Analysis of Exceedance Patterns in Distributional Tails based on Extreme Value Theory

Lixuan An

Supervisors:
Prof. dr. Bernard De Baets
Prof. dr. Stijn Luca

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor (Ph.D.) of Bioscience Engineering: Mathematical Modelling

Department of Data Analysis and Mathematical Modelling
Faculty of Bioscience Engineering, Ghent University
Academic year 2025-2026

Supervisors:

Prof. dr. Bernard De Baets
Department of Data Analysis and Mathematical Modelling,
Ghent University, Belgium

Prof. dr. Stijn Luca
Department of Data Analysis and Mathematical Modelling,
Ghent University, Belgium

Examination committee:

Prof. dr. Peter Goethals
Department of Animal Sciences and Aquatic Ecology,
Ghent University, Belgium

Prof. dr. Niko Verhoest
Department of Environment,
Ghent University, Belgium

Prof. dr. Nico Van de Weghe
Department of Geography,
Ghent University, Belgium

Prof. dr. Mathias Verbeke
Department of Computer Science,
KU Leuven

Prof. dr. Olivier Thas
Data Science Institute,
Hasselt University

Dean:

Prof. dr. Els Van Damme

Rector:

Prof. dr. Petra De Sutter

Dutch translation of the title:

Analyse van overschrijdingspatronen in verdelingsstaarten gebaseerd op extreme-waardentheorie

Please refer to this work as follows:

Lixuan An (2025). Analysis of Exceedance Patterns in Distributional Tails based on Extreme Value Theory, PhD Thesis, Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent, Belgium

To my dear parents, with love and gratitude.

Summary

Extreme events have historically played a critical role in shaping the development of human societies. The statistical study of extreme events took form in the early twentieth century. Over successive decades, classical extreme value theory has evolved into an elegant theoretical framework for modeling the asymptotic behaviour of distributional tails, providing quantitative guidance for risk management of extreme events. Despite the rapid advancement of modern machine learning, where models often excel at anomaly detection based on abundant "normal" data, extreme value theory retains an irreplaceable role in robust extrapolation and estimation into regions far beyond the observed data. Nevertheless, classical extreme value theory focuses on individual extremes in univariate settings and is therefore insufficient to analyse the increasingly prevalent form of extreme events—clustered extremes.

In this thesis, clustered extremes are studied from a probabilistic perspective. Building on classical extreme value theory, clustered extremes are represented as a pattern of exceedances, that is, a finite set of observations that lie in the tail of the underlying distribution. Basically, three settings of such patterns of exceedances are considered. First, for independent and identically distributed (i.i.d.) observations, exceedance patterns in the tail are investigated. Second, for observations that are non-i.i.d., such as temperature data with long-term trends and serial dependence, consecutive exceedance patterns are modeled. Third, for high-dimensional data, exceedances no longer correspond to tail events of a univariate variable but rather to points located in low-density regions; patterns

of such low-density points are studied. Taken together, these three settings extend classical extreme value analysis from isolated univariate extremes to structured exceedance patterns within a unified probabilistic framework.

To start with, we outline the framework of extreme value theory, which provides the theoretical foundation for the developments in this thesis. The focus is on the two standard approaches for modeling extremes, namely the block-maxima model and the peaks-over-threshold model. These two approaches define extremes from different perspectives and lead to asymptotic statistical results for block maxima and exceedances, respectively. Furthermore, building on these classical results, a Poisson point process limit for exceedances in the tail is then introduced. The Poisson point process limit integrates the block-maxima and peaks-over-threshold approaches into a unified framework, and it shows explicitly how the distribution of block maxima and the distribution of exceedances arise as two representations of the same limiting tail process. Through this point-process view, the theoretical foundation for analysing more structured forms of extremes is established.

We investigate the first setting of clustered extremes under the standard conditions of classical extreme value theory, that is, for univariate observations that are approximately independent and identically distributed. In this setting, clustered extremes are characterized as a pattern of exceedances in the tail. By exploiting the Poisson point process limit for tail exceedances, we construct a Janossy representation for the pattern of exceedances, which is referred to as the likelihood. This likelihood specifies the probability density for realizations of a pattern of exceedances, rather than only the marginal distribution of individual extremes. Based on it, we further derive an asymptotic cumulative distribution function of the likelihood. This analytical result makes it possible to quantify how unusual a given realization of a pattern of exceedances is within the whole configuration space.

When the standard assumptions are violated, as is often the case for hydrological and climate variables with serial dependence and long-term trends, exceedances no longer occur in isolation. In such non-i.i.d. settings, what is frequently observed and of practical interest is the pattern of consecutive exceedances, for instance a sequence of days with temperature above a high threshold. It becomes essential to provide a probabilistic description of consecutive exceedance patterns for statistical extrapolation. In this case, the presence of dependence and non-identical behaviour impacts the parameters of the likelihood constructed in the i.i.d. case, but it does not change its structural form. This motivates an adaptation of the likelihood components to account for dependence and non-stationarity, while preserving the structure based on point processes. On this basis,

a likelihood for the pattern of consecutive exceedances is derived together with its cumulative distribution function, which allows clustered extremes from non-i.i.d. observations to be evaluated within the same extreme value framework.

The third setting concerns high-dimensional data, where exceedances are points located in low-density regions. In order to leverage classical extreme value theory as the theoretical support, an intermediate measure is required. This intermediate measure needs to fulfill two key conditions: first, it transforms the high-dimensional data into a univariate quantity suitable for tail analysis, and second, it retains the crucial extreme information about the low-density regions. To this end, the estimated density serves as an intermediate, preserving the extreme information from low-density regions to the lower tail of the density variable. For a broad class of distributions, the left tail of this density is known to have a Weibull limit. By applying a negative transformation to the density, attention can then be shifted to the upper tail, where a pattern of low-density region points is characterized. Within this transformed setting, a Janossy representation of the likelihood for the pattern of low-density points is formulated together with its cumulative distribution function, thereby maintaining a unified framework across all three settings.

In addition to develop theoretical work, the three settings described above are further applied to distinct real-world problems. The first setting addresses demand modeling in bike-sharing systems, where the proposed framework captures the unusual behaviour of candidate areas in a comprehensive manner. The second setting focuses on modeling heat waves, defined as consecutive days with high temperature, providing a joint probabilistic description of heat wave duration and intensity, which in turn allows us to propose a novel algorithm for constructing intensity-duration-frequency curves for heat waves. The third setting offers a comprehensive quantification of group anomalies arising in low-density regions. Within the broader context of anomalous group behaviour, where deviations may occur either in low-density regions or in high-density regions with unusual structure, this setting focuses on the former and evaluates their anomalousness within a probabilistic framework. These studies demonstrate the flexibility of the proposed framework in describing diverse extreme patterns across domains. I believe this thesis will provide readers with new insights into understanding exceedance patterns and how they extend the classical view of extremes.

Samenvatting

Extreme gebeurtenissen hebben historisch een belangrijke rol gespeeld in de ontwikkeling van menselijke samenlevingen. Het statistische onderzoek naar extreme gebeurtenissen kwam tot stand in het begin van de twintigste eeuw. In de daaropvolgende decennia heeft de klassieke extreme-waardentheorie zich ontwikkeld tot een elegant theoretisch kader voor het modelleren van het asymptotische gedrag van verdelingsstaarten, waarmee kwantitatieve richtlijnen voor risicobeheer van extreme gebeurtenissen kunnen worden geboden. Ondanks de snelle vooruitgang in de moderne predictieve modellering, waar modellen vaak uitblinken in anomaliedetectie op basis van een overvloed aan normale gegevens, blijft de extreme-waardentheorie onmisbaar voor robuuste extrapolatie en schatting in gebieden die ver buiten de geobserveerde data liggen. Tegelijkertijd richt de klassieke extreme-waardentheorie zich op afzonderlijke extremen in eendimensionale situaties en is zij daardoor niet toereikend voor de analyse van de steeds vaker voorkomende geclusterde extremen.

In dit proefschrift worden geclusterde extremen vanuit een probabilistisch perspectief bestudeerd. Voortbouwend op de klassieke extreme-waardentheorie worden geclusterde extremen voorgesteld als een overschrijdingspatroon, namelijk een eindige verzameling waarnemingen die zich in de staart van de onderliggende verdeling bevinden. Er worden in essentie drie scenario's van het overschrijdingspatroon beschouwd. Ten eerste worden voor onafhankelijke en identiek verdeelde (i.i.d.) waarnemingen overschrijdingspatronen in de staart onderzocht. Ten tweede worden voor niet onafhankelijke en iden-

tiek verdeelde waarnemingen, zoals temperatuurreeksen met langetermijntrends en seriële afhankelijkheid, patronen van opeenvolgende overschrijdingen gemodelleerd. Ten derde wordt voor hoogdimensionale data gekeken naar overschrijdingen die niet langer overeenkomen met staartgebeurtenissen van een eendimensionale variabele, maar met punten die in een gebied van lage densiteit liggen. De patronen van dergelijke lagedichtheidspunten worden dan bestudeerd. Gezamenlijk breiden deze drie scenario's de klassieke extreme-waardentheorie uit van geïsoleerde univariate extremen naar gestructureerde overschrijdingspatronen binnen één verenigd probabilistisch kader.

Allereerst wordt het raamwerk van de extreme-waardentheorie uiteengezet, wat de theoretische basis vormt voor de verdere ontwikkelingen in dit proefschrift. De nadruk ligt op de twee standaardbenaderingen voor het modelleren van extremen, namelijk het blokmaximamodel en het piek-boven-drempelmodel. Deze twee benaderingen definiëren extremen vanuit verschillende invalshoeken en leiden tot asymptotische statistische resultaten voor respectievelijk blokmaxima en overschrijdingen. Voortbouwend op deze klassieke resultaten wordt vervolgens een limiet in termen van een Poissonpuntproces voor overschrijdingen in de staart van een distributie geïntroduceerd. Deze limiet in termen van een puntproces verenigt de benaderingen met blokmaxima en met piek-boven-drempel in één kader en laat expliciet zien hoe de verdeling van blokmaxima en de verdeling van overschrijdingen kunnen worden opgevat als twee representaties van hetzelfde limietproces in de staart. Via deze puntprocesbenadering wordt de theoretische basis gelegd om meer gestructureerde vormen van extremen te analyseren.

Vervolgens wordt het eerste scenario van geclusterde extremen onderzocht onder de standaardvoorwaarden van de klassieke extreme-waardentheorie, dus voor univariate waarnemingen die benaderend onafhankelijk en identiek verdeeld zijn. In dit scenario worden geclusterde extremen beschreven als het overschrijdingspatroon in de staart. Door gebruik te maken van de limiet in termen van een Poissonpuntproces voor staartoverschrijdingen wordt een Janossy-representatie voor het overschrijdingspatroon geconstrueerd, die hier wordt aangeduid als de aannemelijkheid. Deze aannemelijkheid specificeert de kansdichtheid voor realisaties van het overschrijdingspatroon en niet enkel de marginale verdeling van afzonderlijke extremen. Op basis hiervan wordt een asymptotische cumulatieve verdelingsfunctie van de aannemelijkheid afgeleid. Dit analytische resultaat maakt het mogelijk om te kwantificeren hoe uitzonderlijk een gegeven realisatie van het overschrijdingspatroon is binnen de volledige configuratieruimte.

Wanneer de standaardassumpties niet vervuld zijn, wat vaak het geval is bij hydrologische of klimatologische metingen met seriële afhankelijkheid en langetermijntrends, treden

overschrijdingen niet langer geïsoleerd op. In dergelijke niet i.i.d. situaties is juist het patroon van opeenvolgende overschrijdingen relevant en in de praktijk van belang, bijvoorbeeld een reeks opeenvolgende dagen met een temperatuur boven een hoge drempel. Het wordt dan essentieel om een probabilistische beschrijving van dergelijke patronen van opeenvolgende overschrijdingen te geven, zodat statistische extrapolatie mogelijk blijft. In dit geval beïnvloeden de aanwezigheid van afhankelijkheid en niet-stationair gedrag de parameters van de aannemelijkheid die in het i.i.d.-geval werd geconstrueerd, maar zij veranderen de structurele vorm ervan niet. Dit vormt de aanleiding om de componenten van de likelihood aan te passen zodat afhankelijkheid en niet-stationariteit worden meegenomen, terwijl de op puntprocessen gebaseerde structuur behouden blijft. Op deze basis wordt een aannemelijkheid voor het patroon van opeenvolgende overschrijdingen afgeleid, samen met de bijbehorende cumulatieve verdelingsfunctie, zodat geclusterde extremen uit niet-i.i.d. waarnemingen binnen hetzelfde extreme-waarden raamwerk kunnen worden geëvalueerd.

Het derde scenario heeft betrekking op hoogdimensionale data, waarin overschrijdingen worden opgevat als punten die zich in lage-dichtheidsgebied bevinden. Om de klassieke extreme-waardentheorie toch als theoretische onderbouw te kunnen gebruiken, is een intermediaire maat nodig. Deze intermediaire maat moet aan twee voorwaarden voldoen. Ten eerste moet zij de hoogdimensionale data transformeren naar een eendimensionale grootheid die geschikt is voor staartanalyse. Ten tweede moet zij de essentiële informatie over extremen in lage-dichtheidsgebieden behouden. Daartoe fungeert de geschatte dichtheid als intermediair, zodat de extreme informatie uit de lage-dichtheidsregio wordt overgedragen naar de linkerstaart van de dichtheidsvariabele. Voor een brede klasse van verdelingen is bekend dat deze linkerstaart een Weibull-limiet heeft. Door een negatieve transformatie op de dichtheid toe te passen kan de aandacht vervolgens worden verlegd naar de rechterstaart, waar het patroon van punten in lage-dichtheidsgebied wordt gekarakteriseerd. Binnen deze getransformeerde setting wordt een Janossy-representatie van de aannemelijkheid voor het patroon van lage-dichtheidspunten opgesteld, samen met de bijbehorende cumulatieve verdelingsfunctie. Op deze manier blijft één verenigd kader behouden voor alle drie de scenario's.

Naast de theoretische uitwerking worden de drie hierboven beschreven scenario's toegepast op uiteenlopende praktijkproblemen. Het eerste scenario wordt gebruikt voor de vraagmodellering in deelfietssystemen, waar het voorgestelde kader het uitzonderlijke gedrag van kandidaatgebieden op een volledige wijze vastlegt. Het tweede scenario richt zich op het modelleren van hittegolven, gedefinieerd als opeenvolgende dagen met hoge temper-

atuur. Dit scenario biedt een gezamenlijke probabilistische beschrijving van de duur en intensiteit van hittegolven en maakt het mogelijk om een nieuw algoritme voor de constructie van intensiteit-duur-frequentiecurves voor hittegolven voor te stellen. Het derde scenario biedt een volledige kwantificering van groepsafwijkingen die ontstaan in lage-dichtheidsgebieden. In de bredere context van afwijkend groepsgedrag, waar afwijkingen zowel in lage-dichtheidsgebieden als in hoogdichtheidsgebieden met een ongebruikelijke structuur kunnen voorkomen, richt dit scenario zich op het eerste geval en wordt de afwijkendheid binnen een probabilistisch kader geëvalueerd. Deze toepassingen laten zien dat het voorgestelde kader voldoende flexibel is om uiteenlopende extreme patronen in verschillende toepassingsgebieden te beschrijven. Ik ben ervan overtuigt dat dit proefschrift de lezer een vernieuwend inzicht biedt in het begrijpen van overschrijdingspatronen en in de manier waarop deze het klassieke zicht op extremen kunnen uitbreiden.

Contents

List of Abbreviations

AUC	area under the curve
BM	block-maxima
BSSs	bike-sharing systems
Cali-OCSMM	calibrated one-class support measure machine
CDF	cumulative distribution function
CLT	central limit theorem
EVSMM	extreme value support measure machine
EVT	extreme value theory
FPR	false positive rate
GEV	generalized extreme value
GMM	Gaussian mixture model
GPD	generalized Pareto distribution
IDF	intensity–duration–frequency
KDE	kernel density estimation
LDA	latent Dirichlet allocation
MDA	maximum domain of attraction

MGMM	mixture of Gaussian mixture models
MLE	maximum likelihood estimation
MRL	mean residual life
OCSMM	one-class support measure machine
OCSVM	one-class support vector machine
PCA	principal component analysis
PDF	probability density function
POT	peaks-over-threshold
PPM	point process model
PPP	Poisson point process
P–P plot	probability-probability plot
Q-Q plot	quantile-quantile plot
ROC	receiver operating characteristic
SDSS	sloan digital sky survey
SMDD	support measure data description
TPR	true positive rate

List of Symbols

X	univariate ramdom variable
x	realization or sample of the univariate random variable X
$\boldsymbol{X}$	multivariate ramdom variable
$\boldsymbol{x}$	realization or sample of the univariate random variable $\boldsymbol{X}$
$\tilde{\boldsymbol{X}}$	point process or group
$\tilde{\boldsymbol{x}}$	realization of the point process or group $\tilde{\boldsymbol{X}}$
M_n	block maxima
x^{exc}	exceedances
x_F	right-end point of the distribution of X
c_n	location parameter of the GEV distribution
d_n	scale parameter of the GEV distribution

ξ	shape parameter of the block-maxima and peaks-over-threshold models
τ	threshold
σ_τ	threshold-dependent scale parameter
u_n	block-dependent threshold
σ_{u_n}	block-dependent scale parameter
u	normalized threshold
Λ	intensity measure
N	total number of samples in a dataset
$G(\cdot)$	generalized extreme value cumulative distribution function
$F_X(\cdot)$	cumulative distribution function of X
$f_X(\cdot)$	probability density function of X
$\overline{F}_X(\cdot)$	survival function of X
$F_X^{-1}(\cdot)$	left-continuous quantile function of F_X
$\kappa(\cdot,\cdot)$	kernel function
$K(\cdot,\cdot)$	kernel function on probability distributions

Part I

Introduction and Preliminaries

CHAPTER 1

General introduction

1.1 Extreme value analysis

Extreme events can be broadly understood as those unusual observations that have a relatively small likelihood of occurrence. In real life, extreme events often result in severe losses and influence design choices. Extreme high sea levels, which are becoming a growing concern under climate change, are hazards that occur rarely but could have severe consequences. They may flood low-lying areas, erode shorelines, and contaminate freshwater resources with salt. To mitigate these risks, the design of a sea wall becomes a critical defensive measure, one that should withstand the highest expected sea levels throughout its projected lifespan (e.g., 50 or 100 years). The design is generally based on modeling the historical records of extreme sea levels, as it is constrained by a fundamental trade-off: if the wall is too low, overtopping becomes frequent, leading to flooding and faster deterioration; while if the wall is too high, construction and maintenance costs rise.

Another representative example is the heat case. Extreme daily temperatures not only threaten the health of vulnerable populations (the elderly, children, and outdoor workers) but also impose substantial strains on public infrastructure. They largely challenge the stability of electricity grids and disrupt transportation systems. To ensure resilient operation, engineers must design these systems (e.g., power substations and railway lines) to withstand the severe predicted temperatures over their service life. What people do is model historical temperature extremes to estimate the return level—the temperature value expected to occur, on average, once within the system's design period. Based on these return levels, systems are properly designed to be optimized against thermal overloads.

Up to now, the extreme events described above are primarily observations or measurements of a univariate variable, where we focus on the magnitude of the variable regardless of its duration. However, in reality, extreme events can also appear as observations of a point process, indicating we record not just 'how large' an unusual value is but also 'how often' it occurs. In other words, for an extreme event, both the magnitude and the frequency may change. For example, previously, extreme daily temperatures were analysed, which are hot single days. In addition to this, another key focus of study in the field is extreme heat waves. A heat wave is a sequence of consecutive hot days whose overall length can vary. Due to anthropogenic climate change, the frequency and intensity of extreme heat waves have increased globally. This presents escalating damage to human health and infrastructure, attracting the attention of authorities worldwide. In the UK, a heat wave is declared when a location records at least three consecutive days with daily

temperatures above a threshold (Met Office, 2025; UK House of Commons Library, 2025). International guidance likewise describes heat waves as sequences of unusually hot days and nights, with thresholds adapted to local climate (World Meteorological Organization, 2023; United Nations Office for Disaster Risk Reduction, 2025).

The need to model extremes as a point process is also vital in high-stakes environments such as patient monitoring (Luca et al., 2014). In the hospital, acceleration data are recorded from the movements of epilepsy patients by four accelerometer sensors attached to the limbs during the night. In this context, treating the extreme events as univariate observations—single isolated unusual values—is ineffective. This is because normal movements (such as turning over or repositioning a limb) can also generate high acceleration values exceeding an alert threshold. If the system were to trigger an alarm for every such single point value, the resulting high false alarm rate would waste clinical resources and undermine the entire monitoring system's reliability. Indeed, the extreme events must be defined as a temporal process rather than a single point. Effective intervention relies on identifying a pattern-based extreme—a number of exceedances that occur within a short monitoring window. This definition of extreme events for triggering an alarm is able to filter out background noise and maintain a low false alarm rate for clinical intervention.

From the practical examples discussed above, we have a profile of extreme events regarding their importance, scarcity characteristics, and manifestation. In practice, to mitigate their severe consequences, providing robust risk estimates, such as the return level, is critical. This requires us to look far beyond what we have observed; in statistics, this is referred to as extrapolation. Returning to the coastal example, local data on extreme sea levels might be available, say, 10 years. The challenge is that we need the design of a sea wall that is safe for the next 100 years, with only 10 years of local sea-level data in hand. Luckily, extreme value theory (EVT) paves the way to make such extrapolations. EVT is a statistical discipline for modeling the stochastic behaviour of extreme events, providing a principled way for extrapolation beyond observed records together with uncertainty estimation. The development of EVT has always been driven by practical needs. In the late 1920s, Fisher and Tippett showed that sample maxima, after suitable scaling, can converge in distribution to one of only three possible limit forms. This fundamental insight became the theoretical foundation for the block-maxima (BM) approach (Fisher and Tippett, 1928). Gnedenko gave the first rigorous theorem in 1943, establishing the general conditions for those limits (Gnedenko, 1943). In the 1970s, a second breakthrough showed that exceedances above a high threshold have a simple limiting form, which is the generalized Pareto distribution, allowing analysts to use more of the tail

data. This was established by Balkema and de Haan (Balkema and de Haan, 1974) and by Pickands (Pickands, 1975). Pickands also introduced inference tools that helped make the approach practical. In the 2000s, authoritative monographs such as Coles (2001), De Haan & Ferreira (2006), and Embrechts (2013) consolidated these ideas, facilitating their widespread application across engineering, climate, and finance applications (Coles, 2001; de Haan and Ferreira, 2006; Embrechts et al., 1997).

To establish an intuitive understanding of EVT, its classical foundation—the block-maxima model—is often introduced through a useful analogy with the central limit theorem (CLT). This comparison centers on the fact that both theorems describe the asymptotic behaviour of a statistic derived from independent and identically distributed (i.i.d.[1]) data. Just as the CLT describes the behaviour of sample averages, the BM model demonstrates the behaviour of sample maxima. To clarify this essential analogy, the key properties of these two block-based limit theorems are summarized in Table 1.1.

Table 1.1: Two block-based limit theorems.

	Central Limit Theorem	**Block-Maxima Model**
Data assumption	i.i.d.	i.i.d.
Statistic	*Average* of the block	*Maximum* of the block
Normalization	Convergence in distribution after centering and scaling (by $\sqrt{n}$)	Convergence in distribution after linear rescaling
Limit family	Gaussian distribution	Generalized Extreme Value (GEV) family: Gumbel (Type I), Fréchet (Type II), Weibull (Type III)
Interpretation	Averages stabilize to a Gaussian law	Maxima stabilize to a GEV law

Imagine we draw a large number of i.i.d. samples from an underlying distribution. Then these samples are organized by dividing them into blocks. For the CLT, we average the observations within each block to obtain a sequence of block averages. After centering and scaling, the distribution of these averages is asymptotically a Gaussian distribution. Now switch the lens to extremes. In the block-maxima model, we take the maximum of every block, forming a sequence of block maxima. After linear rescaling, block maxima converge in distribution to a generalized extreme value (GEV) family. In summary, under the same

[1]The i.i.d. shorthand is used for simplicity. In many applications, weak dependence or stationarity with standard regularity conditions also suffices.

block structure, the CLT summarizes the central part (averages) of a distribution, whereas the BM model summarizes the tail (extremes) of a distribution. From Table 1.1, we can see two shared features. First, both models work with blocks of approximately i.i.d. observations. Second, both require a normalization step. For a more detailed analysis of the model structure and the statistical inference, we refer the reader to Chapter 2. In practice, the BM model is highly applicable as engineering design and risk assessment are fundamentally governed by the largest value within a given period (e.g., the highest sea level in a year, the peak river flow in a season). Fitting the GEV distribution to the block maxima enables the robust estimation of various extreme risk metrics, guiding design choices for infrastructural protection.

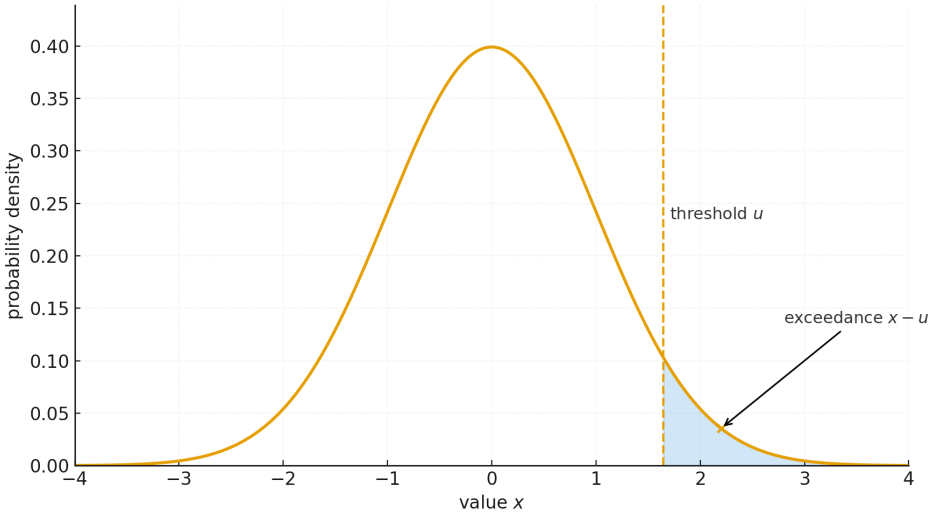


Figure 1.1: Upper-tail exceedances above a high threshold u.

While the block maxima model provides the theoretical foundation for EVT, its approach of keeping only one maximum per block actually wastes extreme information in the tail, especially when blocks are long and extremes are scarce. Consequently, the peaks-over-threshold (POT) model emerged in the 1970s as a crucial alternative. Instead of keeping only the single extreme value in each block, the POT model regards extremes as those observations that exceed a high threshold (referred to as exceedances), thereby making more efficient use of the tail than the block-maxima model. In Figure 1.1, the dashed line marks a high threshold. The light-blue region to the right is the upper tail, where any observation falling in it is an exceedance. The core result of the POT model is that, under standard regularity conditions, the distribution of exceedances is well approximated by a generalized Pareto distribution (GPD). Further details are provided in Chapter 2.

So far, our discussion of the BM and POT models has focused on the upper tail. The lower

tail is handled by duality: for the BM model, use block minima instead of maxima; for the POT model, choose a low threshold and treat observations below it as exceedances. After decades of development, EVT has earned its central place in extreme events analysis. Even in an era rich with machine learning tools, EVT remains indispensable for tail inference. Modern machine learning methods are effective at detecting unusual observations—they learn decision boundaries based on the range covered by training data to label observations as normal or anomalous. Decisions about extreme events in reality, however, require far more than detection: they demand tail extrapolation beyond the observed range together with quantified uncertainty (return levels/periods with confidence intervals). Those tasks are not what machine learning models are designed to deliver, but they are exactly what EVT models are built for through an elegant theoretical framework.

While EVT has been extensively applied to model isolated extreme events, a significant gap remains in the existing study regarding the analysis of clustered extremes. This dissertation is dedicated to advancing this field by focusing on the pattern of exceedances—the structural configurations of extremes—rather than just isolated observations themselves. We develop a novel analytical EVT-based framework to model a pattern of exceedances across multiple settings. Through this work, we present a full characterization of the pattern of exceedances and derive analytical results with theoretical proofs. We will study several real-world applications in which the framework is used to deliver robust, problem-specific solutions for real-world risk management. At the end of the chapter, an overview diagram of the thesis is provided to show how the chapters are organised.

1.2 Contributions and outline

This dissertation is structured as follows. Chapter 2 provides the essential knowledge for the subsequent chapters, presenting an introduction to classical EVT and relevant mathematical methodologies. This chapter is recommended for readers who seek a concise background in EVT. We begin by reviewing the two classical models for extreme value analysis: the BM model and the POT model. Following this, the chapter introduces the crucial concept of the maximum domain of attraction. The Poisson point process formulation is then detailed, which elegantly unifies the BM and POT models and explicitly clarifies the relationship between their limit results. Finally, this chapter provides essential methodological tools—including the one-class support vector machine, the one-class support measure machine, and the uninorm—that are necessary for understanding the

work presented in later chapters.

After the preliminaries, Chapter 3 provides the core analytical result of the pattern of exceedances under conditions consistent with classical EVT, namely for approximately i.i.d. univariate observations. To this end, a likelihood-based characterization is developed to capture the spatial configuration of exceedance patterns in a comprehensive manner. This characterization is motivated by two ideas. First, according to classical EVT, exceedances over a high threshold can be approximated by a Poisson point process. Second, in point process theory, the Janossy densities provide a powerful methodology for describing the structure of a point process. They directly guide the construction of our likelihood variable. Building on this characterization, we derive the asymptotic cumulative distribution function of the likelihood under standard regularity conditions in EVT. This chapter provides an important theoretical framework for this thesis. In subsequent chapters, the proposed theorem is extended to develop further EVT-based methods for analyzing the pattern of exceedances across multiple settings. Concretely, the contributions are:

3.a Develop a likelihood-based characterization of the pattern of exceedances.

3.b Derive the asymptotic cumulative distribution function of the likelihood, and provide a rigorous proof that enables analytical modeling of exceedance patterns.

Chapter 4 presents a direct application of the theoretical framework developed in Chapter 3 to the domain of urban transportation, specifically bike-sharing systems. The primary objective is the identification and ranking of anomalous high-demand areas within a bike-sharing system to support optimal docking station deployment. To achieve this, the objective is formulated as an unsupervised group anomaly detection problem. Each studied region is first divided into subregions. A feature variable (magnitude) is extracted for each subregion, and those with magnitudes exceeding a high threshold are treated as exceedances. Collectively, these exceedances form the region's pattern of exceedances, which is then modeled using the developed likelihood-based approach. This likelihood-based approach benefits the assessment of the anomalousness of candidate subregions with respect to their demand intensity, spatial coverage, and the demand profile in surrounding subregions. We analyse the Münchner Verkehrsgesellschaft bike-sharing system, showing that the approach identifies and ranks candidate areas effectively with a low computational cost. This chapter is mainly based on the peer-reviewed journal paper "Identifying and ranking high-demand areas in a bike-sharing system based on extreme value theory". The detailed derivation and proof of the theorem have been moved to Chapter 3. Concretely, the contributions are:

4.a Develop a data-driven spatial partitioning strategy for the studied area that utilizes existing docking station locations and is designed to be hyperparameter-free.

4.b Apply the likelihood-based model from Chapter 3 to bike-sharing systems, quantifying the anomalousness of candidate subregions through the modeling of their exceedance patterns with a low computational cost.

In Chapter 5, the theoretical framework of Chapter 3 is developed to address data exhibiting non-stationarity and dependence, where this non-i.i.d. setting is often encountered in climate research. Under such non-i.i.d. conditions, the pattern of consecutive exceedances becomes a phenomenon of particular relevance, where we derive its likelihood and subsequently establish the analytical cumulative distribution function of the likelihood. The utility of this framework is demonstrated by the application to heat wave risk analysis, in which an integrated approach is proposed to assess heat waves in terms of intensity, duration, and frequency. Specifically, the likelihood describes heat waves by modeling a sequence of exceedances above a time-varying threshold and explicitly accounting for varying heat wave duration. The analytical cumulative distribution function of the likelihood serves two critical functions. First, it provides a probabilistic tool for consistent comparison of heat waves with different durations and intensity profiles. Second, this framework facilitates the coherent derivation of core risk metrics, such as return periods and duration-specific return levels, which will enable the construction of intensity–duration–frequency curves for comprehensive heat wave risk analysis. This chapter is mainly based on the peer-reviewed journal paper "An integrated approach for assessing the intensity, duration and frequency of heat waves using extreme value theory". Concretely, the contributions are:

5.a Develop a framework to model a pattern of consecutive exceedances under conditions of non-stationarity and dependence, which is broadly applicable to climate research.

5.b Provide a systematic approach for computing intensity–duration–frequency curves.

5.c Demonstrate the practical use of the framework by applying it to real-world climate data from multiple European cities for risk analysis.

In Chapter 6, the pattern of exceedances is investigated in a high-dimensional setting, with the aim of addressing the challenging task of group anomaly detection. In this high-dimensional setting, exceedances are not the points in the tails of a univariate distribution, but rather points that lie in the low-density regions of the feature space. To obtain

the likelihood for the pattern of exceedances, a preliminary step is required, where each high-dimensional point is replaced by its density. In this way, the original multivariate description is expressed on a univariate scale, and the extreme information associated with low-density regions is retained in the lower tail of the density. This provides a principled extreme value basis for modeling points in low-density regions. As the density values are non-negative with a lower endpoint of zero, the lower tail of the random variable 'density at a random draw' is known in existing literature to follow a Weibull limit under mild regularity conditions. Applying the negative logarithm transformation to the density turns its lower tail near zero into an upper tail having a Gumbel limit. We subsequently derive the likelihood for the pattern of low-density region points together with its analytical cumulative distribution function. The model developed in this chapter is designed to comprehensively conduct group anomaly detection. Group anomalous behaviour manifests in two distinct forms: anomalies with an unusual configuration in low-density regions and high-density regions. Our strategy is twofold. First, the likelihood-based model quantifies the anomalousness of a group with respect to points in low-density regions. Second, a calibrated one-class support measure machine is developed to assess the unusual point configuration in high-density regions. These two probabilistic outputs are then integrated to yield an overall measure of group anomalousness. This chapter is mainly based on the published journal paper "The extreme value support measure machine for group anomaly detection". Concretely, the contributions are:

6.a Develop a likelihood framework for high-dimensional settings to quantify the group anomalousness associated with low-density regions.

6.b Calibrate the one-class support measure machine to produce probabilistic outputs, enabling the quantification of group anomalousness with respect to high-density regions.

6.c Construct a uninorm aggregation function to integrate the two probabilistic outputs and preserves strong joint evidence better than averaging, thereby yielding an overall measure of group anomalousness.

Chapter 7 concludes the thesis by summarizing the main findings and outlining future research directions. The appendix provides supplementary material, including the L-moments estimation procedure and additional results for Chapters 5 and 6.

A.1 L-moments estimation: a detailed description of the method together with a worked example.

A.2 Supplementary material for Chapter 5: city-specific threshold values u_n and probability–probability (P–P) plots used for model validation.

A.3 Supplementary material for Chapter 6: receiver operating characteristic (ROC) curves and area under the curve (AUC) boxplots summarizing the model's performance across all experimental cases.

1

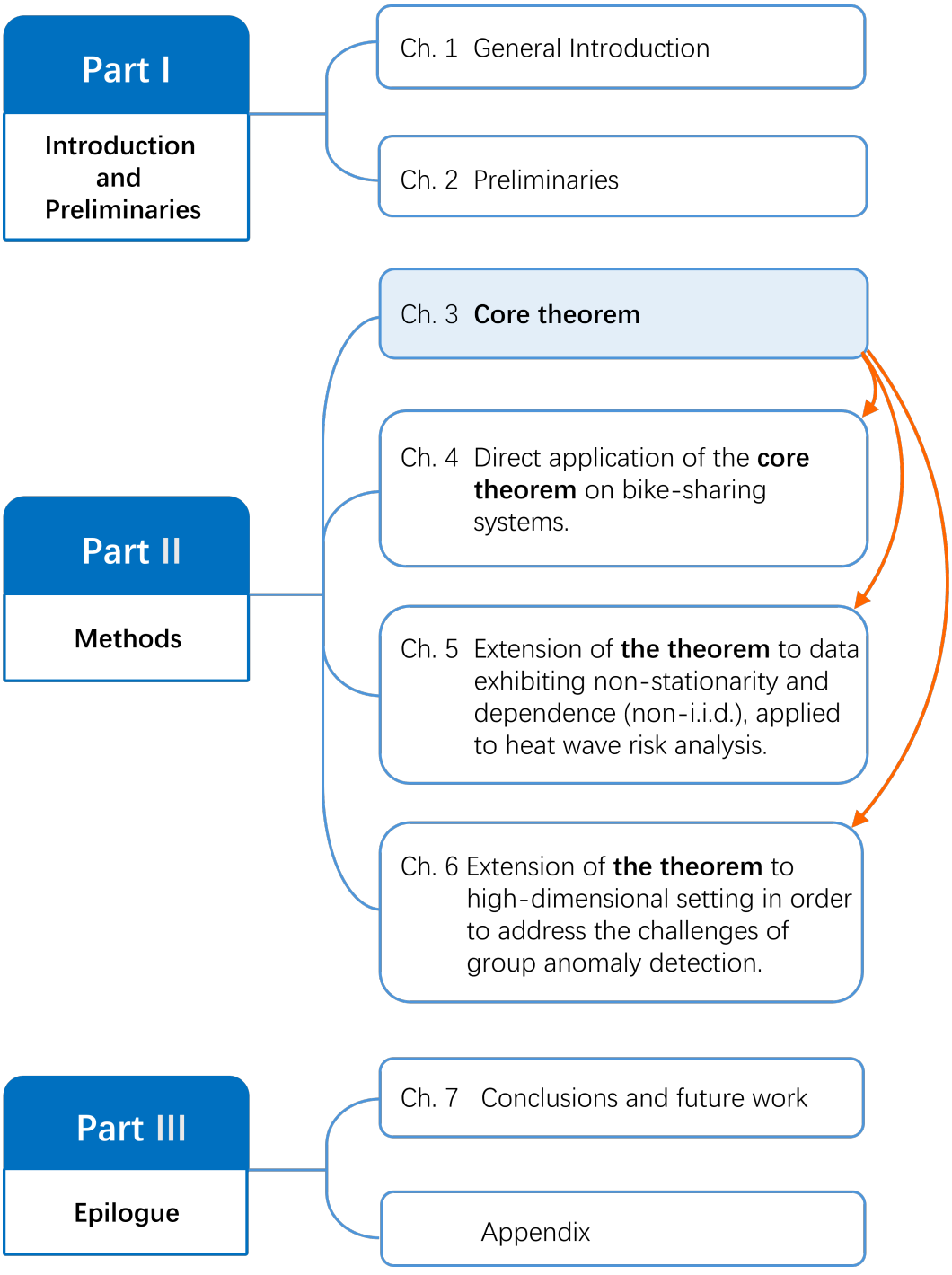
1.3 Mathematical notation

For the sake of clarity and consistency, the following notational conventions are adopted throughout this thesis.

Random variables are denoted by uppercase letters (e.g., X, Y). Observations or sample realizations of random variables are denoted by lowercase letters (e.g., x, y). We use bold letters (e.g., $\boldsymbol{X}$, $\boldsymbol{x}$) to denote variables or observations in a high-dimensional space, which are assumed to be column vectors unless explicitly stated otherwise. Non-bold letters are reserved for univariate quantities. Subscripts, typically integers or letters, are used to denote particular elements within variables or observations (e.g., X_i, x_i). Mathematical spaces are generally denoted by blackboard bold letters (e.g., $\mathbb{R}$ for the space of real numbers).

A set of observations or a general collection of points is enclosed in curly braces and represented as $\tilde{\boldsymbol{x}} = \{x_1, \ldots, x_n\}$. The entire observed dataset or sample space is denoted by calligraphic letters (e.g., $\mathcal{A}$, $\mathcal{T}$). As for probabilistic notations, the probability measure of an event is primarily denoted by $P(\cdot)$, e.g., $P(X \leq x)$. The probability density or mass function is denoted by $f(\cdot)$ or $p(\cdot)$. Finally, the expected value of a random variable is denoted by $E[\cdot]$, e.g., $E[X]$.

Throughout this paper, the term "normal" consistently signifies what is usual and regular. When discussing distributions, we employ the term "Gaussian distribution" instead of "normal distribution".


Part I
Introduction and Preliminaries
Ch. 1 General Introduction
Ch. 2 Preliminaries
Part II
Methods
Ch. 3 Core theorem
Ch. 4 Direct application of the core theorem on bike-sharing systems.
Ch. 5 Extension of the theorem to data exhibiting non-stationarity and dependence (non-i.i.d.), applied to heat wave risk analysis.
Ch. 6 Extension of the theorem to high-dimensional setting in order to address the challenges of group anomaly detection.
Part III
Epilogue
Ch. 7 Conclusions and future work
Appendix

1

CHAPTER 2

2

Preliminaries

This chapter establishes the foundational knowledge necessary for understanding the methods presented throughout this dissertation. The core content of this chapter focuses on extreme value theory. Sections 2.1 to 2.5 introduce a theoretical framework of extreme value theory, which includes the review of two classical models (the block-maxima model and the peaks-over-threshold model), the maximum domain of attraction, and the Poisson point process approach, providing a rigorous foundation for modeling extreme events. Section 2.6 collects other necessary data analysis techniques, including the one-class support vector machine, the one-class support measure machine, and the uninorm.

For all EVT-related concepts and derivations presented in this chapter, we refer to books of Coles (2001) ("An Introduction to Statistical Modeling of Extreme Values") and Embrechts et al. (1997) ("Modelling Extremal Events for Insurance and Finance").

2.1 What are extreme events in EVT?

Extreme events are rare departures from typical behaviour. Such extreme events matter because they often determine socio-economic decisions. A city plans its drainage for the heaviest storms it may face, not the average rain. A river levee is designed for unusually high flows rather than for typical discharge. Insurance pricing reflects the possibility of rare large claims. Financial risk management focuses on sudden market drops more than on day-to-day fluctuations. These extreme events occur rarely but have a large impact in practice.

The necessity of managing extreme events motivated the development of EVT. In classical EVT, an extreme event is an observation (or realization) that is unusually large or small relative to the distribution of a variable, that is, a tail event. The effective application of EVT requires the data to be tied to a clearly specified variable on a specified scale, where the variable defines the quantity of interest (e.g., temperature, strength), and the scale defines the unit of observation (e.g., daily, per position). For example, if we study potential flooding risk, the variable might be rainfall volume, and the scale can be daily. An extreme event here is an observation in the upper tail of the distribution of daily rainfall—an exceptionally heavy daily rainfall. In materials science, an engineer might be interested in the ultimate limits of a new alloy. The variable of interest is the tensile strength, measured on a scale of per specimen. We look at unusually large values in the upper tail to determine the material's maximum capabilities. In finance, when assessing potential losses, the variable is often the loss amount, and the scale might be per trading position under a market scenario. In all these cases, defining the variable and the scale explicitly is crucial because it ensures that the 'large' or 'small' in the tails is consistently interpreted and quantifiable relative to a distribution.

The extreme events lying in the tails typically manifest in two ways that guide how they are modeled. One way keeps only the single most extreme outcome within a period or batch, referred to as a 'block' in EVT, such as the heaviest daily rainfall in a year or the largest claim in a calendar quarter. The other keeps every observation that exceeds a predefined high or low alarm threshold, for example, all days with rainfall above a fixed value or all claims above a chosen retention level. These two viewpoints lead to the two standard forms of extreme value analysis.

A statistical description is now formally introduced. Let X denote a clearly specified variable observed on a fixed scale. The **block-maxima** (BM) approach studies the maximum

observations in non-overlapping blocks of size n:

$$M_n = \max\{X_1, \ldots, X_n\}\,.$$

The **peaks-over-threshold** (POT) approach studies all observations that exceed a high threshold τ, often through the excesses

$$Y = X - \tau, \quad X > \tau.$$

In both cases, the focus is on the upper tail of the distribution of X. The lower tail is handled by applying the same constructions to $-X$. Small values of X correspond to large values of $-X$. The block minima of X are given by:

$$\begin{aligned}\tilde{M}_n &= \min\{X_1, \ldots, X_n\} \\ &= -\max\{-X_1, \ldots, -X_n\}\,.\end{aligned}$$

The exceedances below a low threshold τ^* correspond to:

$$Y^* = \tau^* - X, \quad X < \tau^*.$$

The subsequent sections will present specific models in detail and explain how they describe tail behaviour to support rigorous quantification of risk. Note that the classical EVT is built for the univariate case. Besides, the choice of analyzing the upper or the lower tail is determined entirely by the practical application. For mathematical convenience, the theory is often unified by studying the upper tail of a variable. In this dissertation, we adopt this convention.

2.2 Block-maxima model

The BM model is one of the two fundamental approaches in EVT. This chapter first introduces the core result—the Fisher–Tippett–Gnedenko theorem—which sets the basis for analyzing extremes. We then proceed to discuss the parameter estimation. Finally, the use of the fitted model for risk quantification through return levels and return periods is illustrated.

2.2.1 Model

The central result of the Fisher-Tippett-Gnedenko theorem concerns the limiting distribution of block maxima

$$M_n = \max\{X_1, X_2, \ldots, X_n\}\,, \text{ as } n \to +\infty\,,$$

where $X_1, X_2, \ldots, X_n$ are i.i.d. univariate random variables with a common distribution function F_X.

Theorem 2.1. [**Fisher-Tippet-Gnedenko**] If there exist sequences of normalizing constants c_n and d_n such that the following convergence in distribution holds:

$$P\left(\frac{M_n - c_n}{d_n} \leq x\right) \xrightarrow{d} G_\xi(x)\,, \text{ as } n \to +\infty\,, \tag{2.1}$$

then the non-degenerate limiting distribution $G_\xi(x)$ is a member of the so-called generalized extreme value (GEV) family of distributions:

$$G_\xi(x) = \begin{cases} \exp\left\{-\left[1+\xi x\right]^{-\frac{1}{\xi}}\right\}\,, & \text{if } \xi \neq 0 \\ \exp\{-\exp(-x)\}\,, & \text{if } \xi = 0\,. \end{cases} \tag{2.2}$$

If $\xi \neq 0$, then the domain of the distribution is restricted to the set $\{x \mid 1+\xi x > 0\}$. The parameter ξ controls the tail behaviour of F_X: Fréchet type when $\xi > 0$, Weibull type with a finite upper end-point when $\xi < 0$, and Gumbel type when $\xi = 0$.

This theorem tells us that if block maxima can be normalized so that a non-degenerate limit exists, then the only possible limit law is a member of the GEV family. Consequently, it implies a limiting distribution for block maxima does not exist for every parent distribution F_X. The conditions on the parent distribution F_X required for convergence will be detailed in Section 2.4.

2.2.2 Parameter estimation

Within the BM model, the distribution of block maxima at a chosen block size n is described by three parameters: the location c_n, the scale d_n, and the shape ξ. The shape parameter ξ captures how the tail of the parent distribution F_X behaves and therefore

determines which limiting law within the GEV family applies. The parameters c_n and d_n play the role of centering and scaling constants that put the maxima on a comparable scale. As the block size increases, the maximum of the block tends to be larger due to more opportunities to observe a more extreme value. Therefore, c_n and d_n must depend on n to appropriately center and scale the block maxima. A sequence of $\{c_n\}$ and $\{d_n\}$ track how the distribution of block maxima moves and spreads as n increases, while ξ does not change with n.

In practice, the parameters are typically estimated using maximum likelihood estimation (MLE). The likelihood function is the model's joint probability density evaluated at the observed data, regarded as a function of the parameters. Parameter values with a larger likelihood correspond to a model that is more compatible with the observed data.

For the BM model, the likelihood function is obtained by evaluating the model's density of block maxima as a function of the parameters (c_n, d_n, ξ). Let $x_1^{(M_n)}, \ldots, x_m^{(M_n)}$ denote the observed block maxima. Differentiating the GEV distribution function in Eq. (2.2) with respect to x yields the probability density function (PDF). As a result, the log-likelihood function for the GEV parameters (c_n, d_n, ξ) when $\xi \neq 0$ is

$$\begin{aligned}\ell(c_n, d_n, \xi) = &- m \log d_n - \left(1 + \frac{1}{\xi}\right) \sum_{j=1}^{m} \log\left[1 + \xi\left(\frac{x_j^{(M_n)} - c_n}{d_n}\right)\right] \\ &- \sum_{j=1}^{m} \left[1 + \xi\left(\frac{x_j^{(M_n)} - c_n}{d_n}\right)\right]^{-\frac{1}{\xi}},\end{aligned} \tag{2.3}$$

conditional on

$$1 + \xi\left(\frac{x_j^{(M_n)} - c_n}{d_n}\right) > 0, \quad \text{for } j = 1, \ldots, m\,.$$

In the Gumbel case ($\xi = 0$), the log-likelihood reduces to

$$\ell(c_n, d_n, \xi) = -\, m \log d_n - \sum_{j=1}^{m} \left(\frac{x_j^{(M_n)} - c_n}{d_n}\right) - \sum_{j=1}^{m} \exp\left\{-\left(\frac{x_j^{(M_n)} - c_n}{d_n}\right)\right\}. \tag{2.4}$$

Maximizing the log-likelihood functions in Eqs. (2.3) and (2.4) with respect to (c_n, d_n, ξ) yields the maximum likelihood estimates for the GEV model. However, as the log-likelihood functions have no closed-form maximizer in (c_n, d_n, ξ) (i.e., no analytical solution), the GEV parameters are typically estimated by numerical optimization in MLE.

For practical use, we introduce a widely used method, the L-moments method, for parameter estimation given its robustness and simplicity. This method is based on the idea that the theoretical L-moments of the GEV distribution can be expressed as functions of the parameters (c_n, d_n, ξ). By computing corresponding sample L-moments from data and matching them to the theoretical expressions, one obtains a set of moment-matching equations whose solution yields the parameter estimates. These closed-form estimates of the GEV parameters can be used directly in applications. Besides, they can also serve as effective initial values for numerical MLE. Further details on the L-moments of the GEV distribution, together with examples, are provided in the appendix.

We provide the steps of L-moments estimation. Let $x_{1:m} \leq \cdots \leq x_{m:m}$ be the order statistics of block maxima. The unbiased probability-weighted-moment estimates are then defined as:

$$\hat{b}_0 = \frac{1}{m}\sum_{j=1}^{m} x_{j:m}\,,$$
$$\hat{b}_1 = \frac{1}{m}\sum_{j=1}^{m} \frac{j-1}{m-1}\, x_{j:m}\,,$$
$$\hat{b}_2 = \frac{1}{m}\sum_{j=1}^{m} \frac{(j-1)(j-2)}{(m-1)(m-2)}\, x_{j:m}\,.$$

The sample L-moments are

$$\hat{\ell}_1 = \hat{b}_0\,, \qquad \hat{\ell}_2 = 2\hat{b}_1 - \hat{b}_0\,, \qquad \hat{\ell}_3 = 6\hat{b}_2 - 6\hat{b}_1 + \hat{b}_0\,.$$

Following this, the sample L-skewness $\hat{t}_3 = \hat{\ell}_3/\hat{\ell}_2$ is computed. The shape parameter ξ is estimated by solving the equation:

$$\hat{t}_3 = t_3^{\mathrm{GEV}}(\xi)\,,$$

where $t_3^{\mathrm{GEV}}(\xi)$ is the theoretical L-skewness of the GEV distribution. Since this function is monotone in ξ, the equation has a unique solution. With $\hat{\xi}$ obtained, the location c_n and scale parameters d_n are estimated by equating the first two sample L-moments ($\hat{\ell}_1$ and $\hat{\ell}_2$) to the theoretical L-moments of the GEV distribution. For $\xi = 0$, the resulting closed-form expressions for $\hat{d}_n$ and $\hat{c}_n$ are:

$$\hat{d}_n = \frac{\hat{\ell}_2\hat{\xi}}{(2^{\hat{\xi}}-1)\Gamma(1-\hat{\xi})}\,, \qquad \hat{c}_n = \hat{\ell}_1 - \frac{\hat{d}_n}{\hat{\xi}}(\Gamma(1-\hat{\xi})-1).$$

Here $\Gamma(\cdot)$ is the gamma function. In the Gumbel limit ($\xi \to 0$), by using the continuous

limits we have

$$\hat{d}_n = \frac{\ell_2}{\ln 2}\,, \qquad \hat{c}_n = \ell_1 - \gamma\,\hat{d}_n\,,$$

where $\gamma \approx 0.577$ is the Euler–Mascheroni constant.

For many known parent distributions F_X, a comprehensive summary of asymptotic normalizing constants (c_n, d_n) can be found in Section 3.3 of Embrechts et al. (1997):

- **Gumbel case.** If F_X satisfies the von Mises condition, then the normalized block maxima converge in distribution to a Gumbel limit. A standard choice of normalizing constants is

$$\hat{c_n} = F_X^{-1}\!\left(1 - \frac{1}{n}\right)\,, \qquad \hat{d}_n = a(c_n)\,, \qquad a(x) = \frac{\overline{F}_X(x)}{f_X(x)}\,, \tag{2.5}$$

  where f_X is the PDF of F_X and $\overline{F}_X = 1 - F_X$. F_X^{-1} denotes the (left-continuous) quantile function of F_X, defined by $F_X^{-1}(u) = \inf\{x \in \mathbb{R} : F_X(x) \geq u\}$ for $0 < u < 1$.

- **Fréchet case.** If the block maxima from a parent distribution F_X converge in distribution to a Fréchet distribution, then the common choice of normalizing constants is

$$\hat{c_n} = 0\,, \qquad \hat{d}_n = F_X^{-1}\!\left(1 - \frac{1}{n}\right)\,. \tag{2.6}$$

- **Weibull case.** If the block maxima from a parent distribution F_X converge in distribution to a Weibull distribution, then the common choice of normalizing constants is

$$\hat{c_n} = x_F, \qquad \hat{d}_n = x_F - F_X^{-1}\!\left(1 - \frac{1}{n}\right)\,, \tag{2.7}$$

  where x_F denotes the (finite) right endpoint of F_X.

Note that the scope of this thesis does not relate to the deeper inference of the estimated parameters. Therefore, the discussion of the uncertainty associated with these estimated parameters, such as their standard errors, variances, and confidence intervals, is omitted here. Readers interested are referred to the works by Coles (2001) and Embrechts et al. (1997).

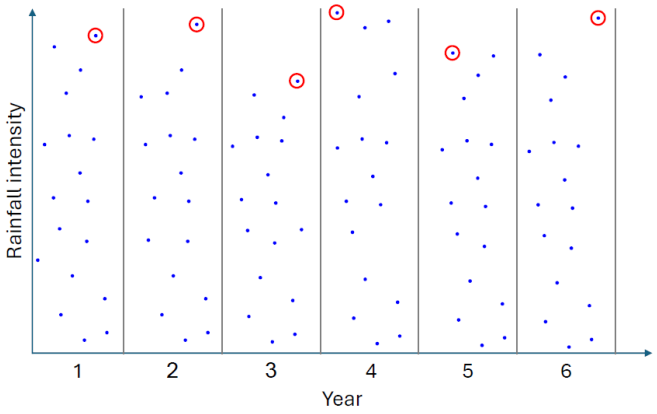


Figure 2.1: Illustration of block maxima. Daily rainfall intensity data (blue dots) partitioned into annual blocks (vertical lines) with the annual maximum (red circles) extracted from each block (Marra, 2024).

2.2.3 Return levels and return periods

Once the BM model has been fitted, it can be used for risk quantification. A key output is the return level, which translates tail probabilities into design criteria. We use an example of daily rainfall to illustrate the concepts of 'return level' and 'return period', as shown in Figure 2.1. The full data record is split into yearly blocks, and the annual maximum is extracted from each block. A BM model is then fitted to these annual maxima, where G is the resulting fitted CDF. The T-year return level x_T is the rainfall intensity value that is expected to be exceeded by the annual maximum with a probability of $\frac{1}{T}$, and the corresponding return period is T years. Equivalently, another description is that the expected waiting time until the next exceedance of x_T is T years. Formally, the return level x_T is the p-th quantile of the GEV distribution, where the non-exceedance probability p is $1 - \frac{1}{T}$:

$$G(x_T) = 1 - \frac{1}{T} \iff x_T = G^{-1}\left(1 - \frac{1}{T}\right).$$

Conversely, for a specified level x, the return period is

$$T(x) = \frac{1}{1 - G(x)}.$$

Given the fitted BM model with $\hat{c_n}$, $\hat{d_n}$ and $\hat{\xi}$, closed-form expressions for the return level x_T are:

If $\xi \neq 0$, then

$$x_T = \hat{c_n} + \frac{\hat{d_n}}{\hat{\xi}}\left\{\left[-\log\left(1 - \frac{1}{T}\right)\right]^{-\hat{\xi}} - 1\right\}.$$

If $\xi = 0$ (Gumbel), then

$$x_T = \hat{c_n} - \hat{d_n} \log\left[-\log\left(1 - \frac{1}{T}\right)\right].$$

2.3 Peaks-over-threshold model

The BM model keeps one value per block (for example, the annual maximum). It uses the GEV limit to model these maxima and performs well when abundant data are available. However, the main drawback is that this method wastes a lot of extreme information in the tails. When a large block size n is chosen, many large observations within each block are discarded, retaining only a single maximum. This process substantially reduces the sample size available for GEV fitting, often leading to unreliable estimates. In contrast, as visually illustrated in Figure 2.2, the POT model addresses this limitation by keeping all observations that exceed some high threshold, which avoids the need for blocking. This typically results in a much more efficient use of tail data from the same dataset, leading to more stable tail inference and risk quantification.

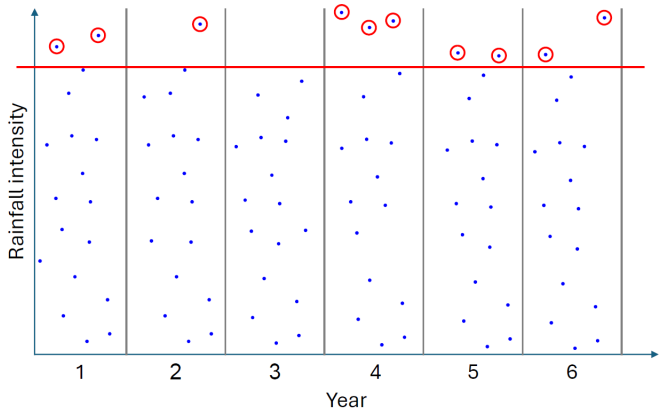


Figure 2.2: Illustration of peaks-over-threshold exceedances. Daily rainfall intensity data (blue dots) are shown with a high threshold (red horizontal line). All observations exceeding the threshold (red circles) are selected as the exceedances (Marra, 2024).

2.3.1 Model

Formally, the tail above a high threshold is characterized by the Pickands–Balkema–De Haan theorem.

Theorem 2.2. [**Pickands–Balkema–De Haan**] Let X be a random variable with distribution F_X. For a high threshold τ, define the excess $Y = X - \tau$ conditional on exceedance $X > \tau$. If Eq. (2.1) holds for some member $G_\xi(x)$ of the GEV family, then there exists a positive, threshold-dependent scale $\sigma_\tau > 0$ such that, for $y \geq 0$ with $1 + \xi \frac{y}{\sigma_\tau} > 0$:

$$P\left(\frac{Y}{\sigma_\tau} > y \mid X > \tau\right) \xrightarrow{d} H_\xi(y)\,, \qquad \text{as } \tau \uparrow x_F\,, \tag{2.8}$$

then the limiting distribution $H_\xi(y)$ is the generalized Pareto distribution (GPD)

$$H_\xi(y) = \begin{cases} 1 - (1 + \xi\, y)^{-\frac{1}{\xi}} & \text{, if } \xi \neq 0 \\ 1 - \exp(-y) & \text{, if } \xi = 0\,. \end{cases} \tag{2.9}$$

In the BM framework, the block maxima from the parent distribution F_X are assumed to converge in distribution to a GEV limit with normalizing parameters c_n and d_n. To achieve a coherent transition from the block maxima formulation to the exceedances formulation, the threshold τ is expressed in terms of normalizing constants of the BM model by setting $u_n = c_n + u\, d_n$, where u is the normalized threshold near zero. Indeed, the block size n in BM and the threshold u_n (or τ) in POT both serve to define the extreme tail region for asymptotic modeling. The Pickands–Balkema–De Haan theorem can also be expressed as follows.

Theorem 2.3. For a given $u \in \mathbb{R}$ sufficiently close to 0, the sequence of thresholds $u_n = c_n + u d_n$ is considered, where c_n and d_n denote normalizing constants of the BM model as defined in Eq. (2.1). If Eq. (2.1) holds for some member $G_\xi(x)$ of the GEV family, then the distribution of the exceedance $X - u_n$, conditional on $X > u_n$, satisfies the limiting property:

$$\lim_{n \to +\infty} P\left(\frac{X - u_n}{\sigma_{u_n}} < x \mid X > u_n\right) = H_\xi(x)\,, \tag{2.10}$$

where σ_{u_n} is the scale parameter for exceedances relative to the threshold u_n. The distribution

$$H_\xi(x) = \begin{cases} 1 - \left(1 + \frac{\xi}{1+\xi u} x\right)^{-\frac{1}{\xi}} & \text{, if } \xi \neq 0 \\ 1 - \exp\left(-x\right) & \text{, if } \xi = 0 \end{cases} \tag{2.11}$$

denotes the family of generalized Pareto distributions with $x \geq 0$ in case $\xi \geq 0$, and $0 \leq x \leq -\frac{1}{\xi}$ in case $\xi < 0$.

2.3.2 Parameter estimation

In the POT model, parameters include a high threshold τ (or u_n), a threshold-dependent scale σ_τ (or σ_{u_n}), and the tail index (shape parameter) ξ. A crucial step in the POT modeling process is the selection of the threshold τ. There is no universally analytic rule that determines an optimal threshold τ from the data. In practice, the most widely used approach is the mean residual life (MRL) plot. The MRL plot procedure involves computing the empirical mean excess for a range of candidate thresholds τ:

$$\hat{e}(\tau) = \frac{1}{N_\tau} \sum_{x_i > \tau} (x_i - \tau) \,,$$

where N_τ is the number of exceedances. The theoretical foundation is that, if the tail of the parent distribution follows the GPD, the mean excess $e(\tau)$ is a linear function of τ. Consequently, the MRL plot should display a stable segment in the region where the GPD approximation is valid. The selection of τ is a bias-variance trade-off: a threshold that is too low may violate the GPD approximation, while one that is too high leaves too few exceedances, making parameter estimations unstable. In general, the threshold is chosen as the smallest value of τ at which the final linear behaviour begins and remains steady over a reasonable range. The selected threshold τ can be further validated through a Q-Q (quantile-quantile) plot, assessing whether the empirical distribution of the threshold exceedances aligns with the fitted GPD.

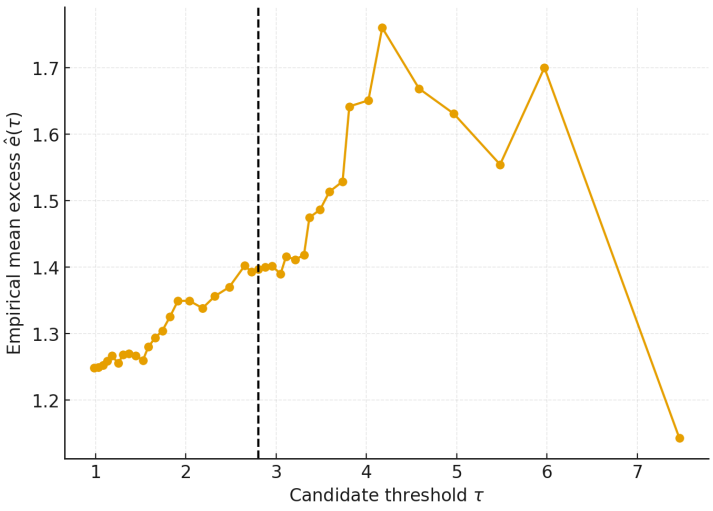


Figure 2.3: Mean residual life plot for threshold selection.

Figure 2.3 shows an example of how to choose a threshold τ based on the MLR plot. We should look for the first point where the curve starts to look approximately linear and

remains stable over a range of nearby thresholds, while τ is still reasonably large but not too large. In this example, the start of linear behaviour is near the vertical dashed line, so τ is set to 2.8.

After fixing the threshold τ, the obtained exceedances are used to estimate other parameters of the POT model. The standard method for estimating these parameters is still the MLE. However, the log-likelihood of GPD in (ξ, σ_τ) has no closed-form maximizer. An exception is the Gumbel case $\xi = 0$, where

$$\hat{\xi} = 0\,, \qquad \hat{\sigma}_\tau = \bar{y} = \frac{1}{N_\tau} \sum_{i=1}^{N_\tau} y_i\,.$$

In practice, consistent with the approach applied in the BM model, the parameters are estimated using the L-moments method.

2.3.3 Return levels and return periods

Once the POT model has been fitted, the model is then used for risk quantification. Let N be the total number of observations and N_τ the number of exceedances over τ. The empirical exceedance probability is:

$$\hat{p}_\tau = P(X > \tau) \approx \frac{N_\tau}{N}\,.$$

If observations are already on an annual scale, the T-year return level x_T satisfies $P(X > x_T) = \frac{1}{T}$. Under the POT model, for $x > \tau$ with $1 + \hat{\xi}\frac{(x-\tau)}{\hat{\sigma}_\tau} > 0\,,$

$$P(X > x) = \hat{p}_\tau \left(1 + \hat{\xi}\left(\frac{x-\tau}{\hat{\sigma}_\tau}\right)\right)^{-\frac{1}{\hat{\xi}}}.$$

Solving $P(X > x_T) = \frac{1}{T}$ for T-year return level x_T gives

$$x_T = \begin{cases} \tau + \dfrac{\hat{\sigma}_\tau}{\hat{\xi}}\left[(T\,\hat{p}_\tau)^{\hat{\xi}} - 1\right] & \text{, if } \hat{\xi} \neq 0\,, \\ \tau + \hat{\sigma}_\tau \log\big(T\,\hat{p}_\tau\big) & \text{, if } \hat{\xi} = 0\,. \end{cases}$$

If the data are more frequent than yearly, then the probability $\hat{p}_\tau$ calculated above is the exceedance probability per observation, not the annual probability. We adjust $\hat{p}_\tau$ to the

annual exceedance probability, $\hat{p}_{\tau,\mathrm{yr}}$. Assuming there are m observations per year (e.g., $m = 365$ for daily data), the annual probability of at least one exceedance above τ is calculated as:

$$\hat{p}_{\tau,\mathrm{yr}} \;=\; 1 - (1 - \hat{p}_\tau)^m \,.$$

If $\hat{p}_\tau$ is small, then the Poisson approximation to the binomial distribution simplifies this to:

$$\hat{p}_{\tau,\mathrm{yr}} \;\approx\; m\hat{p}_\tau \,.$$

Then, the T-year return level x_T is as follows:

$$x_T = \begin{cases} \tau + \dfrac{\hat{\sigma}_\tau}{\hat{\xi}}\Big[(T\,\hat{p}_{\tau,\mathrm{yr}})^{\hat{\xi}} - 1\Big] & , \text{if } \hat{\xi} \neq 0\,, \\ \tau + \hat{\sigma}_\tau \log\big(T\,\hat{p}_{\tau,\mathrm{yr}}\big) & , \text{if } \hat{\xi} = 0\,. \end{cases}$$

For any fixed return level $x > \tau$, the return period $T(x)$ is defined as the inverse of the annual probability of exceeding that level:

$$T(x) = \frac{1}{P(X > x)} = \left[\hat{p}_{\tau,\mathrm{yr}}\left(1 + \hat{\xi}\left(\frac{x-\tau}{\hat{\sigma}_\tau}\right)\right)^{-\frac{1}{\hat{\xi}}}\right]^{-1} .$$

2.4 Maximum domain of attraction

In the previous sections, we discussed two primary models for extreme value analysis: BM and POT. The Fisher–Tippett–Gnedenko theorem and Pickands–Balkema–De Haan theorem establish that the non-degenerate limit distributions (for block maxima and threshold excesses, respectively) exist if and only if there are sequences of normalizing constants (c_n and d_n, or σ_τ) such that the normalized extremes satisfy the convergence. These results imply that limiting distributions do not exist for every parent distribution F_X. Therefore, throughout the previous discussion, we used descriptive expressions such as saying "If the block maxima from a parent distribution F_X converge in distribution to a Weibull distribution" or "If the tail of F_X has a Weibull limit" to assume that this convergence condition was satisfied for the subsequent analysis and derivation. In this section, we will formally specify these necessary and sufficient conditions on F_X and introduce a concept known as the maximum domain of attraction (MDA).

We provide the formal definition of the MDA based on the Fisher–Tippett–Gnedenko

theorem. A distribution function F_X is said to belong to the MDA of a non-degenerate distribution G, denoted $F_X \in \mathrm{MDA}(G)$, if there exist sequences of constants $c_n \in \mathbb{R}$ and $d_n > 0$ such that the convergence in distribution of the normalized block maxima M_n holds:

$$\lim_{n\to+\infty} P\left(\frac{M_n - c_n}{d_n} \leq x\right) = G(x)\,.$$

By the Fisher–Tippett–Gnedenko theorem, the non-degenerate limit $G(x)$ is necessarily a member of the GEV family, $G(x) \equiv G_\xi(x)$. The concept of the MDA establishes a precise classification mechanism: the entire universe of parent distributions F_X is partitioned into three distinct domains of attraction (Fréchet, Weibull, and Gumbel), corresponding exactly to the three possible cases of the shape parameter ξ. Furthermore, the tail bahaviour of the parent distribution F_X is the necessary and sufficient condition determining which of the three domains (GEV types) it is attracted to:

1. **Fréchet Domain (G_ξ with $\xi > 0$)**: This domain is characterized by heavy-tailed parent distributions that possess an infinite upper endpoint ($x_F = \infty$). Distributions in this class are those whose heavy right tail $\overline{F}_X(x)$ (the survival function, $\overline{F}_X(x) = 1 - F_X(x)$) decays like a power function (a property known as regular variation). The necessary and sufficient condition for attraction is that there exists a $\xi > 0$ such that, for all $x > 0$, the following limit holds:

$$\lim_{t\to\infty} \frac{\overline{F}_X(tx)}{\overline{F}_X(t)} = x^{-\frac{1}{\xi}}\,.$$

   Here, $\frac{1}{\xi}$ is known as the tail index. Distributions commonly found in this class include the Pareto and Burr distributions.

2. **Weibull Domain (G_ξ with $\xi < 0$)**: This domain consists of parent distributions characterized by a finite upper endpoint ($x_F < \infty$), meaning their support is bounded from above. The convergence is determined by the rate at which the distribution approaches its endpoint. The necessary and sufficient condition for attraction is that there exists an $\alpha = -\frac{1}{\xi} > 0$ such that, for all $x > 0$, the following limit holds:

$$\lim_{t\to 0^+} \frac{\overline{F}_X(x_F - tx)}{\overline{F}_X(x_F - t)} = x^\alpha\,.$$

   The convergence limit here examines the asymptotic behaviour of the tail as we approach the finite endpoint x_F from the left. Distributions such as the uniform and beta distributions belong to this domain.

3. **Gumbel Domain (G_0 with $\xi = 0$)**: This domain covers parent distributions whose tail decays exponentially or faster, possessing either a finite or infinite upper endpoint. This is the most prevalent domain, often called the exponential domain, and includes standard statistical distributions such as the Gaussian, exponential, gamma, and logistic distributions. For continuous distributions with density f_X, attraction is often verified using the simpler von Mises sufficient condition, which requires that the auxiliary function $a(x) = \overline{F}_X(x)/f_X(x)$ satisfies:

$$\lim_{x \to x_F} \frac{d}{dx} a(x) = 0\,.$$

The MDA framework formally justifies the use of the BM model for analyzing extremes, as it guarantees that any parent distribution encountered in practice, provided its block maxima converge, will yield a limit belonging to one of the three GEV types. Tables 2.1, 2.2, and 2.3 (Embrechts et al., 1997) summarize the properties of common statistical distributions, detailing their corresponding MDA and the explicit normalizing sequences c_n and d_n.

Table 2.1: MDA of the Fréchet distribution.

Fréchet domain: $\frac{M_n - c_n}{d_n} \xrightarrow{d} \exp\{-x^{-\frac{1}{\xi}}\}\,, \quad c_n = 0\,,\, \xi > 0\,.$	
Cauchy distribution	$f(x) = (\pi(1+x^2))^{-1}, \quad x \in \mathbb{R}$ $d_n = \frac{n}{\pi}$
Pareto distribution Burr distribution stable with index $\frac{1}{\xi} < 2$	$\overline{F}(x) \sim K x^{-\frac{1}{\xi}}, \quad K > 0$ $d_n = (Kn)^{\xi}$
Loggamma distribution	$f(x) = \frac{1}{\xi^{\beta}\Gamma(\beta)} x^{-1} (\ln x)^{\beta-1} x^{-\frac{1}{\xi}}, \quad x > 1, \quad \beta > 0$ $d_n = \left((\Gamma(\beta))^{-1} (\ln n)^{\beta-1} n\right)^{\xi}$

Table 2.2: MDA of the Weibull distribution.

Weibull domain: $\frac{M_n - c_n}{d_n} \xrightarrow{d} \exp\{-(-x)^{-\frac{1}{\xi}}\}, \quad c_n = x_F\,,\ \xi < 0\,.$	
Uniform distribution	$f(x) = \frac{1}{b-a}, \quad x \in (a,b)$ (For $a = 0$, $b = 1$, $f(x) = 1$) $c_n = x_F = 1$ $d_n = n^{-1}$
Power law behaviour at x_F (x_F is finite)	$\overline{F}(x_F - x) \sim Kx^{-1/\xi}, \quad K > 0, \xi < 0$ $c_n = x_F$ $d_n = (Kn)^{\xi}$
Beta distribution	$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \quad 0 < x < 1, \quad \text{a, } b > 0$ $c_n = 1$ $d_n = \left(\frac{\Gamma(a)\Gamma(b)}{n\Gamma(a+b)}\right)^{\frac{1}{b}}$

Table 2.3: MDA of the Gumbel distribution.

Gumbel domain: $\frac{M_n - c_n}{d_n} \xrightarrow{d} \exp\{-\exp(-x)\}, \quad \xi = 0.$	
Gaussian distribution	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R}$ $c_n = \mu + \sigma\left(\sqrt{2\ln n} - \frac{\ln(4\pi)+\ln\ln n}{2\sqrt{2\ln n}}\right)$ $d_n = \frac{\sigma}{\sqrt{2\ln n}}$
Exponential distribution	$\overline{F}(x) = \exp\{-\lambda x\}, \quad x > 0, \lambda > 0$ $c_n = \frac{\ln n}{\lambda}$ $d_n = \frac{1}{\lambda}$
Gamma distribution	$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0, \quad \alpha, \beta > 0$ $c_n = \frac{1}{\beta}\left(\ln n + (\alpha - 1)\ln\ln n - \ln\Gamma(\alpha)\right)$ $d_n = \frac{1}{\beta}$
Lognormal distribution	$f(x) = \frac{1}{x\sqrt{2\pi}\sigma} \exp\left\{-\frac{(\ln x-\mu)^2}{2\sigma^2}\right\}, \quad x > 0$ $c_n = \exp\left\{\mu + \sigma\sqrt{2\ln n} - \frac{\ln(4\pi)+\ln\ln n}{2\sqrt{2\ln n}}\right\}$ $d_n = \frac{\sigma}{\sqrt{2\ln n}} c_n$
Weibull distribution	$\overline{F}(x) = \exp\left\{-\left(\frac{x}{\lambda}\right)^k\right\}, \quad x > 0, \quad \lambda, k > 0$ $c_n = \lambda(\ln n)^{\frac{1}{k}}$ $d_n = \frac{\lambda}{k}(\ln n)^{\frac{1}{k}-1}$

2.5 Poisson point process approach

The Poisson point process (PPP) approach provides an elegant and unifying theoretical framework that encompasses both the BM and POT within a single model. This framework is often regarded as a 'novel point of view of the EVT'. The PPP models the extremes as a random collection of points in a two-dimensional space (typically representing the time of occurrence and the magnitude of the event). To illustrate this framework, the subsequent discussion will detail the mathematical properties of the PPP and its dual role in deriving both the GEV and GPD limit distributions.

2.5.1 Point process model

A point process in Euclidean space, often denoted by $\tilde{\boldsymbol{X}}$, is formally defined as a random countable collection (or pattern) of points $\boldsymbol{x}_i$ within some observation space $\mathcal{A} \subset \mathbb{R}^d$ (Illian et al., 2008). A specific realization of a point pattern is given by:

$$\tilde{\boldsymbol{x}} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots\}\,, \quad \boldsymbol{x}_i \in \mathcal{A}\,.$$

A point process model inherently combines stochastic information regarding the number of points and their locations in the feature space. The fundamental tool for characterizing the stochastic properties of a point process is the counting measure $N(\cdot)$. The counting measure maps a point pattern realization $\tilde{\boldsymbol{x}}$ to the number of points that fall within a specific measurable subset A of $\mathcal{A}$ (Daley and Vere-Jones, 2003):

$$N_A(\tilde{\boldsymbol{x}}) = \sum_{i \geq 1} \chi_A(\boldsymbol{x}_i)\,, \tag{2.12}$$

where $\chi_A(\boldsymbol{x}_i)$ is the indicator function, $\chi_A(\boldsymbol{x}_i) = 1$ for $x_i \in A$, and $\chi_A(\boldsymbol{x}_i) = 0$ for $\boldsymbol{x}_i \notin A$. The random nature of a point process $\tilde{\boldsymbol{X}}$ is fully described by the distributions of its counting measures $N_{A_1}, N_{A_2}, \ldots$ across any finite number of non-overlapping subsets $A_t \subset \mathcal{A}$. An important concept associated with a counting measure is its expectation, known as the intensity measure $\Lambda(A)$, which gives the mean number of points in the set A:

$$\Lambda(A) = \mathrm{E}(N_A(\tilde{\boldsymbol{X}}))\,.$$

The intensity measure governs the distribution of point counts, which plays a vital role in characterizing a point process.

2.5.2 A Poisson point process in EVT

In the context of EVT, the focus is modeling the occurrence and magnitude of extremes derived from a sequence of observations that exceed some high threshold u_n (this threshold has been discussed before in Section 2.3, $u_n = c_n + ud_n$). Given a sequence of n approximately i.i.d. variables, $X_1, \ldots, X_n$, and suitable normalizing sequences c_n and d_n, a point process $\tilde{N}_n$ is constructed on the time-magnitude plane $\mathcal{A} = [0,1] \times \mathbb{R}$. This point process, consisting of n points, is defined by:

$$\tilde{N}_n = \left\{ \left(\frac{i}{n+1}, X_i \right) \mid i = 1, \ldots, n \right\} .$$

In this construction, the first coordinate $\frac{i}{n+1}$ represents the scaled time of observation, normalized to the unit interval $[0,1]$, and the second coordinate X_i represents the magnitude of the observation.

The core result of the PPP in EVT is as follows. We are primarily interested in the upper-tail region of the parent distribution F_X, denoted by $A =]0,1] \times]u_n, \infty[$. Under the same asymptotic conditions that lead to the GEV distribution for block maxima (Theorem 2.1) and the GPD for threshold excesses (Theorem 2.3), the point process $\tilde{N}_n$ converges to a PPP $\tilde{N}$ on the upper-tail region $A =]0,1] \times]u_n, \infty[$:

$$\tilde{N}_n \xrightarrow{d} \tilde{N}, \quad \text{as } n \to \infty .$$

The stochastic properties of the limiting Poisson process $\tilde{N}$ are described by its intensity measure Λ, where $\Lambda(A)$ provides the expected number of points falling into any measurable region A of the time-magnitude plane. Specifically, for an upper-tail region $A =]0,1] \times]u_n, \infty[$, the counting measure $N_A(\tilde{N}_n)$ is asymptotically Poisson distributed with the intensity measure

$$\Lambda(A) = \lim_{n \to \infty} \mathrm{E}[N_A(\tilde{N}_n)] = \lim_{n \to \infty} nP\left(X_i > u_n\right) . \tag{2.13}$$

Resulting from the i.i.d. assumption on the original observations, the intensity measure Λ exhibits time homogeneity across the normalized period $[0,1]$. For any subregion $A^* =]t_1, t_2] \times]u_n, \infty[$, where $]t_1, t_2] \subset]0,1]$, the intensity measure is

$$\Lambda(A^*) = \Lambda\left(]t_1, t_2] \times]u_n, \infty[\right) = (t_2 - t_1) \cdot \Lambda(A) .$$

The analytical expression of the intensity measure takes the form

$$\Lambda(A) = \begin{cases} \left[1 + \xi\left(\frac{u_n - c_n}{d_n}\right)\right]^{-\frac{1}{\xi}} = [1 + \xi u]^{-\frac{1}{\xi}} & , \text{for } \xi \neq 0\,, \\ \exp\left(-\frac{u_n - c_n}{d_n}\right) = \exp(-u) & , \text{for } \xi = 0\,. \end{cases} \tag{2.14}$$

The Poisson point process model establishes a unified framework that encompasses both the BM model and POT model for describing extremes. We formally express it as follows:

Theorem 2.4. [**Poisson Point Process Characterization**] Let X be a univariate random variable whose distribution F_X belongs to the MDA of the GEV distribution. The point process of exceedances within the block $\{X_1, \ldots, X_n\}$ that exceed the threshold u_n, *i.e.*:

$$\tilde{X}_n^{exc} = \left\{ \left(\frac{i}{n+1}, X_i \right) \mid X_i > u_n \right\},$$

converges to a Poisson point process as $n \to +\infty$. The corresponding counting measure $N_A(\tilde{X}_n^{exc})$ over region $A =]0, 1] \times]u_n, \infty[$ converges in distribution to a Poisson distribution:

$$\lim_{n \to +\infty} P(N_A(\tilde{X}_n^{exc}) = k) = \frac{\lambda^k}{k!} e^{-\lambda}\,, \tag{2.15}$$

where the intensity measure $\lambda = \Lambda(A)$ is the expectation of the counting measure as shown in Eq. (2.14). For subregion $A^* =]t_1, t_2] \times]u_n, \infty[$, the intensity measure is given by $\lambda^* = (t_2 - t_1)\Lambda(A)$.

2.5.3 Derivation of BM based on PPP

The GEV distribution for block maxima is a direct derivation of the PPP model. The derivation relies on the fundamental property of the Poisson process that the number of points in any subregion A is a Poisson-distributed random variable.

In the BM model, the GEV distribution is the asymptotic distribution of the normalized block maxima. In the PPP framework, the event $M_n \leq x$ over the entire block is identical to the event that the number of points of the process in the region $A_x =]0, 1] \times]x, \infty[$ is zero. Mathematically, this relationship between the limiting distribution of the block maxima and the probability of zero points in the upper-tail region is expressed as:

$$\lim_{n \to \infty} P\left(M_n \leq x\right) = P(N_{A_x} = 0)\,. \tag{2.16}$$

The probability of having zero points in a region A_x is given by the Poisson probability mass function at $k = 0$:

$$\begin{aligned} P(N_{A_x} = 0) &= P\left(N(]0,1]\times]x,\infty[) = 0\right) \\ &= \frac{e^{-\lambda}\lambda^0}{0\,!} \\ &= e^{-\lambda}\,, \end{aligned}$$

where $\lambda = \Lambda(A_x)$ is the intensity measure of the region A_x. According to the analytical result of the intensity measure (Eq. (2.14)), we have:

$$\lambda = \Lambda(A_x) = \begin{cases} \left[1 + \xi\left(\frac{x-c_n}{d_n}\right)\right]^{-\frac{1}{\xi}} & \text{, for } \xi \neq 0\,, \\ \exp\left(-\frac{x-c_n}{d_n}\right) & \text{, for } \xi = 0\,. \end{cases} \tag{2.17}$$

Substituting this result back into the probability equation from Eq. (2.16):

$$\begin{aligned} \lim_{n\to\infty} P\left(M_n \leq x\right) &= P(N_{A_x} = 0) \\ &= \exp(\lambda) \\ &= \begin{cases} \exp\left\{-\left[1 + \xi\left(\frac{x-c_n}{d_n}\right)\right]^{-\frac{1}{\xi}}\right\} & \text{, for } \xi \neq 0\,, \\ \exp\left\{-\exp\left(-\frac{x-c_n}{d_n}\right)\right\} & \text{, for } \xi = 0\,. \end{cases} \end{aligned} \tag{2.18}$$

This final result (Eq. (2.18)) is precisely the cumulative distribution function $G_\xi(x)$ of the GEV family (Eq. (2.2)), thus the BM model is formally derived from the asymptotic properties of the PPP.

2.5.4 Derivation of POT based on PPP

The GPD can be derived from the asymptotic properties of the PPP by determining the conditional probability of the excess magnitude, given that an exceedance occurred. This derivation links the limiting intensity measure of the PPP directly to the GPD's survival function.

The GPD models the distribution of the excess $Y = X - u_n$, conditional on the original variable X exceeding a high threshold u_n. We then derive the survival function $\overline{H}(y)$, defined as the conditional probability $P(Y > y \mid X > u_n)$ for $y > 0$:

$$\begin{aligned}\overline{H}(y) &= P(Y > y \mid X > u_n) \\ &= P(X > u_n + y \mid X > u_n) \\ &= \frac{P(X > u_n + y)}{P(X > u_n)} .\end{aligned} \tag{2.19}$$

In the PPP framework, the probability $P(X_i > x)$ for an arbitrary X_i in the point process $\tilde{N}_n$ is derived from Eq. (2.13), which we restate here for coherence of the derivation:

$$\Lambda(A_x) = \lim_{n\to\infty} \mathrm{E}[N_{A_x}(\tilde{N}_n)] = \lim_{n\to\infty} nP\left(X_i > x\right) ,$$

where $A_x =]0,1]\times]x,\infty[$. Based on the analytical form of $\Lambda(A_x)$ (Eq. (2.14)), the probability of X exceeding a high level (threshold) x is approximated by:

$$P(X > x) \approx \frac{1}{n}\Lambda(A_x) = \frac{1}{n}\left[1 + \xi\left(\frac{x - c_n}{d_n}\right)\right]^{-\frac{1}{\xi}} .$$

By substituting the intensity measure $\Lambda(A_x)$ into the probabilities from Eq. (2.19), the survival function $\overline{H}(y)$ is given by:

$$\overline{H}(y) = \lim_{n\to\infty} P(X > u_n + y \mid X > u_n) = \frac{\frac{1}{n}\Lambda(A_{u_n+y})}{\frac{1}{n}\Lambda(A_{u_n})} , \tag{2.20}$$

where $A_{u_n} =]0,1]\times]u_n,\infty[$ and $A_{u_n+y} =]0,1]\times]u_n + y,\infty[$. The expression of Eq. (2.20) is simplified through substituting the analytical form of $\Lambda(A_x)$:

$$\frac{\Lambda(A_{u_n+y})}{\Lambda(A_{u_n})} = \frac{\left[1 + \xi\left(\frac{(u_n+y)-c_n}{d_n}\right)\right]^{-\frac{1}{\xi}}}{\left[1 + \xi\left(\frac{u_n-c_n}{d_n}\right)\right]^{-\frac{1}{\xi}}} = \left[\frac{1 + \xi\left(\frac{(u_n-c_n)+y}{d_n}\right)}{1 + \xi\left(\frac{u_n-c_n}{d_n}\right)}\right]^{-\frac{1}{\xi}} ,$$

then we have

$$\begin{aligned}\left[\frac{1 + \xi\left(\frac{(u_n-c_n)+y}{d_n}\right)}{1 + \xi\left(\frac{u_n-c_n}{d_n}\right)}\right]^{-\frac{1}{\xi}} &= \left[\frac{1 + \xi\left(\frac{u_n-c_n}{d_n}\right) + \frac{\xi y}{d_n}}{1 + \xi\left(\frac{u_n-c_n}{d_n}\right)}\right]^{-\frac{1}{\xi}} \\ &= \left[1 + \frac{\xi y}{d_n\left[1 + \xi\left(\frac{u_n-c_n}{d_n}\right)\right]}\right]^{-\frac{1}{\xi}} \\ &= \left[1 + \frac{\xi y}{d_n + \xi\left(u_n - c_n\right)}\right]^{-\frac{1}{\xi}} .\end{aligned} \tag{2.21}$$

The threshold-dependent scale parameter is given by $\sigma_{u_n} = d_n + \xi(u_n - c_n)$. Substituting it into the conditional probability Eq. (2.21) yields the survival function:

$$\overline{H}(y) = P(Y > y \mid X > u_n) = \left[1 + \frac{\xi y}{\sigma_{u_n}}\right]^{-\frac{1}{\xi}} .$$

This corresponds precisely to the cumulative distribution function (CDF) of the GPD:

$$H(y) = 1 - \overline{H}(y) = 1 - \left[1 + \frac{\xi y}{\sigma_{u_n}}\right]^{-\frac{1}{\xi}} .$$

This derivation, together with the derivation of the GEV distribution in the previous subsection, establishes the PPP as unifying framework in EVT.

2.6 Other necessary preliminaries

This section shifts to other necessary preliminaries required to understand the subsequent chapters of this thesis, particularly Chapter 6. Specifically, we illustrate two anomaly detection techniques—the one-class support vector machine and its extension, the one-class support measure machine—which are utilized for identifying the unusual points and patterns. Following this, we introduce the concept of the uninorm, a generalized aggregation function that serves as a useful tool for data fusion.

2.6.1 One-class support vector machine

The one-class support vector machine (OCSVM), originally proposed by Schölkopf et al. (2001), is a widely used method for anomaly detection. Unlike classical support vector machines, which separate two classes, OCSVM aims to model the boundary of a single class of normal data. The primary objective of the OCSVM is to find an optimal hyperplane (or boundary) that maximally separates the normal data from the origin. This objective is achieved through two key procedures:

1. **Feature mapping and kernel trick:** The training data is assumed to consist of normal data only. These input data are nonlinearly mapped into a high-dimensional reproducing kernel Hilbert space $\mathbb{H}$, via the mapping function $\phi(\cdot)$. On the one hand, this mapping transforms the data into a space where the normal points can

be linearly separable from the origin. On the other hand, this mapping allows the subsequent optimization problem to be solved efficiently by applying the kernel trick, which only requires computing dot products in the input space without needing the explicit feature mapping.

2. **Optimization:** The optimal decision boundary is determined by solving an optimization problem that balances two goals: maximizing the separation margin between the hyperplane to the origin in space $\mathbb{H}$, and penalizing a small fraction of outliers falling on the wrong side of the boundary.

Consider a training set of N instances $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \mathbb{R}^d$ that can be considered as i.i.d. realizations of a random variable $\boldsymbol{X}$. To find the optimal separating hyperplane, Schölkopf et al. proposed the following optimization problem:

$$\begin{aligned} \min_{\boldsymbol{w},\delta,\rho} \quad & \frac{1}{2}\|\boldsymbol{w}\|^2 - \rho + \frac{1}{\nu N}\sum_{i=1}^{N}\delta_i \\ \text{s.t.} \quad & \langle \boldsymbol{w}, \phi(\boldsymbol{x}_i)\rangle \geq \rho - \delta_i\,, \quad i = 1, \ldots, N \\ & \delta_i \geq 0\,, \quad i = 1, \ldots, N\,, \end{aligned} \tag{2.22}$$

where $\phi(\boldsymbol{x}_i)$ is a feature map that can project each $\boldsymbol{x}_i$ into a high-dimensional feature space. The slack variables δ_i allow points to be at the wrong side of the hyperplane although this is discouraged by the penalizing factor $1/\nu$ in the cost function. The offset ρ represents the distance from the origin to the hyperplane in the transformed feature space and is adjusted during the optimization to maximize the margin while ensuring that the majority of the data lies on or beyond the hyperplane.

The solution of the constrained optimization problem (2.22) results in a decision function $f(\boldsymbol{x}) = \langle \boldsymbol{w}, \phi(\boldsymbol{x})\rangle - \rho$ that can be used to classify a data point as being anomalous or not. In particular, a new data point $\boldsymbol{x}^*$ is classified as normal only when $f(\boldsymbol{x}^*) \geq 0$. Through maximization of a Lagrangian function subject to the so-called Karush–Kuhn–Tucker conditions, the decision function can be computed in terms of a kernel function that operates on pairs of instances:

$$f(\boldsymbol{x}^*) = \sum_{i=1}^{l} \alpha_i \kappa(\boldsymbol{x}_i, \boldsymbol{x}^*) - \rho\,,$$

where l is the number of support vectors ($l \ll N$), and α_i are the Lagrange multipliers

associated with the support vectors. The kernel function is defined as:

$$\kappa(\boldsymbol{x}_i, \boldsymbol{x}^*) = \langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}^*) \rangle \,, \tag{2.23}$$

allowing for the computation of inner products in the high-dimensional feature space without explicitly performing the mapping. This property, known as the kernel trick, significantly reduces computational complexity and enables the model to handle non-linear patterns in the input space effectively. While OCSVM can utilize various kernel functions, such as the linear and polynomial kernels, the Gaussian kernel (often referred to as the radial basis function (RBF) kernel) is by far the most commonly used. The Gaussian kernel is often the default choice in practical anomaly detection problems, given its strong ability to capture the complex, non-linear structure of the normal data.

2.6.2 One-class support measure machine

The one-class support measure machine (OCSMM) is an extension of the OCSVM for the classification of a group $\tilde{\boldsymbol{x}} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ of i.i.d. realizations of a random variable $\boldsymbol{X}$ instead of the classification of an individual point $\boldsymbol{x}$. Compared to OCSVM, whose input is a set of points $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$, the OCSMM operates on a set of groups $\{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_m\}$. In OCSMM, each group $\tilde{\boldsymbol{x}}$ associated with the random variable $\boldsymbol{X}$ is intrinsically represented by a probability distribution $\mathcal{P}$. Consequently, the OCSMM is applied to a space of probability distributions aiming to separate the probability distributions representing normal groups from the probability distributions representing anomalous groups.

To this end, the OCSMM employs two mechanisms analogous to those in OCSVM: probability distributions are nonlinearly mapped into a reproducing kernel Hilbert space $\mathbb{H}$. And, an optimization problem is formulated to find a hyperplane that maximizes the separation margin between the normal probability distributions and the origin in space $\mathbb{H}$. Formally, denote a data space as $\mathcal{X} \subset \mathbb{R}^d$, and let $\mathfrak{P}_\mathcal{X}$ be the set of all probability distributions $\mathcal{P}$ on $\mathcal{X}$. Consider a distribution $F_\mathcal{P}$ on $\mathfrak{P}_\mathcal{X}$. Thus, the set of m i.i.d. realizations of group, i.e. $\{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_m\}$, is denoted by

$$\{\mathcal{P}_1, \ldots, \mathcal{P}_m\} \,,$$

where n_j samples are available from each distribution $\mathcal{P}_j$.

As outlined, the first key mechanism of OCSMM is to map probability distributions into

a high-dimensional feature space. This is achieved through the kernel mean embedding, which maps the probability distributions from the set $\mathfrak{P}_{\mathcal{X}}$ into a reproducing kernel Hilbert space $\mathbb{H}$. Probability distributions are represented as expectation functions in $\mathbb{H}$ (Berlinet and Thomas-Agnan, 2011; Smola et al., 2007). The kernel meaning embedding $\mu : \mathfrak{P}_{\mathcal{X}} \to \mathbb{H}$ is defined with respect to a kernel as:

$$\mu(\mathcal{P}) = \int_{\mathcal{X}} \kappa(\boldsymbol{x}, \cdot) \mathrm{d}\mathcal{P}(\boldsymbol{x}) \,, \tag{2.24}$$

which we will briefly write as $\mu(\mathcal{P}_j) = \mu_{\mathcal{P}_j}$. The mean embedding provides a feature representation of the probability distributions in $\mathfrak{P}_{\mathcal{X}}$. One can show that the inner product on $\mathbb{H}$ is given by:

$$\langle \mu_{\mathcal{P}_j}, \mu_{\mathcal{P}_t} \rangle_{\mathbb{H}} = \iint \kappa(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\mathcal{P}_j(\boldsymbol{x}) \mathrm{d}\mathcal{P}_t(\boldsymbol{y}) \,.$$

The optimization problem of OCSMM can be formulated in $\mathbb{H}$ as follows:

$$\begin{aligned} \min_{\boldsymbol{w}, \delta, \rho} \quad & \frac{1}{2} \|\boldsymbol{w}\|_{\mathbb{H}}^2 - \rho + \frac{1}{\nu l} \sum_{j=1}^{m} \delta_j \\ \text{s.t.} \quad & \langle \boldsymbol{w}, \mu_{\mathbb{P}_j} \rangle_{\mathbb{H}} \geq \rho - \delta_j \,, \quad j = 1, \dots m \\ & \delta_j \geq 0 \,, \quad j = 1, \dots, m \,, \end{aligned} \tag{2.25}$$

where the slack variables δ_j describe margin errors. The hyperparameter ν is a penalty parameter controlling the proportion of anomalous groups. When ν is small, the number of anomalous groups is small compared to the number of normal groups. A large ν implies that there are a considerable number of anomalous groups. In the anomaly detection task, anomalous behaviour is always rare such that ν should be set to a small value.

Similarly to OCSVM, the optimization problem in (2.25) can be solved using the method of Lagrange multipliers leading to a kernel representation of the decision function given by:

$$f(\mathcal{P}) = \langle \boldsymbol{w}, \mu_{\mathcal{P}} \rangle_{\mathbb{H}} - \rho = \sum_{j=1}^{l} \alpha_j K\left(\widehat{\mathcal{P}}_j, \mathcal{P}\right) - \rho \,, \tag{2.26}$$

where l is the number of support vectors ($l \ll m$), and α_j are the Lagrange multipliers associated with the support vectors. The kernel function $K(\cdot, \cdot)$ applied to two distributions $\mathcal{P}_j$ and $\mathcal{P}_t$ is given by:

$$K\left(\mathcal{P}_j, \mathcal{P}_t\right) = \langle \mu_{\mathcal{P}_j}, \mu_{\mathcal{P}_t} \rangle_{\mathbb{H}} \,.$$

For two groups of points $\{\boldsymbol{x}_1^{(j)}, \dots \boldsymbol{x}_{n_j}^{(j)}\}$ and $\{\boldsymbol{x}_1^{(t)}, \dots \boldsymbol{x}_{n_t}^{(t)}\}$ coming from distributions $\mathcal{P}_j$ and $\mathcal{P}_t$ respectively, the empirical form of K can be obtained as:

$$K\left(\widehat{\mathcal{P}}_j, \widehat{\mathcal{P}}_t\right) = \frac{1}{n_j \cdot n_t} \sum_{r=1}^{n_j} \sum_{s=1}^{n_t} \kappa(\boldsymbol{x}_r^{(j)}, \boldsymbol{x}_s^{(t)}) \,. \tag{2.27}$$

2.6.3 Uninorms

The study of aggregation functions is one of the most important subfields of fuzzy set theory (Grabisch, 2009; Beliakov et al., 2007, 2021). Most common aggregation functions are increasing n-ary operations on a real interval that also satisfy some boundary conditions. Of particular importance are associative binary aggregation functions on the real unit interval, as they can be extended unambiguously to a higher number of arguments. Additionally, such associative operations are often assumed to be symmetric and have a neutral element. This leads to the concept of the uninorm (Yager and Rybalov, 1996), which was formally introduced by Yager and Rybalov as a generalization of both t-norms and t-conorms. The significance of uninorms comes from their flexibility and powerful structure, which incorporates both conjunctive and disjunctive bahaviour. Uninorms have increasing importance in a wide range of applications (Calvo et al., 2002; Dombi, 1982; Buchanan and Shortliffe, 1984; Hájek, 1985), including general aggregation tasks, expert systems, and data mining. In this part, we will first review the basic results of t-norms and t-conorms. Subsequently, the formal definition and fundamental properties of uninorms will be introduced. Finally, we will present several representative examples of uninorms.

Definition 2.6.1 (T-norm, Klement et al. (2013))**.** A t-norm is a function $T : [0,1]^2 \to [0,1]$ that satisfies the following four conditions for all $x, y, z \in [0,1]$:

(T1) $T(x,y) = T(y,x)$ (commutativity)

(T2) $T\big(T(x,y), z\big) = T\big(x, T(y,z)\big)$ (associativity)

(T3) $T(x,y) \leq T(x,z)$ whenever $y \leq z$ (monotonicity)

(T4) $T(x,1) = x$ (boundary condition)

Some basic examples of t-norms are presented as follows:

- **minimum** (T_M): $T_M(x,y) = \min(x,y)\,.$

- **product** (T_P): $T_P(x, y) = x\,y\,.$
- **Łukasiewicz t-norm** (T_L): $T_L(x, y) = \max(x + y - 1,\, 0)\,.$
- **drastic product** (T_D):

$$T_D(x, y) = \begin{cases} 0 & \text{, for } (x, y) \in [0, 1]^2, \\ \min(x, y) & \text{, otherwise.} \end{cases}$$

Definition 2.6.2 (T-conorm, Klement et al. (2013))**.** A t-conorm is a function $S : [0, 1]^2 \to [0, 1]$ that satisfies the following four conditions for all $x, y, z \in [0, 1]$:

(S1) $S(x, y) = S(y, x)$ (commutativity)

(S2) $S\big(S(x, y), z\big) = S\big(x, S(y, z)\big)$ (associativity)

(S3) $S(x, y) \leq S(x, z)$ whenever $y \leq z$ (monotonicity)

(S4) $S(x, 0) = x$ (boundary condition)

Some basic examples of t-conorm are presented as follows:

- **maximum** (S_M): $S_M(x, y) = \max(x, y)\,.$
- **probabilistic sum** (S_P): $S_P(x, y) = x + y - x\,y\,.$
- **Łukasiewicz t-conorm** (S_L): $S_L(x, y) = \min(x + y,\, 1)\,.$
- **drastic sum** (S_D):

$$S_D(x, y) = \begin{cases} 1 & \text{, for } (x, y) \in \,]0, 1]^2\,, \\ \max(x, y) & \text{, otherwise.} \end{cases}$$

Moreover, t-norms and t-conorms are dual to each other with respect to the standard negation $N(x) = 1 - x$. A t-conorm S is the dual of a t-norm T if

$$S(x, y) = N\big(T(N(x), N(y))\big) = 1 - T\big(1 - x,\, 1 - y\big)\,.$$

Conversely, a t-norm T is the dual of a t-conorm S if

$$T(x, y) = N\big(S(N(x), N(y))\big) = 1 - S\big(1 - x,\, 1 - y\big)\,.$$

We now formally introduce the definition of a uninorm. Uninorms were proposed to generalize both t-norms and t-conorms into a broader family of aggregation functions, possessing a flexible structure and enhanced modeling capabilities.

Definition 2.6.3 (Uninorm, Yager and Rybalov (1996))**.** A function $U : [0,1] \times [0,1] \to [0,1]$ is called a uninorm if the following properties are satisfied for all $x, y, z \in [0,1]$:

(U1) $U(x,y) = U(y,x)$ (commutativity)

(U2) $U(x,y) \leq U(x,z)$ whenever $y \leq z$ (monotonicity)

(U3) $U\big(x, U(y,z)\big) = U\big(U(x,y), z\big)$ (associativity)

(U4) There exists some element $e \in [0,1]$ called the identity element such that for all $x \in [0,1]$, $U(x,e) = x$. (identity)

The element e is referred to as the neutral element or identity element of the uninorm U. The case $e = 1$ corresponds to t-norms, while the case $e = 0$ corresponds to t-conorms.

The structural properties of uninorms with a non-trivial neutral element $e \in]0,1[$ were extensively studied. A fundamental result is stated below.

Proposition 2.1 (Fodor et al. (1997))**.** Let U be a uninorm with neutral element $e \in]0,1[$. Then

(i) $U(0,1)$ is the zero element of U.

(ii) $U(0,1) \in \{0,1\}$.

(iii) The two associated functions $T_U : [0,1]^2 \to [0,1]$ and $S_U : [0,1]^2 \to [0,1]$ defined by

$$T_U(x,y) = \frac{U(ex,ey)}{e} \tag{2.28}$$

and

$$S_U(x,y) = \frac{U\big(e + (1-e)x,\ e + (1-e)y\big) - e}{1-e} \tag{2.29}$$

are a t-norm and a t-conorm, respectively.

(iv) For any $(x,y) \in [0,1]^2$ such that $\min(x,y) \leq e \leq \max(x,y)$,

$$\min(x,y) \ \leq \ U(x,y) \ \leq \ \max(x,y)\,.$$

(v) U is not continuous.

Every uninorm is bounded by the smallest and largest possible uninorms, denoted $\underline{U}_e$ and $\overline{U}_e$, respectively.

Proposition 2.2 (Fodor et al. (1997))**.** For any uninorm U with neutral element $e \in]0,1[$, the following inequalities hold:

$$\underline{U}_e(x,y) \;\leq\; U(x,y) \;\leq\; \overline{U}_e(x,y)\,,$$

where

$$\underline{U}_e(x,y) = \begin{cases} 0 & \text{, if } x,y \in [0,e[\ , \\ \max(x,y) & \text{, if } x,y \in [e,1]\ , \\ \min(x,y) & \text{, otherwise,} \end{cases}$$

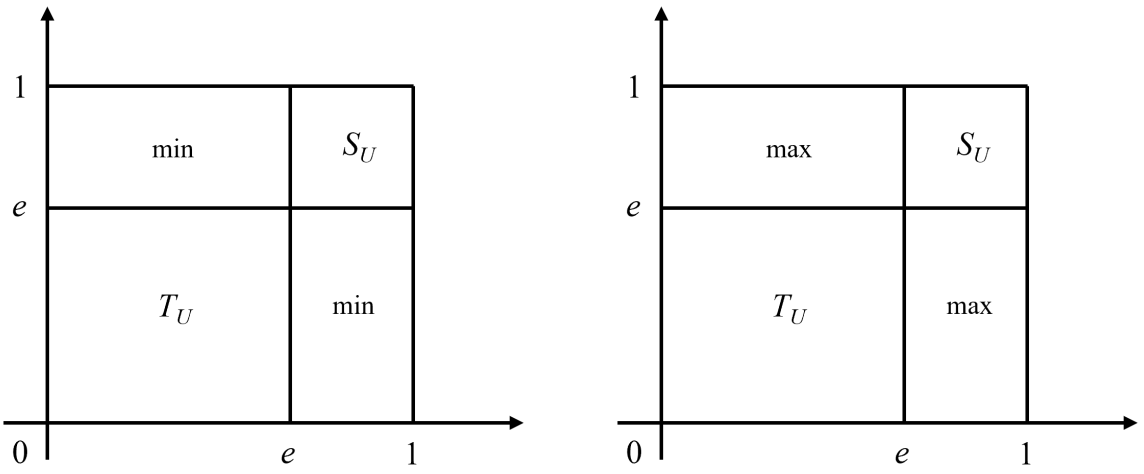
and

$$\overline{U}_e(x,y) = \begin{cases} \min(x,y) & \text{, if } x,y \in [0,e]\ , \\ 1 & \text{, if } x,y \in]e,1]\ , \\ \max(x,y) & \text{, otherwise,} \end{cases}$$

Any uninorm U is either conjunctive (*i.e.*, $U(0,1) = U(1,0) = 0$) or disjunctive (*i.e.*, $U(0,1) = U(1,0) = 1$). The structure of a uninorm with neutral element e is (Fodor et al., 1997): on $[0,e]^2$ it behaves as a rescaled t-norm, thus taking values below the minimum, on $[e,1]^2$ it behaves as a rescaled t-conorm, thus taking values above the maximum, while on the remaining parts of the unit square (*i.e.*, $[0,e] \times [e,1] \cup [e,1] \times [0,e]$) it behaves as a compensatory (also called averaging) operator, taking values between the minimum and the maximum. Given a t-norm and a t-conorm, one can always build a uninorm by considering minimum or maximum as (one of the two extreme) compensatory operator.

The most fundamental classes of uninorms are $\mathcal{U}_{\min}$ and $\mathcal{U}_{\max}$, which utilize the minimum or maximum functions in the mixed region where one input is below e and the other is above e. These classes are characterized by continuity properties on the boundaries, as established in the following theorem.

Theorem 2.5 (Fodor et al. (1997))**.** Suppose that $U : [0,1]^2 \to [0,1]$ is a uninorm with neutral element $e \in]0,1[$, and both functions $x \mapsto U(x,1)$ and $x \mapsto U(x,0)$ are continuous except at point $x = e$. Then U has one of the following forms:



(a) Structure of a uninorm in the class $\mathcal{U}_{\min}$. **(b)** Structure of a uninorm in the class $\mathcal{U}_{\max}$.

Figure 2.4: Visualization of the $\mathcal{U}_{\min}$ (left) and $\mathcal{U}_{\max}$ (right) classes of uninorms, where T_U and S_U denote the underlying t-norm and t-conorm, respectively.

(i) If $U(0,1) = 0$, the uninorm belongs to the class $\mathcal{U}_{\min}$:

$$U(x,y) = \begin{cases} eT_U(\frac{x}{e}, \frac{y}{e}) & \text{, if } x,y \in [0,e]\,, \\ e + (1-e)S_U(\frac{x-e}{1-e}, \frac{y-e}{1-e}) & \text{, if } x,y \in [e,1]\,, \\ \min(x,y) & \text{, if } \min(x,y) \leq e \leq \max(x,y)\,. \end{cases} \tag{2.30}$$

(ii) If $U(0,1) = 1$, the uninorm belongs to the class $\mathcal{U}_{\max}$:

$$U(x,y) = \begin{cases} eT_U(\frac{x}{e}, \frac{y}{e}) & \text{, if } x,y \in [0,e]\,, \\ e + (1-e)S_U(\frac{x-e}{1-e}, \frac{y-e}{1-e}) & \text{, if } x,y \in [e,1]\,, \\ \max(x,y) & \text{, if } \min(x,y) \leq e \leq \max(x,y)\,. \end{cases} \tag{2.31}$$

In both formulas, T_U is a t-norm and S_U is a t-conorm.

2

Part II

Methods

CHAPTER 3

Pattern of exceedances

This chapter presents a core theorem and its proof, proposed by Lixuan An, Bernard De Baets, and Stijn Luca, which establishes the fundamental theoretical framework of this thesis. Chapters 4, 5, and 6 build upon this framework, expanding the theoretical work and demonstrating the corresponding applications.

3.1 The pattern of exceedances

Let X be a univariate random variable with an unknown distribution function F_X. Classical EVT provides a comprehensive description of the asymptotic behaviour of the tails of X. It states that, for a wide class of parent distributions F_X satisfying the convergence condition, the asymptotic distribution of extremes in the tails must belong to the GEV family.

In EVT, extremes typically manifest themselves in two forms. The first manifestation of extremes arises through the block maxima. Consider a block of size n,

$$\tilde{\boldsymbol{X}} = \{X_1, X_2, \ldots, X_n\},$$

where $X_1, X_2, \ldots, X_n$ is a sequence of i.i.d. random variables with a common distribution function F_X. For the block maxima

$$M_n = \max\{X_1, X_2, \ldots, X_n\},$$

if there exist sequences of normalizing parameters c_n and d_n such that the distribution of M_n converges to a non-degenerate limit G as $n \to \infty$, then G belongs to the GEV family

$$\begin{aligned} \lim_{n\to+\infty} P\left(M_n \leq x\right) &= G_\xi(x) \\ &= \exp\left\{-\left[1+\xi\left(\frac{x-c_n}{d_n}\right)\right]^{-\frac{1}{\xi}}\right\}, \end{aligned} \tag{3.1}$$

where ξ is the shape parameter that controls the tail behaviour (Fréchet when $\xi > 0$, Gumbel when $\xi = 0$, Weibull when $\xi < 0$). The details of the estimation of (ξ, c_n, d_n) are discussed in Chapter 2. When the parent distribution F_X belongs to the MDA of the Gumbel distribution, the corresponding shape parameter is $\xi = 0$. In this case, the distribution of the block maxima can be obtained by taking the limit $\xi \to 0$ in $G_\xi(x)$:

$$\begin{aligned} \lim_{\xi\to 0} G_\xi(x) &= G(x) \\ &= \exp\left(-\exp\left(-\frac{x-c_n}{d_n}\right)\right). \end{aligned} \tag{3.2}$$

The second manifestation of extremes moves away from block-based considerations, relying instead on a high threshold to characterize the tail behaviour. Extremes are identified

as those exceedances of X that lie beyond the chosen threshold τ. If the parent distribution F_X satisfies the conditions for the MDA of the GEV distribution, and denoting the right endpoint of F_X by x_F, then the exceedance over a large threshold τ has the following asymptotic distribution derived from Eq. (2.8):

$$\begin{aligned} \lim_{\tau \to x_F} P\left(X \leq x \mid X > \tau\right) &= H_\xi(x) \\ &= 1 - \left(1 + \xi \left(\frac{x-\tau}{\sigma_\tau}\right)\right)^{-\frac{1}{\xi}}, \end{aligned} \tag{3.3}$$

where σ_τ is the threshold-dependent scale. Details on selecting the threshold τ and estimating (ξ, σ_τ) are provided in Chapter 2.

Our analytical framework aims to study the pattern of exceedances that occur within the block. Unlike the separate modeling of block maxima or exceedances discussed above, this perspective places exceedances in a block-based context. Therefore, the exceedances are no longer regarded as individuals, we investigate their collective behaviour within the block. For a block size of n, the pattern of exceedances above threshold τ is

$$\begin{aligned} \tilde{\boldsymbol{X}}^{\text{exc}} &= \{X_1, X_2, \ldots, X_n\} \cap [\tau, +\infty[\\ &= \{X_1^{\text{exc}}, X_2^{\text{exc}}, \ldots, X_K^{\text{exc}}\}, \end{aligned} \tag{3.4}$$

where $K = 0, 1, \ldots$ is a discrete variable denoting the number of exceedances within the block. We use X^{exc} to denote the random variable of the exceedance above τ, and x^{exc} its realization. In the following, we will focus on studying the probabilistic characteristics of the pattern of exceedances within the block.

3.2 Likelihood construction via Janossy densities

To study the pattern of exceedances in a principled way, a formal theoretical framework is required. A good news is that classical EVT establishes the Poisson point process limit, which provides a unified framework connecting both the BM model and the POT model to describe extremes. This result motivates our modeling process and forms the theoretical cornerstone of our work for analyzing exceedance patterns at the block level.

In EVT, a Poisson point process limit links the block maxima and the exceedances by selecting a block-dependent threshold u_n that rises with the block size n. Through this

mapping, the distribution of individual exceedances in Eq. (3.3) takes the form

$$\lim_{n\to+\infty} P\left(X \leq x \mid X > u_n\right) = 1 - \left(1 + \frac{\xi}{1+\xi u}\left(\frac{x-u_n}{\sigma_{u_n}}\right)\right)^{-\frac{1}{\xi}}, \tag{3.5}$$

where $u_n = c_n + ud_n$ and u is a fixed constant near zero. Moreover, the pattern of exceedances above u_n within the block

$$\tilde{\boldsymbol{X}}_n^{\text{exc}} = \{X_1^{\text{exc}}, X_2^{\text{exc}}, \ldots, X_K^{\text{exc}}\} \tag{3.6}$$

converges in distribution to a Poisson point process as $n \to +\infty$, with the intensity controlled by u_n (see Eq. (2.14)).

With this Poisson point process limit in place, we apply the Janossy densities to describe the probabilistic structure of exceedance patterns within a block. The Janossy densities are a fundamental concept in point process theory that specify the probability of observing exactly k points in a region, with those points having a joint spatial density at prescribed locations. It provides a natural way to describe the spatial configuration of exceedance patterns. Building on this idea, we construct a likelihood function for exceedance patterns via their Janossy densities.

To derive this likelihood, we begin by evaluating the density of each observed exceedance. More precisely, the density evaluated at x_j^{exc} is obtained by differentiating the exceedance distribution in Eq. (3.5) with respect to x. This yields the general form of the probability density function of the exceedance. In the special case where the parent distribution F_X lies in the MDA of the Gumbel distribution (i.e., $\xi = 0$), the exceedance distribution in Eq. (3.5) reduces to

$$\lim_{n\to\infty} P(X \leq x \mid X > u_n) = 1 - \exp\left(-\frac{x-u_n}{d_n}\right), \tag{3.7}$$

which corresponds to the exponential distribution with scale parameter $\sigma_{u_n} = d_n$. Differentiating this expression gives

$$p(x_j^{\text{exc}}) \approx \frac{1}{d_n}\exp\left(-\frac{x_j^{\text{exc}}-u_n}{d_n}\right), \tag{3.8}$$

which is the density evaluated at the observed exceedance x_j^{exc}. The Gumbel domain is of particular interest because it covers a broad class of distributions which are commonly encountered in practice, such as the Gaussian, exponential, Weibull, gamma and lognormal

distributions. A detailed discussion of the MDA of the Gumbel distribution has already been provided in Chapter 2.

Having analysed the distribution of individual exceedances, we now shift our focus to the number of exceedances occurring within a block. From the point process perspective established earlier, the collection of exceedances above u_n asymptotically follows a Poisson point process in the limit as $n \to \infty$. Formally, suppose the parent distribution F_X belongs to the MDA of the GEV distribution. Then the number of exceedances within the block, denoted by K, converges in distribution to a Poisson random variable with rate parameter λ. Let μ_k denote the probability of $K = k$, then we have

$$\mu_k = P(K = k) \approx \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k = 0, 1, 2, \dots \tag{3.9}$$

where λ is the expected number of exceedances in the block. In the Gumbel case (i.e., $\xi = 0$), the rate parameter simplifies to $\lambda = e^{-u}$ (the detailed derivation can be found in Chapter 2.5.2).

Based on the above components, we derive the likelihood of the pattern of exceedances via the Janossy representation. The exceedances above u_n are i.i.d. with density $p(x_j^{\text{exc}})$. For any fixed ordering of the k observations $x_1^{\text{exc}}, \dots, x_k^{\text{exc}}$, the joint density is $\prod_{j=1}^{k} p(x_j^{\text{exc}})$. Since a pattern $\tilde{\boldsymbol{x}}^{\text{exc}}$ is an unordered set, the $k!$ permutations of the same configuration must be considered, which introduces the factor $k!$. Combining this with the Poisson probability for the number of exceedances, $P(K = k)$, we obtain the likelihood of a pattern $\tilde{\boldsymbol{x}}^{\text{exc}} = \{x_1^{\text{exc}}, \dots, x_k^{\text{exc}}\}$:

$$\ell_n(\tilde{\boldsymbol{x}}^{\text{exc}}) := \mu_k \, k! \prod_{j=1}^{k} p\big(x_j^{\text{exc}}\big), \qquad k = 0, 1, 2, \dots \tag{3.10}$$

for $k = 0$, the likelihood reduces to

$$\ell_n(\varnothing) = P(K = 0) = \mu_0 = e^{-\lambda}. \tag{3.11}$$

The likelihood $\ell_n(\cdot)$ thus naturally combines the count, magnitude, and permutation components, which is precisely the Janossy representation of a Poisson point process describing exceedances as a pattern within the block.

3.3 Theorem and proof

The likelihood function is derived from the Janossy representation to characterize the patterns of exceedances. In the theorem below, we provide the limiting CDF of the likelihood $\ell_n(\tilde{\boldsymbol{x}}^{\text{exc}})$, forming the framework of developed models in subsequent chapters. The distribution of the likelihood $\ell_n(\tilde{\boldsymbol{x}}^{\text{exc}})$ is asymptotically given by a random variable V of mixed type. Random variables of mixed type are neither discrete nor continuous, but are a mixture of both. The discrete component results from the likelihoods $\mu_0 > 0$ that describe cases when there are no exceedances in $\tilde{\boldsymbol{x}}^{\text{exc}}$. This corresponds to a discontinuity in the CDF that can be described using the Heaviside step function:

$$\mathcal{H}(v) = \begin{cases} 1 & , \text{if } v \geq 0 \\ 0 & , \text{if } v < 0\,. \end{cases} \tag{3.12}$$

When the pattern of exceedances is empty, i.e., for $k = 0$, the block contains no observations above the threshold u_n. In this case, the likelihood reduces to a fixed value

$$\ell_n(\varnothing) = e^{-\lambda}\,. \tag{3.13}$$

Therefore, the discrete component of the mixed-type limiting distribution of V places a point mass at $v = \ell_n(\varnothing)$. The remaining part of the distribution is continuous and is supported on the likelihood values corresponding to patterns with $k \geq 1$. The resulting theorem is stated as follows.

Theorem 3.3. Consider a random variable X with distribution F_X belonging to the MDA of the Gumbel distribution. Then there exist sequences $c_n \in \mathbb{R}$ and $d_n > 0$ such that, the following property holds:

$$\lim_{n\to+\infty} P\left(\frac{X - u_n}{d_n} < x \mid X > u_n\right) = 1 - e^{-x}\,, \tag{3.14}$$

with $u_n = c_n + ud_n$, and $u \in \mathbb{R}$ is the normalized threshold. V_n is the random variable defined as the Janossy likelihood of the pattern of exceedances $\tilde{\boldsymbol{x}}^{\text{exc}} = \{x_1^{\text{exc}}, x_2^{\text{exc}}, \ldots, x_k^{\text{exc}}\}$ with $x_i^{\text{exc}} > u_n$. That is,

$$\begin{aligned} V_n &:= \ell_n(\tilde{\boldsymbol{x}}^{\text{exc}}) \\ &= \mu_k\, k! \prod_{j=1}^{k} p\big(x_j^{\text{exc}}\big)\ . \end{aligned} \tag{3.15}$$

Then V_n converges in distribution to a random variable V with CDF

$$\begin{aligned} G(v) &= \lim_{n\to+\infty} P\left(V_n \leq v\right) \\ &= \mu_0 \mathcal{H}\left(v-\mu_0\right) + \sum_{k\geq 1} \mu_k \Big[1 - \mathcal{E}_{k,1}\Big(-\log \frac{v\cdot (d_n)^k}{\mu_k \cdot k!}\Big)\Big]\,, \end{aligned} \tag{3.16}$$

where $\mathcal{H}(\cdot)$ is the Heaviside step function. $\mathcal{E}_{k,1}(\cdot)$ denotes the CDF of an Erlang distribution with shape parameter k and scale 1, and

$$\begin{aligned} \mu_0 &= \exp\{-\exp\{-u\}\}\,, \\ \mu_k &= \frac{\lambda^k}{k!} e^{-\lambda}\,, \end{aligned} \tag{3.17}$$

where the intensity parameter λ represents the expected number of exceedances in a block, which is $\lambda = e^{-u}$ in the Gumbel case.

Proof. Firstly, we determine the distribution of $W_n = -\log(V_n)$ conditional on the number of exceedances k. Explicitly, we have

$$\begin{aligned} W_n &= -\log\left(\ell_n\left(\tilde{\boldsymbol{x}}\right)\right) \\ &= -\log(\mu_k \cdot k!) + k\log(d_n) + \sum_{j=1}^{k} \frac{x_j^{\text{exc}} - u_n}{d_n}\,. \end{aligned} \tag{3.18}$$

By the limit result in Eq. (3.14), as $n \to +\infty$, the normalized exceedances $\frac{X_i^{\text{exc}}-u_n}{d_n}$ converge in distribution to an exponential distribution, which is characteristic of the MDA of the Gumbel distribution:

$$\frac{X_i^{\text{exc}} - u_n}{d_n} \xrightarrow{d} \text{Exp}(1)\,. \tag{3.19}$$

Therefore, the third term of W_n in Eq. (3.18), given by $\sum_{j=1}^{k} \frac{x_j^{\text{exc}}-u_n}{d_n}$, is the sum of k rescaled exceedances, each with a standard exponential limit distribution.

From the continuous mapping theorem (stating that convergence is preserved by a continuous transformation), the sum of k such independent exceedances converges to the distribution of a sum of k exponential random variables with scale $\sigma = 1$, which is an Erlang distribution with shape parameter k and rate parameter 1. Let the centered

statistic W_n^* be defined by

$$W_n^* := W_n + \log(\mu_k \cdot k!) - k \log(d_n)\,, \tag{3.20}$$

it follows that

$$W_n^* \xrightarrow{d} \text{Erlang}(k, 1)\,. \tag{3.21}$$

As a short summary, the logarithmic transformation of V_n is a core step, as it converts the multiplicative likelihood into an additive form, where the contribution of each exceedance appears linearly in W_n.

The CDF of W_n is derived using the law of total probability. Let $\overline{G}(w)$ denote the limiting CDF of W_n:

$$\begin{aligned}
\overline{G}(w) &= \lim_{n\to+\infty} P\left(W_n \le w\right) \\
&= \lim_{n\to+\infty} \sum_{k\ge 0} \mu_k P\left(W_n < w \mid k\right) \\
&= \sum_{k\ge 1} \mu_k \lim_{n\to+\infty} P\left(W_n^* < w + \log(\mu_k \cdot k!) - k\log(d_n) \mid k\right) \\
&= \mu_0 \mathcal{H}\left(w + \log(\mu_0)\right) + \sum_{k\ge 1} \mu_k \mathcal{E}_{k,1}\left(w + \log(\mu_k \cdot k!) - k\log(d_n)\right)\,,
\end{aligned} \tag{3.22}$$

where $\mathcal{H}(\cdot)$ is the Heaviside step function defined in Eq. (3.12). Transforming back to the original distribution by means of $V_n = e^{-W_n}$, one obtains the desired result for $G(v)$:

$$\begin{aligned}
G(v) &= \lim_{n\to+\infty} P\left(V_n \le v\right) \\
&= 1 - \overline{G}\left(-\log(v)\right) \\
&= \sum_{k\ge 1} \mu_k - \Big[\mu_0 \mathcal{H}\left(-\log(v) + \log(\mu_0)\right) + \sum_{k\ge 1} \mu_k \mathcal{E}_{k,1}\left(-\log(v) + \log(\mu_k \cdot k!) - k\log(d_n)\right)\Big] \\
&= \mu_0\big(1 - \mathcal{H}\left(-\log(v) + \log(\mu_0)\right)\big) + \sum_{k\ge 1} \mu_k\big\{1 - \mathcal{E}_{k,1}\left(-\log(v) + \log(k!\mu_k) - k\log(d_n)\right)\big\} \\
&= \mu_0 \mathcal{H}\left(v - \mu_0\right) + \sum_{k\ge 1} \mu_k\Big[1 - \mathcal{E}_{k,1}\Big(-\log\frac{v \cdot (d_n)^k}{\mu_k \cdot k!}\Big)\Big]\,.
\end{aligned} \tag{3.23}$$

□

CHAPTER 4

Identification and ranking of high-demand areas in a bike-sharing system based on extreme value theory

The content of this chapter is based on the manuscript under peer review "Identifying and ranking high-demand areas in a bike-sharing system based on extreme value theory", jointly written with Bernard De Baets, Lisheng Jiang, and Stijn Luca.

Abstract

Bike-sharing systems provide a convenient and eco-friendly transportation mode in cities worldwide. The continued expansion of urban regions introduces significant challenges in planning docking infrastructure adaptively, which requires accurate identification of emerging high-demand areas. While recent studies have advanced mathematical modeling of demand patterns, they still face limitations in spatial partitioning and computational efficiency. In this work, we formulate the detection of irregular high-demand areas as a group anomaly detection task, and not only address these limitations but also achieve a key advancement over prior studies by enabling the quantitative ranking of candidate areas. We develop a probabilistic framework where candidate areas are characterized by a likelihood model that accounts for their intensity, spatial coverage, and the configuration of their surrounding areas. Unlike existing studies, our spatial partitioning leverages prior knowledge of the operational system, eliminating the need for fine-grained hyperparameter tuning. Furthermore, the model's analytical form enhances its interpretability and practical applicability, and its design ensures low computational complexity. We further conduct a high-level comparison with representative recent methods, demonstrating that our method achieves superior performance. Our approach is applied to nine years (2015–2023) of rental records from the Münchner Verkehrsgesellschaft bike-sharing system. The anomaly extent is quantified across candidate areas, supporting the identification and prioritization of sites for future docking stations, thereby contributing to more data-driven urban mobility planning.

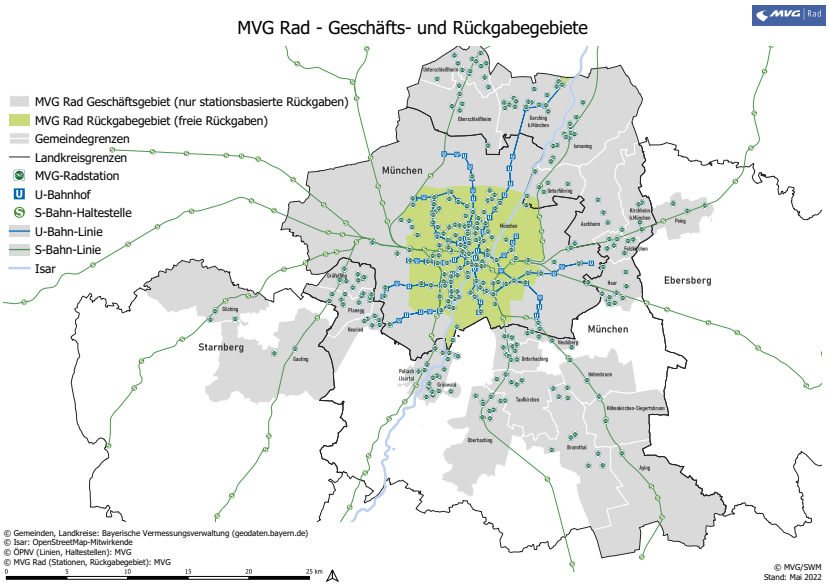
4.1 Introduction

Driven by environmental benefits and the evident enhancement of urban mobility, urban bike-sharing systems (BSSs) have rapidly expanded worldwide over the past decade (Reiss and Bogenberger, 2015; O'brien et al., 2014; Zhu et al., 2024). BSSs have now entered their fourth generation, integrating mobile application support, real-time data, and interoperability with broader mobility networks (Duran-Rodas et al., 2020). These advancements build upon earlier stages, where the first generation introduced free-use bicycles (DeMaio, 2009), the second generation added coin-based access and fixed return stations to improve control (Shen et al., 2018), and the third generation incorporated digital systems enabling user identification and station-based tracking (Fishman, 2016; Du et al., 2019).

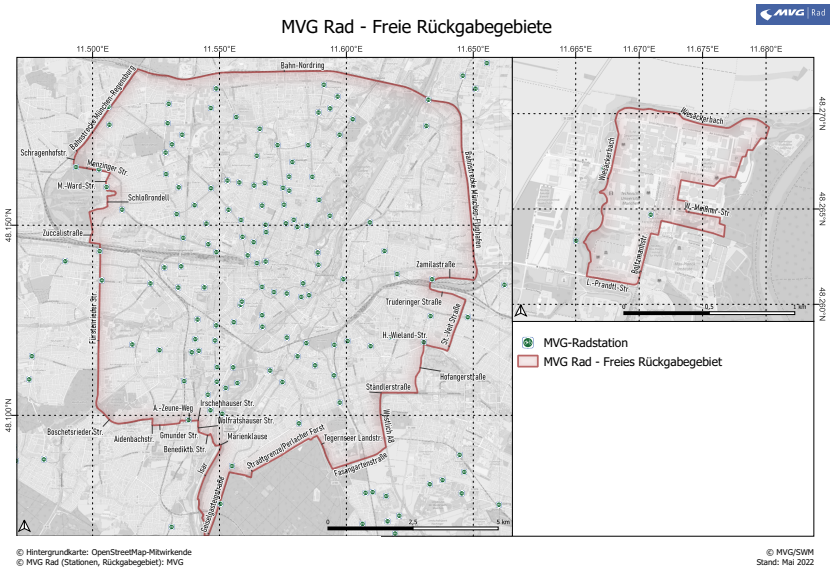
Building upon the aforementioned evolution of BSSs, two dominant operational systems have emerged: docking-based systems and free-floating (or dockless) systems (Zhuang et al., 2025). Docking-based systems were introduced during the early stages of BSSs, with users picking up and returning bikes at designated docking stations. With the development of technology, free-floating systems bring more flexibility by allowing bikes to be parked at arbitrary locations within service regions, without requiring physical docking infrastructure. Despite their convenience, free-floating systems can lead to overoccupation in high-demand areas, as well as higher operational costs due to rebalancing bikes dispersed to remote areas (Du et al., 2019; Shen et al., 2018). Given these limitations, hybrid systems have emerged in recent years and are regarded as a promising alternative for future mobility (Duran-Rodas et al., 2020). These systems are characterized either by integrating designated docking infrastructure into otherwise free-floating models, or by adopting spatially differentiated parking policies across city regions—some allowing for free parking, others requiring docking at specific stations (Zhao and Ong, 2021; Liu and Tian, 2021; Lohrer et al., 2022). Some Chinese cities that were originally dominated by free-floating systems have begun adopting hybrid models by adding docking sites or electronic fences (Wei et al., 2024; Zhao and Ong, 2021). In Europe, a representative hybrid system is the MVG Rad in Munich, operated by the Münchner Verkehrsgesellschaft, which also serves as the use case in this study. As shown in Figure 4.1, bicycles can be picked up and returned freely within the designated green regions (Figure 4.1(a)). Outside these zones, however, users must return bikes to official MVG docking stations (marked by green dots), otherwise they face a penalty.

The effectiveness of a hybrid bike-sharing system is determined by whether newly emerging high-demand areas are served by designated docking stations, rather than being simply incorporated into free-floating regions. Evidence shows that expanding free-floating coverage in such areas often leads to unmanaged bike distribution and operational inefficiencies (Luo et al., 2019; Nikitas, 2019), whereas deploying docking stations in response to emerging demand patterns provides a more sustainable solution (Cheng et al., 2023). These findings motivate our modeling efforts to support strategic planning for station deployment in BSSs.

In this study, we aim to develop an efficient approach not only to identify but also to rank high-demand areas for guiding docking station deployment. This requires modeling demand-driven spatial usage patterns with sufficient granularity and accuracy. When external covariates are not considered, modeling demand-driven spatial patterns typically is an unsupervised learning task. Existing studies have mainly adopted clustering-based



(a) Service area and return regions in the Munich area, with free returns in the inner city and docking station returns elsewhere.



(b) Free-return regions (detail view) with MVG docking stations.

Figure 4.1: Operating areas of the MVG Rad bike-sharing system in Munich (Münchner Verkehrsgesellschaft (MVG)).

approaches to address it (Zhao and Ong, 2021; Wei et al., 2024; Liu and Tian, 2021; Hua et al., 2020; Wang et al., 2022; Du et al., 2019; Zhang et al., 2019), as will be discussed in detail in the literature review (Section 4.2). However, these approaches exhibit multiple critical limitations that reduce their effectiveness in practice:

a) First, as a preprocessing step, the macro spatial partitioning strategies employed

in existing studies—such as uniform grids (Hua et al., 2020; Wei et al., 2024; Chen et al., 2025), administrative boundaries (Du et al., 2019), or the application of k-means clustering to return locations (Wei et al., 2024; Liu and Tian, 2021; Zhao and Ong, 2021)—overlook prior knowledge of existing transportation structures and often rely on hyperparameter tuning (e.g., grid size or the number of clusters).

b) Second, clustering methods such as k-means, DBSCAN, and graph-based algorithms have been widely employed to model spatial demand within partitioned areas, but each comes with notable drawbacks. Among them, DBSCAN—a density-based method—is capable of identifying high-demand areas, but it entails relatively high computational complexity and requires extra hyperparameter tuning (Wei et al., 2024; Liu and Tian, 2021; Zhao and Ong, 2021). Graph-based clustering approaches are also computationally intensive, especially when applied to large-scale spatial networks. This issue is further amplified when such algorithms are integrated with optimization into hybrid frameworks (Wang et al., 2022; Chen et al., 2025). K-means clustering is computationally faster, but its distance-based nature tends to produce clusters of similar size and shape, making it more suitable for spatial partitioning (as previously noted) than for accurate modeling of localized demand patterns.

c) Third, to the best of our knowledge, existing studies focus on detecting anomalous or high-demand areas but do not provide a ranking across candidate areas. In the context of continued urban expansion and constrained resources, the ability to globally rank irregular candidate areas is critical for prioritizing infrastructure investment.

In contrast to existing methods, we propose an analytical framework to model the demand patterns of irregular candidate areas. This study utilizes the geographical location of return records (latitude and longitude) as the primary data source. In contrast to pickup locations, which are heavily affected by operational factors such as bike availability or user proximity to stations, return locations more faithfully reflect users' intended destinations and thus provide a more reliable indication of spatial demand, supporting decisions on permanent infrastructure. Our approach is grounded in EVT, which provides a principled way to characterize the stochastic bahaviour of a univariate process at unusually large (or small) levels. The main contributions of this study are as follows:

1. We introduce an effective partitioning strategy that leverages the prior knowledge of existing docking stations and requires no hyperparameter tuning. Based on this,

we formulate the modeling of demand patterns in irregular areas—an unsupervised learning task—within a group anomaly detection framework.

2. We develop a likelihood-based model using EVT that enables the evaluation of how globally anomalous a candidate area is, considering its joint profile of intensity, spatial coverage, and the configuration of surrounding areas.

3. Our approach achieves not only the identification but also the ranking of irregular candidate areas for new docking station deployment. At the same time, it maintains a low computational complexity, making it suitable for large-scale urban applications.

4. We demonstrate the practical utility of our approach in a case study of the MVG Rad system, a representative hybrid bike-sharing network. The results uncover data-driven patterns in unusual high-demand areas and support infrastructure planning.

The remainder of the Chapter is organized as follows. Section 4.2 reviews the related literature. In Section 4.3, the modeling framework is explained in detail, including an overview of EVT, the construction of the proposed model, and a comparison with representative approaches. Section 4.4 presents the application to the MVG Rad system in Munich. Section 4.5 concludes and outlines directions for future work.

4.2 Literature review

Existing studies on BSSs can be broadly categorized into three directions. The first direction focuses on descriptive analyses of specific bike-sharing systems and associated policy implications. Jensen et al. (2010) provided insights into cycle flows in Lyon. Sarkar et al. (2015) discuss the similarity of bike usage patterns across ten cities. Jonkeren and Kager (2021) conducted a statistical investigation of the parking bahaviour of bicycle–train commuters. The second direction investigates the relationship between demand patterns and various explanatory variables, such as socio-demographic variables and weather conditions. Shen et al. (2018) employed spatial autoregressive models to analyse the impact of the surrounding built environment, access to public transportation, and weather conditions on demand patterns. Du et al. (2019) used random forests to explore the contribution of land-use variables, population, point-of-interest variables and cycling environment variables to demand patterns. Ashqar et al. (2017) predicted the number of available bikes

through regression models from neighborhood availability, temporal factors, and weather conditions. The third research direction, to which our study contributes, emphasizes the development of approaches that directly model demand patterns—without relying on external explanatory variables—to guide future decisions regarding infrastructure design.

Before reviewing specific studies within the third direction, we begin by discussing the role of clustering algorithms, which are among the most widely used techniques in demand pattern modeling. K-means clustering is particularly prevalent due to its low computational complexity and ease of implementation. By assigning each point to the nearest centroid and updating centroids to the mean, it yields compact, roughly convex clusters with similar spread. However, as a distance-based iterative method, k-means clustering can produce clusters with heterogeneous internal densities, which limits its ability to capture spatial density structures. Therefore, k-means clustering has been mainly used to partition the studied area into subregions (Wei et al., 2024; Liu and Tian, 2021; Zhao and Ong, 2021), and, in rare cases, to determine virtual station locations for specific datasets (Hua et al., 2020). In contrast, as a density-based clustering algorithm, DBSCAN is capable of identifying clusters with arbitrary shapes and consistent internal densities. This makes it well suited for detecting high-demand areas. However, due to its higher computational complexity, DBSCAN is generally employed in the final stage of modeling pipelines—after segmentation, preprocessing, and optimization—to identify high-demand areas at a finer spatial scale (Wei et al., 2024; Liu and Tian, 2021; Zhao and Ong, 2021). Other representative methods use graph-based clustering, including community detection, to identify cohesive spatial structures, with computational cost rising on large spatial networks (Liu and Tian, 2021; Wang et al., 2022).

For the third direction, representative studies have modeled demand patterns through unsupervised learning, typically in hybrid frameworks built around clustering algorithms. Caggiani et al. (2018) proposed a spatiotemporal clustering method in which temporal clusters are built first, and subsequently k-means clustering is combined with bilevel optimization to form spatiotemporal clusters. Zhao and Ong (2021) first applied DBSCAN to detect dense clusters, then used k-means clustering to subdivide large clusters under a service-radius constraint to derive geo-fenced parking locations and capacities. Liu and Tian (2021) combined k-means clustering and complex network analysis to conduct macro- and micro-level spatial partitioning, followed by DBSCAN for final demand detection. Wang et al. (2022) constructed a graph representation of the bike-sharing system, where vertices represent bike stations and arcs represent bike flows between stations. A modified fast unfolding algorithm was then applied to this graph to detect communities. Wei et al.

(2024) modeled demand by first partitioning the space with k-means clustering and then applying DBSCAN within each part, and demonstrated the advantage of this approach over using either algorithm independently.

4.3 Model

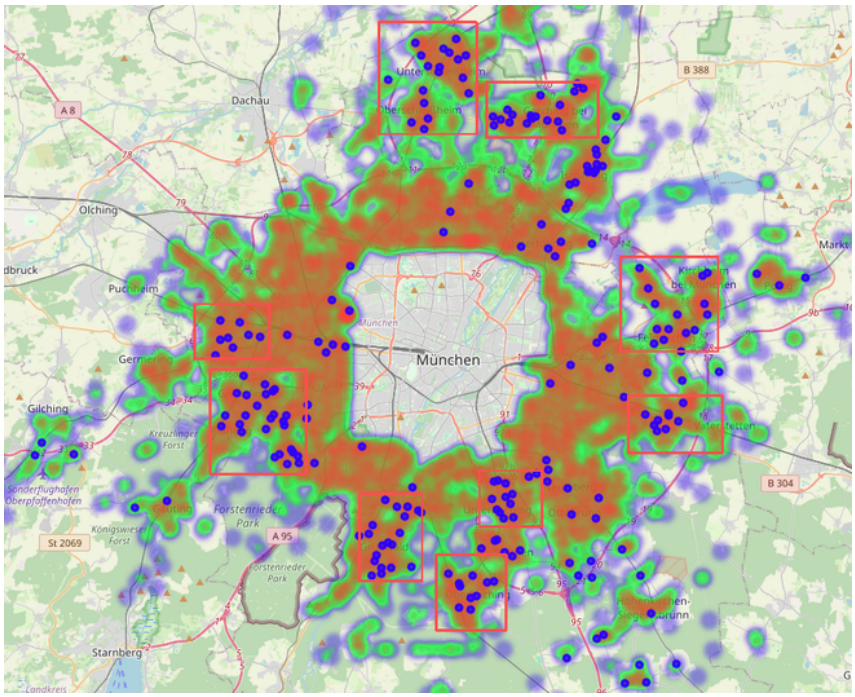
This section presents the proposed model, which identifies and ranks candidate areas based on demand intensity, spatial coverage, and the configuration of their surrounding areas to support docking station deployment. In Section 4.3.1, we formulate the task of modeling demand in candidate areas as a group anomaly detection problem. Then Section 4.3.2 outlines the detailed modeling procedure, including feature extraction, modeling, and anomaly scoring. Finally, in Section 4.3.3, we analyse representative approaches and demonstrate our method's advantage in comprehensively capturing demand characteristics. For illustrative purposes, we use the MVG bike-sharing data from Munich to support the explanation of our modeling framework. Nonetheless, the proposed model is general and not specific to this system.

4.3.1 Formulation of the group anomaly detection task

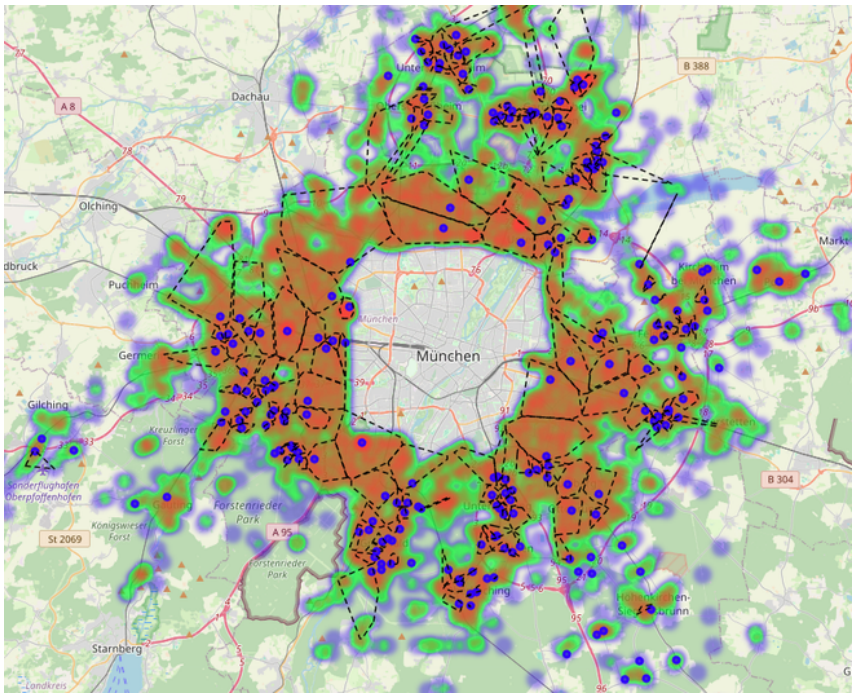
In this subsection, we formulate the problem of detecting areas with unusually high demand in BSSs as a group anomaly detection[1] task. Since direct modeling of the entire citywide area is computationally infeasible, a macro-level partition is required. The resulting partitioned areas are treated as structurally equivalent modeling units so that subsequent modeling can be applied consistently across them. Common partitioning strategies include uniform grids (for example, $50\,\mathrm{m} \times 50\,\mathrm{m}$), administrative boundaries, and k-means clustering. However, these strategies are affected by the modifiable areal unit problem: results can vary with scale and configuration. In practice, the grid size, administrative level, and clustering hyperparameters must be determined through optimization or expert judgment, which increases the computational cost.

In our framework, we propose a data-driven spatial partitioning strategy. The study area

[1] In group anomaly detection, the literature usually distinguishes two forms of anomalous group behaviour: point-based and distribution-based anomalous group behaviour. This chapter focuses on the former and provides a detailed illustration of such bahaviour in the context of the target application. The complete group anomaly detection framework is presented in Chapter 6.



(a) Return activity heatmap with highlighted station-aligned areas.



(b) Spatial partitioning result.

Figure 4.2: Return activity heatmaps and spatial partitioning within non-free return regions of Munich (2015–2023). Blue points denote existing docking stations.

is divided into regions, where each region is defined by a unique docking station and contains all bike return locations closest to it. This strategy is motivated by the station deployment logic. Existing docking stations were deployed to match local demand and

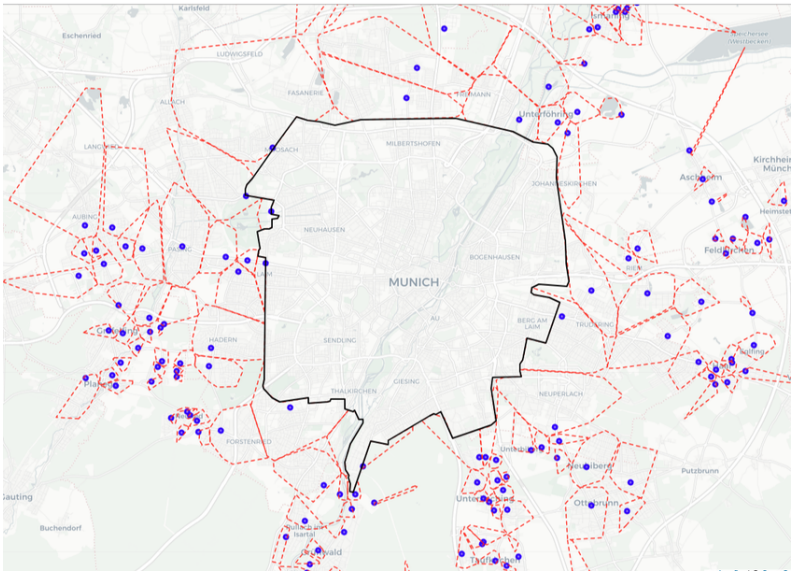
to balance service load across their catchments, which means the layout of stations is expected to reflect the spatial distribution of demand. As a result, the partition yields structurally equivalent modeling units for downstream modeling and avoids complicated hyperparameter tuning required by other approaches. This process is illustrated using the MVG bike-sharing data during the period 2015–2023. The studied area is the non-free return zone. Figure 4.2(a) shows the return activity heatmap within the studied area, with blue points indicating docking stations installed in the past. Figure 4.2(b) presents the partitioning result. Many existing stations remain aligned with local return patterns, for example, those within the red boxes. But in other areas, the current station layout leaves clear gaps in coverage due to the expansion of demand with ongoing urban development.

We now develop a group anomaly detection framework to model demand patterns in candidate areas. In this framework, each region—anchored by a single docking station—is treated as a distinct group. Within each group, bike return locations are further clustered into localized subregions using k-means clustering. This produces spatially coherent subregions that reflect local variation in demand. K-means clustering is adopted for its computational efficiency and its tendency to produce compact, approximately balanced subregions. For each group, the value of k is selected from a candidate set using the elbow criterion. Figure 4.3(a) shows the spatial partitioning of the studied area into groups. Figure 4.3(b) presents an example of a group (Am Hart), with distinct colors denoting the localized subregions.

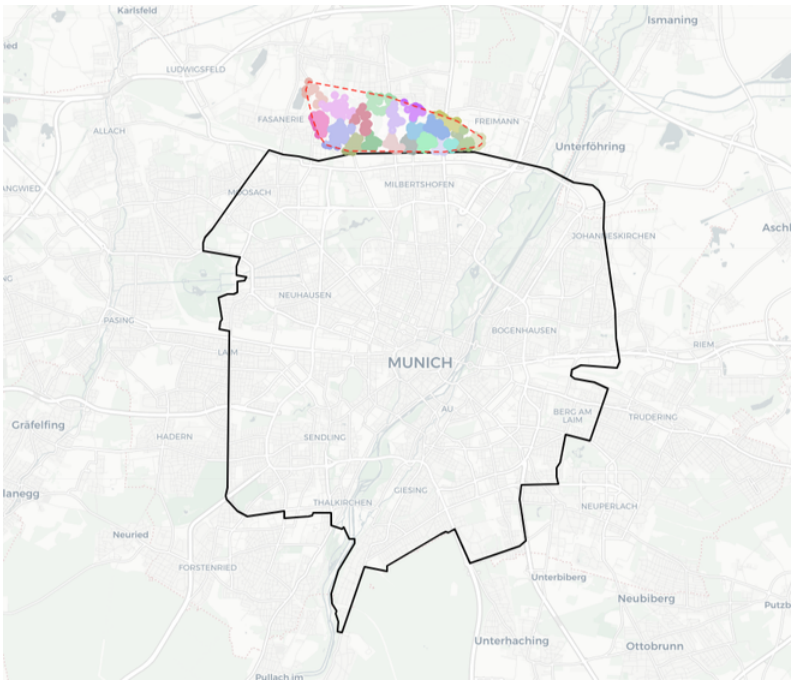
The regions resulting from the spatial partitioning are treated as groups of smaller subregions that are obtained by clustering the geographical coordinates of bike return records. In the next section, the distribution of records across these subregions is modeled to determine whether a region displays anomalous demand patterns and which subregions require additional docking stations.

4.3.2 Subregion-level feature and group anomaly modeling

To describe the magnitude of demand within each subregion, we define a subregion feature. This feature captures both how many return records the subregion contains and how broadly that demand extends in space. These two facets are considered because an unusually high number of return records within a subregion indicates higher demand



(a) Partitioning of non-free return regions into groups anchored by docking stations.



(b) Example group (Am Hart) with localized subregions identified by k-means.

Figure 4.3: Spatial partitioning of non-free return regions into groups and localized subregions. Each group is anchored by a docking station, and within groups, k-means clustering produces subregions.

intensity, and a large spread implies a broad catchment. The feature is defined as:

$$X = \log(C) \cdot \log(E)\,, \tag{4.1}$$

where C denotes the number of returns within a subregion, and E denotes the area of its convex hull. The logarithmic transformation of both intensity and area mitigates disparities in scale and stabilizes variance, preventing either attribute from dominating. The magnitude $X = \log(C) \cdot \log(E)$ amplifies when both intensity and area are high, and it diminishes when either component is small (diminishing marginal effect).

While the feature X summarizes the magnitude of a subregion, assessing a subregion's anomalousness cannot be determined from X alone, as it also depends on how it is embedded in the spatial configuration of surrounding subregions. We illustrate this with an example in Figure 4.4. Suppose the two red-boxed subregions have identical magnitude values X. Their priority for setting new docking stations differs. The upper subregion lies in an area surrounded by many neighboring subregions with high demand but few stations nearby. The lower subregion is surrounded by fewer high-demand neighbors, whereas more stations are located nearby. Therefore, from an urban planning perspective, although their X values are the same, the upper subregion appears more anomalous and would receive higher priority for setting stations.

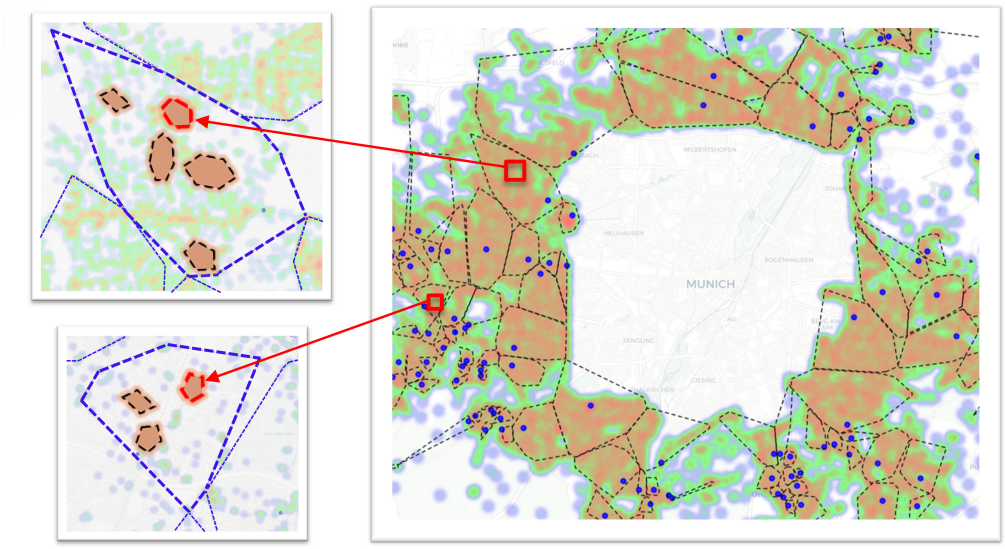


Figure 4.4: Illustration of two subregions with identical magnitude but differing anomalousness due to group-level spatial context. Red boxes indicate their approximate locations.

Next, we will model the configuration within groups using a point process model grounded in EVT to assess subregions' anomalousness. Formally, a group (realization) is represented by a set:

$$\tilde{\boldsymbol{x}} = \{x_1, x_2, \ldots, x_n\}, \tag{4.2}$$

where x_i denotes the magnitude of the i-th subregion within the group. Subregions with magnitude values exceeding a high threshold u_n are considered as areas with an

unusually high demand that need further attention. In EVT, these subregions are referred to as exceedances that capture events in the upper tail of X. The set of exceedances—representing a collective of unusual high-demand subregions within a group—constitutes a realization of a point process on the upper tail of X (see Theorem 2.4). Our model captures the spatial structure of exceedances within a group and enables the computation of the likelihood associated with any observed realization. We further employ the asymptotic distribution of this likelihood, which has been derived in Chapter 3. For the fluency of this chapter, key notations are briefly summarized below, while readers are referred to Chapter 3 for the detailed derivation of the model.

For a group realization $\tilde{\boldsymbol{x}}^{\text{exc}}$,

$$\tilde{\boldsymbol{x}}^{\text{exc}} = \{x_1^{\text{exc}}, x_2^{\text{exc}}, \ldots, x_k^{\text{exc}}\}\,, \tag{4.3}$$

where x_j^{exc} denotes a realization in $\tilde{\boldsymbol{x}}$ that exceeds the threshold u_n. If the tail distribution of X has a Gumbel limit (i.e., $\xi = 0$), the probability density of the j-th exceedance takes the form:

$$p(x_j^{\text{exc}}) = P(X_j^{\text{exc}} = x_j^{\text{exc}}) \approx \frac{1}{d_n} e^{-\frac{x_j^{\text{exc}} - u_n}{d_n}}\,. \tag{4.4}$$

As $n \to \infty$, the exceedances on the upper tail of X converge to a Poisson point process. The number of exceedances K is approximately Poisson with rate λ. Let μ_k denote the probability mass at k:

$$\mu_k = P(K = k) \approx \frac{\lambda^k}{k!} e^{-\lambda}\,.$$

In the Gumbel case (i.e., $\xi = 0$), the Poisson rate parameter simplifies to $\lambda = e^{-u}$, with u being the normalized threshold.

Based on the above components, the likelihood in Janossy representation for observing a group with the aggregate $\tilde{\boldsymbol{x}}^{\text{exc}}$ of k exceedances is expressed as follows:

$$\ell_n(\tilde{\boldsymbol{x}}^{\text{exc}}) := \mu_k k! \prod_{j=1}^{k} p(x_j^{\text{exc}})\,. \tag{4.5}$$

In this expression, μ_k is the probability of observing k exceedances; the factorial term $k!$ corrects for permutations among the k exceedances, ensuring the likelihood is invariant to their ordering; and the product $\prod_{j=1}^{k} p(x_j^{\text{exc}})$ corresponds to the joint likelihood of their observed exceedances under the exponential approximation. The likelihood $\ell_n(\tilde{\boldsymbol{x}}^{\text{exc}})$ serves as a unified measure to reflect the anomalousness of a group in terms of the exceedances

as a whole. The formal statement of the limiting distribution of the likelihood $\ell_n(\tilde{\boldsymbol{x}}^{\text{exc}})$ is presented in the Theorem 3.3, with the proof given in Chapter 3.

In summary, the anomalousness of a group is quantified by the likelihood derived in Eq. (4.5), where lower values indicate greater deviation from expected configurations of exceedances and thus higher anomalousness. Theorem 3.3 provides the asymptotic distribution of the likelihood, enabling us to probabilistically evaluate how rare a given group's likelihood is. Accordingly, we define the anomaly score for a given group $\tilde{\boldsymbol{x}}$ with likelihood value v as the upper-tail probability

$$S(\tilde{\boldsymbol{x}}) = P(V \geq v) = 1 - G(v)\,, \tag{4.6}$$

where $G(v)$ is the limiting cumulative distribution function of the likelihood in Eq. (3.16). The construction of $S(\tilde{\boldsymbol{x}})$ ensures consistency with standard anomaly detection practices, where anomaly scores fall within the range $[0, 1]$, and higher scores correspond to more anomalous bahaviour. For implementation, the scoring procedure is formalized in Algorithm 4.1.

Once the group-level anomaly scores $S(\tilde{\boldsymbol{x}})$ are computed, the proposed framework proceeds to its final stage. Groups are first ranked in descending order of anomalousness to prioritize those most in need of infrastructure improvement. Within each prioritized group, the subregions whose magnitudes exceed the threshold u_n are further sorted based on their magnitudes, allowing for the identification of localized high-demand areas. For each exceedance i in group t, we define the candidate

$$C_{ti} = \big(S(\tilde{\boldsymbol{x}}_t),\, x_{ti}\big)\,.$$

A global ranking of subregions is obtained by ordering the candidates $C_{ti} = \big(S(\tilde{\boldsymbol{x}}_t),\, x_{ti}\big)$ in descending lexicographical order, comparing group scores first and, if equal, subregion magnitudes:

$$C_{ti} \succ_{\text{lex}} C_{t'i'} \iff \Big(S(\tilde{\boldsymbol{x}}_t) > S(\tilde{\boldsymbol{x}}_{t'})\Big) \vee \Big(S(\tilde{\boldsymbol{x}}_t) = S(\tilde{\boldsymbol{x}}_{t'}) \wedge x_{ti} > x_{t'i'}\Big)\,. \tag{4.7}$$

The top R candidates under this order are collected in a set D_R. The overall workflow is illustrated in Figure 4.5, which summarizes the full pipeline from raw return records to actionable deployment sites.

Algorithm 4.1 Computing group anomaly scores from the likelihood of the exceedance set

Input: A collection of m group realizations $\mathbb{A} = \{\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2, \ldots, \tilde{\boldsymbol{x}}_m\}$, where each $\tilde{\boldsymbol{x}}_t$ consists of subregion-level magnitude vectors represented by X. The block size n is set as the maximum number of clusters across groups.

Output: Anomaly scores $\{S(\tilde{\boldsymbol{x}}_1), S(\tilde{\boldsymbol{x}}_2), \ldots, S(\tilde{\boldsymbol{x}}_m)\}$ for all groups.

1: Determine the exceedance threshold u_n using the mean residual life plot constructed from the set of all subregion magnitudes x.
2: Collect all exceedances that satisfy $x > u_n$.
3: Estimate the normalization constants c_n and d_n by fitting an extreme value model to the exceedances.
4: **for** each group $\tilde{\boldsymbol{x}}_t \in \mathbb{A}$ **do**
5: Compute the likelihood $v_t = \ell_n(\tilde{\boldsymbol{x}}_t^{\text{exc}})$ based on Eq. (4.5), and evaluate its cumulative probability $G(v_t)$.
6: Compute the anomaly score $S(\tilde{\boldsymbol{x}}_t) = 1 - G(v_t)$ using Eq. (4.6), which yields a probabilistic measure of anomalousness.
7: **end for**

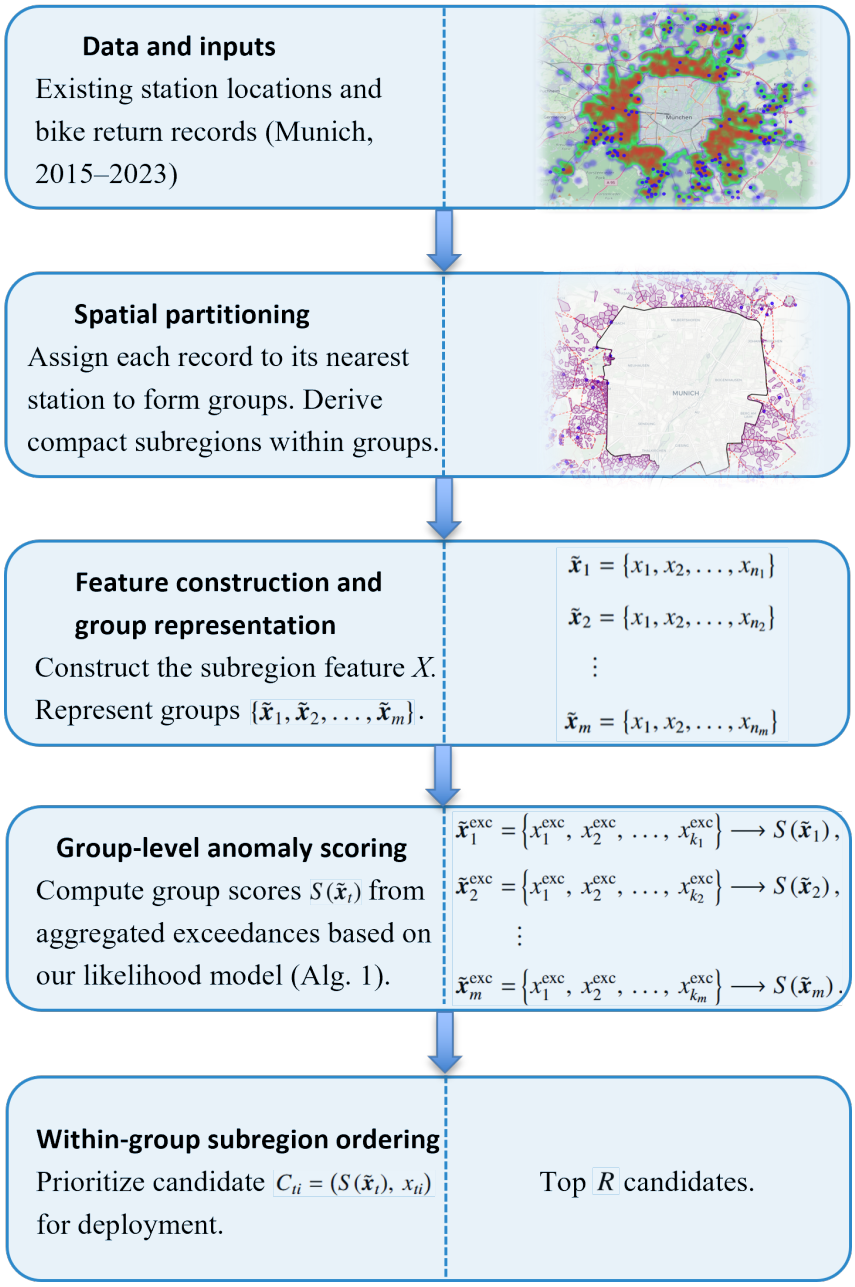


Figure 4.5: Overall workflow of the proposed framework.

4.3.3 Comparison with representative recent methods

We compare our approach with four recent studies on the planning of docking stations. They cover the representative partitioning strategies and typical modeling choices used in practice. Since a direct comparison on a common dataset is not practical, as each study uses a different design of hyperparameters for its own data, we therefore methodologically summarize them in Table 4.1 with respect to four criteria (partitioning strategy, algorithmic structure, ranking capability, and computational complexity).

We first examine how recent studies delineate spatial partitions. Zhang et al. (2019) applies a uniform grid to divide the urban area into predefined spatial units. While this approach is simple, it is fundamentally constrained by the modifiable areal unit problem. The outcome of the model can vary significantly depending on the grid size. In this study, the selected grid resolution ($50\,\mathrm{m} \times 50\,\mathrm{m}$) is determined empirically rather than through a formal optimization process. The other three studies employ data-driven partitioning schemes learned from observations. Zhao and Ong (2021) implemented partitioning using the DBSCAN algorithm, grouping contiguous and demand-similar locations into clusters while treating isolated points as noise. The two DBSCAN hyperparameters—neighborhood radius (ϵ) and minimum points (minPts)—are selected by solving a demand-based optimization problem. Although this enables the partition to reflect local density variations, the exhaustive parameter search entails substantial computational cost. Liu and Tian (2021) and Wei et al. (2024) employ k-means clustering for macro partitioning. In their reported cases, the number of clusters k is fixed *a priori*. However, it is not guaranteed that this way of choosing k generalizes to other applications. Our approach, on the other hand, is data-driven and informed by existing docking stations, leading to a partition that adapts to demand patterns rather than relying on fixed grids. This station-centered strategy eliminates the need to algorithmically derive partitions from data and is computationally efficient.

We next analyse the computational complexity and ranking support of the methods. Zhang et al. (2019) considered fence-site selection as a maximum-coverage problem. The overall complexity is $\mathcal{O}\left(m_g|E|\log|V|\right) + p\mathcal{A} + N_p^2)$. Specifically, they divided the uniform grids into n_g demand cells and m_g candidate cells ($n_g > m_g$). The authors provide a road network $G_r = (V, E)$, where $|V|$ and $|E|$ are the numbers of nodes and edges. The coverage network is constructed by adding an edge between a candidate cell and a demand cell when their network distance (computed on the road network G_r) is within the coverage radius D. This construction process costs $\mathcal{O}\left(m_g|E|\log|V|\right)$. Then selecting p cells from

Table 4.1: Comparison of representative recent methods

Literature	**Partitioning**	**Method**	**Supports Ranking**	**Computational Complexity**
Zhang et al. (2019)	Uniform grids	Maximum coverage selection + DBSCAN	No	$\mathcal{O}\left(m_g\lvert E\rvert \log\lvert V\rvert) + p\mathcal{A} + N_p^2\right)$
Zhao and Ong (2021)	DBSCAN	K-means + Pareto-based tuning	No	$\mathcal{O}\left(GN^2\right)$
Liu and Tian (2021)	K-means clusters	Community detection + DBSCAN	No	$\mathcal{O}\left(N^2\right)$
Wei et al. (2024)	K-means clusters	DBSCAN	No	$\mathcal{O}\left(N^2\right)$
This study	Groups (via stations)	EVT-based model	Yes	$O\left(k^* \log(k^*)\right)$

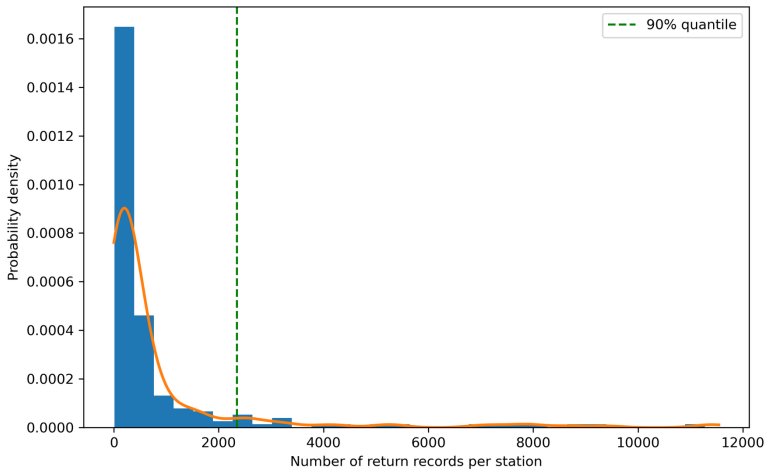
candidate cells by greedy maximum coverage takes $\mathcal{O}(p\mathcal{A})$, where $\mathcal{A}$ is the number of edges in the coverage network. Finally, DBSCAN runs on all return records inside the p selected cells. With N_p records in total, this step costs $\mathcal{O}(N_p^2)$. None of the three terms is guaranteed to be the leading term, so we keep all three. This approach outputs a binary selection of candidate cells, which does not support ranking. Zhao and Ong (2021) applied DBSCAN to all return records to make a macro partition, then ran k-means clustering inside each part. This approach selects ϵ and minPts via a grid search over G settings. Let N be the number of all records. One DBSCAN run costs $\mathcal{O}(N^2)$, trying G settings costs $\mathcal{O}(GN^2)$. The k-means clustering has bounded iterations and is lower-order, so it does not affect the leading $\mathcal{O}(GN^2)$ term. Therefore, the overall complexity is $\mathcal{O}(GN^2)$. This approach compares different settings (ϵ–minPts) rather than ranking areas for a given choice. Although Liu and Tian (2021) and Wei et al. (2024) differ in their intermediate steps, they use the same macro-micro pipeline: k-means clustering partitions the study area into coarse regions, and DBSCAN refines candidates within each region. K-means clustering is of lower-order cost here; the dominant cost is DBSCAN at the micro level. With N total records spread across regions, summing per-region costs gives $\sum_j \mathcal{O}(N_j^2)$, and is upper-bounded by $\mathcal{O}(N^2)$. Neither approach yields a global ranking of candidate areas.

Our method provides a theoretically grounded mechanism for selecting and ranking docking sites with low computational cost. For each group $\tilde{\boldsymbol{x}}_t$ with k_t exceedances, computing the likelihood (Eq. (4.5)) only requires scanning those k_t values once, having $O(k_t)$ complexity. Summed over all groups gives $O(k^*)$, where $k^* = \sum k_t$ denotes the total number of exceedances. The computational complexity is effectively linear in the number of exceedances. The subsequent global ranking of all k^* exceedances by the lexicographical order $\left(S(\tilde{\boldsymbol{x}}_t),\, x_{ti}\right)$ requires $O(k^* \log(k^*))$, which is the dominant term of our method. Compared with other approaches, this cost is substantially lower.

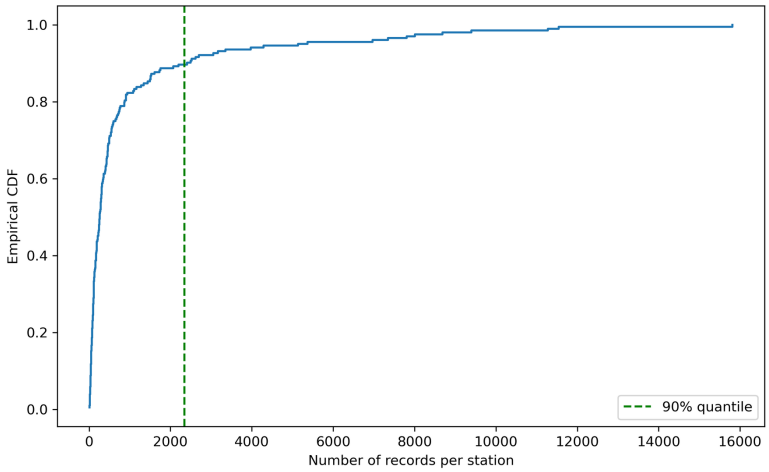
4.4 Case study

In this part the case study of the MVG Rad bike-sharing system in Munich is analysed. Section 4.4.1 describes the dataset. Section 4.4.2 applies the proposed framework to identify and rank anomalous subregions for docking station deployment. Section 4.4.3 examines seasonal demand variation and takes into account season-specific overloads in deployment recommendations.

4.4.1 Dataset and preprocessing



(a) Histogram and kernel density estimate, with dashed line marking the 90-th percentile.



(b) Empirical cumulative distribution function with dashed line marking the 90-th percentile.

Figure 4.6: Distribution of return records per station in the MVG Rad system.

This study uses official trip records from MVG Rad, the public bike-sharing system in Munich operated by the Münchner Verkehrsgesellschaft, accessible at `https://www.mvg.de/ueber-die-mvg/unsere-fahrzeuge/mvg-rad.html`. The dataset covers rides from 2015 to 2023 and contains 4.76 million trip records in total. Annual counts are 49.8 thousand (2015), 342.4 thousand (2016), 419.0 thousand (2017), 435.9 thousand (2018), 753.7 thousand (2019), 721.8 thousand (2020), 619.6 thousand (2021), 709.1 thousand (2022), and 710.1 thousand (2023). Time stamps are recorded in local time (CET/CEST), and locations are given as latitude and longitude in WGS 84. Each trip record includes a unique ride identifier, start and end time stamps, origin and destination coordinates, and

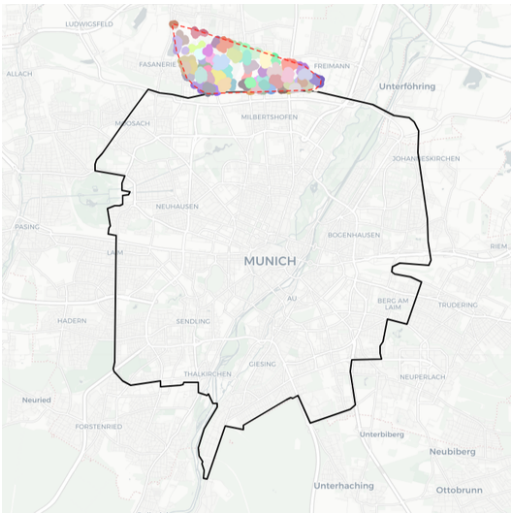
flags indicating whether the ride began or ended at a docking station. The MVG Rad system follows an operational model where bikes can be freely picked up and returned within designated areas, while returns outside these areas are restricted to docking stations, as detailed in Section 4.1.

Preprocessing consists of steps that ensure data integrity and a consistent format for subsequent analyses. Records with missing return coordinates were excluded. Outliers with unrealistic coordinates (e.g., coordinates in polar regions) were removed. The indicators of start and return at docking stations were converted to a consistent binary (0/1) format. Extra spaces in some records were removed to maintain consistent formatting and prevent parsing issues during programming. All analyses were performed in Python 3.10.13 using JupyterLab on Ubuntu Linux. Computations were executed on an Intel Xeon w9-3475X with 36 cores and 72 threads.

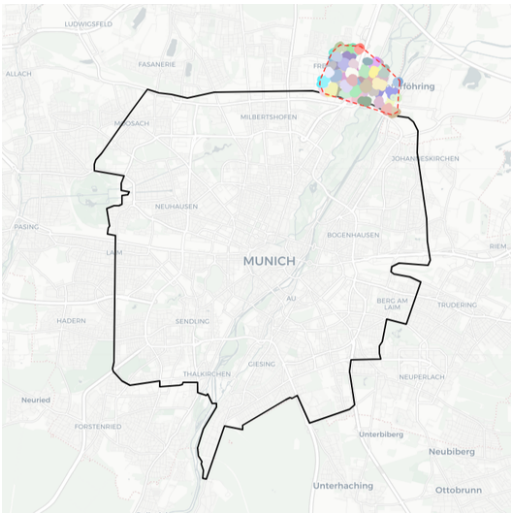
4.4.2 Detection of high-demand subregions and spatial ranking

In our experiments, the studied area is the non-free return region of the MVG Rad system. This system contains 180 docking stations that serve as valid group anchors, with the resulting partition shown in Figure 4.3. Figure 4.6 shows the density and cumulative distribution of the number of return records, obtained across the subregions. The distribution is strongly right-skewed, with a small subset of stations above the 90-th percentile accounting for those with an unusually high number of return records. Given the high demand for this subset of stations, an increased need for docking infrastructure in their neighborhood is expected.

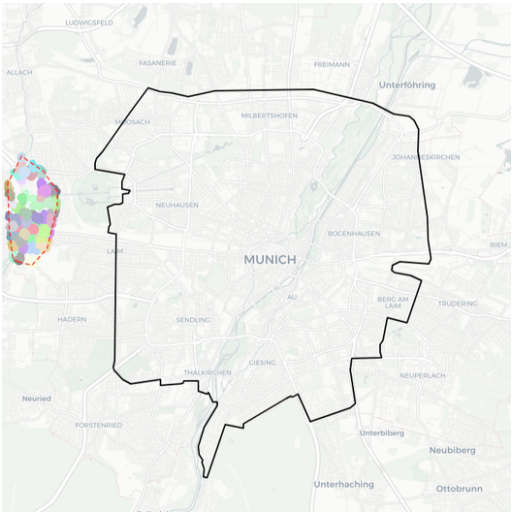
We apply the approach outlined in Section 4.3. To this end, the study area is first partitioned into regions as shown in Figure 4.3, where each existing station belongs to a unique region. Within each region, k-means clustering is performed on the return record data, yielding a set of subregions. We use the elbow method to select the optimal number of clusters. An upper bound on the number of clusters is imposed via a minimum cluster size to prevent over-segmentation. This constraint is an application-level safeguard and is optional, not part of the core method. We use the 10-th percentile of the distribution of returns per station, which equals 45, as the reference minimum cluster size. Limiting the number of clusters in this way leverages the k-means objective to naturally discourage unstable small clusters. Examples of six groups and their subregions are shown in Figure 4.7.



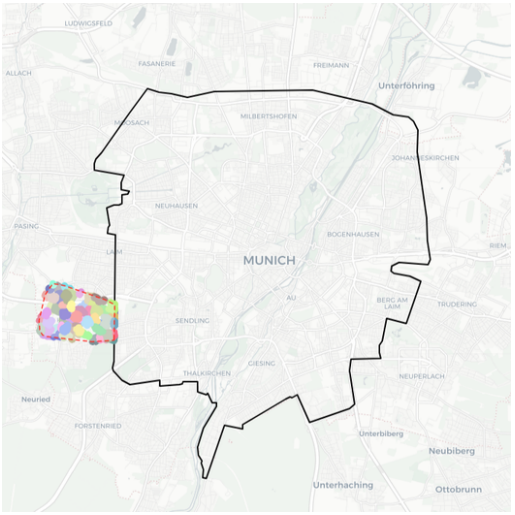
(a) Am Hart



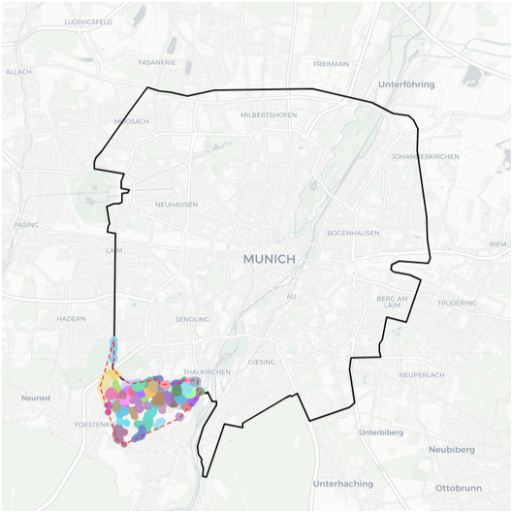
(b) Freimann



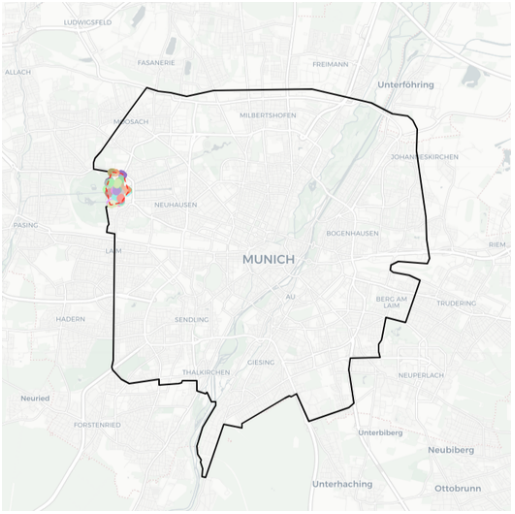
(c) Pasing



(d) Großhadern

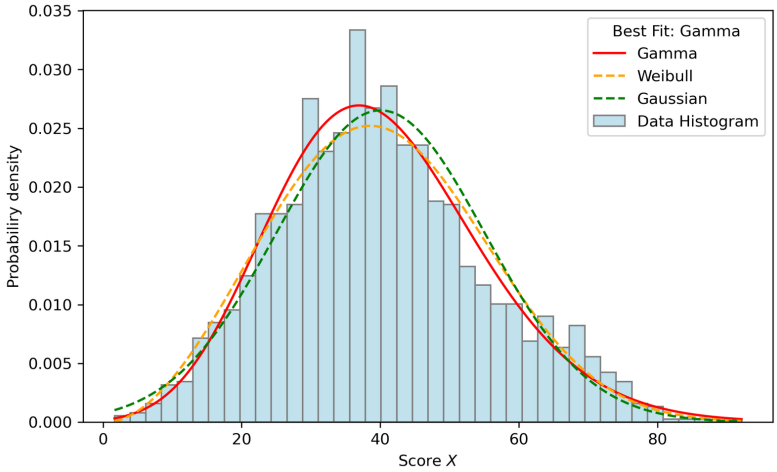


(e) Mobile Station

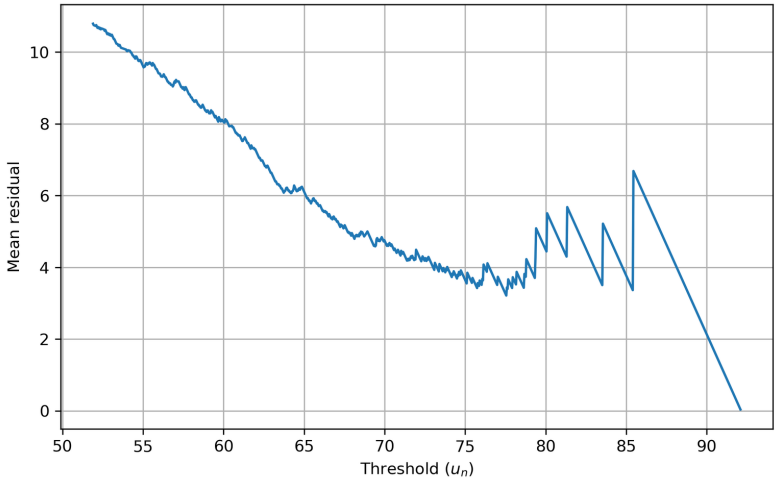


(f) Schloss Nymphenburg

Figure 4.7: Examples of station groups and their subregions obtained by k-means clustering.



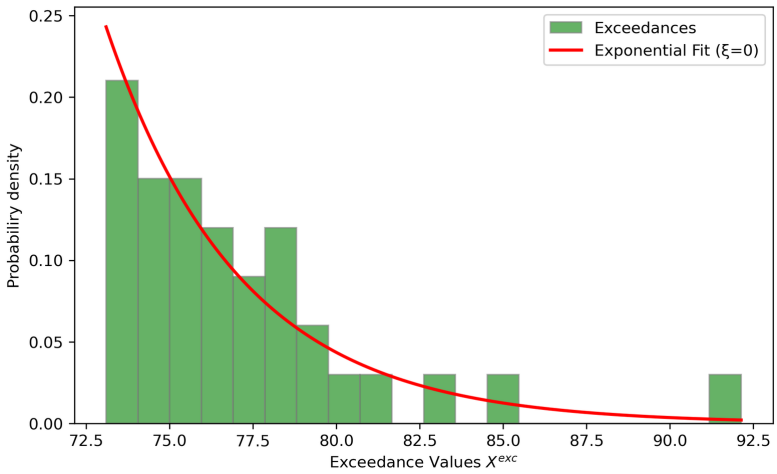
(a) Histogram of X with fitted Gamma, Weibull, and Gaussian distributions.



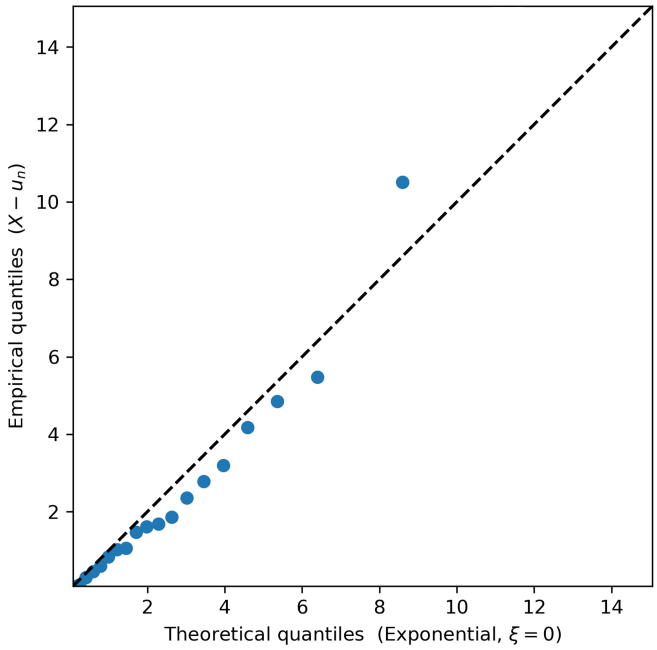
(b) Mean residual life plot.

Figure 4.8: Distribution fitting and threshold selection for X.

The magnitude feature X (Eq. (4.1)) is constructed to characterize each subregion by combining intensity and areal coverage, and it serves as the univariate input for the subsequent modeling. Figure 4.8(a) presents the histogram of X across all subregions. Rather than aiming for a perfect parametric fit, here the goal is to assess whether the tail of X has a Gumbel limit, an assumption underlying our extreme value analysis. To validate this assumption, we examine the distribution of exceedances over threshold u_n using the mean residual life plot (Figure 4.8(b)). For a suitable high threshold, the plot should exhibit an approximately linear trend, which supports the robust estimation of the parameters of the GPD. In Figure 4.8(b), the final linear behaviour begins near 72, so we set $u_n = 72$. Figure 4.9(a) shows that the exceedances agree with a Gumbel fit. As a further check, the Q–Q plot in Figure 4.9(b) compares the empirical quantiles of $X - u_n$



(a) Distribution of exceedances above u_n.

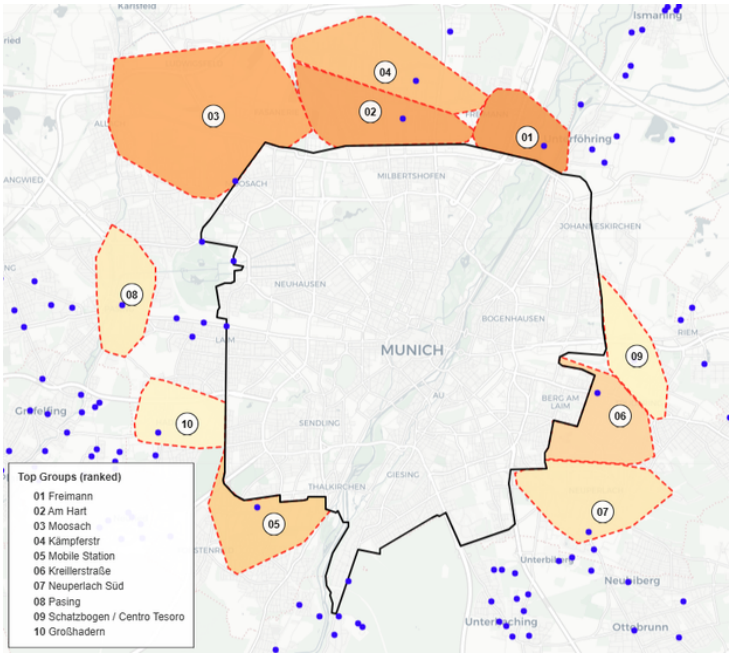


(b) Q-Q plot.

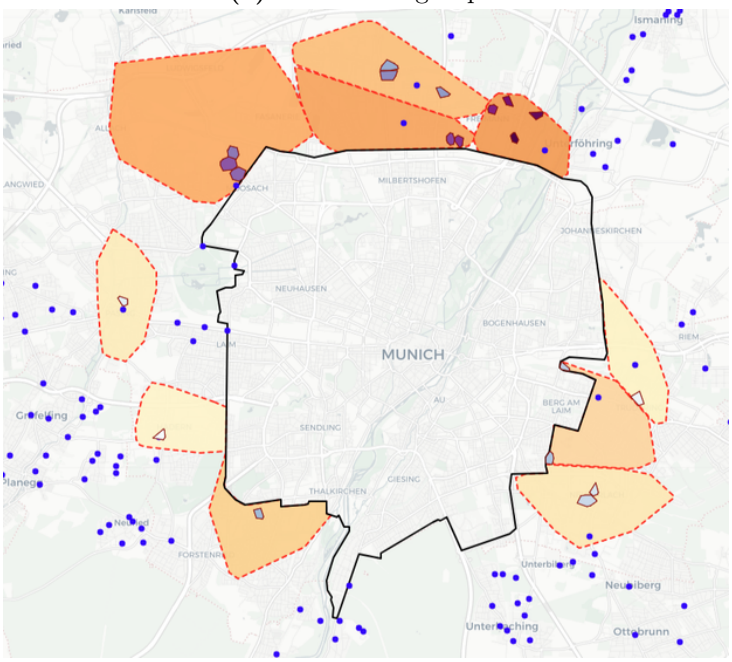
Figure 4.9: Diagnostic plots for assessing the Gumbel domain of attraction of the X.

with the theoretical exponential quantiles. The points lie close to the diagonal, which supports the Gumbel assumption.

Based on the results presented above, we apply the proposed approach to the case study. Figure 4.10 summarizes the ten most anomalous groups. Figure 4.10(a) shows anomalous groups in an orange color scale, where darker shades indicate higher group anomaly scores. To avoid confusion from similar colors on a static map, each group is labeled 1 to 10 to



(a) Anomalous groups.



(b) Anomalous subregions.

Figure 4.10: Detected anomalous groups and subregions prioritized for docking station deployment.

indicate its rank. Figure 4.10(b) displays the anomalous subregions ($X > u_n$) in purple within each orange group. Priority is global and two-level: groups are ranked by the group anomaly score, then subregions (exceedances) are ordered by X within each group. Note that the purple intensity follows this two-level order: darker always means higher priority, and no subregion in a lower-ranked group appears darker than any subregion in a higher-

Table 4.2: Tabular summary of subregions (centroids and names) within the ten highest-ranked anomalous groups.

Rank	Group (Station)	Centroid (lat, lon)	Subregion Name
①	Freimann	48.19315, 11.61520	Freimann, A 9
		48.19917, 11.62391	Kleinlappen, Josef-Wirth-Weg 18
		48.20300, 11.61246	Kieferngarten, Kieferngartenstraße 11
		48.20038, 11.60639	Neufreimann, Heidemannstraße
②	Am Hart	48.19201, 11.59342	Freimann, Maria-Probst-Straße 14
		48.19309, 11.58999	Freimann, Maria-Probst-Straße 15
③	Moosach	48.18679, 11.50307	Moosach-Bahnhof, Teplitzer Weg 3
		48.18402, 11.50660	Alt Moosach, Dachauer Straße 312a
		48.18980, 11.50475	Alt Moosach, Wildermuthstraße 85
④	Kämpferstraße	48.21015, 11.56569	Nordhaide, Golddistelanger 8
		48.20538, 11.58679	Freimann, Ingolstädter Straße
		48.21222, 11.56623	Nordhaide, Graslilienanger 7
⑤	Mobile Station	48.09549, 11.51490	Freiland, Kistlerhofstraße 142
⑥	Kreillerstraße	48.11010, 11.62881	Ramersdorf, Leinbergerstraße
		48.13386, 11.63428	Josephsburg, Truderinger Straße 55
⑦	Neuperlach Süd	48.09877, 11.64357	Neuperlach, Von-Knoeringen-Straße 8
		48.10224, 11.64642	Neuperlach, Thomas-Dehler-Straße 9
⑧	Pasing	48.15105, 11.46202	Kolonie I, Wensauerplatz 10a
⑨	Schatzbogen/Centro Tesoro	48.12508, 11.66361	Trudering - Riem, Truderinger Straße
⑩	Großhadern	48.11583, 11.47643	Neuhadern, Würmtalstraße 136

ranked group. Table 4.2 provides the results in tabular form. It lists the group name, the centroid coordinates of each subregion, and the subregion name. Each subregion is given a readable geographic name constructed from the centroid's administrative context and nearest street. The numerical labels in Figure 4.10 correspond to the Rank column in the table.

Spatially, the highest-ranked groups cluster along Munich' s northern axis (Freimann, Am Hart, and Moosach), with additional high-priority groups along the eastern corridor and a southern outlier at Mobile Station. This pattern is consistent with employment growth, residential expansion, and improved public transport access. The BMW Research and Innovation Center opened in September 2020 in Milbertshofen and Am Hart, strengthening a major commuter attractor (BMW Group, 2020). The area around Freimann has

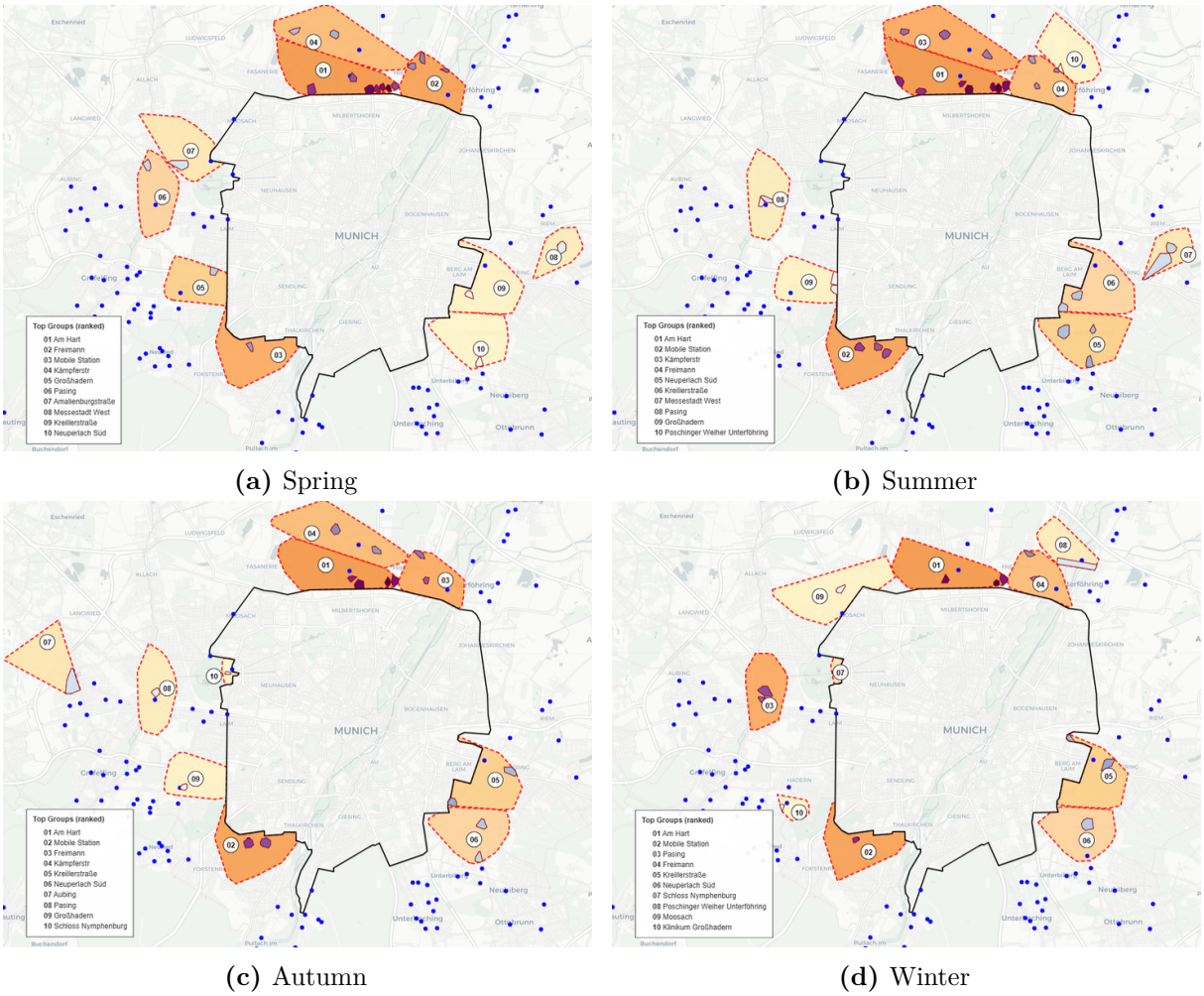
seen substantial residential growth, and the Neufreimann (Bayernkaserne) redevelopment is planned to add about 5,500 dwellings for up to 15,000 residents by 2030 (Landeshauptstadt München, 2025). Accessibility also improved with the opening of the U3 metro line extension to Moosach in 2010, which reinforced Moosach as a hub (Münchner Verkehrsgesellschaft (MVG), 2024). To the east, the pattern aligns with the continued residential development in Messestadt Riem. After a 2021–2022 planning competition, construction of the fifth residential phase began in 2024 and has been progressing since. To the south, renewal of the LMU Klinikum Großhadern campus after the 2021 design competition aligns with the single high-priority group there. Taken together, these employment growth, new housing, and public transport improvements explain why demand concentrates in the north and east and why earlier placement of docking stations no longer matches emerging trip patterns.

At the finer scale, subregions highlight where return records actually cluster. In Freimann, high-ranked subregions lie along the A9 motorway corridor and the Kieferngarten–Heidemannstraße axis (an east–west arterial near Kieferngarten U-Bahn), where returns are more frequent near overpasses and station entrances. In Am Hart, the two Maria-Probst-Straße subregions are situated along a commercial street with shops and offices. In Moosach, Teplitzer Weg and Dachauer Straße are near the Moosach station, so ride endings are concentrated there. In Kämpferstraße, subregions at Golddistelanger and Graslilienanger are situated near exits towards Ingolstädter Straße, and a third subregion lies on the arterial. With limited crossings yet multiple bus and metro access points, ride endings cluster at these gateways. Along the eastern corridor, Ramersdorf–Leinbergerstraße and Josephsburg–Truderinger Straße follow the same arterial-frontage pattern. In the south-east, the two Neuperlach Süd subregions on Von-Knoeringen-Straße and Thomas-Dehler-Straße bridge offices and residential blocks. Schatzbogen/Centro Tesoro lies on Truderinger Straße in the eastern employment belt. Großhadern falls on Würmtalstraße near the university hospital campus.

4.4.3 Seasonal analysis

In the previous sections, the MVG Rad records were analysed across the complete period of nine years to support stable long-term infrastructure planning. As a complement, we will extend our analysis by applying the proposed framework to seasonal data. Unlike divisions based on time of day or weekday cycles, seasonal divisions capture temporal shifts in mobility that are particularly meaningful for planning temporary infrastructure.

This seasonal analysis serves two main purposes: (i) to examine whether the anomalous spatial patterns identified during the full period remain stable across different seasons; and (ii) to detect seasonal mismatches that can inform temporary docking stations or flexible geo-fencing measures.



(a) Spring **(b)** Summer

(c) Autumn **(d)** Winter

Figure 4.11: Seasonal variation in anomalous groups and subregions detected by the proposed framework across spring, summer, autumn, and winter.

Figure 4.11 illustrates the seasonal outcomes of our method, revealing both consistency and fluctuation across the four meteorological seasons. A stable belt in the north from Am Hart to Freimann appears in every season with high ranks. The western fringe remains visible across seasons, although with lower intensity than the northern belt. South-eastern groups such as Kreillerstraße and Neuperlach Süd are present in most seasons but display larger variation in rank and subregion extent. The persistence of Am Hart and Freimann indicates structural demand, as these areas combine residential density, employment sites, and multiple transit access points. Their subregions are compact and recur in all four seasons, which is consistent with regular commuting and interchange traffic rather than

season specific peaks or short-lived event surges.

Some detected subregions exhibit seasonal variability. In summer, subregions extend along the north and the northeast, consistent with longer leisure trips, better cycling conditions, and increased evening activity. Seasonal dock extensions and flexible geo-fencing are recommended near parks and river corridors. During spring and autumn, the overall pattern is similar to the full period baseline, where more western groups are present in the top rank. In winter, Pasing attains its highest rank and Moosach enters the top list. The detection of season-specific anomalies provides actionable guidance for deploying temporary docking stations or adjusting geo-fence boundaries in response to local demand shifts.

4.5 Conclusions and future work

This study presents a principled framework for identifying and ranking irregular high-demand areas for bike-sharing systems, with the aim of supporting docking station deployment. By leveraging EVT and incorporating prior knowledge of the operational system, our method addresses limitations of existing approaches. First, spatial partitioning strategies depend on hyperparameters that require tuning. Second, many detection models incur a high computational cost. Third, most methods do not provide a principled ranking across candidate areas. The proposed likelihood-based model considers demand intensity, spatial coverage, and neighborhood configuration within a unified framework, allowing for both detection and prioritization of candidate areas. Application to the MVG Rad system in Munich demonstrated the model's effectiveness in uncovering meaningful demand patterns and supporting strategic infrastructure planning. Future work could extend the framework to capacity-sizing of candidate sites in the high-priority groups, which involves determining the optimal number of docks and bikes for each station.

CHAPTER 5

An integrated approach for assessing the intensity, duration and frequency of heat waves

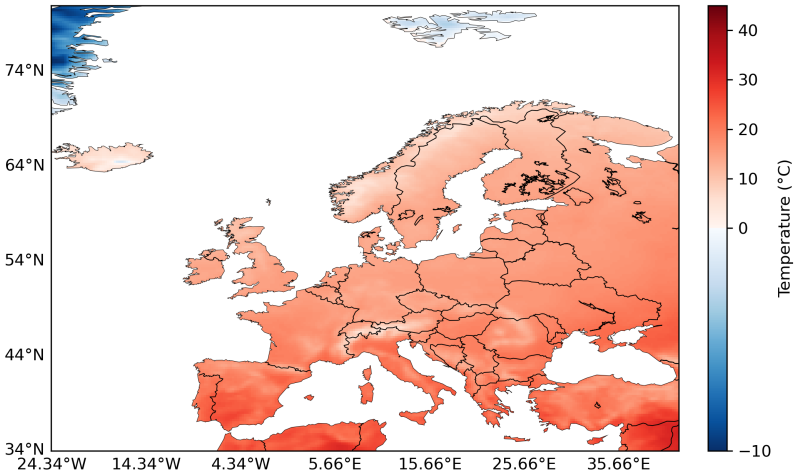
The content of this chapter is based on the manuscript under peer review "An integrated approach for assessing the intensity, duration and frequency of heat waves using extreme value theory", jointly written with Baoying Shan, Bernard De Baets, and Stijn Luca.

Abstract

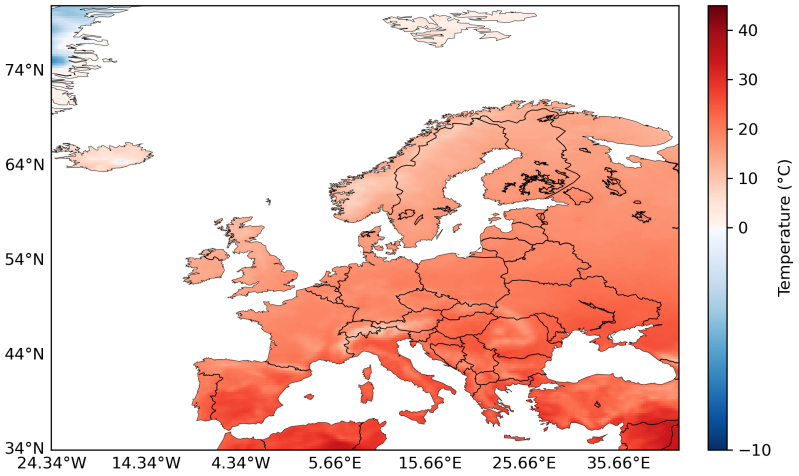
A quantitative understanding of the intensity, duration, and frequency of extreme heat waves is critical for developing policies and designing infrastructure to protect against their impacts. However, such characterization poses several challenges given substantial regional differences in temperature patterns and human acclimatization. On the one hand, even within the same region, comparisons across heat waves are difficult due to variations of day-to-day intensity and overall duration. On the other hand, modeling the frequency of extreme heat waves is challenging because it requires simultaneously considering the sequence of temperatures and the duration, which involve continuous, discrete, and temporal characteristics. In this work, we develop a probabilistic approach to evaluate heat waves in terms of their intensity, duration, and frequency. In contrast to existing studies, this approach is implemented through a novel univariate likelihood model based on extreme value theory, avoiding commonly used multivariate approaches that are less effective in quantifying heat wave risk and are more difficult to interpret. The proposed model is validated through both theoretical proof and experimental evaluation. Building on this, we apply the model to six major European cities using ERA5 data, achieving a systematic analysis and comparison of historical heat waves with varying durations and intensity profiles, and demonstrate how it can be used to construct intensity-duration-frequency curves to support heat wave risk assessment.

5.1 Introduction

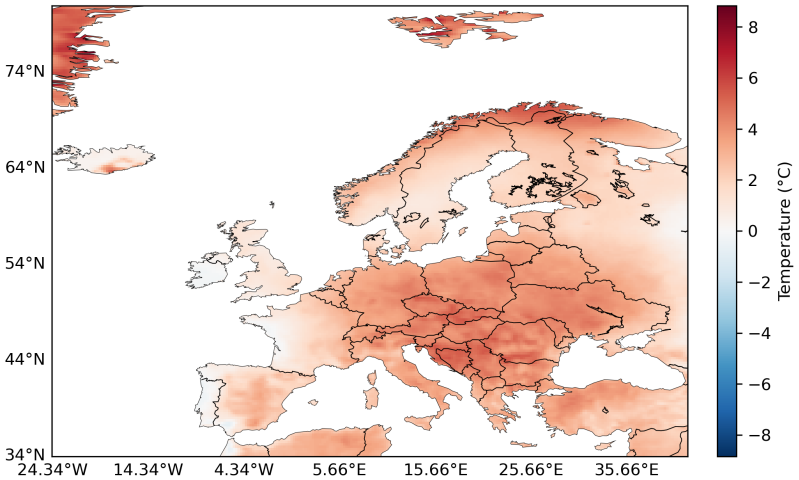
Heat waves are among the most deadly climate-related hazards, posing significant risks to human health and society (Ballester et al., 2023; Thompson et al., 2023). In North America, for instance, extreme heat-related events caused an average of approximately six deaths per 100,000 residents annually between 2000 and 2019 (Thompson et al., 2022; McKinnon and Simpson, 2022). Moreover, driven by anthropogenic climate change and rising summer temperatures, heat waves are projected to become more frequent, more intense, and more prolonged (Lee et al., 2023). Figure 5.1 illustrates the change in summer temperatures in Europe between 1940 and 2024. Therefore, developing reliable tools for risk assessment is essential to support effective mitigation and adaptation policies that address the potential impacts of heat waves. This holds especially for vulnerable regions with large populations unaccustomed to extreme heat events.



(a) Summer mean temperatures in 1940



(b) Summer mean temperatures in 2024



(c) Difference between 2024 and 1940

Figure 5.1: Long-term changes in Europe's summer mean temperatures (1940–2024)

A heat wave is commonly described as a period of consecutive hot days during the summer season, with daily temperatures exceeding a certain threshold. Several climatological studies analyse heat waves using a fixed duration, commonly requiring between three and six consecutive days of unusual high temperatures, with a subsequent analysis of the associated intensity measures (Meehl and Tebaldi, 2004; Stefanon et al., 2012; Schoetter et al., 2015; Ren et al., 2023). However, there is no universally accepted definition regarding either the threshold or the minimum number of days required to constitute a heat wave (Awasthi et al., 2022; Mazdiyasni et al., 2019; Cebrián et al., 2022), which poses challenges for developing robust models for comprehensive risk assessment (Mazdiyasni et al., 2019; Schliep et al., 2021; Cebrián et al., 2022). A typical limitation is the inability to meaningfully compare heat wave events that differ in temporal extent. To address these issues, we formalize heat waves from a statistical perspective to serve as the foundation for the remainder of this work. A heat-wave variable is modeled as a bivariate statistical construct, jointly characterized by intensity (a continuous feature) and duration (a discrete feature). This formulation differs from conventional approaches that evaluate intensity under a fixed-duration constraint, and instead provides a probabilistically coherent basis for the statistical analysis of heat waves.

In this work, we aim at analyzing the statistical characteristics of heat waves in terms of their intensity, duration, and frequency. A wide range of methods have been developed to study these aspects. In fixed-duration settings, intensity is often modeled using summary statistics such as the maximum or mean temperature over the heat-wave period (Thompson et al., 2023; Mazdiyasni et al., 2017; Perkins and Alexander, 2013). Russo et al. (2015) introduced the heat wave magnitude index as an improved measure of intensity by aggregating standardized daily anomalies across the duration, resulting in a cumulative severity score. Some studies incorporated additional meteorological variables, such as humidity, radiation and wind, into composite indices for heat waves (Becker et al., 2022; Awasthi et al., 2022). These indices are developed for the identification of heat waves and are less suitable for diagnosing their severity in terms of the sequence of temperature exceedances. Alternatively, some approaches jointly model intensity and duration without imposing a fixed-duration constraint. These methods often rely on multivariate distributions, such as copulas, to derive intensity–duration–frequency (IDF) relationships (Manning et al., 2019; Mazdiyasni et al., 2019). However, these models face several challenges. One key issue is the need to handle mixed data types, as intensity is typically represented as a continuous variable while duration is discrete. In addition, the statistical dependence between intensity and duration is often weak, which limits the effectiveness of copula-based methods that generally assume strong correlation between continuous variables. These

limitations make copula-based approaches more suitable for precipitation modeling than for temperature. Furthermore, there is no established framework for defining joint return levels, making interpretation difficult.

The statistical analysis of heat waves presents unique challenges due to their rarity and deviation from the bulk of observed temperatures. EVT provides a principled framework for modeling such events by characterizing the tail bahaviour of temperature distributions (Davison and Huser, 2015). Rather than relying on empirical frequencies, EVT allows for extrapolation beyond the range of historical extremes, offering tools to estimate the probability of rare temperature events. EVT has been extensively applied in climate research to study temperature extremes using reanalysis datasets and outputs from climate models (Wehner et al., 2020; Barbosa and Scotto, 2022). Applications include investigating trends in the duration and intensity of heat waves, referred to as hot spells in (Furrer et al., 2010; Photiadou et al., 2014), identifying climatologically vulnerable regions (Thompson et al., 2023), and assessing how risks evolve under different global warming limits (Kharin et al., 2018). These advances demonstrate EVT's utility in deriving robust statistical characterizations of the heat wave bahaviour under limited data conditions.

We propose a statistical approach based on EVT to probabilistically assess the risk of heat waves. This approach quantifies the severity of a heat-wave variable using a likelihood measure that incorporates both its duration and the sequence of temperature exceedances above a chosen threshold. Moreover, we derive the analytical cumulative distribution function (CDF) of the likelihood, which provides a theoretical basis for consistent comparison and evaluation of heat waves with varying durations and intensity levels. This model is further empirically validated using historical temperature data from Brussels by showing that, for each duration, the theoretical CDF of intensity closely matches its empirical distribution. In contrast to multivariate approaches, our method adopts a unified univariate formulation that avoids the associated modeling complexities. In addition, the approach facilitates the construction of IDF curves, which are used to analyse the frequency of extreme heat waves across varying durations. Using ERA5 data, we demonstrate the utility of our method by analyzing historical heat waves in six major European cities: Madrid, Rome, Paris, Brussels, Berlin, and Warsaw.

The remainder of this chapter is organized as follows. In Section 5.2, the proposed method is presented along with its theoretical foundation and practical implementation. Section 5.3 details the validation experiments and provides a comprehensive analysis of historical heat waves in six major European cities. Finally, Section 5.4 concludes this

chapter.

5.2 Model

In this section, we propose a statistical framework for analyzing heat waves. As previously described, a heat wave is characterized as a period of consecutive days with temperatures exceeding a predefined threshold. In Section 5.2.1, we will introduce a likelihood measure to capture the severity of a heat wave by jointly characterizing its duration and the magnitude of each exceedance within the period. We further derive the analytical cumulative distribution function (CDF) of this likelihood measure, providing a theoretical foundation for subsequent analysis. Next, we will demonstrate in Section 5.2.2 how this analytical distribution can be used to quantify the extremeness of a heat wave with respect to its likelihood in probabilistic terms, while accounting for the non-stationarity and temporal dependence properties commonly present in temperature time series. Finally, in Section 5.2.3, we present the algorithmic procedure for constructing IDF curves, which represents a key application of the proposed approach.

5.2.1 The likelihood framework of heat waves

In order to conduct a statistical analysis of heat waves, we consider daily summer temperatures as realizations of a univariate random variable X. These realizations are initially assumed to be independent and identically distributed (i.i.d.), an assumption that will be relaxed in the next section to account for temporal dependence and non-stationarity. Daily temperatures generally exhibit light-tailed or moderately heavy-tailed bahaviour, and are often well approximated by distributions such as the Gaussian, lognormal, or gamma families (Klein Tank and Können, 2003; Haddad, 2021). A unifying characteristic of these distributions is that they belong to the maximum domain of attraction of the Gumbel distribution, corresponding to an extreme value index $\xi = 0$ (Embrechts et al., 1997). This implies that the tail bahaviour of X can be modeled using either the BM or the POT approach, given $\xi = 0$, with normalizing constants c_n, d_n, and the threshold u_n estimated accordingly. The block size used in our study is $n = 92$, as the dataset comprises ERA5 summer temperature records from 1940 to 2024, with each summer containing 92 daily observations (July to September).

A heat wave is then represented by a sequence of k consecutive exceedances above the

threshold u_n:

$$\tilde{\boldsymbol{x}}^{\text{exc}} = \{x_1^{\text{exc}}, x_2^{\text{exc}}, \ldots, x_k^{\text{exc}}\}, \tag{5.1}$$

where k is a realization of a discrete random variable K, denoting the duration (in days) of a heat wave, and each x_i^{exc} represents the magnitude of the i-th exceedance above the threshold u_n. According to the point process limit theory of EVT (as discussed in Section 2.5), the occurrences of exceedances x_i^{exc} can be modeled as a Poisson point process on the region $A = [0,1] \times [u_n, +\infty[$, where the time axis is normalized to $[0,1]$ through scaling time by $\frac{1}{n}$. A heat wave $\tilde{\boldsymbol{x}}^{\text{exc}}$ is interpreted as a realization of a point process of exceedances $\tilde{X}_n^{\text{exc}}$, and the expected number of exceedances falling in A is characterized by the intensity measure $\lambda = \Lambda(A) = e^{-u}$.

We apply the Janossy likelihood to characterize the severity of heat waves, where the likelihood describes the joint density of K consecutive exceedances (Daley and Vere-Jones, 2003). To derive this likelihood, we begin by evaluating the probability density of each observed exceedance. Based on Theorem 2.2, we have that the exceedances over a high threshold u_n are approximately exponentially distributed for large n, with the density of the j-th exceedance given by:

$$p(x_j^{\text{exc}}) = P(X_j^{\text{exc}} = x_j^{\text{exc}}) \approx \frac{1}{d_n} e^{-\frac{x_j^{\text{exc}} - u_n}{d_n}}. \tag{5.2}$$

Next, we consider the probability of observing exactly $K = k$ consecutive exceedances. By Theorem 2.1, the probability of a single exceedance occurring on any given day is approximately:

$$\lambda^* := P(X > u_n) \approx \frac{1}{n} e^{-u} = \frac{\Lambda(A)}{n}.$$

A heat wave of length k requires both temporal continuity and duration, where the threshold u_n must be exceeded on k consecutive days, with non-exceedances on both sides to ensure temporal isolation. The probability of such a configuration is:

$$\begin{aligned} \mu_k &= P(X_t < u_n) P(X_{t+1} > u_n) \cdots P(X_{t+k} > u_n) P(X_{t+k+1} < u_n) \\ &= \lambda^{*k} (1 - \lambda^*)^2. \end{aligned}$$

Finally, the likelihood associated with the heat-wave $\tilde{\boldsymbol{x}}^{\text{exc}}$ of duration k is given by

$$\begin{aligned} \ell_n(\tilde{\boldsymbol{x}}^{\text{exc}}) &:= \mu_k k! \prod_{j=1}^{k} p(x_j^{\text{exc}}) \\ &= k! \lambda^{*k} (1-\lambda^*)^2 \prod_{j=1}^{k} \frac{1}{d_n} e^{-\frac{x_j^{\text{exc}} - u_n}{d_n}}, \end{aligned} \tag{5.3}$$

where the factor $k!$ accounts for the permutation invariance of exceedance patterns, as the point process treats them as unordered collections.

The likelihood $\ell_n(\tilde{\boldsymbol{x}}^{\text{exc}})$ serves as a comprehensive measure to reflect the severity of a heat wave across individual exceedance magnitudes as well as the duration. To interpret this severity measure in probabilistic terms, we next study the distributional bahaviour of $\ell_n(\tilde{\boldsymbol{x}}^{\text{exc}})$. Leveraging the statistical results of EVT, we derive an analytical expression for the limiting distribution of the likelihood $\ell_n(\tilde{\boldsymbol{x}}^{\text{exc}})$. This forms the theoretical foundation of our proposed model, the likelihood pattern of exceedances (LPOE), which will be used in the next section to assess heat-wave risk. The limiting distribution is formalized in the following theorem. While the formulation of the likelihood is adapted to the current modeling setting, the derivation of its asymptotic distribution remains consistent with the analytical approach introduced in Theorem 3.3 of Chapter 3.

Theorem 5.1. Consider a random variable X with distribution F_X asymptotically having a Gumbel distribution for block maxima, meaning that the following property holds:

$$\lim_{n \to +\infty} P\left(\frac{X - u_n}{d_n} < x \mid X > u_n \right) = 1 - e^{-x}, \tag{5.4}$$

with $u_n = c_n + u d_n$, for some $u \in \mathbb{R}$ and c_n, d_n normalizing constants. The likelihoods $V_n = l_n(\tilde{\boldsymbol{x}}^{\text{exc}})$ of patterns of consecutive exceedances $\tilde{\boldsymbol{x}}^{\text{exc}} = \{x_1^{\text{exc}}, x_2^{\text{exc}}, \ldots, x_k^{\text{exc}}\}, x_i^{\text{exc}} > u_n$, will converge in distribution to a random variable V with CDF:

$$\begin{aligned} G(v) &= \lim_{n \to +\infty} P\left(V_n \leq v\right) \\ &= \mu_0 \mathcal{H}\left(v - \mu_0\right) + \sum_{k \geq 1} \mu_k \Big[1 - \mathcal{E}_{k,1}\Big(-\log \frac{v \cdot (d_n)^k}{\mu_k \cdot k!} \Big) \Big], \end{aligned} \tag{5.5}$$

where $\mathcal{E}_{k,1}$ denotes the CDF of an Erlang distribution with shape parameter k and scale 1,

and

$$\mu_0 = e^{-e^{-u}},$$
$$\mu_k = \lambda^{*k}(1-\lambda^*)^2 \quad (k \geq 1),$$
$$\lambda^* = \frac{1}{n}e^{-u}.$$

Here, $\mathcal{H}$ is the Heaviside step function, i.e. $\mathcal{H}(v) = 1$ for $v \geq 0$ and $\mathcal{H}(v) = 0$ for $v < 0$.

5.2.2 Probabilistic assessment of heat waves

Heat waves with unusually high exceedance magnitudes, prolonged durations, or both, are generally regarded as events of high severity. In our framework, the severity of such events is reflected by low likelihood values as expressed in Eq. (5.3). A natural question then arises: given a particular heat wave realization, how extreme is its severity compared to all possible heat wave events. In other words, how much does its likelihood deviate from typical patterns. Theorem 5.1 addresses this question by providing an analytical form of the limiting distribution of the likelihood. This result establishes a probabilistic framework to assess the extremeness of a given heat wave. Specifically, the upper-tail probability

$$P(V \geq v) = 1 - G(v). \tag{5.6}$$

quantifies the probability of observing a heat wave with greater severity. Lower values of this probability correspond to heat waves that are more extreme and hence can be associated with increased hazards.

The theoretical framework developed thus far completes the probabilistic modeling of heat waves under the i.i.d. assumption. In practice, however, temperature data often exhibit non-stationarity and temporal dependence, rendering this assumption unrealistic. We therefore adapt our methodology to accommodate for such properties in temperature time series. A schematic overview of this adaptation is shown in Figure 5.2.

In a first step, we deal with the non-stationarity property. In temperature data, non-stationarity may arise from inner-annual variability (i.e., seasonality) and inter-annual trends (i.e., long-term warming). However, since our analysis is restricted to summer temperature records, inner-annual variability is assumed to be negligible. The inter-annual component, which captures long-term warming trends, is explicitly incorporated into the modeling framework. According to the asymptotic results of EVT (Coles, 2001), long-term trends in the underlying temperature series do not affect the limiting distribution of

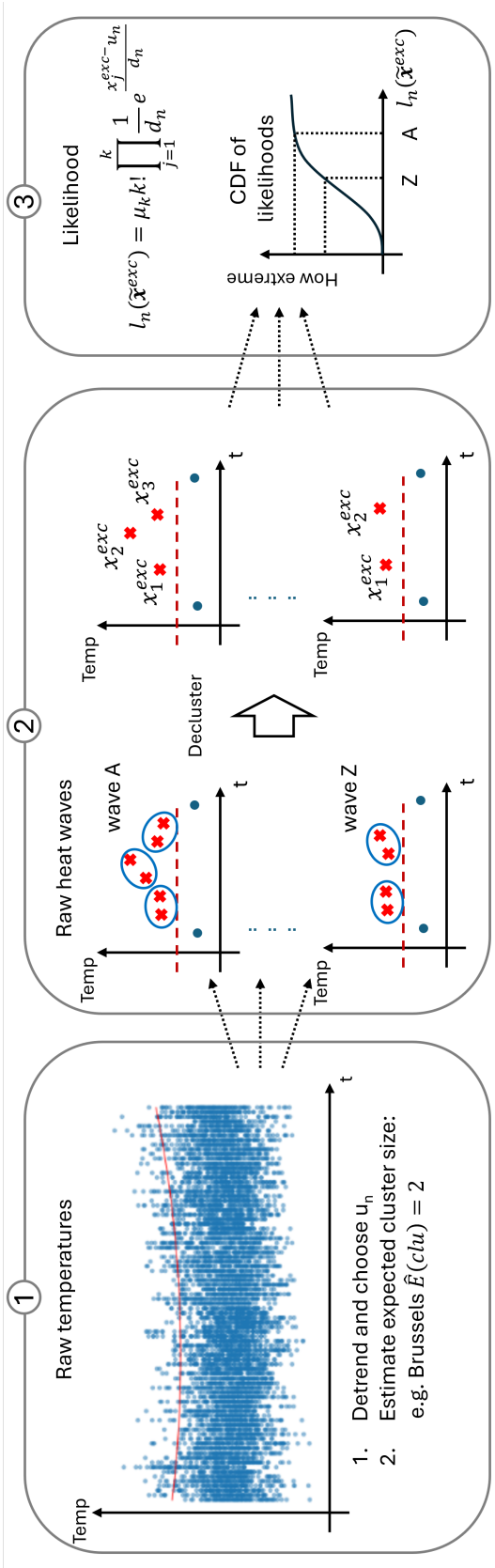


Figure 5.2: Schematic overview of the proposed methodology in three steps: 1) choosing a non-stationary threshold $u_n(t)$ for the temperature time series, 2) applying a declustering strategy to account for temporal dependence among exceedances, and 3) computing the likelihood of the observed heat wave and evaluating its extremeness using the CDF obtained in Theorem 5.1.

exceedances, but can impact the estimation of its parameters. In particular, the threshold $u_n(t)$ is most sensitive to such trends, while the shape parameter ξ and the scaling constants c_n, d_n are typically assumed to remain constant over time (Furrer et al., 2010). To incorporate inter-annual trends into threshold estimation, a temporal model is fitted to u_n in the upper tail of the summer temperature series X, where linear or quadratic trends are considered, as higher-order polynomials are generally deemed unrealistic for temperate-climate data (Twardosz et al., 2021; DelSole and Tippett, 2021). The optimal trend is determined based on standard statistical criteria such as likelihood ratio tests or AIC. The fitted trend is then used to obtain a time-varying threshold $u_n(t)$, ensuring that the asymptotic GPD approximation remains valid under non-stationarity. In Figure 5.2, the fitted threshold $u_n(t)$ is shown for Brussels temperature data series.

In a second step, we address the issue of temporal dependence inherent in temperature data. From EVT, it is known that the asymptotic results in Theorems 2.1 and 2.2 hold as long as exceedances above a high threshold are approximately independent when sufficiently separated in time. This assumption is reasonable for temperature data, where the effect of a hot day on distant future days is limited. Coles (2001) identified declustering as the most widely adopted approach for handling temporal dependence, providing a principled way to select exceedances that are approximately independent for parameter estimation. In this strategy, consecutive exceedances are grouped into clusters, each representing a sequence of exceedances with potential short-term dependence. A heat wave realization may consist of one or several such clusters. The average cluster size, which accounts for the dependence structure, is chosen based on the ratio of the total number of exceedances to the number of heat wave events. Larger cluster sizes indicate stronger short-term dependence, suggesting that a hot day is likely to influence subsequent days within that range. Therefore, within each cluster, only the maximum exceedance is retained as a representative exceedance. These maxima are treated as approximately independent and are used to estimate the normalization constants c_n and d_n in the limit distribution. An illustration of this declustering procedure is shown in Figure 5.2.

As a final step, the methodology LPOE introduced in Section 5.2.1 is applied. Based on the estimated parameters $u_n(t)$, c_n, and d_n, the likelihood $\ell_n(\tilde{\boldsymbol{x}}^{\text{exc}})$ is computed for each observed heat wave, along with the associated cumulative probability $1 - G(v)$, for comprehensive heat-wave risk assessment. Among these, the threshold $u_n(t)$ plays a critical role, as it defines exceedances and thus determines which days are classified as unusually hot. A lower threshold results in more frequent exceedances and longer durations, whereas a higher threshold leads to fewer, shorter events. Thresholds $u_n(t)$

can be selected based either on domain expertise, such as climatological definitions of hot days for each year, or on statistical grounds. In this study, we adopt statistical grounds for each city, as it provides an objective criterion that avoids regional inconsistencies arising from expert judgment. This involves removing the long-term warming trend using a fitted temporal model, followed by selecting an upper quantile using tools such as the mean residual life plot, which is widely used in extreme value analysis. It is worth noting that, from a statistical perspective, however, there is no universally optimal choice for the threshold u_n. Instead, a range of values can be considered reasonable, as our model performs consistently well across this range. To reflect this, the Appendix provides a range of reasonable threshold values u_n (after removing the trend) for each city, together with their corresponding probability-probability (P–P) plots to support model validation.

5.2.3 IDF curves for heat waves

In this section, we show how Theorem 5.1 can be used to construct IDF curves for heat waves. More precisely, given a small probability level p, the corresponding severity threshold v_p is defined by

$$P(V < v_p) = G(v_p) = p\,, \tag{5.7}$$

which implies that the probability of observing a heat wave more extreme than v_p is p, which corresponds to an expected occurrence of once every $1/p$ observations. Under a block size of n, the return period (in years) for heat waves with likelihood v_p is given by

$$t\left(\tilde{\boldsymbol{x}}^{\text{exc}}\right) = \frac{1}{G(v_p)\cdot n} = \frac{1}{p\cdot n}\,. \tag{5.8}$$

For the severity level v_p associated with a return probability p, there exist multiple possible heat wave events with varying durations k and average intensities $\bar{x}_k$, all corresponding to the same likelihood value v_p. Each couple $(k, \bar{x}_k)$ represents a specific manifestation of a heat wave with severity v_p, and together they form points on the IDF curve associated with the return period $t\left(\tilde{\boldsymbol{x}}^{\text{exc}}\right)$. The return level $\bar{x}_k$ corresponding to duration k and severity level v_p can be computed by means of Eq. (5.3) as follows:

$$v_p = k!\lambda^k(1-\lambda)^2\prod_{j=1}^{k}\frac{1}{d_n}e^{-\frac{x_j^{\text{exc}}-u_n}{d_n}} \iff \bar{x}_k = u_n + \frac{d_n}{k}\ln\left(\frac{k!\lambda^k(1-\lambda)^2}{d_n^k v_p}\right)\,. \tag{5.9}$$

This approach avoids the need to model the joint distribution of duration and intensity, which is often complicated due to their differing data types (discrete vs. continuous). The

following Algorithm 5.1 outlines the detailed procedure used to compute the IDF curves. The application results for all study cities are presented in Section 5.3.4.

Algorithm 5.1 Construct IDF curves for heat waves

Input: Summer temperature series $\mathbb{A} = \{x_1, \ldots, x_n, x_{n+1}, \ldots, x_{2n}, \ldots, x_{(m-1)n+1}, \ldots, x_{mn}\}$, where the block size is n and the total number of years is m; A set of studied durations $\mathcal{D} = \{1, 2, \ldots, D\}$; A set of return periods $\mathcal{T} = \{10, 20, 50, 100\}$.

Output: IDF curves corresponding to return periods in $\mathcal{T}$.

1: Fit a temporal model to capture the long-term warming trend in $\mathbb{A}$, and remove this trend from the original data.
2: Determine the exceedance threshold u_n on the detrended data.
3: Decluster the exceedances above threshold u_n, and collect the maxima of all resulting clusters as the set of valid exceedances.
4: Based on Eq. (2.5), estimate the normalization constants c_n and d_n using the selected valid exceedances.
5: **for** each duration $k \in \mathcal{D}$ **do**
6: **for** each return period $t \in \mathcal{T}$ **do**
7: Compute the likelihood quantile v_p using Eq. (5.7), with $\frac{1}{nt}$.
8: Compute the return level $\bar{x}_k$ for duration k based on Eq. (5.9).
9: **end for**
10: **end for**
11: Construct IDF curves for each return period $t \in \mathcal{T}$, using the set of duration–intensity couples $(k, \bar{x}_k)$.

5.3 Results

This section investigates the validation and practical implications of the proposed approach. In Section 5.3.1, we describe the dataset. In Section 5.3.2, we verify how well the theoretical distribution of intensity, derived from Theorem 5.1, aligns with empirical intensity distributions across different heat wave durations. Section 5.3.3 applies the model to assess the severity of historical heat waves in six major European cities, and to examine their vulnerability based on exposure to past extreme heat events. Finally, in Section 5.3.4, we illustrate how the constructed IDF curves can inform heat-wave risk assessment.

5.3.1 Dataset

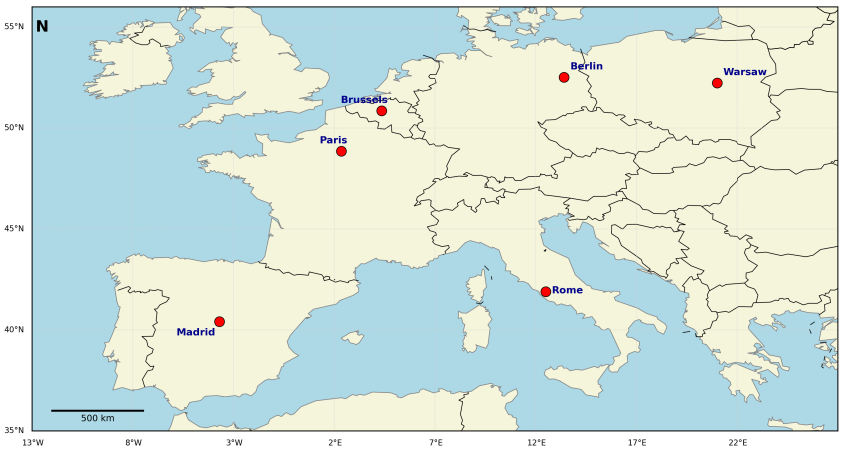


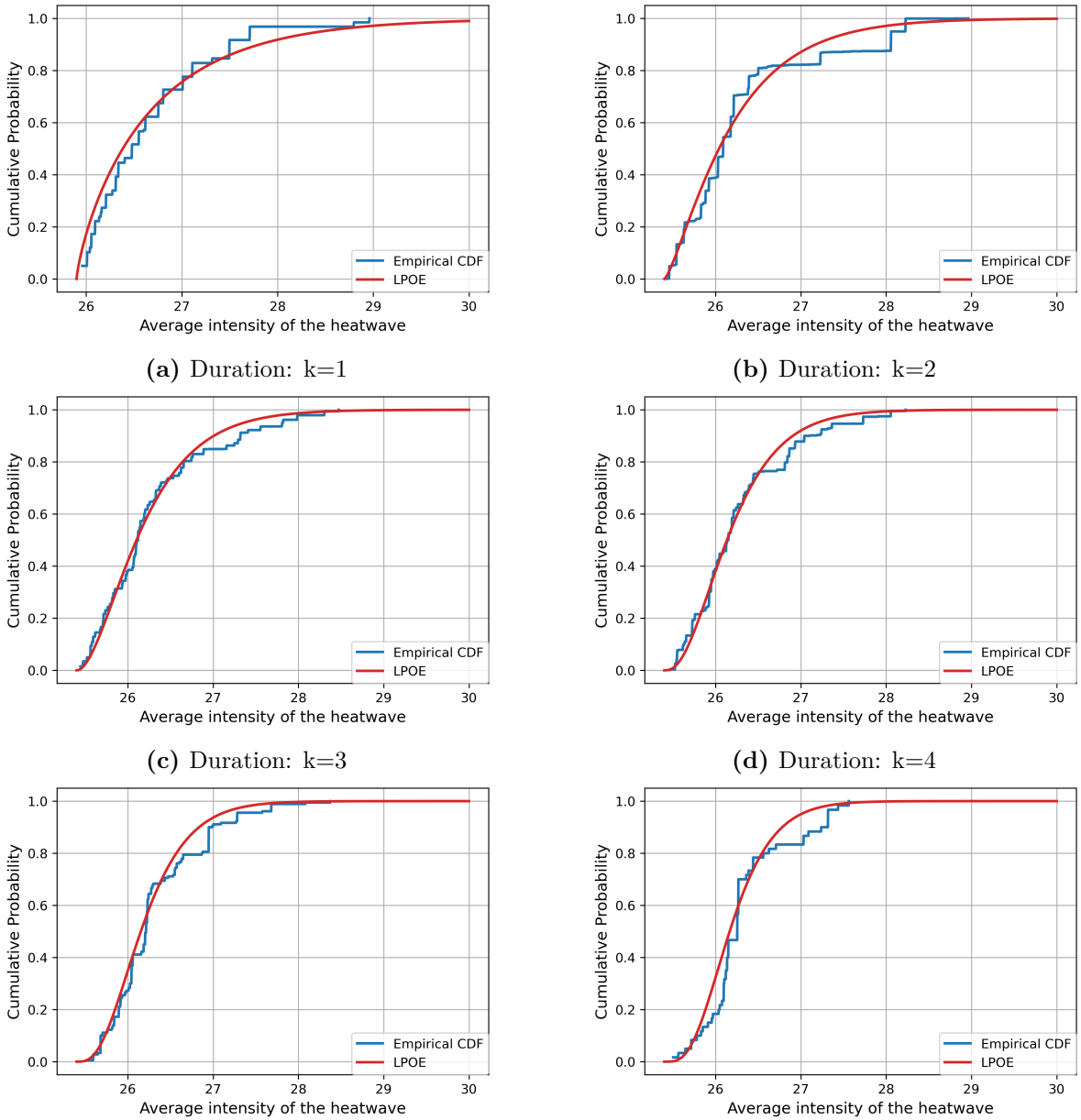
Figure 5.3: European cities considered in our study.

The ERA5 dataset is the fifth-generation atmospheric reanalysis of the global climate that is produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) and covers the period from January 1940 to the present. It provides comprehensive historical climate data through the assimilation of diverse observational sources and advanced atmospheric modeling techniques. In particular, the ERA5 near-surface temperature dataset (t2m) offers high-resolution hourly temperature data measured at two meters above the surface. Its temporal coverage and fine granularity make it especially suitable for studying extreme heat waves.

In this study, temperature data from ERA5 (t2m) spanning the period 1940—2024 is analysed to investigate heat waves in six major European cities (Madrid, Rome, Paris, Brussels, Berlin, and Warsaw), as shown in Figure 5.3. The analysis focuses on the summer season (July through September, covering 92 days). These cities were selected to represent diverse geographic and climatic settings, ensuring a broad representation of European urban environments.

5.3.2 Validation of the proposed model

In addition to the formal proof of Theorem 5.1, this section empirically evaluates its validity. Specifically, we compare the CDF of heat wave intensity derived from Theorem 5.1 with the empirical distribution obtained from historical heat wave records across varying durations.



(a) Duration: k=1

(b) Duration: k=2

(c) Duration: k=3

(d) Duration: k=4

(e) Duration: k=5

(f) Duration: k=6

Figure 5.4: The empirical distributions of heat-wave intensity across various heat-wave durations, along with their analytical CDFs according to Theorem 5.1.

Daily summer temperature data from Brussels (1940–2024) are used for validation. Durations ranging from 1 to 6 days are considered, as events exceeding 6 days are too rare to reliably estimate the empirical intensity distributions. To ensure robust empirical estimation, particularly for longer durations, additional heat wave samples are required. Since the historical series alone does not provide a sufficient number of heat wave events, we simulate 999 synthetic time series of length 85 years by applying the stationary bootstrap method to the detrended Brussels data. This resampling technique preserves key temporal and distributional characteristics of the original annual series. Further details of the resampling procedure, including its implementation and parameter settings, are provided in the Appendix. Together with the original series, this yields 1000 realizations in total. Heat wave samples are extracted within each year based on the criterion of consecutive exceedances above the threshold $u_n(t)$, with non-exceedances on both ends to ensure temporal isolation. Each detected heat wave is assigned uniquely to a duration k without overlap, i.e., a heat wave lasting three consecutive days is assigned exclusively to the $k = 3$ sample set, and not to the sets for $k = 1$ or $k = 2$. Based on these duration-specific samples, the empirical distribution of intensity is computed for each k. Figure 5.4 displays the empirical CDFs of intensity (blue lines) computed for each duration, alongside the corresponding theoretical CDFs derived from Theorem 5.1 (red lines). The horizontal axis represents average intensity, while the vertical axis shows cumulative probability from 0 to 1.

Across all durations considered, the theoretical CDFs derived from our approach align closely with the empirical distributions, suggesting that the proposed model effectively captures the distributional bahaviour of heat wave intensity. The alignment is particularly close for shorter durations (e.g., $k = 1$ to $k = 3$), where empirical variability is lower due to the higher frequency of such events. For longer durations ($k = 4$ to 6), minor deviations are observed, primarily due to the limited number of events and increased sampling variability. Nevertheless, the overall fit remains satisfactory. These results confirm the robustness of the proposed approach across a wide range of durations and demonstrate its suitability for extreme value analysis of real-world heat waves. This demonstrates that the severity-based modeling approach offers consistent performance even for rarer, longer-duration heat wave events. In particular, our approach demonstrates clear advantages in modeling long-duration heat waves, where the limited availability of empirical data typically challenges traditional statistical methods.

5.3.3 Local assessment of historical European heat waves (1940-2024)

In this section, we assess historical heat wave severity in six major European cities using the approach developed in this study. This analysis serves two purposes. First, for each city, we identify the five most severe historical heat wave events, as determined by their likelihood scores. Second, we estimate the corresponding return periods of these events, providing insight into how frequently heat waves of similar severity are expected to recur.

Severity	Date	Day Index	T2m (°C)
1	1976-07-01	1	24.81
	1976-07-02	2	24.99
	1976-07-03	3	25.95
	1976-07-04	4	26.15
	1976-07-05	5	25.71
	1976-07-06	6	25.16
2	1947-08-15	1	25.02
	1947-08-16	2	26.83
	1947-08-17	3	25.60
	1947-08-18	4	25.33
	1947-08-19	5	25.24
3	2022-07-19	1	28.98
4	1957-07-06	1	27.11
5	1947-07-27	1	25.32
	1947-07-28	2	28.79
	1947-07-29	3	27.32

(a) Top 5 heat waves Summary for Brussels.

Severity	Date	Day Index	T2m (°C)
1	2003-08-04	1	27.24
	2003-08-05	2	28.15
	2003-08-06	3	29.01
	2003-08-07	4	28.47
	2003-08-08	5	27.84
	2003-08-09	6	28.17
	2003-08-10	7	28.71
	2003-08-11	8	29.55
	2003-08-12	9	29.41
2	2019-07-23	1	28.33
	2019-07-24	2	30.53
	2019-07-25	3	32.37
3	1947-07-27	1	28.39
	1947-07-28	2	29.77
	1947-07-29	3	26.90
4	2022-07-19	1	30.01
5	2015-07-01	1	29.25

(b) Top 5 heat waves Summary for Paris.

Severity	Date	Day Index	T2m (°C)
1	1994-07-28	1	26.06
	1994-07-29	2	27.02
	1994-07-30	3	27.82
	1994-07-31	4	28.12
	1994-08-01	5	27.63
	1994-08-02	6	26.87
2	1971-07-28	1	25.53
	1971-07-29	2	26.61
	1971-07-30	3	25.89
	1971-07-31	4	26.32
	1971-08-01	5	25.83
	1971-08-02	6	26.05
	1971-08-03	7	26.85
3	1992-08-29	1	28.39
4	1992-08-09	1	27.05
	1992-08-10	2	28.70
5	1971-08-07	1	27.34

(c) Top 5 heat waves Summary for Warsaw.

Severity	Date	Day Index	T2m (°C)
1	1994-07-24	1	26.46
	1994-07-25	2	26.82
	1994-07-26	3	26.66
	1994-07-27	4	26.53
	1994-07-28	5	27.54
	1994-07-29	6	26.87
	1994-07-30	7	27.19
	1994-07-31	8	28.23
	1994-08-01	9	29.31
2	1943-08-02	1	27.83
	1943-08-03	2	29.03
3	2010-07-10	1	28.83
	2010-07-11	2	29.58
	2010-07-12	3	29.64
4	1992-08-09	1	29.64
	1992-08-10	2	27.22
5	2022-07-20	1	29.94

(d) Top 5 heat waves Summary for Berlin.

Severity	Date	Day Index	T2m (°C)
1	1995-07-18	1	31.55
	1995-07-19	2	32.10
	1995-07-20	3	32.58
	1995-07-21	4	32.20
	1995-07-22	5	31.00
2	1995-07-24	1	32.44
3	2018-08-03	1	32.44
4	1967-07-20	1	30.56
5	1987-08-12	1	31.28
	1987-08-13	2	30.63

(e) Top 5 heat waves Summary for Madrid.

Severity	Date	Day Index	T2m (°C)
1	1947-07-31	1	29.49
	1947-08-01	2	29.64
	1947-08-02	3	29.93
	1947-08-03	4	30.86
	1947-08-04	5	29.71
2	1956-08-08	1	31.64
	1956-08-09	2	29.87
	1956-08-10	3	28.70
3	1957-08-12	1	29.41
	1957-08-13	2	29.75
4	2007-08-26	1	30.43
5	1981-08-04	1	29.79
	1981-08-05	2	28.49

(f) Top 5 heat waves Summary for Rome.

Figure 5.6: Tables of the five most extreme heat waves in six European cities: (a) Brussels (Belgium), (b) Paris (France), (c) Warsaw (Poland), (d) Berlin (Germany), (e) Madrid (Spain), and (f) Rome (Italy).

Previous studies on European heat waves often rely on large-scale analyses, using high percentile thresholds applied to spatially aggregated temperature data (Becker et al., 2022). However, such approaches may obscure important regional differences. As noted by Russo et al. (2015), thermal perception varies across regions, meaning that temperatures typical in the south may be experienced as extreme in the north. In our analysis, we assess heat wave severity relative to each city's own historical temperature distribution, ensuring that the rankings reflect locally meaningful extremes. For each heat wave identified in the data, we compute its severity using the likelihood measure introduced in Theorem 5.1, which jointly accounts for both the magnitude of temperature exceedances and the duration. The top five most severe events for each city are reported in Figure 5.6. Cities like Brussels, Paris, and Warsaw tend to experience heat waves with relatively lower temperatures compared to southern cities such as Madrid and Rome, though their durations can still be considerable. For example, Paris experienced its most extreme heat wave in 2003 lasting 9 days with a maximum temperature of 29.41 °C. Likewise, Brussels and Warsaw have experienced prolonged heat waves, including a six-day heat wave in Brussels in 1976 and another six-day heat wave in Warsaw in 1994, with maximum temperatures of 26.15°C and 28.12°C, respectively. On the other hand, Madrid and Rome with their warmer Mediterranean climate experience heat waves characterized by both higher temperatures and long durations. Madrid's most severe heat wave, recorded in July 1995, spanned five days and reached a peak temperature of 32.58 °C, while Rome' s 1947 heat wave also lasted five days, peaking at 30.86 °C. Such prolonged and intense events exacerbate the risks of heat-related illnesses and place sustained pressure on agriculture and economy (Weilnhammer et al., 2021; Brás et al., 2021). The rankings reveal distinct patterns across

the six cities, shaped by their climatic and geographical differences.

Several of the top-ranked heat waves identified by our model correspond to historically recognized events reported in the literature, offering additional support for the validity of our approach. For example, the 2003 Paris heat wave, ranked as the most severe in our analysis, is widely regarded as one of the most devastating in European history, with significant societal and health consequences (García-Herrera et al., 2010). The 2019 event in Paris, also appearing in our top five, has been described as a record-breaking heat wave across Western Europe (Lhotka and Kyselý, 2022). In Berlin, the 1994 heat wave identified by our model has been consistently reported in earlier studies based on ERA5 reanalysis and meteorological station data (Russo et al., 2015; Awasthi et al., 2022; Kyselý, 2010). Likewise, the 1994 and 2003 events in Warsaw, as well as the 1992 event in Berlin, have been highlighted in Central European assessments (Twardosz and Batko, 2012; Kyselý, 2010). Our model also ranks the 2015 Paris heat wave among the city's most severe, which aligns with its characterization in the literature as one of the hottest summers recorded in Central and Eastern Europe (Ionita et al., 2017). While definitions of heat waves vary across studies—often depending on the choice of threshold and exceedance criteria (Xu et al., 2016)—our likelihood-based ranking remains consistent with major events identified through both threshold- and model-based approaches. For example, even though a copula-based framework did not list the 2003 Paris event among the five most severe (Manning et al., 2019), our model does, highlighting its sensitivity to duration-intensity joint structure and its alignment with broader empirical evidence.

While the severity rankings are generally stable, they may vary depending on the choice of the threshold u_n used to define exceedances. As discussed in Section 5.2.2, classical extreme value analysis does not prescribe a single optimal threshold but rather a range of reasonable values (the Appendix provides this range for each city). Relatively lower thresholds yield more heat wave events, often including those with longer durations. These longer-duration events typically exhibit higher severity scores and may affect which events appear among the top-ranked heat waves. Higher thresholds lead to fewer identified heat wave events with shorter durations, generally with shorter durations, but more intense individual exceedances, which may also influence the resulting likelihoods and severity rankings.

Return period offers a probabilistic estimate of how frequently a heat wave of similar severity is expected to recur. Based on Theorem 5.1, Table 5.1 presents the return periods (in years) for the five most severe historical heat waves in each of the six studied cities. The return periods show a clear decreasing trend across the ranks, with more extreme

Table 5.1: Return periods (in years) of the five most extreme heat waves observed in each city.

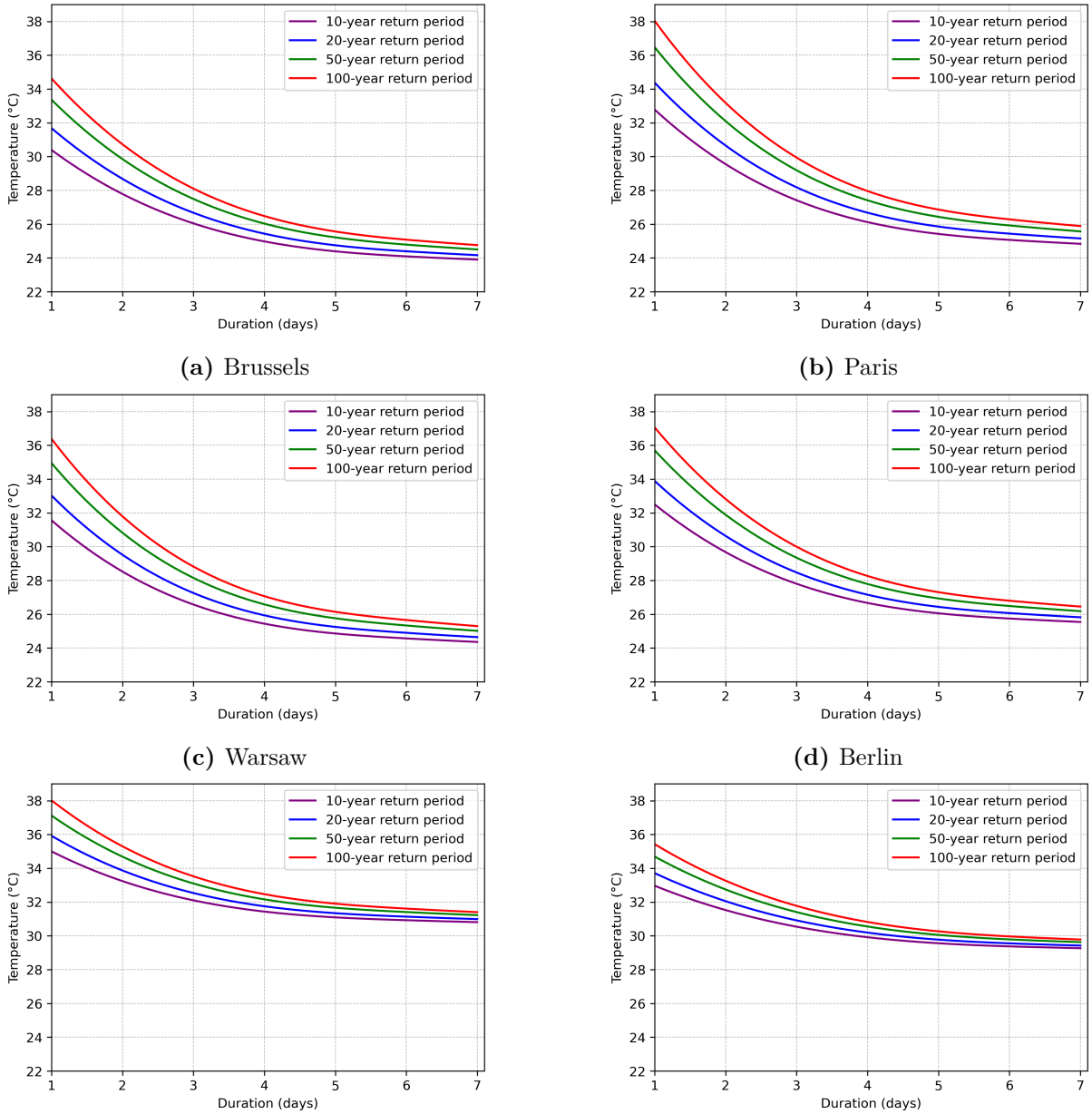
Heat wave severity rank	**Brussels (years)**	**Paris (years)**	**Madrid (years)**	**Berlin (years)**	**Warsaw (years)**	**Rome (years)**
1	171.38	1711.76	262.39	686.91	584.53	26.66
2	129.50	9.34	22.98	214.73	553.83	15.52
3	14.62	7.60	5.25	110.49	9.73	3.47
4	12.13	3.67	4.97	6.57	6.60	2.87
5	10.79	3.25	4.29	5.87	5.44	2.49

events associated with longer expected recurrence intervals. The most severe heat wave in Brussels, which lasted six days with an average temperature of 25.46 °C, is expected to recur approximately once every 171.38 years. In comparison, Paris's most extreme heat wave, lasting nine days with a peak temperature of 29.41 °C, has an estimated return period of 1711.76 years, highlighting its exceptional severity. Cities such as Rome experience extreme heat waves more frequently, with its most severe recorded event—lasting five days with an average temperature of 29.93 °C—expected to occur roughly once every 26.66 years. The substantial variability in the recurrence potential of extreme heat waves across cities underscores the need for adaptation strategies tailored to local climatic conditions and historical exposure.

Return periods also help distinguish cities that have already experienced exceptionally severe heat waves from those that have not. This distinction is crucial for identifying potential vulnerabilities and guiding adaptation efforts (Thompson et al., 2023). Cities like Paris, Berlin, and Warsaw, which have longer return periods for their most severe heat waves, have already encountered exceptionally rare events. This prior exposure indicates that compared to other cities they may be more prepared to handle similar extreme heat waves in the future, informed by lessons learned from past occurrences. In contrast, cities like Rome, with a shorter return period of 26.66 years for its most extreme historical heat wave, face greater vulnerability to unprecedented heat waves of greater severity. The lack of experience with exceptionally rare events underscores the importance of forward-looking strategies that do not rely solely on past observations.

5.3.4 Heat wave IDF curves

This section presents the IDF curves constructed for each of the six cities. These curves serve as a tool for investigating the relationship between the intensity, duration, and frequency of extreme heat waves. To encompass a broad spectrum of heat waves, durations ranging from 1 to 7 days and return periods of 10, 20, 50, and 100 years are considered. The resulting IDF curves are shown in Figure 5.7.



(a) Brussels **(b)** Paris

(c) Warsaw **(d)** Berlin

(e) Madrid **(f)** Rome

Figure 5.7: IDF curves of heat waves for six major cities across Europe (Brussels, Paris, Warsaw, Berlin, Madrid, and Rome) using the proposed approach in Theorem 5.1.

Across all cities, for a given return period, shorter-duration heat waves exhibit higher intensity, whereas longer-duration heat waves correspond to lower intensity. This pattern aligns with both intuitive expectations and historical observations. To illustrate the general bahaviour, we take Brussels as a representative example. The 10-year return period curve reveals that a one-day heat wave with an average temperature of approximately 30.3 °C is expected to occur once every 10 years, corresponding to an annual probability of 0.1. Likewise, a heat wave with a two-day duration at 27.7 °C, a three-day heat wave at 26 °C, and a four-day heat wave at 25 °C exhibit the same probability of occurrence. Besides, a one-day heat wave expected once every 100 years reaches approximately 34.5 °C, significantly higher than its 10-year counterpart. The IDF curves for Brussels reflect its moderate exposure to extreme heat waves, shaped by its oceanic climate, which generally results in milder summers and a lower probability of persistent high-temperature events compared to cities in hotter regions.

IDF curves provide a predictive tool for assessing the potential characteristics of future extreme heat waves by relating their expected intensity and duration to specific recurrence intervals. In cities such as Paris, Brussels, Berlin, and Warsaw, high-intensity heat waves with temperatures exceeding 30 °C tend to be limited in duration, often lasting only a few days. This indicates a greater tendency toward short-lived but intense heat wave occurrences. In contrast, Rome and Madrid show a different pattern. Although extremely hot single days are less frequent, these cities are more prone to heat waves that persist over longer periods with moderate to high temperatures.

5.4 Conclusion and future work

This study introduced a novel statistical framework for characterizing heat waves by integrating both intensity and duration within a probabilistic setting based on point processes. Unlike many existing methods that rely on heuristically chosen thresholds, our approach allows for statistically validated threshold selection. It also offers a tractable alternative to more complex multivariate models traditionally used to account for duration and intensity simultaneously.

We derived the asymptotic distribution of the likelihood measure used to quantify heat wave severity and validated the model empirically using historical data from Brussels. This approach enables the ranking of historical heat waves and the estimation of their return periods based on a unified likelihood metric. Furthermore, we demonstrated how

this methodology could be applied to construct IDF curves for heat waves to study the relationship between their duration, intensity and frequency. Our approach was illustrated using the ERA5 near-surface temperature data in six major European cities, providing a practical basis for extreme heat risk evaluation and city-specific adaptation planning.

Future work will focus on extending the analytical model with spatial structure to allow for assessments at more granular geographic scales. Additionally, our framework can be adapted to study other climate extremes, such as heavy precipitation, prolonged droughts, and strong winds, thereby broadening its applicability.

CHAPTER 6

The extreme value support measure machine for group anomaly detection

The content of this chapter is based on the published journal paper:
Lixuan An, Bernard De Baets, and Stijn Luca. "The Extreme Value Support Measure Machine for Group Anomaly Detection." *Mathematics*, 13(11): 1813, 2025.

Abstract

Group anomaly detection is a subfield of pattern recognition that aims at detecting anomalous groups rather than individual anomalous points. Existing approaches, however, mainly target the unusual aggregate of points in high-density regions. In this way, unusual group bahaviour with a number of points located in low-density regions is not fully detected. In this paper, we propose a systematic approach based on extreme value theory (EVT), a field of statistics adept at modeling the tails of a distribution where data are sparse, and one-class support measure machines (OCSMMs) to quantify anomalous group bahaviour comprehensively. Firstly, by applying EVT to a point process model, we construct an analytical model describing the likelihood of an aggregate within a group with respect to low-density regions, aimed at capturing anomalous group bahaviour in such regions. This model is then combined with a calibrated OCSMM, which provides probabilistic outputs to characterize anomalous group bahaviour in high-density regions, enabling improved assessment of overall anomalous group bahaviour. Extensive experiments on simulated and real-world data demonstrate our method outperforms existing group anomaly detectors across diverse scenarios, showing its effectiveness in quantifying and interpreting various anomalous group bahaviour.

6

6.1 Introduction

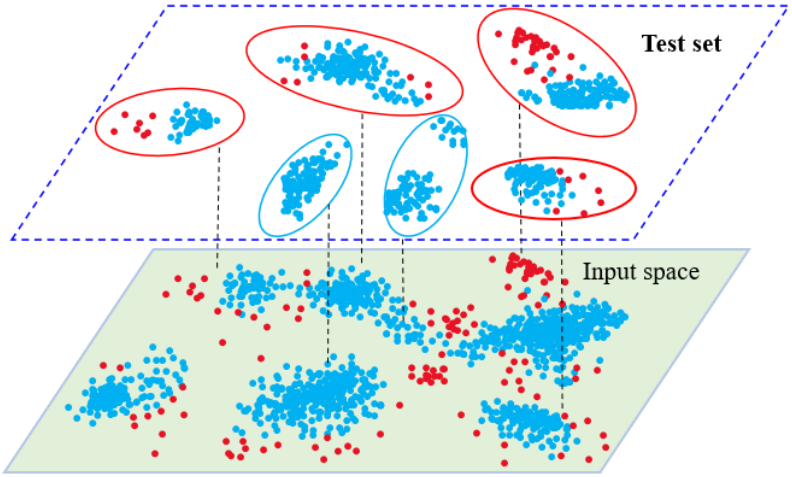
Anomaly detection is a critical task in data analysis that differs substantially from traditional classification. While classification typically involves modeling two or more classes with labeled data and comparable sample sizes, anomaly detection focuses on distinguishing between "normal" and "anomalous" samples, often characterized by an imbalanced class distribution, where anomalous samples are rare. A key assumption in anomaly detection is that the training data comprises only normal samples (Lee et al., 2022). Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ be a multivariate random variable representing d feature components with distribution function $F_{\boldsymbol{X}}$. Traditional approaches to anomaly detection focus on a point model $M(\boldsymbol{x})$ in the feature space of $\boldsymbol{X}$ that captures distribution characteristics of normal point samples (which typically reside in high-density regions) to identify anomalies deviating from this model (the so-called pointwise anomaly detection).

Group anomaly detection extends pointwise anomaly detection to scenarios where a sample is not an individual point $\boldsymbol{x}$, but a group of points $\tilde{\boldsymbol{x}} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, representing

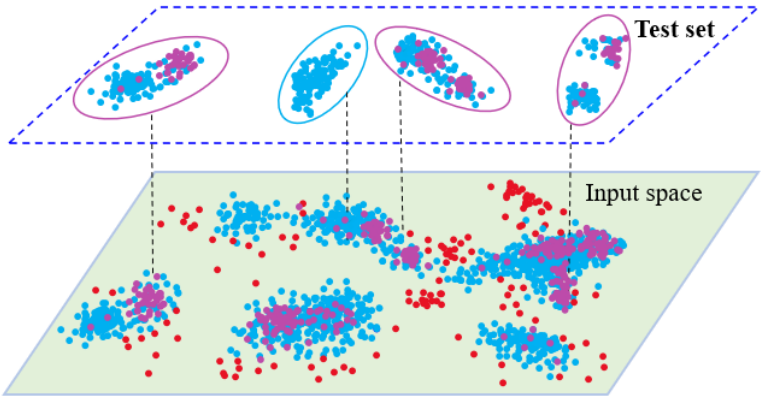
a sample of the group variable $\tilde{\boldsymbol{X}} = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$, where $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ share a common distribution function $F_{\boldsymbol{X}}$. Group anomalies can arise in diverse contexts, such as irregular sequences of measured heartbeats (Luca et al., 2018). When applied to detecting wind turbine failures, group anomalies are observed as shifts of signal sequences (Jin et al., 2022). Group anomalies also appear in the usage patterns of urban bicycle rental systems (Lohrer et al., 2022). Although pointwise anomaly detectors could theoretically be applied in group scenarios if a representative feature set for the group variable were extractable (Chandola et al., 2009), these methods face significant challenges in practical applications: (i) constructing a representative feature set that adequately encapsulates diverse group-level characteristics is often difficult, especially when the number of training groups is limited and the variable $\boldsymbol{X}$ comprising the group is high-dimensional; (ii) achieving interpretability of the group-level feature set in terms of the spatial configuration of points in the original feature space of $\boldsymbol{X}$ remains nontrivial. As a result, specialized methods for group anomaly detection have been developed, which bypass group feature extraction and allow for direct analysis within the original feature space of $\boldsymbol{X}$.

This paper presents a statistical approach for group anomaly detection, aiming to construct a framework to analytically quantify various types of anomalous group bahaviour. Given normal groups in the training set, the points across all normal groups collectively define a common, potentially complex distribution with density $y = p(\boldsymbol{x})$. Thresholding the density $p(\boldsymbol{x})$ partitions the feature space of $\boldsymbol{X}$ into high-density and low-density regions, thereby enabling all possible anomalous group bahaviour to be systematically classified into two types: (i) *point-based anomalous group bahaviour*, where a group exhibits unusual collective bahaviour among its points labeled as belonging to low-density regions; (ii) *distribution-based anomalous group bahaviour*, referring to a group whose distribution pattern of points labeled as belonging to high-density regions differs from that in any normal group in the training set. Such categorization is in line with existing work (Xiong et al., 2011a; Muandet and Schölkopf, 2013; Chalapathy et al., 2018; Song et al., 2020; Kasieczka et al., 2021) but is statistically underpinned through the density function $y = p(\boldsymbol{x})$. Figure 6.1 provides a visual illustration of the two types of group anomalies. The blue points in the two-dimensional input space represent points from all normal groups and collectively characterize the underlying point distribution, where densely clustered areas correspond to high-density regions and sparse areas indicate low-density regions. Figure 6.1(a) shows groups with point-based anomalous group bahaviour, where each group with red borders in the test set contains an unusual aggregate of points located in low-density regions. Figure 6.1(b) illustrates groups with distribution-based anomalous group bahaviour. Although the purple points are situated in high-density

regions, the groups with purple borders exhibit distribution patterns that substantially differ from any of the patterns observed among the normal training groups. Groups exhibiting both bahaviour are considered anomalous on both fronts.



(a) Point-based group anomalies



(b) Distribution-based group anomalies

Figure 6.1: Illustration of two types of anomalous group bahaviour. The red points represent low-density region points, and the purple points are high-density region points. The boundaries of individual normal groups are not explicitly shown in the figure.

Existing models for group anomaly detection face two critical limitations, which motivate the development of our novel statistical approach. First, while a wide range of popular models excel at detecting distribution-based anomalous group bahaviour, they exhibit a significant weakness in identifying point-based anomalous group bahaviour. For example, models like one-class support measure machine (OCSMM) (Muandet and Schölkopf, 2013), support measure data description (SMDD) (Guevara et al., 2015) and the neural network model by Chalapathy et al. (2018) primarily assess how the distribution pattern

of points (labeled as belonging to high-density regions) in a test group deviates from those of normal training groups, while largely neglecting the crucial aggregation bahaviour of points in low-density regions within this test group. In domains like health monitoring, both types of group anomalies are crucial, and the failure to effectively detect point-based anomalous group bahaviour can lead to spurious alarms. Although models like mixture of Gaussian mixture model (MGMM) (Xiong et al., 2011a) and flexible genre model (FGM) (Xiong et al., 2011b) attempt to address both types of group anomalies, they often exhibit higher sensitivity towards distribution-based anomalous group bahaviour. This sensitivity imbalance will be further illustrated in Sections 6.2 and 6.4.

The second limitation lies in modeling point-based anomalous group bahaviour, where the strategy of most models that combine (pointwise) anomaly scores is flawed. The main reason is that low-density region points need to be considered as a whole to capture their aggregate bahaviour, otherwise the multiple hypothesis testing problem will manifest. The point-based group anomaly detection problem can be stated as follows:

$$H_0\text{: } \tilde{\boldsymbol{x}} \textit{ is a group of points drawn from the distribution of } \boldsymbol{X}\,,$$
$$H_1\text{: } \tilde{\boldsymbol{x}} \textit{ is an anomalous group with respect to the distribution of } \boldsymbol{X}\,.$$

Clearly, it is a multiple hypothesis testing problem where more than one hypothesis is being tested, more specifically one hypothesis for each individual point being anomalous or not. For the group size $n = 1$, this simplifies to classic (pointwise) anomaly detection where the likelihood $p(\boldsymbol{x}_1)$ can be directly used as an anomaly score (Tarassenko et al., 1995). However, for $n > 1$, the risk of wrongly classifying a group $\tilde{\boldsymbol{x}}$ as anomalous considerably increases. Indeed, while the type-I error for $\boldsymbol{x}_i \in \tilde{\boldsymbol{x}}$ can be controlled at a significance level α, the error of making at least one type-I error among all n hypotheses corresponds to $P(\text{type-I error}) = 1 - (1 - \alpha)^n > \alpha$. As a result, the decisions made by these models become unreliable.

This work aims to develop an effective method to address the aforementioned limitations and accurately evaluate the anomalousness of a group. We integrate a likelihood model based on EVT with a calibrated OCSMM within a unified framework. Each component is designed to specifically target a distinct form of anomalous group bahaviour, with the likelihood model being employed to characterize the low-density regions and avoid the multiple hypothesis testing problem, whereas the calibrated OCSMM is intended to capture deviations in the distributional structure of high-density regions across groups. Leveraging the complementary information captured by each component, these two com-

ponents are subsequently combined through a constructed uninorm to generate an overall anomaly score. Specifically, the contributions of this paper are:

1. We offer a novel framework where EVT is applied to a point process model (PPM) in order to fully capture the spatial configuration that is hidden in the data points situated in low-density regions.

2. An analytical result is proven to approximate the distribution of the PPM. Based on this result, a point-based anomaly score is defined to evaluate the anomaly extent of a group with respect to low-density regions, effectively addressing the multiple hypothesis testing problem.

3. We extend the existing OCSMM framework to yield probabilistic outputs, thereby enabling the construction of a distribution-based anomaly score that exclusively assesses the anomaly extent of a group with respect to high-density regions. OCSMM is chosen for this purpose because it effectively captures the distributional characteristics of normal groups in high-density regions, without redundantly incorporating information from low-density regions, thus avoiding overlap with the previously constructed point-based anomaly score.

4. A uninorm is constructed to effectively combine the above two scores, each focusing on the anomaly extent of one of the group's two complementary aspects. The resulting overall anomaly score provides a comprehensive measure of the group's anomalousness.

The remainder of this Chapter is structured as follows. In Section 6.2, the existing approaches for group anomaly detection are reviewed and analysed. Subsequently, Section 6.3 introduces our novel method for group anomaly detection. In Section 6.4, the method is evaluated on both synthetic and real-world datasets. Section 6.5 concludes this chapter.

6.2 Related work

As our study is concerned with group anomaly detection, we focus our literature review on this area rather than (pointwise) anomaly detection methods. Following classic pointwise anomaly detection, early developments in group anomaly detection methods combined the individual predicted scores generated by pointwise anomaly detectors into a single one for a

statistical decision (Dietterich, 2002). For instance, in the image analysis research of Hazel (2000), a pointwise detector was used to identify anomalous pixels that were subsequently gathered into anomalous groups. Das et al. (2008) classified groups according to a high fraction of individual anomalous points within a group. Belhadi et al. (2022) combined different data mining techniques to identify groups of sequence outliers. Mohod and Janeja (2022) adapted density-based clustering methods to discover spatial anomalous groups (windows) of arbitrary shape. Das et al. (2009) also introduced a group anomaly detection framework based on scan statistics to detect clusters with increased counts of spatial data. More recently, Pehlivanian and Neill (2023) extended scan statistics to detect contiguous group anomalies in ordered data. However, scan statistics are based on the number of instances within a region but discard their relative location in space. Moreover, these methods are ineffective in addressing the multiple hypothesis testing problem.

With the advancement of statistical techniques, there is a growing interest in detecting unusual group bahaviour, especially in high-density regions. Some representative models have been proposed. Muandet and Schölkopf (2013) introduced OCSMM, which extends the one-class support vector machine (OCSVM) (Schölkopf et al., 2001) to groups. By adapting the kernel in OCSVM for mapping points into feature space to the kernel mean embedding mapping groups into feature space, OCSMM extracts distribution information from normal groups and separates them from the origin. SMDD (Guevara et al., 2015) also utilizes the kernel mean embedding technique, but it is based on the idea of constructing a minimum enclosing ball (Tax and Duin, 2004). Ting et al. (2023) proposed the isolation distribution kernel to quantify distribution similarities within groups. Fisch et al. (2022) introduced a statistical approach for detecting both collective and point anomalies with a linear computational cost. Other approaches, such as those from Yu et al. (2015), Song et al. (2020), Chalapathy et al. (2018) are tailored for specific application areas. However, these models consistently fail to handle scenarios where both point-based and distribution-based anomalous group bahaviour occur within the same group. Some efforts have been made to detect both types of anomalous group bahaviour, such as MGMM (Xiong et al., 2011a) and FGM (Xiong et al., 2011b), which are strongly related to the topic model latent Dirichlet allocation (LDA) (Blei et al., 2003). Topic models are often used in natural language processing to classify a collection of documents, each of which is typically concerned with multiple topics and where each topic is represented by a cluster of multiple words. MGMM considers each group as a mixture of Gaussian-distributed topics, while FGM defines flexible structures called “genres” to model topic distributions. Scores are then defined to measure point-based and distribution-based anomalous group bahaviour and combined using a weight function. However, these models tend to focus primarily on

the distribution of points in high-density regions, as the topic generation process is less affected by points in low-density regions, particularly when these points are sparse. As a result, they exhibit significant bias in detecting point-based anomalous group bahaviour.

In recent years, to develop a suitable model for point-based anomalous group bahaviour, researchers have explored considering low-density region points as an integral component. Luca et al. (2014, 2016, 2018) investigated the use of EVT to extract features characterizing the aggregate of points in low-density regions (also termed extreme risk regions) of $p(\boldsymbol{x})$. This involved summarizing aggregation information from extreme points in terms of the mean, maximum excess of $p(\boldsymbol{x})$, or the number of extremes with respect to some threshold. While these models address the multivariate hypothesis testing problem by employing such features to characterize aggregated extreme bahaviour, questions remain about how to identify a comprehensive feature set that effectively captures this bahaviour.

6.3 A statistical framework for modeling group bahaviour

This section introduces the extreme value support measure machine (EVSMM), a novel statistical method for comprehensively quantifying how anomalous a group is compared to training normal groups. The foundation of EVSMM involves characterizing the group-level normal bahaviour of training normal groups within the d-dimensional point feature space $\mathbb{R}^d$. We quantitatively analyse group bahaviour by modeling the statistical characteristics of two complementary parts within the group: the low-density region part and the high-density region part. The EVSMM operates through a two-step process. First, two scores are constructed to quantify how unusual the parts of a group situated respectively in the low-density and high-density regions are. Second, an effective way to combine these two scores is proposed by using a constructed uninorm aggregation function, whose roots trace back to the fusion of certainty factors in highly successful early days expert systems. The following sections will detail the rationale and necessity behind the construction steps of EVSMM.

6.3.1 Two key scores

We start with the construction of two scores to quantify the anomaly extent of a group's low-density and high-density region parts, respectively.

The first score is based on a PPM framework that models the spatial configuration of exceedances with respect to a decision boundary in its full generality. We then derive an asymptotic distribution, valid as $n \to \infty$, which is used as an approximation to model the aggregate of exceedances.

Let $\boldsymbol{X}$ be a d-dimensional random variable, with no distributional assumptions imposed on $\boldsymbol{X}$. Consider a training set of normal groups $\mathbb{A} = \{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_l\}$, where each group $\tilde{\boldsymbol{x}}_\ell = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{n_\ell}\} \subset \mathbb{R}^d$ have varying size n_ℓ, with $\mathrm{E}(n_\ell) = n$. The objective is to comprehensively quantify the unusual bahaviour of a group with respect to low-density regions, which entails characterizing the aggregate bahaviour of exceedances, determined by a certain number of exceedances and their various locations within low-density regions. The aggregate of exceedances within the group $\tilde{\boldsymbol{x}}$ is given by:

$$\{\boldsymbol{x} \in \tilde{\boldsymbol{x}} \mid p(\boldsymbol{x}) \leq e^{-u_n}\},$$

for large values of u_n. The low-density regions are defined by the decision boundary $p(\boldsymbol{x}) \leq e^{-u_n}$, where u_n is the threshold in the POT model introduced in Theorem 2.2.

As previously described, this study begins by analyzing the distribution of points across all training groups, which describes the probability space of d-dimensional variable $\boldsymbol{X}$. As our framework is not restricted to a particular density estimator, any suitable method could be adopted to estimate the probability density function $p(\boldsymbol{x})$. To illustrate this flexibility, We mention several representative alternatives, such as parametric models like Gaussian mixture models (GMMs), non-parametric approaches like projection pursuit density estimation (PPDE), and random projection-based techniques for high-dimensional data. In our experiments, we adopt kernel density estimation (KDE), a widely used non-parametric method that does not require prior assumptions about the data distribution. We explicitly clarify that KDE is used solely as an example and is not necessarily the optimal choice for all scenarios. Next, the transformation $Y = p(\boldsymbol{X})$ is applied, mapping the multivariate random variable $\boldsymbol{X}$ to the univariate random variable Y. Consequently, within each group $\tilde{\boldsymbol{x}}$, the samples $\boldsymbol{x}_i$ are transformed into samples y_i within $\tilde{\boldsymbol{y}}$ as follows:

$$\tilde{\boldsymbol{y}} = \{y_1, y_2, \ldots, y_n\} = \{p(\boldsymbol{x}_1), p(\boldsymbol{x}_2), \ldots, p(\boldsymbol{x}_n)\} \ .$$

Therefore the exceedances of $\tilde{\boldsymbol{x}}$, whether they are extremely large (the right tail) or small (the left tail) values in case of a univariate X, or occupy diverse locations within low-density regions in case of a multivariate $\boldsymbol{X}$, correspond solely to the left tail of Y. As Y is the probability density, it is bounded below by zero, which guarantees the existence of

an extreme value distribution in its left tail. Moreover, Clifton et al. (2011) proved the results that the distribution of minima $y^{\text{ext}} := \min\{y_1, y_2, \ldots, y_n\}$ can be approximated by a Weibull distribution. The Weibull distribution, with a lower bound at zero, describes the bahaviour of small values of densities near zero (Coles, 2001). To correct the skewness in the distribution of Y near zero, a logarithmic transformation $Z = -\log(Y)$ is applied:

$$\tilde{\boldsymbol{z}} = \{z_1, z_2, \ldots, z_n\} = \{-\log(y_1), -\log(y_2), \ldots, -\log(y_n)\}$$

mapping the short tail of the Weibull distribution near zero to the right tail of a Gumbel distribution for maxima:

$$P\left(\max\{Z_1, \ldots, Z_n\} \leq z\right) \approx G(z)\,, \qquad \text{as } n \to \infty\,,$$

where $G(z)$ denotes a cumulative distribution within the Gumbel family. In this way, the multivariate complex low-density regions $\{\boldsymbol{x} \mid p(\boldsymbol{x}) \leq e^{-u_n}\}$ are transformed into regions $\{z \mid z \geq u_n\}$ on the real line. The significance of transforming $\boldsymbol{X}$ to Z not only lies in mapping low-density regions from a high-dimensional space to the right tail in a one-dimensional space, but also in making it possible to exploit the well-established theoretical results for the Gumbel case ($\xi = 0$) in EVT. Within a group $\tilde{\boldsymbol{z}} = \{z_1, z_2, \ldots, z_n\}$ of n observations of Z, we study those points that exceed the threshold u_n and consider them as an aggregate:

$$\tilde{\boldsymbol{z}}^{\text{exc}} = \left\{z_1^{\text{exc}}, \ldots, z_{K_n}^{\text{exc}}\right\}\,,$$

where K_n denotes the number of exceedances in $\tilde{\boldsymbol{z}}$ with respect to u_n. In what follows, we derive an analytical result to determine the asymptotic distribution of such aggregate of exceedances as $n \to \infty$.

To tackle the above problem in its full generality, $\tilde{\boldsymbol{z}}^{\text{exc}}$ is viewed as a realized non-empty point pattern of a PPM. First, the probabilistic model of individual exceedances is derived using the limiting property Eq. (2.10) in Theorem 2.2 for $\xi = 0$ (*i.e.*, the Gumbel case). As a result, for large n, the likelihood of an exceedance z_i^{exc} can be approximated by:

$$p_n\left(z_i^{\text{exc}}\right) \approx \frac{1}{d_n} e^{-\frac{z_i^{\text{exc}} - u_n}{d_n}}\,, \tag{6.1}$$

where $u_n = c_n + u d_n$ and c_n and d_n are defined in Eq. (2.5). Second, the aggregating bahaviour of these exceedances is analysed. The aggregate of exceedances with respect to low-density regions is fully characterized by the counting measure introduced in Section 2.5. It follows that:

$$P(K_n = k) \approx \frac{\lambda^k}{k!} e^{-\lambda},$$

where $\lambda = e^{-u}$, and $u \geq 0$. These approximations motivate the definition of a probability measure, referred to as the PPM of exceedances, with the following likelihood:

$$f_n\left(\tilde{\boldsymbol{z}}^{\text{exc}}\right) := k! \mu_k^n \prod_{i=1}^{k} \frac{1}{d_n} e^{-\frac{z_i^{\text{exc}} - u_n}{d_n}}, \tag{6.2}$$

where k denotes the observed size of $\tilde{\boldsymbol{z}}^{\text{exc}}$ and $\mu_k^n = P(K_n = k)$. In the theorem below, we will present that the distribution of the likelihoods $f_n\left(\tilde{\boldsymbol{z}}^{\text{exc}}\right)$ is asymptotically given by a random variable V^{exc} of mixed type (the derivation of its asymptotic distribution remains consistent with the analytical approach introduced in Theorem 3.3 of Chapter 3). Random variables of mixed type are neither discrete nor continuous, but are a mixture of both (Soong, 2004). The discrete component results from the likelihoods $\mu_0^n > 0$ that describe cases when there are no exceedances in $\tilde{\boldsymbol{z}}$. This corresponds to a discontinuity in the CDF that can be described using the Heaviside step function:

$$\mathcal{H}(v) = \begin{cases} 1 & \text{, if } v \geq 0 \\ 0 & \text{, if } v < 0\,. \end{cases} \tag{6.3}$$

Theorem 6.1. Consider a random variable Z that satisfies the following limiting property:

$$\lim_{n \to +\infty} P\left(\frac{Z - u_n}{d_n} < z \mid Z > u_n\right) = 1 - e^{-z}, \tag{6.4}$$

where $u_n = c_n + u d_n$, with $u \geq 0$, is determined by a sequence of thresholds with:

$$\begin{aligned} c_n &= \inf\left\{z \mid P(Z \leq z) \geq 1 - \frac{1}{n}\right\}, \\ d_n &= E(Z - c_n \mid Z > c_n)\,. \end{aligned} \tag{6.5}$$

The random variables V_n^{exc} of likelihoods $f_n\left(\tilde{\boldsymbol{z}}^{\text{exc}}\right)$ as defined in (6.2) will converge in distribution to a random variable V^{exc} with CDF:

$$G(v^{\text{exc}}) = \mu_0 \mathcal{H}(v^{\text{exc}} - \mu_0) + \sum_{k \geq 1} \mu_k \left[1 - \mathcal{E}_{k,1}\left(-\log \frac{v^{\text{exc}} \left(d_n\right)^k}{k! \mu_k}\right)\right], \tag{6.6}$$

where $\mathcal{E}_{k,1}$ denotes the CDF of an Erlang distribution with shape parameter k and scale 1,

and $\mu_k = \frac{\lambda^k}{k!}e^{-\lambda}$. In particular, it holds that

$$\lim_{n\to+\infty} P\left(V_n^{\text{exc}} \leq v^{\text{exc}}\right) = G(v^{\text{exc}})\,. \tag{6.7}$$

As a brief summary, we propose the likelihood measure v^{exc} to evaluate group samples. The likelihood measure of a group is constructed based on the pattern of its exceedances—taking into account both the individual locations of low-density region points and their aggregate bahaviour. The derived analytical form $G(v^{\text{exc}})$ describes the distribution of this likelihood measure and plays a central role in quantifying the group's probabilistic anomaly extent. Notably, similar to the central limit theorem, a key advantage of EVT—the foundation of our framework—is that while the asymptotic theory requires large n, in practice, the method often performs well even for moderate values of n.

Based on the above analytical result, the first score is constructed using the expression of the CDF $G(v^{\text{exc}})$ in Eq. (6.6), which assesses the anomaly extent of a group $\tilde{\boldsymbol{x}} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ with respect to low-density regions. Addressing the multiple hypothesis problem is critical, because as the number of points forming a group increases, the exceedances are expected to become more extreme, and their number is expected to grow. The proposed likelihood concept encompasses both the probability of each individual exceedance location and the probability of the size of the aggregate of exceedances within a group. Through the likelihood concept, all exceedances within a group are treated as a unified whole. Accordingly, the first score, termed the point-based group anomaly score, is defined as:

$$\alpha(\tilde{\boldsymbol{x}}) = 1 - G(v^{\text{exc}})\,, \tag{6.8}$$

where v^{exc} denotes the likelihood, as defined in Eq. (6.2), of the aggregate of exceedances $z_i = -\log p(\boldsymbol{x}_i)$ with respect to the threshold u_n. A high anomaly score $\alpha(\tilde{\boldsymbol{x}})$ indicates that there is a small probability of observing a group with a lower likelihood of aggregate of exceedances than v^{exc} in any other group. We summarize the practical implementation of the model in Algorithm 6.1.

The second score, a distribution-based group anomaly score, is constructed by calibrating the OCSMM to quantify the anomaly extent of the group with respect to high-density regions. Basically, the OCSMM extends OCSVM by classifying groups $\tilde{\boldsymbol{x}}$ instead of individual points $\boldsymbol{x}$ (Schölkopf and Smola, 2002). The core concept behind OCSMM involves applying the kernel trick to map the mean embeddings of training groups into a feature space, where they are separated from the origin by a hyperplane. For more details about OCSMM, we refer to Section 2.6.

Algorithm 6.1 PPM of exceedances: point-based group anomaly score

Input: A set of normal groups $\mathbb{A} = \{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_l\}$; a test group set $\mathbb{T} = \{\tilde{\boldsymbol{x}}_1^{(t)}, \ldots, \tilde{\boldsymbol{x}}_{l'}^{(t)}\}$; the expected group size n;

Output: The point-based anomaly score $\alpha(\mathbb{T})$ of the test group set $\mathbb{T}$.

1: Select an appropriate density estimator to estimate $p(\boldsymbol{x})$ based on all point samples $\boldsymbol{x}$ from the normal groups in $\mathbb{A}$. Dimensionality reduction may optionally be applied prior to estimation, depending on the chosen estimator.
2: Apply the transformations $Y = p(\boldsymbol{X})$ and $Z = -\log(Y)$ to all point samples $\boldsymbol{x}$ (collected from the normal groups in $\mathbb{A}$), yielding samples z of the variable Z.
3: Use the mean residual life plot based on the sample z to select a threshold u_n for the variable Z.
4: Estimate the location parameter c_n and scale parameter d_n based on samples z using Eq. (2.5), and compute the normalized threshold $u = \frac{u_n - c_n}{d_n}$.
5: **for** each $\tilde{\boldsymbol{x}}_j^{(t)}$ in the test group set $\mathbb{T}$ **do**
6: Transform each point $\boldsymbol{x}_i \in \tilde{\boldsymbol{x}}_j^{(t)}$ to $z_i = -\log p(\boldsymbol{x}_i)$, and obtain $\tilde{\boldsymbol{z}}_j^{(t)}$.
7: Extract the exceedances $\tilde{\boldsymbol{z}}^{\text{exc}} = \{z_i \in \tilde{\boldsymbol{z}}_j^{(t)} : z_i > u_n\}$ and compute the likelihood $v^{\text{exc}} = f_n(\tilde{\boldsymbol{z}}^{\text{exc}})$ using Eq. (6.2).
8: According to Theorem 6.1, compute the point-based anomaly score as $\alpha(\tilde{\boldsymbol{x}}_j^{(t)}) = 1 - G(v^{\text{exc}})$.
9: **end for**
10: Obtain the point-based group anomaly scores $\alpha(\mathbb{T})$ for the test set as $\alpha(\mathbb{T}) = \{\alpha(\tilde{\boldsymbol{x}}_1^{(t)}), \ldots, \alpha(\tilde{\boldsymbol{x}}_{l'}^{(t)})\}$.

There are several key reasons for selecting OCSMM. First, it is specifically designed to capture the distribution characteristics within each normal group and demonstrates robustness to outliers in low-density regions, providing a stable and reliable representation of distribution information for the training group set with respect to high-density regions. This capability complements the model in the previous section, which focuses on low-density regions. Second, OCSMM leverages the kernel trick, enabling efficient modeling of the decision boundaries for complex group structures. By using the kernel mean embeddings, OCSMM computes inner products in high-dimensional spaces without requiring explicit group feature transformations. Third, OCSMM is firmly grounded in statistical learning theory, kernel mean embedding theory, and optimization theory. Its formulation as a convex optimization problem, with clearly defined constraints, ensures that a unique global optimum can be reliably determined.

Although OCSMM provides a flexible and mathematically robust framework, its output is inherently binary, producing classifications without any probability estimates. While the decision function generates continuous scores, classification is determined solely by the sign of the output. These scores indicate relative distances from the decision boundary,

but they do not provide a direct probabilistic interpretation.

As far as we are aware, there has been no prior research that specifically focuses on the probabilistic calibration of OCSMM. A critical challenge in this process is generating anomalous group samples, a necessary step for calibration due to the inherent scarcity of anomalous samples in both classical anomaly detection and group anomaly detection tasks. The following section details our calibration procedure, including the choice of the calibration method and the steps for generating anomalous groups.

We argue that, among various calibration methods, sigmoid fitting is particularly suitable for calibrating OCSMM. Unlike binning methods and isotonic regression, which often introduce discontinuities and risk overfitting, sigmoid fitting is specifically designed for binary classification problems and performs robustly on imbalanced data. In addition, sigmoid fitting offers flexibility and computational efficiency, which makes it advantageous for handling large datasets. Furthermore, sigmoid fitting assumes a monotonic relationship between decision scores and class probabilities, an assumption that aligns well with the bahaviour of OCSMM's decision scores. Based on these considerations, sigmoid fitting was chosen as the calibration method.

Denote the decision function of the OCSMM model trained on the training group set as $\gamma(\tilde{\boldsymbol{x}})$. Let $\beta(\tilde{\boldsymbol{x}})$ represent the calibrated score of a group $\tilde{\boldsymbol{x}}$ based on the output of $\gamma(\tilde{\boldsymbol{x}})$. Then, the sigmoid fitting performs the following fit:

$$P\left(\beta(\tilde{\boldsymbol{x}}) = 1 \mid \tilde{\boldsymbol{x}}\right) \approx J_{g,q}\left(\gamma(\tilde{\boldsymbol{x}})\right) \equiv \frac{1}{1 + \exp\left(g\gamma(\tilde{\boldsymbol{x}}) + q\right)}, \tag{6.9}$$

where g and q are parameters estimated by solving the following regularized unconstrained optimization problem:

$$\max_{g,q} \quad \sum_{j=1}^{l^*} \zeta_j \log\left(J_j\right) + \left(1 - \zeta_j\right) \log\left(1 - J_j\right), \tag{6.10}$$

where $J_j = J_{g,q}\left(\gamma(\tilde{\boldsymbol{x}}_j)\right)$ and

$$\zeta_j = \begin{cases} \frac{N_+ + 1}{N_+ + 2} & \text{, if } \beta_j = 1 \\ \frac{1}{N_- + 2} & \text{, if } \beta_j = -1 \end{cases},$$

for $j = 1, \ldots, l^*$. The number of samples labeled 1 ($\beta_j = 1$) and -1 ($\beta_j = -1$) are denoted by N_+ and N_-, respectively. For the calibration related to a pointwise classifier (e.g., SVM,

OCSVM), Platt used the Levenberg–Marquardt (LM) algorithm to solve (6.10) (Press et al., 2007). Later, this was improved by Lin et al. (2007) who demonstrated that (6.10) is a convex optimization problem and presented a more robust algorithm with proven theoretical convergence to solve it.

To calibrate the group classifier OCSMM, it is essential to generate anomalous groups rather than individual anomalous points. Inspired by previous work on anomalous points (Clifton et al., 2014; Tax and Duin, 2001; Markou and Singh, 2006), we have adapted these ideas for groups. For generating anomalous points, the central assumption is that spatially all such points lie outside high-density regions. Building upon this assumption, previous studies focus on constructing anomalous points for an optimal calibration, where a threshold on the density $y = p(\boldsymbol{x})$ is determined to classify a point as anomalous or not. However, anomalous groups cannot be generated by the same approach as anomalous points, as they exhibit more complex structural bahaviour. Unlike anomalous points, anomalous groups may not only deviate spatially from the high-density regions but can also significantly overlap with normal groups. These simulated groups are specifically constructed to introduce mild deviations from normal patterns near the decision boundary, serving as informative references for calibrating the OCSMM score. Based on the above analysis, we propose the following algorithm for generating anomalous groups.

Algorithm 6.2 starts from generating the group centers. A key difference between Algorithm 6.2 and the pointwise calibration methods discussed earlier is that the threshold, a hyperparameter used to define the boundary of high-density regions, is removed. This modification allows group centers to be located in both high-density and low-density regions. The next step focuses on ensuring that anomalous groups exhibit as much diversity in their distribution characteristics as possible. To achieve this, the covariance matrix is generated randomly under constraints derived from the covariance matrices of the normal groups in the training set. Then a new anomalous group for the calibration is generated by sampling from a multivariate Gaussian distribution, using the group center as the location and the generated covariance matrix as the scale.

Algorithm 6.2 avoids generating extreme groups that would be too easily detected, and eliminates the need for the intricate threshold determination step required in methods for generating anomalous points. This design provides a pragmatic approach to addressing the scarcity of labeled group anomalies, and aims to improve probabilistic interpretability rather than explicitly simulating specific unknown anomalous bahaviour. After obtaining the anomalous groups, the sigmoid calibration technique is applied. We then define a distribution-based anomaly score as the calibrated OCSMM score $\beta(\tilde{\boldsymbol{x}})$ to quantify how

Algorithm 6.2 Generate a set of anomalous groups $\mathbb{B}$ for the calibration

Input: A set of normal groups $\mathbb{A} = \{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_l\}$; the number of anomalous groups h; the expected size $\lambda^{(l)}$ of each anomalous group;

Output: A set of anomalous groups $\mathbb{B}$ of size h.

1: Compute the centers $\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_l$ and covariance matrices $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_l$ of the normal groups in the training set. Compute the center $\boldsymbol{c}^*$ of all the centers $\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_l$ and the average Euclidean distance r from the centers $\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_l$ to $\boldsymbol{c}^*$.

2: Generate a set of h points $B = \{\boldsymbol{b}_{l+1}, \ldots, \boldsymbol{b}_j, \ldots, \boldsymbol{b}_{l+h}\}$ uniformly within a hypersphere with center $\boldsymbol{c}^*$ and radius $2r$.

3: **for** each $\boldsymbol{b}_j$ in B **do**

4: Identify the nearest normal group in $\mathbb{A}$, where the distance $\Delta_{\boldsymbol{b}_j}$ is computed as the minimal Euclidean distance between $\boldsymbol{b}_j$ and the centers in $\mathbb{A}$. A symmetric covariance matrix $\boldsymbol{v}_{\boldsymbol{b}_j}$ is generated based on the covariance matrix of the nearest group. Each element is randomly sampled from a uniform distribution between 0 and the corresponding element of the nearest group's covariance matrix. The resulting matrix is symmetrized by adding it to its transpose and dividing by two.

5: Use $\boldsymbol{b}_j$ as the center and the covariance matrix $\boldsymbol{v}_{\boldsymbol{b}_j}$ to generate a group $\tilde{\boldsymbol{x}}_{\boldsymbol{b}_j}$ of points randomly from a multivariate Gaussian distribution, with the number of points drawn from a Poisson distribution with mean $\lambda^{(l)}$.

6: **end for**

7: Obtain a set of anomalous groups $\mathbb{B} = \left\{\tilde{\boldsymbol{x}}_{l+1}, \ldots, \tilde{\boldsymbol{x}}_{\boldsymbol{b}_j}, \ldots, \tilde{\boldsymbol{x}}_{l+h}\right\}$.

anomalous the high-density region part of a group is. The overall steps for obtaining the distribution-based anomaly score $\beta(\tilde{\boldsymbol{x}})$ using calibrated OCSMM are summarized in Algorithm 6.3.

Algorithm 6.3 The calibrated OCSMM: distribution-based group anomaly score

Input: A set of normal groups $\mathbb{A} = \{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_l\}$; a test group set $\mathbb{T} = \{\tilde{\boldsymbol{x}}_1^{(t)}, \ldots, \tilde{\boldsymbol{x}}_{l'}^{(t)}\}$; the number of anomalous groups h; the expected size n of each group;

Output: The distribution-based anomaly score $\beta(\mathbb{T})$ of the test group set $\mathbb{T}$.

1: Train an OCSMM model based on the normal group set $\mathbb{A} = \{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_l\}$.

2: Using Algorithm 6.2, generate h anomalous groups, $\mathbb{B} = \left\{\tilde{\boldsymbol{x}}_{l+1}, \ldots, \tilde{\boldsymbol{x}}_{\boldsymbol{b}_j}, \ldots, \tilde{\boldsymbol{x}}_{l+h}\right\}$, with each group size drawn from a Poisson distribution with expectation parameter n.

3: Compute the OCSMM scores for the normal groups, $\gamma(\mathbb{A}) = \{\gamma(\tilde{\boldsymbol{x}}_1), \ldots, \gamma(\tilde{\boldsymbol{x}}_l)\}$, and the anomalous groups, $\gamma(\mathbb{B}) = \{\gamma(\tilde{\boldsymbol{x}}_{l+1}), \ldots, \gamma(\tilde{\boldsymbol{x}}_{\boldsymbol{b}_j}), \ldots, \gamma(\tilde{\boldsymbol{x}}_{l+h})\}$, using the decision function $\gamma(\cdot)$ of the trained OCSMM model.

4: Fit a sigmoid model to $\gamma(\mathbb{A})$ and $\gamma(\mathbb{B})$ by estimating the parameters g and q.

5: Apply the trained OCSMM model to compute the OCSMM scores for the test set, $\gamma(\mathbb{T})$, and then use the trained sigmoid model to obtain the calibrated scores as $\beta(\mathbb{T}) = \{\beta(\tilde{\boldsymbol{x}}_1^{(t)}), \ldots, \beta(\tilde{\boldsymbol{x}}_{l'}^{(t)})\}$.

6.3.2 Extreme value support measure machine

In previous sections, we introduced two group anomaly scores to evaluate two complementary parts of the group: a point-based anomaly score for a group's low-density region part, denoted as $\alpha(\tilde{\boldsymbol{x}})$, and a distribution-based anomaly score for a group's high-density region part, denoted as $\beta(\tilde{\boldsymbol{x}})$. Specifically, $\alpha(\tilde{\boldsymbol{x}})$ assesses the anomalousness of the aggregate entity of exceedances within the low-density regions. In contrast, $\beta(\tilde{\boldsymbol{x}})$ evaluates whether the distributional characteristics of points, considered as an aggregate within the high-density regions deviate significantly from those observed in the training groups. The next step is to develop an integrated framework that effectively leverages both probabilistic scores to capture the overall level of anomalousness.

For this purpose, we take inspiration from the way certainty factors are combined, encountered in the early days of expert systems such as PROSPECTOR (Duda et al., 1976) and MYCIN (Buchanan and Shortliffe, 1984) capable of reasoning under uncertainty. Functions such as Van Melle's combining function (De Baets and Fodor, 1999), typically of the type $[-1,1]^2 \to [-1,1]$, were well-thought functions designed for that purpose. Only in the nineties (De Baets and Fodor, 1999), after linearly rescaling to functions of the type $[0,1]^2 \to [0,1]$, it was realized that they were examples of a broader class of aggregation functions on the unit interval called uninorms. Uninorms operate on the unit interval $[0,1]$, providing a systematic way to combine probabilistic scores. In our case, the two probabilistic scores, each ranging from $[0,1]$, align particularly well with uninorms, which are binary operations capable of meeting these engineering demands.

We now argue that uninorms with neutral element e, which acts as a hyperparameter, are ideally suited for our purpose. In engineering applications with multiple measures assessing different aspects of a system, it is essential that when more than one measure exceeds a given threshold (triggering alarms for specific aspects), the overall measure should reflect a more severe status of the system. For instance, in health monitoring, when a patient exhibits multiple anomalous health indicators, we tend to consider the overall health condition to be worse. Constructing a single measure that exceeds any anomalous health indicator, when assessed on the same scale, provides more informative insights into the patient's condition. In such cases, traditional weighting methods, such as weighted averaging and t-conorms (e.g., maximum), either dilute dominant signals or fail to model mutual reinforcement. Uninorms offer a principled mechanism to capture both synergistic and antagonistic effects, which is critical in group anomaly scenarios involving dual sources of evidence.

A uninorm is a symmetric, associative, increasing binary operation on $[0, 1]$ with a neutral element $e \in [0, 1]$. The case $e = 1$ corresponds to the well-known t-norms (including the minimum operation), and the case $e = 0$ to the well-known t-conorms (including the maximum operation). Due to its associativity, any uninorm U is either conjunctive (*i.e.*, $U(0, 1) = U(1, 0) = 0$) or disjunctive (*i.e.*, $U(0, 1) = U(1, 0) = 0$). The structure of a uninorm with neutral element e is (Fodor et al., 1997): on $[0, e]^2$ it behaves as a rescaled t-norm, thus taking values below the minimum, on $[e, 1]^2$ it behaves as a rescaled t-conorm, thus taking values above the maximum, while on the remaining parts of the unit square (*i.e.*, $[0, e] \times [e, 1] \cup [e, 1] \times [0, e]$) it behaves as a compensatory (also called averaging) operator, taking values between the minimum and the maximum. Given a t-norm and a t-conorm, one can always build a uninorm by considering minimum or maximum as (one of the two extreme) compensatory operator.

In our case, suppose that exactly one of the anomaly scores takes a value higher than e, then one would consider this as an indication of the group being anomalous, irrespective of the other score, and thus consider it as the overall anomaly score. Hence, to construct a uninorm, we consider the maximum as compensatory operator a viable choice. However, if both anomaly scores exceed e, then they could be seen as reinforcing one another and act synergistically, yielding an overall score exceeding the maximum of the individual scores. Similarly, if both anomaly scores take a value lower than e, then they should act antagonistically, yielding an overall score lower than their minimum. The symmetry of the uninorm expresses that both anomaly scores play the same role.

The most famous uninorm having the desired synergistic and antagonistic bahaviour is the 3Π operator $U_{3\Pi} : [0, 1]^2 \to [0, 1]$ with neutral element $\frac{1}{2}$:

$$U_{3\Pi}(a, b) = \frac{ab}{ab + (1 - a)(1 - b)}$$

with either $U_{3\Pi}(0, 1) = U_{3\Pi}(1, 0) = 0$ or $U_{3\Pi}(0, 1) = U_{3\Pi}(1, 0) = 1$. This operation is nothing else but a rescaled version of the function used for combining certainty factors in the rule-based system PROSPECTOR. Without going into detail, we mention that all members of the important class of representable uninorms are isomorphic to this iconic 3Π operator, illustrating its central role.

Recalling our argumentation above that on $[0, e] \times [e, 1] \cup [e, 1] \times [0, e]$ the selected uninorm should preferably act as the maximum, we can modify the 3Π operator (keeping the underlying t-norm and t-conorm), resulting in the uninorm U_{ano} used in our final

aggregation method. It is given by

$$U_{ano}(a,b) = \begin{cases} 2e\, U_{3\Pi}(\frac{a}{2e}, \frac{b}{2e}) & , \text{ if } (a,b) \in [0,e]^2 \\ e + (1-e)\big[2U_{3\Pi}(\frac{1}{2} + \frac{a-e}{2-2e}, \frac{1}{2} + \frac{b-e}{2-2e}) - 1\big] & , \text{ if } (a,b) \in]e,1]^2 \\ \max(a,b) & , \text{ elsewhere}\,. \end{cases} \tag{6.11}$$

The detailed derivation of the uninorm U_{ano} based on the 3Π operator $U_{3\Pi}$ is provided in the appendix. As a result, the combination of the two anomaly scores is completed. We present the full procedure of the proposed model EVSMM in Algorithm 6.4.

Algorithm 6.4 Extreme value support measure machine (EVSMM)

Input: A set of normal groups $\mathbb{A} = \{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_l\}$; A test group set $\mathbb{T} = \{\tilde{\boldsymbol{x}}_1^{(t)}, \ldots, \tilde{\boldsymbol{x}}_{l'}^{(t)}\}$;
Output: The anomaly scores $\phi(\tilde{\boldsymbol{x}}^{(t)})$ of the test group set $\mathbb{T}$.

1: Apply KDE on the normal groups in $\mathbb{A}$ to have an estimation of the density of each point. Then transform the normal training groups $\tilde{\boldsymbol{x}}$ to $\boldsymbol{z}$ as shown in Section 6.3.1. Compute c_n, d_n through Eq. (2.5), respectively.
2: Train an OCSMM model based on the training set $\mathbb{A}$. Then apply Algorithm 6.2 and the objective function (6.10) to get a sigmoid model $\beta(\tilde{\boldsymbol{x}})$ for the calibration of the OCSMM score.
3: **for** each $\tilde{\boldsymbol{x}}_j^{(t)}$ in $\mathbb{T}$ **do**
4: (**Probabilistic point-based score**) The likelihood of a group with respect to an aggregate of exceedances in low-density regions is obtained from Eq. (6.2). Then anomaly score $\alpha(\tilde{\boldsymbol{x}}_j^{(t)})$ is determined from Eq. (6.8) and Eq. (6.6) based on the analytical asymptotic CDF of the likelihood.
5: (**Probabilistic distribution-based score**) Compute the OCSMM score $f(\tilde{\boldsymbol{x}}_j^{(t)})$. Then calibrate the OCSMM score through the trained sigmoid model Eq. (6.9) to obtain the probabilistic distribution-based score $\beta(\tilde{\boldsymbol{x}}_j^{(t)})$.
6: (**Final score**) Compute the final anomaly score of the group $\tilde{\boldsymbol{x}}_j^{(t)}$ as $\phi(\tilde{\boldsymbol{x}}_j^{(t)}) = U_{ano}\Big(\alpha(\tilde{\boldsymbol{x}}_j^{(t)}), \beta(\tilde{\boldsymbol{x}}_j^{(t)})\Big)$ by applying (6.11).
7: **end for**

6.4 Experiments

In this section, we present extensive experiments to demonstrate the effectiveness of the proposed EVSMM in detecting unusual group bahaviour. The performance of the EVSMM is compared with five other group anomaly detectors: the discriminant model

SMDD[1] (Guevara et al., 2015), the OCSMM[2] (Muandet and Schölkopf, 2013) and its calibrated output based on a sigmoid transformation (Cali-OCSMM), the generative model MGMM[3] (Xiong et al., 2011a), and an extreme value model (EVM) based on extreme value theory (Luca et al., 2018).

6.4.1 Experiments on synthetic data

We will perform an experiment using synthetic data with a set-up comparable to the one used in (Xiong et al., 2011a; Muandet and Schölkopf, 2013; Guevara et al., 2015; Chalapathy et al., 2018; Song et al., 2020). Following these references, the normal groups are simulated from two Gaussian mixture distributions in the plane that each consists of four components. The mixing proportions for the Gaussian mixtures are respectively given by $(0.22, 0.64, 0.03, 0.11)$ and $(0.22, 0.03, 0.64, 0.11)$. To generate a normal group, probabilities $(0.48, 0.52)$ are assigned to determine which Gaussian mixing proportion is applied. The selected Gaussian mixing proportion is then used to generate the group. The components of both mixtures are centered at $(-1, -1)$, $(1, -1)$, $(0, 1)$ and $(1, 1)$ and share the covariance matrix $\boldsymbol{\Sigma} = 0.15 \times \boldsymbol{I}_2$, with $\boldsymbol{I}_2$ the 2×2 identity matrix.

Three types of anomalous groups are considered: (i) point-based anomalous groups, (ii) distribution-based anomalous groups and (iii) groups that are anomalous as a mixture of unusual low-density and high-density region parts. The threshold for identifying low-density regions is determined based on a mean residual plot. To generate groups with varying extents of point-based anomalous bahaviour, the number of low-density region points is randomly chosen according to a Poisson distribution with mean $\lambda^{\text{ext}} = 3$, truncated at zero to ensure the presence of at least one anomalous point. For a more detailed investigation of the impact of λ^{ext}, additional experiments were conducted with varying λ^{ext} (as presented later). These experiments reveal that as λ^{ext} increases, other group anomaly detectors, such as MGMM, begin to achieve performance comparable to our model. This phenomenon occurs because an increase in the number of exceedances gradually transforms point-based anomalous group bahaviour into distribution-based anomalous group bahaviour. Groups exhibiting various distribution-based anomalous bahaviour are simulated through a Gaussian mixture distribution with the same centers as the normal groups but different mixing proportions. The proportions $(0.6, 0.1, 0.07, 0.23)$, as used

[1] https://github.com/jorjasso/SMDD-group-anomaly-detection
[2] https://github.com/kdgutier/ocsmm
[3] https://www.cs.cmu.edu/ lxiong/gad/gad.html

by Muandet and Schölkopf (2013), are complemented by including three additional configurations: $(0.80, 0.01, 0.06, 0.13)$, $(0, 0, 0.06, 0.94)$ and $(0.64, 0.22, 0.03, 0.11)$. A mixed anomalous group contains both an aggregate of points in low-density regions and an aggregate of points randomly simulated from unusual mixing proportions in high-density regions.

To evaluate our method, with varying extents of distribution-based and point-based anomalous bahaviour, nine distinct cases were designed to ensure a diverse representation of anomalous group configurations in the test set. The corresponding experiments consider varying fractions of the three types of anomalous groups, defined as $[1 : 0 : 0]$, $[0 : 1 : 0]$, $[0 : 0 : 1]$, $[0.5 : 0.5 : 0]$, $[0 : 0.5 : 0.5]$, $[0.5 : 0 : 0.5]$, $[0.5 : 0.25 : 0.25]$, $[0.25 : 0.5 : 0.25]$, and $[0.25 : 0.25 : 0.5]$ (the first component representing point-based anomalous groups, the second distribution-based ones and the third one a mixture thereof). In each of the experiments, we inject 60 anomalous groups into the test set according to the corresponding proportions.

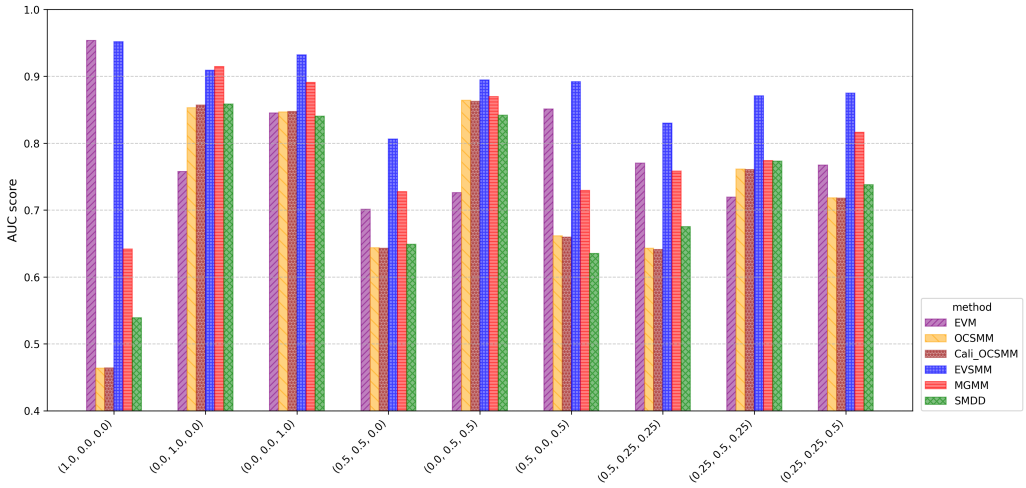


Figure 6.2: Average AUC scores (10-fold cross-validation) for nine synthetic cases. EVSMM consistently achieves a strong performance across all cases, while EVM excels in detecting point-based group anomalies. In contrast, OCSMM, MGMM, and SMDD are more effective for distribution-based group anomalies, as reflected in cases with higher proportions of such groups.

We use 1000 normal groups for training. The training involves a 10-fold cross-validation: in each run, the dataset is randomly partitioned into ten parts, of which seven are used for training, two for hyperparameter optimization in a validation step, and one for testing. In the proposed EVSMM approach, two hyperparameters are subject to tuning: γ in the Cali-OCSMM model controlling the kernel width, and e, the threshold in the uninorm that determines the aggregation bahaviour when combining anomaly scores designed to evaluate complementary aspects of the group. These hyperparameters are optimized using

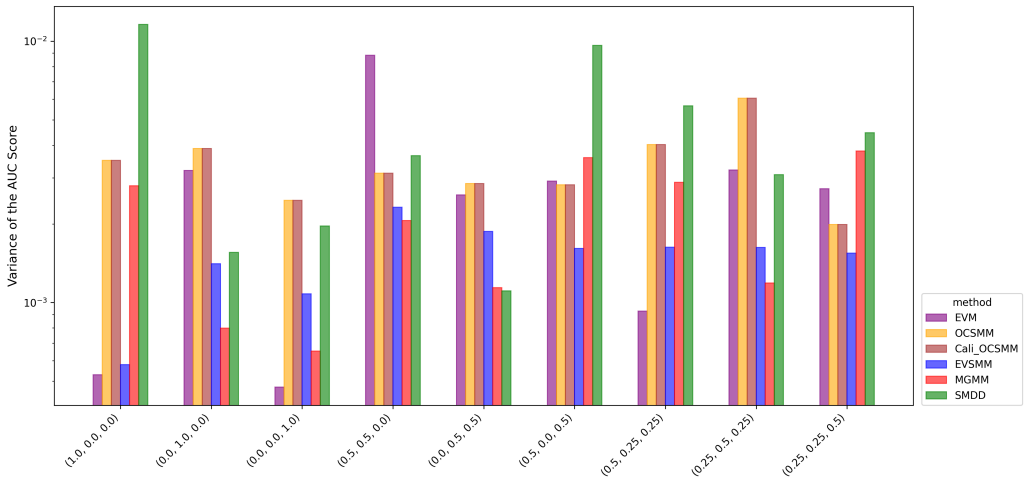


Figure 6.3: Variance of AUC scores (logarithmic scale) based on 10-fold cross-validation for nine synthetic cases. EVSMM exhibits stable performance with low variance across most cases. MGMM shows lower variance than EVSMM in cases with few point-based group anomalies.

a grid search strategy guided by nested cross-validation where $\gamma \in \{0.5, 0.8, 1.1, 1.4\}$ and $e \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. The group size is simulated using a Poisson distribution with mean $\lambda^{(l)} = 50$. Model performances are analysed based on the ROC curve and the AUC metric. The detailed results are presented in Figure 6.2–6.4.

In contrast to the other approaches, from the results shown in Figure 6.2, it follows that EVSMM achieves consistently competitive performance across all cases. MGMM is also relatively effective compared to other models, achieving AUC scores exceeding 0.75 in 8 out of 9 cases. However, it is observed that MGMM is less performant in evaluating point-based anomalous groups that rather have a limited number of exceedances. Although MGMM is designed to detect group anomalies of different types, its ability to detect point-based anomalous group bahaviour requires a sufficient number of exceedances, which will be examined further through experiments varying λ^{ext}. From the results in cases $[1:0:0]$ and $[0:1:0]$, it is clear that the EVM is particularly suited to effectively detect point-based group anomalies. Although half of the groups in case $[0.5:0:0.5]$ exhibit varying extents of distribution-based anomalous group bahaviour, EVM stands out for its ability to quantify point-based anomalous group bahaviour across all groups. However, when compared with the outcomes from case $[0:0.5:0.5]$, it becomes evident that EVM is unable to address the distribution-based anomalous group bahaviour present in half of the groups in case $[0.5:0:0.5]$. SMDD, MGMM, and OCSMM are rather suited for groups with an unusual aggregate of points in high-density regions, as shown in cases $[0:1:0]$ and $[0:0.5:0.5]$, where every group exhibits a varying extent of distribution-based anomalous bahaviour, resulting in higher AUC values for the these models. Also in

cases [0.25 : 0.5 : 0.25] and [0.25 : 0.25 : 0.5] with a high proportion of distribution-based group anomalies, three models show their detection capabilities. In the Appendix, we show comprehensive ROC curves for all nine cases allowing to compare the performances of various approaches.

Figure 6.3 shows the variance of the AUC scores on a logarithmic scale. Overall, EVSMM shows a satisfactory stability of the AUC scores. In cases with a relatively low proportion of point-based group anomalies, the MGMM has a lower variance than EVSMM. The OC-SMM and Cali-OCSMM have the same variance in all cases confirming that the sigmoid calibration is effective. The Appendix shows an alternative graphical representation using boxplots to compare the AUC scores among various methods for each case.

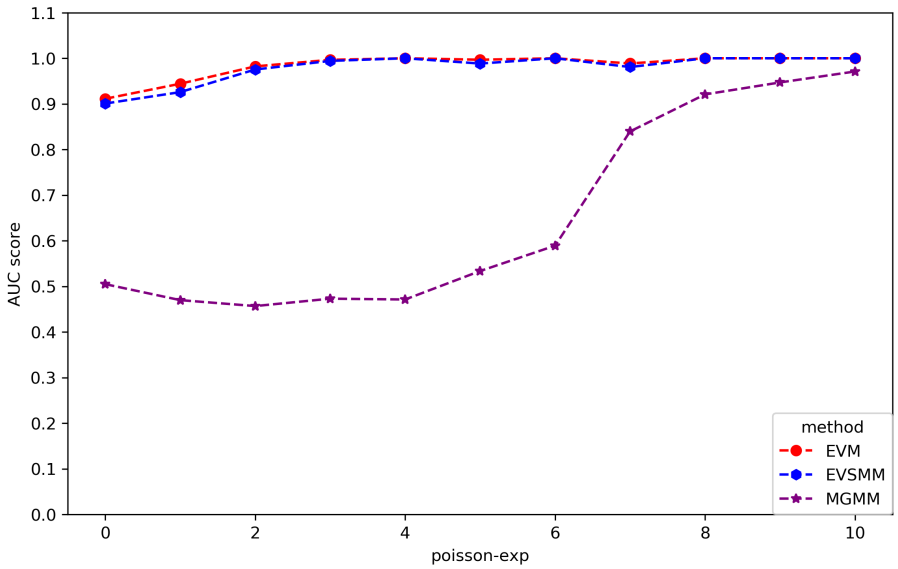


Figure 6.4: Average AUC score (10-fold cross-validation) for the synthetic data where the mean number λ^{ext} of anomalous points in each group is varied between 0 and 10. EVSMM and EVM maintain a stable performance across the entire range, and MGMM shows clear performance improvement as λ^{ext} increases, reflecting its sensitivity to the number of exceedances when quantifying point-based group anomalies.

Furthermore, we conduct an experiment that focuses exclusively on point-based group anomalies, specifically by varying the number of points located in low-density regions within groups. In our previous experiments, we assumed that the number of points in low-density regions is governed by a Poisson distribution with expectation parameter $\lambda^{\text{ext}} = 3$. Here, we vary this parameter from 0 to 10 and look at the effect on the performance of MGMM, EVM, and EVSMM, all of which possess the capability to evaluate point-based anomalous group bahaviour. Figure 6.4 reveals that as the number of points in low-density regions increases, the performance of MGMM is comparable with those of EVM and EVSMM. However, when the number is lower, MGMM fails to quantify point-

based anomalous bahaviour. This can be attributed to the gradual transition of point-based anomalous group bahaviour into distribution-based anomalous group bahaviour as the number of exceedances increases. From the previous results, although the overall performance of MGMM is not as strong as that of EVSMM, MGMM exhibits greater sensitivity to this transition compared to other models that excel in detecting distribution-based group anomalies.

6.4.2 Experiments on the sloan digital sky survey data

In this section, we evaluate the EVSMM on the sloan digital sky survey (SDSS) dataset. This dataset includes information on millions of celestial objects, such as stars, galaxies, and various other astronomical phenomena. It provides detailed measurements of the positions, brightness, and spectra of these objects. Our focus is on studying the subset of galaxies, which consists of a large volume of 4000-dimensional spectra of about 7×10^5 galaxies. We adopt the 1000-dimensional feature vectors derived from the original 4000-dimensional spectra, as provided in the publicly available benchmark dataset used in prior studies (Xiong et al., 2011a; Muandet and Schölkopf, 2013; Guevara et al., 2015; Song et al., 2020). We then apply principal component analysis (PCA) to further reduce the data to 4 dimensions. The resulting four principal components retain approximately 85% of the total variance. Specifically, PC1, PC2, PC3, and PC4 account for 77.2%, 5.8%, 1.6%, and 0.5% of the total variance, respectively. This reduced representation is subsequently used as input for density estimation. Specifically, KDE is performed on the resulting 4-dimensional features, thereby mitigating the impact of high dimensionality on the estimation process.

Xiong et al. (2011a) were the first to utilize the SDSS dataset for group anomaly detection. They identified 505 spatial groups of galaxies using a constructed neighborhood graph, where each connected component in the graph represents a group. Then they conducted group anomaly detection on these 505 groups without prior knowledge of any group anomaly labels. The results gained positive recognition from astronomers. Furthermore, in order to get a statistically meaningful comparison among various methods, artificial anomaly injections were employed to address the absence of labels, a method that was commonly adopted in subsequent research (Póczos et al., 2011; Xiong et al., 2011a; Muandet and Schölkopf, 2013; Song et al., 2020; Guevara et al., 2015).

Building on the injection method used in previous studies, we increase the complexity

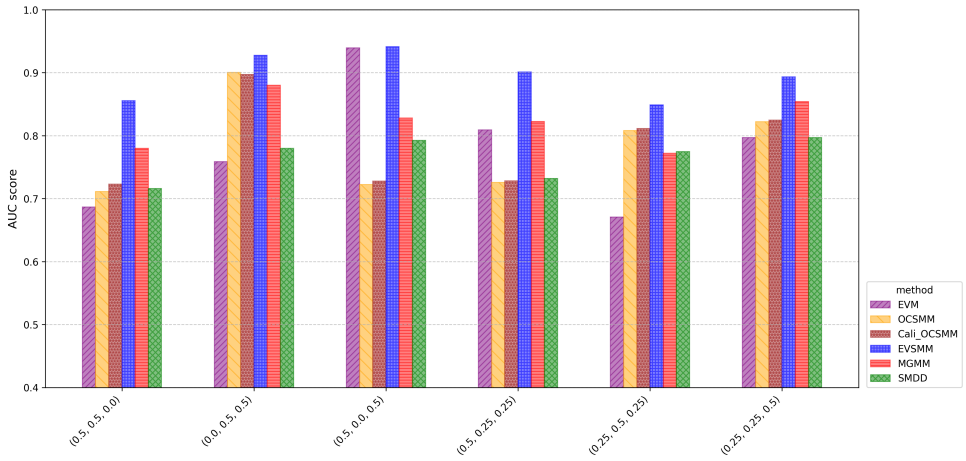


Figure 6.5: Average of the AUC scores (10-fold cross-validation) for the SDSS data across different anomaly configurations. EVSMM consistently achieves high AUC values and outperforms other methods across all cases. MGMM shows uneven performance depending on the anomaly configuration.

of the injected anomalous groups to avoid configurations that are trivial to distinguish. Point-based, distribution-based, and mixed anomalous groups are constructed by randomly selecting existing points from the training group set. Specifically, point-based group anomalies are generated by aggregating galaxies located in low-density regions, with aggregate sizes following a Poisson distribution with $\lambda^{\text{ext}} = 3$, whereas distribution-based group anomalies are constructed by randomly selecting galaxies from high-density regions, with the constraint that the selected galaxies are not all drawn from the same group. We consider six cases, where in each case a total of 50 anomalous groups are injected with proportions of point-based, distribution-based, and mixed anomalous groups set to $[0.5 : 0.5 : 0]$, $[0 : 0.5 : 0.5]$, $[0.5 : 0 : 0.5]$, $[0.5 : 0.25 : 0.25]$, and $[0.25 : 0.5 : 0.25]$, respectively.

From Figure 6.5, EVSMM demonstrates consistently high AUC scores across all cases, outperforming or being comparable to other methods. Considering Figure 6.6, which shows the variance of the AUC scores on a logarithmic scale, we observe that EVSMM stands out for both high AUC scores and low variance, showcasing its capability to effectively and reliably handle various types of group anomalies. While MGMM achieves high AUC scores in many cases, its performance is less stable in configurations dominated by point-based group anomalies, as reflected in its higher variance. EVM and Cali-OCSMM perform well at detecting distribution-based group anomalies but exhibit less stability compared to EVSMM. On this real-world dataset, we observe that SMDD exhibits notable variability in its AUC score variances across different data configurations. This may suggest that, compared to other methods, the stability of SMDD is more affected by the

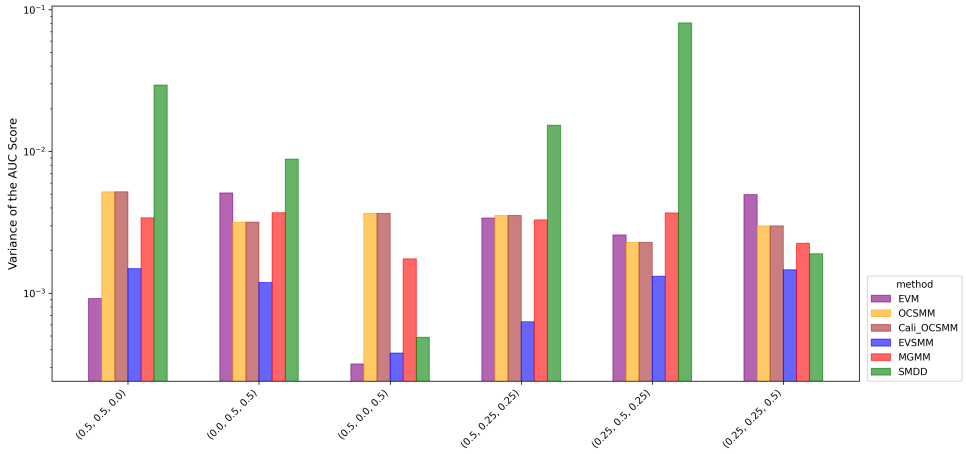


Figure 6.6: Variance of the AUC scores (shown on a logarithmic scale) (10-fold cross-validation) for the SDSS data. EVSMM consistently exhibits a low variance, indicating strong stability across all cases. SMDD shows fluctuating variances, particularly under configurations involving point-based or mixed group anomalies.

dataset properties, resulting in marked discrepancies in variance values.

The performances of OCSMM and Cali-OCSMM are very similar indicating that the calibration method for the OCSMM is effective for the SDSS data. In the Appendix, we show for each case the ROC curves and boxplots of the AUC scores. These results highlight the effectiveness and robustness of EVSMM, particularly in scenarios with mixed anomaly types with varying extents of anomalous bahaviour. While MGMM shows strong potential, its sensitivity to specific configurations (e.g., $[0.5 : 0 : 0.5]$, $[0.5 : 0.25 : 0.25]$) warrants further investigation to enhance its stability and performance.

6.5 Conclusions and future work

In this work, we have proposed an efficient method for comprehensively detecting group anomalies, termed EVSMM. The method integrates a PPM based on EVT and a calibrated OCSMM, which respectively evaluate complementary aspects of the target group, with the scores subsequently being integrated using a constructed uninorm. Indeed, through the PPM we formally construct the likelihood for an aggregate of low-density region points within a group, and more importantly, we derive the analytical distribution of the likelihood. The calibrated OCSMM quantifies the aggregate of points within high-density regions, addressing the limitation of its raw output which provides only classification without probability estimates. Through this calibration, a novel approach is

introduced for generating diverse anomalous groups, which offers a reference for future research on the calibration and validation of group models. A uninorm is constructed to combine two anomaly scores with the ability to locally synergize, antagonize, and weigh these scores relative to the threshold e.

Extensive experiments show that existing group anomaly detection approaches such as OCSMM, MGMM, and SMDD are mainly focusing on the detection of distribution-based group anomalies, but lack the capability to comprehensively quantify diverse group anomalies. Alternative approaches that are built upon methods centered around the detection of individual anomalous points tend to inaccurately model the point-based anomalous group bahaviour, and face challenges managing the multiple hypothesis problem. The proposed EVSMM effectively quantifies various anomalous group bahaviour and, unlike existing methods, provides a probability-based anomaly score that enhances interpretability and utility in practical applications.

Future work opens several promising directions. We will keep exploring the model's ability to detect varying extents of anomalous bahaviour, extending research beyond evaluating the test set accuracy. Specifically, our plan is to study the relationship between the entropy of groups and the anomaly scores to gain further insights into the model's capabilities. Moreover, we intend to explore more meaningful real-world group anomaly detection tasks to assess the practical utility and robustness of our approach. These directions are expected to provide a clearer understanding of the model's strengths.

6

Part III

Epilogue

CHAPTER 7

Conclusions and future perspectives

7.1 Conclusions

Motivated by growing attention for clustered extreme events in diverse fields, this thesis provided new insights into modeling such events in a probabilistic lens. Basically, individual extreme events were characterized as exceedances lying in distributional tails above some threshold. Clustered extreme events were then statistically represented as realizations of a pattern of exceedances, varying in both their length and magnitude. Building on a classical discipline in statistics, the extreme value theory, this thesis developed an analytical framework to model exceedance patterns across diverse settings. The main contributions are summarized below.

In this thesis, three settings of the pattern of exceedances were investigated to demonstrate our theoretical framework:

- patterns of exceedances, where exceedances lie in the distributional tails of independent and identically distributed (i.i.d.) univariate variables.
- patterns of consecutive exceedances, where exceedances lie in the distributional tails of temporally dependent and non-identically distributed (non-i.i.d.) univariate variables.
- patterns of points located in low-density regions of multivariate variables.

Collectively, these investigations extended traditional extreme value analysis from isolated extremes to structured exceedance patterns, providing a unified probabilistic understanding of clustered extreme events.

In Chapter 2, the core results of classical extreme value theory were reviewed. The block-maxima and peaks-over-threshold models, which define extremes from different perspectives, were demonstrated, and their asymptotic results were demonstrated from the viewpoint of tail modeling. The notion of the maximum domain of attraction was then presented in order to clarify the conditions under which these asymptotic results hold. Finally, the Poisson point process limit for extremes was introduced, which unified the block-maxima and peaks-over-threshold models by revealing that both arise from the same limiting point process in the tails. This chapter also introduced several methodological tools that supported subsequent model development. As a result, this chapter established the theoretical basis for developing the analytical models throughout this thesis.

In Chapter 3, a theoretical framework for modeling a pattern of exceedances was proposed under conditions consistent with classical extreme value theory, that is, for univariate observations that are approximately independent and identically distributed. Exceedances in the upper tail were modeled by means of a Poisson point process limit. Subsequently, by leveraging the Janossy densities, which describe the probability densities for finite point configurations of a point process, we constructed a likelihood representation to characterize a pattern of exceedances, accounting for the randomness in both the number and the magnitude of the individual exceedances. A key theorem was then derived, providing a closed-form cumulative distribution function of this likelihood. This cumulative distribution function explicitly quantifies how unusual a given realization of a pattern of exceedances is within the sample space.

Chapter 4 presented the first application of the proposed theorem in the transportation domain, focusing on the critical challenge of infrastructure deployment in bike-sharing systems. Although bike-sharing systems offer an eco-friendly transport mode, challenges exist in adaptively identifying emerging high-demand areas for docking station or electronic fence deployment. Existing studies have advanced mathematical modeling of demand patterns based on illegal return activities, they still face limitations in spatial partitioning and computational efficiency. In this chapter, a novel partitioning strategy for the studied area was described, which requires no hyperparameter tuning. For each irregular subregion as a modeling unit, high-demand was quantified by a joint consideration of the number of returns, the spatial coverage, and the demand profile of its neighborhood subregions. While recent research applying graph-based models offers valuable insights, these approaches typically require intensive parameterization and have high computational costs. The proposed theorem in Chapter 3 was then applied to investigate the collective behaviour of unusual subregions within each part of a partition, achieving a global ranking of candidate subregions to determine to what extent an area requires new docking infrastructure. A methodology-level comparison demonstrated the computational efficiency of our approach compared to other recently proposed methods. A case study on the Münchner Verkehrsgesellschaft bike-sharing system illustrated the utility of our approach in the hybrid bike-sharing system.

Chapter 5 focused on the non-i.i.d. data and studied a more complex object than in Chapter 3, namely a pattern of consecutive exceedances in the distributional tails. In practice, non-i.i.d. data are commonly encountered in climate research, where observations of temperature, rainfall, and sea level exhibit dependence and non-stationarity. The risk analysis of extreme events remains a crucial task in climate research. This chapter

investigated heat waves for statistical extrapolation and risk estimation. The heat wave is a typical instance of the pattern of consecutive exceedances, with each realization defined as a period of consecutive hot days. To this end, the framework developed in Chapter 3 was extended to the tail of non-i.i.d. data, where the likelihood of the pattern of consecutive exceedances was derived with its cumulative distribution function. Within this framework, heat waves were assessed in terms of intensity, duration, and frequency, where the intensity was measured by the magnitude of exceedances, the duration was reflected by the size of the pattern, and the frequency was statistically measured by our derived cumulative distribution function. Finally, a novel method was proposed for computing the intensity–duration–frequency curves. Unlike existing approaches that compute return levels at different durations of heat waves on disparate scales, the proposed method provided a unified framework for calculating return levels and enabled global comparison across heat waves with differing durations and exceedance magnitudes.

In Chapter 6, the pattern of exceedances was investigated in a high-dimensional setting, within the context of group anomaly detection. In the univariate case, classical extreme value theory formalizes exceedances as points in the distributional tails. In high-dimensional settings, however, exceedances are considered as points lying in low-density regions. In order to derive the likelihood for patterns of exceedances, a mapping of data from high-dimensional to one-dimensional space was performed via the density, with the objective of preserving extreme information of low-density regions. Leveraging existing results of the density variable, which has a Weibull limit for its tail, a likelihood for the pattern of low-density points was derived, together with its cumulative distribution function. The derived analytical model was designed to assess anomalous group behaviour in low-density regions; this constitutes one of the two main classes of unusual group behaviour, the other being group anomalies in high-density regions. The proposed model was combined with a calibrated one-class support measure machine, which provided probabilistic outputs to assess the unusual point configuration in high-density regions, achieving an improved assessment of overall anomalous group behaviour.

7.2 Future perspectives

Future work encompasses methodological extensions and application studies. The pattern-of-exceedances framework established in this thesis can be extended to a broad range of practical contexts beyond those investigated here. In parallel, further theoretical work

can be developed to advance the asymptotic distributional results of classical extreme value theory.

When it comes to Chapter 4, future studies may incorporate richer operational information in the construction of the likelihood. Such information includes expected population and land-use changes in surrounding areas, service level requirements imposed by the operator, and other planning-related attributes. Incorporating these factors would allow the likelihood-based model to facilitate the determination of an appropriate capacity for each docking site. A second direction is to move from mainly spatial analysis to spatio-temporal analysis. Exceedance patterns can be defined over sliding time windows in order to detect short-term demand peaks that support timely rebalancing operations. Finally, the approach developed in Chapter 4 can be explored in broader contexts that share a similar structural setting. Investigating such applications would further demonstrate the generality of the proposed framework.

As for Chapter 5, another representative application is the rainfall in hydrology, where event intensity and duration vary nearly continuously in time, unlike heat waves whose duration is typically discretized by days. In this context, the framework can be extended by defining a realization of a pattern of exceedances as a continuous period during which the process remains above a chosen threshold. This would preserve tail extreme information while accommodating continuous durations. Copula-based models are widely used to capture dependence between rainfall intensity and duration. The proposed likelihood-based approach may offer an alternative that produces an explicit cumulative distribution for the joint tail behaviour. A focused comparison can be designed under common baselines, evaluating goodness of fit, extrapolation accuracy, and computational cost on matched datasets.

For the group-based work in Chapter 6, the proposed framework is useful for assessing collections of extremes in a high-dimensional space. At the same time, there are several aspects that deserve further consideration. One aspect that remains open for improvement is that the performance of the framework is linked to the density-based intermediate measure. Although a variety of density estimation techniques has been developed, efficiently obtaining densities in high-dimensional settings remains challenging. Future research may therefore investigate alternative intermediates that preserve the extreme-related information but are less sensitive to the curse of dimensionality. Another aspect that deserves further attention is that practical applications that fully exploit this framework are still relatively scarce in the existing literature. Identifying problem domains whose data are naturally organised in groups and whose anomaly structure matches the idea of collec-

tions of exceedances in high-dimensional settings would be an important direction. Such studies would not only demonstrate the practical utility of the proposed approach but also clarify the contexts under which its advantages are most evident.

I am passionate about contributing to further theoretical developments in extreme value analysis as well. Although classical extreme value theory provides an elegant and systematic framework, its analytical results are limited to the tails of one-dimensional distributions. In multivariate settings, distribution-level asymptotic results remain less developed. Many studies construct a one-dimensional intermediate measure from high-dimensional data, where classical univariate extreme value theory can then be applied. For example, some works treat the Mahalanobis distance as an intermediate. Our Chapter 6 also follows this route by using the density variable as an intermediate construct. Currently, we are working toward investigating asymptotic results in the high-dimensional space directly from the original distribution without using any intermediates. Given the structural complexity of multivariate settings, our plan starts from foundational cases, such as the multivariate Gaussian variable. The aim is to derive the limit representation for the extreme distribution directly in the data space. To that end, we are developing a kernel of extremes approach. The core idea is to make use of the geometric similarity between the extreme region and the distribution's level sets. For multivariate Gaussian variables, the extreme density contour has the same geometric form as the contours of the parent distribution. This alignment can be used to construct a contour-based representation of the tail region.

APPENDIX A

Appendix

A.1 L-moments method

A.1.1 Model

The L-moments method is an alternative to the MLE method for fitting a BM model for data. It is often preferred in practice to estimate (c_n, d_n, ξ) for a GEV distribution because of two reasons:

- It provides closed-form estimates for the GEV distribution with simplicity and robustness.
- The MLE method does not provide an analytical solution, which requires iterative numerical optimization.

The core idea of the L-moments method is to compute the first three sample L-moments $(\hat{\ell}_1, \hat{\ell}_2, \hat{\ell}_3)$, and subsequently the sample L-moment ratios: the L-scale ratio ($\hat{t_2} = \hat{\ell}_2/\hat{\ell}_1$) and the L-skewness ($\hat{t}_3 = \ell_3/\hat{\ell}_2$). This method then equates these sample L-moment ratios to their corresponding theoretical L-moment ratios of the GEV distribution. Specifically, the shape parameter ξ is first solved from the non-linear equation $\hat{t}_3 = t_3(\xi)$. The location $\hat{c_n}$ and scale $\hat{d_n}$ can then be determined in closed form once $\hat{\xi}$ is determined. We detail the procedures as follow.

Before calculating the L-moments, it is necessary to first compute the probability-weighted moments (PWMs). They were introduced as linear combinations of the order statistics by Greenwood et al. (1979) and later refined by Hosking (1990). Crucially, L-moments are linear combinations of PWMs. Since the GEV distribution has three parameters (c_n, d_n, ξ), the method relies on the first three L-moments (ℓ_1, ℓ_2, ℓ_3), which requires the calculation of the first three PWMs (b_0, b_1, b_2).

After sorting samples of block maxima M_n in the ascending order: $x_{1:m} \leq x_{2:m} \leq \cdots \leq x_{m:m}$, the unbiased estimates of the first three PWMs $(\hat{b}_0, \hat{b}_1, \hat{b}_2)$ are calculated as:

$$\hat{b}_0 = \frac{1}{m}\sum_{j=1}^{m} x_{j:m}\,,$$

$$\hat{b}_1 = \frac{1}{m}\sum_{j=1}^{m} \frac{j-1}{m-1}\, x_{j:m}\,,$$

$$\hat{b}_2 = \frac{1}{m}\sum_{j=1}^{m} \frac{(j-1)(j-2)}{(m-1)(m-2)}\, x_{j:m}\,.$$

Then, the sample L-moments $(\hat{\ell}_1, \hat{\ell}_2, \hat{\ell}_3)$ are calculated as linear combinations of these PWM estimates:

(i) $\hat{\ell}_1$ (L-location):

$$\hat{\ell}_1 = \hat{b}_0 \,.$$

(ii) $\hat{\ell}_2$ (L-scale):

$$\hat{\ell}_2 = 2\hat{b}_1 - \hat{b}_0 \,.$$

(iii) $\hat{\ell}_3$ (third L-moment):

$$\hat{\ell}_3 = 6\hat{b}_2 - 6\hat{b}_1 + \hat{b}_0 \,.$$

The sample L-skewness ratio $\hat{t}_3$ is computed:

$$\hat{t}_3 = \frac{\hat{\ell}_3}{\hat{\ell}_2}$$

This is a pivotal step of the L-moments method. The theoretical L-skewness ratio t_3 for the GEV distribution is a function of the shape parameter ξ only, being independent of the location c_n and scale d_n parameters, which is a major advantage of L-moments method. We estimate ξ by solving the equation:

$$\hat{t}_3 = t_3^{\mathrm{GEV}}(\xi) \,.$$

For $\xi \neq 0$, the theoretical L-skewness ratio is given by:

$$t_3^{\mathrm{GEV}}(\xi) = 2\,\frac{1 - 3^{-\xi}}{1 - 2^{-\xi}} - 3 \,.$$

Once the shape parameter $\hat{\xi}$ is determined, the scale parameter d_n and the location parameter c_n are estimated by equating the first two sample L-moments $\hat{\ell}_1$ and $\hat{\ell}_2$ to their theoretical GEV counterparts ℓ_1^{GEV} and ℓ_2^{GEV}.

The theoretical second L-moment for the GEV distribution ℓ_2^{GEV} is a function of both the scale parameter d_n and the shape parameter ξ:

$$\ell_2^{\mathrm{GEV}}(d_n, \xi) = \frac{d_n}{\xi}(2^{\xi} - 1)\Gamma(1 - \xi) \,.$$

By setting

$$\hat{\ell}_2 = \ell_2^{\mathrm{GEV}}(d_n, \hat{\xi}),$$

for $\hat{\xi} \neq 0$, $\hat{d}_n$ is obtained:

$$\hat{d}_n = \frac{\hat{\ell}_2 \hat{\xi}}{(2^{\hat{\xi}} - 1)\Gamma(1 - \hat{\xi})} \,. \tag{A.1}$$

Finally, the theoretical first L-moment ℓ_1^{GEV} is a function of all three parameters:

$$\ell_1^{\mathrm{GEV}}(c_n, d_n, \xi) = c_n + \frac{d_n}{\xi}(\Gamma(1 - \xi) - 1) \,.$$

We equate the sample first L-moment $\hat{\ell}_1$ to the theoretical ℓ_1^{GEV}:

$$\hat{\ell}_1 = \ell_1^{\mathrm{GEV}}(c_n, \hat{d}_n, \hat{\xi}) \,,$$

for $\hat{\xi} \neq 0$, $\hat{c}_n$ can be solved for in closed-form:

$$\hat{c}_n = \hat{\ell}_1 - \frac{\hat{d}_n}{\hat{\xi}}(\Gamma(1 - \hat{\xi}) - 1) \,. \tag{A.2}$$

If the estimated shape parameter is $\hat{\xi} \approx 0$ (the Gumbel case), Eq. (A.1) and Eq. (A.2) are evaluated by taking the limit as $\xi \to 0$:

$$\begin{aligned} \hat{d}_n &= \frac{\hat{\ell}_2}{\ln 2} \,, \\ \hat{c}_n &= \hat{\ell}_1 - \gamma \, \hat{d}_n \,, \end{aligned}$$

where $\gamma \approx 0.577$ is the Euler–Mascheroni constant.

A

A.1.2 Example

To illustrate the L-moments estimation procedure, we will use a sample of $m = 10$ block maxima $\{71.0, 76.9, 65.1, 79.2, 70.3, 62.0, 72.8, 83.1, 74.5, 67.5\}$. The ordered sample $\boldsymbol{x}^{M_n}$ are given by:

$$\boldsymbol{x}^{M_n} = \{62.0, 65.1, 67.5, 70.3, 71.0, 72.8, 74.5, 76.9, 79.2, 83.1\} \,.$$

The estimation of the GEV parameters $(\hat{c}_n, \hat{d}_n, \hat{\xi})$ follows the three steps outlined in Section A.1.1. The unbiased estimates of the first three PWMs $(\hat{b}_0, \hat{b}_1, \hat{b}_2)$ are computed from

the ordered sample $\boldsymbol{x}^{M_n}$:

$$\hat{b}_0 = \frac{1}{10}\sum_{j=1}^{10} x_{j:10} = 72.24\,,$$
$$\hat{b}_1 = \frac{1}{10}\sum_{j=1}^{10} \frac{j-1}{9}\, x_{j:10} \approx 38.06\,,$$
$$\hat{b}_2 = \frac{1}{10}\sum_{j=1}^{10} \frac{(j-1)(j-2)}{72}\, x_{j:10} \approx 26.04\,.$$

The sample L-moments $(\hat{\ell}_1, \hat{\ell}_2, \hat{\ell}_3)$ are calculated using the closed-form linear combinations of the PWMs:

$$\hat{\ell}_1 = \hat{b}_0 = 72.24\,,$$
$$\hat{\ell}_2 = 2\hat{b}_1 - \hat{b}_0 \approx 2(38.06) - 72.24 = 3.88\,,$$
$$\hat{\ell}_3 = 6\hat{b}_2 - 6\hat{b}_1 + \hat{b}_0 \approx 6(26.04) - 6(38.06) + 72.24 = 0.12\,.$$

The sample L-skewness ratio $\hat{t}_3$ is then:

$$\hat{t}_3 = \frac{\hat{\ell}_3}{\hat{\ell}_2} = \frac{0.12}{3.88} \approx 0.03\,.$$

The shape parameter $\hat{\xi}$ is obtained by solving the non-linear equation $\hat{t}_3 = t_3^{\mathrm{GEV}}(\xi)$. Using $\hat{t}_3 \approx 0.03$ and solving the GEV L-skewness function numerically yields the estimate:

$$\hat{\xi} \approx 0.23\,.$$

The scale parameter $\hat{d}_n$ is obtained using the closed-form solution (Eq. (A.1)):

$$\hat{d}_n = \frac{\hat{\ell}_2\hat{\xi}}{(2^{\hat{\xi}} - 1)\Gamma(1-\hat{\xi})} = \frac{3.88 \cdot 0.23}{(2^{0.23} - 1)\cdot 1.2} \approx 4.30\,.$$

The location parameter $\hat{c}_n$ is obtained using the closed-form solution (Eq. (A.2)):

$$\hat{c}_n = \hat{\ell}_1 - \frac{\hat{d}_n}{\hat{\xi}}(\Gamma(1-\hat{\xi}) - 1) = 72.24 - \frac{4.30}{0.23}(1.231 - 1) \approx 67.92\,.$$

The L-moments method provides the following GEV parameter estimates for the given sample:

$$(\hat{c}_n, \hat{d}_n, \hat{\xi}) \approx (67.92, 4.30, 0.23)\,.$$

A.2 Appendix of the risk analysis of heat waves

A.2.1 City-specific ranges of reasonable thresholds u_n

In this section, we provide an overview of the range of reasonable thresholds u_n for each city, based on the agreement between the empirical distribution of the heat wave likelihood—computed from all historical heat wave records—and the analytical likelihood distribution as given by Theorem 3.3. The results are illustrated in the left panel of each figure, where the x-axis represents a selection of candidate threshold values, expressed as quantiles of the daily temperature distribution, and the y-axis shows the mean squared error (MSE), which quantifies the distance between the empirical and analytical likelihood distributions. Lower values on the y-axis indicate a better agreement between the empirical data and the theoretical model.

For each city, a range of threshold quantiles corresponding to relatively low MSE values can be identified. These quantiles define a reasonable interval for u_n that can be adopted for practical analysis or serve as a theoretical reference in future studies. Notably, the model demonstrates high accuracy in characterizing heat wave likelihoods throughout this interval, indicating that the modeling performance remains robust across a broad range of threshold values, rather than being sensitive to the choice of a single optimal threshold.

Furthermore, in the right panel of each figure, we present the probability-probability (P–P) plot corresponding to a selected threshold u_n (from the identified interval) used in the main manuscript. This plot illustrates the agreement between the empirical distribution of the likelihood, constructed from historical heat wave data, and the analytical likelihood distribution based on the chosen u_n, thereby providing a further validation of our threshold selection.

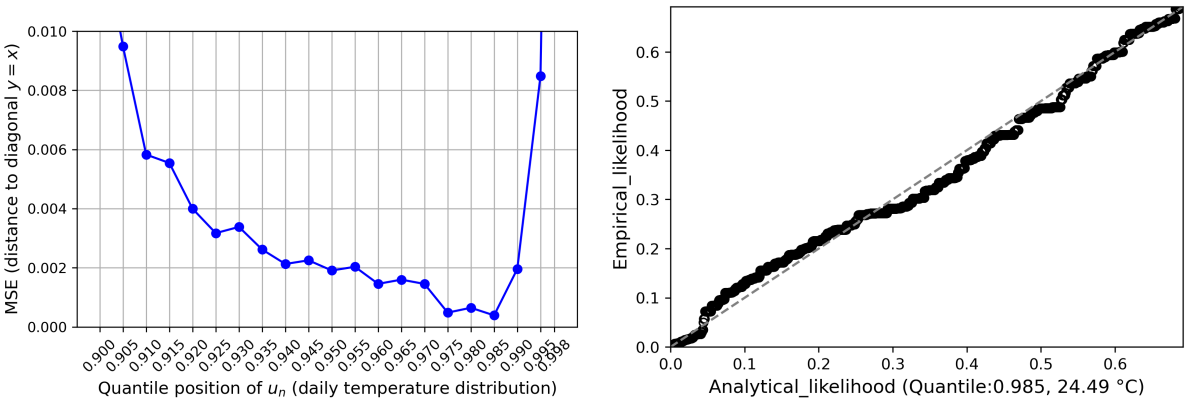


Figure A.1: Brussels

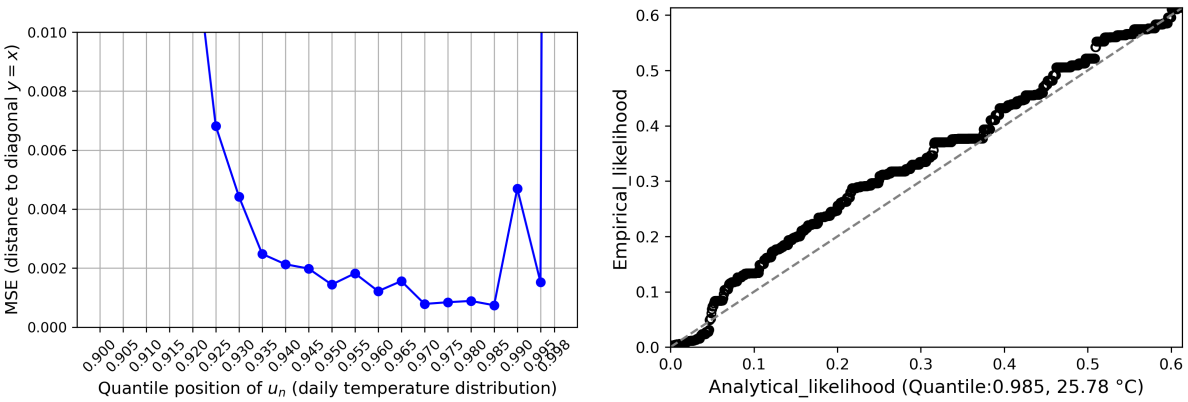


Figure A.2: Paris

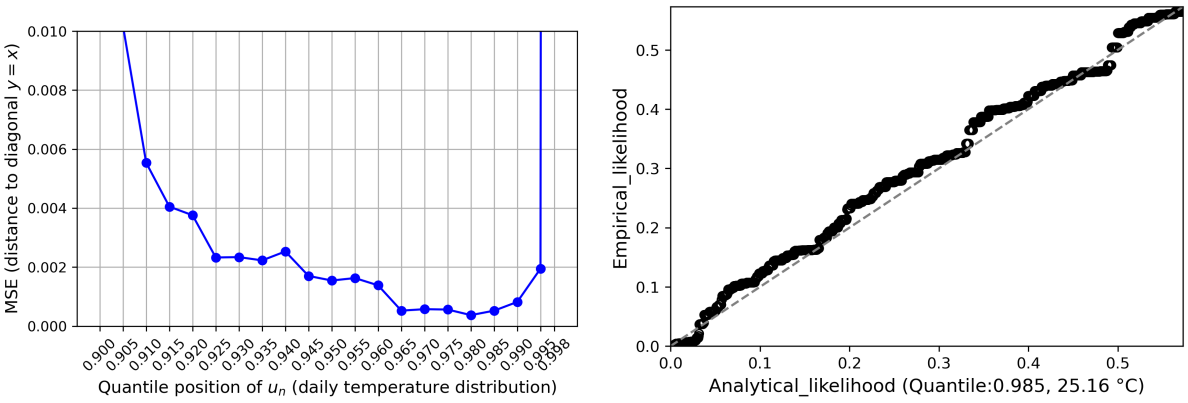


Figure A.3: Warsaw

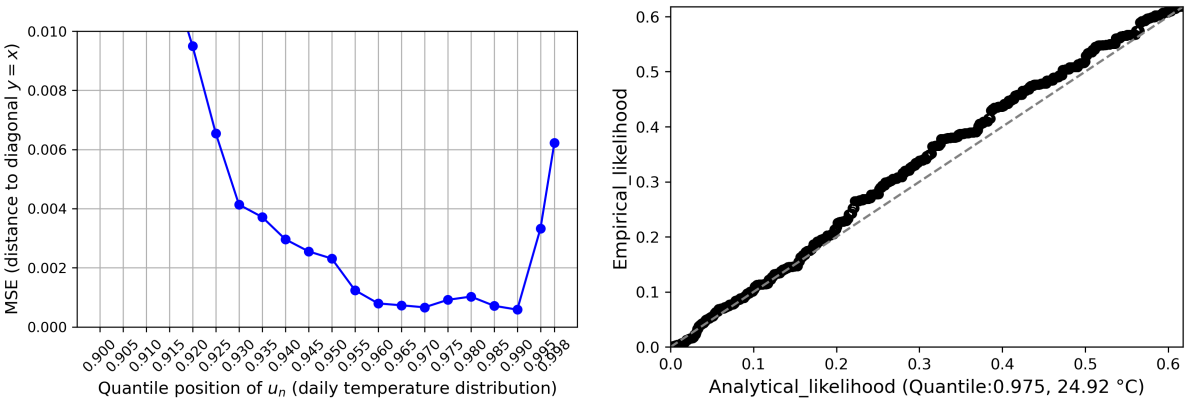


Figure A.4: Berlin

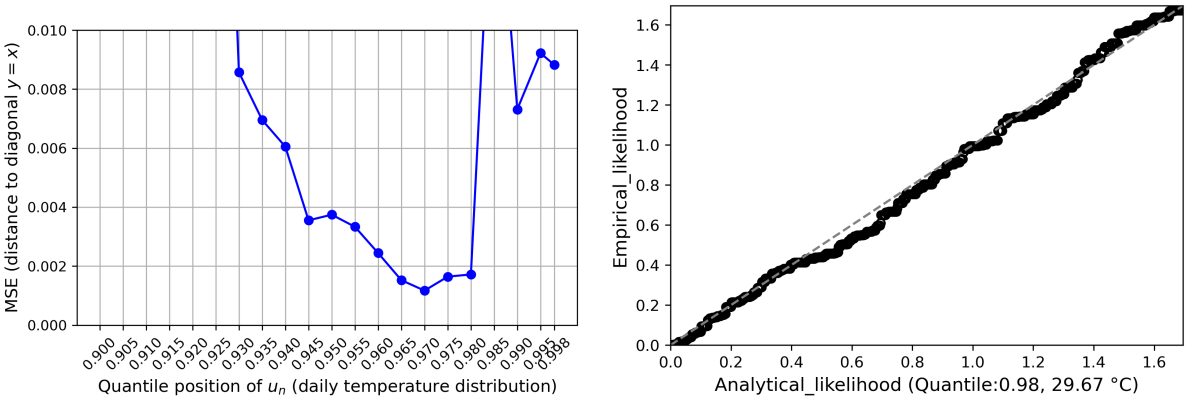


Figure A.5: Madrid

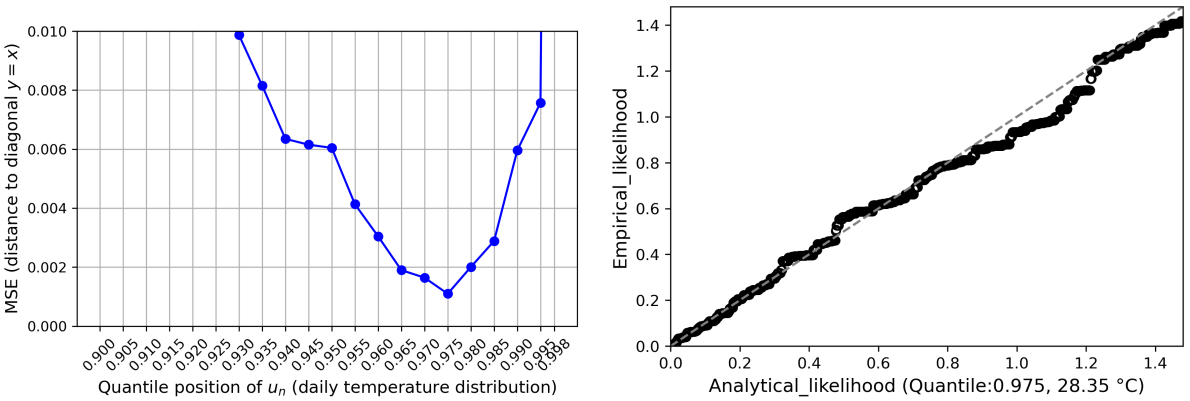


Figure A.6: Rome

A.2.2 Time series simulation via the stationary bootstrap method

To address the limited number of observed heat wave events in the Brussels historical temperature records, particularly for longer event durations, we generate synthetic daily temperature series through resampling. This augmentation is necessary to ensure sufficient empirical data for robust estimation of heat wave intensity distributions across various durations.

We use the stationary bootstrap method proposed by Politis and Romano (1994), which is designed to generate synthetic time series that preserve key features of the original data, including stationarity, marginal distribution, and short-range temporal dependence. Unlike block bootstrap methods with fixed block lengths, stationary bootstrap introduces randomness in both block location and length, enhancing its ability to mimic natural variability and reducing artifacts such as periodicity and boundary effects. In our context, the stationary bootstrap method is applied to the detrended daily summer temperature series. The step-by-step procedure is described below:

1. For each year from 1940 to 2024, the 92-day summer temperature sequence is extracted, resulting in an 85-row matrix of annual summer series.
2. The stationary bootstrap method is applied independently to each year's 92-day sequence to generate 1000 bootstrap samples per year, using randomly selected starting points and geometrically distributed block lengths (with parameter $p = \frac{1}{n^{1/3}}$).
3. For each bootstrap iteration i (where $i = 1, \ldots, 1000$), a time series realization is constructed by concatenating the i-th bootstrap sample from each of the 85 years in chronological order.
4. This process yields 1000 synthetic realizations, each consisting of 85 years $\times$ 92 days = 7820 daily values, maintaining both the year-wise seasonal structure and intra-seasonal dependence.

A.3 Appendix of the group anomaly detection

A.3.1 Derivation of the uninorm U_{ano} based on $U_{3\Pi}$

The most famous uninorm having the desired synergistic and antagonistic behavior is the 3Π operator $U_{3\Pi} : [0,1]^2 \to [0,1]$ with neutral element $\frac{1}{2}$:

$$U_{3\Pi}(a,b) = \frac{ab}{ab + (1-a)(1-b)}$$

with either $U_{3\Pi}(0,1) = U_{3\Pi}(1,0) = 0$ or $U_{3\Pi}(0,1) = U_{3\Pi}(1,0) = 1$.

Proposition A.1. Consider a uninorm U with neutral element $e \in]0,1[$. The operations $T_U, S_U : [0,1]^2 \to [0,1]$ defined by (Klement et al., 2013):

$$\begin{aligned} T_U(a,b) &= \frac{1}{e}U(ea, eb)\,, \\ S_U(a,b) &= \frac{1}{1-e}\Big(U\big(e + (1-e)a, e + (1-e)b\big) - e\Big)\,. \end{aligned} \tag{A.3}$$

are a t-norm and a t-conorm, respectively.

By Proposition A.1, we obtain the following t-norm and t-conorm induced by $U_{3\Pi}$:

$$\begin{aligned} T_{U_{3\pi}}(a,b) &= 2U_{3\pi}(\frac{a}{2}, \frac{b}{2})\,, \\ S_{U_{3\pi}}(a,b) &= 2U_{3\pi}\big(\frac{1}{2} + \frac{a}{2}, \frac{1}{2} + \frac{b}{2}\big) - 1\,, \end{aligned} \tag{A.4}$$

Proposition A.2. Consider a t-norm T, a t-conorm S and $e \in]0,1[$. The operation $U_{T,S,e} : [0,1]^2 \to [0,1]$ defined by (Klement et al., 2013):

$$U_{T,S,e}(a,b) = \begin{cases} eT(\frac{a}{e}, \frac{b}{e}) & , \text{ if } (a,b) \in [0,e]^2 \\ e + (1-e)S(\frac{a-e}{1-e}, \frac{b-e}{1-e}) & , \text{ if } (a,b) \in (e,1]^2 \\ \max(a,b) & , \text{ elsewhere.} \end{cases} \tag{A.5}$$

is a uninorm with neutral element e.

By Proposition A.2, the uninorm U_{ano} induced by $T_{U_{3\pi}}$ and $S_{U_{3\pi}}$:

$$U_{ano}(a,b) = \begin{cases} 2e\, U_{3\Pi}(\frac{a}{2e}, \frac{b}{2e}) & , \text{ if } (a,b) \in [0,e]^2 \\ e + (1-e)\Big(2U_{3\Pi}(\frac{1}{2} + \frac{a-e}{2-2e}, \frac{1}{2} + \frac{b-e}{2-2e}) - 1\Big) & , \text{ if } (a,b) \in]e,1]^2 \\ \max(a,b) & , \text{ elsewhere.} \end{cases} \tag{A.6}$$

A.3.2 Details of experiments

In this part, we present ROC curves to compare the performances of the various methods applied to the synthetic data and the SDSS data. Furthermore, we show boxplots of the AUC scores obtained in the different runs of the 10-fold cross-validation.

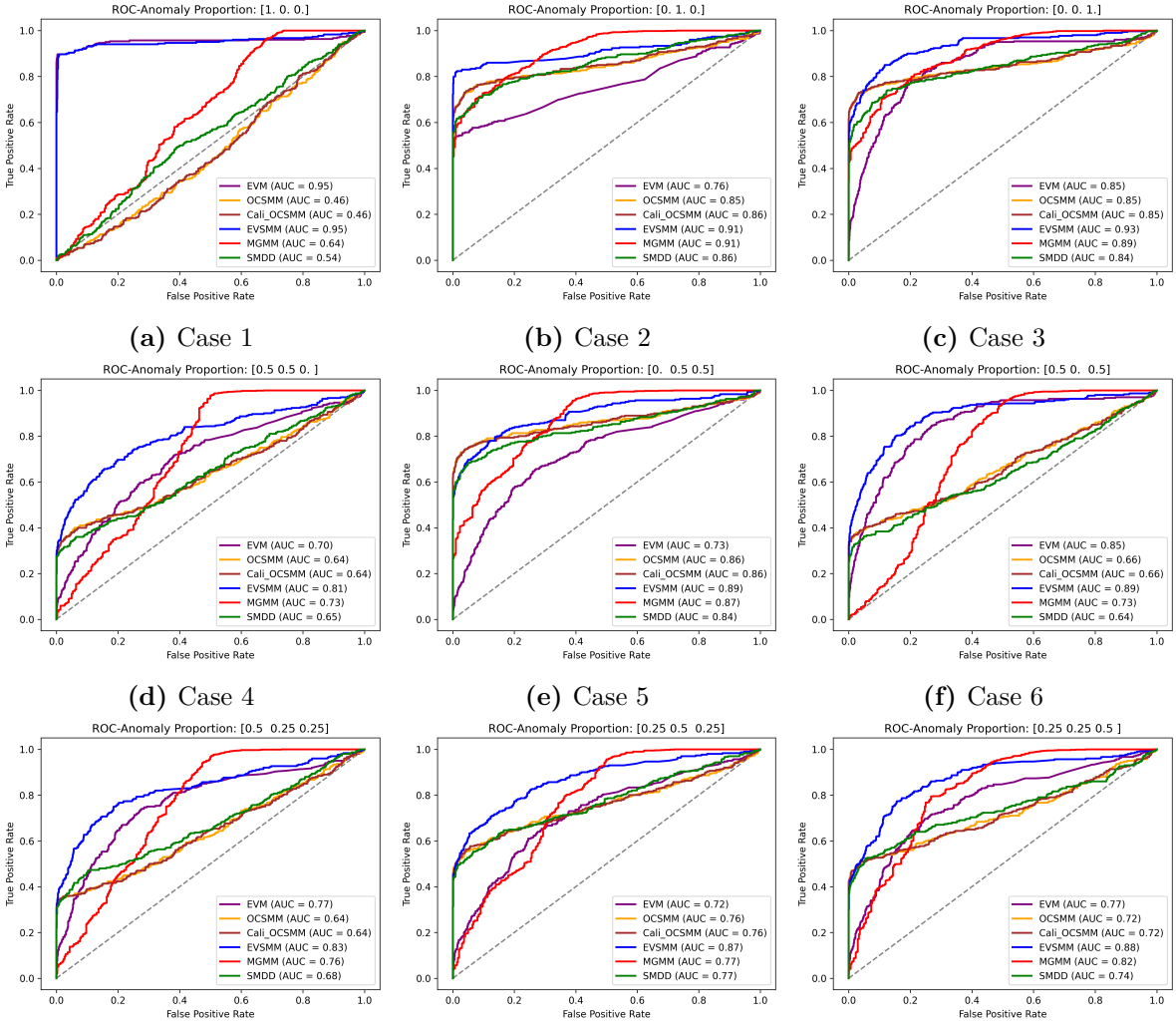
The receiver operating characteristic (ROC) plot is a widely used tool to evaluate the performance of a classifier. It shows the true positive rate (TPR) as a function of the false positive rate (FPR) for various classification thresholds. This provides a visual representation of the trade-offs between TPR (also known as sensitivity) and specificity. An ROC curve is a graphical evaluation tool that avoids the choice of a classification threshold which generally depends on the specific needs of an application. Based on the ROC curve, one can calculate the area under the curve (AUC) which is a performance metric of the classifier algorithm. The AUC score is equal to the probability that a randomly selected positive example is ranked higher by the classifier than a randomly selected negative example. A good classifier has an AUC close to 1.0, while a random classifier has an AUC near 0.5. The AUC score indicates the classifier's ability to distinguish between positive and negative classes, where higher scores indicate better performance.

The ROC plots shown in Figures A.7 and A.9 display the performance of our method compared to other methods for group anomaly detection on the synthetic data and the SDSS data. The ROC curves of our method, represented in blue, consistently demonstrate superior detection performance across all cases, as evidenced by their higher TPR and lower FPR compared to competing methods. The optimal thresholds of the models are determined by selecting the largest TPR while minimizing the FPR.

Figures A.8 and A.10 show boxplots of the AUC scores for the synthetic data and the SDSS data in the different runs of the 10-fold cross-validation experiments. The results underscore the reliable performance of our method EVSMM, as reflected by its consistently high AUC scores and narrow variability across all cases. Furthermore, the boxplots reveal

that EVSMM exhibits lower variance in AUC scores compared to other methods, such as MGMM, which also performs quite well in overall AUC scores shown previously in Figures A.7 and A.9. This demonstrates not only EVSMM's effectiveness in detecting anomalies but also its robustness and reliability compared to other methods.

The hyperparameters γ and e were tuned through a grid search strategy within a nested 10-fold cross-validation. For each fold, the grid search was employed to identify the best combination of hyperparameters γ and e that optimizes the performance of the proposed EVSMM. Across all cases, γ consistently resulted in 1.4 during the grid search process. Tables A.1 and A.2 explicitly summarize the values of e selected for each fold based on the best model performance identified through grid search, with Table A.1 showing the results for the synthetic data and Table A.2 presenting the results for the SDSS data. The final rows of the tables present the most frequently selected values as the representative choices across the 10 folds for each case, providing a useful reference for practitioners using EVSMM. It is noteworthy that while grid search identifies the value of e corresponding to the best model performance for each fold, the performance differences among various e values are remarkably small. This demonstrates that EVSMM achieves consistent performance across a range of hyperparameter configurations, reducing the dependency on precise tuning and further enhancing its practical reliability.

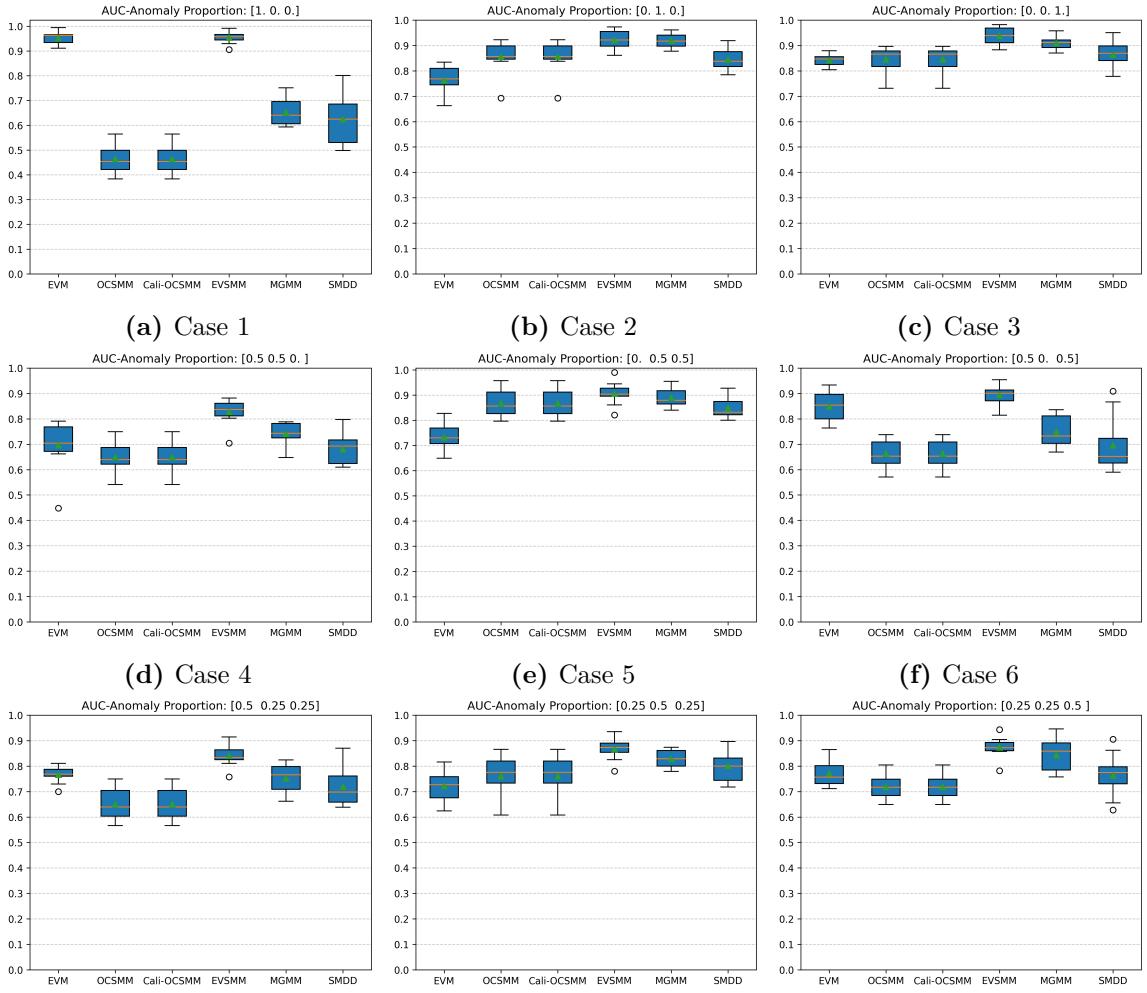


(a) Case 1 **(b)** Case 2 **(c)** Case 3

(d) Case 4 **(e)** Case 5 **(f)** Case 6

(g) Case 7 **(h)** Case 8 **(i)** Case 9

Figure A.7: The ROC curves of the experiments on synthetic data



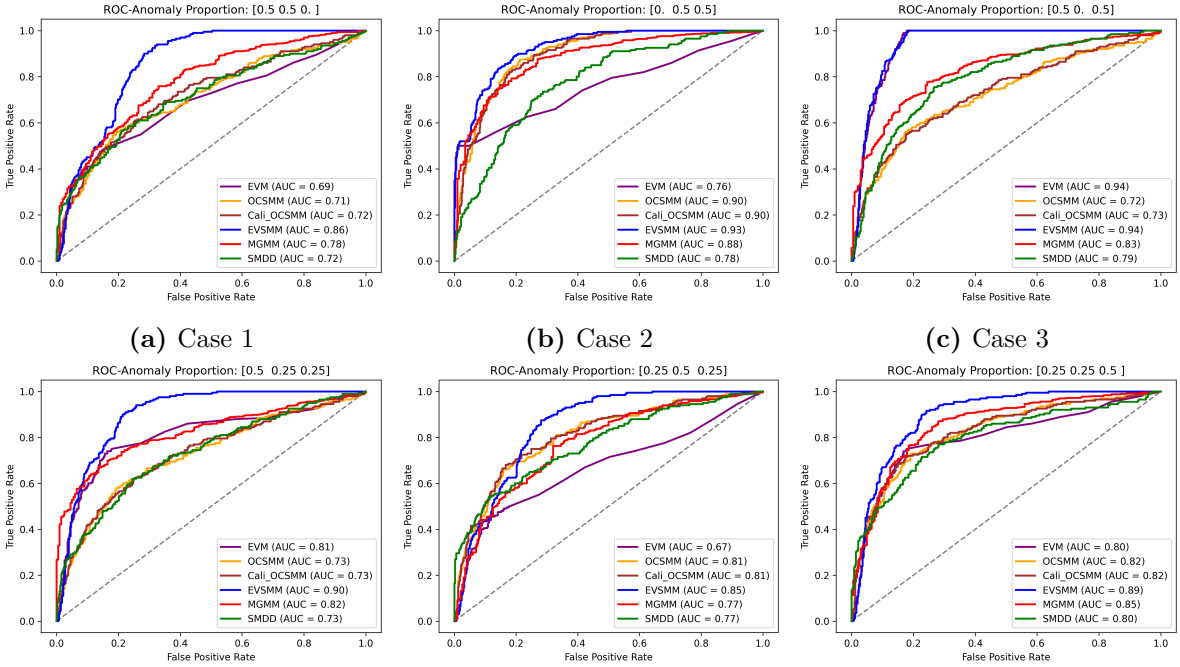
(a) Case 1 **(b)** Case 2 **(c)** Case 3

(d) Case 4 **(e)** Case 5 **(f)** Case 6

(g) Case 7 **(h)** Case 8 **(i)** Case 9

Figure A.8: Boxplots of the AUC scores across different runs of the experiments on synthetic data.

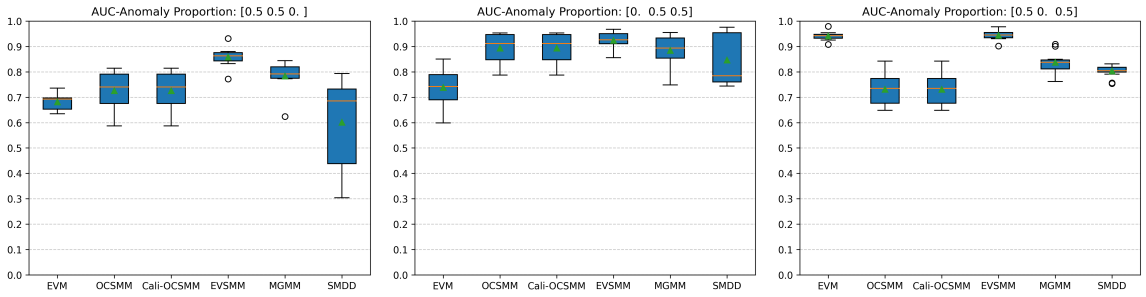
A



(a) Case 1 **(b)** Case 2 **(c)** Case 3

(d) Case 4 **(e)** Case 5 **(f)** Case 6

Figure A.9: ROC curves of the experiments on the SDSS data



(a) Case 1 **(b)** Case 2 **(c)** Case 3

(d) Case 4 **(e)** Case 5 **(f)** Case 6

Figure A.10: Boxplots of the AUC scores across different runs of the experiments on the SDSS data.

Table A.1: Hyperparameter e values across 10-fold cross-validation for synthetic data cases

Fold \ Case	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9
Fold 1	0.4	0.1	0.1	0.1	0.1	0.1	0.7	0.2	0.4
Fold 2	0.2	0.6	0.8	0.8	0.1	0.1	0.6	0.1	0.1
Fold 3	0.5	0.1	0.2	0.7	0.3	0.1	0.8	0.1	0.3
Fold 4	0.1	0.6	0.1	0.2	0.2	0.7	0.2	0.1	0.1
Fold 5	0.1	0.1	0.1	0.1	0.5	0.7	0.8	0.1	0.5
Fold 6	0.2	0.1	0.4	0.7	0.3	0.8	0.1	0.2	0.1
Fold 7	0.2	0.7	0.1	0.2	0.9	0.1	0.1	0.2	0.1
Fold 8	0.6	0.6	0.8	0.1	0.8	0.6	0.7	0.1	0.7
Fold 9	0.5	0.1	0.9	0.9	0.8	0.2	0.9	0.2	0.1
Fold 10	0.2	0.6	0.9	0.1	0.9	0.1	0.6	0.1	0.2
Optimal	0.2	0.1	0.1	0.1	0.3	0.1	0.7	0.1	0.1

Table A.2: Hyperparameter e values across 10-fold cross-validation for SDSS data cases

Fold \ Case	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Fold 1	0.5	0.1	0.9	0.9	0.5	0.1
Fold 2	0.7	0.1	0.9	0.9	0.9	0.9
Fold 3	0.9	0.2	0.9	0.9	0.9	0.1
Fold 4	0.6	0.1	0.8	0.6	0.9	0.9
Fold 5	0.8	0.4	0.9	0.9	0.1	0.9
Fold 6	0.6	0.7	0.3	0.9	0.9	0.1
Fold 7	0.9	0.1	0.1	0.9	0.9	0.1
Fold 8	0.1	0.1	0.1	0.1	0.1	0.1
Fold 9	0.3	0.1	0.9	0.9	0.4	0.9
Fold 10	0.8	0.1	0.9	0.9	0.7	0.9
Optimal	0.9	0.1	0.9	0.9	0.9	0.1

Bibliography

Met Office. What is a heatwave?, 2025. URL https://weather.metoffice.gov.uk/learn-about/weather/types-of-weather/temperature/heatwave. Accessed on 2025-09-22.

UK House of Commons Library. Heatwaves in the uk, July 2025. URL https://commonslibrary.parliament.uk/heatwaves-in-the-uk/. Accessed on 2025-09-22.

World Meteorological Organization. Heatwave, 2023. URL https://wmo.int/topics/heatwave. Accessed on 2025-09-22.

United Nations Office for Disaster Risk Reduction. Heatwave (mh0501) —hazard information profile, 2025. URL https://www.undrr.org/understanding-disaster-risk/terminology/hips/mh0501. Accessed on 2025-09-22.

Luca, S., Karsmakers, P., and Vanrumste, B. Anomaly detection using the poisson process limit for extremes. In *Proceedings of the 2014 IEEE International Conference on Data Mining*, pages 370–379, 2014.

Fisher, R. A. and Tippett, L. H. C. "Limiting forms of the frequency distribution of the largest or smallest member of a sample". *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190, 1928.

Gnedenko, B. "Sur la distribution limite du terme maximum d'une série aléatoire". *Annals of Mathematics*, 44(3):423–453, 1943.

Balkema, A. A. and de Haan, L. "Residual life time at great age". *The Annals of Probability*, 2 (5):792–804, 1974.

Pickands, J. "Statistical inference using extreme order statistics". *The Annals of Statistics*, 3 (1):119–131, 1975.

Coles, S. *An Introduction to Statistical Modeling of Extreme Values*. Springer, London, 2001.

de Haan, L. and Ferreira, A. *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006.

Embrechts, P., Klüppelberg, C., and Mikosch, T. *Modelling Extremal Events: for Insurance and Finance*. Springer Science & Business Media, New York, 1997.

Marra, F. Extreme value analyses and precipitation extremes. Lecture 1 presented at the University of Lausanne, 2024.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley & Sons, Chichester, 2008.

Daley, D. J. and Vere-Jones, D. *An Introduction to the Theory of Point Processes Volume I: Elementary Theory and Methods*. Springer Science & Business Media, New York, 2003.

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., and Williamson, R. C. "Estimating the support of a high-dimensional distribution". *Neural Computation*, 13(7):1443–1471, 2001.

Berlinet, A. and Thomas-Agnan, C. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer Science & Business Media, New York, 2011.

Smola, A., Gretton, A., Song, L., and Schölkopf, B. A hilbert space embedding for distributions. In *Proceedings of the International Conference on Algorithmic Learning Theory*, pages 13–31, 2007.

Grabisch, M. *Aggregation Functions*. Cambridge University Press, Cambridge, 2009.

Beliakov, G., Pradera, A., and Calvo, T. *Aggregation Functions: A Guide for Practitioners*. Springer-Verlag Berlin, Heidelberg, 2007.

Beliakov, G., James, S., Kolesárová, A., and Mesiar, R. "Cardinality-limiting extended pre-aggregation functions". *Information Fusion*, 76:66–74, 2021.

Yager, R. R. and Rybalov, A. "Uninorm aggregation operators". *Fuzzy Sets and Systems*, 80(1): 111–120, 1996.

Calvo, T., Mayor, G., and Mesiar, R. *Aggregation Operators: New Trends and Applications*, volume 97. Springer Science & Business Media, 2002.

Dombi, J. "Basic concepts for a theory of evaluation: the aggregative operator". *European Journal of Operational Research*, 10(3):282–293, 1982.

Buchanan, B. and Shortliffe, E. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Stanford University, Stanford, 1984.

Hájek, P. "Combining functions for certainty degrees in consulting systems". *International Journal of Man-Machine Studies*, 22(1):59–76, 1985.

Klement, E. P., Mesiar, R., and Pap, E. *Triangular Norms*, volume 8. Springer Science & Business Media, 2013.

Fodor, J. C., Yager, R. R., and Rybalov, A. "Structure of uninorms". *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 5(4):411–427, 1997.

Reiss, S. and Bogenberger, K. Gps-data analysis of munich's free-floating bike sharing system and application of an operator-based relocation strategy. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, pages 584–589. IEEE, 2015.

O'brien, O., Cheshire, J., and Batty, M. "Mining bicycle sharing data for generating insights into sustainable transport systems". *Journal of Transport Geography*, 34:262–273, 2014.

Zhu, B., Hu, S., Kaparias, I., Zhou, W., Ochieng, W., and Lee, D.-H. "Revealing the driving factors and mobility patterns of bike-sharing commuting demands for integrated public transport systems". *Sustainable Cities and Society*, 104:105323, 2024.

Duran-Rodas, D., Villeneuve, D., and Wulfhorst, G. "Bike-sharing: the good, the bad, and the future-an analysis of the public discussion on twitter-.". *European Journal of Transport & Infrastructure Research*, 20(4), 2020.

DeMaio, P. "Bike-sharing: History, impacts, models of provision, and future". *Journal of Public Transportation*, 12(4):41–56, 2009.

Shen, Y., Zhang, X., and Zhao, J. "Understanding the usage of dockless bike sharing in singapore". *International Journal of Sustainable Transportation*, 12(9):686–700, 2018.

Fishman, E. "Bikeshare: A review of recent literature". *Transport Reviews*, 36(1):92–113, 2016.

Du, Y., Deng, F., and Liao, F. "A model framework for discovering the spatio-temporal usage patterns of public free-floating bike-sharing system". *Transportation Research Part C: Emerging Technologies*, 103:39–55, 2019.

Zhuang, C., Li, S., Zhuang, H., and Liu, X. "Mining individual daily commuting patterns of dockless bike-sharing users: A two-layer framework integrating spatiotemporal flow clustering and rule-based decision trees". *Sustainable Cities and Society*, 118:105985, 2025.

Zhao, D. and Ong, G. P. "Geo-fenced parking spaces identification for free-floating bicycle sharing system". *Transportation Research Part A: Policy and Practice*, 148:49–63, 2021.

Liu, Y. and Tian, L. "A graded cluster system to mine virtual stations in free-floating bike-sharing system on multi-scale geographic view". *Journal of Cleaner Production*, 281:124692, 2021.

Lohrer, A., Binder, J. J., and Kröger, P. Group anomaly detection for spatio-temporal collective behaviour scenarios in smart cities. In *Proceedings of the 15th ACM SIGSPATIAL International Workshop on Computational Transportation Science*, pages 1–4, 2022.

Wei, Z., Ma, H., and Li, Y. "A multiscale approach for free-float bike-sharing electronic fence location planning: A case study of shenzhen city". *Journal of Advanced Transportation*, 2024 (1):1783038, 2024.

Münchner Verkehrsgesellschaft (MVG). Mvg rad-Geschäfts-und Rückgabegebiete / Freies Rückgabegebiet. URL `https://www.mvg.de/dam/jcr:1fd3dfde-19cb-4be0-ab54-9011d256b984/geschaeftsgebiet-mvg-rad.pdf`. Map, accessed on 2025-08-19.

Luo, H., Kou, Z., Zhao, F., and Cai, H. "Comparative life cycle assessment of station-based and dock-less bike sharing systems". *Resources, Conservation and Recycling*, 146:180–189, 2019.

Nikitas, A. "How to save bike-sharing: An evidence-based survival toolkit for policy-makers and mobility providers". *Sustainability*, 11(11):3206, 2019.

Cheng, L., Huang, J., Jin, T., Chen, W., Li, A., and Witlox, F. "Comparison of station-based and free-floating bikeshare systems as feeder modes to the metro". *Journal of Transport Geography*, 107:103545, 2023.

Hua, M., Chen, X., Zheng, S., Cheng, L., and Chen, J. "Estimating the parking demand of free-floating bike sharing: A journey-data-based study of Nanjing, China". *Journal of Cleaner Production*, 244:118764, 2020.

Wang, Y.-J., Kuo, Y.-H., Huang, G. Q., Gu, W., and Hu, Y. "Dynamic demand-driven bike station clustering. *Transportation Research Part E: Logistics and Transportation Review*, 160: 102656, 2022.

Zhang, Y., Lin, D., and Mi, Z. "Electric fence planning for dockless bike-sharing services". *Journal of Cleaner Production*, 206:383–393, 2019.

Chen, W., Liu, X., Chen, X., Cheng, L., and Chen, J. "Deciphering flow clusters from large-scale free-floating bike sharing journey data: a two-stage flow clustering method". *Transportation*, 52(1):155–184, 2025.

Jensen, P., Rouquier, J.-B., Ovtracht, N., and Robardet, C. "Characterizing the speed and paths of shared bicycle use in lyon". *Transportation Research Part D: Transport and Environment*, 15(8):522–524, 2010.

Sarkar, A., Lathia, N., and Mascolo, C. "Comparing cities' cycling patterns using online shared bicycle maps". *Transportation*, 42:541–559, 2015.

Jonkeren, O. and Kager, R. "Bicycle parking at train stations in the netherlands: Travellers' behaviour and policy options". *Research in Transportation Business & Management*, 40: 100581, 2021.

Ashqar, H. I., Elhenawy, M., Almannaa, M. H., Ghanem, A., Rakha, H. A., and House, L. Modeling bike availability in a bike-sharing system using machine learning. In *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pages 374–378. IEEE, 2017.

Caggiani, L., Camporeale, R., Ottomanelli, M., and Szeto, W. Y. "A modeling framework for the dynamic management of free-floating bike-sharing systems". *Transportation Research Part C: Emerging Technologies*, 87:159–182, 2018.

BMW Group. Building for the future: BMW group enters a new era of r&d with FIZ nord. `https://www.press.bmwgroup.com/global/article/detail/T0317727EN/building-for-the-future%3A-bmw-group-enters-a-new-era-of-r-d-with-fiz-nord?language=en`, September 2020. Press release, accessed on 2025-08-26.

Landeshauptstadt München. Neufreimann (ehemalige bayernkaserne). `https://stadt.muenchen.de/infos/neufreimann.html`, 2025. City planning project page, accessed on 2025-08-26.

Münchner Verkehrsgesellschaft (MVG). U-umgebungsplan moosach. `https://www.mvg.de/aushangfahrplan/P8_H_MO_0.pdf`, February 2024. Accessed on 2025-08-26.

Ballester, J., Quijal-Zamorano, M., Méndez Turrubiates, R. F., Pegenaute, F., Herrmann, F. R., Robine, J. M., Basagaña, X., Tonne, C., Antó, J. M., and Achebak, H. "Heat-related mortality in Europe during the summer of 2022". *Nature Medicine*, 29(7):1857–1866, 2023.

Thompson, V., Mitchell, D., Hegerl, G. C., Collins, M., Leach, N. J., and Slingo, J. M. "The most at-risk regions in the world for high-impact heatwaves". *Nature Communications*, 14(1): 2152, 2023.

Thompson, V., Kennedy-Asser, A. T., Vosper, E., Lo, Y. E., Huntingford, C., Andrews, O., Collins, M., Hegerl, G. C., and Mitchell, D. "The 2021 western North America heat wave among the most extreme events ever recorded globally". *Science Advances*, 8(18):eabm6860, 2022.

McKinnon, K. A. and Simpson, I. R. "How unexpected was the 2021 pacific northwest heatwave?". *Geophysical Research Letters*, 49(18):e2022GL100380, 2022.

Lee, H., Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P., Trisos, C., Romero, J., Aldunce, P., Barret, K., et al. Climate change 2023: Synthesis report. contribution of working groups I, II and III to the sixth assessment report of the intergovernmental panel on climate change. Technical report, Intergovernmental Panel on Climate Change (IPCC), Geneva, Switzerland, 2023.

Meehl, G. A. and Tebaldi, C. "More intense, more frequent, and longer lasting heat waves in the 21st century". *Science*, 305(5686):994–997, 2004.

Stefanon, M., D' Andrea, F., and Drobinski, P. "Heatwave classification over europe and the mediterranean region". *Environmental Research Letters*, 7(1):014023, 2012.

Schoetter, R., Cattiaux, J., and Douville, H. "Changes of western European heat wave characteristics projected by the cmip5 ensemble". *Climate Dynamics*, 45(5):1601–1616, 2015.

Ren, Y., Liu, J., Zhang, T., Shalamzari, M. J., Arshad, A., Liu, T., Willems, P., Gao, H., Tao, H., and Wang, T. "Identification and analysis of heatwave events considering temporal continuity and spatial dynamics". *Remote Sensing*, 15(5):1369, 2023.

Awasthi, A., Vishwakarma, K., and Pattnayak, K. "Retrospection of heatwave and heat index". *Theoretical and Applied Climatology*, 147:589–604, 2022.

Mazdiyasni, O., Sadegh, M., Chiang, F., and AghaKouchak, A. "Heat wave intensity duration frequency curve: A multivariate approach for hazard and attribution analysis". *Scientific Reports*, 9(1):14117, 2019.

Cebrián, A. C., Asín, J., Gelfand, A. E., Schliep, E. M., Castillo-Mateo, J., Beamonte, M. A., and Abaurrea, J. "Spatio-temporal analysis of the extent of an extreme heat event". *Stochastic Environmental Research and Risk Assessment*, 36(9):2737–2751, 2022.

Schliep, E. M., Gelfand, A. E., Abaurrea, J., Asín, J., Beamonte, M. A., and Cebrián, A. C. "Long-term spatial modelling for characteristics of extreme heat events". *Journal of the Royal Statistical Society Series A: Statistics in Society*, 184(3):1070–1092, 2021.

Mazdiyasni, O., AghaKouchak, A., Davis, S. J., Madadgar, S., Mehran, A., Ragno, E., Sadegh, M., Sengupta, A., Ghosh, S., Dhanya, C., and Niknejad, M. "Increasing probability of mortality during indian heat waves". *Science Advances*, 3(6):e1700066, 2017.

Perkins, S. E. and Alexander, L. V. "On the measurement of heat waves". *Journal of climate*, 26(13):4500–4517, 2013.

Russo, S., Sillmann, J., and Fischer, E. M. "Top ten European heatwaves since 1950 and their occurrence in the coming decades". *Environmental Research Letters*, 10(12):124003, 2015.

Becker, F., Fink, A., Bissolli, P., and Pinto, J. "Towards a more comprehensive assessment of the intensity of historical European heat waves (1979–2019)". *Atmospheric Science Letters*, 23(11), 2022.

Manning, C., Widmann, M., Bevacqua, E., Van Loon, A. F., Maraun, D., and Vrac, M. "Increased probability of compound long-duration dry and hot events in europe during summer (1950–2013)". *Environmental Research Letters*, 14(9):094006, 2019.

Davison, A. and Huser, R. "Statistics of extremes". *Annual Review of Statistics and Its Application*, 2(1):203–235, 2015.

Wehner, M., Gleckler, P., and Lee, J. "Characterization of long period return values of extreme daily temperature and precipitation in the CMIP6 models: Part 1, model evaluation". *Weather and Climate Extremes*, 30:100283, 2020.

Barbosa, S. and Scotto, M. G. "Extreme heat events in the iberia peninsula from extreme value mixture modeling of ERA5-land air temperature". *Weather and Climate Extremes*, 36:100448, 2022.

Furrer, E. M., Katz, R. W., Walter, M. D., and Furrer, R. "Statistical modeling of hot spells and heat waves". *Climate Research*, 43(3):191–205, 2010.

Photiadou, C., Jones, M., Keellings, D., and Dewes, C. "Modeling European hot spells using extreme value analysis". *Climate Research*, 58(3):193–207, 2014.

Kharin, V., Flato, G., Zhang, X., Gillett, N., Zwiers, F., and Anderson, K. "Risks from climate extremes change differently from 1.5°C to 2.0°C depending on rarity". *Earth's Future*, 6(5): 704–715, 2018.

Klein Tank, A. and Können, G. "Trends in indices of daily temperature and precipitation extremes in europe, 1946–99". *Journal of climate*, 16(22):3665–3680, 2003.

Haddad, K. "Selection of the best fit probability distributions for temperature data and the use of L-moment ratio diagram method: a case study for NSW in australia". *Theoretical & Applied Climatology*, 143:1261–1284, 2021.

Twardosz, R., Walanus, A., and Guzik, I. "Warming in europe: recent trends in annual and seasonal temperatures". *Pure and Applied Geophysics*, 178(10):4021–4032, 2021.

DelSole, T. and Tippett, M. K. "Comparing climate time series–Part 2: A multivariate test". *Advances in Statistical Climatology, Meteorology and Oceanography*, 7(2):73–85, 2021.

Weilnhammer, V., Schmid, J., Mittermeier, I., Schreiber, F., Jiang, L., Pastuhovic, V., Herr, C., and Heinze, S. "Extreme weather events in Europe and their health consequences –A systematic review". *International Journal of Hygiene and Environmental Health*, 233, 2021.

Brás, T., Seixas, J., Carvalhais, N., and Jägermeyr, J. "Severity of drought and heatwave crop losses tripled over the last five decades in Europe". *Environmental Research Letters*, 16(6), 2021.

García-Herrera, R., Díaz, J., Trigo, R., Luterbacher, J., and Fischer, E. "a review of the European summer heat wave of 2003". *Critical Reviews in Environmental Science and Technology*, 40(4):267–306, 2010.

Lhotka, O. and Kyselý, J. "The 2021 European heat wave in the context of past major heat waves". *Earth and Space Science*, 9(11):e2022EA002567, 2022.

Kyselý, J. "Recent severe heat waves in central Europe: how to view them in a long-term prospect?". *International Journal of Climatology*, 30(1):89–109, 2010.

Twardosz, R. and Batko, A. "Heat waves in Central Europe (1991–2006)". *International Journal of Global Warming*, 4(3-4):261–272, 2012.

Ionita, M., Tallaksen, L., Kingston, D., Stagge, J., Laaha, G., Van Lanen, H., Scholz, P., Chelcea, S., and Haslinger, K. "The European 2015 drought from a climatological perspective". *Hydrology and Earth System Sciences*, 21(3):1397–1419, 2017.

Xu, Z., FitzGerald, G., Guo, Y., Jalaludin, B., and Tong, S. "Impact of heatwave on mortality under different heatwave definitions: A systematic review and meta-analysis". *Environment International*, 89:193–203, 2016.

Lee, W., McCormick, T. H., Neil, J., Sodja, C., and Cui, Y. "Anomaly detection in large-scale networks with latent space models". *Technometrics*, 64(2):241–252, 2022.

Luca, S., Pimentel, M. A., Watkinson, P. J., and Clifton, D. "Point process models for novelty detection on spatial point patterns and their extremes". *Computational Statistics & Data Analysis*, 125:86–103, 2018.

B

Jin, H., Yin, G., Yuan, B., and Jiang, F. "Bayesian hierarchical model for change point detection in multivariate sequences". *Technometrics*, 64(2):177–186, 2022.

Chandola, V., Banerjee, A., and Kumar, V. "Anomaly detection: A survey". *ACM Computing Surveys (CSUR)*, 41(3):1–58, 2009.

Xiong, L., Póczos, B., Schneider, J., Connolly, A., and VanderPlas, J. Hierarchical probabilistic models for group anomaly detection. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, pages 789–797, 2011a.

Muandet, K. and Schölkopf, B. One-class support measure machines for group anomaly detection. In *Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence*, pages 449–458, 2013.

Chalapathy, R., Toth, E., and Chawla, S. Group anomaly detection using deep generative models. In *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 173–189, 2018.

Song, W., Dong, W., and Kang, L. "Group anomaly detection based on bayesian framework with genetic algorithm". *Information Sciences*, 533:138–149, 2020.

Kasieczka, G., Nachman, B., and Shih, D. New methods and datasets for group anomaly detection from fundamental physics. In *Proceedings of the 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1–11, 2021.

Guevara, J., Canu, S., and Hirata, R. Support measure data description for group anomaly detection. In *ODDx3 Workshop on Outlier Definition, Detection, and Description at the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015.

Xiong, L., Póczos, B., and Schneider, J. Group anomaly detection using flexible genre models. In *Advances in Neural Information Processing Systems*, pages 1071–1079, 2011b.

Tarassenko, L., Hayton, P., Cerneaz, N., and Brady, M. Novelty detection for the identification of masses in mammograms. In *Proceedings of the 4th International Conference on Artificial Neural Networks*, pages 442–447, 1995.

Dietterich, T. G. Machine learning for sequential data: A review. In *Proceedings of the Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition and Structural and Syntactic Pattern Recognition*, pages 15–30, 2002.

Hazel, G. G. "Multivariate Gaussian MRF for multispectral scene segmentation and anomaly detection". *IEEE Transactions on Geoscience and Remote Sensing*, 38(3):1199–1211, 2000.

Das, K., Schneider, J., and Neill, D. B. Anomaly pattern detection in categorical datasets. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 169–176, 2008.

Belhadi, A., Djenouri, Y., Srivastava, G., Cano, A., and Lin, J. C.-W. "Hybrid group anomaly detection for sequence data: Application to trajectory data analytics". *IEEE Transactions on Intelligent Transportation Systems*, 23(7):9346–9357, 2022.

Mohod, P. and Janeja, V. P. "Density-based spatial anomalous window discovery". *International Journal of Data Warehousing and Mining*, 18(1):1–23, 2022.

Das, K., Schneider, J., and Neill, D. B. *Detecting Anomalous Groups in Categorical Datasets*. Carnegie Mellon University, Pittsburgh, 2009.

Pehlivanian, C. A. and Neill, D. B. "Efficient optimization of partition scan statistics via the consecutive partitions property". *Journal of Computational and Graphical Statistics*, 32(2): 712–729, 2023.

Tax, D. M. and Duin, R. P. "Support vector data description". *Machine Learning*, 54:45–66, 2004.

Ting, K., Xu, B., Washio, T., and Zhou, Z. "Isolation distributional kernel: a new tool for point and group anomaly detection". *IEEE Transactions on Knowledge and Data Engineering*, 35 (3):2697–2710, 2023.

Fisch, A. T., Eckley, I. A., and Fearnhead, P. "A linear time method for the detection of collective and point anomalies". *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 15(4):494–508, 2022.

Yu, R., He, X., and Liu, Y. "Glad: group anomaly detection in social media analysis". *ACM Transactions on Knowledge Discovery from Data*, 10(2):1–22, 2015.

Blei, D. M., Ng, A. Y., and Jordan, M. I. "Latent dirichlet allocation". *Journal of Machine Learning Research*, 3:993–1022, 2003.

Luca, S., Clifton, D. A., and Vanrumste, B. "One-class classification of point patterns of extremes". *Journal of Machine Learning Research*, 17(191):1–21, 2016.

Clifton, D. A., Hugueny, S., and Tarassenko, L. "Novelty detection with multivariate extreme value statistics". *Journal of Signal Processing Systems*, 65(3):371–389, 2011.

Soong, T. T. *Fundamentals of Probability and Statistics for Engineers*. John Wiley & Sons, Chichester, 2004.

Schölkopf, B. and Smola, A. J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT press, 2002.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge, 2007.

Lin, H. T., Lin, C. J., and Weng, R. C. "A note on Platt' s probabilistic outputs for support vector machines". *Machine Learning*, 68:267–276, 2007.

Clifton, L., Clifton, D. A., Zhang, Y., Watkinson, P., Tarassenko, L., and Yin, H. "Probabilistic novelty detection with support vector machines". *IEEE Transactions on Reliability*, 63(2): 455–467, 2014.

Tax, D. M. and Duin, R. P. "Uniform object generation for optimizing one-class classifiers". *Journal of Machine Learning Research*, 2:155–173, 2001.

Markou, M. and Singh, S. "A neural network-based novelty detector for image sequence analysis". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1664–1677, 2006.

Duda, R. O., Hart, P. E., and Nilsson, N. J. Subjective Bayesian methods for rule-based inference systems. In *Proceedings of the June 7-10, 1976, National Computer Conference and Exposition*, pages 1075–1082, 1976.

De Baets, B. and Fodor, J. "Van Melle's combining function in MYCIN is a representable uninorm: An alternative proof". *Fuzzy Sets and Systems*, 104(1):133–136, 1999.

Póczos, B., Xiong, L., and Schneider, J. Nonparametric divergence estimation with applications to machine learning on distributions. In *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence*, pages 599–608, 2011.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R. "Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form". *Water Resources Research*, 15(5):1049–1054, 1979.

Hosking, J. R. "L-moments: analysis and estimation of distributions using linear combinations of order statistics". *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 52(1):105–124, 1990.