefore, I nourish the expectancy that we can cross-fertilize landscape ecology and spatiotemporal EVT. For example, we could obtain useful definitions of relevant coefficients of extreme excursion sets from landscape ecology for EVT, and we could use EVT-based characterizations of the geometry of extreme excursions in environmental and climatic conditions to drive and improve ecological analyses.

This body of work will be conducted by a PhD student starting his thesis in the spring of 2021, in collaboration with Elena di Bernardino (3IA Côte d'Azur, Nice University).

11.3.2 Marked point process analyses for local extrema

By adopting a bird's eye view on the large-scale spatial and spatiotemporal behavior of extremes, an extreme episode can be considered as a point indexed by space and time (*e.g.*, the barycenter of an excursion set, or a local maximum of the process). A point may carry additional information as point marks, for instance the value of the exceedance of the local maximum, or the area of the excursion set. Therefore, we obtain a spatiotemporal marked point pattern, and we can apply techniques from stochastic geometry to explore and model the spatiotemporal trends in the occurrence intensity, in the spatiotemporal distribution of marks conditional to the occurrence pattern, and in the local interactions between points and marks. The point process representation can be viewed as a meta-modeling approach that still preserves a lot of valuable information about the extreme-event structure, but it may lead to a lower-dimensional representation of gridded data through a comparatively small set of marked points. This approach is of interest when considering excursion sets and local hotspots of the raw data without marginal transformations, but it could also be used to study such properties in anomalies obtained after marginally normalizing observations with respect to local climate properties (*e.g.*, mean, variance, tail parameters). Moreover, it can be considered as a spatiotemporal declustering method for peaks-over-threshold analysis.

In work in progress with colleagues at BioSP and the Mathematics department (LMA) of

Avignon University, we consider local maxima in gridded reanalyses of US temperatures, and we develop a hybrid of Gibbs processes and LGCPs using the INLA-SPDE approach. Due to our extraction algorithm of hotspots, these cannot occur in adjacent pixels, *i.e.*, there is some form of local spatial inhibition between the points. We take into account local inhibition of points by capturing them through "mechanistic" interactions in the Gibbs energy, and large-scale aggregation patterns are captured through Gaussian random effects in the trend function of the Gibbs model.

11.4 Modeling compound extremes and multiple risks

My current work encompasses promising projects geared towards the spatiotemporal analysis of climatic, environmental, ecological and agro-epidemiological risks. While the deployment of cutting-edge techniques of statistical modeling has led to important progress in the mapping and prediction of high-impact events of specific types (*e.g.*, landslides, wildfires, windstorms, heatwaves, agricultural pest hotspots), the study of the potential effects of concurrence and interaction of multiple event types is still largely underexplored. The understanding, modeling and probabilistic prediction of such multiple events has been identified as a major methodological challenge, at the international level within the Sendai framework of the United Nations Office for Disaster Risk Reduction, but also at the national French level in prospective outlooks of research institutes (INSU OSU, INRAE...). The presence of interdependence, or more specifically of cascading effects, may lead to a nonlinear scaling of event magnitudes, occurrence probabilities and joint impacts as compared to a separate, compartmentalized analysis of events with an "additive" treatment of separate risks. The notion of *compound events* has been coined for climate-related events comprising a combination of several types of risk drivers or risks (Leonard et al., 2014; Zscheischler et al., 2018). I plan to develop models and methods for the joint probabilistic analysis of multiple events by using relevant spatial and temporal scales at the observation and process level within Bayesian hierarchical modeling, so far deployed mostly in the setting of single event types. Moreover, developments in multivariate spatial and spatiotemporal extreme-value analysis will be important to characterize the interplay of extreme conditions in drivers and risks. A crucial task is to achieve accurate propagation of uncertainties between model components, especially between predictors associated to the different response variables (*i.e.*, to different event types). Moreover, the projection of joint risks in the setting of possible changes in climate and Land Cover Land Use (LULC) is important for long-term prospective risk analysis and the deployment of resilience strategies.

To give an example of cascading multiple events, let us consider the system composed of drought, wildfires, extreme precipitation and landslides. Metereological drought is a key condition for the outbreak of wildfires. Wildfires consume combustibles such as trees and other vegetation, laying bare the soil and making it sensitive to atmospheric conditions as well as erosive and sliding processes. In particular, tree roots that have previously stabilized slopes with non-solid surfaces may be burnt or may rot, thus creating instabilities and channels for precipitation run-off. If this situation arises during the years following a large wildfire event, then major landsliding events may be triggered by extreme precipitation, and in particular landslides may occur at intermediate slope angles where the soil structure has been destabilized. Therefore, joint modeling of drought conditions, precipitation extremes, wildfire and landslide risk is

of interest. It involves modeling of two event cascades (droughts $\rightsquigarrow$ wildfires, precipitation $\rightsquigarrow$ landslides), and it could further be modified to make allowance for the explicit modeling of soil properties. Another nexus of compound events involving wildfires concerns the emission of large quantities of greenhouse gases and atmospheric pollution. In this respect, one may even observe feedback loops at relatively large scales (*e.g.*, global) when greenhouse gas emissions further exacerbate climate change, which in turn may lead to even stronger meteorological extremes, especially drought conditions (Jones et al., 2020).

In the ecological context, these multi-risk developments will contribute to the field of predictive approaches in biology, ecology and agriculture *i.e.*, to the development of methods, often data-driven or simulation-based, that allow predicting how biological or ecological systems may react to changes such as modified external forcings. For example, one may study the risk of crop yield losses in an environment characterized by an increase of extreme climatic risks, or one may study the risk of invasive species (which could be hosts of pathogens of cultivated crops) propagating as a consequence of climatic or land-use changes.

Two more specific topics that I wish to explore are listed in the following.

11.4.1 Risk functionals and risk measures for compound extremes

The term *compound extremes* has been coined in the climate science literature (*e.g.*, Zscheischler and Seneviratne, 2017; Zscheischler et al., 2018). Compound extremes arise when several risk variables or drivers of such variables are simultaneously in a state that can be considered as extreme. This notion is therefore closely related to multivariate extremes. In the current literature, compound extremes of variables defined over spatiotemporal supports are typically studied by using an aggregation functional to summarize spatially indexed observations of a variable into a scalar value, and then the multivariate dependence of such scalar variables can be studied (*e.g.*, Zscheischler et al., 2021). The extension and application of generalized Pareto processes or conditional extremes models to a multivariate spatial or spatiotemporal setting could improve the knowledge about the spatiotemporal variability when compound phenomenons occur. In particular, the construction of aggregation functionals for multivariate spatiotemporal risk drivers, whose exceedances appropriately identify situations of high risk (*i.e.*, of extreme states arising in one, or jointly in several response variables), could be formalized through novel methodological developments and can be considered as a statistical learning problem.

Moreover, the notion of risk measures, or more specifically the class of *coherent risk measures* satisfying certain axioms (*e.g.*, Delbaen, 2002), is commonly used in the actuarial context for univariate or multivariate variables (Albrecht, 2014). Risk measures are deterministic scalar quantities that give a measure for the severity of a risk (*i.e.*, of some random quantity). Extensions to the multivariate spatiotemporal context are still rare (some simple ad-hoc defined risk measures were used in our work Palacios-Rodriguez et al. (2020)), but would be relevant for many applications. Some ideas have been developed for univariate spatial processes, *e.g.*, for Gaussian processes (Ahmed et al., 2016), max-stable processes (Koch, 2017) or more general processes used to model maxima data (Ahmed et al., 2020), and would merit extensions to more general settings.

11.4.2 Joint modeling of multiple risks

Beyond the multivariate modeling of components of extreme risks, there is so far no standard approach for the joint modeling of multiple risks, especially in the spatial and spatiotemporal setting. Models will become more complex when several risk components have to be modeled jointly in a realistic manner, and accurate modeling of their potentially very complex interactions is challenging. The propagation of different types of uncertainties (natural variability in the modeled processes, model uncertainty, estimation uncertainty of parameters, uncertainty due to different scenarios of climate change...) among model components and towards the resulting predictions for risks is particularly difficult to handle. Hierarchical modeling approaches, especially Bayesian hierarchical modeling, are a promising lead to achieve sound inferences, predictions and uncertainty assessments. Especially the modeling framework of INLA-SPDE allows for incorporating multi-source multi-scale data into multi-component models for one or several response variables, where complex covariate effects and spatial-temporal random effects can be captured. The propagation of uncertainties between different components can be achieved through correlated random effects, or sharing of random effects from one component to others using a sharing coefficient to be estimated; *i.e.*, we put a random effect $W(s,t)$ in one component, and it can be copied and rescaled to appear in other components as $\beta_{\text{comp}}W(s,t)$, with the hyperparameter β_{comp} to be estimated.

Simulation-based approaches also play an important role, especially when we deploy a coupling of bio-geo-physical procerss models and stochastic/statistical models. The work on agroecological risks (see §10.1.3) is a good example of coupling stochastic landscape models (with statistical inference of landscape parameters) and spatially explicit population dynamics models, numerically simualted according to a system of partial differential equations.

11.5 Spatiotemporal modeling of complex ecological processes

I will invest in collaborative efforts to tackle the following challenges in statistical ecology. Current state-of-the-art approaches in quantitative ecology still often disregard important spatial variations in ecological phenomena (*e.g.*, by using mean-field approaches where variables are aggregated over space and/or time). Moreover, the development of spatiotemporal modeling, *i.e.*, the inclusion of temporal dynamics, is still in its infancy in many ecological fields including species distribution modeling. Through my collaborations with quantitative ecologists at BioSP and beyond, I aim to help contribute new methodological tools for spatiotemporal analysis of ecological phenomena as outlined in the following.

11.5.1 Marked point process meta-models of dynamical systems

As already mentioned in §§11.3.2,10.1.3, marked point processes and stochastic geometry tools could be used for studying the geometry of extremes in spatiotemporal stochastic processes and gridded datasets. A similar analysis can be envisioned for the numerical simulations of spatially explicit dynamical systems, for instance systems of Lotka–Volterra-like equations for population dynamics. Then, a marked point process describing key events of the spatiotemporal dynamics

(*e.g.*, exceedance of a threshold of a pest species in an agricultural landscape) can be viewed as a "meta-model" for the full dynamical system, and sensitivity analyses of the system with respect to its parameters can be carried out through models and tools for point patterns and stochastic geometry. In the project conducted by Patrizia Zamberletti in her PhD thesis at BioSP, we apply this approach to simulations of prey-predator systems in agricultural landscapes. A parametric stochastic landscape simulator is used to generate a multitude of landscape supports (crop and non-crop fields, presence and absence of hedges on field boundaries), and then a 1D-2D population dynamics model is simulated according to parameters controling the dynamics of prey and predator species with respect to different habitat types (1D = hedge, 2D = crop or semi-natural). When the pest density in a field exceeds a predefined threshold, then a pesticide treatment is applied, reducing the local pest density to a very small value. We consider the local pest hotspots leading to pesticide treatments as key agroecological events, and we model the locations of the corresponding local maxima and threshold excesses as a spatiotemporal marked point process. Local and global landscape properties, and population dynamics parameters, are considered as predictors in a system of two regression equations for the point process intensity and the size of the excess. This approach provides new insights into the spatiotemporal dynamics of pest hotspots, and it shows how biological control measures can be designed to keep pest densities below an economic threshold where pesticide treatments have to be applied.

While we do not use asymptotic models (*e.g.*, GPDs) in this analysis, a further integration of extreme-value techniques could be explored in future research, and the approach could be used more generally for sensitivity analyses of outputs of more general spatially indexed dynamical systems in ecology and other disciplines.

11.5.2 Intervention events in space-time point patterns

In ongoing work, I adapt tools of stochastic geometry (especially space-time variants of Ripley's K function) to assess how the occurrence of one type of event, which we may call an *intervention*, influences the occurrence intensity of a second type of event. This provides insights into dynamical structures of processes and can be linked to causal modeling. It is also related to the concept of Before-After-Control-Impact (BACI) analyses, as established by Stewart-Oaten et al. (1986), although BACI is usually not carried out in the context of multi-type point processes. To make this abstract concept more tangible, let us consider two examples that I am currently studying with partner institutes. The first example concerns wolf attacks on sheep herds, which have become increasingly frequent over recent years due to the recolonization of France by wolf packs. An important yet unanswered question concerns the effect of wolf removals (*i.e.*, by shooting them = intervention) on the occurrence pattern of wolf attacks on sheep herds. I am currently investigating this problem with colleagues from the French Biodiversity Office (OFB) and CEFE-CNRS in Montpellier. The second example concerns Asian hornets, an invasive species in France and Europe that inflicts major damage on native bee populations. Based on citizen science data collected by beekeepers and other stakeholders, we want to explore the effect of deploying traps for capturing the Asian hornets that found new nests during the spring season (= intervention) on the space-time point pattern of observed locations of Asian hornet nests. I am coordinating and implementing the statistical modeling efforts within a partnership between BioSP, the ITSAP institute and the Museum National d'Histoire Naturelle (MNHN). Finally, the pesticide treatments in the pest-predator systems discussed in §11.5.1 could also be seen

as intervention events, and a BACI analysis of the spatiotemporal point pattern of pesticide hotspots would provide a novel way of analyzing these model outputs of numerical simulations of dynamical systems.

11.5.3 Spatiotemporal preferential sampling in species distribution data

It is a challenging task to model observation processes (*i.e.*, what typically corresponds to the top layer in a hierarchical model) when the observation effort is not uniform and not known from a precisely defined observation protocol. The rise of crowdsourced data collection, *e.g.*, in the context of ecological projects for the observation of animal or vegetal species, but also the availability of large historical datasets without strict observation protocols collected by national parks in France, call for innovations in quantitative modeling through appropriate statistical models and methods. Some statistical tools already allow us to mitigate or even fully control potential biases in statistical analyses, often thanks to the availability of smaller data subsets of protocoled data (Pennino et al., 2019; Gelfand and Shirota, 2019). One of my research perspectives, in collaboration with colleagues at BioSP and the CisStats-network of INRAE (for statistics in the context of Citizen Science), aims to explore and model such preferential sampling effects and their space-time dynamics. We seek to disentangle preferential sampling effects from species distribution dynamics by exploiting multi-source data with different types of observation protocols (or absence thereof). Strong connections to space-time modeling of marked point processes will arise. We will implement such approaches within the INLA-SPDE framework enabling Bayesian modeling and the complex assessment of uncertainties.

Another important extension of state-of-the-art species distribution models concerns multi-species modeling with a moderate to large number of species (say, between 2 to 100), for which variants of parsimoniously defined coregionalization models could be developed, implemented and compared using the INLA-SPDE approach. For instance, Choiruddin et al. (2020) have developed frequentist regularized estimation of cross-correlation parameters of log-Gaussian Cox processes in a highly multivariate setting, and we aim to transfer their approach to a Bayesian setting where parameters, point process intensities and associated uncertainties can be estimated more holistically.

Moreover, preferential sampling effects could also be explored in the context of extreme values of continuous variables. For example, air pollution measurement networks are often denser in urban or industrial regions exposed to extreme concentrations. For this goal, point process models with extreme marks, similar to the models presented for wildfire modeling in §9.3, could be used.

11.5.4 Stochastic geometry for hidden elements in forest remote sensing

In work started during the Master project of Thomas Minotto at BioSP in 2018, we have worked on the problem of the detectability of trees through remote sensing. When using aerial photos and airborne laser scanning to identify individual trees and their properties, and to estimate population measures such as total biomass, biases arise since some trees may not be

correctly identifiable or may be hidden by others. By considering a forest as a marked point pattern, we have started to explore the influence of spatial interaction among trees (aggregation, repulsiveness, complete randomness) on the detection probability of trees. Points represent tree stems, and marks can be defined as the radius of the tree crown. Based on the K-function used to characterize pairwise interaction in point processes, and using certain assumptions on the mark distribution, we have obtained first theoretical results on the detection probabilities of trees. These results could be used to extend the estimation approaches developed in the PhD thesis of Kansanen (2020) under the assumption of complete spatial randomness.

In another context, regarding the estimation of the Leaf Area Density of trees in a forest (an important ecological indicator of the forest cover) terrestrial LiDAR sensors (light detection and ranging) can be deployed to sample the forest canopy structure around the location of the LiDAR instrument. The LiDAR.records the reflection of laser beams that it emits. Forest canopy elements farther away from the instrument are often occluded by nearer elements. Therefore, estimation of the forest canopy structure, especially of Leaf Area Density, must take into account the occlusion effects, which can be considered as a variant of preferential sampling. In Soma et al. (2020), colleagues from INRAE Avignon have proposed a kriging method that takes into account the censoring of laser beams when predicting the 3D Leaf Area Density structure from the LiDAR data. To extend the conceptual framework and to further improve estimations, the stochastic geometry of the laser beams and occluding objects could be taken into account in greater detail. Specifically, we could draw from the developments around the detectability of forest trees and spatiotemporal preferential sampling outlined previously, and we could develop Bayesian kriging techniques such as the INLA-SPDE approach.

11.6 Towards spatiotemporal data science of rare events

An overarching goal of my work is to significantly contribute to the development of a branch of data science that focuses on space-time data, and especially on observations at low occurrence frequency but presenting high impacts on the Earth system with its human societies and ecosystems. At BioSP, spatiotemporal data science has been identified as a thematic priority to be developed through joint working groups and new recruitments in our lab.

In my projects, the goal will consist of combining tools and relative benefits from machine learning and probabilistic space-time modeling in order to achieve powerful predictive approaches, where model outputs allow for probabilistic interpretation and decision support. The strengths of machine learning are rooted in its capability to treat high-dimensional "big" datasets, to use regularisation techniques, to work with simple-to-compute cost functions that differ from unwieldy classical likelihoods, and to operate with only few modeling hypotheses. By contrast, probabilistic spatiotemporal modeling places strong emphasis on explicit spatial-temporal dependence structures, seeks to control and accurately assess uncertainties, allows for attribution of response variations to covariates and risk factors through statistical inference, provides good interpretability of results, and enables robust problem-specific solutions grounded in theoretical domains including extreme-value theory and stochastic geometry. Merging the best of both worlds would offer huge benefits.

In particular, embedding machine learning tools such as boosting, regularization, graphical (Markovian) models or neural networks in a natural way into stochastic modeling of spatially

indexed rare event data is among the big challenges that I wish to help tackle. For instance, probabilistic Gaussian-based machine learning models could be embedded into the conditional extremes framework, where Gaussian processes offer high flexibility for capturing extremal dependencies. As to the modeling of discretely observed rare events (*i.e.*, point patterns), the estimation of spatiotemporal regression equations (see Chapter 7) presents a promising vantage point to apply classification algorithms from machine learning for the estimation of point process intensities and interaction potentials.

Therefore, future work should aspire to a powerful fusion of Bayesian and frequentist statistics with machine learning techniques, while making allowance for the specifics of spatial and spatiotemporal data.

Appendix A

Latent Gaussian modeling using the INLA-SPDE framework

A.1 The integrated nested Laplace approximation (INLA)

The integrated nested Laplace approximation (INLA) has been proposed as a tool for analytical approximation of posterior distributions in the wide class of Bayesian generalized additive regression models with Gaussian process priors, also referred to as *latent Gaussian models*. Since its inception in Rue et al. (2009), INLA has today become the prime inference approach for spatial Bayesian models thanks to its implementation in the `INLA` package of the `R` software. It also strongly benefits from its integration with the SPDE approach of Lindgren et al. (2011), which provides numerically advantageous Gauss–Markov representations of the Matérn covariance function, see §A.2.1. This section gives an overview of how approximations are carried out and highlights some of the available models. It draws strongly from the presentation in Opitz (2017).

A.1.1 Latent Gaussian models

Here we provide some general details on the latent Gaussian approach for spatial modelling, denoting the observed data generically by $\boldsymbol{y} = (y_1, \ldots, y_n)$. Following Bayesian convention, we use lower-case notations for parameters and random variables, and we use $\pi(\cdot)$ as the generic symbol to refer to probability densities (and the associated distributions). In hierarchical modeling with latent Gaussian processes, we define a latent, unobserved multivariate Gaussian vector $\boldsymbol{x} = (x_1, \ldots, x_d)$, and we assume conditional independence of the observations $\boldsymbol{y}$ with respect to $\boldsymbol{x}$. We use the so-called *observation matrix* $A \in \mathbb{R}^{n \times d}$ to define a linear predictor

$$\boldsymbol{\eta} = \boldsymbol{\eta}(\boldsymbol{x}) = A\boldsymbol{x} \tag{A.1}$$

that linearly combines the latent variables in $\boldsymbol{x}$ into components η_i associated with y_i, $i = 1, \ldots, n$. For instance, $\boldsymbol{x}$ may contain the values of a spatial field at nodes $\tilde{s}_1, \ldots, \tilde{s}_d$, and A has i-th line $A_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ if the observation location s_i of y_i coincides with one of the nodes $\tilde{s}_{j_0}$, and the 1-entry is at the j_0th position. Otherwise, several entries of A_i could have non-zero weight to implement interpolation between the nodes. The distribution of $\boldsymbol{\eta}$ is

also multivariate Gaussian, by analogy with $\boldsymbol{x}$. The univariate probability distribution of y_i, often referred to as the *likelihood model*, can be non-Gaussian and is parametrized by the linear predictor η_i, and potentially other hyperparameters related to the shape of the distribution. The vector of hyperparameters (*i.e.*, of parameters that are not components of one of the Gaussian vectors $\boldsymbol{x}$ and $\boldsymbol{\eta}$), such as parameters related to variance, spatial dependence range, or the smoothness of a spline curve, is denoted by $\boldsymbol{\theta}$. The hierarchical model is structured as follows:

$$\begin{aligned} \boldsymbol{\theta} &\sim \pi(\cdot) && \text{hyperparameters,} \\ \boldsymbol{x} \mid \boldsymbol{\theta} &\sim \mathcal{N}_m(\mathbf{0}, Q(\boldsymbol{\theta})^{-1}) && \text{latent Gaussian components,} \\ y_i \mid \boldsymbol{x}, \boldsymbol{\theta} &\overset{ind.}{\sim} \pi(\cdot \mid \eta_i, \boldsymbol{\theta}), && \text{likelihood of observations.} \end{aligned}$$

The matrix $Q(\boldsymbol{\theta})$ denotes the precision matrix of the latent Gaussian vector $\boldsymbol{x}$, whose variance-covariance structure may depend on some of the hyperparameters in $\boldsymbol{\theta}$ that we seek to estimate. For example, in the simplest case of observations y_i having a Gaussian distribution, the conditional variance σ^2 of y_i given η_i is a hyperparameter, and we define

$$y_i \mid \eta_i, \sigma^2 \sim \mathcal{N}(\eta_i, \sigma^2), \quad i = 1, \ldots, n.$$

Distributions available for the likelihood in `R-INLA` are the binomial, Poisson, negative binomial, log-gaussian, skew-gaussian, beta, exponential, gamma, Weibull, GEVD and GPD, among many others. The linear predictor is usually linked to one of the parameters of the distribution through a link function, *e.g.*, the log-link in the log-Gaussian exponential, Poisson cases, or the logistic link (logit) for the binomial and Beta cases.

A major benefit of the construction with latent variables is that the dimension d of the latent vector $\boldsymbol{x}$ is not directly determined by the number of observations n. The computational complexity and stability of matrix operations (*e.g.*, determinants, matrix products, solution of linear systems) arising in the likelihood calculations for the above Bayesian hierarchical model is therefore mainly determined by the tractability of the precision matrix $Q(\boldsymbol{\theta})$, whose dimension can be controled independently from the number of observations. Such matrix operations can be implemented very efficiently if precision matrices are sparse (Rue, 2005). If data are replicated many times with dependence between replications, such as spatial data observed at regular time steps in spatiotemporal modeling, the sparsity property can be preserved in the precision matrix of the latent space-time process $\boldsymbol{x}$, especially by using separable dependence structures where replications are linked together through a precision matrix that is again sparse, such as the precision matrix of a temporal Gaussian auto-regressive process.

A.1.2 Laplace approximation

The joint posterior density of latent variables $\boldsymbol{x}$ and hyperparameters $\boldsymbol{\theta}$ in a latent Gaussian model is

$$\pi(\boldsymbol{x}, \boldsymbol{\theta} \mid \boldsymbol{y}) \propto \exp\left(-0.5\boldsymbol{x}'\boldsymbol{Q}(\boldsymbol{\theta})\boldsymbol{x} + \sum_i \log \pi(y_i \mid \eta_i, \boldsymbol{\theta}) + \log \pi(\boldsymbol{\theta})\right). \tag{A.2}$$

This density over a high-dimensional space does usually not characterize one of the standard multivariate families and is therefore difficult to interpret and to manipulate directly. In this

setting, Laplace approximation is useful to derive posterior densities and estimations of practical interest. In practice, the main interest lies in the marginal posteriors of hyperparameters θ_j, of latent variables x_i and of the resulting predictors η_i, where the latter can be included into $\boldsymbol{x}$ for notational convenience. Calculation of these univariate posterior densities requires integration with respect to $\boldsymbol{\theta}$ and $\boldsymbol{x}$:

$$\pi(\theta_j \mid \boldsymbol{y}) = \int\int \pi(\boldsymbol{x}, \boldsymbol{\theta} \mid \boldsymbol{y}) \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{\theta}_{-j} = \int \pi(\boldsymbol{\theta} \mid \boldsymbol{y}) \mathrm{d}\boldsymbol{\theta}_{-j}, \tag{A.3}$$

$$\pi(x_i \mid \boldsymbol{y}) = \int\int \pi(\boldsymbol{x}, \boldsymbol{\theta} \mid \boldsymbol{y}) \mathrm{d}\boldsymbol{x}_{-i}\, \mathrm{d}\boldsymbol{\theta} = \int \pi(x_i \mid \boldsymbol{\theta}, \boldsymbol{y}) \pi(\boldsymbol{\theta} \mid \boldsymbol{y})\, \mathrm{d}\boldsymbol{\theta}. \tag{A.4}$$

The classical Laplace approximation used for latent Gaussian models was proposed by Tierney and Kadane (1986). Typically, one seeks to evaluate an integral $\int_{\mathbb{R}^d} f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}$, where the positive integrand function f, here written as $f(\boldsymbol{x}) = \exp(kg(\boldsymbol{x}))$ with a scale variable $k \geq 1$, is defined over a high-dimensional space with large d and is "well-behaved" in the sense that it satisfies some minimal regularity requirements, is unimodal and its shape is not too far from the symmetric Gaussian bell-shape; for instance, requiring strict log-concavity of f is useful, see Saumard and Wellner (2014). Since the integral value is mainly determined by the behavior around the mode of g, a second-order Taylor approximation of g can be substituted for g to calculate an approximate value of the integral. Assuming that $\boldsymbol{x}^\star$ is the unique global maximum of g, we get $g(\boldsymbol{x}) \approx g(\boldsymbol{x}^\star) + 0.5(\boldsymbol{x} - \boldsymbol{x}^\star)' \boldsymbol{H}(g)(\boldsymbol{x}^\star)(\boldsymbol{x} - \boldsymbol{x}^\star)$ for values $\boldsymbol{x}$ close to $\boldsymbol{x}^\star$ with the Hessian matrix $\boldsymbol{H}(g)(\boldsymbol{x}^\star)$. Notice that $-\boldsymbol{H}(g)(\boldsymbol{x}^\star)$ is positive definite. An approximate value of the integral can be calculated using the fact that a multivariate Gaussian density integrates to 1. The resulting following integral approximation in dimension d is expected to become more and more accurate for higher values of k, *i.e.*, when the area below the integrand $\exp(kg(\boldsymbol{x}))$ becomes concentrated more and more closely around the mode (Tierney and Kadane, 1986). :

$$\int_{-\infty}^{\infty} f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = \int_{-\infty}^{\infty} \exp(kg(\boldsymbol{x}))\, \mathrm{d}\boldsymbol{x} \tag{A.5}$$

$$\overset{k\to\infty}{\sim} \int_{-\infty}^{\infty} \exp(kg(\boldsymbol{x}^\star) + 0.5k(\boldsymbol{x} - \boldsymbol{x}^\star)' \boldsymbol{H}(g)(\boldsymbol{x}^\star)(\boldsymbol{x} - \boldsymbol{x}^\star))\, \mathrm{d}\boldsymbol{x}$$

$$= \left(\frac{2\pi}{k}\right)^{d/2} |\boldsymbol{H}(g)(\boldsymbol{x}^\star)|^{-1/2} \exp(kg(\boldsymbol{x}^\star)). \tag{A.6}$$

In statistical practice, k may represent the number of i.i.d. replications, each of which has density $\exp(g(\boldsymbol{x}))$. Higher values of k usually lead to better approximation, and more detailed formal results on the quality of approximation have been derived (Tierney and Kadane, 1986; Rue et al., 2009). Many of the models commonly estimated with INLA have no structure of strictly i.i.d. replication, but the Laplace approximation remains sufficiently accurate in most cases since there usually still is a structure of internal replication; ideally, for each latent variable x_{i_0} we have at least several observations y_i which contain information about x_{i_0}. Recall that observations are conditionally independent with respect to $\boldsymbol{x}$ by construction of the model.

In the context of INLA, the following observation will be interesting and useful. Fix $k = 1$ in (A.5) and suppose that $f(\boldsymbol{x}) = \exp(g(\boldsymbol{x})) = \pi(\boldsymbol{x}, \boldsymbol{\theta})$, where $\pi(\boldsymbol{x}, \boldsymbol{\theta})$ is the joint probability density of a random vector $(\boldsymbol{x}, \boldsymbol{\theta})$. Then, in (A.6), the term $\exp(g(\boldsymbol{x}^\star))$ is the value of π at

its mode $\boldsymbol{x}^\star$ for fixed $\boldsymbol{\theta}$, whereas $(2\pi)^{d/2}\,|\boldsymbol{H}(g)(\boldsymbol{x}^\star)|^{-1/2}$ is $1/\pi_G(\boldsymbol{x}^\star \mid \boldsymbol{\theta})$ with π_G a Gaussian approximation with mean vector $\boldsymbol{x}^\star$ to the conditional density of $\boldsymbol{x} \mid \boldsymbol{\theta}$. In practice, we can determine the mean $\boldsymbol{\mu}^\star = \boldsymbol{x}^\star$ and the precision matrix $\boldsymbol{Q}^\star = -\boldsymbol{H}(g)(\boldsymbol{x}^\star)$ of π_G through an iterative Newton–Raphson optimization. Starting from the joint posterior (A.2) of our latent Gaussian model, we set $g(\boldsymbol{x}) = -0.5\boldsymbol{x}'\boldsymbol{Q}(\boldsymbol{\theta})\boldsymbol{x} + \sum_i \log \pi(y_i \mid \eta_i, \theta)$. We further write $g_i(x_i) = \log \pi(y_i \mid x_i, \boldsymbol{\theta})$ and calculate its second-order Taylor expansion $g_i(x_i) \approx g_i(\mu_i^{(0)}) + b_i x_i - 0.5 c_i x_i^2$. Without loss of generality, we here assume that the linear predictor $\boldsymbol{\eta}$ corresponds to the latent Gaussian vector $\boldsymbol{x}$. We start the iterative optimization with initial values $\boldsymbol{Q}^{(1)} = \boldsymbol{Q} + \mathrm{diag}(\boldsymbol{c})$ and $\boldsymbol{\mu}^{(1)}$, where $\boldsymbol{Q}^{(1)}\boldsymbol{\mu}^{(1)} = \boldsymbol{b}$. We then iterate this procedure until convergence such that $\boldsymbol{\mu}^{(j)} \to \boldsymbol{\mu}^\star = \boldsymbol{x}^\star$ and $\boldsymbol{Q}^{(j)} \to \boldsymbol{Q}^\star = \boldsymbol{Q} + \mathrm{diag}(\boldsymbol{c}^\star)$, $j = 1, 2, \ldots$, $j \to \infty$, where an appropriate convergence criterion must be used. Notice that the conditional independence assumption of observations y_i with respect to $(\eta_i, \boldsymbol{\theta})$ allows preserving the sparse structure in $\boldsymbol{Q}^\star$. Moreover, a strictly log-concave likelihood function $x_i \mapsto \pi(y_i \mid x_i, \boldsymbol{\theta})$ ensures $c_i > 0$ such that $\boldsymbol{Q}^{(j)}$ are valid precision matrices and local curvature information around $\mu_i^{(j)}$ can be used for constructing a useful Gaussian approximation. It is further possible to impose linear constraints $\boldsymbol{M}\boldsymbol{x} = \boldsymbol{e}$ onto $\boldsymbol{x}$ and $\boldsymbol{x}^\star$ with given matrix $\boldsymbol{M}$ and vector $\boldsymbol{e}$ by using the approach of conditioning through kriging (Rue et al., 2009).

To calculate posterior marginal densities of hyperparameters,

$$\pi(\theta_j \mid \boldsymbol{y}) = \int\int \pi(\boldsymbol{x}, \boldsymbol{\theta} \mid \boldsymbol{y}) \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{\theta}_{-j} = \int \pi(\boldsymbol{\theta} \mid \boldsymbol{y}) \mathrm{d}\boldsymbol{\theta}_{-j}, \tag{A.7}$$

we use the Laplace approximation for the inner integral $\int \pi(\boldsymbol{x}, \boldsymbol{\theta} \mid \boldsymbol{y}) \mathrm{d}\boldsymbol{x} = \pi(\boldsymbol{\theta} \mid \boldsymbol{y})$ such that the approximated density $\tilde{\pi}$ satisfies

$$\tilde{\pi}(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\pi_G(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})} \,|_{\boldsymbol{x}=\boldsymbol{x}^\star(\boldsymbol{\theta})} \tag{A.8}$$

with $\boldsymbol{x}^\star(\boldsymbol{\theta})$ the mode of the joint density $\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})$ for fixed $(\boldsymbol{\theta}, \boldsymbol{y})$ and a Gaussian density π_G that approximates $\pi(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})$:

$$\pi_G(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y}) = (2\pi)^{n/2} |\boldsymbol{Q}^\star(\boldsymbol{\theta})|^{1/2} \exp\left(-0.5(\boldsymbol{x} - \boldsymbol{x}^\star(\boldsymbol{\theta}))'\boldsymbol{Q}^\star(\boldsymbol{\theta})(\boldsymbol{x} - \boldsymbol{x}^\star(\boldsymbol{\theta}))\right). \tag{A.9}$$

Notice that the Gaussian approximation π_G is exact if the likelihood model $\pi(y_i \mid \eta_i, \boldsymbol{\theta})$ itself is Gaussian. An approximation of the posterior marginal of θ_j in (A.7) is now obtained through a numerical integration with a set of integration nodes $\boldsymbol{\theta}_\ell$ chosen from a numerical exploration of the surface of the density $\tilde{\pi}(\boldsymbol{\theta}_{-j}, \theta_j \mid \boldsymbol{y})$ (with θ_j held fixed). This yields

$$\tilde{\pi}(\theta_j \mid \boldsymbol{y}) = \sum_{\ell=1}^{L} \omega_\ell \tilde{\pi}(\boldsymbol{\theta}_\ell \mid \boldsymbol{y}) \tag{A.10}$$

with weights ω_ℓ (which are chosen to be equal in the approaches implemented in `R-INLA`). In `R-INLA`, $\boldsymbol{\theta}_\ell$ can either be chosen as a grid around the mode of $\tilde{\pi}(\boldsymbol{\theta} \mid \boldsymbol{y})$ (`int.strategy="grid"`, the most costly variant), or through a simpler so-called *complete composite design* which is less costly when the dimension of $\boldsymbol{\theta}$ is relatively large (`int.strategy="ccd"`, the default approach), or we may use only one integration node given as the mode value (`int.strategy="eb"`, corresponding to the idea of an *empirical Bayes* approach).

For calculating posterior marginal densities of the latent Gaussian field, *i.e.*, the marginal density $\pi(x_i \mid \boldsymbol{y})$ of a latent variable x_i, we lean on representation (A.4). Numerical integration is performed with respect to the integral $\boldsymbol{\theta}$, and the Laplace approximation (A.8) allows approximating $\pi(\boldsymbol{\theta} \mid \boldsymbol{y})$. It thus remains to (approximately) evaluate $\pi(x_i \mid \boldsymbol{\theta}, \boldsymbol{y})$. A simple and fast solution would be to use the univariate Gaussian approximation resulting from the multivariate Gaussian approximation (A.9) whose mean value is $x_i^\star(\boldsymbol{\theta})$ and whose variance can easily and quickly be calculated from a partial inversion of the precision $\boldsymbol{Q}^\star(\boldsymbol{\theta})$ (Rue, 2005) (`strategy="gaussian"` in `R-INLA`). However, this Gaussian approximation often fails to capture skewness behavior and can generate nonnegligible bias in certain cases – an important exception to this issue being the case where the data likelihood is Gaussian. In the general case, using again a Laplace-like approximation

$$\frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\pi_G(\boldsymbol{x}_{-i} \mid x_i, \boldsymbol{\theta}, \boldsymbol{y})} \Big|_{\boldsymbol{x}_{-i}=\boldsymbol{x}^\star_{-i}(x_i,\boldsymbol{\theta})} \tag{A.11}$$

with mode $\boldsymbol{x}^\star_{-i}(x_i, \boldsymbol{\theta})$ of $\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})$ for fixed $(x_i, \boldsymbol{\theta}, \boldsymbol{y})$ would be preferable, but is relatively costly (`strategy="laplace"` in `R-INLA`). Instead, Rue et al. (2009) propose a so-called *simplified Laplace approximation* based on third-order Taylor developments of numerator and denominator in (A.11) that satisfactorily remedies location and skewness inaccuracies of the Gaussian approximation (`strategy="simplified.laplace"` in `R-INLA`, the default). Notice that the "Nested" in INLA refers to this second Laplace-like approximation. Another recent extension implemented in `R-INLA` is known as the `adaptive` approximation and uses the simplified Laplace approximation for latent Gaussian components composed of a number of latent Gaussian variables below a user-defined maximum (*e.g.*, 100), whereas the Gaussian approximation is applied to larger components. This ensures a good approximation for important model components such as fixed effect coefficients, while memory and time are saved through the potentially cruder approximation of larger components.

A.2 Gauss–Markov models

We say that a random vector $\boldsymbol{x} \mid \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}^{-1})$ is Gauss–Markov if the number of non null entries of its $n \times n$ precision matrix $\boldsymbol{Q} = (q_{ij})_{1 \le i,j \le n}$ is $\mathcal{O}(n)$. Such sparse precision matrices allow efficient numerical computation of matrix operations such as LR-decomposition (with sparse factors L and R), determinant calculation, and matrix-vector products. For instance, complexity of matrix inversion decreases from $\mathcal{O}(n^3)$ for matrices without any structural constraints to around $\mathcal{O}(n^{3/2})$ for sparse matrices. Using Gauss–Markov structures fundamentally shifts the dependence characterization from covariance matrices $\boldsymbol{Q}^{-1}$ to precision matrices $\boldsymbol{Q}$. Notice that the conditional expectation is easily expressed through the regression $\mathbb{E}(x_i \mid \boldsymbol{x}_{-i}) = -\sum_{j \neq i}(q_{ij}/q_{ii})x_j$ where only a small number of the sum terms, also called the neighborhood of x_i, are non-zero owing to the sparse structure of $\boldsymbol{Q}$. The conditional variance is $\mathbb{V}(x_i \mid \boldsymbol{x}_{-i}) = 1/q_{ii}$. Efficient algorithms for sparse matrix calculations (Rue and Held, 2005) are used in `R-INLA`. For fast and memory-saving calculations, it is important to make the precision matrix $\boldsymbol{Q}$ "as diagonal as possible" by reordering variables to regroup nonzero elements as close as possible to the diagonal. `R-INLA` has implemented several of those reordering strategies; see Rue and Held (2005) for more details on reordering algorithms. If certain Gauss–Markov

models exist for spatially indexed graphs (*e.g.*, the classical conditionally autoregressive model of Besag, 1974), useful covariance functions defined over continuous space $\mathbb{R}^d$ and leading to Gauss–Markov covariance matrices are difficult to establish. An exception is the very flexible approximate Gauss–Markov representation of Matérn-like covariances based on certain stochastic partial differential equations – the *SPDE approach* – which is also implemented in `R-INLA`; see the following §A.2.1.

A.2.1 The SPDE approach

The latent variable framework used by INLA allows us to choose the spatial resolution of the latent model separately from the design of observed locations. We can use the results of Lindgren et al. (2011), known as the stochastic partial differential equation (SPDE) approach, to work with numerically convenient Markovian approximations to the Matérn covariance function, leading to sparse precision matrices. This modeling framework is most often used for Gaussian processes in 2D space, but it also works for spline functions defined in 1D space. The SPDE is given by

$$\left(\kappa^2 - \Delta\right)^{\alpha/2} \tau W(s) = B(s), \quad \alpha = \nu + D/2, \quad s \in \mathbb{R}^D, \tag{A.12}$$

with the Laplace operator $\Delta y = \sum_{j=1}^{D} \partial^2 y / \partial^2 x_j$ (involving a so-called *fractional Laplacian* if $\alpha/2 \notin \mathbb{N}$), and a standard Gaussian white noise process $B(s)$. It has a unique stationary solution given by a zero-mean Gaussian process $W(s)$ with Matérn covariance function. Here, ν is the shape parameter of the Matérn, with $\nu = 0.5$ yielding the exponential covariance model. The marginal variance is $\Gamma(\nu)/(\Gamma(\nu+D/2)(4\pi)^{D/2}\kappa^{2\nu}\tau^2)$, and the *empirical range*, where a correlation of approximately 0.1 is attained between two points, is approximately at distance $\sqrt{8\nu}/\kappa^2$. This range parameter is different from the range in the classical Matérn parametrization (Fuglstad et al., 2018).

In practice, the domain is finite, *i.e.*, different from $\mathbb{R}^D$, and appropriate boundary conditions must be imposed to ensure a unique solution. An approximate solution is constructed through the representation $W(s) = \sum_{j=1}^{m} \tilde{W}_j \Psi_j(s)$ with locally supported basis functions $\Psi_j(s)$ (*e.g.*, linear or quadratic B-splines for $D = 1$, and finite elements for $D = 2$). The solution of the SPDE in the subspace spanned by the linear combination $W(s)$ then yields $\tilde{\boldsymbol{W}} = (\tilde{W}_1, \ldots, \tilde{W}_m)^T \sim \mathcal{N}_m(0, Q_W^{-1})$ with precision matrix Q_W known in analytical form. The vector $\tilde{\boldsymbol{W}}$ contains the variables used to represent a single replicate of the Gaussian process.

The value of α determines how the approximate solution of the SPDE is constructed (Lindgren et al., 2011), and it must be fixed when estimating the model with INLA. The INLA implementation currently supports using $\alpha \in [1, 2]$, *i.e.*, $\nu \in [0, 1]$ for $D = 2$.

A.2.2 Space-time Gauss-Markov models

Spatiotemporal modeling is usually based on replicated observations of a spatial process. In this subsection, we detail how to combine Kronecker products of precision matrices to generate the latent variable representations of space-time processes featuring a sparse precision matrix, *e.g.*, space-time processes with first-order temporal autoregression.

In a setting with $\ell \geq 1$ independent and identically distributed copies of spatial Gaussian fields, the joint precision matrix of the ℓ fields can be represented as the Kronecker product

$Q_{ST} = Q_T \otimes I_n$, where I_ℓ is the $\ell \times \ell$ identity matrix. More general time-stationary and temporally dependent sparse precision matrices are possible using the assumption of separable space-time dependence. Given sparse precision matrices Q_S and Q_T representing the purely spatial and purely temporal covariance structure, respectively, the precision matrix for ℓ time steps of the space-time process corresponds to the Kronecker product $Q_{ST} = Q_T \otimes Q_S$. The precision matrix Q_T corresponds to a stationary Gaussian time series (e.g., of a first-order auto-regressive process), assumed to have variance 1 for the sake of identifiability of variance parameters.

With `R-INLA`, the standard choice for modeling spatiotemporal dependence is temporal auto-correlation for Q_T. Using discrete and equidistant time steps, we consider the stationary space-time process, with auto-correlation parameter $\rho \in (-1, 1)$, given as

$$\begin{aligned} W(s,1) &= \varepsilon_1(s), \\ W(s,t+1) &= \rho W(s,t) + \sqrt{1-\rho^2}\varepsilon_{t+1}(s), \quad t = 1, 2, \ldots, \end{aligned} \tag{A.13}$$

where ε_t, $t = 1, 2, \ldots$ are Gaussian random fields with Matérn covariance, and $W(s,t)$ and $\varepsilon(s,t)$ possess the same variance. Then, the space-time precision matrix for the Cartesian product of a collection of sites and times corresponds to the Kronecker product of the corresponding purely spatial Matérn precision matrix Q_W described in §A.2.1, and the purely temporal $\ell \times \ell$ precision matrix Q_{AR1} of a stationary first-order auto-regressive process with marginal variance 1, defined as follows for $\ell \geq 1$ time steps:

$$Q_{\text{AR1}} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & & & & & \\ -\rho & 1+\rho^2 & -\rho & & & & \\ & -\rho & 1+\rho^2 & -\rho & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -\rho & 1+\rho^2 & -\rho & \\ & & & & -\rho & 1 \end{pmatrix}.$$

Using the spatial observation matrices A as introduced in Equation (A.1), the observation matrix A^{ST} for the space-time process observed over the same set of locations at each time step is given by the block-diagonal matrix

$$A^{ST} = \begin{pmatrix} A^S & 0 & \cdots & \cdots & 0 \\ 0 & A^S & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & A^S & 0 \\ 0 & \cdots & \cdots & 0 & A^S \end{pmatrix}$$

with ℓ blocks on the diagonal.

Appendix B

Details for extreme-value analysis

B.1 Elliptical distributions and processes

A random vector $\boldsymbol{X} \in \mathbb{R}^d$ is said to follow an elliptical distribution if it can be written as

$$\boldsymbol{X} = RA\boldsymbol{U} + \boldsymbol{\mu}, \tag{B.1}$$

with R a nonnegative random variable, A a $d \times d$ deterministic nonsingular matrix defining the dispersion matrix $\Sigma = AA^T$ (*i.e.*, a covariance matrix), $\boldsymbol{U}$ a random vector independent of R and distributed uniformly on the Euclidean unit sphere $\{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x}^T\boldsymbol{x} = 1\}$ and $\boldsymbol{\mu} \in \mathbb{R}^d$ a deterministic shift vector. The restriction to nonsingular square matrices A excludes some special cases of minor practical importance and ensures invertibility of Σ. Examples of elliptical distributions are the multivariate Gaussian and the multivariate elliptical t distributions. As an extension of (B.1), a random process $\{X(s)\}$ is called elliptical if all its finite-dimensional distributions are elliptical with the entries of the dispersion matrices Σ defined through a correlation function.

Elliptical distributions can be viewed as a random scaling of a uniform random vector residing on the unit sphere defined with respect to the Mahalanobis norm $\|\boldsymbol{x}\|_\Sigma = \sqrt{\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x}}$. Using elliptic theory (Cambanis et al., 1981), one can write the multivariate density of an elliptical random vector $\boldsymbol{X}$ as

$$g(\boldsymbol{x}) = |\Sigma|^{-1/2} h_d\left(\|\boldsymbol{x}\|_\Sigma^2\right), \qquad x \in \mathbb{R}^d,$$

for some function $h_d : [0,\infty) \to [0,\infty)$. The d-variate Gaussian distribution is characterized by $h_d(t) = (2\pi)^{-d/2}\exp(-t/2)$, $t \geq 0$. A change of variables gives the density f_R of the radial component R as $f_R(r) = A_d r^{d-1} h_d(r^2)$, $r > 0$, where A_d denotes the surface area of the unit ball in $\mathbb{R}^d$, *i.e.*, $A_1 = 1$ and $A_d = 2\pi^{d/2}\{\Gamma(d/2)\}^{-1}$, $d > 1$, with the gamma function $\Gamma(\cdot)$.

B.2 Breiman's lemma

The following Lemma 1 is widely known as *Breiman's lemma* and is useful in many contexts of extreme-value modeling, especially with random scale constructions.

Lemma 1 (Breiman's lemma, see Breiman (1965) and Pakes (2004), Lemma 2.1)**.** *Suppose $X \sim F$, $Y \sim G$ are independent random variables. If $\overline{F} \in RV^{\infty}_{-\alpha}$ with $\alpha \geq 0$ and $Y \geq 0$ with $E(Y^{\alpha+\varepsilon}) < \infty$ for some $\varepsilon > 0$, then*

$$\overline{F}_{XY}(x) \sim E(Y^{\alpha})\ \overline{F}(x), \qquad x \to \infty.$$

Equivalently, if $F \in \mathrm{ET}_{\alpha}$ and $E(e^{(\alpha+\epsilon)Y}) < \infty$, then $\overline{F}_{X+Y}(x) = \overline{F \star G}(x) \sim E\left(e^{\alpha Y}\right)\ \overline{F}(x)$.

B.3 Relationships among univariate tail classes and examples

.

The following lemma summarizes important relationships between the tail classes listed in §2.2.3. We refer to the class of heavy-tailed distributions by HT, and to superheavy-tailed distributions by SHT.

Lemma 2 (Relationships between tail classes)**.** *The following relationships between distribution classes hold:*

1. $\mathrm{RV}^{\infty}_{\alpha} \subset \mathrm{CE}_0$ *for* $\alpha > 0$,
2. $\mathrm{ET}_0 \subsetneq \mathrm{HT}$.
3. *For* ET_{α} *with* $\alpha > 0$, *we have:*
   - $F(\exp(\cdot)) \in \mathrm{ET}_{\alpha} \Leftrightarrow F \in \mathrm{RV}^{\infty}_{\alpha}$,
   - $\mathrm{CE}_{\alpha} \subset \mathrm{ET}_{\alpha}$,
   - $\mathrm{ET}_{\alpha,\beta>-1} \cap \mathrm{CE} = \emptyset$.
4. *For* WT_{β}, *we have:*
   - $\mathrm{WT}_1 \subset \bigcup_{\alpha>0} \mathrm{ET}_{\alpha}$,
   - $\mathrm{WT}_{\beta} \subset \mathrm{CE}_0$ *for* $\beta < 1$,
   - $\mathrm{LWT}_{\beta} \subset \mathrm{SHT}$ *for* $\beta < 1$.
5. *By denoting $F_1 \prec F_2$ if $\overline{F}_1(x)/\overline{F}_2(x) \to 0$ for $x \to \infty$, we have:*
   - *If* $\tilde{\alpha} < \alpha$, *then* $\mathrm{WT}_{\beta>1} \prec \mathrm{ET}_{\alpha} \prec \mathrm{ET}_{\tilde{\alpha}} \prec \mathrm{WT}_{\beta<1} \prec \mathrm{LWT}_{\beta>1} \prec \mathrm{RV}^{\infty}_{\alpha>0} \prec \mathrm{RV}^{\infty}_{\tilde{\alpha}} \prec \mathrm{SHT}$.
   - $\mathrm{CE}_{\alpha>0} \prec \mathrm{ET}_{\tilde{\alpha},\beta}$ *for* $\tilde{\alpha} \leq \alpha$ *and any* $\beta > 0$.

We recall the membership in tail classes for well-known parametric distribution families in Table B.1, see Johnson et al. (1994, 1995) for reference about parameters. Here we abstract away from the usual parameter symbols of these distributions to avoid conflicting notations with general tail parameters. Given a random variable X, we refer parameters as scl and loc if $\mathrm{scl} \times X + \mathrm{loc}$ has scale scl and location loc, where X has scale 1 and location 0. Another parameter shp may be related to shape for some distributions.

	RV_α^0	WT_β	ET_α	HT	LWT_β	$\mathrm{RV}_\alpha^\infty$
normal		$\beta = 2$				
log-normal				✓	$\beta = 2$	
exponential		$\beta = 1$	$\alpha = \mathrm{scl}$			
gamma		$\beta = 1$	$\alpha = \mathrm{scl}$			
inverse normal		$\beta = 1$	$\alpha = \frac{\mathrm{shp}}{2\,\mathrm{mean}^2}$			
logistic		$\beta = 1$	$\alpha = \mathrm{scl}$			
log-logistic		$\beta = 1$		✓		$\alpha = \mathrm{shp}$
Gumbel		$\beta = 1$	$\alpha = \mathrm{scl}$			
Weibull		$\beta = \mathrm{shp}$	$\mathrm{shp} = 1, \alpha = \mathrm{scl}$	$\mathrm{shp} < 1$		
t				✓	$\beta = 1$	$\alpha = \mathrm{shp}$
Pareto				✓	$\beta = 1$	$\alpha = \mathrm{shp}$
Fréchet				✓	$\beta = 1$	$\alpha = \mathrm{shp}$
stable		$\beta = \mathrm{shp} = 2$		$\mathrm{shp} < 2$	$\mathrm{shp} < 2, \beta = 1$	$\alpha = 1/\mathrm{shp} > 1/2$
$F(\mathrm{shp1}, \mathrm{shp2})$				✓		$\alpha = 2/\mathrm{shp2}$
uniform	$\alpha = 1$					
Beta(shp1, shp2)	$\alpha = \mathrm{shp2}$					
triangular	$\alpha = 2$					
GEV	$\alpha = 1/\mathrm{shp} < 0$	$\mathrm{shp} = 0, \beta = 1$	$\mathrm{shp} = 0, \alpha = \mathrm{scl}$	$\mathrm{shp} > 0$	$\mathrm{shp} > 0, \beta = 1$	$\alpha = 1/\mathrm{shp}$

Table B.1: Membership in tail classes (columns) for distribution families (rows). The column RV_α^0 refers to the behavior of $\overline{F}(x^\star - \cdot)$ when $x^\star < \infty$. All heavy-tailed distributions in this table are also subexponential. All distributions in ET_α listed in this table are in $\mathrm{ET}_{\alpha,\beta}$ except for the inverse normal; the inverse normal is in CE_α. The parameter shp of the stable distributions is here chosen as their stability parameter.

Bibliography

Adler, R. J. (1981). *The geometry of random fields*, volume 62. SIAM.

Adler, R. J. (2010). *The geometry of random fields*. SIAM.

Ahmed, M., Maume-Deschamps, V., Ribereau, P., and Vial, C. (2016). Spatial risk measure for Gaussian processes. *arXiv preprint arXiv:1612.08280*.

Ahmed, M., Maume-Deschamps, V., Ribereau, P., and Vial, C. (2020). Spatial risk measures for max-stable and max-mixture processes. *Stochastics*, 92(7):1005–1020.

Albrecht, P. (2014). Risk measures. *Wiley StatsRef: Statistics Reference Online*.

Amaral-Turkman, M. A., Turkman, K. F., Le Page, Y., and Pereira, J. M. C. (2011). Hierarchical space-time models for fire ignition and percentage of land burned by wildfires. *Environmental and Ecological Statistics*, 18(4):601–617.

Arendarczyk, M. and Debicki, K. (2011). Asymptotics of supremum distribution of a Gaussian process over a Weibullian time. *Bernoulli*, 17(1):194–210.

Bacro, J.-N., Gaetan, C., Opitz, T., and Toulemonde, G. (2019). Hierarchical Space-Time Modeling of Asymptotically Independent Exceedances With an Application to Precipitation Data. *Journal of the American Statistical Association*, 115(530):555–569.

Baddeley, A., Berman, M., Fisher, N. I., Hardegen, A., Milne, R. K., Schuhmacher, D., Shah, R., Turner, R., et al. (2010). Spatial logistic regression and change-of-support in Poisson point processes. *Electronic Journal of Statistics*, 4:1151–1201.

Baddeley, A., Coeurjolly, J.-F., Rubak, E., and Waagepetersen, R. (2014). Logistic regression for spatial Gibbs point processes. *Biometrika*, 101(2):377–392.

Baddeley, A., Rubak, E., and Turner, R. (2015). *Spatial point patterns: methodology and applications with R*. CRC press.

Baddeley, A. and Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns. *Australian & New Zealand Journal of Statistics*, 42(3):283–322.

Baddeley, A., Turner, R., Mateu, J., and Bevan, A. (2013). Hybrids of Gibbs point process models and their implementation. *Journal of Statistical Software*, 55(11):1–43.

Barndorff-Nielsen, O. E., Lunde, A., Shephard, N., and Veraart, A. E. D. (2014). Integer-valued Trawl Processes: A Class of Stationary Infinitely Divisible Processes. *Scandinavian Journal of Statistics*, 41(3):693–724.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley.

Belzile, L. R. and Nešlehová, J. G. (2017). Extremal attractors of Liouville copulas. *Journal of Multivariate Analysis*, 160:68–92.

Berman, M. and Turner, T. R. (1992). Approximating point process likelihoods with GLIM. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 41(1):31–38.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225.

Bopp, G. P., Shaby, B. A., and Huser, R. (2020). A hierarchical max-infinitely divisible spatial model for extreme precipitation. *Journal of the American Statistical Association*, pages 1–14.

Bortot, P. and Gaetan, C. (2014). A Latent Process Model for Temporal Extremes. *Scandinavian Journal of Statistics*, 41(3):606–621.

Breiman, L. (1965). On some limit theorems similar to the arc-sin law. *Theory of Probability & Its Applications*, 10(2):323–331.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78:1–3.

Cambanis, S., Huang, S., and Simons, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11(3):368–385.

Cameletti, M., Lindgren, F., Simpson, D., and Rue, H. (2013). Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Advances in Statistical Analysis*, 97(2):109–131.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The Annals of Statistics*, 14(3):1171–1179.

Carreau, J. and Bengio, Y. (2009). A hybrid Pareto model for asymmetric fat-tailed data: the univariate case. *Extremes*, 12(1):53–76.

Castro-Camilo, D. and Huser, R. (2020). Local likelihood estimation of complex tail dependence structures, applied to US precipitation extremes. *Journal of the American Statistical Association*, 115(531):1037–1054.

Castro-Camilo, D., Mhalla, L., and Opitz, T. (2020). Bayesian space-time gap filling for inference on hot spots: an application to Red Sea surface temperatures. *Extremes*, 24:105–128.

Chavez-Demoulin, V. and Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 54(1):207–222.

Chiu, S. N., Stoyan, D., Kendall, W. S., and Mecke, J. (2013). *Stochastic Geometry and Its Applications*. John Wiley & Sons, third edition.

Choiruddin, A., Cuevas-Pacheco, F., Coeurjolly, J.-F., and Waagepetersen, R. (2020). Regularized estimation for highly multivariate log Gaussian Cox processes. *Statistics and Computing*, 30(3):649–662.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer.

Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for extreme value analyses. *Extremes*, 2(4):339–365.

Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the royal statistical society (Series B)*, 53(2):377–392.

Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102(479):824–840.

Cox, D. R. and Isham, V. (1980). *Point processes*, volume 12. CRC Press.

Cui, W. and Perera, A. H. (2008). What do we know about forest fire size distribution, and why is this knowledge useful for forest management? *International Journal of Wildland Fire*, 17(2):234–244.

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes*. Springer.

Davison, A. C., Huser, R., and Thibaud, E. (2019). Spatial extremes. In Gelfand, A. E., Fuentes, M., Hoeting, J. A., and Smith, R. L., editors, *Handbook of Environmental and Ecological Statistics*. CRC press, Boca Raton.

Davison, A. C., Padoan, S., and Ribatet, M. (2012). Statistical Modelling of Spatial Extremes (with Discussion). *Statistical Science*, 27(2):161–186.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 52(3):393–442.

de Fondeville, R. and Davison, A. C. (2018). High-dimensional peaks-over-threshold inference. *Biometrika*, 105(3):575–592.

De Haan, L. (1984). A spectral representation for max-stable processes. *The Annals of Probability*, 12(4):1194–1204.

de Haan, L. and Ferreira, A. (2007). *Extreme Value Theory: An Introduction*. Springer.

De Zea Bermudez, P., Mendes, J., Pereira, J. M., Turkman, K. F., and Vasconcelos, M. J. (2009). Spatial and temporal extremes of wildfire sizes in Portugal (1984–2004). *International Journal of Wildland Fire*, 18(8):983–991.

Delbaen, F. (2002). Coherent risk measures on general probability spaces. In *Advances in finance and stochastics*, pages 1–37. Springer.

Dombry, C., Engelke, S., and Oesting, M. (2016). Exact simulation of max-stable processes. *Biometrika*, 103(2):303–317.

Dombry, C., Éyi-Minko, F., and Ribatet, M. (2013). Conditional simulation of max-stable processes. *Biometrika*, 100(1):111–124.

Dombry, C. and Ribatet, M. (2015). Functional regular variations, Pareto processes and peaks over threshold. *Statistics and Its Interface*, 8(1):9–17.

Dunson, D. B. (2018). Statistics in the big data era: Failures of the machine. *Statistics and Probability Letters*, 136:4–9.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling extremal events*. Springer.

Engelke, S., Opitz, T., and Wadsworth, J. L. (2019). Extremal dependence of random scale constructions. *Extremes*, 22:623—-666.

Fahrmeir, L. and Tutz, G. (2013). *Multivariate statistical modelling based on generalized linear models*. Springer, 2nd edition.

Fargeon, H., Martin-StPaul, N., Pimont, F., De Cáceres, M., Ruffault, J., Opitz, T., Allard, D., and Dupuy, J.-L. (2018). Assessing the increase in wildfire occurrence with climate change and the uncertainties associated with this projection. In *8th International conference on forest fire research*.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, 27(8):861–874.

Ferreira, A. and De Haan, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli*, 20(4):1717–1737.

Fougeres, A.-L. and Mercadier, C. (2012). Risk measures and multivariate extensions of Breiman's theorem. *Journal of Applied Probability*, 49(2):364–384.

Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. *Annales de la Société Polonaise de Mathématiques*, 6:93–116.

Frigessi, A., Haug, O., and Rue, H. (2003). A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes*, 5(3):219–235.

Fuglstad, G.-A., Simpson, D., Lindgren, F., and Rue, H. (2018). Constructing priors that penalize the complexity of Gaussian random fields. *Journal of the American Statistical Association*, 114(525):445–452.

Gabriel, E., Opitz, T., and Bonneu, F. (2017). Detecting and modeling multi-scale space-time structures: the case of wildfire occurrences. *Journal of the French Statistical Society (Special Issue on Space-Time Statistics)*, 158(3).

Gelfand, A. E. and Shirota, S. (2019). Preferential sampling for presence/absence data and for fusion of presence/absence data with presence-only data. *Ecological Monographs*, 89(3):e01372.

Gelman, A. (2008). Objections to Bayesian statistics. *Bayesian Analysis*, 3(3):445–449.

Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and computing*, 24(6):997–1016.

Genz, A. and Bretz, F. (2009). *Computation of multivariate normal and t probabilities*. Springer.

Giglio, L., Randerson, J. T., and Van Der Werf, G. R. (2013). Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database gfed4. *Journal of Geophysical Research: Biogeosciences*, 118(1):317–328.

Giné, E., Hahn, M. G., and Vatan, P. (1990). Max-infinitely divisible and max-stable sample continuous processes. *Probability theory and related fields*, 87(2):139–165.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of mathematics*, pages 423–453.

Gneiting, T. and Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1:125–151.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378.

Guan, Y. (2006). A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association*, 101(476):1502–1512.

Hall, W. and Wellner, J. A. (2020). Estimation of mean residual life. In *Statistical modeling for biological systems*, pages 169–189. Springer.

Hashorva, E. and Hüsler, J. (2003). On multivariate Gaussian tails. *Annals of the Institute of Statistical Mathematics*, 55(3):507–522.

Hastie, T. and Tibshirani, R. (1990). *Generalized additive models*. CRC Press.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):497–546.

Horvitz, D. G. and Thompson, D. J. (1952). A Generalization of Sampling Without Replacement from a Finite Universe. *Journal of the American Statistical Association*, 47(260):663–685.

Huang, S. T. and Cambanis, S. (1979). Spherically invariant processes: Their nonlinear structure, discrimination, and estimation. *Journal of Multivariate Analysis*, 9(1):59–83.

Hult, H. and Lindskog, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied probability*, 34(3):587–608.

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.

Huser, R., Opitz, T., and Thibaud, E. (2021). Max-infinitely divisible models and inference for spatial extremes. *Scandinavian Journal of Statistics*, 48(1):321–348.

Huser, R. and Wadsworth, J. L. (2019). Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, 114:434–444.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*, volume 70. John Wiley & Sons.

Illian, J. B., Sørbye, S. H., and Rue, H. (2012). A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *The Annals of Applied Statistics*, pages 1499–1530.

Joe, H. (2014). *Dependence modeling with copulas*. CRC Press.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous univariate distributions. Vol. 1. 2nd ed.* Chichester: Wiley, 2nd ed. edition.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). *Continuous univariate distributions. Vol. 2. 2nd ed.* New York, NY: Wiley, 2nd ed. edition.

Jones, M. W., Smith, A., Betts, R., Canadell, J. G., Prentice, I. C., and Le Quéré, C. (2020). Climate change increases risk of wildfires. *ScienceBrief Review*.

Kabluchko, Z., Schlather, M., and De Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, 37(5):2042–2065.

Kansanen, K. (2020). *Horvitz–Thompson-like estimators based on stochastic geometry for forest remote sensing*. PhD thesis, Faculty of Science and Forestry, School of Computing, University of Eastern Finland.

Klüppelberg, C. and Kuhn, G. (2009). Copula structure analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):737–753.

Klüppelberg, C., Kuhn, G., and Peng, L. (2008). Semi-parametric models for the multivariate tail dependence function–the asymptotically dependent case. *Scandinavian Journal of Statistics*, 35(4):701–718.

Klüppelberg, C. and Resnick, S. I. (2008). The Pareto copula, aggregation of risks, and the emperor's socks. *Journal of Applied Probability*, 45(1):67–84.

Koch, E. (2017). Spatial risk measures and applications to max-stable processes. *Extremes*, 20(3):635–670.

Kotz, S., Balakrishnan, N., and Johnson, N. L. (2005). *Continuous Multivariate Distributions, Models and Applications: Second Edition*, volume 1. Jon Wiley & Sons.

Krainski, E. T., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., and Rue, H. (2018). *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*. Chapman and Hall/CRC.

Krupskii, P., Huser, R., and Genton, M. G. (2018). Factor copula models for replicated spatial data. *Journal of the American Statistical Association*, 113(521):467–479.

Lalancette, M., Engelke, S., and Volgushev, S. (2020). Rank-based estimation under asymptotic dependence and independence, with applications to spatial extremes. *arXiv preprint arXiv:2008.03349*.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.

Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 59(2):475–499.

Leininger, T. J., Gelfand, A. E., et al. (2017). Bayesian inference and model assessment for spatial point patterns using posterior predictive samples. *Bayesian Analysis*, 12(1):1–30.

Leonard, M., Westra, S., Phatak, A., Lambert, M., van den Hurk, B., McInnes, K., Risbey, J., Schuster, S., Jakob, D., and Stafford-Smith, M. (2014). A compound event framework for understanding extreme impacts. *Wiley Interdisciplinary Reviews: Climate Change*, 5(1):113–128.

Lindgren, F. and Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63(19).

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary mathematics*, 80(1):221–239.

Loader, C. (2006). *Local regression and likelihood*. Springer Science & Business Media.

Lombardo, L., Opitz, T., Ardizzone, F., Guzzetti, F., and Huser, R. (2020). Space-Time Landslide Predictive Modelling. *Earth Science Reviews*, 209.

Lombardo, L., Opitz, T., and Huser, R. (2018). Point process-based modeling of multiple debris flow landslides using INLA: an application to the 2009 Messina disaster. *Stochastic environmental research and risk assessment*, 32(7):2179–2198.

Lombardo, L., Opitz, T., and Huser, R. (2019). Numerical Recipes for Landslide Spatial Prediction Using R-INLA: A Step-by-Step Tutorial. In *Spatial Modeling in GIS and R for Earth and Environmental Sciences*, pages 55–83. Elsevier.

Mariethoz, G. and Caers, J. (2014). *Multiple-point geostatistics: stochastic modeling with training images*. John Wiley & Sons.

Mariethoz, G., McCabe, M. F., and Renard, P. (2012). Spatiotemporal reconstruction of gaps in multivariate fields using the direct sampling approach. *Water Resources Research*, 48(10).

Mariethoz, G., Renard, P., and Straubhaar, J. (2010). The Direct Sampling method to perform multiple-point geostatistical simulations. *Water Resources Research*, 46(11).

Mhalla, L., Opitz, T., and Chavez-Demoulin, V. (2019). Exceedance-based nonlinear regression of tail dependence. *Extremes*, 22(3):523–552.

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian journal of statistics*, 25(3):451–482.

Moller, J. and Waagepetersen, R. P. (2003). *Statistical inference and for spatial point processes*. CRC Press.

Mornet, A., Opitz, T., Luzi, M., and Loisel, S. (2015). Index for Predicting Insurance Claims from Wind Storms with an Application in France. *Risk Analysis*, 35(11):2029–2056.

Mornet, A., Opitz, T., Luzi, M., Loisel, S., and Bailleul, B. (2017). Wind storm risk management: sensitivity of return period calculations and spread on the territory. *Stochastic Environmental Research and Risk Assessment*, 31(8):1977–1995.

Naveau, P., Huser, R., Ribereau, P., and Hannart, A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*, 52(4):2753–2769.

Naveh, Z. and Lieberman, A. S. (2013). *Landscape ecology: theory and application*. Springer Science & Business Media.

Nikoloulopoulos, A. K., Joe, H., and Li, H. (2009). Extreme value properties of multivariate t copulas. *Extremes*, 12(2):129–148.

Northrop, P. J. and Jonathan, P. (2011). Threshold modelling of spatially-dependent non-stationary extremes with application to hurricane-induced wave heights (with discussion). *Environmetrics*, 22:799–809.

Nzali, M. D. T., Bringay, S., Lavergne, C., Mollevi, C., and Opitz, T. (2017). What patients can tell us: topic analysis for social media on breast cancer. *JMIR Medical Informatics*, 5(3).

Opitz, T. (2013). Extremal t processes: Elliptical domain of attraction and a spectral representation. *Journal of Multivariate Analysis*, 122:409–413.

Opitz, T. (2016). Modeling asymptotically independent spatial extremes based on Laplace random fields. *Spatial Statistics*, 16:1–18.

Opitz, T. (2017). Latent Gaussian modeling and INLA: A review with focus on space-time applications. *Journal of the French Statistical Society (Special Issue on Space-Time Statistics)*, 158(3).

Opitz, T., Allard, D., and Mariethoz, G. (2021). Semi-parametric resampling with extremes. *Spatial Statistics*, 42:100445.

Opitz, T., Azé, J., Bringay, S., Joutard, C., Lavergne, C., and Mollevi, C. (2014). Breast cancer and quality of life: medical information extraction from health forums. In *Medical Informatics Europe Conference 2014*, pages 1070–1074.

Opitz, T., Bacro, J.-N., and Ribereau, P. (2015). The spectrogram: A threshold-based inferential tool for extremes of stochastic processes. *Electronic Journal of Statistics*, 9(1):842–868.

Opitz, T., Bakka, H., Huser, R., and Lombardo, L. (2020a). High-resolution Bayesian mapping of landslide hazard with unobserved trigger event. *arXiv preprint arXiv:2006.07902*.

Opitz, T., Bonneu, F., and Gabriel, E. (2020b). Point-process based modeling of space-time structures of forest fire occurrences in Mediterranean France. *Spatial Statistics*, 40:100429.

Opitz, T., Huser, R., Bakka, H., and Rue, H. (2018). INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles. *Extremes*, 21(3):441–462.

Padoan, S. A. (2013). Extreme dependence models based on event magnitude. *Journal of Multivariate Analysis*, 122:1–19.

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.

Pakes, A. G. (2004). Convolution equivalence and infinite divisibility. *Journal of Applied Probability*, 41(02):407–424.

Palacios-Rodriguez, F., Toulemonde, G., Carreau, J., and Opitz, T. (2020). Generalized Pareto processes for simulating space-time extreme events: an application to precipitation reanalyses. *Stochastic Environmental Research and Risk Assessment*, 34:2033—-2052.

Papastathopoulos, I. and Tawn, J. A. (2013). Extended generalised Pareto models for tail estimation. *Journal of Statistical Planning and Inference*, 143(1):131–143.

Pennino, M. G., Paradinas, I., Illian, J. B., Muñoz, F., Bellido, J. M., López-Quílez, A., and Conesa, D. (2019). Accounting for preferential sampling in species distribution models. *Ecology and Evolution*, 9(1):653–663.

Pereira, J. M. C. and Turkman, K. F. (2019). Statistical models of vegetation fires: Spatial and temporal patterns. In *Handbook of Environmental and Ecological Statistics*, pages 401–420. Chapman and Hall/CRC.

Pimont, F., Fargeon, H., Opitz, T., Ruffault, J., Barbero, R., Martin-StPaul, N., Rigolot, E. I., Rivière, M., and Dupuy, J.-L. (2021). Prediction of regional wildfire activity with a probabilistic Bayesian framework. *Ecological applications (In press)*.

Preisler, H. K., Brillinger, D. R., Burgan, R. E., and Benoit, J. (2004). Probability based models for estimation of wildfire risk. *International Journal of wildland fire*, 13(2):133–142.

Rathbun, S. L. (2013). Optimal estimation of Poisson intensity with partially observed covariates. *Biometrika*, 100(1):277–281.

Rathbun, S. L., Shiffman, S., and Gwaltney, C. J. (2007). Modelling the effects of partially observed covariates on Poisson process intensity. *Biometrika*.

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., and Guzzetti, F. (2018). A review of statistically–based landslide susceptibility models. *Earth-Science Reviews*, 180:60–91.

Reiss, R.-D. and Thomas, M. (2007). *Statistical Analysis of Extreme Values*. Birkhäuser, Basel, 3rd edition.

Resnick, S. I. (1987). *Extreme values, regular variation and point processes*. Springer.

RESSTE Network (2017). Analyzing spatio-temporal data with R: Everything you always wanted to know – but were afraid to ask. *Journal of The French Statistical Society (Special Issue on Space-Time Statistics)*, 158(3).

Ribatet, M. (2013). Spatial extremes: Max-stable processes at work. *Journal of Société Française de Statistique (Special edition on extreme value theory)*, 154(2):156–177.

Rootzén, H., Segers, J., and L. Wadsworth, J. (2018). Multivariate peaks over thresholds models. *Extremes*, 21(1):115–145.

Rootzén, H., Segers, J., and Wadsworth, J. L. (2018). Multivariate generalized Pareto distributions: Parametrizations, representations, and properties. *Journal of Multivariate Analysis*, 165:117–131.

Rootzén, H. and Tajvidi, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli*, 12(5):917–930.

Rue, H. (2005). Marginal variances for Gaussian Markov random fields. Technical report, NTNU.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC Press.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society (Series B)*, 71(2):319–392.

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., and Lindgren, F. K. (2017). Bayesian computing with INLA: a review. *Annual Review of Statistics and Its Application*, 4:395–421.

Saumard, A. and Wellner, J. A. (2014). Log-concavity and strong log-concavity: a review. *Statistics Surveys*, 8:45–114.

Scarrott, C. and MacDonald, A. (2012). A Review of Extreme Value Threshold Estimation And Uncertainty Quantification. *REVSTAT*, 10(1):33–60.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, 5(1):33–44.

Schlather, M. and Tawn, J. A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, 90(1):139–156.

Simpson, D., Rue, H., Riebler, A., Martins, T. G., Sørbye, S. H., and Others (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, 32(1):1–28.

Simpson, E., Opitz, T., and Wadsworth, J. L. (2020). High-dimensional modeling of spatial and spatio-temporal conditional extremes using INL and the SPDE approach. *arXiv preprint arXiv:2011.04486*.

Simpson, E. S. and Wadsworth, J. L. (2020). Conditional modelling of spatio-temporal extremes for Red Sea surface temperatures. *arXiv preprint arXiv:2002.04362*.

Smith, R. L. (1990). Max-stable processes and spatial extremes. Technical report.

Soma, M., Pimont, F., Allard, D., Fournier, R., and Dupuy, J.-L. (2020). Mitigating occlusion effects in Leaf Area Density estimates from Terrestrial LiDAR through a specific kriging method. *Remote Sensing of Environment*, 245:111836.

Soubeyrand, S. and Haon-Lasportes, E. (2015). Weak convergence of posteriors conditional on maximum pseudo-likelihood estimates and implications in ABC. *Statistics & Probability Letters*, 107:84–92.

Stewart-Oaten, A., Murdoch, W. W., and Parker, K. R. (1986). Environmental impact assessment: "pseudoreplication" in time? *Ecology*, 67(4):929–940.

Tahmasebi, P. (2018). Multiple point statistics: A review. In *Handbook of Mathematical Geosciences: Fifty Years of IAMG*, pages 613–643. Springer International Publishing.

Taillardat, M., Fougères, A.-L., Naveau, P., and de Fondeville, R. (2019). Extreme events evaluation using CRPS distributions. *arXiv preprint arXiv:1905.04022*.

Tapi Nzali, M. D., Aze, J., Bringay, S., Lavergne, C., Mollevi, C., and Optiz, T. (2019). Reconciliation of patient/doctor vocabulary in a structured resource. *Health Informatics journal*, 25(4):1219–1231.

Tapi Nzali, M. D., Bringay, S., Lavergne, C., Opitz, T., Azé, J., and Mollevi, C. (2015). Construction d'un vocabulaire patient/médecin dédié au cancer du sein à partir des médias sociaux. *26. Journées Francophones d'Ingénierie des Connaissances (IC), Rennes*.

Tawn, J., Shooter, R., Towe, R., and Lamb, R. (2018). Modelling spatial extreme events with environmental applications. *Spatial Statistics*, 28:39–58.

Taylor, B. M. and Diggle, P. J. (2014). INLA or MCMC? A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes. *Journal of Statistical Computation and Simulation*, 84(10):2266–2284.

Thibaud, E. and Opitz, T. (2015). Efficient inference and simulation for elliptical Pareto processes. *Biometrika*, 102(4):855–870.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography*, 46:234–240.

Tokdar, S. T. and Kass, R. E. (2010). Importance sampling: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(1):54–60.

Turkman, K. F., Amaral Turkman, M. A., and Pereira, J. M. (2010). Asymptotic models and inference for extremes of spatio-temporal data. *Extremes*, 13(4):375–397.

van Niekerk, J., Bakka, H., Rue, H., and Schenk, L. (2019). New frontiers in Bayesian modeling using the INLA package in R. *arXiv preprint arXiv:1907.10426*.

van Wagner, C. (1977). Conditions for the start and spread of crown fire. *Canadian Journal of Forest Research*, 7(1):23–34.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42.

Waagepetersen, R. and Guan, Y. (2009). Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):685–702.

Wadsworth, J., Tawn, J., et al. (2013). A new representation for multivariate tail probabilities. *Bernoulli*, 19(5B):2689–2714.

Wadsworth, J. L. and Tawn, J. (2019). Higher-dimensional spatial extremes via single-site conditioning. *arXiv preprint arXiv:1912.06560*.

Wadsworth, J. L. and Tawn, J. A. (2012). Dependence modelling for spatial extremes. *Biometrika*, 99(2):253–272.

Wadsworth, J. L. and Tawn, J. A. (2014). Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika*, 101(1):1–15.

Wadsworth, J. L., Tawn, J. A., Davison, A. C., and Elton, D. M. (2017). Modelling across extremal dependence classes. *Journal of the Royal Statistical Society (Series B)*, 79(1):149–175.

Wickham, H. and Grolemund, G. (2017). *R for Data Science: Visualize, Model, Transform, Tidy, and Import Data*. O'Reilly Media.

Wolpert, R. L. and Ickstadt, K. (1998). Simulation of Lévy random fields. *Practical nonparametric and semiparametric Bayesian statistics*, pages 227–242.

Xi, D. D., Taylor, S. W., Woolford, D. G., and Dean, C. (2019). Statistical models of key components of wildfire risk. *Annual review of statistics and its application*, 6:197–222.

Yadav, R., Opitz, T., and Huser, R. (2020). Spatial hierarchical modeling of threshold exceedances using rate mixtures. *Environmetrics (In press)*.

Youngman, B. D. (2019). Generalized additive models for exceedances of high thresholds with an application to return level estimation for US wind gusts. *Journal of the American Statistical Association*, 114(528):1865–1879.

Zamberletti, P., Papaïx, J., Gabriel, E., and Opitz, T. (2021a). Landscape allocation: stochastic generators and statistical inference. *Annals of Applied Statistics (In press)*.

Zamberletti, P., Papaïx, J., Gabriel, É., and Opitz, T. (2021b). Spatio-temporal point processes as meta-models for population dynamics in heterogeneous landscapes. *(Submitted)*.

Zamberletti, P., Sabir, K., Opitz, T., Bonnefon, O., Gabriel, É., and Papaïx, J. (2021c). More pests but less treatments: ambivalent effect of landscape complexity on Conservation Biological Control. *(Submitted)*.

Zhong, P., Huser, R., and Opitz, T. (2020). Assessing Non-Stationary Heatwave Hazard with Magnitude-Dependent Spatial Extremal Dependence. *arXiv preprint arXiv:2006.01569*.

Zhong, P., Huser, R., and Opitz, T. (2021). Exact simulation of max-infinitely divisible processes. *arXiv preprint arXiv:2103.00533*.

Zscheischler, J., Naveau, P., Martius, O., Engelke, S., and Raible, C. C. (2021). Evaluating the dependence structure of compound precipitation and wind speed extremes. *Earth System Dynamics*, 12(1):1–16.

Zscheischler, J. and Seneviratne, S. I. (2017). Dependence of drivers affects risks associated with compound events. *Science advances*, 3(6):e1700263.

Zscheischler, J., Westra, S., Van Den Hurk, B. J., Seneviratne, S. I., Ward, P. J., Pitman, A., AghaKouchak, A., Bresch, D. N., Leonard, M., Wahl, T., et al. (2018). Future climate risk from compound events. *Nature Climate Change*, 8(6):469–477.