Graphical Modeling of Spatially Distributed Extreme Events

Master Thesis

In partial fulfillment of the requirements for the degree

Master of Science in Statistics

Humboldt University Berlin

Supervised by:
Dr. Matthias Eckardt
Prof. Dr. Jan Marcus

Submitted by:
Marie Isabelle Bernière

Student ID:
620758

Berlin, 27.01.2025

Abstract

Motivated by the recent definition and research around graphical models for extremes (Engelke and Hitz 2020), this thesis explored its applicability to spatial point processes. In particular, some extreme events can be represented as random points across space and time. These can be either seen as extreme in magnitude or extreme in the sense of rare in time. Graphical models on their end, are used in multivariate statistics to visually represent conditional dependencies. Using the example of wildfires with associated ignition day and burned land, this work presents the existing graphical modeling approach defined for threshold exceedances (Engelke and Hitz 2020), discusses its applicability for spatial point processes and proposes an extension of this model incorporating a distance scaling. A simulation study confirms the utility of this approach in depicting spatial dependencies in the burned land from wildfires. However, an application to real-life wildfire data highlighted some challenges. The limited data can lead to unreliable extrapolations, resulting in a potential overestimation of the dependencies. Despite these issues, the findings contribute to the understanding of spatial extremes and offer insights and ideas for improving dependency modeling. This work can also motivate further development of methodologies that integrate spatial information in the analysis of spatial extremes.

Contents

List of Abbreviations iv

List of Figures v

1. Introduction 1
- 1.1. Motivation . . . 1
- 1.2. Literature Review . . . 2

2. Theoretical Foundations 6
- 2.1. Spatial Point Processes . . . 6
- 2.2. Extreme Value Theory . . . 10
- 2.3. Graphical Models . . . 16

3. Extending Graphical Models for Spatial Extremes 24
- 3.1. Hüsler-Reiss Graphical Models . . . 25
- 3.2. Proposed Methodology and Extension . . . 28
- 3.3. Advantages and Challenges . . . 31
- 3.4. Estimation . . . 33

4. Simulation Exercise 37
- 4.1. General Set Up . . . 37
- 4.2. Simulation Scenarios . . . 40
- 4.3. Results . . . 41
- 4.4. Discussion . . . 48

5. Application - Forest Fires in Spain 50
- 5.1. Data Description . . . 50
- 5.2. Preliminary Analysis . . . 51
- 5.3. Graphical Modeling . . . 56

6. Concluding Remarks and Outlook 60

A. Simulation Set-Up 62

B. Additional Simulation Results 63

C. Additional Results Fire Data 68

D. List of R Packages used **71**

Bibliography **72**

Declaration of Academic Honesty **75**

List of Abbreviations

AIC	Akaike Information Criteria
BIC	Bayesian Information Criteria
EVT	Extreme Value Theory
FWI	Fire Weather Index
GEV	Generalized Extreme Value
glasso	Graphical Lasso
GPD	Generalized Pareto Distribution
ha	hectares
HR	Hüsler-Reiss
NS	Neighbor Selection

List of Figures

2.1. Single realizations of a spatial Poisson process in a unit square with intensity (a) 5, (b) 15 and (c) 30. R Package used: *Spatstat* (Baddeley and Turner 2005) . . 9
2.2. Visualization of the density function of the Generalized Pareto Distribution for (a) varying shape and fixed scale and (b) varying scale and fixed shape. `R` Package used: `evd` (A. G. Stephenson 2002) 13
2.3. Examples of undirected graphs with four nodes (examples taken from Fig 1 in Engelke, Lalancette, and Volgushev (2021)). (a) is an example of a tree graph, (b) of a decomposable graph and (c) of a non-decomposable graph 18

4.1. Histogram of n=10000 sample from a GPD($\mu = 1, \sigma = 1, \xi = 0.5$). `R` Package used: `evd` package by A. G. Stephenson (2002). 39
4.2. Scenario 1 - Mean values per location for (a) whole grid and (b) a grid subset. . 40
4.3. Scenario 2 - Mean values per location for (a) whole grid and (b) a grid subset. . 41
4.4. Scenario 3 - Mean values per location 41
4.5. Estimated HR graphical model for scenario 1 using `eglearn` algorithm in (a) (Engelke, Lalancette, and Volgushev 2021) and a scaled version in (b) using the maximum distance of the estimates obtained via `Kest()` function (Baddeley and Turner 2005). The graph in (a) minimizes BIC, using a 90% quantile threshold and the neighbor selection based algorithm. The $\hat{\chi_{ij}}$ are the estimated extremal correlations between two nodes. 44
4.6. Portion of the 29 locations in Scenario 1 with strictly positive quantile value for different levels. 45
4.7. Estimated HR graphical model for scenario 2 as detailed in Figure 4.5. 45
4.8. Estimated HR graphical model for scenario 3 as detailed in Figure 4.5. In addition, the red points indicate the locations of the trigger locations. 47

5.1. Marked point process `clmfires` for (a) fire cause and (b) burnt area (log scale). 51
5.2. Marked point process `clmfires` for burnt land (a) below 1 ha and (b) below 10ha. 52
5.3. Histogram of burned area, capped at different values. 52
5.4. Probability (a) and Quantile (b) plot for a fitted GPD using the 60% quantile value (1ha) as threshold. 53
5.5. Probability (a) and Quantile (b) plot for a fitted GPD using the 95% quantile value (26ha) as threshold. 53
5.6. Fitted return periods for daily aggregates of burned hectares of land over the region of Castilla-La Mancha - based on GPD fit with threshold of 138.6 ha. . . 54

5.7. Fire Activity over the years (a) and over the month (b). 55
5.8. Estimate of the K function against true K values of a random Poisson point process. The 'best' fit is chosen, which corresponds to Ripley's isotropic edge correction (see Baddeley and Turner (2005) for more details) 56
5.9. HR graphical model (a) and scaled version (b) with extremal correlation estimates between locations connected by an edge 57
5.10. Scaled HR graphical model, mapped to nearest city, with extremal correlation estimates between locations connected by an edge 58

A.1. Overview of the simulation set-up 62

B.1. Penalization parameters against BIC of the corresponding estimated graphs - Used are, a 90% quantile threshold and the NS based algorithm of `eglearn` (Engelke, Lalancette, and Volgushev 2021). Package used: `graphicalExtremes` (Engelke, Hitz, et al. 2024) 63
B.2. Scenario 1 - Estimated HR graphical model using `eglearn` algorithm in (a) (Engelke, Lalancette, and Volgushev 2021) and a scaled version in (b) using the maximum distance of the estimates obtained via `Kest()` function (Baddeley and Turner 2005). The graph in (a) minimizes BIC, using a 90% quantile threshold and the neighbor selection based algorithm. 64
B.3. Scenario 1 - Estimated HR graphical model using `eglearn` algorithm in (a) (Engelke, Lalancette, and Volgushev 2021) and a scaled version in (b). The graph in (a) minimizes BIC, using a 90% quantile threshold and the glasso based algorithm. 64
B.4. Scenario 1 - Estimated HR graphical model using `eglearn` algorithm in (a) (Engelke, Lalancette, and Volgushev 2021) and a scaled version in (b). The graph in (a) minimizes BIC, using a 70% quantile threshold and the NS based algorithm. 65
B.5. Scenario 2 - Estimated HR graphical model (a) and its scaled version (b) as described in Figure B.2. 65
B.6. Scenario 2 - Estimated HR graphical model using `eglearn` algorithm in (a) (Engelke, Lalancette, and Volgushev 2021) and a scaled version in (b). The graph in (a) minimizes BIC, using a 90% quantile threshold and the glasso based algorithm. 66

B.7. Scenario 2 - Estimated HR graphical model using `eglearn` algorithm in (a) (Engelke, Lalancette, and Volgushev 2021) and a scaled version in (b). The graph in (a) minimizes BIC, using a 70% quantile threshold and the NS based algorithm. 66
B.8. Scenario 3 - Estimated HR graphical model (a) and its scaled version (b) as described in Figure B.2 with trigger locations marked in red. 67

C.1. 200 pixel images of (a) elevation, (b) slope and (c) landuse of the administrative region of Castilla-La Mancha (white outline). Images available in R Package *spatstat* under `clmfires.extra`. 68
C.2. Diagnostic plots for the GPD fit on daily aggregated burned land in the region Castilla La Mancha, using the 95% quantile value (138.6 ha) as a threshold. . . 69
C.3. Subset of points in the northern region of Castilla-La Mancha. 69
C.4. Number of months with fires in the spatially aggregated locations in the northern region. 70
C.5. Scatterplot of monthly area burned between two connected locations. 70

1. Introduction

1.1. Motivation

The study of extreme events has become increasingly important in the context of climate change. In fact, natural disasters like earthquakes, floods and windstorms are increasing in Europe and can have significant impacts on the economy, communities, and the environment (Beniston et al. 2007). As a consequence from long droughts and heatwaves, other events like wildfires are also growing in frequency. For example, the annual area of forest burned in France has largely increased from just over 1,500 hectares in 2008 to more than 20,000 hectares in 2023 (Statista.com)[1]. This rising trend is not specific to France, as wildfires in southern Europe are expected to increase further in the coming years (Dupuy et al. 2020).

A branch of statistics, known as Extreme Value Theory, focuses on analyzing the behavior of data in the tails of its distribution, representing the upper or lower extremes, depending on the context (Coles et al. 2001; Haan and Ferreira 2006). Unlike classical statistics, which typically focuses on the average behavior, extreme value theory is specifically suited to understanding rare events. Beyond the analysis of single variables (e.g. extreme temperatures), accurately characterizing extremal dependencies between climate variables is of particular interest (Davison and Huser 2015). In addition to dependencies between climate variables, extreme events can exhibit spatial dependencies. For example, a wildfire in one part of the forest could be predictive of a fire in another part due to a shared environment such as rainfall exposure, slope or vegetation type. Thus, there is an interest in characterizing the extremal dependence not only between climate variables but also across a geographic space. Some spatial events can exhibit point process-like characteristics, meaning that they occur sporadically in time and space (Cressie 2015; Baddeley, Bárány, and Schneider 2007). The question is then, how does the presence of an event at one location, extreme in the sense of rare in time and/or magnitude, relate to the presence of an event in another geographic location.

These considerations form the basis of this thesis, which focuses on modeling spatial dependencies of moderate and extreme events using techniques from extreme value theory and graphical modeling. Graphical models are widely used in multivariate statistics due to their ability to visually represent conditional dependency structures in an interpretable and sparse way (Lauritzen 1996). More specifically, they represent conditional dependence relationships between random

[1] https://fr.statista.com/statistiques/1321001/hectares-forets-brules-france/

variables using a graph. The nodes in the graph represent variables (e.g. locations of interest) and are connected using edges indicating conditional dependencies. A further motivation for the use of graphical models for the modeling of spatial extreme events is their recent adaptation to the context of extremes (Engelke and Hitz 2020) and their application and extension in spatial contexts (Cisneros, Hazra, and Huser 2024). However, their application to spatial point processes, where events might occur sporadically across time and space, remains to be further explored.

By combining the principles of extreme value theory with graphical modeling to spatial point processes, this thesis aims to develop a framework and methodology for identifying dependencies in spatial extreme events. Wildfires and their associated burned land are used as an example throughout this thesis, but the results should not be restricted to that case. More specifically, it addresses the following research questions.

- Can a graphical modeling approach for extreme values as in Engelke and Hitz (2020) effectively capture the spatial dependency structures in extreme events represented as spatial point processes ?
- Can an extension of these models, incorporating a spatial constrain, improve the dependency modeling ?

The results of this study could have implications in several areas. By providing an example of how extreme events as spatial point processes can be modeled using graphical models, this research can contribute to further research ideas in this field. For fire management communities, the discovery of conditional dependency structures between different areas of a forest can help improve resource allocation during fire seasons.

1.2. Literature Review

Wildfires present complex challenges for modeling and prediction and their associated risks have previously been studied under different angles and methodologies (see Dupuy et al. (2020) for a review). One approach leverages meteorological information and studies their impact on fire activity. Weather variables (e.g. precipitation, temperature) are combined to a fire index to establish potential fire weather danger (Dowdy et al. 2009; Cisneros, Hazra, and Huser 2024). The Fire Weather Index (FWI) system is the one used worldwide and studies the impact of meteorological variables on fuel moisture (Wagner (1987) cited in Dupuy et al. (2020)). Fire activity can then be modeled using such index via correlation methods (Amatulli, Camia, and San-Miguel-Ayanz (2013) and Turco et al. (2014) cited in Dupuy et al. (2020)). Another approach

in modeling wildfires discussed by Dupuy et al. (2020), focuses explicitly on capturing fire spread dynamics. They discuss studies that incorporate wind patterns and other meteorological conditions to simulate fire propagation. These models aim to provide a detailed understanding of how environmental factors influence fire behavior.

Spatial point processes provide a natural framework for studying wildfire occurrences and sizes by modeling them as discrete events distributed across a continuous spatial domain. For example, Genton et al. (2006) and Hering, Bell, and Genton (2009) used spatial point process models to analyze wildfire occurrences and their deviation from complete spatial randomness in Florida, highlighting the importance of clustering patterns in understanding fire behavior. Koh et al. (2023) further model, not only the occurrence of fires but also their burned area using spatial point process and extreme value methods. In some cases, wildfire event data in point process form is aggregated using larger grid resolution. For example, Serra et al. (2012) and Bivand, Gómez-Rubio, and Rue (2015) explored this approach by aggregating wildfire occurrences in defined grid cells. Cisneros, Richards, et al. (2024) further constructed an areal data set using spatial and temporal aggregation of fire burned area in Australia. Although they acknowledge the potential loss of spatial detail, they point out the advantage of such dataset as it represents regular temporal intervals and consistent spatial locations.

Extreme Value Theory (EVT) provides a robust mathematical framework for analyzing rare and extreme events, which are crucial for understanding environmental risks such as wildfires. As outlined by Coles et al. (2001), EVT addresses the limitations of the scarcity of observed extreme values by offering theoretical tools to extrapolate beyond the observed data. Two primary methodologies are prevalent in EVT: block maxima methods (modeling maximum value within a block of values) and threshold exceedance models (modeling values above a set threshold). Additionally, multivariate EVT extends these concepts to model dependencies between extremes, which is particularly relevant for spatial applications (Davison and Huser 2015). EVT has also been adapted to spatial settings using max-stable and latent spatial processes, which can allow the estimation of dependencies between extreme values at different locations (Coles et al. 2001).

However, there are recent discussions about the appropriate use of max-stable processes in environmental contexts. In particular, Huser, Opitz, and J. Wadsworth (2024) criticize the widespread reliance on these processes and advocate for more flexible approaches using threshold exceedances in environmental applications. Another discussed topic is that the underlying assumption of asymptotic dependence might not be suitable (Davison and Huser 2015; Huser and J. L. Wadsworth 2019; Huser, Opitz, and J. Wadsworth 2024) and can lead to inaccurate

modeling. Another limitation highlighted by Huser and J. L. Wadsworth (2022) is the computational difficulty of some models. They further discuss recent advances in this area and address the limitations of existing models.

The applications of EVT in environmental modeling span a wide range of extreme phenomena. For example, bridging EVT and spatial point processes, Koh et al. (2023) modeled moderate and extreme wildfires in the south of France using FWI covariates to predict risk patterns. Similarly, Yadav et al. (2023) investigated landslides counts and sizes through a combination of EVT and spatial methods. Both studies highlight the versatility of EVT when integrated with spatial modeling approaches and using different grid resolutions. Spatial dependencies are particularly important in the context of extreme events, as nearby locations might exhibit similar behaviors. Recent advances have focused on understanding and modeling dependencies specifically in extreme scenarios. For example, Engelke and Hitz (2020) used elements from multivariate EVT and graphical modeling to analyze dependencies in river water levels for flood risks using threshold approaches. Based on their methodology, Cisneros, Hazra, and Huser (2024) further developed a modeling strategy of spatial dependencies for a fire danger index in Australia using a mixture of trees.

Graphical models provide a complementary framework for analyzing dependency structures, especially in high-dimensional datasets. As detailed in Lauritzen (1996), these models represent conditional dependencies in multivariate contexts and aim to provide a sparse probabilistic view of high-dimensional data. This is particularly relevant in spatial statistics, where the number of locations can lead to high-dimensional challenges. In the multivariate normal case, conditional independence is well defined via the entries of the inverse of the covariance matrix, and dependency graphs can easily be derived (Lauritzen 1996). In an attempt to define graphical models for extremes, Engelke and Hitz (2020) introduced a new definition of conditional independence for threshold exceedances. This definition enables them to further define graphical models tailored for extreme values following multivariate Pareto distributions. Subsequent research developed learning algorithms and estimation strategies for tree structures and more general graphical models (Engelke and Volgushev 2022; Engelke, Lalancette, and Volgushev 2021). These graphical models for extremes have proven to be valuable in uncovering sparse dependency structures in high-dimensional datasets and interpretation of extreme event behavior (Engelke and Hitz 2020; Cisneros, Hazra, and Huser 2024).

This short literature review shows the interplay between spatial modeling, EVT, and graphical models, highlighting their respective roles in advancing our understanding of wildfires and

other extreme environmental phenomena. The integration of these methodologies provides powerful tools to analyze dependencies, predict risks, and inform decision-making in complex spatial contexts. The remainder of this thesis is structured as follows. Chapter 2 introduces the theoretical background and key statistical concepts. Chapter 3 describes the proposed model, methodology and estimation procedure. Chapter 4 presents a simulation study, followed by results and implications. Chapter 5 applies the model to a wildfire dataset from Castilla-La Mancha in Spain. Finally, Chapter 6 concludes this thesis with a summary and suggestions for future work.

2. Theoretical Foundations

This chapter aims to introduce the theoretical background needed to understand the problem and to set some notation that will be reused in the following chapters. The goal of this thesis is to explore the graphical modeling of dependence structures of spatially distributed extreme events. This can be broken down into three theoretical blocks for which some background is needed.

- First, I am interested in spatially distributed extreme events. In this setting, the data could exhibit a structure similar to point processes. I will introduce spatial point processes with associated marks, mainly following Baddeley and Turner (2005).
- Then an added complexity to this setting is the extreme component. In fact, specific theory and distributions were developed for analyzing distribution tails. This section will thus provide relevant theoretical background on the analysis of extremes and motivate the use of threshold methods following the introductory book by Coles et al. (2001) and the multivariate notation in Engelke and Hitz (2020).
- Finally, to model the dependency structure of these extreme events using graphical models, I will motivate and introduce their use in the third section. In particular, I will provide the general definitions, vocabulary, and properties based on Lauritzen (1996), as well as the particularities when dealing with extremes as described in Engelke and Hitz (2020).

2.1. Spatial Point Processes

Spatially distributed extreme events can take different forms. For example, some climatic measurements can be taken over a geographic region at regular times (e.g. hourly or daily temperatures, humidity levels, or air pollutant concentrations). In that case, an extreme event can be defined as extreme measurement levels either locally in one location or more broadly (extreme measures over several regions). Other climatic events can arise more sporadically in space and time and are not necessarily subject to constant measurements per se (e.g Earthquakes, Wildfires...). These events are typically modeled via spatial point processes. This thesis focuses on the latter. In this section, unless otherwise stated, the theory, concepts, definitions and formulas relating to (spatial) point processes are citations from Baddeley, Bárány, and Schneider (2007). The notation might slightly differ in some cases, to ease reading of this thesis. More in depth theory on spatial data including point processes can also be found in the well-cited book by Cressie (2015).

Marked Point Processes

A point process X is a statistical representation of randomly located points in a d-dimensional space. Each point is characterized by its location in the given space and can be associated with one or more values, in which case they are referred to as *marked* point processes. In the one-dimensional case, the points are ordered on a line (e.g across time). A spatial point processes lives on an at least 2 dimensional space such that a location corresponds to a geographic point (e.g. longitude/latitude or UTM coordinates). Formally, Baddeley, Bárány, and Schneider (2007) refer to the space of locations S and the space of marks M. They further define Y as a marked point process on S with marks in M, where a single point is then defined by the pair (x, m) with $x \in S$ and $m \in M$. The marks can be numerical values, categorical or even more complex objects. The underlying point process X on S represents the process without marks. As opposed to the one-dimensional case, the points of a multi-dimensional process do not follow a natural ordering which leads to more general mathematical formulations. In the rest of this thesis, I will focus on the case of spatial point processes and assume $S \subseteq \mathbb{R}^2$.

Distribution and First Moments

The distribution of a point process X can be defined via the joint distribution of the *count variable* $N(B)$. The latter can be defined as a measure and represents the number of points in a specified area defined by a subset $B \subseteq S$. The count variable satisfies quite natural and intuitive properties for which I will not go into details but a few examples discussed by Baddeley, Bárány, and Schneider (2007) are:

- For two disjoint subsets $A, B \in S$, $N(A \cup B) = N(A) + N(B)$. In words, the number of points in two separate (non-overlapping) areas can be simply added together.
- $N(\emptyset) = 0$ meaning that the empty set cannot count a positive number of points. Geographically, this makes sense: it is not possible to count events in a non existent geographical location.
- $N(B) \geq 0$ i.e no negative number of events are possible.

The associated point process is also generally assumed to be finite, meaning $N(B) < \infty$ with probability 1. A point process is simple if all points are distinct, meaning no duplicates of a given location exist. If $N(B)$ is known for all subsets B, then the point process can be fully retrieved. Therefore, a point process can be defined by a collection of the count variable $N(B)$.

The count variable is also crucial for defining the first and second moments of a point process X. The first moment is called *intensity* and is defined as:

$$\nu(B) = E(N_X(B)), \quad B \subset S \tag{2.1}$$

where $v(B) < \infty$ for all compact B. The second moment measure is generally defined as:

$$\nu_2(A \times B) = E(N_X(A)N_X(B)), \quad A \times B \subset S \times S \tag{2.2}$$

The definition of intensity for a marked point process Y is a natural extension where $\nu(U) = E(N_Y(U))$ with $U \subset S \times M$. It corresponds, for example, to the number of points in an area with marks within a specified range. Formally:

$$\begin{aligned} \nu(B) &= E(N_X(U)), \quad U \subset S \times M && (2.3)\\ &= E(N_X(B \times C)), \quad B \subset S, C \subset M && (2.4)\\ &= E\big(\sum_{(x,m)\in \mathbf{Y}} \mathbf{1}\{x \in B\}\mathbf{1}\{m \in C\}\big) && (2.5) \end{aligned}$$

A *reduced* form of the second moment measure can be derived for stationary point process. A point process X in $\mathbb{R}^2$ is stationary if the first (and second) moment measures are invariant under (simultaneous) shifts. This means that:

$$E(N_X(B + v)) = E(N_X(B)) \tag{2.6}$$

$$E(N_X(A + v)N_X(B + v)) = E(N(A)N(B)) \tag{2.7}$$

for all $v \in \mathbb{R}^2$. The definition of the reduced second moment can be found in the literature (Baddeley, Bárány, and Schneider 2007, Theorem 2.4). Important here is to know that this reduced form exists, which is used to then define the *K-function*. The latter is a fundamental function used for (spatial) point processes and gives information on the expected number of points within a specified distance t from any given point in the process. Hence, it provides first information on the underlying dependency structure. Ripley (1976) cited by Baddeley, Bárány, and Schneider (2007) introduced the following function:

$$\beta K(t) = \mathcal{K}(b(0, t)), \quad t \geq 0 \tag{2.8}$$

where $\mathcal{K}$ is the reduced second moment measure and β is the intensity of the process X. The area $b(0, t)$ represents a circle of radius t around a point x in the process. Put into words, $\beta K(t)$ represents the expected number of *other* points in this circle. The function `Kest` of the `R`-package

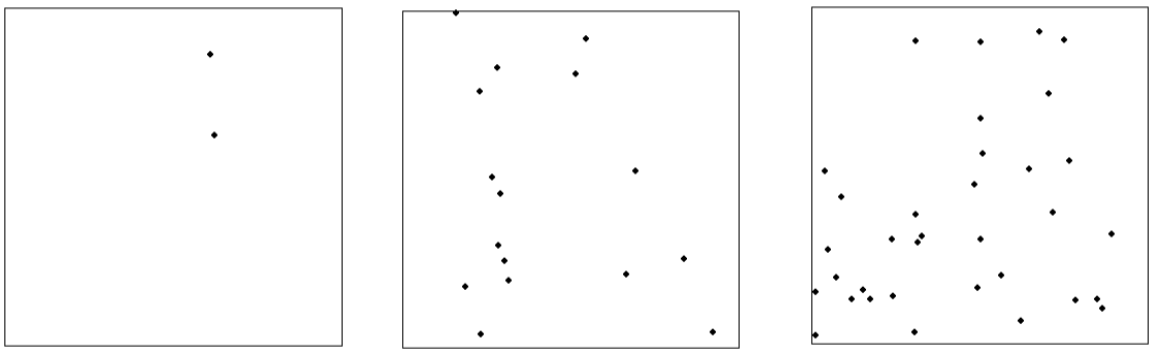
`spatstat` (Baddeley and Turner 2005) estimates this K-function from the data and compares it to the theoretical values of a Poisson process of constant intensity. The latter can be seen as a process without any interaction or structure. If the estimates are above the theoretical K-function, it suggests clustering in the dataset.

Point Process Example: Spatial Poisson Process

An example of spatial point processes is the Poisson process in $\mathbb{R}^2$. It assumes that the count variable $N(B)$ follows a Poisson distribution $Pois(\beta\mathcal{A}(B))$, where β is the expected number of points per unit area. $\mathcal{A}(B)$ refers to the area of B. Figure 2.1 shows realizations of Poisson process in a 2-dimensional space for different intensities. Further, $N(B_1), .., N(B_m)$ are independent if $B_1, ..., B_m$ are disjoint regions (non-overlapping).

The Poisson process can then easily be derived if assuming the following:

- The expected number of points in an area $\mathcal{A}(B)$ is equal to the expectation of the Poisson process. Formally, $E(N(B)) = \beta\mathcal{A}(B)$.
- The probability that a small area will count a positive number of points should be low. Formally, for a small area, the authors write $P(N(B) > 1) = o(\mathcal{A}(B))$.



(a) (b) (c)

Figure 2.1.: Single realizations of a spatial Poisson process in a unit square with intensity (a) 5, (b) 15 and (c) 30. R Package used: *Spatstat* (Baddeley and Turner 2005)

Wildfires as Spatial Point Processes

As previously mentioned, spatially distributed events like wildfires can naturally be seen as a spatial point process. In fact, the location of fires can be seen as random on a spatial two-dimensional

grid. For example, Koh et al. (2023) worked on modeling wildfires, using the spatial point processes framework. In their study, they considered fire locations and associated burned land and ignition date as a realization of a marked Poisson point process. The counts $N(B)$ are therefore assumed to follow a Poisson distribution with expectation $E(N(B)) = E(\sum_{i=1}^{N} 1(x_i \in B))$. One particularity is that they distinguish between *moderate* and *extreme* marks i.e. burned land of different sizes.

2.2. Extreme Value Theory

Extreme Value Theory (EVT) plays a critical role in analyzing rare but impactful events, such as wildfires. These events are challenging to study due to the scarcity of data in the distribution's tail, leading to increased uncertainty when analyzing extremes (Davison and Huser 2015). This uncertainty intensifies as one moves further into the tail, where observations become sparser. In such cases, EVT offers a robust theoretical framework for studying extreme values. However, the reliability of EVT-based analysis depends on a clear understanding its underlying assumptions, which directly influence the accuracy of results. To provide the necessary foundation, this section summarizes key concepts of extreme value theory and presents important distribution families that will be used in later chapters. In the following, I present key information, definitions, and notations from the introductory EVT book by Coles et al. (2001). Unless otherwise specified, all concepts, definitions and equations are citations from this source. In some cases, the notation will differ slightly, to ensure consistency across this thesis. Another introductory book by Haan and Ferreira (2006) is also available. For a more in-depth lecture, the reader is advised to look into these sources.

Peak-Over-Threshold vs. Block Maxima

There are two main approaches in extreme value theory: the block maxima and the threshold approach. They directly refer to the characterization of extreme data. Where the block maxima methods focus on modeling maximum values within a block of values, the threshold method focus on all values above a certain threshold. Suppose a collection of measurements (extreme and not extreme) $X_1, ..., X_n$.

- For a block maxima approach, one defines a *block* (e.g a day, a week, a year...) for which only the maximum value is kept. This leads to a collection of maxima $M_1, ..., M_k$. Then, limiting distributions for the maxima can be derived, commonly a Generalized Extreme Value (GEV) distribution.

- For the threshold approach, or also called *Peak-Over-Threshold*, one is interested in all values exceeding a certain threshold u. Thus, are taken into account all values considered as high, and not only the maximum value. The values over the threshold are then typically modeled via a Generalized Pareto Distribution (GPD). Davison and Huser (2015) and Coles et al. (2001) qualify this approach as less wasteful than the block maxima approach, as it naturally incorporates more data (e.g second highest value). In addition, in a recent publication, Huser and J. L. Wadsworth (2019) discussed their use in modeling environmental data and strongly advised the use of threshold methods as they offer greater flexibility and seem to be better suited for such applications.

Distributions of the Univariate Extremes

The two main distribution families to model univariate extreme values are, as mentioned above, the Generalized Extreme Value distribution for maxima and the Generalized Pareto distribution for threshold exceedances. Although the focus of this thesis application is primarily on threshold exceedances, the two distribution families are connected. Therefore, I will present both.

First, consider a sequence of n independent and identically distributed variables $X_1, ..., X_n$ (e.g. hourly temperature measurements). Denote the distribution and density function $F_X(x)$ and $f_X(x)$ respectively. If one is interested in the distribution of the maximum value $M_n = max(\{X_1, ..., X_n\})$, it can be directly derived from the distribution F_X:

$$P(M_s \leq z) = F_X(z)^n \tag{2.9}$$

However, since F is usually unknown, this derivation if not desirable, as errors in its characterization can have a big impact on the distribution of Maxima. In addition, for stabilization purposes, it is preferable to model a rescaled version of the maxima M_n by means of two suitable constants $a_n > 0$ and b_n. If a sequence of such constants exists such that the limiting distribution of the scaled Maxima is a non-degenerate distribution function, then, by Theorem 3.1 in Coles et al. (2001), it is a Generalized Extreme Value distribution. Formally, for suitable constants, the authors write:

$$P(M^* \leq z) = P\Big(\frac{M_n - b_n}{a_n} \leq z\Big) \rightarrow G(z) \quad as\ n \rightarrow \infty \tag{2.10}$$

The GEV distribution function is said to be max stable and is defined as:

$$G(z) = exp\Big(- \big(1 + \xi\Big(\frac{z-\mu}{\sigma}\Big)\Big)^{-\frac{1}{\xi}}\Big) \tag{2.11}$$

To fit such a distribution, the data is divided into blocks, and the maximum value from each

block is extracted, resulting in a sequence of maxima—hence the name block maxima method. The fitting process is then performed on this sequence.

The distribution of the values exceeding a set threshold u can be defined via the probabilities of the values given that they exceed u. Formally:

$$P(X_i > u + y | x > u) = \frac{1 - F_X(u + y)}{1 - F_X(u)} \tag{2.12}$$

However, as F_X is typically unknown, the following approach is needed. If the distribution of the maxima M_n follow a GEV distribution i.e.

$$P(M \leq z) \simeq G(z) \tag{2.13}$$

for large enough n, then, the distribution function of the threshold exceedances $y = (x - u)$ can be approximated for large u by the family of GPD (Theorem 4.1 in Coles et al. (2001)):

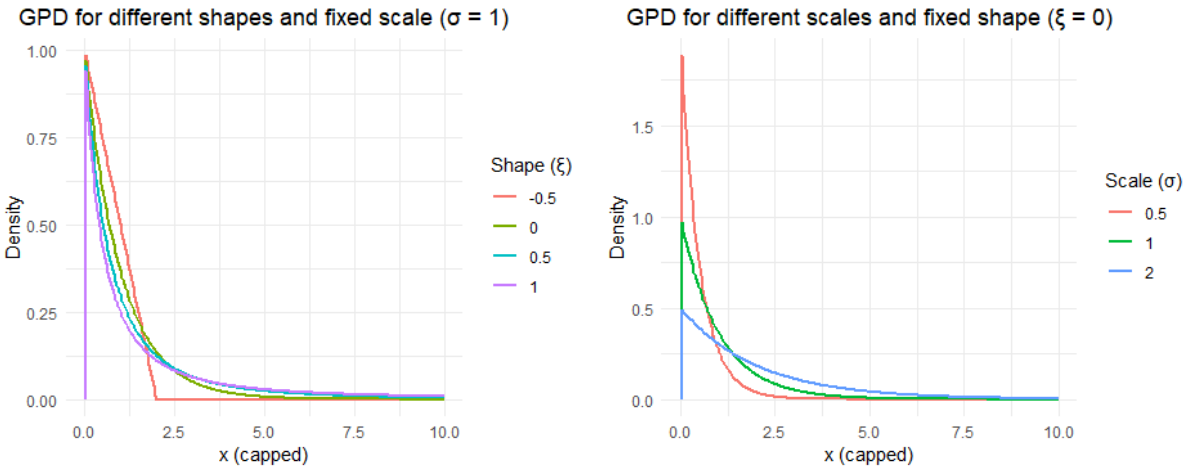
$$H(y) = 1 - \left(1 + \xi \frac{y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}} \tag{2.14}$$

for $y > 0$ and $(1 + \xi \frac{y}{\tilde{\sigma}}) > 0$. This lead to the following conditional probability formulation:

$$P(X > x | X > u) = \left(1 + \xi \left(\frac{x - u}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}}\right) \tag{2.15}$$

which relates somehow to the GEV distribution in (2.11).

Overall, these are important results as they establish a connection between the GEV and the GPD. Specifically, if the distribution of the maxima can be approximated by a GEV for sufficiently large n, then the threshold exceedances follow a related GPD. In fact, note that the GEV and GPD have a common parameter ξ which is referred to as the *shape* parameter and is invariant to the block size. The second parameter of the GPD is $\tilde{\sigma}$ which is the *scale*. The figure 2.2 displays the density of the GPD for different shape and scale parameters. It appears clear that the shape parameter defines the tail thickness. For negative values, the tail becomes finite, whereas for large positive values, the tail becomes heavy. The scale rather affects the spread of the distribution i.e. how much the values are concentrated on the x-axis.



(a) (b)

Figure 2.2.: Visualization of the density function of the Generalized Pareto Distribution for (a) varying shape and fixed scale and (b) varying scale and fixed shape. R Package used: `evd` (A. G. Stephenson 2002)

Multivariate Distributions

So far, we have focused on existing distribution families to characterize the univariate extremes. However, the key aspect of this thesis is the dependency analysis of extreme values, making the multivariate characterization particularly important. Even though multivariate extreme value theory is discussed by Coles et al. (2001), I will build upon the framework presented in Engelke and Hitz (2020). Throughout this subsection, I will introduce their multivariate notation and, using the relevant formulas from this publication, present the connection between the multivariate GEV and GPD via the exponent measure. The following are thus combined citations from this publication. Slight differences in notations might be used to better blend within this thesis.

First, the authors consider a d-dimensional vector $\mathbf{X} = (X_1, ..., X_d)$ and assume the interested lies in the tail. Let $\mathbf{X}_i$ for $i = 1, ..., n$ denote independent copies of the random vector $\mathbf{X}$. In addition, to simplify the analysis, they assume that each component $X_k, k \in \{1, ..., d\}$ has been normalized to a standard Pareto distribution. The univariate maxima approach extends naturally to the multivariate case by considering the component-wise maxima. Formally they consider:

$$\mathbf{M_n} = (M_{1n}, ..., M_{dn}), \quad where\ M_{kn} = max_{i=1}^{n} X_{ik} \tag{2.16}$$

Specifically, for large n and appropriate normalizing constants $a_{kn} > 0$ and b_{kn}, each component Maxima M_{kn} follows a GEV distribution as introduced before.

Without delving into the full theoretical development, Engelke and Hitz (2020) provide with the following result for the joint distribution of the maxima:

$$\lim_{n\to\infty} \mathbb{P}(M_{1n} \leq nz_1, ..., M_{dn} \leq nz_d)) = \mathbb{P}(\mathbf{Z} \leq \mathbf{z}) \tag{2.17}$$

where $\mathbf{Z}$ is a d-dimensional max stable vector with standard Fréchet marginals. The latter is a special case of a GEV distribution, with $\mu = 0, \sigma = 1, \xi = 1$ (Coles et al. 2001). For any $\mathbf{z} = (z_1, ..., z_d)$, its distribution is given by Engelke and Hitz (2020) as:

$$\mathbb{P}(\mathbf{Z} \leq \mathbf{z}) = exp\{-\Lambda(\mathbf{z})\}, \quad \mathbf{z} \in [0, \infty)^d \setminus \{0\} \tag{2.18}$$

Here $\Lambda(\cdot)$ is known as the exponent measure and is fundamental to understanding the link with the multivariate exceedances. In fact, the authors further write the multivariate Pareto distribution of threshold exceedances $\mathbf{Y}$ as a function of $\Lambda(\cdot)$:

$$\mathbb{P}(\mathbf{Y} \leq \mathbf{z}) = \frac{\Lambda(\mathbf{z} \wedge \mathbf{1}) - \Lambda(\mathbf{z})}{\Lambda(\mathbf{1})} \quad \mathbf{z} \in [0, \infty)^d \setminus \{0\} \tag{2.19}$$

and the corresponding density as:

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{\lambda(\mathbf{y})}{\Lambda(\mathbf{1})}, \quad \mathbf{y} \in \mathcal{L} \tag{2.20}$$

where $\lambda(\cdot)$ is the density of the exponent measure $\Lambda(\cdot)$.

Thus, the exponent measure $\Lambda(\cdot)$ and its density $\lambda(\cdot)$ mutually define both the multivariate Pareto distribution (used in modeling of threshold exceedances) and the distribution of maxima (used in block maxima modeling). This connection will be reused later, when introducing the Hüsler-Reiss distributions (Hüsler and Reiss (1989) as cited in Engelke and Hitz (2020)). In the context of Hüsler-Reiss distributions, the exponent measure encapsulates the dependency structure between the components of the threshold exceedance vector $\mathbf{Y}$.

Having established the foundational concepts of univariate and multivariate extreme value theory, we now shift focus to two important tools for describing extreme events: return levels and extremal dependence. These measures help understanding the intensity and frequency of rare events, as well as the relationships between extreme values across multiple dimensions. These descriptive tools will be used in Chapter 5, when analyzing forest fires in spain.

Return Levels

Return levels are a standard univariate measure used to quantify the rarity and magnitude of extreme events. Formally, for a GPD with parameter ξ and $\tilde{\sigma}$, Coles et al. (2001) express return levels through the following, which builds on the equation (2.15):

$$P(X > x) = P(X > u)P(X > x|X > u) = \zeta_u\Big(1 + \xi\Big(\frac{x-u}{\tilde{\sigma}}\Big)^{-\frac{1}{\xi}}\Big) \tag{2.21}$$

In this equation, $P(X > x)$ represents the probability that the random variable X exceeds some value x. It can be retrieved via multiplying the probability that the variable exceeds some threshold u, $P(X > u)$, with the probability that X exceeds x given that it has already exceeded u, $P(X > x|X > u)$. Building on this, Coles et al. (2001) define the "*m-observation return level*", x_m, which corresponds to the value that is expected to be exceeded only once in m observations. In other words, the value x_m is the level that has a $1/m$ probability of being exceeded, which directly relates the return level to a specific recurrence period:

$$P(X > x_m) = \frac{1}{m} \tag{2.22}$$

For example, in the context of annual data, a 10-year return level corresponds to a value that is statistically expected to be exceeded once every 10 years.

Extremal Correlation

We now turn to the multivariate case, where we consider not just individual components of a vector $\mathbf{X} = (X_1, \ldots, X_d)$ but also their joint tail behavior. In this context, understanding the dependence between extreme values across multiple dimensions is crucial, as it provides insights into how extreme events in different components might occur simultaneously.

One important measure of dependence in the extremes is extremal correlation. In the two-dimensional case, Coles et al. (2001) define extremal correlation between two random variables X_i and X_j as:

$$\chi_{ij} := \lim_{z \to z_+} \mathbb{P}(X_i > z | X_j > z) \tag{2.23}$$

Here, χ_{ij} quantifies the limiting probability that one variable exceeds a threshold z given that the other variable also exceeds the same threshold. As z approaches the end of the joint distribution, χ_{ij} reflects the degree of asymptotic dependence between the variables. A value of $\chi_{ij} = 1$ indicates that the two variables are perfectly dependent in the extremes, meaning that when one variable is extreme, the other is likely to be extreme as well. Conversely, $\chi_{ij} = 0$ suggests

asymptotic independence, implying that extreme values of one variable do not correspond with extremes in the other.

In the case of threshold exceedances, where the random vector $\mathbf{Y}$ follows a multivariate Pareto distribution with standardized marginals, Engelke and Volgushev (2022) provide a direct formulation of extremal correlation as: $\chi_{ij} = \mathbb{P}(Y_i > 1 | Y_j > 1)$. This formulation assumes that the exceedances have been standardized such that the threshold is set to 1, simplifying the analysis. Specifically, this indicates that the extremal correlation reflects the conditional probability that one component exceeds a threshold (in this case, 1) given that another component has already exceeded it.

This measure plays a key role in the later stages of the thesis, particularly when exploring how dependencies between the marginals are modeled using distributions such as the Hüsler-Reiss distribution (Hüsler and Reiss 1989), which we will introduce in Chapter 3.

Wildfires and Extreme Value Theory

Having discussed the foundational concepts of extreme value theory, we can now explore how these tools can be applied to the study of extreme events, such as wildfires. The threshold exceedance method, for instance, offers a structured approach that uses all values above a predefined threshold, minimizing data loss compared to block maxima approaches. This method has already been applied to wildfire data, particularly in modeling the burnt area of wildfires, as demonstrated by Koh et al. (2023). The threshold approach also provides customized and flexible ways of defining extreme events. In fact, Koh et al. (2023) incorporate both moderate and extreme values, which is particularly useful in the study of wildfires. The inclusion of moderate values (e.g., smaller fires) can help reduce uncertainty by increasing data volume, thereby enhancing the robustness of extreme event analyses.

2.3. Graphical Models

In the analysis of spatial extreme events and their prediction, it is crucial to understand their dependence structure. In fact, one might be interested in how an extreme event in one location can help predict an event in another. For example, a forest fire might be spreading in a specific direction due to the type of vegetation or the geographical slope. Or, nearby areas might share similar weather conditions prone to fire risk, resulting in simultaneous wildfires in different locations.

In the previous section, we introduced the basic framework of extreme value theory. We primarily discussed the common distributions and we briefly introduced multivariate distributions and the extremal dependence coefficient. This serves as a first step in characterizing tail dependencies between components of a multivariate vector. New research has been conducted on visualizing extremal dependencies using graphical models, also in a spatial context (Engelke and Hitz 2020; Cisneros, Hazra, and Huser 2024). Graphical models are built to provide a visual representation of complex dependency structures between different components or variables. They offer a powerful framework to effectively picture dependency structures using definitions of conditional independence. It can offer a first step in entangling a much bigger problem by enforcing sparsity and simple probabilistic structures, especially in high-dimensional settings. It can also be used to further investigate causal structures (see Glymour, Zhang, and Spirtes (2019) for a review).

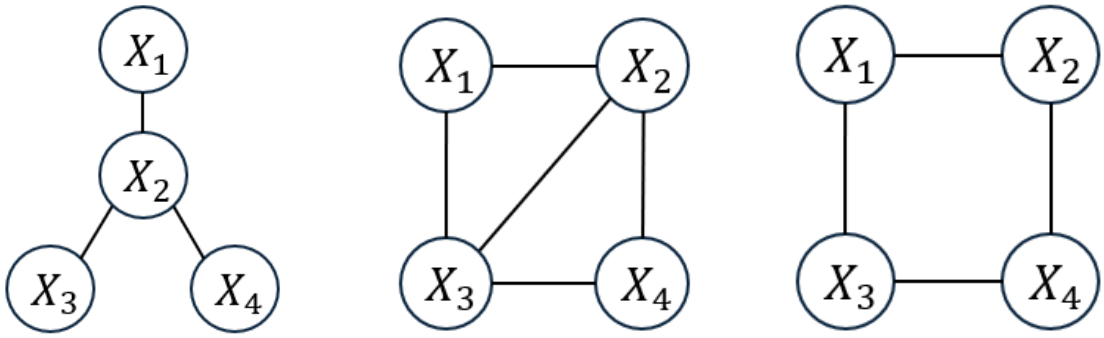
In this section, I will introduce the theoretical foundations, terminology, and properties of graphical models. If not stated otherwise, all definitions, concepts, notation and formulas are combined citations from Lauritzen (1996). In addition, I will introduce some specificities when graphically modeling extremes using the work of Engelke and Hitz (2020). Note that the following are only brief overviews; the reader is advised to look in the corresponding literature for more detailed definitions.

2.3.1. Graphical Models - Basic Concepts and Terminology

Nodes, Edges, and Graph Structure

A Graph is composed of *nodes*, also called *vertices*, representing the variables connected via lines or arrows called *edges*. Formally, one writes $G = (V, E)$ where V is the set of nodes and E the set of edges. The edge set is a subset of all pair combinations of nodes. They can be directed (commonly represented by an arrow) or undirected (simple line). Consider a random vector of multiple components $\mathbf{X} = (X_1, ...X_d)$. The set of nodes then typically consists of all components of the vector, i.e. $V = \{X_1, ...X_d\}$ or sometimes written $V = \{1, ...d\}$. An edge connecting two nodes X_i and X_j is undirected if both (X_j, X_i) and $(X_i, X_j) \in E$. If only one of them is present, the edge is directed. Two nodes are *neighbors* if they are connected via an undirected edge. A graphical model is undirected if all edges are undirected. Examples for such graphs can be found in figure 2.3. I will not go into details on graphs with directed edges as they are not the focus of this thesis.

The notions of clique and graph completeness are also central to describing and defining



(a) (b) (c)

Figure 2.3.: Examples of undirected graphs with four nodes (examples taken from Fig 1 in Engelke, Lalancette, and Volgushev (2021)). (a) is an example of a tree graph, (b) of a decomposable graph and (c) of a non-decomposable graph

graphical models. Quite simply, if a graph is fully connected i.e. all nodes are connected to each other via edges, then it is said to be complete. The example graphs in figure 2.3 are, for example, not complete. If an additional edge connected X_1 to X_4, then graph (b) would be complete. If a subset of the graph is complete, then the set of nodes forming this sub-graph is called a *clique*. For example, in graph (b), the set $\{X_1, X_2, X_3\}$ is a clique.

Separators and Graph Decomposability

Separability and decomposability are key structural properties of graphical models. These notions refer to the property that the set of nodes can be partitioned into disjoint subsets, enabling the graph to be decomposed into distinct components. Consider, for example, the graph (a) in figure 2.3. We can identify 3 disjoint sets: $A = \{X_1\}$, $B = \{X_3, X_4\}$ and $C = \{X_2\}$. One can then notice, that to go from a node in set A to a node in the set B, it is necessary to pass through the node in C. In this scenario, we say that A is *separated* from B by C. C is called a seperator set. If, in addition, the union of these three sets corresponds to V and C is fully connected, then the three sets A, B and C are said to form a *weak decomposition* of the graph:Then, An undirected graph is decomposable if a decomposition in the above sense exists. The graph G is then decomposable into two subgraphs on $A \cup C$ and $B \cup C$. In figure 2.3, the graphs (a) and (b) are decomposable, whereas the graph (c) is not.

Conditional Independence

An edge represents a dependency between two nodes, given the rest of the nodes in the graph. As importantly, the non-existence of an edge indicates a lack of dependency between two nodes given all the others. The notion of conditional independence thus plays a fundamental role in graphical modeling. Consider a simple model with three continuous random variables X, Y and Z. X is said to be conditionally independent of Y given Z if a factorization of the joint density of the following form exists:

$$X \perp\!\!\!\perp Y|Z \iff f_{XY|Z}(x, y|z) = f_{X|Z}(x, z)f_{Y|Z}(y|z) \tag{2.24}$$

$$\iff f_{XYZ}(x, y, z)f_Z(z) = f_{XZ}(x, z)f_{YZ}(y, z) \tag{2.25}$$

Put into words, if, for instance, the variables represent knowledge extracted from books, Lauritzen (1996, p.30) interpret the conditional independence $X \perp\!\!\!\perp Y|Z$ intuitively as: *"Knowing Z, reading Y is irrelevant for reading X"*.

Pairwise and Global Markov Property

Graphical models hold an important property called global Markov property which relates to the notions of separation and conditional independence. Consider an undirected graph G where the set of vertices V is composed of a collection of random variables. Then, the graph is said to possess the global Markov property if, for three disjoint subsets A, B and C where C separates A from B:

$$A \perp\!\!\!\perp B|C \tag{2.26}$$

This property allow us to read the graph in the following way: If two sets of nodes A and B are separated by a third separator set C then, the random variables in A and in B are independent, given the random variables in the set C. A weaker property is the pairwise Markov property, which holds if for any pair of non-connected nodes (X_i, X_j):

$$X_i \perp\!\!\!\perp X_j|V \setminus \{X_i, X_j\} \tag{2.27}$$

It appears clear that the Markov properties are directly related to the conditional independence and thus to the factorization of a probability measure. A well-known theorem in the field of graphical modeling is the Hammersley and Clifford theorem; see Theorem 3.9 in Lauritzen (1996). This theorem proves the equivalence between the pairwise Markov property of an undirected graph and the factorization of the density f of a probability distribution on the graph. Meaning

that, if the density f of a probability distribution factorizes according to the graph, then the pairwise Markov property holds and conversely. These properties allow us to read the graph in terms of conditional dependencies.

Conditional independence for an undirected graphical model can be approached, as we have seen, via the factorization of the joint probability density of the random variables. In particular:

$$f(\mathbf{x}) = \prod_{c \in \mathcal{C}} \psi_c(\mathbf{x}) \tag{2.28}$$

The $\mathcal{C}$ is here the set of cliques, meaning that the density factorizes over all cliques of the graph. In the example graph (a) in figure 2.3, this would mean the density factorizes in the following way:

$$f(\mathbf{x}) = g(x_1, x_2)h(x_2, x_3)i(x_2, x_4) \tag{2.29}$$

Note that, if the graph G is additionally decomposable by the three sets (A, B, C) defined as before, then the joint density can be expressed in terms of the following marginal densities:

$$f_x(x) = \frac{f_{A \cup C}(x_{A \cup C}) f_{B \cup C}(x_{B \cup C})}{f_C(x_C)} \tag{2.30}$$

2.3.2. Graphical Models for Extremes

Having established the basic terminology and core properties of graphical models, we now focus on the case of extremes. Specifically, we focus on the work of Engelke and Hitz (2020), who defined a graphical model to represent the conditional independence structure, not of the random variables in $\mathbf{X}$, but of their extreme values, quantified through threshold exceedances $\mathbf{Y}$. In their work, they provide a formal definition of conditional independence for threshold exceedances, which is essential for properly defining a graphical model for extremes. In this section, we discuss their approach and their definition of conditional independence in this context.

Conditional Independence for threshold exceedances

In fact, given a multivariate vector of threshold exceedances $\mathbf{Y}$ of d components following a multivariate Pareto distribution, the notion of conditional independence differs from the classical case. The definition of conditional independence in (2.25) lies on product spaces, whereas multivariate Pareto distributions have support on $\mathcal{L}$-shaped space (Engelke and Hitz 2020), which they write as:

$$\mathcal{L} = \{\mathbf{x} \in \mathcal{E} : \|x\|_\infty > 1\} \tag{2.31}$$

where $\mathcal{E} = [0, \infty)^d \setminus \{\mathbf{0}\}$ and $\|x\|_\infty$ denotes the sup norm. The latter represents the largest absolute value of the vector $\mathbf{x}$. The space $\mathcal{L}$ thus consists of all $\mathbf{x}$ in $\mathcal{E}$ where at least one component exceeds 1 in absolute value. This is the case since we now condition on the fact that the values exceed a certain threshold (here 1). Thus, a definition on the product space as before is not possible. However, the authors found a way around it by restricting $\mathbf{Y}$ to product spaces. In particular, they consider a vector $\mathbf{Y}^k$ which corresponds to $\mathbf{Y}$ given that the element Y_k exceeds the threshold 1. Then $\mathbf{Y}^k$ has support on the product space $\mathcal{L}^k$ which is a subspace of $\mathcal{L}$ with the additional condition that $x_k > 1$. The density of $\mathbf{Y}^k$ can then be written as (Engelke and Hitz 2020):

$$f^k(\mathbf{y}) = \frac{f_{\mathbf{Y}}(\mathbf{y})}{\int_{\mathcal{L}^k} f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y}} = \lambda(\mathbf{y}), \qquad \mathbf{y} \in \mathcal{L}^k \tag{2.32}$$

This then enables the authors to define extremal conditional independence in the following sense. For a multivariate Pareto distribution $\mathbf{Y}$ with positive and continuous density $f_{\mathbf{Y}}$ on the $\mathcal{L}$-shaped space introduced above, they first define the set $V = \{1, ..., d\}$ and three disjoint, non-empty subsets $A, B, C \subset V$ where $A \cup B \cup C = V$. Then, the authors define the extremal conditional independence between $\mathbf{Y}_A$ and $\mathbf{Y}_B$ given $\mathbf{Y}_C$ as (Engelke and Hitz 2020, Definition 1):

$$\forall k \in V : \mathbf{Y}_A^k \perp\!\!\!\perp \mathbf{Y}_B^k | \mathbf{Y}_C^k \tag{2.33}$$

In short, they write: $\mathbf{Y}_A \perp_e \mathbf{Y}_B | \mathbf{Y}_C$ where the subscript e emphasizes the extremal setting.

Engelke and Hitz (2020) further show that this definition is equivalent to the factorization of the exponent measure $\lambda(y)$. Formally, they write, for $\mathbf{y} \in \mathcal{L}$:

$$\mathbf{Y}_A \perp_e \mathbf{Y}_B | \mathbf{Y}_C \iff \lambda(\mathbf{y}) = \frac{\lambda_{A \cup C}(\mathbf{y}_{A \cup C}) \lambda_{C \cup B}(\mathbf{y}_{C \cup B})}{\lambda_C(\mathbf{y}_C)} \tag{2.34}$$

Thus, if we compare to the non-extreme case where the joint density can be expressed in terms of its marginal densities, here it is not the case. In fact, the joint multivariate Pareto density of $\mathbf{Y}$ factorizes in terms of the marginal densities of its exponent measure.

Graphical Modeling

Now that a definition for conditional independence for threshold exceedances has been properly defined, Engelke and Hitz (2020) further define graphical models for such extremes. Let the setting be the same as up to now, meaning, assume $\mathbf{Y}$ is a multivariate Pareto distribution with continuous density $f_{\mathbf{Y}}$ on $\mathcal{L}$. Let $G = (V, E)$ be an undirected graph as before, with the set of nodes V being the subscripts of the components in $\mathbf{Y}$ i.e. $V = \{1, ..., d\}$. Then, the authors

rewrite the pairwise Markov property (2.27) in terms of extremal conditional independence, meaning:

$$Y_i \perp_e Y_j | \mathbf{Y}_{\setminus\{i,j\}}, \qquad (i,j) \notin E \tag{2.35}$$

If this property holds, they call the corresponding graph and extremal graphical model.

Similar to the non-extreme case, the authors show that, for a decomposable graph, the pairwise Markov property for threshold exceedances is equivalent to the factorization of the density $f_{\mathbf{Y}}$ into the marginal densities of the exponent measure (Engelke and Hitz 2020, Theorem 1). Formally, for $\mathbf{y} \in \mathcal{L}$:

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{\Lambda(1)} \frac{\prod_{C \in \mathcal{C}} \lambda_C(\mathbf{y}_C)}{\prod_{D \in \mathcal{D}} \lambda_D(\mathbf{y}_D)} \tag{2.36}$$

$\mathcal{C}$ denotes again the sequence of cliques and $\mathcal{D}$ the sequence of separators. In summary, the difference from the non-extreme case lies mainly in the factorization of the density. Specifically, in this context, the density is not factorized on the basis of the marginal distributions. Instead, it is factorized using the marginal density of the exponent measure.

In addition, the authors note that the definition of conditional independence in (2.33) is not extendable to a definition of general independence. In graphical models, general independence is implied by a node not being connected to the rest of the graph. Hence, graphical models using the definition of conditional independence are necessarily connected.

All in all, the newly defined conditional independence for threshold exceedances enables Engelke and Hitz (2020) to introduce two types of extremal graphical models:

- *Tree graphical models*: a special type of graphical model that is decomposable, connected and has the particularity of having no cycles (e.g graph (a) in figure 2.3). A cycle is present in the graph if a path exists that begins and ends at the same node (Lauritzen 1996).
- *Hüsler-Reiss graphical models*: Graphical models for the Hüsler-Reiss Pareto distribution, which is a type of multivariate Pareto distribution. The authors argue that the latter can be seen as the Gaussian distribution in the extreme context, as it shares some similarities. They are parametrized by a variogram matrix that can be used to derive the graphical structure. More details follow in Chapter 3.

To conclude this section and chapter, let us summarize the key concepts introduced so far. This chapter introduced the basic theory and building blocks from three different fields in statistics, starting with spatial point processes, extreme value theory, and finally, graphical modeling. It

also bridged a first gap between extreme value theory and graphical modeling, discussing the theoretical framework developed by Engelke and Hitz (2020). These concepts and definitions lay the foundation for the next chapter, where we introduce a methodology to model the dependencies of spatial point processes using Hüsler-Reiss graphical models. The next chapter will focus on the specific construction and estimation of suited graphical models, building on the concepts introduced so far.

3. Extending Graphical Models for Spatial Extremes

As previously mentioned, understanding spatial dependencies in extreme events is crucial for assessing risks associated with rare climatic events. This chapter focuses on introducing and extending the methodological framework of extremal graphical models as defined by Engelke and Hitz (2020) to further capture underlying spatial structure of spatial point processes. I will be using the example of dependencies between burned land from wildfires, but the framework should be applicable to similar type of events.

Graphical models are powerful tools for representing dependencies between variables due to their sparse probabilistic structure and interpretability (Lauritzen 1996). In the context of extreme value theory, graphical models can depict conditional dependencies for threshold exceedances, as introduced by Engelke and Hitz (2020). Specifically, they define a Hüsler-Reiss graphical model which provides a framework to model conditional dependencies via the estimation of a variogram and related precision matrices. However, its application to spatial settings where dependencies are influenced by geographical proximity has not been fully explored.

The primary objective of this chapter is to present an extension of this Hüsler-Reiss graphical model to incorporate spatial constraints as well as a methodology for its application to spatial point processes. This involves the aggregation of the data on a lower spatial resolution, similar to the alternative modeling approach of spatial point processes suggested in Yadav et al. (2023). In addition, to incorporate meaningful spatial information into the graphical model, the proposed approach involves modifying the model's precision matrices, enabling the detection of conditional extremal dependencies in a geographical plausible way. Specifically, building on the theoretical foundations discussed in Chapter 2, the proposed approach includes:

- A framework for the pre-processing of spatial point processes, including spatial and temporal aggregation of the data and the inclusion of artificial "0-events".
- A spatial constraint on the precision matrix: defining an appropriate maximum distance for which a dependency between two locations is judged as plausible.

The remainder of this chapter is structured as follows. The first section introduces in more detail the Hüsler-Reiss graphical model defined by Engelke and Hitz (2020) and its key components. Section 2 introduces the proposed approach, detailing the data pre-processing steps, spatial scaling methodology, and graphical model construction. Section 3 discusses the advantages of this

framework, particularly in comparison to existing methods, and highlights potential challenges. Finally, Section 4 outlines the estimation procedure, including practical considerations for implementation.

3.1. Hüsler-Reiss Graphical Models

The Hüsler-Reiss distributions (Hüsler and Reiss 1989) can be used to model tail dependencies in a multivariate setting. Engelke and Hitz (2020) refer to the Hüsler-Reiss distribution as the analogue in the extreme setting to Gaussian distributions. They are parametrized by a variogram matrix Γ which encapsulated the dependency structure between all components of a random vector $\mathbf{Y}$. Formally, the variogram matrix Γ is a symmetric, strictly conditionally negative definite matrix with 0 on the diagonal and non-negative entries $\Gamma_{ij,i\neq j}$. Specifically, Γ_{ij} describes the strength of dependence between Y_i and Y_j (Engelke and Hitz 2020). The associated conditional dependencies can be identified via non-zero entries in suited precision matrices (Engelke and Hitz 2020; Hentschel, Engelke, and Segers 2024)

Hüsler-Reiss Distribution

I provide an overview of the functions, equations, and concepts related to the multivariate Hüsler-Reiss distribution, as presented in Engelke and Hitz (2020). The following are thus, unless otherwise specified, combined citations from this source. Throughout, I adhere to their original notation for consistency. First, recall that the continuous density $f_{\mathbf{Y}}$ of a multivariate Pareto distribution can be expressed as a function of the the density function of the exponent measure (see equation 2.20):

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{\lambda(\mathbf{y})}{\Lambda(\mathbf{1})}, \qquad \mathbf{y} \in \mathcal{L} \tag{3.1}$$

For the particular class of Hüsler-Reiss multivariate Pareto distributions, the density of the exponent measure is the following:

$$\lambda(\mathbf{y}) = y_k^{-2} \prod_{i \neq k} y_i^{-1} \phi_{d-1}\Big(\tilde{\mathbf{y}}_{\setminus k}; \Sigma^{(k)}\Big), \tag{3.2}$$

for $\mathbf{y} \in [0,\infty)^d] \setminus \{\mathbf{0}\}$ and some $k \in \{1, ..., d\}$ [1]. In the above expression, $\phi_p(\cdot)$ is the density of a multivariate Normal distribution of dimension p, parametrized by a covariance matrix. The

[1] The density of the exponent measure (3.2) is actually independent of the choice of k (Engelke and Hitz 2020).

density is taken over the following transformed values:

$$\tilde{\mathbf{y}} = \left\{ log\left(\frac{y_i}{y_k}\right) + \frac{\Gamma_{ik}}{2} \right\}_{i=1,\dots,d} \tag{3.3}$$

and, the covariance $\Sigma^{(k)}$ refers to the $(d-1)$-dimensional covariance matrix of the components $\mathbf{Y} \setminus Y_k$. The latter can be expressed in terms of the variogram entries, with

$$\Sigma_{ij}^{(k)} = \frac{1}{2}\{\Gamma_{ik} + \Gamma_{jk} - \Gamma_{ij}\} \quad for\, i, j \neq k \tag{3.4}$$

To provide a clearer understanding of the concept represented by the variogram, without going into too much detail, consider the following. The authors discuss how the Hüsler-Reiss distribution can be represented as component-wise maxima derived from a Poisson point process and a centered multivariate normal distribution $\mathbf{W}$ with covariance Σ. In particular, as they explain, the entries of the variogram matrix Γ correspond to the squared differences between the two corresponding components of $\mathbf{W}$. Formally:

$$\Gamma_{ij} = \mathbb{E}(W_i - W_j)^2 \tag{3.5}$$

This expression can also directly be written in term of the covariance entries (Hentschel, Engelke, and Segers 2024):

$$\Gamma_{ij} = \Sigma_{ii} + \Sigma_{jj} - 2\Sigma_{ij} \tag{3.6}$$

It is then not hard to see that in case of complete dependence i.e. $W_i = W_j$, the corresponding entry Γ_{ij} is zero. Conversely, the greater the difference between the two components, the larger the variogram value. Engelke and Hitz (2020) note that several covariance matrices can correspond to a single variogram matrix. As a consequence, the conditional dependence structure cannot be retrieved from the inverse of the covariance matrix, as is typically done in the regular multivariate normal case (Lauritzen 1996, Chapter 5).

The variogram also relates to the extremal correlation χ. In fact, the following relation exists between the two (Engelke and Volgushev 2022):

$$\chi_{ij} = 2 - 2\Phi\left(\sqrt{\Gamma_{ij}}/2\right) \tag{3.7}$$

with Φ corresponding to the standard normal distribution. This is useful as we can derive from an estimate of Γ, the corresponding extremal correlation matrix which provides information on the strength of the tail dependence.

Hüsler-Reiss Graphical Model

Hüsler-Reiss distributions are referred to by Engelke and Hitz (2020) as sharing similarities with Gaussian distributions. Nevertheless, they note that the analysis of conditional dependence of Hüsler-Reiss distributions differs. In fact, as mentioned above, for multivariate normal distributions, the conditional dependence structure can be read of the inverse of its invertible covariance matrix called the precision matrix (Lauritzen 1996). However, for Hüsler-Reiss Pareto distribution, the graphical structure is retrieved from the set of $(d-1) \times (d-1)$ matrices (Engelke and Hitz 2020):

$$\Sigma^{(k)}; \quad k \in \{1, ..., d\} \tag{3.8}$$

as defined in (3.4) and the corresponding set of precision matrices:

$$\Theta^{(k)} = (\Sigma^{(k)})^{-1}; \quad k \in \{1, ..., d\} \tag{3.9}$$

Further described by the authors, for a given k, the precision matrix $\Theta^{(k)}$ encodes the conditional dependence structures of all pair of nodes $i, j \in V \setminus k$. The conditional dependence of the node k with all other nodes can be then retrieved by either the corresponding entries of a precision matrix $\Theta^{(k')}$ with $k' \neq k$, or via row or column sum of $\Theta^{(k)}$. Formally, they write:

$$Y_i \perp_e Y_j | \mathbf{Y}_{\setminus\{i,j\}} \iff \begin{cases} \Theta^{(k)}_{ij} = 0 & \text{if i,j} \neq \text{k,} \\ \sum_{l \neq k} \Theta^{(k)}_{lj} = 0 & \text{if i = k, j} \neq \text{k,} \\ \sum_{l \neq k} \Theta^{(k)}_{il} = 0 & \text{if i} \neq \text{k, j = k} \end{cases} \tag{3.10}$$

To avoid this representation, Hentschel, Engelke, and Segers (2024) define a single precision matrix that contains all the information of the matrices $\Theta^{(k)}$. The latter enables the characterization of the Hüsler-Reiss graphical model, using a single matrix. Formally, they define a precision matrix Θ of a d-dimensional variogram Γ, for $d \geq 3$ as follows:

$$\Theta_{ij} = \Theta^{(k)}_{ij}, \qquad \text{for some } k \in V \setminus \{i, j\} \tag{3.11}$$

This is possible, as $\Theta^{(k)}_{ij} = \Theta^{(k')}_{ij}$ for distinct $k, k' \in V \setminus \{i, j\}$ (Lemma 1, Engelke and Hitz (2020) cited in Hentschel, Engelke, and Segers (2024)).

In summary, for the Hüsler-Reiss graphical model Y on $G = (V, E)$ defined by Engelke and Hitz (2020), the missing edges $(i, j) \notin E$ indicate independence in the extreme sense between nodes i and j given the rest of the graph. These relationships can be identified using the representation in (3.10) or via the precision matrix (3.11). Hentschel, Engelke, and Segers

(2024) formally write:

$$(i,j) \notin E \iff Y_i \perp_e Y_j | Y_{\setminus\{i,j\}} \tag{3.12}$$

$$\iff \Theta_{ij} = 0 \tag{3.13}$$

A limitation of the standard Hüsler-Reiss graphical model is its inability to explicitly incorporate the spatial structure of multivariate data in a geographic context. In many applications, particularly environmental and climate modeling, geographic proximity plays a crucial role in shaping dependency structures (Cressie 2015). To address this, I suggest scaling the precision matrix Θ using distance-based weights to refine the graphical structure and better capture spatial dependencies. The next section details the proposed methodology for implementing this extension in the case of spatial point processes.

3.2. Proposed Methodology and Extension

3.2.1. Setting and Aggregation Problem

To begin, let us outline the setting. Consider a marked spatial point process representing wildfires in a 2-dimensional space $S \subset \mathbb{R}^2$ (e.g. lon/lat, UTM coordinates, etc.). Let $s_i \in S$ be the location associated to the i-the point in the process. Further, let each point have associated marks $\{(t,x) \,|\, t \in \mathbb{N}, x \in \mathbb{R}_{>0}\}$ where

- t is an index recording at which time unit the event took place. For simplicity, let the index represent a day.
- x is a value associated reflecting the magnitude of the event (e.g. hectares of burned land). Let x follow a generalized Pareto distribution.

In this setting, we have a collection of points representing an event. The event is here a wildfire at a specific location. Typically, the location of such event is fairly precise (high spatial resolution) such that duplicates in position are very unlikely[2]. While spatial point processes are used to model the occurrence of individual wildfire events, an alternative approach in spatial statistics is the use of areal data, which consists in aggregated values over a predefined grid (Yadav et al. 2023; Cisneros, Richards, et al. 2024). In particular, Yadav et al. (2023) investigate an approach where landslide counts and sizes are aggregated (summed up) on a smaller spatial resolution. Cisneros, Richards, et al. (2024) used a similar approach to create a data set of

[2]For example, the spatial point process dataset `clmfires` by Prof. Jorge Mateu in Baddeley and Turner (2005) contains no duplicates in point locations.

aggregated monthly burned areas on an irregular spatial grid in Australia that corresponds to meaningful community-level locations. In this way, the authors argue that the impact can be studied at the community level. Both studies note that this type of spatial aggregation results in not a spatial point process anymore but in so-called areal or lattice data. The latter are observations at a fixed set of locations (on regular or irregular grids) (Cressie 2015). Since we are interested in the dependency modeling of wildfires between different locations, it appears sensible to rely on the aggregation of the underlying spatial point process per time unit.

3.2.2. Data Pre-Processing

The suggested data pre-processing steps consist of two parts: First a spatial aggregation and, second, inclusion of 0-events. In addition, I briefly discuss potential temporal aggregation.

As introduced, I first suggest to smooth spatially the point process, aggregating the points to a fixed set of d locations, similar to the studies mentioned above (Yadav et al. 2023; Cisneros, Richards, et al. 2024). In practical terms, each event of the spatial point process is assigned a modified location s_k^c corresponding to the centroid of the grid cell in which it falls. The subscript k represents one of the d locations. This approach reduces the dimensionality of the data while increasing the number of observations per location, by assigning multiple fires to a single location on the grid. The spatially smoothed data should preserve underlying patterns and relationships. If choosing an appropriate grid resolution, this aggregation process should not suffer from too much information loss. In other words, one must balance between increasing the number of observations per location and keeping the grid cell size appropriately small. In cases where two events fall in the same grid cell on the same day i.e. $t_i = t_j$, they can be aggregated to a single event with a magnitude mark $x_i + x_j$. In other words, the aggregation is first done spatially, keeping distinct time and magnitude marks except in cases where the time and corresponding cell centroid coincide.

Then, a key step is to extend the transformed spatial point process, defining "0-events" between the first and last observed wildfire. Let n denote the number of time units considered. Artificially, points with 0 magnitude marks at locations $k \in \{1, ..., d\}$ at time $t \in \{1, ..., n\}$ may be added if no events are recorded. Let X_k be the vector of all magnitudes (original + the added magnitude 0-events) for a given location k. Thus, each location counts n observations.

The objective of these first preprocessing steps is to obtain consistent values over time for all d locations in the grid. In this context, the focus is not on conventional measurements,

but rather on the magnitude values associated with the event, including "0-events". This approach enables the study of dependencies as the vectors of marks X_k can be summarized in a $n \times d$ matrix $\mathbf{X} = (X_1, ..., X_d)$. In the following, s_k^c denotes the fixed grid location of a given location $k \in \{1, ..., d\}$. As highlighted by Yadav et al. (2023), after aggregation, the dataset is no longer a spatial point process but instead becomes an areal dataset. In the following I will distinguish between the original spatial point process data and the transformed areal dataset.

Remark on temporal aggregation: The constructed areal data for a specified grid resolution can be, if needed, further aggregated across time. We considered in our example n daily values but the spatially aggregated data can be further aggregated across time if one is, for instance, interested in the dependency of fires on a weekly or monthly basis rather than daily. For example, Cisneros, Richards, et al. (2024) aggregate fire burned land per month on a spatial grid. Thus, a similar approach can be taken here.

3.2.3. Graphical Modeling

Since the magnitude marks in the original point process setting described above are assumed to follow a generalized Pareto distribution, it can be reasonable to assume that the underlying tail distribution in $\mathbf{X}$ follow a multivariate Pareto distribution. In fact, $\mathbf{X}$ consists of a lot of zeros (induced by the inclusion of 0-events) and of the positive original marks which are assumed to follow a generalized Pareto distribution. Thus, modeling the tail of $\mathbf{X}$ captures extremes not only in terms of high magnitudes (i.e., high positive marks) but also in the sense of rare in time, as the 0-events reflect periods with no recorded activity. Depending on the chosen threshold, small to moderate fires can be included.

All in all, for an appropriately high threshold, such that the positive marks in $\mathbf{X}$ are considered[3], we can fit a Hüsler-Reiss graphical model (Engelke and Hitz 2020) using the learning algorithm developed by (Engelke, Lalancette, and Volgushev 2021). The latter enables the estimation of a variogram matrix $\widehat{\Gamma}$ which fits the learned graphical structure (Hentschel, Engelke, and Segers 2024). Then, one can derive an estimate $\widehat{\Theta}$ the precision matrix Θ in (3.11), which encodes the graphical structure. In the following, $\mathbf{Y} = (Y_1, ..., Y_d)$ represents the standardized threshold exceedances of $\mathbf{X} = (X_1, ..., X_d)$.

I propose to further spatially scale the precision matrix $\hat{\Theta}$ of the estimated Hüsler-Reiss graphical model to incorporate spatial information into the estimated graph. A maximum distance

[3] If, for a location on the grid, too little positive events are recorded, it can be removed from the modeling process.

$dist^{MAX}$ for an accepted dependency can be estimated using tools specifically designed for the original spatial point process. In particular, one can use the default maximum distance considered in the estimation of the K-function by the `Kest()` command in R Package `spatstat` (Baddeley and Turner 2005). Recall that the K-function introduced in Chapter 2 provides information on the clustering behavior of an underlying point process. Applying this maximum distance allows information from the original spatial point process to be integrated into the graph estimated from the aggregated areal data.

The missing edges can then be identified in a rescaled matrix, which I define as follows:

$$\widehat{\Theta}^s = W \circ \widehat{\Theta} \tag{3.14}$$

where $\circ$ is an element-wise multiplication and W is a $d \times d$ matrix with following binary entries:

$$W_{ij} = \begin{cases} 0 & \text{if dist}(\mathrm{s}_i^c, s_j^c) > dist^{MAX} \\ 1 & \text{if dist}(\mathrm{s}_i^c, s_j^c) \leq dist^{MAX} \end{cases} \tag{3.15}$$

for $i, j \in \{1, ..., d\}$. The entries of $\widehat{\Theta}^s$ are then:

$$\widehat{\Theta}^s_{ij} = W_{ij} * \widehat{\Theta}_{ij} \tag{3.16}$$

The superscript s highlights the modified precision matrix through spatial scaling. The final decision, on whether to integrate an edge between two components of $\mathbf{Y}$ can thus be adjusted to the following:

$$(i, j) \notin E \iff \Theta^s_{ij} = W_{ij} * \Theta_{ij} = 0 \tag{3.17}$$

Here, I impose an additional constraint on conditional dependencies by excluding edges in the graph for points that are sufficiently far away in a geographic sense. Compared to the graphical model induced by $\widehat{\Theta}$, this modified graph using $\widehat{\Theta}^s$ is expected to be sparser, with fewer edges.

3.3. Advantages and Challenges

The proposed methodology offers several advantages compared to existing approaches for modeling spatial dependencies in extreme events. First, like Engelke and Hitz (2020) and Koh et al. (2023), this methodology adopts a threshold-based perspective, which is a well-established and effective tool for modeling extremes. It also aligns with the recent recommendations of Huser, Opitz, and J. Wadsworth (2024), of adopting threshold approaches when possible.

Another advantage of the suggested approach is its ease of use. In fact, unlike the complex modeling set-up proposed by Koh et al. (2023), this approach does not rely on additional covariates, making it more accessible for scenarios where latent data may be unavailable. In addition, in comparison, the graphical modeling framework simplifies the modeling process by avoiding the intricate hierarchical structures of Bayesian models for point processes. Also, the sparsity induced by the graphical model simplifies dependency structures, making the results easier to interpret. This is a key strength of graphical models in general and is maintained in the proposed approach.

Furthermore, by incorporating spatial constraints, the proposed extension aligns with the recommendations of incorporating physical knowledge when modeling extremes in environmental applications (Huser, Opitz, and J. Wadsworth 2024). This extension of the Hüsler-Reiss graphical models includes spatial considerations directly into the conditional dependency structure. Even though the distance used is based on statistical results rather than exact physical knowledge, it sill reflects realistic spatial interactions and ensures the captured dependency remains geographically plausible. It is, however, possible to adjust the spatial scaling rule to specific applications where expert knowledge is available.

Despite its advantages, the methodology also presents certain challenges. First, extreme events are, by definition, rare, which may lead to limited data for estimating the variogram, particularly in high-dimensional spatial grids. In addition, the accuracy of the proposed method strongly relies on the assumption that the underlying spatial extremes follow a Hüsler-Reiss multivariate Pareto distribution. Deviations from this assumption could impact the validity of the results. Thus, selecting an appropriate threshold can be challenging and may influence the results. It is crucial to find a good balance between large enough sample size (i.e. lower threshold) while taking into account the assumptions on the tail distribution.

Moreover, a difficulty can arise when selecting meaningful spatial constraints for the dependencies. The latter can have a big impact on the estimated conditional dependency structures as entries in the precision matrix are directly adjusted. All in all, this can require domain-specific knowledge and additional statistical analysis to validate the chosen constraint.

A final challenge is the potential computational complexity. In fact, the number of geographical locations in environmental applications can be quite large. The estimation of variogram matrices and precision matrices can thus become computationally intensive when dealing with large spatial

grids.

3.4. Estimation

Having laid out the theoretical framework and the proposed approach, we now turn to the practical implementation which is done using the `R` package `graphicalExtremes` (Engelke, Hitz, et al. 2024). The latter provides with all the essential tools and functions for graph-based modeling of extremes. It also includes all transformations between Γ, $\Theta^{(k)}$ and χ.

This section is structured as follows: The first step in the estimation process is the standardization of the data and the threshold selection for the exceedances. This is important because we want to go from the data as in $\mathbf{X}$, to standardized threshold exceedances $\mathbf{Y}$. The second step is the empirical estimation of the variogram from the data. Using the transformations detailed earlier, this estimated variogram can be used to derive the precision matrix and, thus, the graphical structure. Following, I will summarize the Hüsler-Reiss graph learning algorithm developed by Engelke, Lalancette, and Volgushev (2021) which uses graphical lasso and nearest neighbor methods, to learn a sparse graph from the data. Finally, I will briefly present how the resulting estimated sparse graph can be used to retrieve a variogram associated with that graph, as detailed in Hentschel, Engelke, and Segers (2024).

Standardization and Threshold Selection

The data is first standardized to a multivariate Pareto scale to ensure comparability across locations and consistency with extreme value theory. This standardization is done via the command `data2mpareto()` and involves the following steps (Engelke, Hitz, et al. 2024): First, for a chosen threshold based on quantiles (e.g. the 90% quantile), the data is standardized column-wise. Each column of the matrix (here representing a location) is standardized to standard Pareto distributions, a special form of the generalized Pareto distribution. Following, the standardized data is filtered. Observations where all components fall below the specified threshold quantile are removed. These will not contribute to the analysis of extremes. Finally, the observations left are normalized by dividing them by their respective quantile threshold.

Given the matrix $\mathbf{X}$ of dimension $n \times d$ as introduced above, this process results in a $n_e \times d$ matrix $\mathbf{Y}$ where $n_e \leq n$ which represents the standardized extreme values, filtered and rescaled, for d locations. Let $\mathbf{Y}$ be in the following, standardized as such, meaning that 1 is the threshold for all components. Note that, the notion of extreme is thus different for each location.

Empirical Variogram Estimation

Engelke and Volgushev (2022) define the variogram $\Gamma^{(k)}$ for each root node $k \in V$ as follows:

$$\Gamma_{ij}^{(k)} = \mathbb{V}ar\{logY_i^{(k)} - logY_j^{(k)}\}, \;\; i,j \in V \tag{3.18}$$

This variogram corresponds to the variance of the log difference between components of $\mathbf{Y}^{(k)} = (\mathbf{Y}|Y_k > 1)$. The latter is the data conditioned on the k-th component exceeding a threshold. For the Hüsler Reiss distribution, the authors additionnally note that:

$$\Gamma_{ij} = \Gamma_{ij}^k, \quad for\ any\ k \in \{1,..,d\}\ and\ i,j \in V \tag{3.19}$$

This means, that the variogram parametrizing the Hüsler-Reiss distribution can be derived from the variogram rooted at any node. To estimate this variogram, the authors introduces the following estimator, as a function of the empirical distribution $\tilde{F}_i$ of the underlying data $X_i \in \mathbf{X}$:

$$\widehat{\Gamma}_{ij}^{(k)} := \widehat{Var}\Big(log\Big(1 - \tilde{F}_i(X_{ti})\Big) - log\Big(1 - \tilde{F}_j(X_{tj})\Big) : \tilde{F}_k(X_{tk}) > 1 - \frac{n_k}{n}\Big) \tag{3.20}$$

Here, $\widehat{Var}()$ is the sample variance and the sample size n_k corresponds to the number of extreme observations of the k-th component. The condition $\tilde{F}_k(X_{tk}) > 1 - \frac{n_k}{n}$ ensures only observations corresponding to extreme values of the k-th component are used.

In practice, the function `emp_vario()` of the package `graphicalExtremes` (Engelke, Hitz, et al. 2024) calculates an empirical estimate of the variogram and is used in the following learning algorithm.

Learning Algorithm

Engelke, Lalancette, and Volgushev (2021) introduce an algorithm `Eglearn` which is also implemented in the `R` package `graphicalExtremes` (Engelke, Hitz, et al. 2024). It aims at learning the graphical structure of a Hüsler-Reiss model through L^1 penalization. The detailed algorithm set-up can be found in the publication and the following aims at summarizing the main idea. The starting point of their algorithm consists in estimating, from an estimate of the variogram, the covariance $\widehat{\Sigma}^{(k)}$. This is done by replacing Γ with $\widehat{\Gamma}$ in the transformation (3.4). Then, the edge set $\hat{E}$ is estimated using what the authors call a "base learner algorithm" which estimates for sparse patterns of the inverse of $\widehat{\Sigma}^{(k)}$. Sparsity is enforced via a penalty on the absolute values of the entries in the precision matrix $(\widehat{\Sigma}^{(k)})^{-1}$, effectively shrinking the entries. The strength of the penalization is tuned using a parameter $\rho > 0$. Higher values imply stronger penalization i.e more sparsity. The base learner returns a matrix $\widehat{Z}^{(k)}$ which encapsulates the sparsity pattern of

$(\widehat{\Sigma}^{(k)})^{-1}$. This process is iterated for each k, resulting in a collection of d matrices $\tilde{Z}^{(k)}$. The latter are d-dimensional matrices corresponding to the zero augmented matrices $\widehat{Z}^{(k)}$. In particular, the entries of each $\tilde{Z}^{(k)}$ can then be seen as weights or votes for or against an edge. The decision on whether to include an edge in $\hat{E}$ is based on this collection.

The authors introduce two "base learner algorithms" which are the neighborhood selection (NS) and graphical lasso (glasso) although the neighborhood selection worked best in their simulations. The authors cite Meinshausen and Bühlmann (2006) and Friedman, Hastie, and Tibshirani (2008) as original sources respectively.

Matrix Completion and Model Selection

Although the learning algorithm described above identifies a sparse graph structure for a given tuning parameter ρ, it does not provide an associated variogram matrix that aligns with the graph. To address this limitation, Hentschel, Engelke, and Segers (2024) developed a theoretical framework for estimating the corresponding variogram and is directly implemented in `Eglearn` in the `R` package `graphicalExtremes` (Engelke, Hitz, et al. 2024). In my understanding, following the work of Hentschel, Engelke, and Segers (2024), for a given estimated graph and an original variogram estimate, a variogram which fits the graph can be estimated. The latter is composed of the entries of the original variogram estimate for the existing edges and for the entries corresponding to the non-existing edges, the entries are estimated such that the corresponding precision matrix entry would be zero (see also the function documentation in Engelke, Hitz, et al. (2024)).

This result has important use for Engelke, Lalancette, and Volgushev (2021), as the latter can be used for model selection. In fact, the authors mention that, typically, graphs are learned for different penalization parameters ρ resulting in different levels of sparsity in the graphical structure. They denote the precision matrix associated to the estimated graph as $\widehat{\Theta}^{\hat{E}}$. The selection of the "best" model can then be done minimizing infromation criteria AIC and BIC which stand for Akaike and Bayesian information criteria respectively. In the context of graph learning (Engelke, Lalancette, and Volgushev 2021; Engelke, Hitz, et al. 2024):

$$AIC = -L(\widehat{\Theta}^{\hat{E}}) + 2|\hat{E}| \tag{3.21}$$

$$BIC = -L(\widehat{\Theta}^{\hat{E}}) + log(n_e)|\hat{E}| \tag{3.22}$$

Here, $L(\widehat{\Theta}^{\hat{E}})$ represents the log-likelihood of the estimated precision matrix, $|\hat{E}|$ is the number of edges in the graph (graph size), and n_e is the number of extreme observations. Both criteria balance goodness-of-fit via the log-likelihood and model complexity in terms of graph size.

In summary, the `Eglearn` algorithm estimates a graph for a given variogram estimate and penalization parameter ρ, which tunes the amount of sparsity to be induced. For a series of penalization parameters and corresponding series of estimated graphs, using the matrix completion problem detailed in Hentschel, Engelke, and Segers (2024), the best model fit can be chosen minimizing a chosen information criteria. The suggested additional spatial scaling can then be performed on the best precision matrix estimate.

4. Simulation Exercise

4.1. General Set Up

Is the graphical model for extremes introduced by Engelke and Hitz (2020) capable of detecting dependencies of spatially distributed extreme events ? The simulation exercise in this chapter aims at answering this question. For simplicity, I will directly simulate the aggregated areal data on a spatial grid with an underlying dependency structure. To do so, I set up 3 different simulation scenarios, starting with a very simple spatial dependency structure and increasing in every step the complexity. A hypothetical case, where a relatively clear dependency structure can be modeled, is the case of a spatial spread over time and serves as starting point of this simulation. For example, a fire might occur on day one in a location, and a next day a fire is recorded in neighboring location. In the following, I will refer to such spatial spread as a spatial epidemic. All scenarios are simulated using a common set-up described visually in figure A.1 in the Appendix. The simulation were performed using the statistical software R (R Core Team 2022) and all code can be found in the supplementary material. In addition, Appendix D contains the references of all packages used for this present chapter as well as for Chapter 5.

Grid and Trigger Locations

The idea is to start with a simple square grid of dimension $d \times d$ representing a 2-dimensional geographic space. Depending on the case, I will consider a 10x10 or 20x20 grid. Each grid cell represents a geographic location and is characterized by its centroid using vertical (i) and horizontal (j) placement. In total, the process counts $D = d \times d$ distinct locations. On this grid, I determine possible locations for an event to first occur to which I will refer as trigger locations. Depending on the scenario, a fixed set of trigger locations is randomly sampled. Thus, the trigger locations represent the possible start of an event that will spread with each time iteration to neighboring locations. The set S of all locations on the grid is then:

$$S = \{(i, j);\ \ i, j \in (1, ..., d)\} \tag{4.1}$$

and the set of all trigger locations L be randomly sampled from S:

$$L \subset S \tag{4.2}$$

Time Iterations and Triggered Events

The simulation of events on the grid will follow simple daily iterations i.e. simple time units. Each simulation will iterate over $N = 10000$ days, which corresponds to a little over 27 years. I define a blank *simulation matrix* of dimension $N \times D$ for which, every iteration, a row is filled. Each column represents a location in the process as described above.

The probability of an events to be triggered at time t is 0.2 and is modeled using a Bernoulli distribution. Formally, let R be a binary random variable indicating if an event is triggered. Then,

$$R \sim Bernoulli(0.2) \tag{4.3}$$

In general, only one event can be triggered per day t. This means that if an event is triggered, a random location from the subset of trigger locations L is sampled as the event start.

Spatial Epidemic and Mark Simulation

Per iteration t, every location is assigned a value. In our wildfire case, it would represent the area of land burned. Practically speaking, a row in the simulation matrix is completed. Now there are two scenarios:

- No event is triggered: In that case, all locations for which no value is assigned yet, are set to 0. This 0 mark represents the 0-events introduced earlier i.e. the absence of a fire.
- An event is triggered: the location selected as the event start will be assigned a positive mark at time t. Then, an epidemic-like structure follows, and neighboring sites will also be assigned a positive mark at future times. A direct neighbor might be assigned one at time $t + 1$ and following, their direct neighbor at time $t + 2$ and so on. This mark is sampled from a Generalized Pareto Distribution with location and scale $\mu = 1, \sigma = 1$ and shape $\xi = 0.5$ (see Figure 4.1 for a visualisation). The latter is naturally lower-bounded by 1.

In the second case, the sites affected by the event spread are determined using the distance and direction to the event start location. The distances between sites is calculated using the Euclidean distance. Let $s_1 = (s_{11}, s_{12})$ and $s_2 = (s_{21}, s_{22})$ be two different sites in S. Then, the distance between the points is[1]:

$$d(s_1, s_2) = \sqrt{(s_{11} - s_{21})^2 + (s_{12} - s_{22})^2} \tag{4.4}$$

[1]The formulation of the euclidean distance is available in Cressie (2015) p.6 for example.

First, a spread distance is sampled i.e. how far the given triggered locations will spread across the grid. The spread distance is at least 1 but cannot exceed 3. Then, the spread direction is defined inspired by the cardinal directions. Knowing how far the event spreads and in which direction, the neighboring sites affected by the event can be assigned a positive mark in a next time iteration. The locations at a distance between $(k-1, k]$ and within a specified direction will be assigned a positive value at time $t+k$.

For example, let the event spread in all directions with a distance of 2. Then,

- The event start location gets assigned a positive mark at time t
- All sites at a distance (0,1] get assigned a positive mark at time $t+1$
- All sites at a distance (1,2] get assigned a positive mark at time $t+2$

When only a subset of directions is specified, the locations within the set direction are assigned a value like above. The locations that are within distance but not within direction, are randomly assigned a value with probability 0.3.

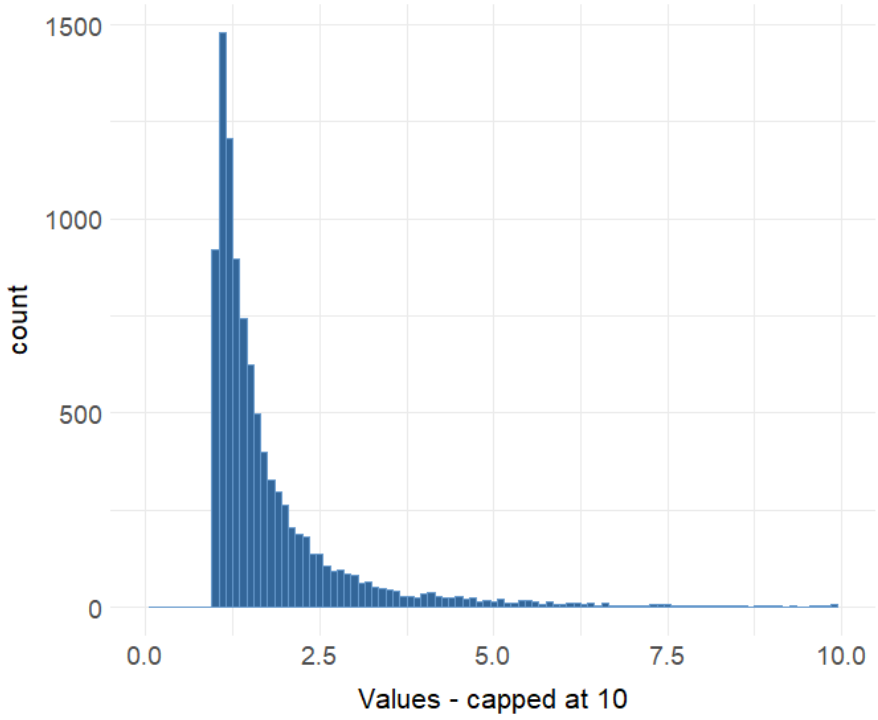
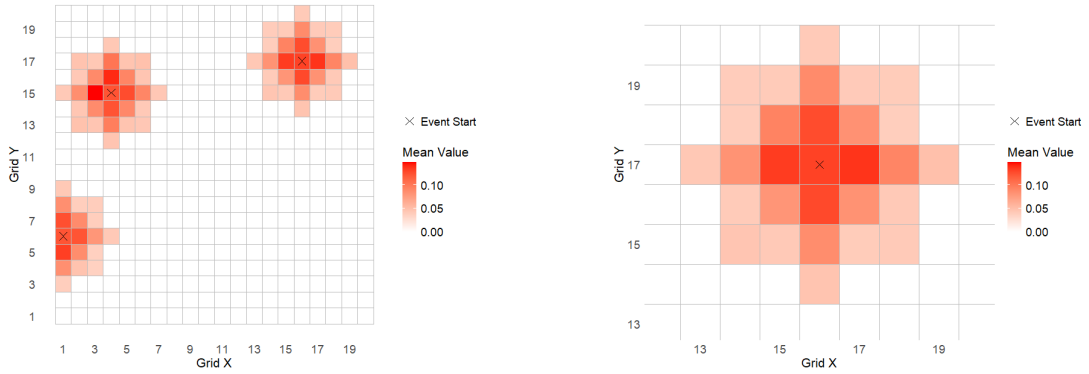


Figure 4.1.: Histogram of n=10000 sample from a GPD($\mu = 1, \sigma = 1, \xi = 0.5$). `R` Package used: `evd` package by A. G. Stephenson (2002).

4.2. Simulation Scenarios

In total, I simulate 3 different datasets, corresponding to different scenarios. Each scenario aims to test the graphical model of Engelke and Hitz (2020) and its learning algorithm (Engelke, Lalancette, and Volgushev 2021) as well as to compare it to the proposed spatially scaled graph build using the rule in (3.17) .

Scenario 1. The first scenario is fairly simple. An event spreads in all directions equally and only 3 trigger locations are used on a 20×20 grid. Here is expected a graph that shows connections across all locations near a location used as event start. The suggested graph extension should further entangle the dependencies to better recognize this spread-like structure. Figure 4.2 shows the mean values in each site for the whole grid and a subset. Naturally, some sites never experience an event and some more often than others. For example, a site at distance 3 from the event start location will less often have an event than a site at distance 2.

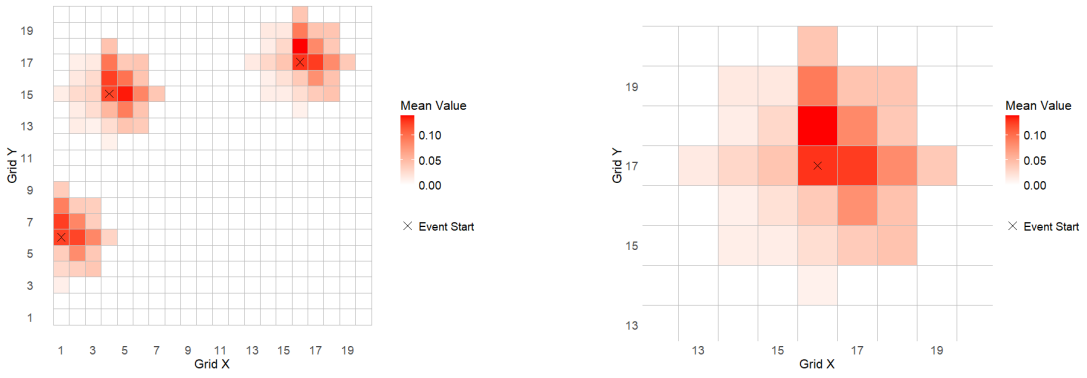


(a) (b)

Figure 4.2.: Scenario 1 - Mean values per location for (a) whole grid and (b) a grid subset.

Scenario 2. The second scenario is identical to the first one, with the difference that the spread is asymmetrical. In fact, the spread will always go to the right of the start location as depicted in figure 4.3. In this scenario, the strength of dependence left from the event start should be clearly different from the right. Although some connections on the left side could be expected as they still occur around the same time but with a less clear conditional dependency structure. The estimated graphical models should be more strongly connected on the right side.

Scenario 3. The third scenario differs from the very simple setting of scenario 1 and 2. Here, the grid size is lower (10×10) grid and 6 trigger locations are sampled. This means that there is more overlap between the different events, as this might be the case in real life. The figure 4.4 visualizes the simulated grid data by means. Similar to scenario 2, the spread is asymmetrical to the



(a) (b)

Figure 4.3.: Scenario 2 - Mean values per location for (a) whole grid and (b) a grid subset.

right for each triggered event. Here the dependency structure is less clear but the estimated graphical model should be able to identify conditional dependencies on the right of each trigger location.

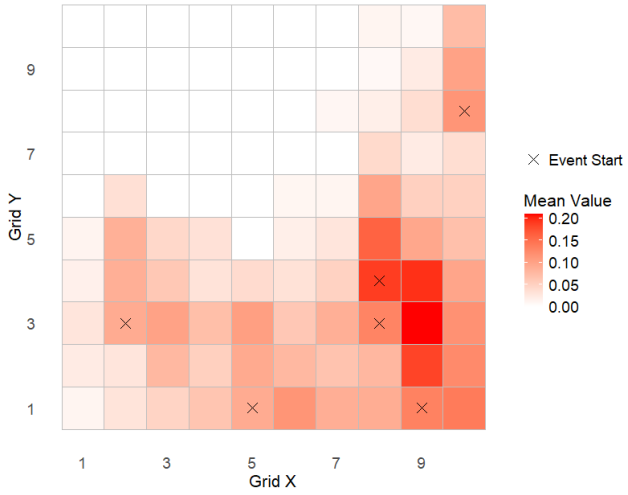


Figure 4.4.: Scenario 3 - Mean values per location

4.3. Results

For all simulation scenarios, the same estimation and pre-processing routine is performed. Specifically, the simulated data is first aggregated to weekly values, allowing the model to capture the dependency structure of weekly burned land across locations. Then, for each scenario, 2 graphs are estimated: A Hüsler-Reiss (HR) graphical model using the learning algorithm from Engelke, Lalancette, and Volgushev (2021) described in Chapter 3, minimizing BIC criteria for

a given range of penalization parameters. And, secondly, a scaled version of that graph using the maximum distance. The visualization of the BIC criteria against penalization parameters for each scenario can be found in the Appendix B. I choose to visualize the resulting graphical models similar to the visualization of Cisneros, Hazra, and Huser (2024). This means, that in all following graphical models, the placements of the nodes are forced to follow the underlying geographical space. In all cases, the presence of an edge indicates a conditional dependency. This means that if two nodes are connected, knowing that an extreme event occurred at one location provides information on the likelihood of an extreme event at a connected location. This is still to be seen with caution, as an edge connecting two nodes can pass visually through different nodes that are actually not connected. In addition, I choose, again inspired from Cisneros, Hazra, and Huser (2024), to color code the edges connecting the nodes according to their estimated extremal correlation retrieved by the transformation (3.7). This enables the joint visualization of the estimated strength of extremal correlation between two locations and their conditional dependence structure. The corresponding graphs, without color-coding and displaying only the conditional dependencies, can be found in Appendix B. All the following results and figures are derived using the `graphicalExtremes` (Engelke, Hitz, et al. 2024) package for the Hüsler-Reiss graphical model and the `spatastat` package (Baddeley and Turner 2005) for the estimation of a maximum distance. The code to reproduce the figures is provided in the supplementary material.

Overall, the evaluation of the results will mainly rely on a visual diagnosis of the graphs, as the true precision matrix is unknown. However, the simulation setting ensures a certain known pattern. To provide a unified framework for discussing the results, I focus on the following criteria:

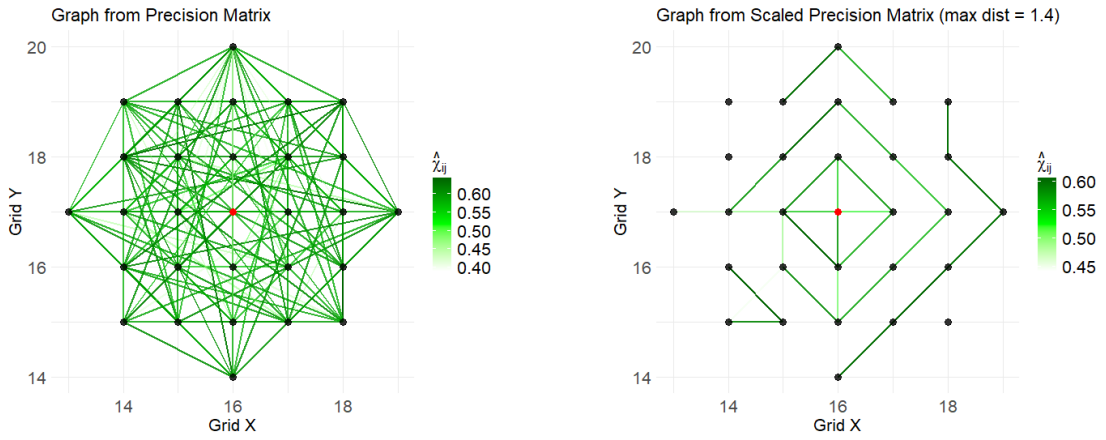
- **Model Fit**: Does the detected graph structure align with the dependency structure implied by the simulation? Does the graph identify clusters of locations that exhibit dependencies? Do the extremal correlation estimates make sense?
- **Model Sparsity**. Are the fitted graphs sparse ? Sparsity is desirable result for graphical models as it simplifies interpretation and helps avoid over-fitting.
- **Sensitivity Analysis**: The robustness of the resulting graphs is tested against key parameters. This includes an evaluation of how the graph structure varies with changes in threshold quantiles and a comparison of learning algorithms (e.g., NS versus glasso).
- **Limitations and Computational Considerations**: For each graph, the potential issues encountered during the computation process and other data limitations will be discussed if relevant.

Remark on the interpretation of an edge connecting two nodes : In the following discussion, I mention an edge connecting two locations as an indicator of fire occurrence dependence. Specifically, I interpret the presence of an edge as suggesting that the occurrence of a fire at one location is conditionally dependent on the occurrence of a fire at another location. However, this interpretation is somewhat imprecise. The graphical model is constructed based on the associated magnitude marks, meaning the conditional dependencies are estimated for these magnitudes rather than the occurrences themselves. That said, the simulation did not induce dependency in fire sizes, given that a fire occurred. In addition, the inclusion of 0-events means that the tail does not only represent extreme fire magnitudes but can also account for the occurrence of a fire itself. Therefore, I choose to frame the interpretation in terms of fire occurrence dependence. Additionally, due to the nature of the Pareto distribution in figure 4.1, most values are concentrated around a similar range, making it likely that magnitudes at two locations are comparable when events occur. Note however, that extreme in one location does not mean the same extreme in another location. The notion of extreme is to be seen individually in each of the locations.

Scenario 1 and 2 - Modeling Results

For these scenario, I decide to focus on one of the extreme event hot spots. This is motivated by the fact, that the Hüsler-Reiss graph is necessarily connected (Engelke and Hitz 2020) and the set up is such that two events ”starts” are not triggered at the same time at two different locations i.e. no dependency is to be expected. This means that we would in fact have 3 different graphs on the grid. Hence, I focus on the subset shown in figure 4.2 (b) and 4.3 (b). A total of 29 sites exhibit at least once in the process a fire and each of these sites count 1429 weekly observations. The results for the two estimated graphs, choosing a quantile threshold of 90% and using the NS based algorithm (Engelke, Lalancette, and Volgushev 2021), are displayed in figure 4.5 and 4.7 for the scenario 1 and scenario 2 respectively. I refer to the Hüsler-Reiss graphical model as the Graph A and its scaled version as Graph B.

A first observation concerns the general connectivity patterns. The Graph A of scenario 1 shows high connectivity between the nodes due to the triggered events across all locations. This results in a dense and less readable graph. Similarly, in the second scenario, the Graph A shows connections across the whole graph. If considering the extremal correlations, stronger connections are localized in the top-right cluster. These connections reflect the spread direction defined in the simulation. It is thus sensible to say that in both scenarios, the Graph A detects dependencies in line with the simulation pattern. However, it also highlights the benefit of using additional measure to interpret the strength of the depicted conditional dependencies. In fact, the directional spread in Scenario 2 is not visible when purely considering the absence / presence



(a) (b)

Figure 4.5.: Estimated HR graphical model for scenario 1 using `eglearn` algorithm in (a) (Engelke, Lalancette, and Volgushev 2021) and a scaled version in (b) using the maximum distance of the estimates obtained via `Kest()` function (Baddeley and Turner 2005). The graph in (a) minimizes BIC, using a 90% quantile threshold and the neighbor selection based algorithm. The $\hat{\chi}_{ij}$ are the estimated extremal correlations between two nodes.

of an edge (see Figure B.5). In addition, the strong connectivity across the entire Graph A in Scenario 1, shows the importance of using sparser models like Graph B, as finer dependency structures are more visible.

In both scenarios, some strong connections exist between distant nodes: Consider for example the edge connecting the node (16,20) and (17,17) in scenario 1 and 2. This is expected due to the simulated structure, where events at locations of similar distance to the triggered center, occur always simultaneously. Although this makes sense given the simulation setup, it might not be the most plausible edge in a geographic sense. A sensible question is then: is an edge representing conditional dependence between these two edges plausible ? Or does it maybe make more sense that the conditional dependence lies only at geographically neighboring locations ? Graph B resolves this by limiting connections to proximate nodes. In both scenarios, the Graph B thus improves the geographic plausibility of the edges and improving clarity, especially in Scenario 1 where edges are densely packed.

The other advantage of the additional sparsity induced by the spatial scaling in Graph B, is the detection of finer clusters. In particular, it helps detecting the spread like structure. In fact, given the graphs, one can detect a cluster of all locations which are at distance $(1, 2]$ in the first scenario. This is to be expected, as they the events occur simultaneously in this setting. Similar clusters can

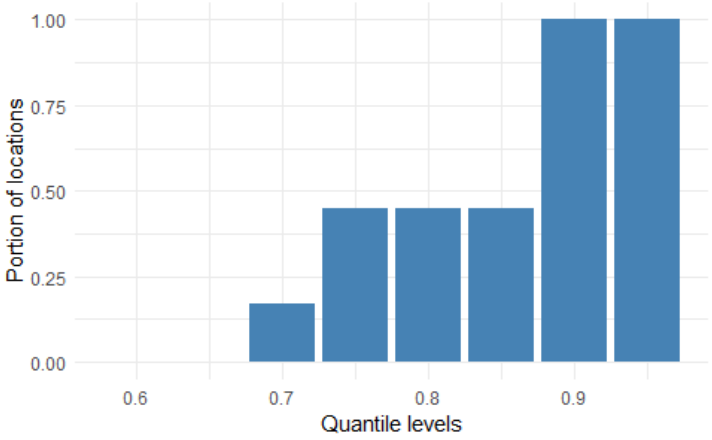
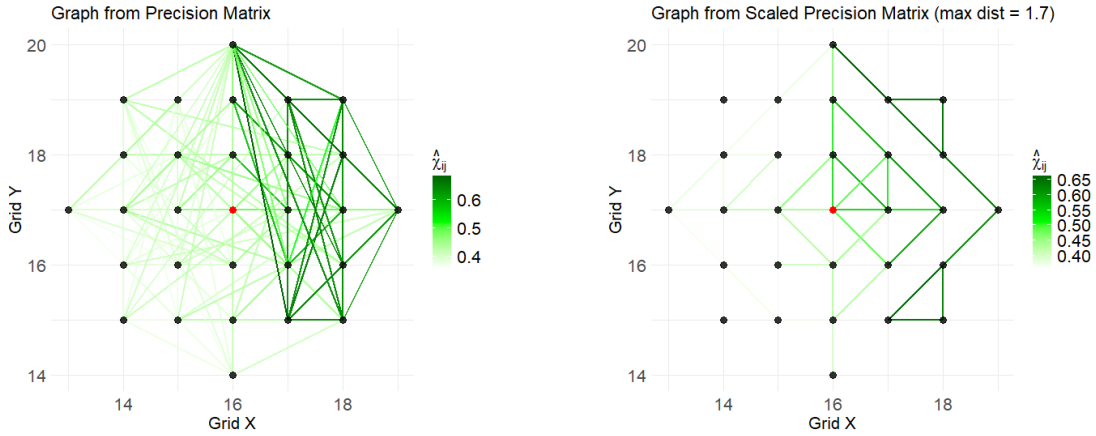


Figure 4.6.: Portion of the 29 locations in Scenario 1 with strictly positive quantile value for different levels.



(a) (b)

Figure 4.7.: Estimated HR graphical model for scenario 2 as detailed in Figure 4.5.

be observed in the second scenario. The conditional dependency here makes sense, as knowing that an event occurred in a location of this cluster suffices to predict the event occurrence at other locations of this cluster. Thus, even though they are related to the occurrence of an event in the center, if conditioning on the other locations, these are independent.

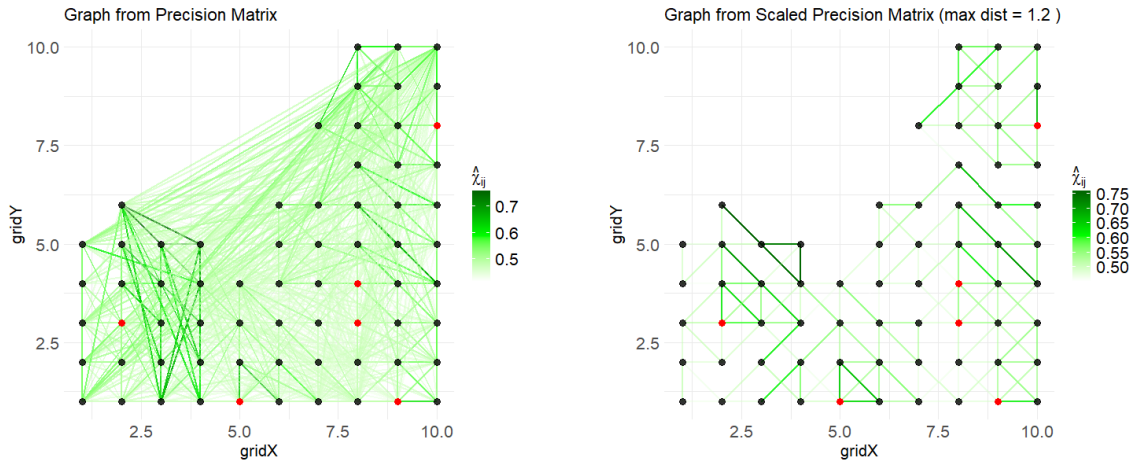
However, it is important to note that some expected edges are not present. For example, I expected a visual connection between the node (16,14) and (15,16) in the Graph B of the first scenario and it is not the case here. A potential explanation could be the limited data in the tail of these locations, as they are less often triggered. This does not fully explain it, as their tail should be similar to other parts of the graph.

The estimated Graphs A in both scenarios offer a sparser view of the overall dependency structure. In fact, a total of 406 edges are possible (number of pair combinations), but only 122 are kept in the first scenario and 121 in the second. Per design, Graphs B are even more sparse, counting 30 (41) edges for Scenario 1 (2).

The graphs obtained via glasso learner algorithm can be found in the appendix. In both scenarios the Graph A is less sparse (210 edges for scenario 1, 215 for scenario 2) compared to the one obtained with the NS selection algorithm. Visually however, it is hard to distinguish between them, at least in that representation choice, as the additional edges tend to have relatively low associated extremal correlations. In their respective scaled version (Graph B) a few additional edges are visible. For example, in scenario 1, the missing visible edge between (16,14) and (15,16) mentioned earlier is, in that case, present. When choosing a lower quantile threshold of 70%, the size of the Graphs A is slightly higher in both scenarios (145 edges for scenario 1 and 143 in the scenario 2). A lower quantile threshold allows more fire data to be included. In Scenario 1, the edges present in the center of the graph are similar using a 70 % or 90% quantile threshold, but some additional edges are visible in the 3rd row of the spread. This confirms, that the lack of connections in the first setting where selecting a threshold of 90% might be due to the lack of data for these locations for high quantiles. However, the lower thresholds should be used knowing the consequences, as the 0's in the process might be wrongly incorporated during the standardization process. The figure 4.6 illustrates for scenario 1, the number of locations with positive quantile value. We can see that for the 90% quantile, all locations have a strictly positive threshold value which ensures that only the positive magnitudes are treated as threshold exceedances. This might not be true if choosing a 70% threshold where only a little less than a quarter of the locations have a positive quantile value. Overall, the resulting graph is relatively stable across the selection algorithms and threshold selection. The graphs corresponding to the 70% threshold quantile choice and the ones based on the glasso algorithm can be found in the Appendix B.

Scenario 3 - Modeling Results

While in the first two scenarios, the triggered event starts are well separated on the grid, it is not the case in this scenario. Although the individual spread pattern remains the same to those observed in Scenario 2, the overlapping nature of the patterns introduces additional complexity. In total, this grid counts 68 locations with weekly observations that have at least one non-zero event. The two estimated graphs are shown in figure 4.8. Again, I refer to Graph A for the estimated Hüsler-Reiss graphical model and Graph B for its scaled version.



(a) (b)

Figure 4.8.: Estimated HR graphical model for scenario 3 as detailed in Figure 4.5. In addition, the red points indicate the locations of the trigger locations.

A first observation from Graph A is the dominating number of edges with relatively low associated extremal correlations. In fact, only a few strong edge connections around the event start (highlighted in red) are recognizable. However, this is still a good result, as the dependency lies exactly in these areas. The spread-like structure identified on the graphs in the second scenario is again here visible, although not completely and not everywhere. For example, around location (2,3), the upper right spread is quite visible but is not for location (5,1) where only the spread at distance 1 is identifiable.

As in previous scenarios, the scaled version in Graph B provides improved readability due to the additional sparsity introduced by the scaling. In comparison, the Graph A counts 383 edges, whereas the Graph B counts only 157 edges.

There is a lack of dependency around the event start (8,3). This is explained by the overlap of the patterns around location (8,4). The missing edges here are thus not surprising, as a fire in one of these locations occurs in two cases within this scenario, creating less clear dependency structures. For example, a fire in location (9,3) is not predictive of a fire at location (8,4) as they can occur in two different settings. This highlights a limitation of the graphical model in distinguishing overlapping dependencies in such cases.

A graphical lasso estimation led to similar visual results, but, as in the first and second scenarios, the number of edges is much higher (701 in this case). However, these additional edges have low associated extremal correlations, which explains why visually the difference almost

unnoticeable. In addition, the computation time was higher for glasso estimation in this setting. Glasso required approximately 5 minutes for estimation, whereas the NS method produced results within 1 minute[2].

Using a 70% threshold did not yield results with the NS algorithm. In fact, in this setting, the completed variogram matrix did not match the graph, which means that the matrix completion did not converge. This is problematic, since the estimation of Graph B relies on the estimate of a variogram associated to the Graph A.

4.4. Discussion

This simulation study aimed to explore the conditional dependency structures in simulated extreme events with a spread-like pattern and to test out the graphical models as a tool to detect these dependencies. The analysis tested two graphical model estimates: the Hüsler-Reiss Graphical Model estimated from the learning algorithm (Graph A) and its scaled version (Graph B), across three scenarios with varying conditions. The diagnosis of the models ability to detect the underlying spatial patterns determined during the simulation was mainly visual and led to following observations:

- In the first two scenarios, where event triggers were well seperated across the grid, Graph A exhibited dense connectivity in Scenario 1, with relatively strong extremal correlations across all locations as the spread was symmetric around the trigger. In contrast, Scenario 2 highlighted more localized connections, particularly in the top-right cluster, reflecting the specified simulated spread direction of the event. These results suggest that Graph A effectively captured the spread-like structure of the dependencies defined by the simulation, although it became harder to interpret in the densely connected graph of Scenario 1. However, the directional spread is not directly identifiable if considering only the absence/presence of an edge.

- The scaled version, Graph B, introduced sparsity that improved the clarity and geographical plausibility of the model, especially in Scenario 1. By limiting connections to geographically proximate nodes, Graph B provided a more intuitive representation of the spread, identifying finer clusters of dependent locations. This helped isolate localized patterns and facilitated better interpretation of the conditional dependencies.

- Scenario 3 introduced an additional layer of complexity by allowing event spreads to

[2]This is for estimating a graph for 10 different penalization terms (0.01, 0.03, ..., 0.19, 0.21)

overlap. Here, the graphical models revealed some limitations, particularly around the locations where spreads from different event starts intersected, although the results were still coherent.

- Graphical lasso yielded results similar to the NS method but with significantly more edges. These additional edges were associated with relatively low extremal correlations, which did not visually impact the interpretation, but instead added visual blur. Glasso also required more computation time for the specified penalization parameters (approximately 5 minutes) compared to the NS method (1 minute) in scenario 3, which might be an important consideration in some applications.

- The sensitivity of the model to threshold choice was observed, particularly when using the 70% quantile threshold. In fact, the estimation failed to retrieve an variogram associated to the graph in the third scenario. Also, using lower thresholds should be used with caution and positive quantile threshold values for the locations to model are preferable.

- Another note can be made on the visualization of the edges. I already mentioned this point but I think it is worth to emphasizing. Using this representation choice, it is possible that two locations appear conditionally dependent, even if they are in fact, not. This is due to the forced grid-like structure, resulting in seemingly connected nodes if they happen to be in the way of a connection between two other nodes.

In conclusion, the study demonstrates the effectiveness of graphical models in capturing event dependencies, with notable improvements through spatial scaling and the visualization of extremal correlations. However, challenges remain, particularly in handling overlapping event patterns. In addition, the model failed at recognizing some patterns. Further refinement of the model and consideration of alternative approaches may help address these limitations.

5. Application - Forest Fires in Spain

As mentionned in the previous chapter, all code is done using the statistical software `R` (R Core Team 2022) and all packages used are listed in Appendix D. The code for the analysis below is provided in the supplementary materials.

5.1. Data Description

The data used in this application is a marked spatial point process dataset `clmfires` by Professor Jorge Mateu, publicly available in the R Package `spatstat` (Baddeley and Turner 2005). The description provided here is based on the package documentation, which offers detailed information about the dataset (`https://rdrr.io/cran/spatstat.data/man/clmfires.html`). The dataset can be easily accessed in `R` using the command `data(clmfires)`. It contains reports of forest fires between 1998 and 2007 in the Castilla-La Mancha region in Spain. In addition to the main dataset, are included a list of images (`clmfires.extra`) illustrating underlying covariates and physical aspects (See figure C.1 in the Appendix).

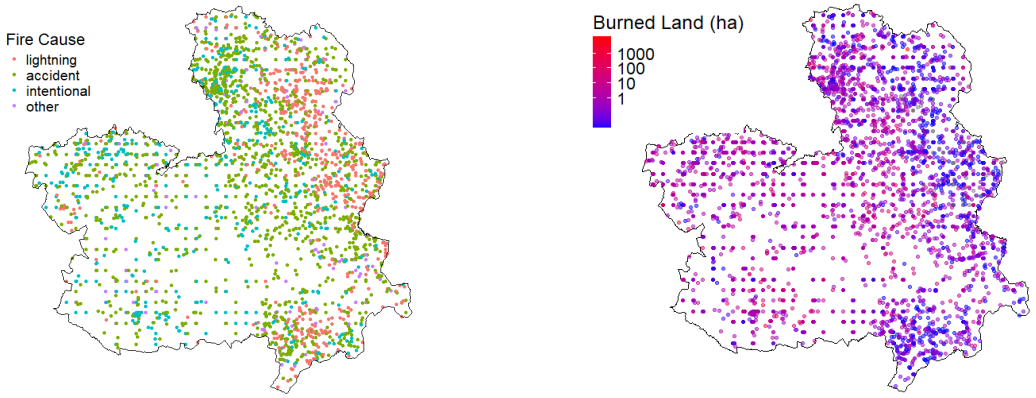
Castilla-La Mancha is a region South-East of Madrid, and is mainly flat (figures C.1 (a) and (b)) with a few mountainous areas in the North and South East. It is the second region with the highest farm holdings (Instituto National de Estadística)[1]. Nevertheless, the region is also covered in natural vegetation like forests, bushes, grass lands etc. (Image (c) in figure C.1) which makes it prone to fire risk, especially in the Southern areas and in the North.

The dataset is a point process object, containing a total of 8488 points in a 2-dimensional coordinate system, indicating the location of the fire centroids. The coordinates are recorded in kilometers, covering approximately 400x400km. Each point has 4 marks, providing additional information on each fire:

- *cause*: Categorical variable indicating the fire cause reported. This can be lightening, accident, intentional or other/unknown causes. The majority are reported as accidents.
- *burnt.area*: Numerical variable indicating the total area burned in hectares (ha)
- *date* and *julian.date* records the date of the fire start in a standard YYY-MM-DD format and in number of days since January 1st 1998.

[1] as of 2016, source: Instituto National de Estadística - `https://www.ine.es/`

All plots and figures in this chapter are based on this dataset and can be reproduced using the code in the supplementary material. For example, the plots in figure 5.1 visualize the spatial point process for two marks. In order to conduct some of the analysis below, the dataset was transformed in a data frame format and contains the same information. The latter is composed of 8488 observations with 6 columns: 4 marks and 2 x/y columns indicating the points location. In addition, I matched each fire location to its corresponding municipality and province using a list of lat- and longitudinal administrative division of Spain. The list was developed by Urquizu (2010) using Google maps and INE (Instituto National de Estadística) data (see `https://www.businessintelligence.info/varios/longitud-latitud-pueblos-espana.html`). Since the fires were recorded in a UTM coordinate system rather than standard longitude and latitude, I converted the lon/lat values to this spatial point pattern dataset system and matched them to the nearest fire location. This conversion is however approximate.



(a) (b)

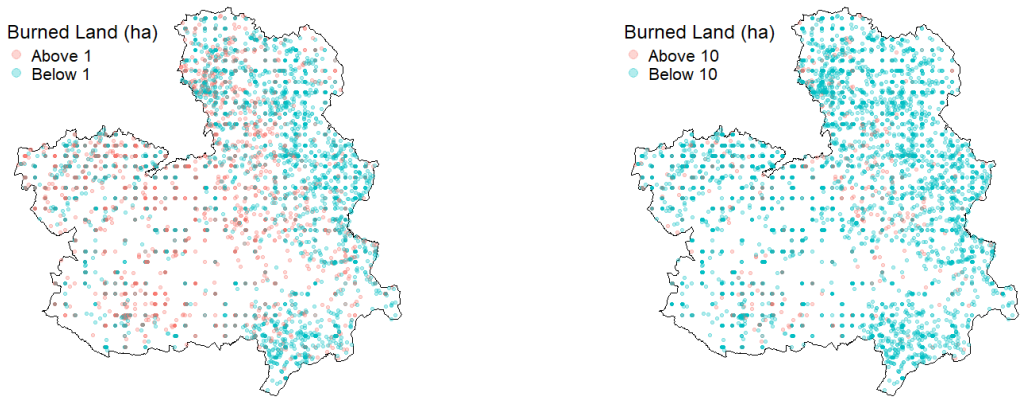
Figure 5.1.: Marked point process `clmfires` for (a) fire cause and (b) burnt area (log scale).

5.2. Preliminary Analysis

Small, Moderate and Large Fires

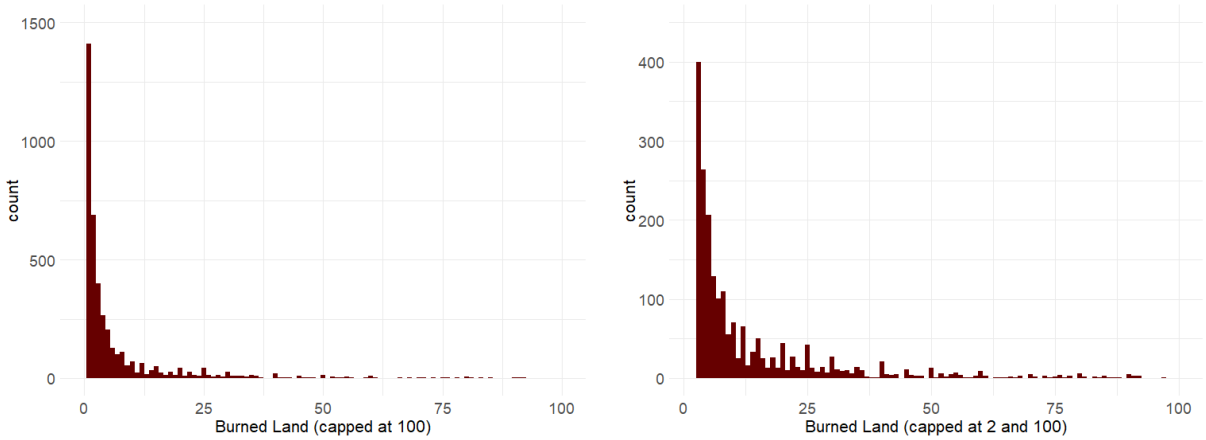
A first observation of the marked point process, is the largely dominating small fires (figure (b) 5.1). In fact, over half of the reported fires burned less than one hectare of land and nearly 90% burned less than 10 hectares (figure 5.2). This is consistent with the observations by Koh et al. (2023) who observed a large portion of small to moderate fires and only a few very large fires dominated the total burned land. The histograms of burned land in figure 5.3 further show the distribution of the data. The values were capped for better readability as the tail is very long (122

fires burned over 100ha which corresponds to 1.44% of all fires).



(a) (b)

Figure 5.2.: Marked point process `clmfires` for burnt land (a) below 1 ha and (b) below 10ha.



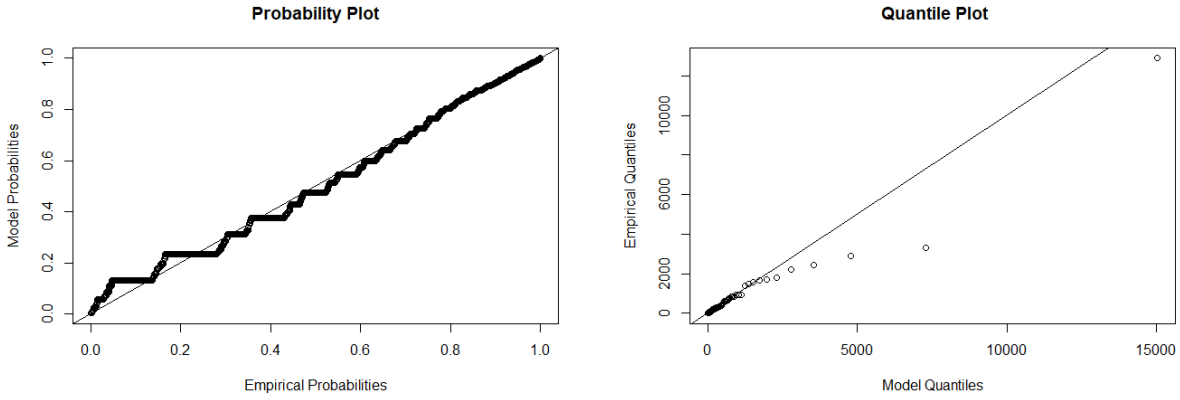
(a) (b)

Figure 5.3.: Histogram of burned area, capped at different values.

Univariate Modeling of Burned Land

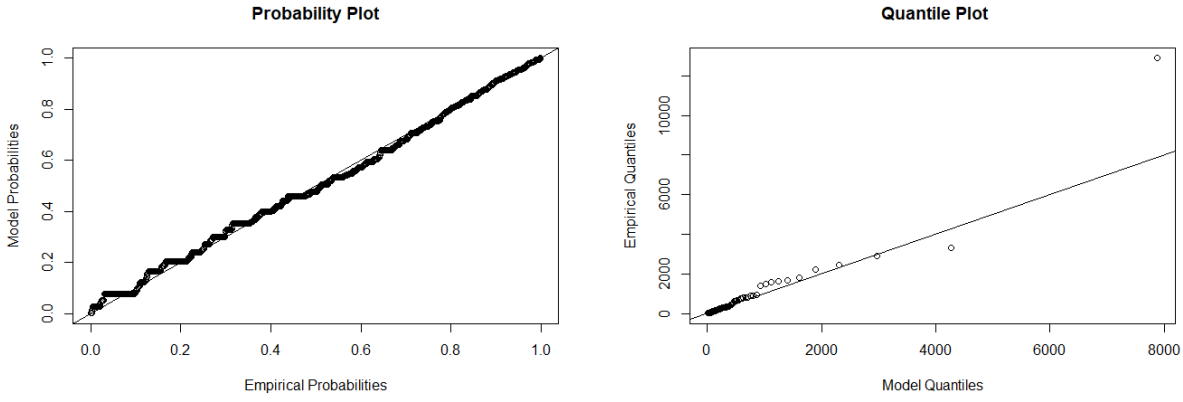
In their study, Koh et al. (2023) model extreme values of burned land using a GPD. Similarly, I fitted a GPD distribution using the `fevd()` command from the package `extRemes` (Gilleland and Katz 2016) to explore how well it can represent the data. The function fits a distribution using maximum likelihood estimates and provides with diagnostic plots, which represent the modeled vs. empirical probabilities and quantiles. Figure 5.4 and 5.5 show these plots for a GPD fit using, respectively the 60% and 95% quantile value as threshold.

In both cases, based on these plots, the fit looks pretty good for lower probabilities and quantiles. However, using the 60% quantile value as a threshold (corresponds to 1 ha), the modeled GPD overestimates higher quantiles. This is improved using the 95% quantile value (26 ha). The fit shows that the GPD provides a good approximation of the burned area in this region, especially further in the tail. It also performs reasonably well for smaller fires, particularly for lower quantile levels.



(a) (b)

Figure 5.4.: Probability (a) and Quantile (b) plot for a fitted GPD using the 60% quantile value (1ha) as threshold.



(a) (b)

Figure 5.5.: Probability (a) and Quantile (b) plot for a fitted GPD using the 95% quantile value (26ha) as threshold.

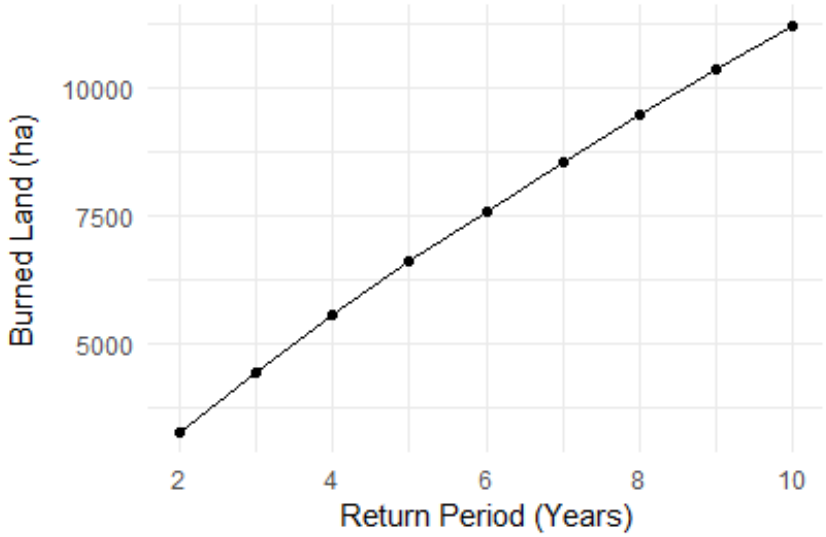


Figure 5.6.: Fitted return periods for daily aggregates of burned hectares of land over the region of Castilla-La Mancha - based on GPD fit with threshold of 138.6 ha.

Return Levels

In the previous section, univariate GPD were fitted to the tail of burned area associated to individual fires. In order to estimate return values, I fitted a GPD on aggregated daily burned area over the whole region of Castilla La-Mancha. For days without recorded fires, I included zero values to ensure continuity in the dataset. I fitted a GPD using the 95% quantile value as threshold, which corresponds to approximately 140 ha of burned land. The corresponding probability and quantile plots can be found in the appendix. Figure 5.6 visualizes the corresponding estimated return levels. For example, the model suggests that a daily burned area of approximately 3,000 hectares is expected to occur once every two years in this region. This gives us only an idea on the frequency of major wildfires, measured on a daily level. However, this statistic only provides information on a regional level and doesn't say where it might occur. Similar estimations can be done on more localized regions.

Time Development of Fire Activity and Seasonality

The dataset enables to analyse changes in intensity and frequency of fires over almost 10 years (Figure 5.7 (b)). There seems to be a consistent increase in the number of fires until 2004, after which the number of fires is declining. No clear trend appears for the total burned land per year, even though the largest peak is observed in 2005 with over 20000 burned hectares. Figure 5.7 (b) shows the frequency of wildfires per month from January to December and, consistent with the application of Koh et al. (2023), most of the fire activity is recorded in the summer months.

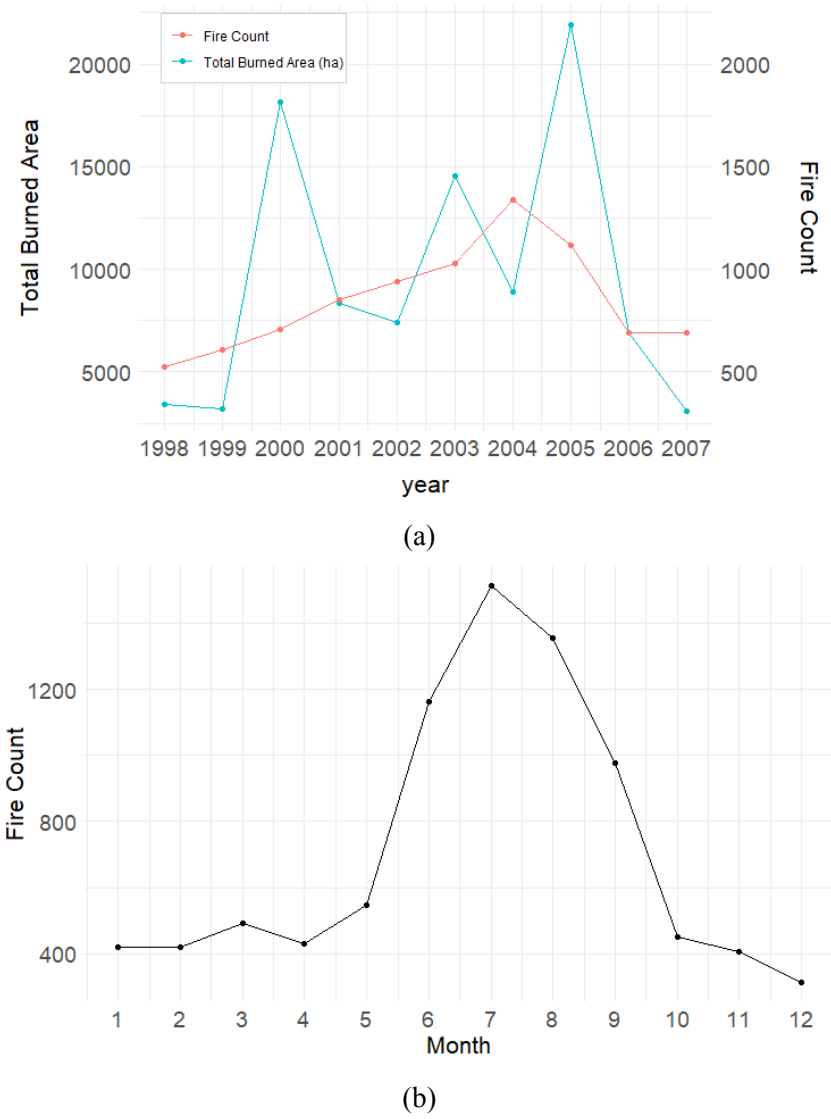


Figure 5.7.: Fire Activity over the years (a) and over the month (b).

K-function

The package `spatstat` enables the estimation of the reduced second moment function as introduced in Chapter 2 (2.8). As mentioned, the estimates can be first indicators for clustering behavior between points. Figure 5.8 shows the results of a 'best' K-function estimate[2] for the northern region of Castilla-La Mancha. This sub-region appeared in previous figures to be the one with most fire activity. The subset of considered points are visualized in Figure C.3 in the appendix. Clearly, the deviations from the random Poisson point process suggest clustering behavior between the fire's locations in this part of the region.

[2] According to the `Kest()` function documentation in `R`, this corresponds to Ripley's isotropic correction. More information can be found in Baddeley and Turner (2005)

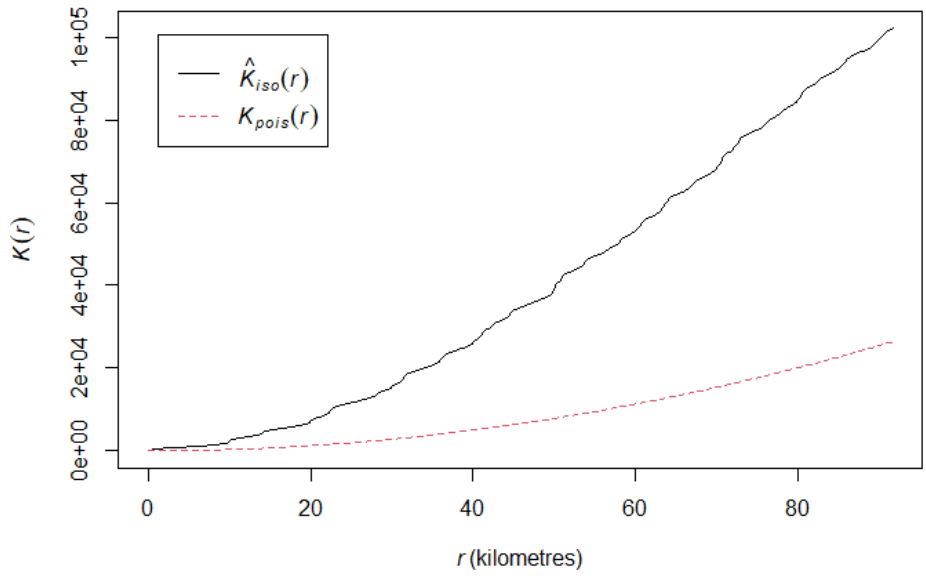


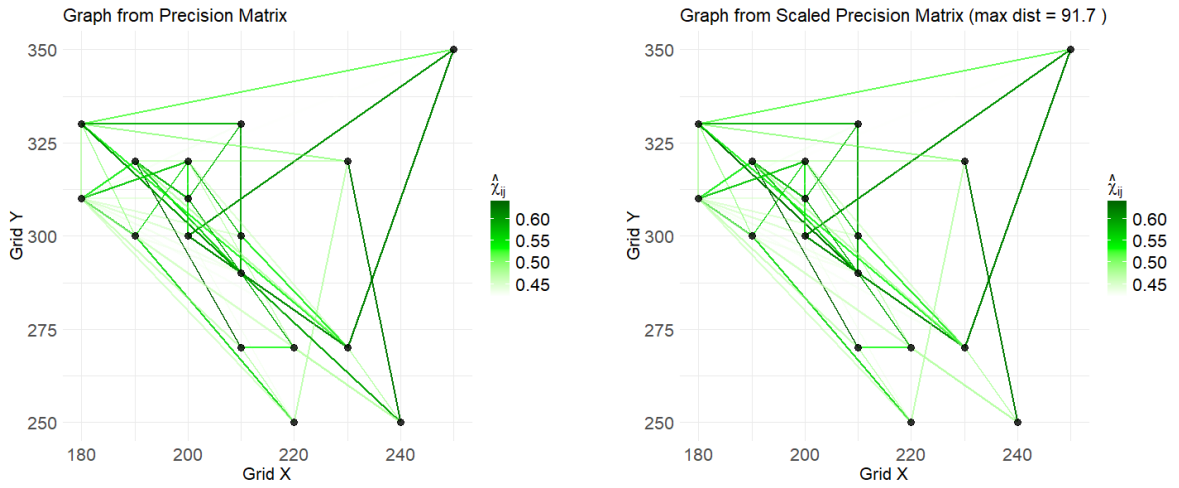
Figure 5.8.: Estimate of the K function against true K values of a random Poisson point process. The 'best' fit is chosen, which corresponds to Ripley's isotropic edge correction (see Baddeley and Turner (2005) for more details)

5.3. Graphical Modeling

For the graphical modeling, I choose to focus on the northern region depicted in Figure C.3. First, the values are, as described in Chapter 3, spatially aggregated to a lower grid resolution and the 0-events are added. I choose to use regular 10x10km grid cells, which results in a total of 181 distinct locations. I further aggregate the values monthly, such that the interest lies in the dependency between monthly burned area across the different locations. I also only use the summer month (from May to October) where fire activity is the highest. In total, 60 monthly values per location are considered. The number of locations is quite high and most of them only experience a few months with fires across the entire 9 years (Figure C.4). Therefore, I choose to fit the graphical models only on the locations, which have, for a given quantile threshold, an empirical quantile value above 1ha. This is also motivated by the study of Koh et al. (2023) who modeled fires from 1ha of burned land on. I choose to focus on a threshold quantile of 90%, for which 17 locations have a 90% quantile value above 1.

5.3.1. Results and Discussion

The estimated graphical model, minimizing the BIC criteria and using the neighbor selection based algorithm, as well as the scaled version of the graph, using the maximum distance considered in the K-function, are shown in Figure 5.9. To better relate to the actual underlying locations, Figure 5.10 provides a visualization of the scaled graph, mapped to the nearest city. Note that from the 181 original distinct grid locations in this area, only 17 have a high enough quantile threshold value. These 17 locations are all concentrated in the western part of the considered region.



(a) (b)

Figure 5.9.: HR graphical model (a) and scaled version (b) with extremal correlation estimates between locations connected by an edge

My general observations of the graphs are the following: First, the spatial scaling did not impact the graph here as the maximum distance used in the K-function was quite high (91.7 km). Second, the associated estimated extremal correlations are rather moderate than high. For its interpretation, the graph in Figure 5.10 can be read in the following way. Consider, for example, the edge connecting Huete to Barriopedro. The monthly burned land in near Huete is, given the fires around the other cities, dependent of the fires near Barriopedro. This would suggest that a fire might occur within the same month near both of the cities. However, I do question the reliability of this graphical model on this data and this for several reasons:

- The Hüsler-Reiss graphical model is, as introduced in Chapter 3, necessarily connected even though this might not be the case here. This means, that locations which are independent from all the other locations cannot be separated from the graph. However, while the

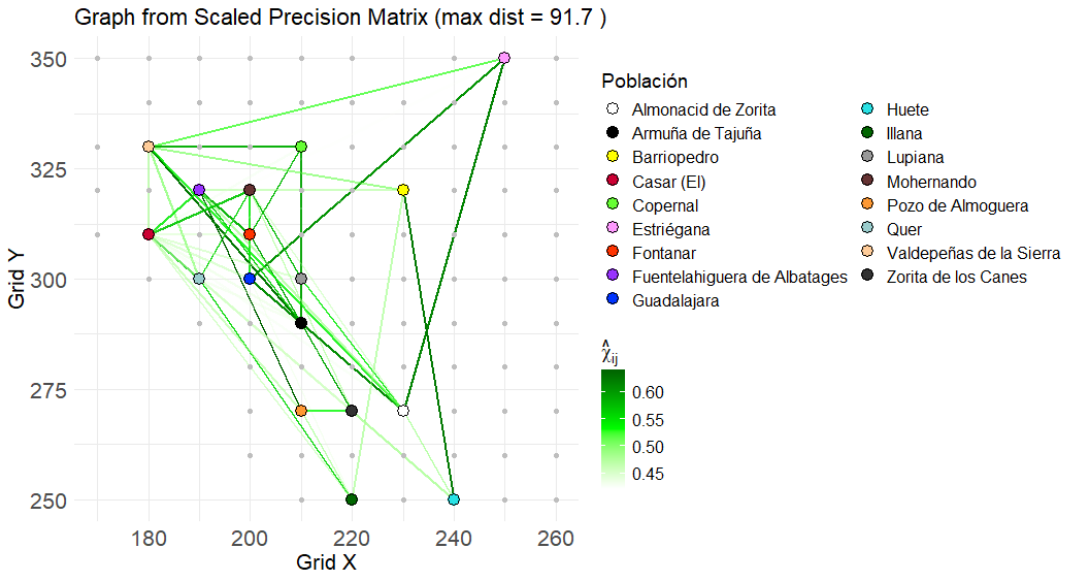


Figure 5.10.: Scaled HR graphical model, mapped to nearest city, with extremal correlation estimates between locations connected by an edge

variogram and other parameter matrices associated to the graph can then not be estimated, the learning algorithm `Eglearn` should still be able to estimate such graphical structure.

- The graphical model aims at detecting sparser dependency structures in the tails, which is useful when the dependencies are high and complex between the locations. This is, however, not observed in the data at first. For example, simple scatterplots of the monthly burned land between two seemingly dependent locations on the graph, do not show clear strong correlated behavior (see examples in Figure C.5). In fact, these scatterplots show a few simultaneous positive monthly burned areas rather than consistent correlations. The reason as to why the model still estimated dependencies could be the following: if two locations have simultaneous positive burned land even only once, the extrapolation might use it as an indicator of correlated behavior, despite its absence in most months.

- On a last note, the analysis was performed only on 60 monthly values per location, and from those, only 42 are kept after standardization to multivariate Pareto scale. As a comparison, nearly 600 observation of threshold exceedances were used in the simulation (Scenario 2). Thus, this data limitation might result in unreliable model estimation.

Overall, based on the results and analysis, my reflections on the use graphical models in this context are the following. First, graphical models mainly make sense when applied to datasets where the dependency structure is high and dense. If the data lacks preliminary indications of dependencies, their utility is limited. Second, a blind reliance on graphical models can lead to

unreliable interpretations. Thus, it is important to validate plausibility of these models through additional exploratory analysis, ensuring that the estimated dependencies align with the data. Finally, the model does not incorporate additional covariates, such as precipitation, temperature or other relevant factors. While it has the advantage of requiring less data, it also weakens the support for the results. The inclusion of additional covariates could help provide more reliable interpretations of the conditional dependencies depicted on the graph.

6. Concluding Remarks and Outlook

To conclude, this thesis focused on the graphical modeling of spatial point processes using methods from extreme value theory. The example of wildfires, characterized by their location in space, with ignition day and magnitude marks (burned land), was used. Burned land associated with wildfires has previously been modeled using the Generalized Pareto Distribution (Koh et al. 2023) which motivates the use of extremal distributions. First I introduced the relevant background in spatial point processes, extreme value theory and graphical modeling. In particular, I discussed the graphical modeling framework for multivariate Hüsler-Reiss Pareto distributions of Engelke and Hitz (2020). In fact, the authors show that the conditional dependency structure of this distribution class can conveniently be read of zero entries of a suited precision matrix. The objective was then to investigate how spatial dependency structures in wildfire burned areas could be modeled using this framework. In addition, I proposed the incorporation of a spatial scaling by defining a maximum distance, derived from point process methods, to limit the conditional dependency allowed between two locations. I evaluated both approaches using a simulation study and assessed their applicability to real-world wildfire data. The main findings are the following.

The simulation study demonstrated that the Hüsler-Reiss graphical model produced plausible graphs when combined with joint visualization of edges and extremal correlations between locations. The scaled graph enabled the detection of finer dependency structures and improved overall readability. However, the application of the Hüsler-Reiss graphical models to the spatial point pattern dataset `clmfires` of the package `spatstat` (Baddeley and Turner 2005) revealed clear limitations. In particular, it suggested that the extrapolation process may overestimate the dependency structures. Additionally, the HR models are necessarily connected, which means that the corresponding parameter matrices can only be retrieved if the estimated graph is connected. Furthermore, the spatial and temporal aggregation of the spatial point process to lattice-like data may result in information loss. Finally, the visualization approach used in this study does not offer a clear distinction of edges that pass through intermediate nodes. This creates a chain-like visualization which could result in misleading interpretation of connections between two nodes.

Overall, this thesis can highlight several directions for future research in the field of graphical modeling of spatial extremes. For example, one could think of extending the learning algorithm developed by Engelke, Lalancette, and Volgushev (2021), directly incorporating spatial information as opposed to the scaling after estimation. Distance between locations can be used as weights on the decision to incorporate an edge or not. In addition, the suggested graphical modeling framework does not address potential asymptotic independence. In particular, Huser

and J. L. Wadsworth (2019) discuss how spatial dependencies can become more weak as events become more extreme. This results in asymptotic independence which is not considered in models based on generalized Pareto processes. Thus, further research in that context could be interesting. Another area for future study, already identified by Engelke, Lalancette, and Volgushev (2021), is to quantify the uncertainty associated with an estimated graph. In fact, in this thesis's application, no uncertainty measure was associated to the estimated graphical structure. This can lead to erroneous conclusions if the fit is not properly assessed.

A. Simulation Set-Up

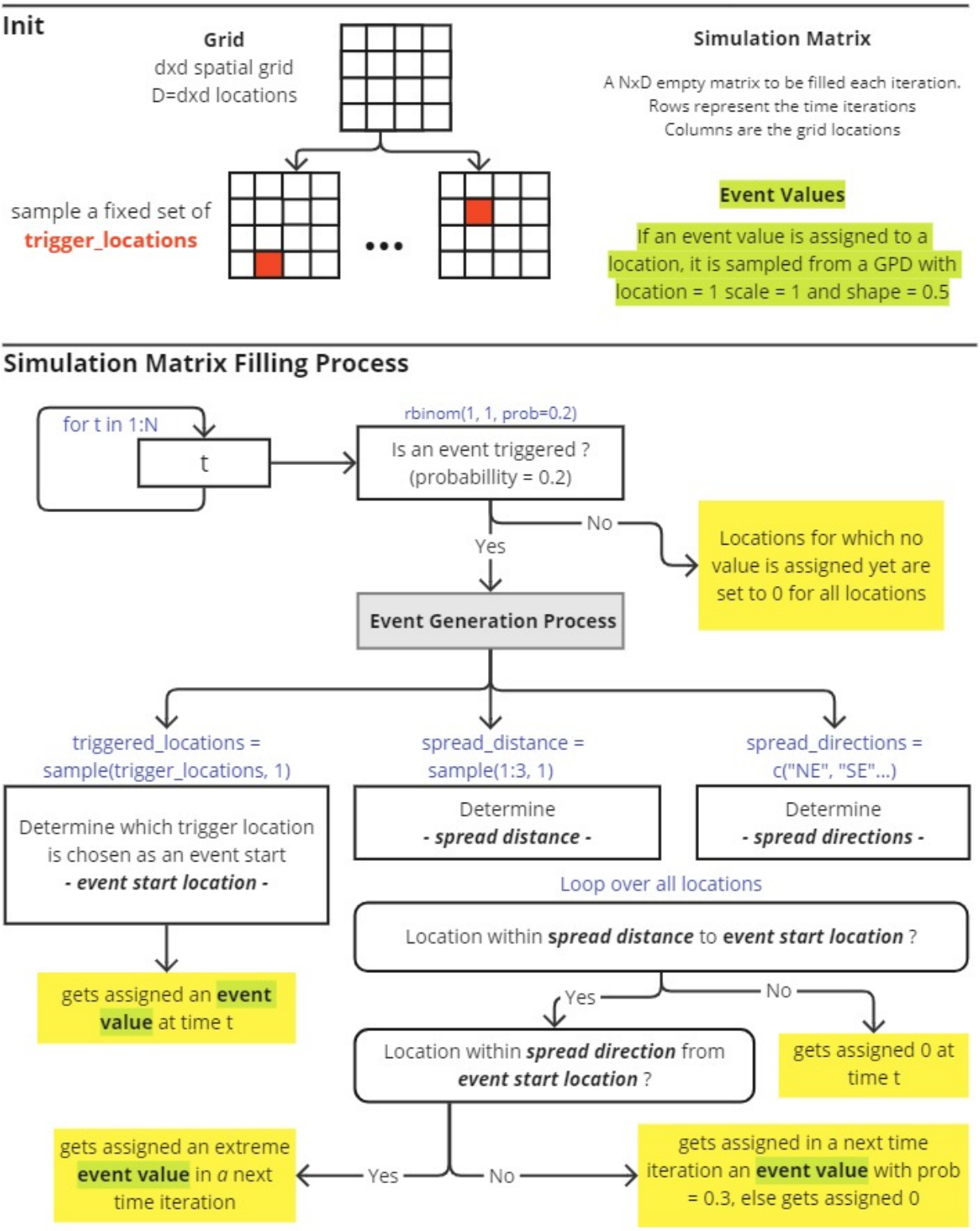
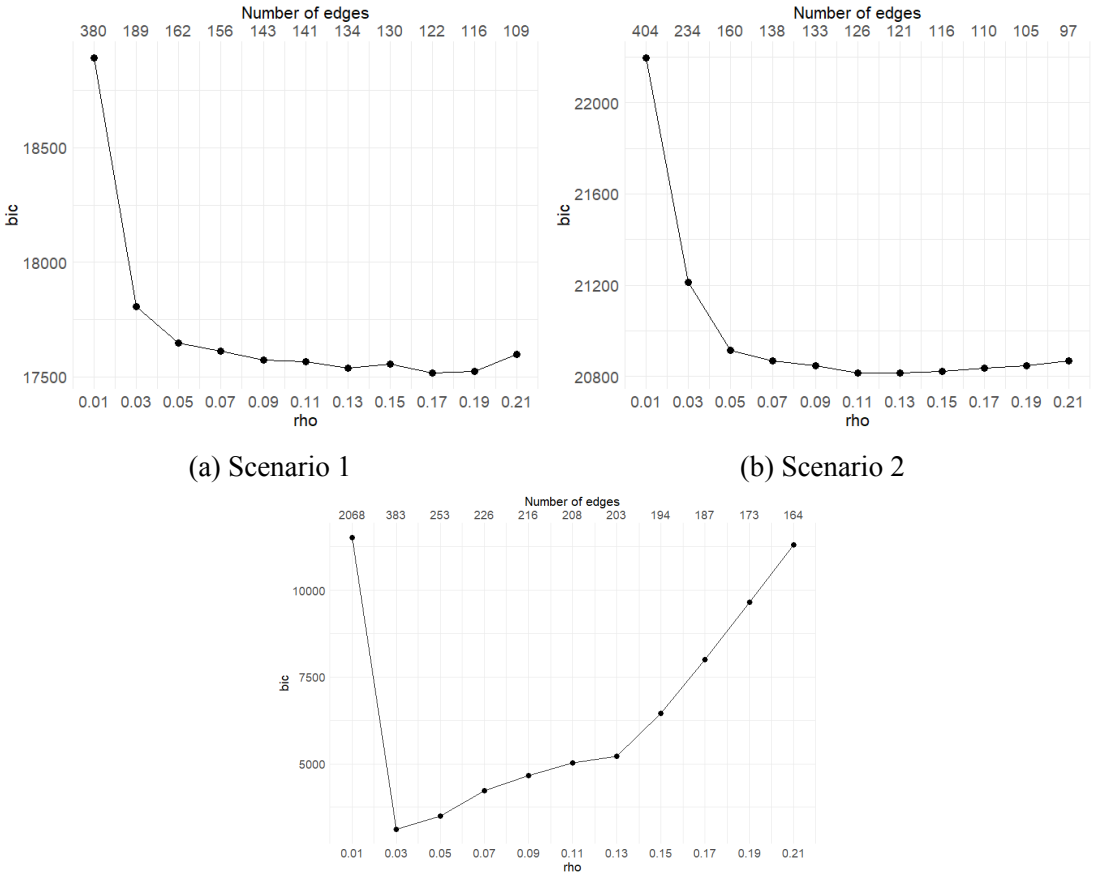


Figure A.1.: Overview of the simulation set-up

B. Additional Simulation Results

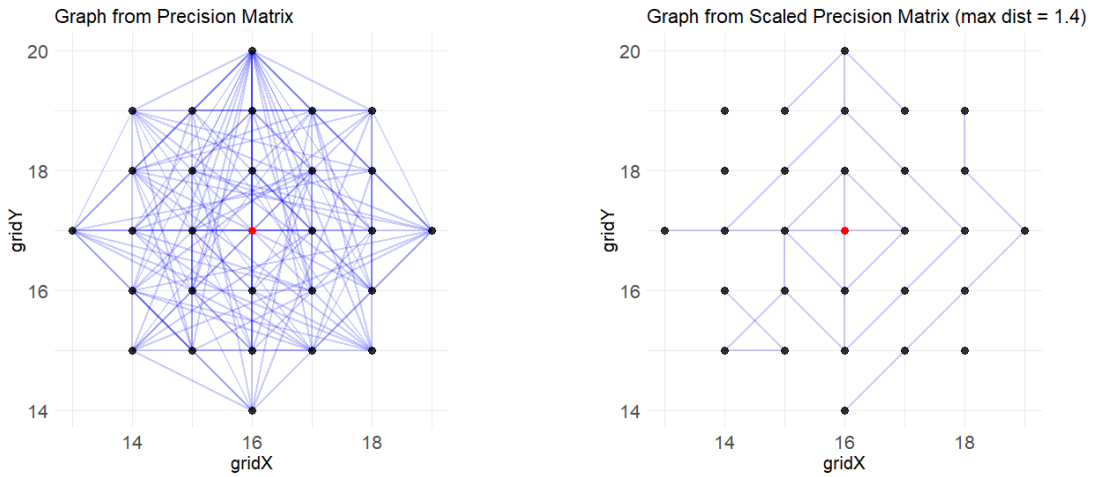


(a) Scenario 1

(b) Scenario 2

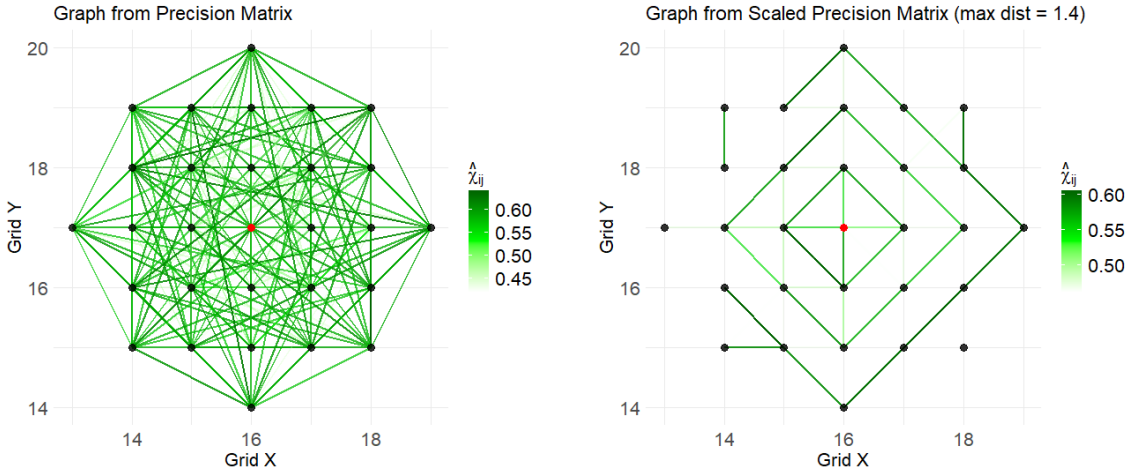
(c) Scenario 3

Figure B.1.: Penalization parameters against BIC of the corresponding estimated graphs - Used are, a 90% quantile threshold and the NS based algorithm of `eglearn` (Engelke, Lalancette, and Volgushev 2021). Package used: `graphicalExtremes` (Engelke, Hitz, et al. 2024)



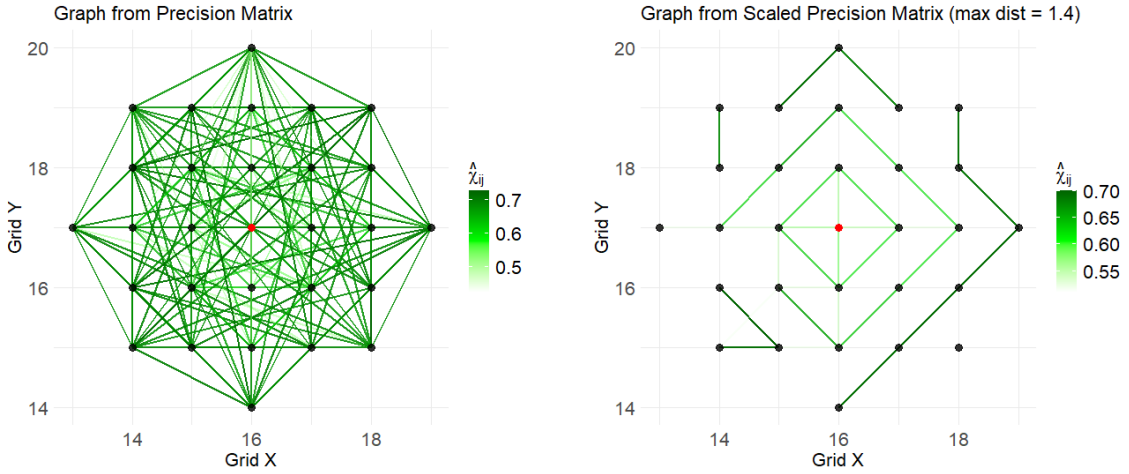
(a) (b)

Figure B.2.: Scenario 1 - Estimated HR graphical model using `eglearn` algorithm in (a) (Engelke, Lalancette, and Volgushev 2021) and a scaled version in (b) using the maximum distance of the estimates obtained via `Kest()` function (Baddeley and Turner 2005). The graph in (a) minimizes BIC, using a 90% quantile threshold and the neighbor selection based algorithm.



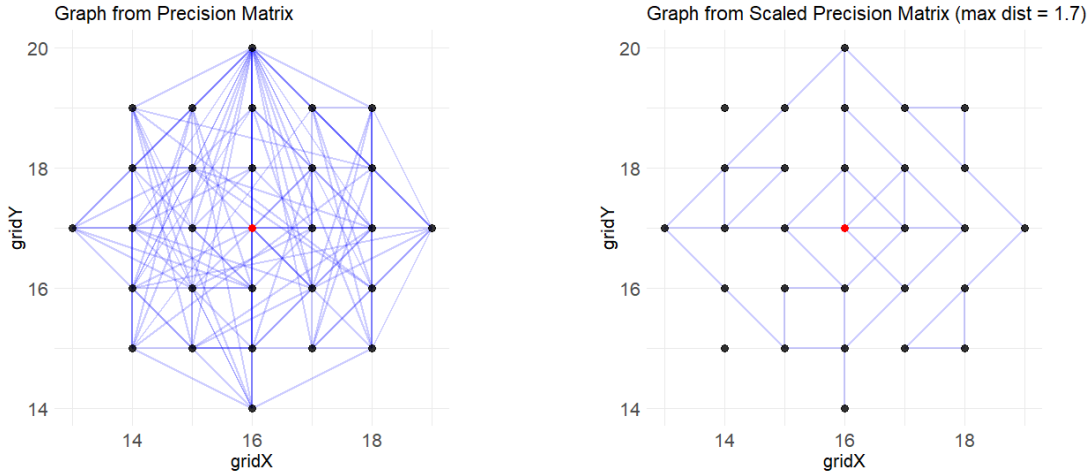
(a) (b)

Figure B.3.: Scenario 1 - Estimated HR graphical model using `eglearn` algorithm in (a) (Engelke, Lalancette, and Volgushev 2021) and a scaled version in (b). The graph in (a) minimizes BIC, using a 90% quantile threshold and the glasso based algorithm.



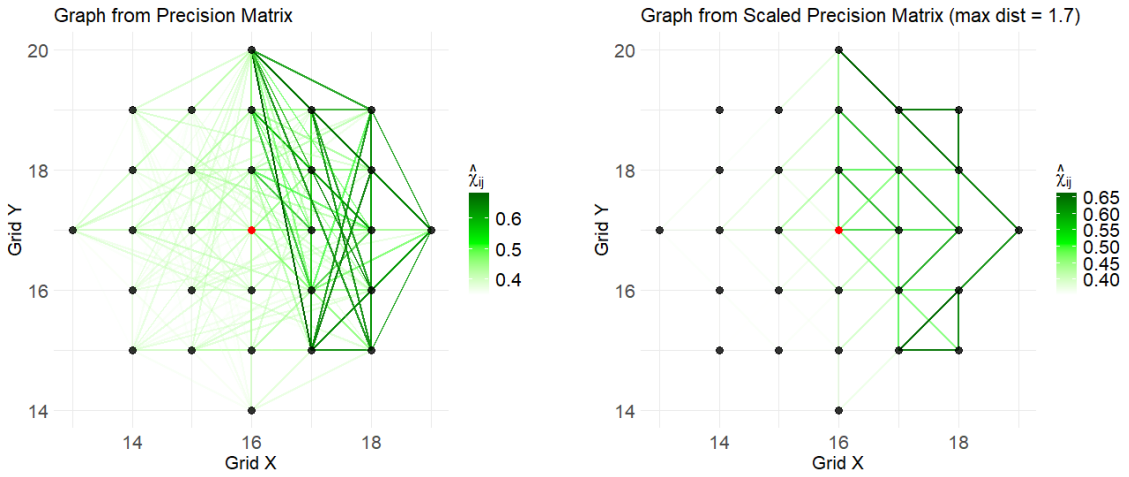
(a) (b)

Figure B.4.: Scenario 1 - Estimated HR graphical model using `eglearn` algorithm in (a) (Engelke, Lalancette, and Volgushev 2021) and a scaled version in (b). The graph in (a) minimizes BIC, using a 70% quantile threshold and the NS based algorithm.



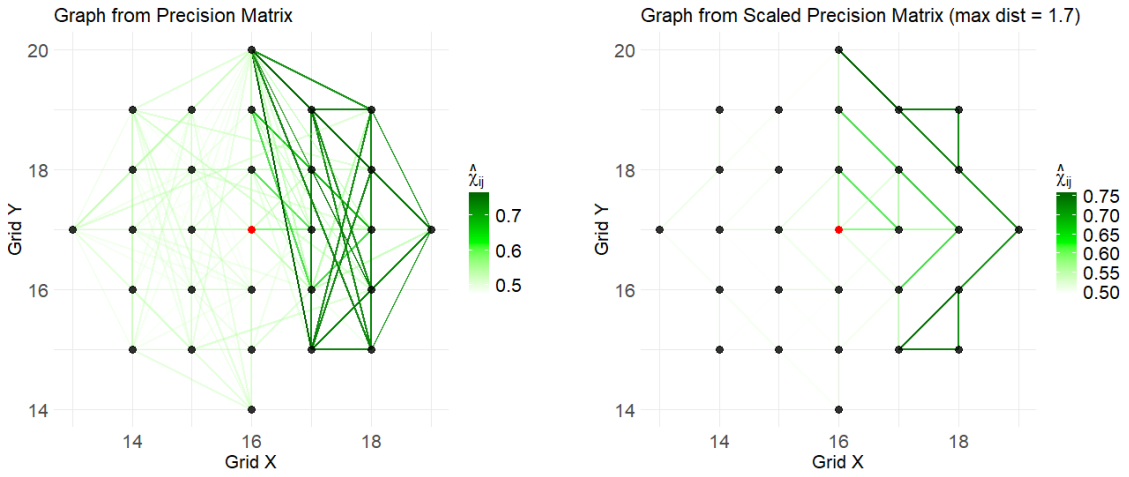
(a) (b)

Figure B.5.: Scenario 2 - Estimated HR graphical model (a) and its scaled version (b) as described in Figure B.2.



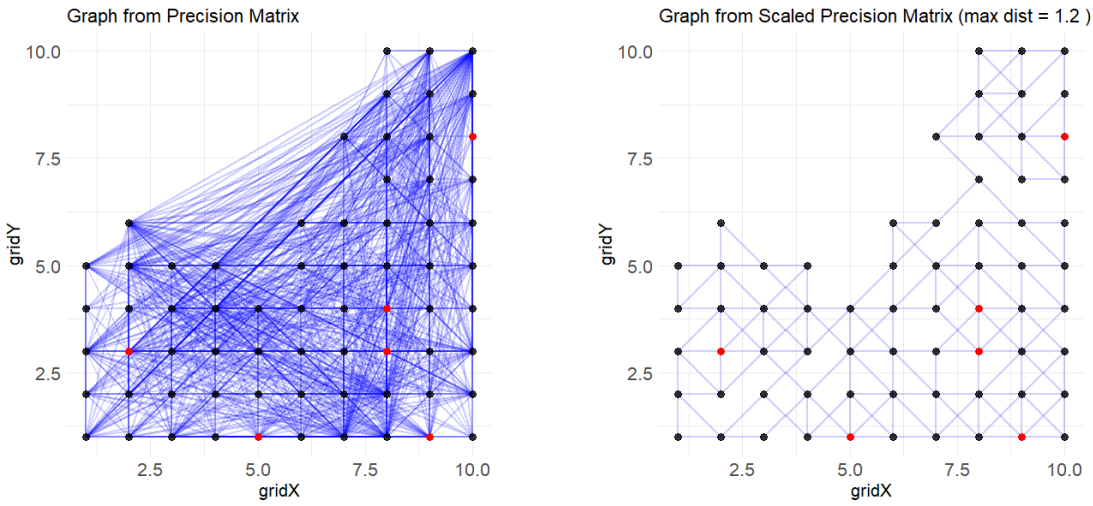
(a) (b)

Figure B.6.: Scenario 2 - Estimated HR graphical model using `eglearn` algorithm in (a) (Engelke, Lalancette, and Volgushev 2021) and a scaled version in (b). The graph in (a) minimizes BIC, using a 90% quantile threshold and the glasso based algorithm.



(a) (b)

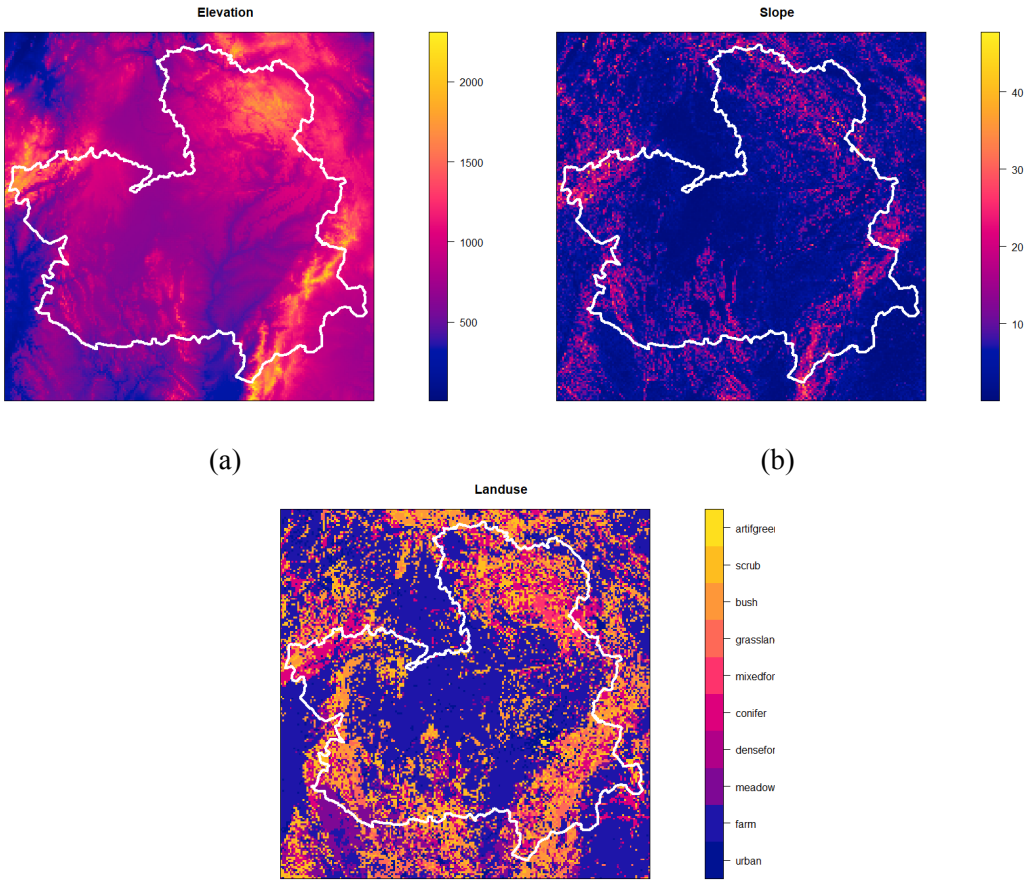
Figure B.7.: Scenario 2 - Estimated HR graphical model using `eglearn` algorithm in (a) (Engelke, Lalancette, and Volgushev 2021) and a scaled version in (b). The graph in (a) minimizes BIC, using a 70% quantile threshold and the NS based algorithm.



(a) (b)

Figure B.8.: Scenario 3 - Estimated HR graphical model (a) and its scaled version (b) as described in Figure B.2 with trigger locations marked in red.

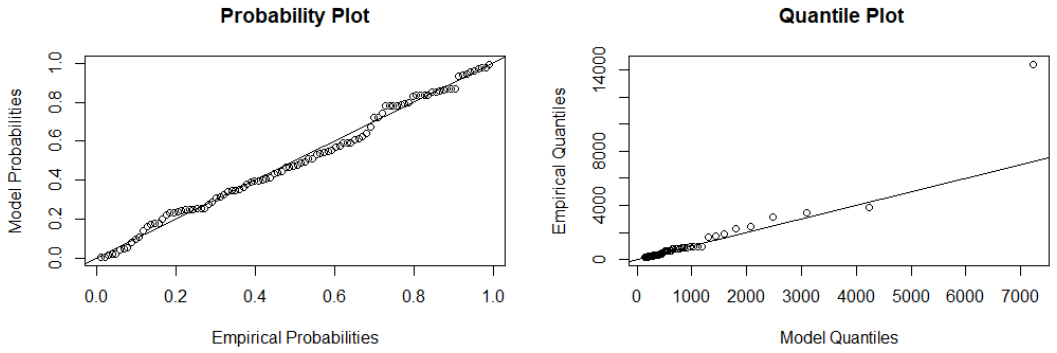
C. Additional Results Fire Data



(a) (b)

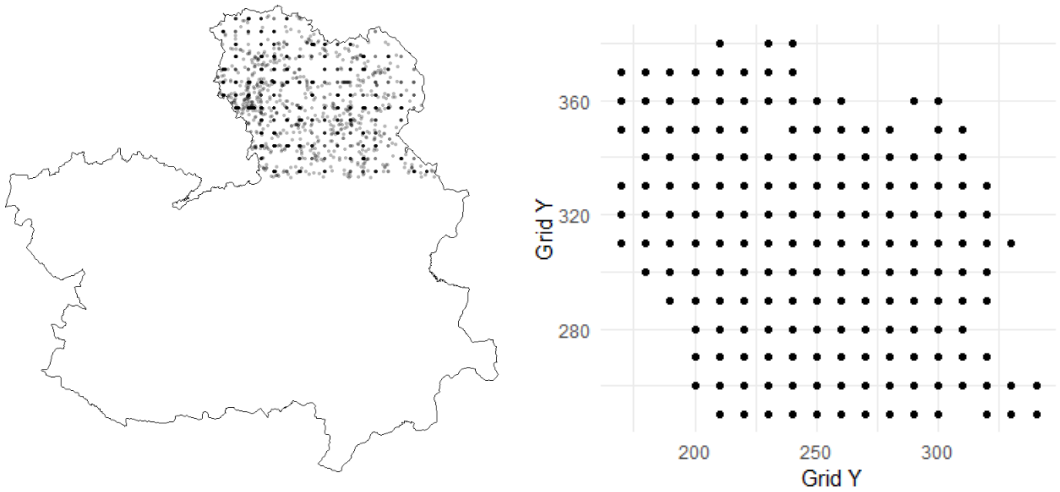
(c)

Figure C.1.: 200 pixel images of (a) elevation, (b) slope and (c) landuse of the administrative region of Castilla-La Mancha (white outline). Images available in R Package *spatstat* under `clmfires.extra`.



(a) Probability Plot (b) Quantile Plot

Figure C.2.: Diagnostic plots for the GPD fit on daily aggregated burned land in the region Castilla La Mancha, using the 95% quantile value (138.6 ha) as a threshold.



(a) Original spatial point process (b) spatially aggregated points

Figure C.3.: Subset of points in the northern region of Castilla-La Mancha.

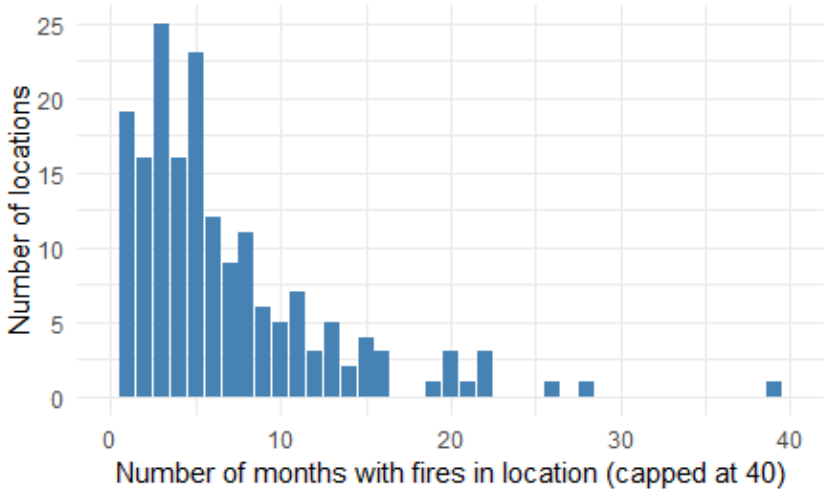
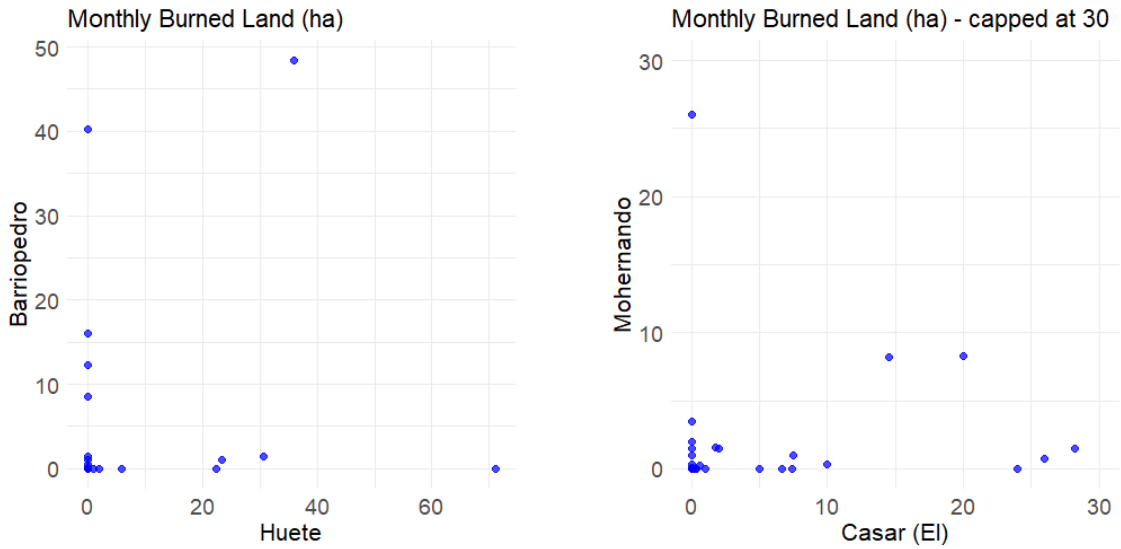


Figure C.4.: Number of months with fires in the spatially aggregated locations in the northern region.



(a) (b)

Figure C.5.: Scatterplot of monthly area burned between two connected locations.

D. List of R Packages used

This appendix lists all the R (R Core Team 2022) packages used in this thesis and serves as acknowledgment for the work of their developers. These packages provide functions and features that were essential for data analysis and visualization. To ensure transparency and reproducibility, the complete code is provided in the supplementary material. The packages listed here can thus be traced back to specific parts of the thesis.

- Most of the plots where created using `ggplot2` (Wickham 2016)
- Data cleaning and manipulation was done using both `tidyr` (Wickham, Vaughan, and Girlich 2024) and `dplyr` (Wickham, François, et al. 2023).
- For handling longitude/latitude values and converting them to utm coodinates, the package `sf` (Pebesma 2018) was used.
- For the graphical modeling aspects, the general package `igraph` (Csárdi et al. 2025) and the more specific package `graphicalExtremes` (Engelke, Hitz, et al. 2024) were used.
- Other extreme value analysis was conducted using the packages `evd` (A. G. Stephenson 2002) and `extRemes` (Gilleland and Katz 2016).
- Spatial point pattern analysis was performed using `spatstat` (Baddeley and Turner 2005).
- Finally, `lubridate` (Grolemund and Wickham 2011) was used to handle dates and `readxl` (Wickham and Bryan 2023) to load excel files.

Bibliography

Amatulli, G., Camia, A., and San-Miguel-Ayanz, J. (2013): "Estimating future burned areas under changing climate in the EU-Mediterranean countries." In: *Science of the total environment* 450, pp. 209–222.

Baddeley, A., Bárány, I., and Schneider, R. (2007): "Spatial point processes and their applications." In: *Stochastic Geometry: Lectures Given at the CIME Summer School Held in Martina Franca, Italy, September 13–18, 2004*, pp. 1–75.

Baddeley, A. and Turner, R. (2005): "Spatstat: an R package for analyzing spatial point patterns." In: *Journal of statistical software* 12, pp. 1–42.

Beniston, M., Stephenson, D. B., et al. (2007): "Future extreme events in European climate: an exploration of regional climate model projections." In: *Climatic change* 81, pp. 71–95.

Bivand, R., Gómez-Rubio, V., and Rue, H. (2015): "Spatial data analysis with R-INLA with some extensions." In: American Statistical Association.

Cisneros, D., Hazra, A., and Huser, R. (2024): "Spatial Wildfire Risk Modeling Using a Tree-Based Multivariate Generalized Pareto Mixture Model." In: *Journal of Agricultural, Biological and Environmental Statistics* 29.2, pp. 320–345.

Cisneros, D., Richards, J., et al. (2024): "Deep graphical regression for jointly moderate and extreme Australian wildfires." In: *Spatial Statistics*, p. 100811.

Coles, S., Bawa, J., et al. (2001): *An introduction to statistical modeling of extreme values*. Vol. 208. Springer.

Cressie, N. (2015): *Statistics for spatial data*. John Wiley & Sons.

Csárdi, G., Nepusz, T., et al. (2025): *igraph: Network Analysis and Visualization in R*. R package version 2.0.3. DOI: `10.5281/zenodo.7682609`. URL: `https://CRAN.R-project.org/package=igraph`.

Davison, A. C. and Huser, R. (2015): "Statistics of extremes." In: *Annual Review of Statistics and its Application* 2.1, pp. 203–235.

Dowdy, A. J., Mills, G. A., et al. (2009): "Australian fire weather as represented by the McArthur forest fire danger index and the Canadian forest fire weather index." In.

Dupuy, J.-l., Fargeon, H., et al. (2020): "Climate change impact on future wildfire danger and activity in southern Europe: a review." In: *Annals of Forest Science* 77, pp. 1–24.

Engelke, S. and Hitz, A. S. (2020): "Graphical models for extremes." In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82.4, pp. 871–932.

Engelke, S., Hitz, A. S., et al. (2024): "Package 'graphicalExtremes': Statistical Methodology for Graphical Extreme Value Models." In: `https://github.com/sebastian-`

`engelke/graphicalExtremes`. URL: `https://CRAN.R-project.org/package=graphicalExtremes`.

Engelke, S., Lalancette, M., and Volgushev, S. (2021): "Learning extremal graphical structures in high dimensions." In: *arXiv preprint arXiv:2111.00840*. `https://doi.org/10.48550/arXiv.2111.00840`.

Engelke, S. and Volgushev, S. (2022): "Structure learning for extremal tree models." In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84.5, pp. 2055–2087.

Friedman, J., Hastie, T., and Tibshirani, R. (2008): "Sparse inverse covariance estimation with the graphical lasso." In: *Biostatistics* 9.3, pp. 432–441.

Genton, M. G., Butry, D. T., et al. (2006): "Spatio-temporal analysis of wildfire ignitions in the St Johns River water management district, Florida." In: *International Journal of Wildland Fire* 15.1, pp. 87–97.

Gilleland, E. and Katz, R. W. (2016): "extRemes 2.0: an extreme value analysis package in R." In: *Journal of Statistical Software* 72, pp. 1–39.

Glymour, C., Zhang, K., and Spirtes, P. (2019): "Review of causal discovery methods based on graphical models." In: *Frontiers in genetics* 10, p. 524.

Grolemund, G. and Wickham, H. (2011): "Dates and Times Made Easy with lubridate." In: *Journal of Statistical Software* 40.3, pp. 1–25. URL: `https://www.jstatsoft.org/v40/i03/`.

Haan, L. and Ferreira, A. (2006): *Extreme value theory: an introduction*. Vol. 3. Springer.

Hentschel, M., Engelke, S., and Segers, J. (2024): "Statistical inference for Hüsler–Reiss graphical models through matrix completions." In: *Journal of the American Statistical Association*, pp. 1–13.

Hering, A. S., Bell, C. L., and Genton, M. G. (2009): "Modeling spatio-temporal wildfire ignition point patterns." In: *Environmental and Ecological Statistics* 16, pp. 225–250.

Huser, R., Opitz, T., and Wadsworth, J. (2024): "Modeling of spatial extremes in environmental data science: Time to move away from max-stable processes." In: *arXiv preprint arXiv:2401.17430*.

Huser, R. and Wadsworth, J. L. (2019): "Modeling spatial processes with unknown extremal dependence class." In: *Journal of the American statistical association* 114.525, pp. 434–444.

– (2022): "Advances in statistical modeling of spatial extremes." In: *Wiley Interdisciplinary Reviews: Computational Statistics* 14.1, e1537.

Hüsler, J. and Reiss, R.-D. (1989): "Maxima of normal random vectors: between independence and complete dependence." In: *Statistics & Probability Letters* 7.4, pp. 283–286.

Koh, J., Pimont, F., et al. (2023): "Spatiotemporal wildfire modeling through point processes with moderate and extreme marks." In: *The annals of applied statistics* 17.1, pp. 560–582.

Lauritzen, S. (1996): *Graphical models*. Clarendon Press.

Meinshausen, N. and Bühlmann, P. (2006): "High-dimensional graphs and variable selection with the lasso." In: *The Annals of Statistics* 34.3. DOI: 10.1214/009053606000000281, pp. 1436–1462.

Pebesma, E. (2018): "Simple Features for R: Standardized Support for Spatial Vector Data." In: *The R Journal* 10.1, pp. 439–446. DOI: 10.32614/RJ-2018-009. URL: https://doi.org/10.32614/RJ-2018-009.

R Core Team (2022): *R: A Language and Environment for Statistical Computing*. R version 4.2.1. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Ripley, B. D. (1976): "The second-order analysis of stationary point processes." In: *Journal of applied probability* 13.2, pp. 255–266.

Serra, L., Saez, M., et al. (2012): "Spatio-temporal modelling of wildfires in Catalonia, Spain, 1994–2008, through log-Gaussian Cox processes." In: *WIT Transactions on Ecology and The Environment* 158. doi:10.2495/FIVA120041.

Stephenson, A. G. (June 2002): "evd: Extreme Value Distributions." In: *R News* 2.2, pp. 31–32. URL: https://CRAN.R-project.org/doc/Rnews/.

Turco, M., Llasat, M.-C., et al. (2014): "Climate change impacts on wildfires in a Mediterranean environment." In: *Climatic Change* 125.3, pp. 369–380.

Urquizu, P. (2010): *Longitud y latitud de los municipios de España*. URL: https://www.businessintelligence.info/varios/longitud-latitud-pueblos-espana.html.

Wagner, C. V. (1987): *Development and structure of the Canadian forest fire weather index system.* 35. https://www.cabidigitallibrary.org/doi/full/10.5555/19910646918.

Wickham, H. (2016): *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN: 978-3-319-24277-4. URL: https://ggplot2.tidyverse.org.

Wickham, H. and Bryan, J. (2023): *readxl: Read Excel Files*. R package version 1.4.2. URL: https://CRAN.R-project.org/package=readxl.

Wickham, H., François, R., et al. (2023): *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4. URL: https://CRAN.R-project.org/package=dplyr.

Wickham, H., Vaughan, D., and Girlich, M. (2024): *tidyr: Tidy Messy Data*. R package version 1.3.1. URL: https://CRAN.R-project.org/package=tidyr.

Yadav, R., Huser, R., et al. (2023): "Joint modelling of landslide counts and sizes using spatial marked point processes with sub-asymptotic mark distributions." In: *Journal of the Royal Statistical Society Series C: Applied Statistics* 72.5, pp. 1139–1161.

Declaration of Academic Honesty

I, Marie Isabelle Bernière, hereby declare that I have not previously submitted the present work for other examinations. I wrote this work independently. All sources, including sources from the Internet, that I have reproduced in either an unaltered or modified form (particularly sources for texts, graphs, tables and images), have been acknowledged by me as such. I understand that violations of these principles will result in proceedings regarding deception or attempted deception.

Berlin, 27.01.2025