ABSTRACT

Title of Dissertation: DEVELOPMENT AND APPLICATION OF PROPINQUITY MODELING FRAMEWORK FOR IDENTIFICATION AND ANALYSIS OF EXTREME EVENT PATTERNS

Vitaly Kholodovsky
Doctor of Philosophy, 2023

Dissertation Directed by: Professor Xin-Zhong Liang
Department of Atmospheric and Oceanic Science

Extreme weather and climate events such as floods, droughts, and heat waves can cause extensive societal damage. While various statistical and climate models have been developed for the purpose of simulating extremes, a consistent definition of extreme events is still lacking. Furthermore, to better assess the performance of the climate models, a variety of spatial forecast verification measures have been developed. However, in most cases, the spatial verification measures that are widely used to compare mean states do not have sufficient theoretical justification to benchmark extreme events. In order to alleviate inconsistencies when defining extreme events within different scientific communities, we propose a new generalized Spatio-Temporal Threshold Clustering method for the identification of extreme event episodes, which uses machine learning techniques to couple existing pattern recognition indices with high or low threshold choices. The method consists of five main steps: construction of essential field quantities, dimension reduction, spatial domain mapping, time series clustering, and threshold selection. We develop and apply this method using a gridded daily precipitation dataset derived from rain gauge stations over the contiguous United States. We observe changes in the distribution of conditional frequency of extreme precipitation from large-scale, well-connected spatial patterns to smaller-scale, more isolated rainfall clusters, possibly leading to more localized droughts and heatwaves, especially during the summer months.

Additionally, we compare empirical and statistical probabilities and intensities obtained through the Conventional Location Specific methods, which are deficient in geometric interconnectivity between individual spatial pixels and independent in time, with a new Propinquity modeling framework. We integrate the Spatio-Temporal Threshold Clustering algorithm and the conditional semi-parametric Heffernan and Tawn (2004) model into the Propinquity modeling framework to separate classes of models that can calculate process level dependence of large-scale extreme processes, primarily through the overall extreme spatial field. Our findings reveal significant differences between Propinquity and Conventional Location Specific methods, in both empirical and statistical approaches in shape and trend direction. We also find that the process of aggregating model results without considering interconnectivity between individual grid cells for trend construction can lead to significant variations in the overall trend pattern and direction compared with models that do account for interconnectivity. Based on these results, we recommend avoiding such practices and instead adopting the Propinquity modeling framework or other spatial EVA models that take into account the interconnectivity between individual grid cells.

Our aim for the final application is to establish a connection between extreme essential field quantity intensity fields and large-scale circulation patterns. However, the Conventional Location Specific Threshold methods are not appropriate for this purpose as they are memoryless in time and not able to identify individual extreme episodes. To overcome this, we developed the Feature Finding Decomposition algorithm and used it in combination with the Propinquity modeling framework. The algorithm consists of the following three steps: feature finding, image decomposition, and large-scale circulation patterns connection. Our findings suggest that the Western Pacific Index, particularly its 5th percentile and 5th mode of decomposition, is the most significant teleconnection pattern that explains the variation in the trend pattern of the largest feature intensity.

DEVELOPMENT AND APPLICATION OF PROPINQUITY MODELING FRAMEWORK FOR IDENTIFICATION AND ANALYSIS OF EXTREME EVENT PATTERNS

by

Vitaly Kholodovsky

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:
Professor Xin-Zhong Liang, Chair/Advisor
Project Scientist Dr. Eric Gilleland, Co-Advisor
Professor James Carton
Assistant Professor Maria Molina
Assistant Professor Jonathan Poterjoy
Senior Scientist Thomas R. Knutson
Professor Michael N. Evans, Dean's Representative

There will always be one (or more)

value that will exceed all others.

Emil J. Gumbel

Dedication

This doctoral dissertation is dedicated to my parents, Luba and Anatoly; my wife, Shaylem; my advisor, Professor Xin-Zhong Liang; my co-advisor, Dr. Eric Gilleland; and other scientists who work tirelessly to discover secrets of the universe. This would not be possible without your support, encouragement, patience, and understanding.

Acknowledgments

First and foremost, I'd like to thank my advisor, Professor Xin-Zhong Liang, for giving me an opportunity to work in his group. I also would like to thank him for his support, patience, and understanding throughout my studies. Our discussions and debates have been enlightening, and I have learned a lot from them. It has been an honor to learn and collaborate with such an exceptional scholar and person.

I also want to thank Dr. Eric Gilliland, my co-advisor, for his time and efforts in teaching me about extreme value theory and spatial forecast verification. I am grateful for his hosting me at NCAR through their ASP program.

Special thanks to my mentor, Mr. Tom Knutson, for hosting me at GFDL and spending his valuable time teaching me about his approach to the detection and attribution of extreme events.

I would like to thank all my teammates at ESSIC, classmates, and teachers in the AOSC department for their willingness to share their knowledge with me. Many thanks to Dr. Chao Sun and Dr. Sijal Dangol for their assistance with data and engaging discussions.

This research thesis was supported by the National Oceanic and Atmospheric Administration, Educational Partnership Program with Minority-Serving Institution, U.S. Department of Commerce, under agreement no. NA16SEC4810006. Additional support came from NSF Innovations at the Nexus of Food, Energy and

Water Systems under grant nos. EAR1639327 and EAR1903249 as well as from NRT-INFEWS: UMD Global STEWARDS (STEM Training at the Nexus of Energy, WAter Reuse and FooD Systems) grant no. 1828910.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AO	Arctic Oscillation
BHM	Bayesian Hierarchical Model
BIC	Bayesian Information Criterion
CESM	Community Earth System Model
CLST	Conventional Location Specific Threshold
CONUS	Contiguous United States
CPC	NOAA Climate Prediction Center
CRU	Climatic Research Unit, University of East Anglia
df	Distribution Function
DJF	Winter
EA	Eastern Atlantic
EFQ	Essential Field Quantity
EP	East Pacific
EVA	Extreme Value Analysis
EVD	Extreme Value Distribution
EVT	Extreme Value Theory
GEV	Generalized Extreme Value
GHCN	Global Historical Climatology Network
GP	Generalized Pareto
HT2004	Heffernan and Tawn (2004)
HT	Heffernan and Tawn
ICP	Spatial Forecast Verification Intercomparison Project
IMA	Integrative Modeling Approach
IMF	Intrinsic Mode Function
IMFx	Intrinsic Mode Function, where x is the mode number from 1 to 5
IPCC	Intergovernmental Panel on Climate Change

JISAO	Joint Institute for the Study of the Atmosphere and Ocean
JJA	Summer
LDA	Linear Discriminant Analysis
MAM	Spring
MCMC	Markov Chain Monte Carlo
MED	Mean Error Distance
MLE	Maximum-Likelihood Estimation
MSE	Mean Square Error
NAO	North Atlantic Oscillation
NEF	Negative Extreme Field
NCEI	NOAA National Centers for Environmental Information
Niño 3	Eastern Tropical Pacific SST
Niño 3.4	East Central Tropical Pacific SST
NP	North Pacific Oscillation
ONI	Oceanic Niño Index
PAM	Partitioning Around Medoids
PCA	Principal Component Analysis
PDO	Pacific Decadal Oscillation
PEF	Positive Extreme Field
PNA	Pacific North American Index
POT	Peaks Over Threshold
PP	Point Process
PQ	Propinquity
PRISM	Parameter-Elevation Regression on Independent Slopes Model
RMSE	Root Mean Square Error
SOI	Southern Oscillation Index
SON	Fall
STTC	Spatio-Temporal Threshold Clustering
TNI	Trans-Niño Index
XGBoost	Extreme Gradient Boosting algorithm

VGAM	Vector Generalized Additive Model
VGLM	Vector Generalized Linear Model
WMO	World Meteorological Organization
WP	Western Pacific Index
WR	Western Russia

Chapter 1: A generalized Spatio-Temporal Threshold Clustering method for identification of extreme event patterns

Extreme events of essential climate variables (Bojinski et al., 2014), such as heavy rainfall, high temperatures, strong winds, and their derivatives, such as droughts and heat waves, are a major source of risk to society and the environment. The purpose of risk management is to design and implement a set of common procedures for predicting, measuring, and managing such high-impact events. The critical step is identifying these extreme processes. However, there is no uniformly accepted definition of extreme events. More often than not, the process of identifying extreme events has been typically based on classifying them into groups with characteristics such as frequency of occurrence, intensity, temporal duration, and timing (Stephenson, 2008). In addition, the evaluation of the multidimensional nature of extreme events, in most cases, has been confined to individual grid points or stations, thus overlooking embedded spatial dependency. The Intergovernmental Panel on Climate Change (IPCC) defines a climate extreme as "the occurrence of a value of a weather or climate variable above (or below) a threshold value near the upper (or lower) ends of the range of observed values of the variable" (IPCC, 2012). Typically, the choice of a high (or low) threshold depends on the conventions of specific scientific disciplines. For example, the climate science community tends to select location-specific thresholds based on either categories (e.g., Dai, 2001; Dai et al., 2017; Sun et al., 2007), quantiles (e.g., Groisman et al., 2005; Lau et al., 2013; Pendergrass and Hartmann, 2014), combinations (e.g., Kunkel et al., 2010, 2012, 2003, 2007), or indices (e.g., Alexander et al., 2006; Donat et al., 2013a,b; Zhang et al., 2011). Hereafter, we refer to these selection methods as Conventional Location Specific Threshold (CLST) methods.

In the realm of statistical modeling, under appropriate conditions, excesses above (below) a high (low) threshold are often modeled using the generalized Pareto (GP) distribution (Balkema and de Haan, 1974; Pickands, 1975), where requirements for a chosen threshold is to be high (low) enough to satisfy chosen goodness of fit test. Although threshold choices often overlap across disciplines, it is still difficult to compare extreme value analysis results between different scientific studies if the threshold selection methodologies differ. Even when the same forecast verification strategy is used (e.g., Jolliffe and Stephenson, 2003; Wilks, 2011, Chapter 8), the results will be different. Thus, a unified approach for threshold selection (i.e., universal extreme event definition) is essential because it standardizes the inference process for extreme event analysis, leading to further consistency and transparency when comparing results between different scientific findings. Unification, standardization, and transparency are the key pillars of disciplined risk management.

While many univariate methods have been proposed to automate the threshold choice (e.g., Bader et al., 2017; Fukutome et al., 2015; Scarrott and MacDonald, 2012), there are no clear-cut criteria, and in practice, threshold choice tends to be determined by either exploration methods or by stability assessments of parameter estimates, and in most cases, the threshold selection is disconnected from the geometrical properties of the desired spatial field. Here, we express geometrical properties by geometric attributes that measure the texture/dispersiveness of the spatial field. Ultimately, the choice of threshold is always an interplay between bias and variance (Smith, 1987). For high extreme thresholds, if the chosen threshold is not high enough, the GP distribution likely will not have a good fit to the excesses above the threshold, leading to approximation bias. Conversely, if the chosen threshold is too high, only a small number of exceedances will be generated, and consequently, there will be high variance in the estimators (in the case of low extreme thresholds, these conditions are reversed). A more comprehensive evaluation of univariate extreme value theory (EVT) is given in (Coles, 2001a).

Analysis of extreme events must also incorporate the spatial nature of the key climate variables since many single quantities, such as precipitation or temperature, are measured at multiple locations that may also be teleconnected via atmospheric circulations or hydrological cycles. Implementation of spatial extreme value analysis has to be done using an integrative modeling approach (IMA), in which different scientific disciplines are combined into one holistic process, joined by a common modeling factor(s) subject to uniform assumptions. For example, in statistics, it may involve the integration of multivariate EVT (e.g., Cooley et al., 2012b), geostatistics (e.g., Banerjee et al., 2015), and spatial forecast verification (e.g., Friederichs and Thorarinsdottir, 2012), and in atmospheric science, it would encourage the use of high-resolution climate models (e.g., to resolve localized extreme events) combined with a spatio-temporal threshold selection algorithm and multivariate EVT to capture spatial dependencies. While a myriad of other coupling combinations is possible, the challenging aspect of this integration lies in maintaining suitable and consistent statistical assumptions across all modeling blocks.

This manuscript proposes a new generalized Spatio-Temporal Threshold Clustering (STTC) method for extreme events within the IMA framework. By generalized, we mean the algorithm's applicability to the wide variety of essential climate and non-climate variables (e.g., air pollution, crop yield, streamflow) and their derived quantities, hereafter, essential field quantities (EFQs). The algorithm, as objectively as possible, detects large-scale extreme spatial patterns that occupy fairly extensive geographical areas. Those patterns contain extreme fields (i.e., for high or low extreme thresholds, only fields from the respective upper or lower quantile of the distribution are selected) in both temporal and spatial variations. They can be linked to individual extreme episodes such as flash floods, heat waves, hurricanes, and droughts.

The threshold selection process is conditioned on these extreme patterns and enables one to detect latent spatial dependencies with the help of geometric indices from digital topology. Furthermore, to evaluate areas with the largest conditional frequency values, we apply the STTC methodology that uses a time series clustering procedure for a number of geometric indices. This procedure facilitates the algorithm's ability to classify a spatial pattern of an image (e.g., the conditional frequency of positive or negative extreme EFQ) by several geometrical constructs described in Table 1.1.

Pixel	Smallest unit of a digital image
Isolated Structure	Collection of adjacent pixels with value of one
Area	Number of non-zero pixels in the structure
Convex Hull	Smallest convex polygon that contains the structure
Perimeter	Length of the outside boundary of the structure

Table 1.1: Geometrical constructs used in geometric indices.

As far as we are aware, this machine learning IMA is a new contribution in atmospheric science that links the threshold selection process conditioned on extreme patterns with the clustering of multivariate series of geometric properties. The remainder of the Chapter 1 is structured as follows. Section 1.1 reviews the conceptual framework and details all steps for the STTC algorithm. Section 1.2 describes the dataset for illustrating the algorithm development and evaluation. Section 1.3 and its subsections present underlying methods. Section 1.4 highlights an application and comparative analysis utilizing daily precipitation data and underlines essential findings. Section 1.5 gives the conclusions.

1.1 Conceptual framework

The STTC algorithm consists of five main steps:

1. EFQ construction closely matched to the timeframe of interest;
2. Dimension reduction based on statistical quantiles;
3. Spatial domain mapping represented by the geometric indices;
4. Time series clustering applied to the multivariate series of geometric properties;

5. Threshold selection linked to the time series clustering.

The first step is the most flexible and difficult to apply. The user must decide whether to convert (e.g., normalize, standardize) the raw data to best represent an extreme pattern of interest. Large-scale patterns are usually described by gridded data, such as precipitation, temperature, geopotential height, or vorticity, to name a few. The patterns can be analyzed either at the right or left tail of the distribution. Ultimately, the choice of EFQ is based on the problem at hand and individual user preference.

Next, the dimension reduction or conditioning step identifies extreme processes that are spread out over a relatively extended portion of the spatial domain. This step is necessary to narrow spatio-temporal space to fit the class of events of interest. For example, the conditioning can be performed utilizing the algorithm to find either positive (i.e., wet-day) or negative (i.e., dry-day) extreme fields. Rather than considering extreme values at individual locations and their temporal dependence, we consider an overall spatial field that is conditioned on being extreme (i.e., not all individual grid cells within the field have to be extreme). In this case, it is possible to depict large-scale spatial extreme patterns independent of whether or not individual grid cells in space are extreme. The concept is similar to the high field energy used to identify severe storm environments in Gilleland et al. (2013, 2016a). They used the product of the 0-6km wind shear and maximum potential wind speed of updrafts and selected its upper quartile in space greater than its 90th percentile over time to define the high field energy. Here, we apply this framework by first identifying a data-specific quantile in space (i.e., the quantile function is applied over the entire spatial domain) and then in time. We also replace the concept of high field energy with more general positive and negative extreme fields (see Section 1.3.2) that can accommodate a variety of extreme field definitions, including different quantile values to meet the specific needs of different datasets.

The classification and evaluation of the graphical properties of the spatial fields, depicted in the previous step, is the purpose of the spatial domain mapping phase of the algorithm. The user selects a vector of initial thresholds where evaluation takes place. The threshold selection framework is integrated with methods from digital topology. Values greater or equal to a chosen threshold are assigned to one, and values below zero. It is called image digitizing and is a widely used technique in computer imaging. The digitized image can then be mapped to several geometric attributes that represent a particular graphical property of the image. As the threshold varies, so do the values of the geometric attributes. As such, one can create a threshold series that is mapped to the corresponding multivariate series of geometric properties. The mapping process can be repeated for all geometric attributes, creating a non-linear dependence between the threshold series and the desired geometrical properties.

The fourth step of the algorithm consists of applying time series clustering to the multivariate series of geometric properties derived in the previous step. This unsupervised machine learning approach separates the multivariate series of geometric properties into individual clusters by minimizing the average dissimilarity between each cluster's centrally-located representative object and any other object in the same cluster. Each of these representative objects is associated with a specific threshold from the threshold series.

The last threshold selection step selects a threshold out of the threshold series based on the clustering analysis of the multivariate series of geometric properties. The position of the representative object in the multivariate series of geometric properties is then matched to the same position in the threshold series, producing a cluster-specific threshold.

Thus, the overall objective of the algorithm is to detect latent spatial dependencies for an EFQ of interest within large-scale extreme episodes and to automate threshold selection for the extreme spatio-temporal processes. Ultimately, the identified spatial extremes can be linked to individual weather patterns because the corresponding occurrence times of such extreme events are tracked during the identification process. The objective threshold choice for extreme event modeling can help to deepen our understanding of the underlying processes in forming those patterns, that were possibly overlooked in the CLST methods. The STTC algorithm incorporates spatial and temporal dependencies in one holistic modeling framework and enables future spatio-temporal analysis of extreme events based on either a single quantity (e.g., precipitation) or a composite index of multiple quantities (e.g., the Palmer Drought Severity Index). Moreover, the extreme threshold value that was estimated through the unsupervised learning approach and the resultant extreme spatial field can be further incorporated into a spatial forecast verification process or independently applied in statistical modeling utilizing multivariate EVT. The use of this method can thus ensure consistency between the extreme threshold values and spatial fields selected during the modeling and forecast verification steps. A detailed summary of the entire STTC algorithm can be found in the Appendix.

1.2 Data

To illustrate the algorithm development and evaluation, we considered station-based precipitation measurements in mm/day obtained from the Global Historical Climatology Network-Daily (GHCN-Daily; Menne et al., 2012a) collected over the period 01/01/1961 to 12/31/2016 (56 years). While the GHCN-Daily dataset was subject to a rigorous quality assurance procedure (Menne et al., 2012b), it was not adjusted for artificial discontinuities such as changes in station location, instrumentation, and time of observation (Donat et al., 2013b). To deal with these issues, we only selected stations with at least 40% data coverage, yielding 8516 stations. In addition, as described in Liang et al. (2004), we performed topographical adjustments to account for the strong elevation dependence of the rain gauge sites by using the Parameter-Elevation Regression on Independent Slopes Model (PRISM) (Daly et al., 1994, 1997). We then interpolated the GHCN-Daily dataset using the Cressman objective analysis scheme (Cressman, 1959) to a 30 km by 30 km grid covering the contiguous United States (CONUS). As a result, we obtained approximately 27,000 spatial points, with each point having over 20,000 days of temporal rainfall coverage.

1.3 Methods

1.3.1 EFQ construction

A fundamental decision in any extreme value analysis is to choose the types of extreme events to analyze (e.g., flash floods, persistent droughts, or other natural disasters). Depending on the choice, our algorithm allows adjusting an accumulation window (W). The proposed EFQ derivation involves several steps, which can be modified or entirely skipped depending on a specific application:

1. Remove seasonal patterns by subtracting climatological mean values over the entire time length from the raw data;
2. Calculate W-period rolling accumulated anomalies;
3. Standardize accumulated anomalies by the corresponding high (or low) temporal quantile.

In this research thesis, we select no accumulation window, high quantile (0.95), and use daily precipitation as the field of study. That is, our interest during the algorithm testing lies in short-term extreme rainfall. These short-duration impactful precipitation events can produce a major natural hazard, such as a flash flood in some areas. To illustrate the general application, we also later present examples using more extended Ws for extreme EFQs. By design, EFQ has no units.

1.3.2 Dimension reduction

Conventional temporal dimension reduction methods such as principal component analysis (PCA) and linear discriminant analysis (LDA), which rely heavily on mean and covariance estimation, are not very informative for extremes. Although a method similar to principal component analysis, based on extreme value analysis, was proposed in Cooley and Thibaud (2019) and used in Jiang et al. (2020), it is not clear how it can be generalized to accommodate all types of extreme value dependence structures (Coles et al., 1999a). Additionally, the benefits of this method over traditional PCA are also not well understood. The block-maxima approach is often used to analyze extremes but potentially excludes relevant observations (Coles, 2001a). The challenge is, therefore, to subsample a relatively small number of observations to represent the tail of the distribution and select an aggregate measure that accurately describes extreme events. Gilleland et al. (2013) considered, for this purpose, a conditional frequency with quantile-based dimension reduction methodology in analyzing severe storms. In the present context, we modify this method to apply for temporal and spatial quantile values more appropriate in our case (see Section 1.4). We further formalize this method with the following definitions.

Definition 1.3.2.1. Let $Y(\mathbf{s}, t)$ be a spatio-temporal EFQ with a number of spatial pixels $\mathbf{s} \in \mathcal{D}$, where $\mathcal{D}$ is the spatial domain, then at each time point $t = [1, \ldots, T]$, the ηth-quantile in space $y_{t,\eta}$, is given by $\mathbb{P}\{Y(\mathbf{s}, t) \leq y_{t,\eta}\} = \eta$, where $\eta \in [0, 1]$.

The quantile $y_{t,\eta}$ in Definition 1.3.2.1 is taken over space at each point in time so that a univariate time series of spatial quantiles results. We denote this series by $\mathbf{g}_t = g_1, \ldots, g_T = y_{1,\eta}, \ldots, y_{T,\eta}$. In addition, we introduce a function $\lambda[\cdot]$ with possible forms in Table 1.2.

Name	Value
Fixed Value	$\lambda[x_t] = constant$
Average	$\lambda[x_t] = \frac{1}{T}\sum\limits_t x_t = u$
Temporal Quantile	$\lambda[x_t] = x_\tau$, where $\mathbb{P}\{x_t \leq x_\tau\} = \tau$ and $\tau \in [0, 1]$

Table 1.2: Some possible values for function $\lambda[x_t]$.

Definition 1.3.2.2. A space-time process $Y(\mathbf{s}, t)$ is called a *Positive Extreme Field* (PEF) at any time point t_i when $\{\mathbf{g}_{t_i} > \lambda[\mathbf{g}_t], i = [1, \ldots, N_{\mathrm{H}}], N_{\mathrm{H}} \leqslant T\}$ denote to be the $\mathrm{H}(\mathbf{s}, t_i)$ to be the subset of $Y(\mathbf{s}, t)$ when $Y(\mathbf{s}, t)$ is a PEF.

Definition 1.3.2.3. A space-time process $Y(\mathbf{s}, t)$ is called a *Negative Extreme Field* (NEF) at any time point t_j when $\{\mathbf{g}_{t_j} < \lambda[\mathbf{g}_t], j = [1, \ldots, N_{\mathrm{L}}], N_{\mathrm{L}} \leqslant T\}$ denote to be the $\mathrm{L}(\mathbf{s}, t_j)$ to be the subset of $Y(\mathbf{s}, t)$ when $Y(\mathbf{s}, t)$ is a NEF.

The PEF/NEF concepts are general constructs that can be applicable to both short and long-term impactful extreme events. For instance, if we express PEF and NEF in terms of precipitation rates and try to detect natural disasters such as documented in the NOAA National Centers for Environmental Information (NCEI) report titled "U.S. Billion-Dollar Weather and Climate Disasters (2018)", we can identify a multitude of high profile cases, a few of which are presented in Figure 1.1.

Figure 1.1a shows Superstorm Sandy, where high PEF values (> 6) are observed over the Mid-Atlantic and New England regions and northern Ohio. The Great Flood of 1993, which is found with a much longer accumulation window

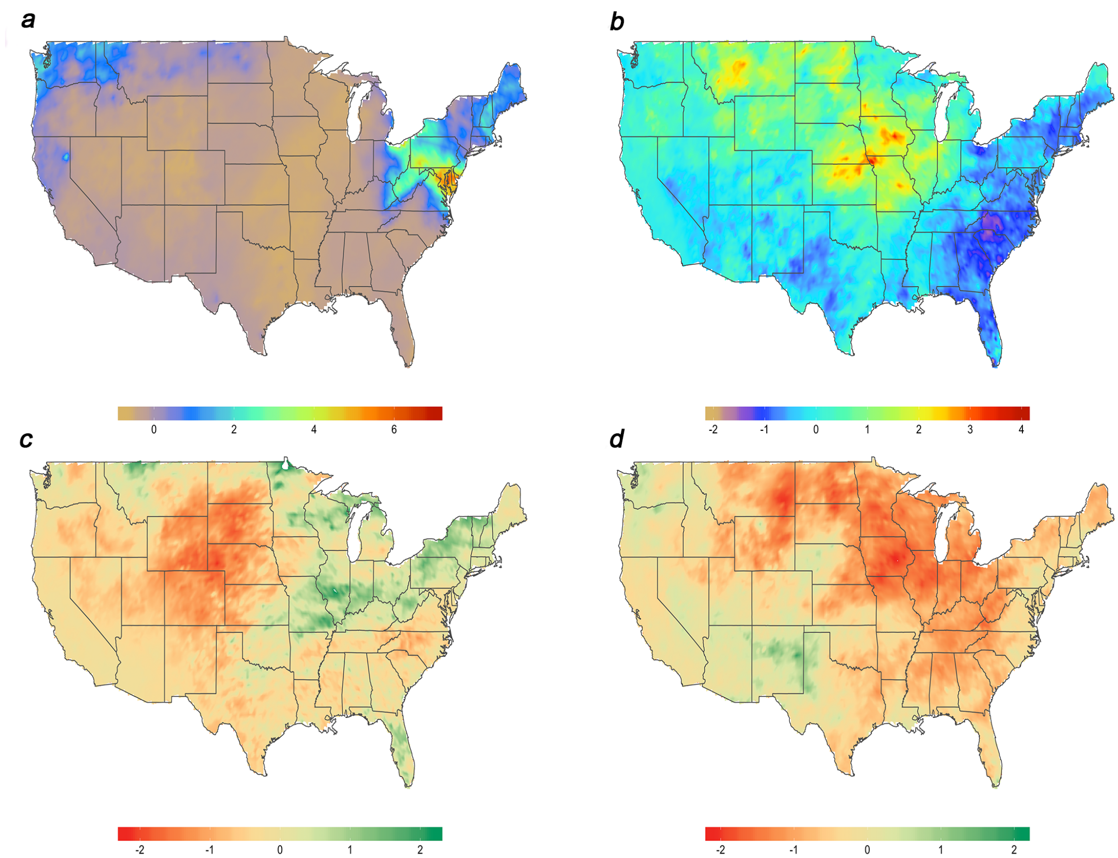

Figure 1.1: Examples of PEF: (a) October 2012 Superstorm Sandy (W= 4); (b) May to October Great Flood of 1993 (W= 180). Examples of NEF: (c) 2002 Drought, April - June (W= 90); (d) 1988 Drought, March - July (W= 150). The PEF and NEF were derived from precipitation measurements obtained from the gridded Global Historical Climatology Network-Daily (GHCN-Daily) dataset (see Section 1.2 for more information) and expressed in terms of standardized anomalies. The function $\lambda[.]$ from Table 1.2 was set to the temporal quantile definition with $\eta = 0.99$ for PEF ($\eta = 0.01$ for NEF) for the univariate time series g_t with $\tau = 0.95$ for PEF ($\tau = 0.05$ for NEF).

(W=180), is displayed in Figure 1.1b. The areas with PEF $>$ 2 can be seen along the Missouri and Mississippi rivers and their tributaries. In addition, several heavy rainfall events are spotted in Montana during the same time period. Results for Drought of 2002 are shown in Figure 1.1c. The Rocky Mountains and Midwest regions are particularly impacted by very dry conditions. Figure 1.1d shows a similar graph but for 1988 Drought and with W=150. The effects of low negative NEF ($<$ -1) are seen in Mid-Atlantic states, the southeastern United States, the midwestern United States, and the western United States. These are all examples of types of extreme processes $\mathrm{H}(\mathbf{s}, t_i)$ ($\mathrm{L}(\mathbf{s}, t_j)$) that define a spatio-temporal subset of a dataset used to define a conditional frequency of extreme EFQ.

The conditional frequency of an EFQ $Y(\mathbf{s}, t)$ is given by (i.e., conditioned on) $\mathrm{H}(\mathbf{s}, t_i)$ can be defined as a function of high threshold θ_h at any spatial pixel $\mathbf{s} \in \mathcal{D}$ in the following way

$$f(\mathbf{s}; \theta_\mathrm{h}) = \frac{\sum_{i=1}^{N_\mathrm{H}} \mathbb{1}_{\{\mathrm{H}(\mathbf{s},t_i) > \theta_\mathrm{h}\}}}{N_\mathrm{H}} \tag{1.1}$$

where $\mathbb{1}_{\{\cdot\}}$ is the indicator function.

Alternatively, we can delineate a similar quantity but conditioned on $\mathrm{L}(\mathbf{s}, t_j)$ if a low threshold θ_l is of interest

$$f(\mathbf{s}; \theta_\mathrm{l}) = \frac{\sum_{j=1}^{N_\mathrm{L}} \mathbb{1}_{\{\mathrm{L}(\mathbf{s},t_j) < \theta_\mathrm{l}\}}}{N_\mathrm{L}} \tag{1.2}$$

In terms of additional notations, let $\theta = [\theta_1, \ldots, \theta_\mathrm{n}]$ be a set of initial thresholds such as $\theta_1 < \theta_2 <, \ldots, < \theta_\mathrm{n}$. We can think of the series $f(\mathbf{s}; \theta_1), \ldots, f(\mathbf{s}; \theta_\mathrm{n})$ as a procedure of partitioning the conditional frequency image into several overlapping homogeneous fields, where each field is mapped into the corresponding geometric properties series described in the next section.

1.3.3 Spatial domain mapping

Geometric indices employed here were first introduced by AghaKouchak et al. (2010) and applied to validate radar data against satellite precipitation estimates and weather prediction models. Gilleland (2017a) complemented geometric indices with mean-error and mean-square-error distances, to introduce new diagnostic plots within a spatial forecast verification framework. The diagnostics were applied to a number of cases from the spatial forecast verification intercomparison project (ICP, `http://www.ral.ucar.edu/projects/icp`) in Ahijevych et al. (2009) and Gilleland et al. (2009a, 2010). The indices vary between zero and one and describe the connectivity, shape, and area of the image pixels for a predefined threshold. The connectivity index is defined as

$$C_{\text{index}} = 1 - \frac{n-1}{\sqrt{m}+n}, \tag{1.3}$$

where n is the number of isolated structures, and m is the number of pixels with a value of one (i.e., above a chosen threshold). Here, an isolated structure is a group of adjacent non-zero pixels. The connectivity index shows how the structures within the image are interconnected. The higher (lower) the index, the more connected (dispersed) the fields are within the image. The shape index is given by

$$S_{\text{index}} = \frac{P_{\text{min}}}{P}, \tag{1.4}$$

where P is the perimeter of the pixels above a given threshold, and P_{min} is the theoretical minimum perimeter of an m-pixel pattern, which is attained if the pattern of non-zero pixels was formed closest to a perfect circle. Mathematically, it is defined as

$$P_{\text{min}} = \begin{cases} 4 \times \sqrt{m}, & \text{if } \lfloor \sqrt{m} \rfloor = \sqrt{m} \\ 2 \times \left(\lfloor 2 \times \sqrt{m} \rfloor + 1 \right), & \text{otherwise} \end{cases},$$

where $\lfloor \cdot \rfloor$ is the floor function. The near-one values of S_{index} imply an approximately circular pattern of non-zero pixels. For a more detailed discussion about these indices, see Gilleland (2017a). Finally, to measure object complexity, we introduce another geometric index defined in Bullock et al. (2016) as follows

$$\bar{A}_{\text{index}} = 1 - A_{\text{index}} \equiv 1 - \frac{A}{A_{\text{convex}}}, \tag{1.5}$$

where A is the area of the pixels (i.e., the number of non-zero grid points) above a given threshold, A_{convex} is the area of the convex hull around those pixels, and A_{index} defined as their ratio. The index values close to zero are representative of more structured image patterns, whereas fairly dispersive image patterns imply values near one.

As previously stated, past applications of geometric indices were performed primarily for model validation. However, we adapt this approach to observed data to evaluate geometric index values for different thresholds. The aim is to determine specific geometrical properties that are relevant to the conditional frequency of extreme EFQ, where high (or low) threshold values are derived via an unsupervised clustering procedure. That is, for every threshold θ_i from the set of initial thresholds θ, geometric properties are estimated and then grouped into the multivariate series of geometric properties for time series clustering depicted in the following section. For brevity, we represent these geometric properties series as $\mathrm{G}_\alpha[f(\mathbf{s};\theta)]$, where $\alpha = [1,\dots,4]$ and $\mathrm{G}_\alpha[\cdot]$ is an index operator for geometric attribute α of the conditional frequency f at threshold θ. We set $\mathrm{G}_1[\cdot] = C_{\text{index}}[\cdot]$, $\mathrm{G}_2[\cdot] = S_{\text{index}}[\cdot]$, $\mathrm{G}_3[\cdot] = \bar{A}_{\text{index}}[\cdot]$ and $\mathrm{G}_4[\cdot] = A[\cdot]$.

Figure 1.2 shows a typical example of the evolution of geometric indices as the functions of a threshold series.

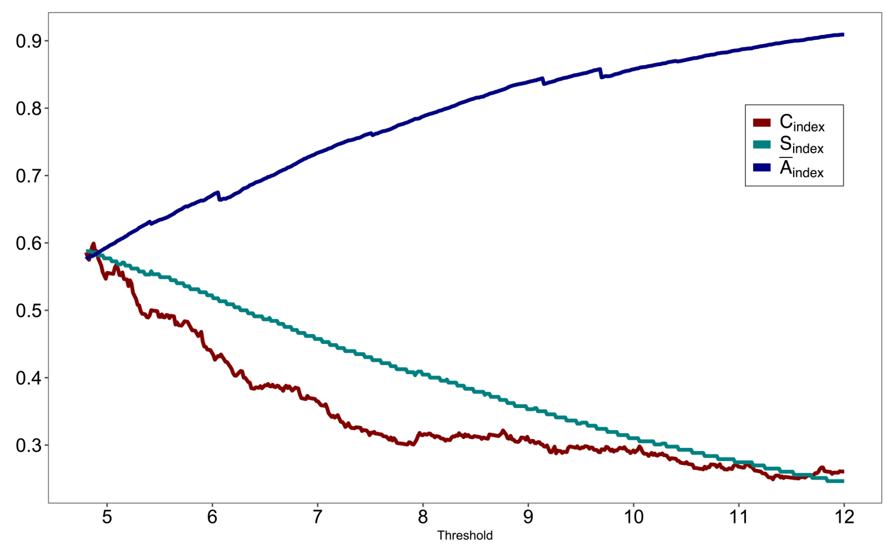

Figure 1.2: Geometric indices evolution as a function of threshold for summer EFQ (GHCN-Daily dataset) (see Section 1.2 for more information).

The connectivity index starts high, near 0.6, with a few isolated non-zero grid points. As the threshold increases, the index drops as more contiguous clusters are segmented out and finally levels off at around 0.2 with more scattered spatial fields. The shape index is almost monotonically decreasing as a function of threshold series, with values near one indicating an approximately circular pattern of non-zero grid cells. The complexity index is almost an inverse mirror image of the shape index, with values closer to unity indicating higher dispersiveness of the spatial patterns. Another way to look at it is to consider how the frequency of high extreme events changes as a function of threshold. At low threshold values, the probability of simultaneous extreme events at two different spatial locations is relatively high, and at high threshold values, the same probability is relatively low. As such, both connectivity and shape indices will decrease, but the complexity index will increase (area will decrease). Analyses were performed using the `SpatialVx` (Gilleland, 2017b) package in R.

1.3.4 Time series clustering

Clustering is a statistical process applied in machine learning to group unlabeled data into homogeneous segments or clusters. The clusters are formed through the segmentation of the data to maximize both inter-group dissimilarity and intra-group similarity according to objective criteria. The segmentation process entirely depends on a distance or dissimilarity metric, which measures how far away two objects are from each other. The degree of dissimilarity (or similarity) between the clustered objects is of significant importance in cluster analysis. Common dissimilarity metrics such as Euclidean and Manhattan are not suitable for time series clustering because they ignore serial correlations within the time series. Our main aim is to investigate the topological features of sequences of conditional frequency images represented by a multivariate series of geometric properties. These series may exhibit high degrees of serial correlation at different lag times (Figure 1.2). Therefore, the first fundamental task in applying a clustering method is the selection of a dissimilarity metric that can capture these subtle characteristics. In particular, we are interested in discovering clusters of multivariate series of geometric properties with similar profiles representing large-scale extreme spatial patterns that can be buried under noise. For this purpose, several authors have suggested using dissimilarity measures based on the estimated autocorrelation functions of the time series (e.g., Caiado et al., 2006; Galeano and Peña, 2000). Here, we apply this approach to the multivariate series of geometric properties defined in the following way. Given $f(\mathbf{s};\theta)$ mapping to $\mathrm{G}_\alpha[f(\mathbf{s};\theta)]$, where $\alpha = [1,\dots,4]$, we can alternatively map $f(\mathbf{s};\theta)$ to $\mathrm{G}_\beta[f(\mathbf{s};\theta)]$, where $\beta = [1,\dots,4]$. We let $\mathrm{G}_\alpha[f(\mathbf{s};\theta)]$ and $\mathrm{G}_\beta[f(\mathbf{s};\theta)]$ be two geometric properties series, that is $\mathrm{G}_\alpha[f(\mathbf{s};\theta)] = \mathrm{G}_\alpha[f(\mathbf{s};\theta_1)],\dots,\mathrm{G}_\alpha[f(\mathbf{s};\theta_n)]$ and $\mathrm{G}_\beta[f(\mathbf{s};\theta)] = \mathrm{G}_\beta[f(\mathbf{s};\theta_1)],\dots,\mathrm{G}_\beta[f(\mathbf{s};\theta_n)]$ for $\alpha \neq \beta$. We denote the vectors of the estimated autocorrelation coefficients of the geometric properties series G_α and G_β by $\widehat{\rho}_{G_\alpha} = (\widehat{\rho}_{1,G_\alpha}\dots\widehat{\rho}_{R,G_\alpha})$ and $\widehat{\rho}_{G_\beta} = (\widehat{\rho}_{1,G_\beta}\dots\widehat{\rho}_{R,G_\beta})$ for some R such as $\widehat{\rho}_{i,G_\alpha} \approx 0$ and $\widehat{\rho}_{i,G_\beta} \approx 0$ for $i > R$. Dissimilarity between G_α and G_β is measured by the following Euclidean distance

$$d_{\mathrm{ACF}}(G_\alpha, G_\beta) = \sqrt{\sum_{i=1}^{R}(\widehat{\rho}_{i,G_\alpha} - \widehat{\rho}_{i,G_\beta})^2} \tag{1.6}$$

Analogously, we define $\widehat{\phi}_{ii,G_\alpha}$ and $\widehat{\phi}_{ii,G_\beta}$ as the vectors of the estimated partial autocorrelation coefficients of G_α and G_β. A corresponding dissimilarity metric is given by

$$d_{\text{PACF}}(G_\alpha, G_\beta) = \sqrt{\sum_{i=1}^{R}(\widehat{\phi}_{ii,G_\alpha} - \widehat{\phi}_{ii,G_\beta})^2} \tag{1.7}$$

In the seasonal analysis that follows, the d_{ACF} and d_{PACF} measures vary between 0 and 0.001, indicating further how delicate the autocorrelation differences between geometric properties series can be. To calculate both distances, we used R package `TSclust` (Montero and Vilar, 2014) .

The choice of the number of clusters K to form is another important task in the clustering algorithm. It typically involves iteratively searching for the number of clusters that optimize a variety of criterion measures. Generally, cluster selection is still an unresolved problem requiring multiple iterations. However, in our case, it is somewhat simpler.

Previously, we described that a balance between bias and variance was necessary when determining a threshold in EVT. A threshold choice for extremes in digital topology is also a challenge. A prudent question to pose is, "What are desirable geometrical properties of the conditional frequency of extreme EFQ?" While it is difficult to generalize the shape and complexity indices, we postulate that choosing a threshold value with the desirable geometrical properties is an interplay between the connectivity index and the area of non-zero points (referred to henceforth as connectivity/area trade-off). In cases involving both relatively low and very high threshold values, the connectivity index is close to one (see Equation 1.3). The higher the threshold value, the lower the probability of simultaneous extreme events at two different spatial locations and the smaller the area of non-zero grid points. To balance this trade-off, we need to find a threshold value that is high enough to be practical to determine large-scale spatial patterns but not too high to cause the connectivity index to get close to one and the area of the non-zero pixels to be close to zero. Thus, there are three clusters of interest: the first cluster has a multivariate series of geometric properties linked to the thresholds that are still high but may not be high enough to provide the desired geometric properties for the conditional frequency of extreme EFQ; the second cluster provides the high thresholds linked to desirable geometric properties; the third cluster produces the thresholds that are too high and impractical for our purpose. Therefore, we can observe if we have chosen more than three clusters, our task to deal with connectivity/area trade-off becomes very daunting. That is why we decided to concentrate our analysis on the second cluster to minimize potential connectivity/area impacts.

After selecting the dissimilarity measures and the number of clusters, it is necessary to select a clustering algorithm. The most popular partitioning clustering approaches are k-means and partitioning around medoids (PAM) (Kaufman and Rousseeuw, 1987). K-means creates cluster centers by averaging points within the cluster. However, this averaging process can be sensitive to outliers and breaks the max-stability property so that the mean of two maxima is no longer a maximum, which is an essential assumption in EVT. We, therefore, chose PAM, implemented in the `cluster` (Maechler et al., 2017) package in R, which provides a more robust outcome where each cluster is identified by its most representative object (called a medoid). Additionally, PAM preserves the observational features of the dataset, which is essential for clustering extremes. The algorithm has two phases, BUILD and SWAP:

1. In the first phase, k representative objects are selected to form an initial clustering set;
2. In the second phase, an attempt is made to improve clustering by exchanging selected and unselected objects.

The objective of PAM is for all selected clusters to minimize average dissimilarity between their centrally located representative object and any other object in the same cluster. Further details can be found in Kaufman and Rousseeuw (1990). The quality of the resulting clusters and the choice of the number of clusters can be assessed using the "silhouette coefficient" (Rousseeuw, 1987), a statistic to determine how similar an object is to its own cluster (cohesion) compared to other clusters (separation). For an object i it is defined as follows,

$$Sil(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \tag{1.8}$$

where $a(i)$ is the average dissimilarity of the object i to all other objects in its cluster, and $b(i)$ is the minimum average distance of object i to all other objects in the given cluster, not containing i. The value of the silhouette coefficient $Sil(i)$ ranges between -1 and 1, where a negative value is undesirable. If $Sil(i) \approx 1$, a strong structure has been found, which means that the inter-cluster distance is much larger than the intra-cluster distance. Conversely, if $Sil(i) < 0.25$, no substantial structure has been found. The average silhouette coefficient $\overline{Sil}(K)$ can be used to evaluate the quality of segregation into K clusters, such as, for any finite K find $\max_{K} \overline{Sil}(K)$.

1.3.5 Spatio-temporal threshold selection

Threshold selection is commonly used in the process of image segmentation (e.g. Sahoo et al., 1988; Sezgin et al., 2004), that is, the process of using thresholds to partition an image into meaningful regions. This may include computer vision tasks such as finding a single threshold to separate an object in an image from its corresponding background or separation of light and dark regions. However, these methods are not informative about spatial patterns (Kaur and Kaur, 2014) and are thus not appropriate for complex images (i.e. color images where multiband thresholding may be necessary) (Yogamangalam and Karthikeyan, 2013) such as those involved in this study. The problems we aim to solve involve identifying distinct extreme episodes that spread out over a large spatial domain and thus require more sophisticated methods.

Our threshold choice is directly linked to the outcome of the time series clustering step. The clustering solution produces results for the three clusters and their medoids, where every member is associated with the threshold and corresponding values of the geometric properties series. As mentioned in the previous section, we address the connectivity/area trade-off by selecting members from the second cluster only. The threshold selection is implemented as follows. First, we select maximum average silhouette coefficient between $\overline{Sil}_{\mathrm{ACF}}$ and $\overline{Sil}_{\mathrm{PACF}}$ and called it

$\overline{Sil}_{\text{best}}$. This is necessary to select the best clustering solution between two distance measures. Next, the medoid from the second cluster is matched by the $\overline{Sil}_{\text{best}}$ to correspond to our threshold choice. Finally, the medoid index in the multivariate series of geometric properties is used to determine the corresponding geometric attributes. Equipped with both the high spatio-temporal threshold value and the spatial domain snapshot represented by geometric properties series, the user can make a more informative choice on a threshold appropriate for the particular class of extreme events and the needed applications.

1.3.6 Comparative analysis

To better understand the difference in spatial patterns produced by the two threshold selection methodologies, we compare geometrical properties between the STTC and CLST methods. The former is represented by a binary image of $f(\mathbf{s};\theta)$ conditioned on the fields are PEFs, whereas the latter is a binary image of usual location-specific frequency. However, in this form, the two frequencies are not compatible over space because they represent different subsets of the temporal domain. To put them on an equal footing, we start with a definition of a location-specific frequency. Let $Y(\mathbf{s},t_j)$ be a spatio-temporal EFQ such that $Y(\mathbf{s},t_j) > 0$ for any $j = [1,\dots,M]$, $M \leqslant T$ (i.e., our interest lies in "wet" EFQs in this comparative study). Then for any spatial pixel $\mathbf{s} \in \mathcal{D}$, we can write

$$f_\tau(\mathbf{s}) = \frac{\sum\limits_{j=1}^{M} \mathbb{1}_{\{Y(\mathbf{s},t_j) > \lambda[Y(\mathbf{s},t_j)]\}}}{M} \tag{1.9}$$

where $\lambda[\cdot]$ is the temporal quantile defined in Table 1.2 such as $\lambda[Y(\mathbf{s}, t_j)] = Y_\tau(\mathbf{s})$ with $\tau = 0.95$. Note that $f_\tau(\mathbf{s})$ is a single spatial field representing the frequency of exceeding the $\tau = 0.95$ quantile at each spatial location.

By its definition, the binary area calculated for all spatial pixels $\mathbf{s}$ and derived from A_{index} for this quantity will have more non-zero pixels than the segmented area of $f(\mathbf{s}; \theta)$ and therefore will still not be adequate for the comparison. To reach a compromise, we threshold the spatial field for $f_\tau(\mathbf{s})$ to get as close to the area of $f(\mathbf{s}; \theta)$ as possible. We can write this area-matched indicator as $f_\tau(\mathbf{s}; \theta')$ for some threshold θ'. Mathematically, we are searching for threshold θ' such that $A[f_\tau(\mathbf{s}; \theta')] \approx A[f(\mathbf{s}; \theta)]$. Henceforth, geometrical properties comparison is performed between $f_\tau(\mathbf{s}; \theta')$ and the algorithmically-determined $f(\mathbf{s}; \theta)$.

1.3.7 Mean error distance

To complete the above comparative analysis, we now introduce a localized performance measure for binary image comparison described in Peli and Malah (1982) for image detection in computer vision and applied to evaluate weather forecasting model performance (e.g., Gilleland, 2017a; Zhu et al., 2011).

Suppose $B1$ and $B2$ represent two binary images, and N_{B1} and N_{B2} correspond to the number of non-zero pixels in space. Let $d(x, B2)$ be the shortest Euclidean distance from the pixel $x \in B1$ to the set $B2$ of one-valued pixels. Then, the mean error distance (MED) is defined by

$$\underset{B1 \to B2}{\text{MED}} = \frac{1}{N_{B1}} \sum_{x \in B1} d(x, B2) \tag{1.10}$$

If *B1* and *B2* do not have identical geometrical properties, then $\underset{B1 \to B2}{\mathrm{MED}} \neq \underset{B2 \to B1}{\mathrm{MED}}$. This asymmetrical property is used to calculate a difference between two binary images denoted by

$$\Delta_{\mathrm{MED}}(B1, B2) = \underset{B1 \to B2}{\mathrm{MED}} - \underset{B2 \to B1}{\mathrm{MED}} \tag{1.11}$$

When Δ_{MED} is positive, it indicates that *B1* has more non-zero grid cells than *B2*, and vice versa, when the difference is negative. The larger the difference, the more material this distinction is.

1.4 Results

We apply the STTC algorithm to the EFQ of daily precipitation over CONUS. Throughout, our analysis describes the spatial pattern and temporal evolution of the conditional frequency of extreme, short-duration (i.e., daily) EFQ based on a high threshold. That is, our goal is to select a high threshold θ_{h} for the frequency of extreme EFQ conditioned on the fields that are PEFs based on daily precipitation rates. Based on empirical tests not shown here, we set $\eta = 0.99$ for the univariate time series g_t and apply the temporal quantile definition for the function $\lambda[\cdot]$ from Table 1.2 with $\tau = 0.95$.

Heavy rainfall is seasonally dependent and varies in space and time. Thus, to understand the complete evolution of $f(\mathbf{s}; \theta)$ and its spatial extent, we stratify the dataset by seasons defined as winter (DJF), spring (MAM), summer (JJA), and fall (SON). For every season, the clustering algorithm determines a high threshold value using the multivariate series of geometric properties clustering procedure. Each cluster strength is represented via its average silhouette coefficient; values near one suggest strong clustering structures. Every threshold value is mapped to the corresponding set of geometric properties series, revealing the geometrical properties of the spatial image.

We demonstrate the performance of the STTC algorithm using inter-cluster comparison and determine important graphical properties of the three clusters. Also, we compare the results of our algorithm to the area-matched frequency (see Equation 1.9) and identify the key differences between these two methods.

1.4.1 Inter-cluster comparison

To evaluate the performance of the clustering algorithm and the resultant geometrical properties of the conditional frequency of extreme EFQ, we characterize the results in the following manner:

Cluster 1. Low thresholds. Expected both high values for the connectivity index and the area of non-zero points. The probability of simultaneous extreme events at two different spatial locations is relatively high.

Cluster 2. Intermediate thresholds. Expected intermediate values (i.e., between first and third clusters) for the area of non-zero points. The connectivity index is greater than zero and less than one-half. The threshold values from this cluster can be potentially used for modeling in EVT in conjunction with other statistical diagnostic tools. The probability of simultaneous extreme events at two different spatial locations is decreasing.

Cluster 3. High thresholds. Expected small values for the area of non-zero points. The connectivity index is greater than zero. The probability of simultaneous extreme events at two different spatial locations is relatively low.

Tables 1.3 - 1.5 display the outcome of this process.

Months	θ	$\overline{Sil}(K)$	C_{index}	S_{index}	$\bar{A}_{\text{index}}$	A
DJF	4.05	0.73	0.66	0.59	0.55	4901
MAM	6.12	0.80	0.64	0.62	0.48	4793
JJA	8.42	0.71	0.39	0.49	0.70	2930
SON	8.52	0.70	0.42	0.53	0.61	3904

Table 1.3: Results of the Spatio-Temporal Threshold Clustering algorithm for the first (low thresholds) cluster.

Months	θ	$\overline{Sil}(K)$	C_{index}	S_{index}	$\bar{A}_{\text{index}}$	A
DJF	7.22	0.46	0.42	0.41	0.72	2008
MAM	9.86	0.46	0.31	0.40	0.75	1800
JJA	11.44	0.49	0.30	0.35	0.84	1379
SON	11.80	0.50	0.34	0.36	0.81	1668

Table 1.4: Results of the Spatio-Temporal Threshold Clustering algorithm for the second (intermediate thresholds) cluster.

Months	θ	$\overline{Sil}(K)$	C_{index}	S_{index}	$\bar{A}_{\text{index}}$	A
DJF	9.17	0.65	0.36	0.32	0.82	1128
MAM	12.20	0.58	0.27	0.31	0.85	940
JJA	13.57	0.58	0.26	0.27	0.89	829
SON	13.70	0.65	0.33	0.31	0.86	1104

Table 1.5: Results of the Spatio-Temporal Threshold Clustering algorithm for the third (high thresholds) cluster.

Not surprisingly, the first *low thresholds* cluster has a higher connectivity and shape index and a larger areal extent, and smaller complexity than the other two clusters. The average silhouette coefficient for this cluster is between 0.70 and 0.80. This cluster is useful if it is necessary to sample lower threshold values than the other two clusters. However, as we pointed out earlier, it may not be appropriate for extreme value analysis. As expected, the third *high thresholds* cluster has the lowest values for the connectivity and shape index and the smallest non-zero area among all other clusters. The average silhouette coefficient for the third cluster varies between 0.58 and 0.65. Finally, the second *intermediate thresholds* cluster has values of the geometric indices and the non-zero area between the first and second clusters. In addition, further examination of Figure 1.2 reveals that the location of the second cluster falls within a low volatility regime for the connectivity index, which is desirable for extreme value analysis as it helps to deal with the connectivity/area trade-off. However, compared with the other two clusters, it results in a lower $a(i)$ (i.e., the average dissimilarity of the object i to all other objects in its cluster) from Equation 1.8 and, consequently, a lower average silhouette coefficient, which fluctuates between 0.46 and 0.50. However, it is still relatively high, as in all other clusters (i.e., above 0.25). Clearly, our clustering process characterizes the common features of a multivariate series of geometric properties for the three clusters reasonably well. Hence, selecting a threshold from a different cluster for the conditional frequency of extreme EFQ can lead to different spatial forecast verification outcomes and, ultimately, to different statistical inferences regarding factors that explain extreme weather and climate events.

Figure 1.3 provides a graphical representation of the above results.

The *intermediate thresholds* cluster has fewer isolated and more contiguous

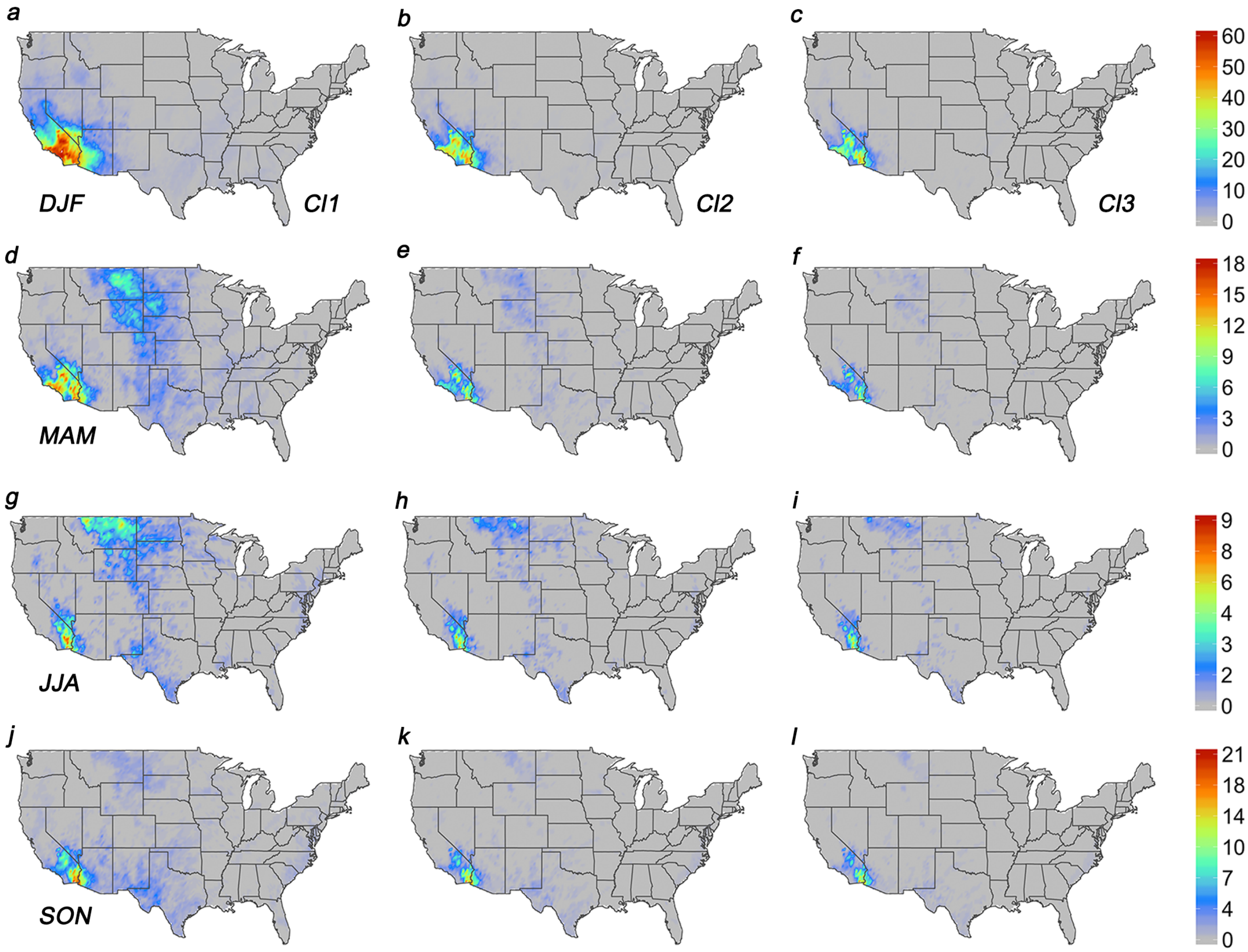

Figure 1.3: Frequency of extreme daily EFQ conditioned on PEFs for the three clusters: *low thresholds* cluster with threshold values from Table 1.3 (Cl1); *intermediate thresholds* cluster with threshold values from Table 1.4 (Cl2); *high thresholds* with threshold values from Table 1.5 (Cl3), for every season: (a)-(c) Winter (DJF); (d)-(f) Spring; (g)-(i) Summer; and (j)-(l) Fall. All frequencies are stated in percent form.

clusters (i.e., higher connectivity index), which are approximately circular in shape (i.e., higher shape index) and less scattered (i.e., lower complexity index). It occupies a broader spatial domain than the *high thresholds* cluster. The main advantage of having higher connectivity and larger areal extent attributes in spatial forecast verification for extremes is that it provides a more coherent representation of extreme patterns in space and, as a result, allows for a more informed decision about the causes, severity, and impact of the extremes under climate change, assuming climate models can simulate these graphical configurations reasonably well. These graphical qualities make the *intermediate thresholds* cluster the most likely choice for high threshold selection for the conditional frequency of extreme EFQ.

The threshold selection methodology adopted here for $f(\mathbf{s};\theta)$ illustrates a fundamental difference with other conventional approaches for defining extreme event frequency. It is a common practice for threshold choice in univariate EVT to carry out an exploratory analysis or stability assessment of estimated parameters in the temporal domain without considering any spatial dependency. In our case, the threshold selection process has been conditioned on the fields are PEFs, which allows analysis of temporal processes within a reasonably large spatial field, notwithstanding that many of the individual grid cells may not be necessarily extreme.

1.4.2 Comparison with conventional approach

We stratify our dataset annually and by seasons, and Table 1.6 contains a summary of the large differences for connectivity, shape, and complexity indices between $f_\tau(\mathbf{s};\theta')$ and $f(\mathbf{s};\theta)$. Based on the experimental analysis (not shown here), we define a large difference for connectivity and shape indices as a difference smaller than -0.2, and for the complexity index, the corresponding large difference is defined as greater than 0.2. In the table, we count those differences, and the resulting sum is divided by the number of samples (56 in our case). Analysis of the table clearly reveals that the smallest differences between $f_\tau(\mathbf{s};\theta')$ and $f(\mathbf{s};\theta)$ are observed in complexity and shape indices. This is because $f_\tau(\mathbf{s};\theta')$ is designed to closely match the area of $f(\mathbf{s};\theta)$. The $\bar{A}_{\text{index}}$ is related to the area in A_{index}, and the S_{index} is related to the square root of the area in $P_{\min}$. On the other hand, it is apparent the largest differences are confined to the connectivity index. That is the algorithmically-determined $f(\mathbf{s};\theta)$ has more connected and contiguous regions than the area-matched $f_\tau(\mathbf{s};\theta')$.

Months	C_{index}	S_{index}	$\bar{A}_{\text{index}}$
DJF	82	2	5
MAM	89	2	2
JJA	68	0	0
SON	57	2	2

Table 1.6: Percent difference in geometric properties series between $f_\tau(\mathbf{s};\theta')$ and $f(\mathbf{s};\theta)$ when difference is large.

The differences between $f_\tau(\mathbf{s};\theta')$ and $f(\mathbf{s};\theta)$ are further verified with Figure 1.4 displaying MED values from the $f_\tau(\mathbf{s};\theta')$ to $f(\mathbf{s};\theta)$ against the $f(\mathbf{s};\theta)$ to $f_\tau(\mathbf{s};\theta')$ direction.

Perfect agreement between the two directions of MED would result in a point at the origin. Fields with non-zero MED that fall precisely on the 1-1 line (i.e.,

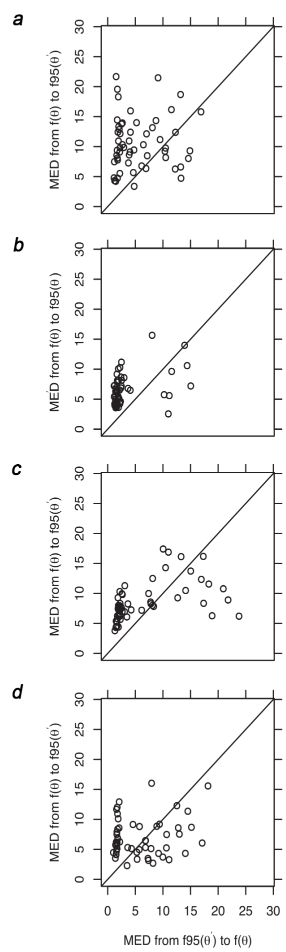

Figure 1.4: Plot of $\underset{f_\tau(\mathbf{s};\theta')\to f(\mathbf{s};\theta)}{\text{MED}}$ against $\underset{f(\mathbf{s};\theta)\to f_\tau(\mathbf{s};\theta')}{\text{MED}}$ with 1-1 line: (a) Winter, (b) Spring, (c) Summer, (d) Fall. For ease of notation, we abbreviate $f_\tau(\mathbf{s};\theta')$ to $f95(\theta')$ and $f(\mathbf{s};\theta)$ to $f(\theta)$ in the axis labels.

$\underset{B1 \to B2}{\text{MED}} = \underset{B2 \to B1}{\text{MED}}$) mean they contain spatial areas with exactly the same shape in roughly the same places of the respective fields; neither field will have non-zero grid points in an area where the other does not. For all seasons, we observe that most of the points lay above the 1-1 line. This is a strong indication that, on average, the area-matched $f_\tau(\mathbf{s};\theta')$ had fewer non-zero pixels than the algorithmically-determined $f(\mathbf{s};\theta)$. However, there were some years with larger magnitudes primarily displayed for MED in the $f_\tau(\mathbf{s};\theta')$ to $f(\mathbf{s};\theta)$ direction (see Figure 1.4c), possibly implying more scattered spatial representation of $f_\tau(\mathbf{s};\theta')$. Simply put, for large-scale, well-structured spatial patterns $f(\mathbf{s};\theta)$ will have a more prominent graphical representation than $f_\tau(\mathbf{s};\theta')$. Moreover, when MED values from $f_\tau(\mathbf{s};\theta')$ to $f(\mathbf{s};\theta)$ direction were small, there is a clear grouped area in MED from the $f(\mathbf{s};\theta)$ to $f_\tau(\mathbf{s};\theta')$ direction. Since this behavior is consistent across all seasons, we argue that it can play a crucial role in explaining the differences between the two MED directions.

So far, while we have seen a clear difference in the two MED directions and have determined its key feature, we could only speculate about geometrical properties of spatial patterns for both $f_\tau(\mathbf{s};\theta')$ and $f(\mathbf{s};\theta)$. Thus, it is imperative to evaluate this phenomenon further.

As before, we stratify our dataset annually and by seasons, but this time, to understand the underlying spatial processes that drive the difference between $f_\tau(\mathbf{s};\theta')$ and $f(\mathbf{s};\theta)$, we used a decision tree approach, which is non-parametric machine learning modeling method that splits the tree into branches based on the predictor variables according to the local optimal decision rule to predict the variable of interest (Breiman, 2017). The algorithm selects the best splitting criterion among selected attributes (i.e., C_{index}, S_{index} and $\bar{A}_{\text{index}}$) to predict the optimal value for $\Delta_{\text{MED}}[f_\tau(\mathbf{s};\theta'), f(\mathbf{s};\theta)]$ under these conditions. Figure 1.5 shows results for Δ_{MED}, with the differences of the three geometric indices between $f_\tau(\mathbf{s};\theta')$ and $f(\mathbf{s};\theta)$ as predictor variables.

For every season, we select a regression tree split, which results in the most observations in the terminal nodes. The constructed trees show that the first split in the regression trees is always based on the difference of C_{index} between $f_\tau(\mathbf{s};\theta')$ and $f(\mathbf{s};\theta)$, but it varied by season. For all seasons but SON, the left terminal node (DJF (< -0.27), MAM (< -0.25), JJA (< -0.16), SON (< -0.27)) is where the most observations are located. The corresponding percentages of the observations are 71%, 88%, 71%, and 48%. The SON terminal nodes observations are almost evenly split between left and right nodes (i.e., 48% vs 52%). Thus, the majority of the observations contained in these regression splits have relatively large negative differences of C_{index} between $f_\tau(\mathbf{s};\theta')$ and $f(\mathbf{s};\theta)$. As a result, if we continue to follow to the terminal nodes, the decision tree algorithm predicted fairly large negative corresponding Δ_{MED} values.

Equally important, Figure 1.6 plots Δ_{MED} for $f_\tau(\mathbf{s};\theta')$ and $f(\mathbf{s};\theta)$ against the differences of the three geometric indices for the same fields. It is apparent that the two frequency fields have very similar differences in shape and complexity indices. However, after selecting a decision node that contained most of the observations, they have very distinguishing properties in the difference in connectivity index with relatively large negative Δ_{MED}. A principal feature of the decision tree method is their ability to estimate a measure of the importance of each predictor variable in

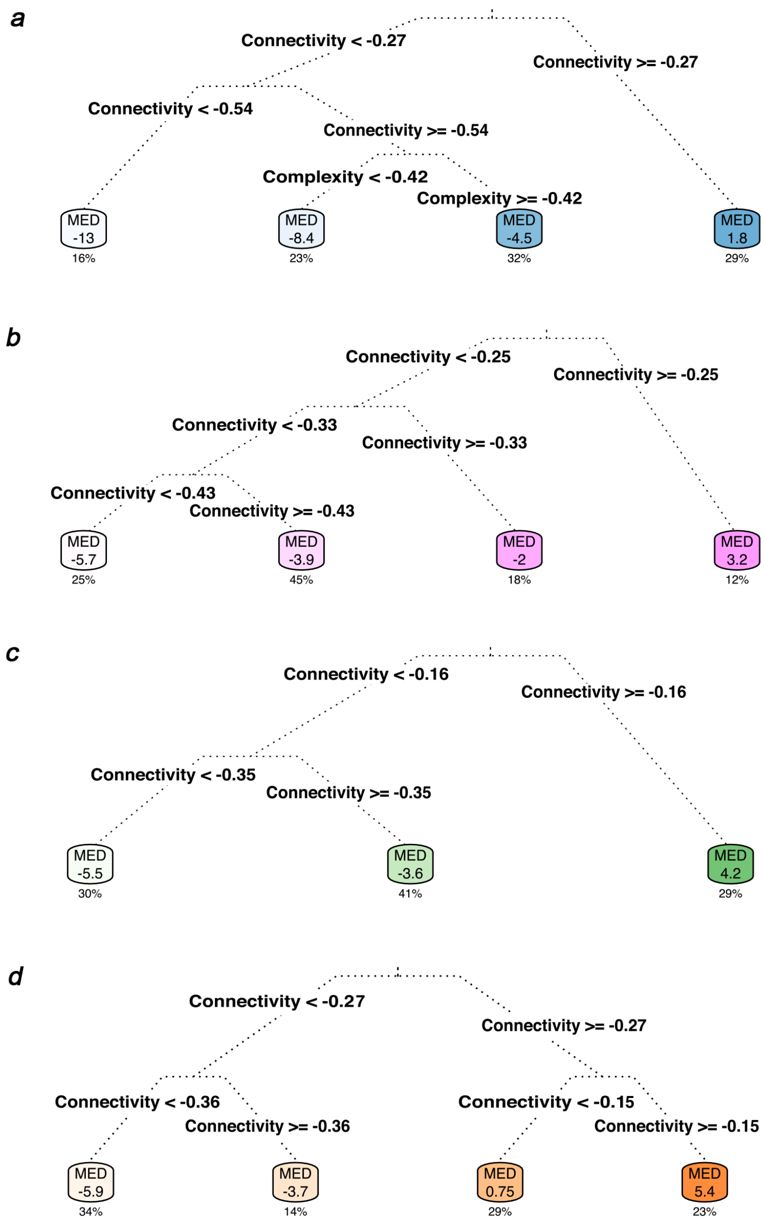

Figure 1.5: Regression tree for $\Delta_{\text{MED}}[f_\tau(\mathbf{s};\theta'), f(\mathbf{s};\theta)]$ derived from the differences of C_{index}, S_{index} and $\bar{A}_{\text{index}}$ between $f_\tau(\mathbf{s};\theta')$ and $f(\mathbf{s};\theta)$: (a) Winter, (b) Spring, (c) Summer, (d) Fall. Every non-terminal node displays a splitting criterion. If the split condition is "true," it goes to the left. If "false" to the right. Terminal nodes depict predicted values for Δ_{MED} (inside the cylinders) and percentages of observations in the nodes (outside the cylinders). Computation was done with `rpart` (Therneau et al., 2017) package in R.

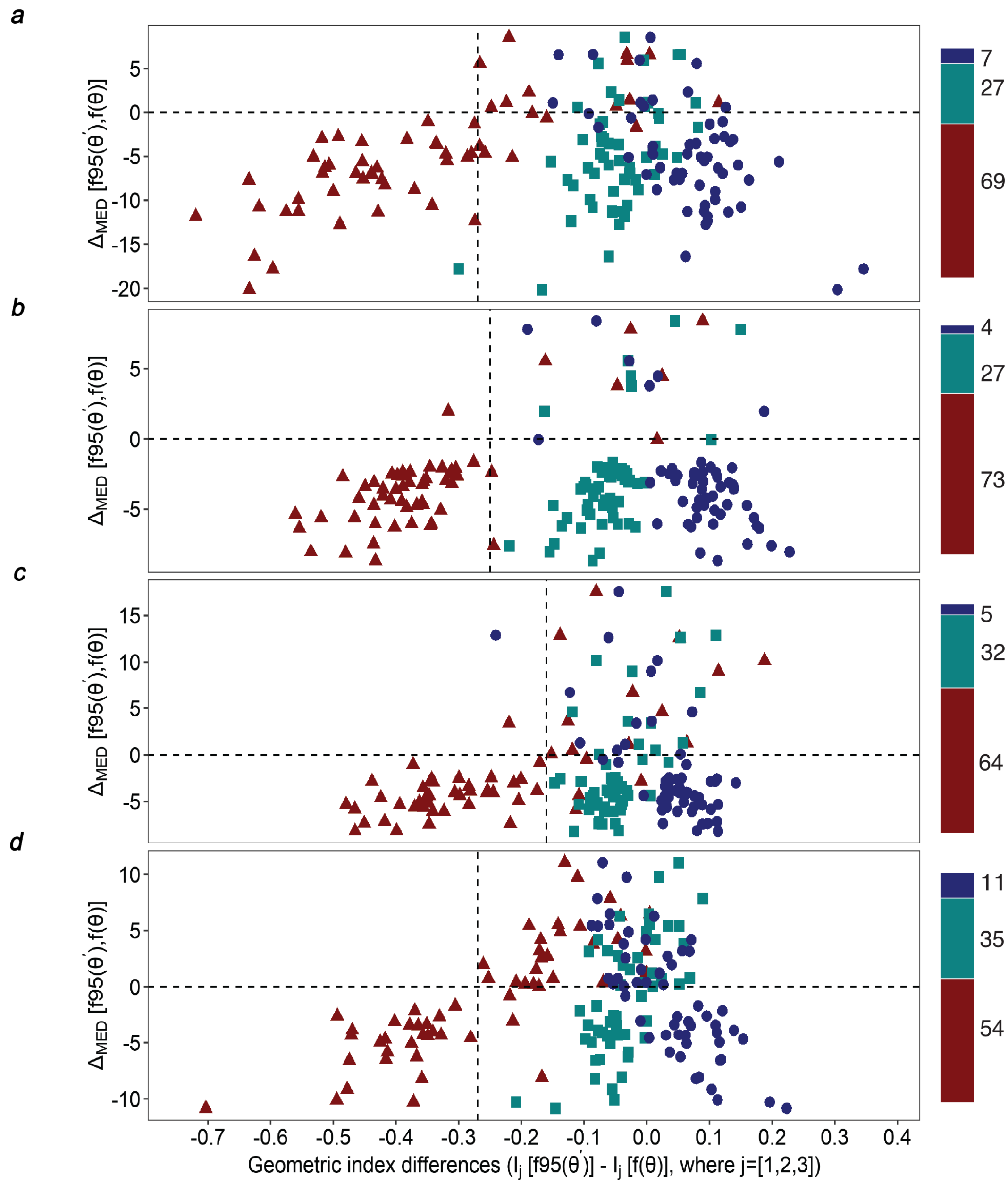

Figure 1.6: $\Delta_{\mathrm{MED}}[f_\tau(\mathbf{s};\theta'), f(\mathbf{s};\theta)]$ is plotted against the difference of $I_1 = C_{\mathrm{index}}$ (red triangles), $I_2 = S_{\mathrm{index}}$ (green squares) and $I_3 = \bar{A}_{\mathrm{index}}$ (blue circles) between $f_\tau(\mathbf{s};\theta')$ and $f(\mathbf{s};\theta)$: a) Winter, (b) Spring, (c) Summer, (d) Fall. For ease of notation, we abbreviate $f_\tau(\mathbf{s};\theta')$ to $f95(\theta')$ and $f(\mathbf{s};\theta)$ to $f(\theta)$ in the axis labels. Vertical dotted lines show the split on the difference of C_{index} for the majority of observations from the decision regression tree (see Figure 1.5). Stacked bars show the importance of differences in geometric indices expressed in percent.

modeling the response variable. The stacked bars show over 50% importance for the difference in C_{index} between $f_\tau(\mathbf{s};\theta')$ and $f(\mathbf{s};\theta)$ in the same figure.

In addition, Figure 1.7 displays a representative example with a large connectivity index difference equal to -0.27. Figures 1.7a and b display results for $f(\mathbf{s};\theta)$ and $f_\tau(\mathbf{s};\theta')$ respectively.

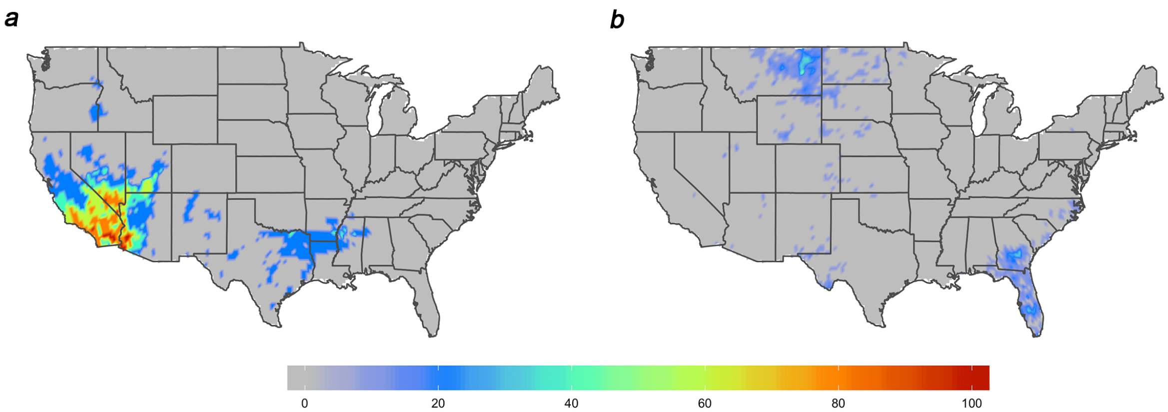

Figure 1.7: Examples of (a) $f(\mathbf{s};\theta)$ and (b) $f_\tau(\mathbf{s};\theta')$ for Winter of 2001 when the difference in C_{index} = -0.27, S_{index} = -0.12 and $\bar{A}_{\text{index}}$ = 0.1 between $f_\tau(\mathbf{s};\theta')$ and $f(\mathbf{s};\theta)$. All frequencies are stated in percent form.

It is clear that the STTC algorithm has captured real spatial relationships by having low complexity a greater number of contiguous (though a few were isolated), approximately-circular-in-shape clusters. One of these regions covers most of California, southern Nevada, Utah, and western Arizona, while another one starts at eastern Texas and stretches northeast along the Mississippi and Ohio rivers (see Figure 1.7a). By comparison, the CLST method may or may not pick up important synoptic systems because it displays spatial fields with high complexity, including more isolated and disconnected clusters. For example, there are spatial areas with large $f_\tau(\mathbf{s};\theta')$ located in northeastern Montana, southern Georgia, and central Florida (see Figure 1.7b).

To understand why $f(\mathbf{s};\theta)$ and $f_\tau(\mathbf{s};\theta')$ spatial representations have such different graphical properties, we consider the individual extreme event cases that we use to calculate these two frequencies. In the case of $f(\mathbf{s};\theta)$, we map these cases (Figure 1.8; see caption) to the extreme weather episodes described in National Weather Service and local weather station reports. Figures 1.8a and b show high PEF values in south-central California from the substantial rainfall and considerable snow accumulation in the mountains.

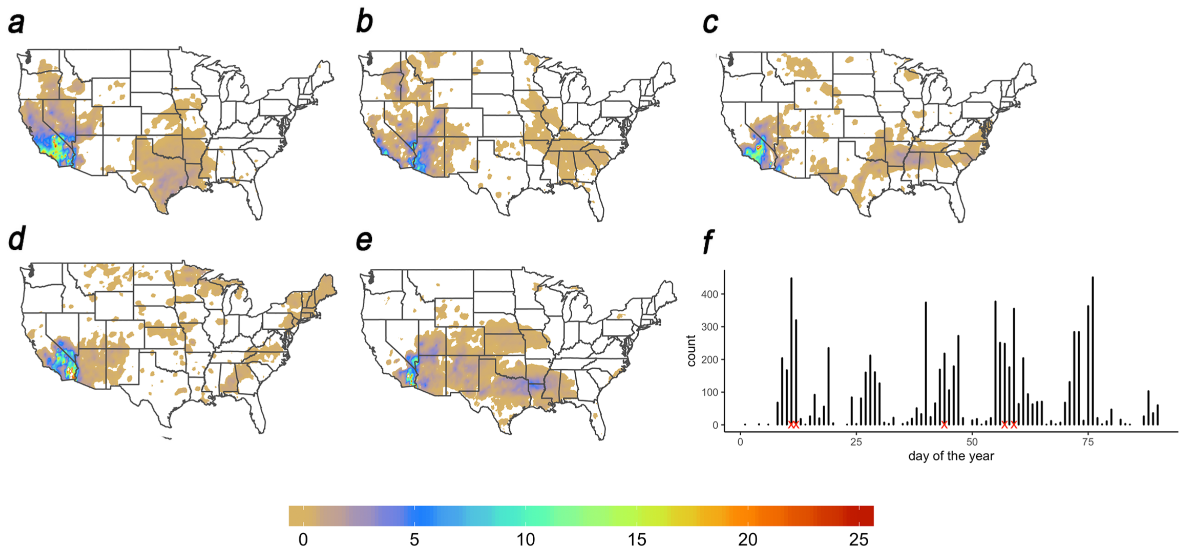

Figure 1.8: Actual days for the individual PEFs that were used to calculate $f(\mathbf{s};\theta)$ from Figure 1.7a: (a) 01/11/2001, (b) 01/12/2001, (c) 02/24/2001, (d) 02/28/2001, (e) 12/17/2001. (f) Barplot for frequency counts of the days of the year $f_\tau(\mathbf{s};\theta')$ from Figure 1.7b. Red crosses mark days from $f(\mathbf{s};\theta)$.

Two significant snowstorms in California are displayed in Figures 1.8c and d. Severe storms that spawn at least one tornado causing ample rainfall and flooding in northeastern Texas and Arkansas are shown in Figure 1.8e. Conversely, since $f_\tau(\mathbf{s};\theta')$ calculations are done for every grid cell, we present a frequency distribution of the extreme cases in Figure 1.8f. For every calendar day of the studied period from 1/1/2001 to 12/31/2001, we assign a corresponding sequential number from 1 to 90. We display this number on the x-axis. On the y-axis, we show a frequency distribution of only extreme days that we use in $f_\tau(\mathbf{s};\theta')$ calculation (Equation 1.9) across the entire spatial domain. High-frequency values imply a large contribution to $f_\tau(\mathbf{s};\theta')$ over $\mathbf{s}$, whereas low values imply a small contribution. The red crosses mark days for the extreme cases from $f(\mathbf{s};\theta)$ calculation (Equation 1.1). While the same days for the extreme cases appear in both frequencies, there were many other local extreme events present in $f_\tau(\mathbf{s};\theta')$ that were eliminated from $f(\mathbf{s};\theta)$ during the dimension reduction step. This explains why $f(\mathbf{s};\theta)$ and $f_\tau(\mathbf{s};\theta')$ have such different spatial representations in Figure 1.7.

Overall, while helpful for a location-specific analysis, the CLST method displays extremes at every non-empty grid cell and, in most cases, is deficient in geometric interconnectivity between spatial pixels. At the same time, it has similar shape and complexity indices to the STTC algorithm. These graphical properties highlight the main advantage of the methodology adopted here - it represents larger-scale weather patterns such as storms and high- or low-pressure systems more accurately than the CLST method by removing noisy data in both space and time, which is clearly desirable for spatial forecast verification of extreme events.

1.5 Conclusions and discussion

The lack of a universal definition of the concept of "extreme" makes it difficult to compare different scientific studies of extreme events. It also makes statistical inference much harder and obfuscates transparency and risk management processes. The current work was an attempt to formalize this important concept. We introduced a new generalized Spatio-Temporal Threshold Clustering method (STTC) for extreme events using an integrative modeling approach framework. Step by step, we described the interworking of an algorithm that is applicable to a wide variety of essential field quantities (EFQs). We used EFQ constructed from a gridded GHCN-Daily dataset comprised of 8516 stations from 1961 to 2016 across the CONUS. We applied a quantile-based dimension reduction methodology in space and time to identify extreme patterns for rainfall and droughts (Figure 1.1). These precipitation patterns can be asymmetric for some storms due to wind shear and are thus not fully detectable using conventional location-specific methods. To overcome this, we evaluated an entire spatial field comprised of relatively large-scale extreme patterns regardless of whether individual grid cells in space were extreme or not. Conditioned on these extreme patterns, the STTC method for the frequency of extreme EFQ allowed us to study not only individual local extreme episodes but also the interconnectivity, shape, complexity, and spatial extent of multiple extremes.

The threshold selection process, which is necessary for the conditional frequency calculation, was based on a multivariate series of geometric properties clustering analysis. We used the silhouette coefficient to measure the quality of the resulting clusters. In Tibshirani et al. (2001), a variety of other metrics were used to decide on the number of clusters for non-extreme events, and it was found that the silhouette method did not have the best performance. However, in our case, we performed clustering of the multivariate series of geometric properties for the extreme events, and we predetermined a number of clusters and set them to three.

We analyzed the output of the clustering algorithm, and we demonstrated that the *intermediate thresholds* cluster was the most suitable choice to deal with connectivity/area trade-off for analysis of extreme events when compared to the *low thresholds* and *high thresholds* clusters. In addition, we established that the STTC algorithm had more spatial connectivity and, in most cases, similar complexity and shape when compared to the CLST method. This finding is particularly important for spatial forecast verification of high-resolution climate models.

Our new threshold selection algorithm has a number of potential benefits. It objectively automates the threshold selection process through the clustering algorithm and can ultimately be used in conjunction with spatial forecast verification and modeling of extreme events. It is adaptable to model extremes with both high and low threshold choices. It incorporates spatial and temporal dependence in one holistic modeling framework, thus opening an opportunity for future analysis of statistical inference of extreme events for univariate and possibly multivariate EFQ in space and time, which is not currently possible using conventional location-specific methods. It is less sensitive to the data grid size when performing areal mean interpolation (Chen and Knutson, 2008) and can thus provide a consistent spatial pattern across different grid resolutions. It also links the chosen threshold to desired geometrical properties, which offers users more informed and flexible choices in selecting extreme events. The algorithm is relatively fast for large datasets and could be embedded within climate modeling systems (global or regional) to identify synoptic-scale spatial patterns that are linked to individual extreme episodes. However, it entails selecting a range of initial thresholds, which could be time-consuming for some users. The proposed threshold selection method requires spatially and temporally varying data and is not applicable in modeling location-specific single extreme events (Cattiaux and Ribes, 2018). The algorithm is quite stable to small changes in the range of the initial thresholds necessary for the time series clustering step. However, it is sensitive to the choice of $\lambda[\cdot]$ as well as to spatial and temporal quantile values, especially when the spatio-temporal data size is small. The user must decide what form of $\lambda[\cdot]$ is best suitable for the underlying application and what quantile values are capturing essential extreme episodes for the specific domain of interest.

The algorithm can be used to compare biases between global and regional climate models (Liang et al., 2012b) and possibly uncover the underlying physical mechanisms of these biases. Furthermore, an interesting perspective of this work is to determine whether or not one can capture dynamic and thermodynamic properties of the spatio-temporal structure of the conditional frequency of extreme EFQ from an ensemble of simulations in a single climate model with multiple physics configurations (Sun and Liang, 2020) or a set of different climate models (Min et al., 2011). Another potential application is to investigate how the observed spatio-temporal processes of the conditional frequency of extreme EFQ have changed and the causes of those changes, utilizing the most recent detection and attribution framework (e.g., Knutson et al., 2017; Ribes et al., 2017).

We foresee that our novel threshold selection approach could lead to new insights into spatial trend analysis in patterns of extreme EFQs. Figure 1.9 shows a linear downward trend for the connectivity index (C_{index}) of $f(\mathbf{s};\theta)$ from 1961–2016, where the conditional frequencies are objectively and independently derived for each year.

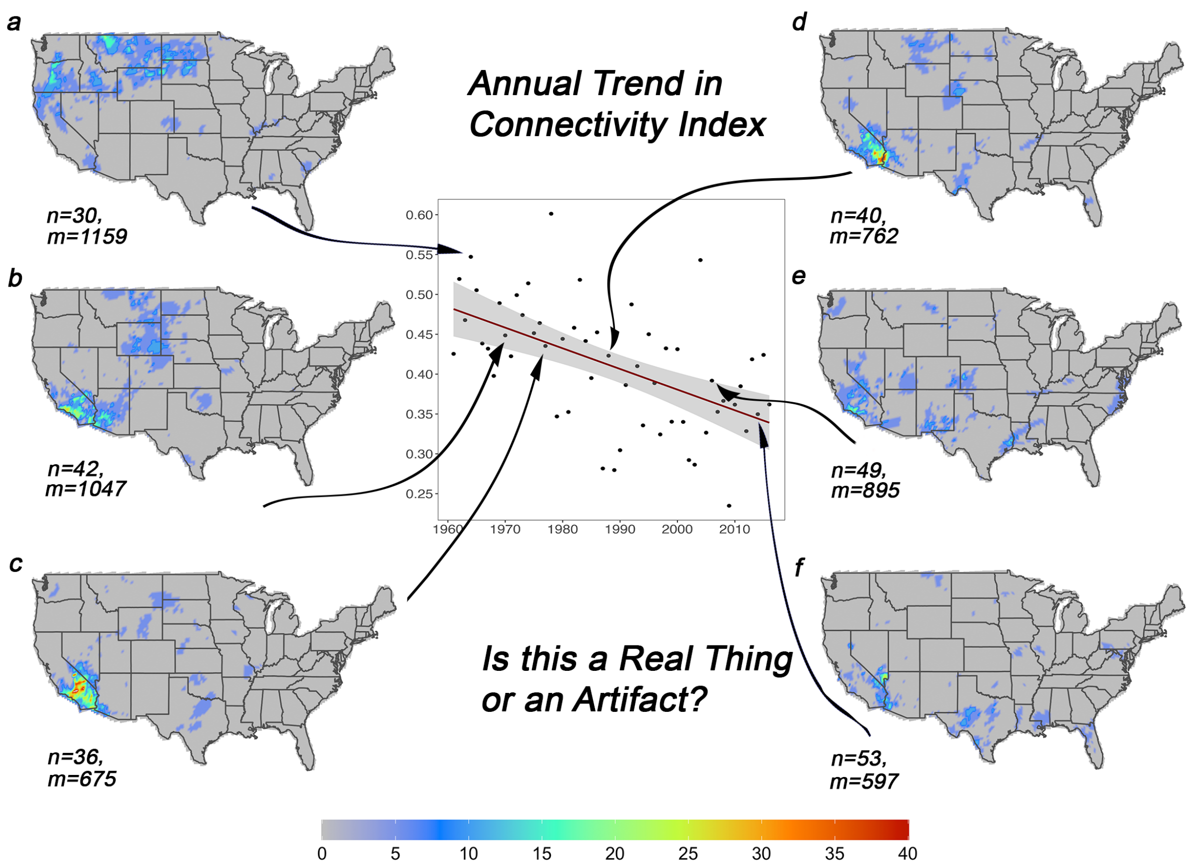

Figure 1.9: Annual linear trend for C_{index} of $f(\mathbf{s};\theta)$ from 1961–2016. Arrows show selected years, where n is the number of isolated structures, and m is the number of non-zero grid cells (see Eq. (1.3)). List of the years: (a) 1963, (b) 1970, (c) 1988, (d) 1998, (e) 2004; (f) 2014. All frequencies are stated in percent form.

The model fit ($intercept = 5.57 \pm 1.07, p << 0.001; slope = -0.003 \mp 0.0005, p << 0.001; R^2 = 0.29$) exhibited about the same variability between 1961–1988 (interquartile range ≈ 0.07) and 1989–2016 (≈ 0.08). The downward trend in C_{index} is primarily caused by the upward trend (not shown here) in the parameter n (i.e., the number of isolated structures). We found no trend in the parameter m (i.e., the number of non-zero grid cells) (see Equation 1.3). Figures 1.9a–f display $f(\mathbf{s};\theta)$ for selective years with decreasing n and corresponding m values. To the untrained eye, it is difficult to see any periodic changes in the number of isolated structures. However, the numerical values we supplied should substantiate our claim.

An important caveat of C_{index} is that it is a binary index. As such, it only calculates all components based on the number of non-zero grid cells in the spatial patterns by setting values below a threshold to zero and those above to one. It does not provide any information about the original raw intensities. However, evaluation of this index at different threshold values can provide some information about error intensity (Gilleland, 2017a).

The upward trend in the number of isolated structures implies that the distribution of conditional frequency of extreme precipitation is changing from large-scale, well-connected spatial patterns to smaller-scale, less contiguous rainfall clusters. In order to understand why we observe the distributional change, it is important to establish the linkages between changes in circulation and other essential climate variables (e.g., temperature and precipitation). If the trend continues, we expect to see more localized droughts and heatwaves, especially during the summer months. It would also create a tremendous challenge for weather forecasting and climate adaptation efforts as the forecast skill drops significantly at smaller scales. Thus, the STTC offers a unique way to perform diagnostic analysis of the large-scale weather patterns specifically associated with these extreme episodes, and such analysis for the detected trend is in progress.

Chapter 2: A review of statistical models for extreme value analysis

In this chapter, we introduce a new statistical modeling framework for spatial extreme events. Since the type of conditional spatial model proposed here estimates spatial dependence somewhat differently than the classical spatial extreme value models, it is beneficial first to provide a high-level overview of extreme value statistical methodologies applicable to univariate and multivariate extreme value analysis (EVA) to understand their conceptual differences and limitations. Following this, our proposed model and overall modeling framework are described in detail. The practical applications of this framework will be discussed in the next chapter.

2.1 Univariate EVA background

EVA is a sub-field of statistics that focuses on analyzing the very rare values of a process, such as the maximum or values near the maximum. The theory for univariate EVA is well established, but spatial EVA is a much more active area of research. A very brief background follows where considerably more detail can be found in the texts Beirlant et al. (2004); Coles (2001b); de Haan and Ferreira (2006); Reiss and Thomas (2007), as well as in Gilleland and Katz (2016); Katz et al. (2002); Palutikof et al. (1999); Smith (2003).

There are three main modeling approaches available in univariate EVA, and all of them are interrelated. One is to model the maxima over long blocks of time where the generalized extreme value (GEV) distribution function (df) has theoretical support for its use. Another is to model excesses over a high threshold conditioned on the variable of interest having exceeded the threshold. In this setting, the generalized Pareto (GP) df is appropriate, which approximates the tail part of the GEV df. The third approach is to model the extreme values as a point process (PP), which can be performed either as a marked Poisson process where the marks are GP distributed or as a two-dimensional PP where the frequency and intensity components are modeled simultaneously. The latter approach can be re-characterized as a GEV df, and if the temporal information about the data is available, then it is possible to characterize the PP as a GEV df in terms of the same blocks (usually taken to be annual) as the block-maxima approach. When utilizing the GP or PP frameworks, an event is determined when a variable exceeds a high threshold; collectively, these two approaches are known as peaks over threshold (POT) methods.

For independent and identically distributed random variables $X_1, \ldots, X_{mn}$ that follow a df F, called the parent df, the GEV df is used to model $Z_1 = M_{1n} = \max\{X_1, \ldots, X_n\}, \ldots, Z_m = M_{mn} = \max\{X_{mn-n+1}, \ldots, X_{mn}\}$, and is given by

$$\mathbb{P}\mathrm{r}\{Z \leq z\} = G(z) = \exp\left[- \left\{ 1 + \xi \left(\frac{z-\mu}{\sigma} \right) \right\}^{-1/\xi} \right], \tag{2.1}$$

defined when $1 + \xi(z-\mu)/\sigma > 0$, $-\infty < \mu < \infty$, and $\sigma > 0$, $-\infty < \xi < \infty$. Equation 2.1 encompasses three types of df's depending on the sign of the shape parameter, ξ. The heavy-tail Fréchet df results when $\xi > 0$, the upper-bounded Weibull df when $\xi < 0$ with upper bound at $\mu - \sigma/\xi$ and when $\xi = 0$ defined by continuity, Eq (2.1) is the light-tailed Gumbel df given by $G(z) = \exp[-\exp\{-(z-\mu)/\sigma\}]$. The parameters μ and σ are the location and scale parameters of the GEV df, respectively.

Of particular interest for the GEV df are its quantiles, z_p where $G(z_p) = 1-p$, because they can be interpreted as the value expected to be exceeded on average once every $1/p$ periods (again, usually a period is one year). Unlike the normal df, these quantiles can be solved analytically from Equation 2.1. Of interest in this work, however, is the GP characterization for extreme values. The GP df is given by

$$\mathbb{Pr}\{X - u > x | X > u\} = H(x) = \left[1 + \xi\left(\frac{x-u}{\sigma_u}\right)\right]^{-1/\xi}, \tag{2.2}$$

defined where the part in brackets is positive and where u is a high threshold, the scale parameter, $\sigma_u > 0$, depends on the threshold choice, and the shape parameter, $-\infty < \xi < \infty$, determines which of three types of df describes the threshold excesses. In this case, the threshold u, which is not usually taken to be a parameter, plays the role of the location parameter leaving only two parameters for the GP df. If $\xi > 0$, then Equation 2.2 is heavy-tailed Pareto, and if it is less than zero, it is a bounded upper tail beta df. Finally, when $\xi = 0$ is again defined by continuity, the result is a light-tailed exponential df.

It is more convenient to interpret extreme value models through quantiles or return levels rather than analyzing individual parameter values. Estimating the return levels for a GP df is not as simple as merely determining its quantiles because it is also necessary to estimate the frequency of an event's occurrence, ζ_u, with the u subscript to emphasize its dependence on the threshold. The value x_k that is exceeded on average once every k observations for the GP df is given by

$$x_k = \begin{cases} u + \frac{\sigma_u}{\xi}\left[(k\zeta_u)^{\xi} - 1\right], & \xi \neq 0, \\ u + \sigma_u \log(k\zeta_u), & \xi = 0. \end{cases} \tag{2.3}$$

In the case of a Poisson random variable, N, defined to be the number of *events* determined by when $X > u$, $\mathbb{P}\mathrm{r}\{N = 0\} = \mathbb{P}\mathrm{r}\{\max(X_1, \ldots, X_{mn}) \leq u\}$, which illustrates the connection between the Poisson POT approach and the block maxima type of analysis. The connection can also be seen by examining the forms in Equation 2.1 and (2.2) where $H(x)$ is very similar to the exponent part of $G(z)$ except with u instead of μ. Further, if $Z_1, \ldots, Z_m$ follow a GEV df with parameters μ, σ and ξ, then the corresponding GP df for $X_1, \ldots, X_{mn}$ has parameters $\sigma_u = \sigma + \xi(u - \mu)$ and ξ.

Parameter estimation for fitting the GEV or GP df's to data can be performed in a variety of ways, with maximum-likelihood estimation (MLE) arguably the most popular. Collectively, the df's (2.1), (2.2), and the PP characterization thereof are known as the extreme value distributions (EVD's).

The Poisson df is a very important df for EVA as it ties the entire theory together because it approximates the binomial df under the condition that its rate parameter remains constant as the threshold defining an event increases towards infinity. The exponential df facilitates probabilities of the waiting time between events for the Poisson df, and this df has the peculiar memoryless property, which means that if an event occurs or does not occur and it is desired to estimate the probability of another event occurring, that probability is independent of the previous event's occurrence. That is, $\mathbb{P}\mathrm{r}\{X_{t+1} > u | X_t > u\} = \mathbb{P}\mathrm{r}\{X_{t+1} > u\}$.

Moreover, the GEV df, and subsequently the extreme value distributions more generally, are related to the Poisson df and possess either a max-stability property in the case of the GEV or a similar POT-stability property in the case of the POT models. These stability properties are similar to the memoryless property of the exponential df and are important components of the theoretical justification for their use in making inferences about extreme values. For example, the max-stability property essentially states that the exact df for the maximum based on the parent df after appropriate scaling and transposition is again the same df with appropriately transformed parameters. In other words, the exact df for $\mathbb{P}\mathrm{r}\{\max(X_1, \dots, X_n) \leq x\}$ is $F^n(x)$, but this df is difficult to estimate and quickly tends to zero as the events become rare. Additionally, since F must be estimated, small errors are raised to the n-th power, leading to large errors. Therefore, the game is to find sequences of constants $a_n > 0$ and b_n such that $F^n(a_n x + b_n)$ tends to a non-degenerate df as $n \longrightarrow \infty$, and that df is the GEV df, and the only df's for which this max-stable property holds is the GEV family of df's.

Given the relationship of the EVD's with the Poisson and this relationship with the exponential df, it should not be surprising that estimating multivariate and, subsequently, spatial extreme values is difficult. The appropriate df's essentially impose an independence criterion on the processes of interest, but many such processes are nevertheless dependent on their extreme values. Indeed, the classical multivariate extreme value theory begins by modeling the extremes of the different variates taken separately. The next two sections give a brief overview of classical multivariate extreme value theory. The material is germane to many of the spatial modeling approaches that have been proposed in the literature.

2.2 Bivariate EVA

For completeness, a brief review of bivariate EVA is given. The extension to multiple dimensions is relatively straightforward, though difficult to implement if the dimension is even moderately large. In the bivariate case, much of the theory requires first deciding between the type of dependence that the two series exhibit: asymptotic independence or dependence. Classical bivariate EVA uses asymptotically dependent models (cf. Coles and Tawn, 1991; de Haan and de Ronde, 1998). For higher dimensions, it is more complicated, and for example, Wadsworth and Tawn (2013) summarized k-dimensional joint tail dependence using $\sum_{i=0}^{k-2}\binom{k}{i}$ limits of the form $\chi = \lim_{u \to 1} \mathbb{P}\mathrm{r}\{F_i(X_i) > u | F_j(X_j) > u\}$, where $X_i \sim F_i, X_j \sim F_j$. In general, empirical estimates of χ for two sequences of random variables is often used to summarize extremal dependence between them.

Many methods for modeling bivariate extreme values jointly have been proposed, but as is evidenced by the form of χ, they effectively constitute a copula-type of modeling procedure where usually only the extremes of the variables are used, and a copula that is appropriate for extreme values is chosen; although exceptions exist in the literature, they are not generally considered proper for EVA. In most cases, it is necessary, or at least convenient, to first transform the variables so that they both follow the same marginal df, which puts them on the same scale.

In a geostatistical context, for a spatial random field $Z(\mathbf{s})$ indexed by spatial coordinates, $\mathbf{s} \in \mathcal{D}$, a version of the structure function (e.g., the variogram is a special case; cf. Harris et al., 2001; Yaglom, 1987) is the madogram, which for a spatial lag h is defined as

$$\nu(h) = \frac{1}{2}E|Z(\mathbf{s}+h) - Z(\mathbf{s})|. \tag{2.4}$$

In the asymptotic dependence case, arbitrary norms are used to separate the transformed pair of random variables into their pseudo radial $R = \|X_i, X_j\|_p$ and angular $\mathbf{W} = (X_i, X_j)/R$ components, which become independent in the limit $\lim_{c\to\infty} \mathbb{P}\mathrm{r}\{\mathbf{W} \in B, R > c(\eta+1)|R > c\} = H(B)(\eta+1)^{-1}$, where $B \subset \mathcal{S}^p := \{\boldsymbol{\omega} \in \mathbb{R}^2_+ : \|\boldsymbol{\omega}\|_p = 1\}$ for continuity sets of the limit measure H. While this limit holds for asymptotic independence as well, $H(\cdot)$ is a discrete two-point df placing atoms of probability on the endpoints of the continuous arc, $\mathcal{S}^p$ because, under this paradigm, the absolute largest values of X_i and X_j occur singly so that all of the mass is pushed to one or the other. Subsequently, the model is not useful in this setting. Ancona-Navarrete and Tawn (2002) introduced a measure of pairwise extremal dependence for spatial processes that are marginally invariant, which enables decisions to be made about whether a spatial process is independent, asymptotically dependent or asymptotically independent. Bacro et al. (2010) used the madogram to derive a statistic in order to test for asymptotic independence of bivariate maxima vectors.

A max-stable process (e.g., de Haan, 1984; Ribatet, 2013) provided for a spatial analysis framework that mirrors the more customary geostatistical approach for averages where a covariance (or variogram) model is fit to pair-wise data typically as a function separated by lags, h. Instead of fitting a variogram or covariance model, a max-stable model is fit to the madogram of Equation 2.4. The approach utilizes pairwise likelihoods as a surrogate to fit the entire multivariate df to the data.

2.3 Multivariate models for EVA

Buishand et al. (2008), Turkman et al. (2010), Cooley et al. (2012a), Davison et al. (2012), Ribatet (2013) and Huser and Wadsworth (2022) all gave thorough reviews of spatial EVA, so only a brief review is given here for clarity and how the proposed model fits into the overall landscape. Analyzing extreme values spatially is a challenging problem and an area of active research. Specific model choice is largely determined by the aim of the spatial analysis, such as (i) making inferences about return levels over space while incorporating spatial information either for pooling data to reduce the variance of the estimates or to account for physical dependences that could lead to extreme values at multiple locations, (ii) analyzing extreme value behavior over space, (iii) summarizing extreme values across space, (iv) interpolating (or extrapolating) extreme values at unobserved spatial locations.

Perhaps the simplest form of the spatial extreme value model involves varying the parameters of one of the EV df's described in Section 2.1. The approach is actually a form of vector generalized linear/additive modeling (VGLM/VGAM, cf. Diggle and Ribeiro, 2007; Gregersen et al., 2013; Staid et al., 2014; Yee and Stephenson, 2007; Yee and Wild, 1996). Gilleland et al. (2006) employed such a model where they impose a Gaussian process on the scale parameters of the GP df and pool the stations to have a single shape parameter, which is the most difficult parameter to estimate. Beguería and Vicente-Serrano (2006); Bocci et al. (2012) used a similar approach for high-magnitude rainfall. Hanel et al. (2009) allowed some parameters to vary while others are held constant over space in order to pool data in a similar fashion as is achieved with the index flood assumption in hydrology (cf. Fowler et al., 2005). Jonathan and Ewans (2011) applied a similar approach with the non-homogeneous Poisson process with the inclusion of directional information. Katz et al. (2003) borrowed strength spatially by constraining the transition matrix of a (temporal) Markov Chain process to be the same for all sites within a region.

A popular, more robust extension of the VGLM/VGAM approach is to employ a Bayesian hierarchical model (BHM), which allows for considerable flexibility for adventitious features, for example, for tracking spatial variation of marginal df's Casson and Coles (1999); Coles and Casson (1998); Cooley et al. (2006a, 2007); Cooley and Sain (2010); Eastoe (2009); Fawcett and Walshaw (2006); Gaetan and Grigoletto (2007); Heaton et al. (2011); Hrafnkelsson et al. (2012); Kottas and Sansó (2007); Mendes et al. (2010); Oliver et al. (2014); Sang (2008); Sang and Gelfand (2009); Schliep et al. (2010); Tye and Cooley (2015). These models made a conditional independence assumption that is not always met so that they cannot always reproduce simulations of extreme events (cf. Davison et al., 2012; Sang and Gelfand, 2010; Thibaud et al., 2016; Turkman et al., 2010). Moreover, the dependence modeled is on the parameters of the marginal df's and not at the process level, so that dependence information may be lost. One approach to counter this issue is to utilize a hierarchical model that also incorporates dependence at the process level. Sang and Gelfand (2010) and Fuentes et al. (2013) employed a Gaussian copula model (cf. Coles et al., 1999b; Nelsen, 2006; Poon et al., 2004; Schlather and Tawn, 2003) in order to capture the smoothness of the process level at the finer scales. The Gaussian copula may not be able to capture the dependence at the most extreme levels as it is asymptotically independent (Davison et al., 2013). A similar framework was used by Ghosh and Mallick (2008).

Ribatet et al. (2012) and Shaby (2014) utilized composite likelihood Varin et al. (2011) functions within a Markov chain Monte Carlo (MCMC; see also Huerta and Sansó, 2007) algorithm with Bayesian inference using max-stable models. Shaby and Reich (2012) proposed a max-stable spatial hierarchical model that incorporated a logistic copula (cf. Stephenson, 2009) that is generally more appropriate for extreme values than the Gaussian. Thibaud et al. (2016) introduced a Bayesian hierarchical model based on the Brown-Resnick process (e.g., Engelke et al., 2014; Huser and Davison, 2013) using the joint distribution of component-wise maxima and their partitions in terms of individual events (Stephenson and Tawn, 2005). Oesting et al. (2017) introduced a bivariate framework within a spatial Brown-Resnick process.

Of course, max-stable processes Blanchet and Davison (2011); Coles (1993);

Davis and Mikosch (2008); Davison and Gholamrezaee (2011); de Haan (1984); de Haan and Pereira (2006); de Haan and Zhou (2008); Genton et al. (2015); Hainy et al. (2015); Huser and Davison (2014); Kabluchko et al. (2009); Padoan et al. (2010); Schlather (2002); Shang et al. (2011); Smith and Stephenson (2009); Smith (1990); Strokorb et al. (2015); Wang and Stoev (2010) are a popular choice for modeling extreme values spatially (e.g., Asadi et al., 2015; Buhl and Klüppelberg, 2015; Coles and Tawn, 1996; Dieker and Mikosch, 2015; Dombry et al., 2013; Genton et al., 2011; Koch, 2017; Olinda et al., 2014; Stephenson et al., 2014; Thibaud and Opitz, 2015; Westra and Sisson, 2011). Coles and Walshaw (1994) proposed a max-stable process model appropriate to the topology of a circle in order to additionally account for directional dependence. Huser and Genton (2016) developed non-stationary max-stable dependence structures where covariates can be easily incorporated with inference using pairwise likelihoods. Davis et al. (2013a,b) extend the framework to incorporate time as well as space. Reich and Shaby (2012) proposed a random effects model that accounts for spatial dependence and demonstrated that its distribution leads to a max-stable process with the Gaussian extreme value process as a limiting case. Typically, max-stable processes are for use with block maxima data, but more recent work has shown that it can be applied with the POT approach (cf. Thibaud et al., 2013). Davison et al. (2013); Wadsworth and Tawn (2012) employed max-stable processes to accommodate asymptotic dependence and inverted max-stable processes for asymptotic independent cases. Zheng et al. (2015) compared the performance of the independence method (i.e., not accounting for spatial dependence) against three different max-stable process dependence models appropriate for spatial modeling on marginal parameter estimation, and they concluded that the independence method is preferred when only the marginal distributions are of interest.

Wadsworth and Tawn (2014) proposed methods for circumventing issues with max-stable processes by considering random fields in the domain of attraction of the class of max-stable processes constructed through manipulation of log-Gaussian random functions. They exploited the d-dimensional multivariate Poisson process with a censoring scheme in order to also include information below the threshold so that only one component of the multivariate vector needs to be extreme. They also used the methods developed in Stephenson and Tawn (2005) that use occurrence times to simplify the likelihood.

Aberg and Guttorp (2007) studied the distribution for extremes of ambient air quality standards at unobserved locations by a Rice method (Azais and Wschebor, 2005; Mercadier, 2006) relying on a generalization of upcrossings of a level in one to two dimensions. Mannshardt-Shamseldin et al. (2010) compared the results between return levels derived from GEV df's fitted to point locations against those fitted to gridded data and developed a family of regression relationships between them while also accounting for spatial variation.

Apputhurai and Stephenson (2013) incorporated temporal trends for precipitation maxima in the location and scale parameters of the GEV df and incorporate spatial features through anisotropic Gaussian random fields while also incorporating site-specific explanatory variables. Cheng and Schwartzman (2015) derived a general formula for the distribution of the height of local maxima for a smooth, non-stationary, Gaussian random field. Economou et al. (2014) introduced a flexible spatio-temporal model for analyzing extreme extra-tropical cyclones using spatial regularization with latitude as a covariate along with spatial random effects.

Bonazzi et al. (2012) applied extreme value copula modeling to express changes in climate variability in terms of storm frequency, local intensity, and the spatial structure of storms in order to assess risk in Europe from extreme winds. Föllmer (2014) investigated convex risk measures in the spatial setting for financial data. Cooley et al. (2006b) proposed different estimators of extremal coefficients for stationary max-stable random fields with the aim of capturing the spatial structure using the madogram. Naveau et al. (2009); Vannitsem and Naveau (2007) assessed pairwise dependence of temporal maxima in a similar way.

Bador et al. (2015); Bernard et al. (2013) applied multivariate EVA to cluster analysis in space. Burauskaite-Harju and Grimvall (2013) employed traditional multivariate extremes methods to time-lagged data in order to diagnose risk for events that move across time. Butler et al. (2007) employed extreme value methods in a local likelihood approach to provide a non-parametric description of temporal and spatial variations in the magnitude and frequency of storm surge events. Coelho et al. (2008) used neighborhood pooling of data along with extreme value models and potentially explanatory variates on the distributions for extreme events of climate variables, including an investigation of clustering and spatial dependence of extremes. French and Davis (2013) used analogs of the conditions of Hsing et al. (1996) to show that the maxima of two-dimensional Gaussian random fields exhibit clustering in the limit.

Northrop and Jonathan (2010) considered setting a covariate-dependent threshold using quantile regression could be compatible with extreme value parameterizations, and they adjust the uncertainty for spatial dependence using polynomial basis functions. Reich et al. (2013) investigated ground-level ozone control strategies by modeling their behavior under different emission scenarios from a reduced form model using quantile regression for the center of the distribution and EVD's for the tail, incorporating space by way of a BHM.

Rietsch et al. (2013) investigated network design for impact studies from hydrology and climatology using EVT and machine learning via neural networks and a "Query-by-Committee" approach.

Huser et al. (2017) proposed flexible Gaussian scale mixture models within a parsimonious parametric copula to smoothly interpolate from asymptotic dependence to independence.

Other relevant spatial models that attempt to make inferences about larger values away from the mean, but not necessarily extreme by EVA standards, include: Craigmile et al. (2005); Gilleland and Nychka (2005); Reich et al. (2011); Zhang et al. (2008).

More recently Rootzén et al. (2018a,b) introduced multivariate GP distribution as an extension of univariate GP distribution and demonstrated how conditioning of the extreme events on threshold exceedance results in four basic representations of the GP distribution family.

2.3.1 Limitations

The classical models of spatial extremes mentioned earlier have a number of limitations:

1. Computationally expensive for high dimensions;
2. Inability to capture a wide range of dependence structures;
3. Noticeable tail dependence even as the difference between locations increases (i.e., it should be more independent with increased distance).

Heffernan and Tawn (2004) (henceforth, HT2004) addressed some of these limitations in their conditional extreme model, which we describe next.

2.4 Heffernan and Tawn conditional EVA model

The conditional EVA model introduced by HT2004 and further clarified by Heffernan and Resnick (2007) is a natural model for the PQ framework because it provides a model for the dependence of the process at each spatial location, i.e., $X_1, \ldots, X_n$ conditioned on Y such that $Y > u$, where u is a very high threshold. In particular, for suitably transformed random vectors $\mathbf{X} = X_1, \ldots, X_n$ and random variable Y assume that

$$\mathbb{P}\mathrm{r}\left\{ \frac{X_1 - a_1(y)}{b_1(y)} \leq z_1, \ldots, \frac{X_n - a_n(y)}{b_n(y)} \leq z_n, Y - u > y \middle| Y > u \right\} \underset{u \to \infty}{\longrightarrow} G(\mathbf{z}) \exp(-y), \tag{2.5}$$

where $\mathbf{z} = z_1, \ldots, z_n$ and $\lim_{z_i \to \infty} G(z_i) = 1$ for all i. Model (2.5) is henceforth called the HT model. By *suitably transformed* X and Y, any appropriate transformation may be used. Here, a Laplace transformation is employed when using the HT model.

When using the Laplace transformation, the models for $a_i(y)$ and $b_i(y)$ have a simple parametric form regardless of whether X_i is positively, negatively or not associated with Y in the tail (Keef et al., 2013a,b). In particular,

$$a_i(y) = \alpha_i y, \text{ and } b_i(y) = y^{\beta_i} \text{ for } y > u, \tag{2.6}$$

which means that there are two parameters, $-1 \leq \alpha_i \leq 1$ and $-\infty < \beta_i < 1$, that must be estimated for each location $\mathbf{s}_i$. In the PQ modeling framework, it means that $2n$ parameters must be estimated for the HT dependence part of the model, where n is the number of spatial locations. The model (2.6) was shown by HT2004 to be valid for variables from a wide family of copula dependence models (including all of those investigated in Joe, 1997; Nelsen, 2006).

The Laplace transformation is achieved for a variable X^* by $\log 2\hat{F}(X^*)$ if $X^* < \hat{F}^{-1}(1/2)$ and by $-\log[2(1-\hat{F}(X^*))]$ when $X^* \geq \hat{F}^{-1}(1/2)$, where $\hat{F}$ is the estimated df for all values (extreme or not) of X. HT2004 suggested a hybrid model (Coles and Tawn, 1991, 1994) achieved by fitting a GP to excesses over a high threshold, u, so that

$$\hat{F}(x) = \begin{cases} 1 - (1 - \tilde{F}(x))(1 + \xi(x-u)/\sigma)^{-1/\xi} & \text{for } x > u \\ \tilde{F}(x) & \text{for } x \leq u \end{cases} \tag{2.7}$$

where $\tilde{F}(x)$ is the empirical df. If $\mathbf{X}$ is the Laplace-transformed random vector of $\mathbf{X}^*$, then the marginal df for a given variate X_i is

$$F_{X_i}(x) = \mathbb{P}\mathrm{r}\{X_i < x\} = \begin{cases} \exp(x)/2 & \text{for } x < 0 \\ 1 - \exp(-x)/2 & \text{for } x \geq 0 \end{cases}$$

so that both upper and lower tails of X_i are exactly exponentially distributed with mean 1.

A result of the assumption (2.5) combined with the Laplace transform so that the result (2.6) holds for both types of association is that for Laplace-transformed $\mathbf{X}$ and $Y > u$

$$X_i = \alpha_i Y + Y^{\beta} Z_i \tag{2.8}$$

where Z_i are the residuals from $(X_i - \alpha_i Y)/Y^{\beta_i}$, which follow the df G, which does not have a simple closed-form expression unless a specific dependence model is assumed, but with knowledge of the marginal GP parameters for X_i^*, σ_i, ξ_i for each spatial location $i = 1, \ldots, n$, and for the conditioning variable (here, the PEF) Y, σ_Y and ξ_Y, as well as α_i and β_i, it can be simulated from.

The parameters α_i and β_i determine the dependence structure between each X_i and Y. Generally, positive association implies $\alpha_i \in (0, 1]$ and negative association $\alpha_i \in [-1, 0)$. In the special case of independence, $\alpha_i = \beta_i = 0$, $i = 1, \ldots, n$ and G factorizes into $n-1$ Laplace df's. If there is asymptotic positive dependence between X_i and $Y|Y > u$, then $\alpha_i = 1$ and $\beta_i = 0$, and for negative asymptotic dependence, $\alpha_i = -1$ and $\beta_i = 0$. For asymptotic independence between X_i and $Y|Y > u$, then

$-1 < \alpha_i < 1$.

2.4.1 Estimation and inference

HT2004 initially proposed a complicated but sound, semi-parametric estimation and inference strategy for their model. The marginal df's employ the hybrid model (2.7) so that the df for values below the threshold are estimated non-parametrically via the empirical df to obtain $\tilde{F}_{X_i^*}(x)$ for each spatial location $i = 1, \ldots, n$ and similarly $\tilde{F}_{Y^*}(y)$ for the conditioning variate, Y^*. For the upper (or lower) tail, the GP df is fit to excesses over an appropriately high threshold u, usually using standard maximum likelihood estimation (MLE), which requires numerical optimization. Using these estimates for $F_{X_i^*}(x)$ and $F_{Y^*}(y)$, a transformation is then applied, which is free of any further parameters to be estimated, in order to obtain X_i and Y. Again, any appropriate transformation can be applied, but following Keef et al. (2013a,b), a Laplace transformation is used in this work in order to ensure that Equation 2.6 holds regardless of the type of association each X_i has with Y.

The next part of this estimation scheme utilizes representation (2.8) and requires a false working assumption that each Z_i is independent and normally distributed. Keef et al. (2013a) also tested using a Laplace df assumption in place of the normal but found the normal df to perform better in general. If Z_i has mean η_i and variance ϕ_i^2, then the conditional (on $Y > u$) means and standard deviations for each X_i are $\alpha_i Y + \eta_i Y^{\beta_i}$ and $\phi_i Y^{\beta_i}$, respectively. MLE can now be used to estimate

α_i, β_i and the nuisance parameters η_i and ϕ_i. Given these estimates, non-parametric estimation of G using the empirical joint df on the residuals $\mathbf{z} = (\mathbf{x} - \boldsymbol{\alpha} y)/y^{\boldsymbol{\beta}}$, $y > u$.

Estimation of probabilities for multivariate extreme events, conditional on $Y > u$, is obtained by way of Monte Carlo methods using the opening assumption (2.5). Inference is carried out for these procedures using a semi-parametric bootstrap procedure.

2.4.2 Advantages, limitations, and applications

The main advantages of the HT model are as follows:

(1) the model is capable of handling complex multivariate extreme value distributions;

(2) it takes into account the variety of dependence structures between variables, and it applies regardless of whether the variables are asymptotically dependent or asymptotically independent;

(3) the model can be used to estimate the conditional distribution of a multivariate variable given one of its components is large;

(4) it is practical as it can handle large datasets in space and time dimensions.

However, there are also some limitations to the HT model, such as:

(i) the need for large amounts of data to estimate the parameters of the model accurately;

(ii) the sensitivity of the model to the choice of the conditioning variables;

(iii) the model assumes that the joint distribution of the variables is stationary, which may not hold in practice.

The HT model has been successfully applied in environmental applications to gauge the risk of extreme events. For example, Mendes and Pericchi (2008) employed the HT conditional model to assess flooding risk from five rivers. Keef et al. (2009b) introduced a methodology that extends the HT approach by adopting an asymptotically justified dependence model over sites and time that accounts for time lags between variables at different sites (see also, Keef et al., 2009a, 2013b). Jonathan et al. (2014b) used the HT conditional EVA model to estimate design contours by way of a constant probability of exceeding a high threshold using measured and hindcast data from the Northern North Sea, Gulf of Mexico, and the North West Shelf of Australia. They quantified their uncertainties using a bootstrap analysis. Jonathan et al. (2014a) utilized the HT conditional model with covariate information included. More recently, Wadsworth and Tawn (2022) proposed the spatial conditional extreme value model as an extension of the HT model to continuous space. Although this model is more parsimonious than the HT model, the computation of the d-dimensional Gaussian process as part of the likelihood makes it computationally prohibitive for more than several hundred spatial locations.

Overall, the conditional HT model is a valuable tool for modeling multivariate extreme value distributions, but it is essential to thoroughly evaluate its limitations and potential pitfalls when applying it to real-world data.

In our application of the HT model to the trend analysis, we account for potential non-stationarity over time by adopting a fixed-size rolling window approach. At each time step, we estimate the model parameters. If the parameters remain constant throughout the entire sample, the estimates obtained from the rolling window analysis should be relatively similar. However, if the parameters change at any point in time, the rolling estimates should detect this instability.

To the best of our knowledge, the HT model has never been used with the rolling window approach before. Another innovation in our work is that we utilize the HT model with the STTC algorithm within the scope of a new modeling framework to project conditional probabilities and return levels for extreme EFQ of the daily precipitation over the CONUS domain. The processing time of the HT model in such a setting is improved because we run the model after the dimension reduction step in the STTC algorithm. Essentially, instead of running the HT model with the entire dataset, which could be a time-consuming process, we only use PEF fields as input for the model. The succeeding section introduces this new framework.

2.5 Propinquity modeling framework

Most spatial statistical models directly measure spatial dependence between variables at different spatial locations, usually by distance separation or a Markov process. The current work is distinct from previous research because it examines the spatial aspect of EFQ conditioned on the occurrence of extreme events somewhere in the field. Although some spatial fields may not encounter any extreme events over time, applying PEF/NEF concepts suggests one or more extreme regions will exist.

We refer to this modeling technique as the Propinquity (henceforth, PQ) modeling framework. The term propinquity is mostly used in the field of social psychology, where it is considered a primary driver for interpersonal attraction. For example, neighbors who live near the staircase tend to have friends from other floors than those living further from the staircase, whereas neighbors otherwise have more familiarity with neighbors from the same floor (`https://en.wikipedia.org/w/index.php?title=Propinquity&oldid=834329079`, last visited 10 August 2023). Its use in this framework emphasizes the difference in how it measures spatial dependence where the PEF or NEF serves as the staircase through which the modeling of spatial dependence at the process level of extreme EFQ is facilitated. We further formalize this modeling framework with the following definition.

Definition 2.5.0.1. Let $\mathcal{V}_t(\mathrm{Z}(\mathbf{s_1}, t), \ldots, \mathrm{Z}(\mathbf{s_n}, t))$=$\mathrm{Q_t}$ be the univariate time series variable measuring the PEF of the propinquity model, where $\mathrm{Z}(\mathbf{s_1}, t), \ldots, \mathrm{Z}(\mathbf{s_n}, t))$=$\mathbf{Z}(\mathbf{s}, t)$ are the spatial fields at time $t = [1, \ldots, T]$ and indexed by spatial coordinates $\mathbf{s} \in \mathcal{D}$, the spatial domain. Call $\mathrm{Q_t}$ the staircase variable for brevity and ease of understanding. Then the propinquity model models the spatio-temporal fields conditional on $\mathrm{Q_t} > \theta_\mathrm{h}$ for some high threshold θ_h. In general, the propinquity model is

$$\mathbb{P}[\mathrm{Z}(\mathbf{s_1}, t) \leq z_1, \ldots, \mathrm{Z}(\mathbf{s_n}, t) \leq z_n, \mathrm{Q_t} \leq z | \mathrm{Q_t} > \theta_\mathrm{h}] \tag{2.9}$$

The propinquity threshold is generally chosen so that the univariate generalized Pareto model can be reasonably applied, or it can be chosen based on the graphical properties of the extreme events as in STTC.

Empirically, the subset of time points from $\mathbf{Z}(\mathbf{s}, t)$ determined to be those times when $\mathcal{V} > \theta_{\mathrm{h}}$ is chosen, and any distributional properties can be analyzed, such as the frequency that each individual location exceeds a high value as in Gilleland et al. (2016b), the mean median or other quantiles at each location. However, in order to allow for extrapolation to extreme values that have not yet been seen, as well as to obtain uncertainty information about, for example, changes in the extreme values over space and/or time, a model is needed, and the HT model is a natural fit in this setting.

The PQ modeling framework aims to capture spatial dependence by examining the relationship between the process variable at each grid cell within the PEF(NEF) over time. This relationship is expressed as a function of all spatial locations and is a univariate function in time. For instance, the PEF(NEF) can be represented by a high (low) quantile, which indicates that dependence is observed when a certain percentage of the grid cells have high (low) values. Our focus in this study is on PEFs defined with high temporal and spatial quantiles (refer to Section 1.3.2) so that the dependence reflects cases where a smaller proportion of grid cells have significantly high values.

The PQ framework definition is flexible enough to include various empirical and parametric models that measure process-level dependence, primarily through the overall extreme spatial field. For instance, the PQ modeling framework has been used in Gilleland et al. (2013) for a large-scale severe-weather indicator variable. In that setting, the process of interest served as a surrogate for an inability to accurately measure the true variable of interest, severe weather, at the fine spatial scales on which those processes were resolved. Gilleland et al. (2016b) also used the PQ framework for the same type of severe-storm indicator variables as in Trapp et al. (2007) and Tippett and Cohen (2016) but not for their extremes but rather for their larger values. There, simple empirical frequencies of the variable conditioned on high field energy were analyzed without any need or interest in modeling the very rare extremes. Neither Gilleland et al. (2013) nor Gilleland et al. (2016b) used the term *propinquity* to describe these models as it is first suggested here to identify the type of spatial modeling framework. In this thesis, the data-driven STTC algorithm and the semi-parametric HT model also fall within the PQ modeling framework.

Chapter 3: Applications

In this chapter, we will demonstrate the practical application of the univariate GP model to the Ellicott City flood. Furthermore, we will apply the PQ modeling framework to trend analysis and compare it with conventional methods. Lastly, we will describe a new Feature Finding algorithm that links extreme precipitation patterns to large-scale circulations.

3.1 2016 Ellicott City flood

On July 30, 2016, Ellicott City experienced heavy rainfall, resulting in severe flash flooding. The National Weather Service (NWS) reported that the area received 5.96 inches of rainfall in just two hours, with a total of 6.60 inches recorded for the entire event. Flooding caused extensive damage to houses, roads, and infrastructure, resulting in two fatalities and over 100 rescues. The purpose of this application is to estimate return levels and return periods for this flood and also to see if those quantities are projected to change in the future. This is the perfect application for the univariate GP model conditioned on the high thresholds described in Section 2.1.

The rainfall data for Ellicott City was gathered from the Ellicott City meteorological station, which is situated at latitude $39°15'58.3''$N, longitude $76°51'25.2''$W, and an altitude of 125 meters (GHCN-Daily dataset). For future projections, we use the Climate-Weather Research and Forecasting (CWRF) model (Liang et al., 2012a) coupled with RCP8.5 projections of NCAR's Community Earth System Model (CESM) model.

We fit the GP model with the threshold of 30mm to present and future model runs. The threshold choice was justified upon testing a variety of threshold ranges to balance the bias and variance trade-off described in Chapter 1. The projections were made for a wide range of return periods from 2 to 800 years. The results are shown in Figure 3.1. Upon analyzing this figure, it can be observed that rainfall events of similar intensity are likely to be more severe and increase from 162 to 200mm, with return periods shortening from 450 to 150 years.

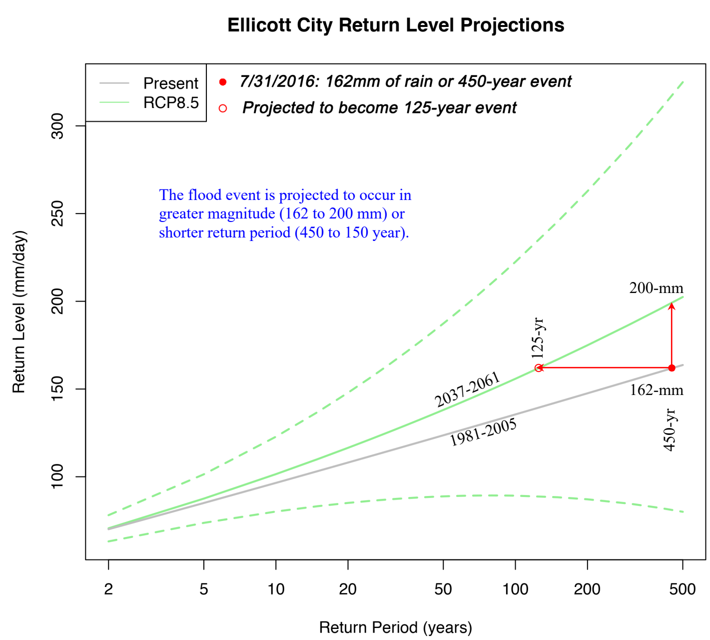

Figure 3.1: Ellicot City rainfall projections. The solid gray and green lines indicate return level projections from observations and future model runs, respectively. Dotted lines display 95% confidence intervals for future model runs.

3.2 Comparing the PQ modeling framework to traditional extreme value models for trend analysis

In Section 1.4.2, we conducted a comparison of the geometrical properties of the STTC and CLST methods. Our analysis involved the use of the STTC algorithm to calculate the frequency of extreme daily EFQ, conditioned on the fields being PEFs. Additionally, we applied the CLST method to determine the usual location-specific frequency, which was area-matched to the STTC frequency. The main distinction between these two methods was that the CLST method, in the majority of cases, lacked geometric interconnectivity between individual grid cells. Our objective in this section is to explore the impact of this graphical property on spatial trend analysis. Specifically, we will compare the empirical and actual probabilities and intensities obtained through the PQ modeling framework with the STTC algorithm (henceforth, STTC) against those obtained through the univariate CLST method. Additionally, we compare outputs from the PQ modeling framework with the HT model coupled with the STTC algorithm (henceforth, HT) against the GP model.

We assess the following quantities:

1. The empirical probabilities of extreme EFQ conditioned on PEFs based on the STTC algorithm and the empirical probabilities calculated by the CLST method;
2. The empirical intensities of extreme EFQ conditioned on PEFs based on the

STTC algorithm and empirical intensities calculated by the CLST method;

3. 20-year probability projections simulated from the HT model conditioned on PEFs and 20-year probability projections derived from the GP model conditioned on the high threshold;

4. 20-year return level projections simulated from the HT model conditioned on PEFs and 20-year return level projections derived from the GP model conditioned on the high threshold.

All those quantities are aggregated in space, and the following scientific questions for extreme EFQ are addressed:

1. Is it an accurate procedure to aggregate the CLST method and the GP model results over space to infer trend direction?

2. How sensitive are the trend patterns to the quantile-based temporal threshold choices?

In order to address the first question, we examine the trend pattern outcomes of PQ and traditional univariate extreme value models (both empirical and statistical) for conditional probabilities and intensities. We focus on how these patterns differ when individual grid cells are combined (i.e., averaged) in space across the entire CONUS domain.

To answer the second question, we perform temporal threshold sensitivity analysis at the 95th, 98th, and 99th percentiles.

We also fit third-degree polynomials to spatially averaged time series of extreme quantities for trend pattern identification. We select the degree of the polynomial based on the Bayesian information criterion (BIC) and mean square error (MSE) measures. However, we limit the highest degree to three to prevent overfitting. For almost all analyzed empirical quantities, the third-degree polynomial is the best fit, and we maintain this consistency across all the time series.

3.2.1 Data

We continue using the GHCN-Daily dataset described in detail in Section 1.2, but we expanded it from 1 January 1951 to 31 December 2020 (70 years). We are still adhering to the strict quality assurance process and making necessary PRISM adjustments. We utilized natural neighbor interpolation (Sibson, 1981) to create a 30 km by 30 km grid covering the contiguous United States (CONUS) from the station-based GHCN-Daily dataset. This resulted in around 27,000 spatial points, each containing over 25,000 days of rainfall data.

3.2.2 Baseline period selection

One commonly used climatological baseline period is a 30-year "normal" period, defined by the World Meteorological Organization (WMO). Although the WMO has changed this period for climate data from 1961-1990 to 1991-2020, it still recommends using the former as a fixed reference for calculating and monitoring global climate anomalies for historical comparison and climate change tracking. The 1961-1990 baseline period also served as a standard reference for numerous impact studies and has better observational climate data coverage and availability compared to earlier periods. However, it is important to note that some areas may show anthropogenic climate changes during this period compared to earlier periods.

3.2.3 EFQ construction

EFQ construction was performed with the following steps:

1. Eliminate seasonal patterns by subtracting climatological mean values over the entire time length from the raw data;
2. Set no accumulation window;
3. Standardize accumulated anomalies by the corresponding high temporal quantile;
4. Subset dataset for 30-year periods with a 1-year rolling step.

We use the daily precipitation dataset and select a high quantile of 0.95.

3.2.4 Rolling analysis

The final step in constructing EFQ involves the application of rolling analysis, a commonly used technique in time series modeling, to evaluate the stability of the model parameters over time (Zivot and Wang, 2006). The rolling analysis is also applied to account for non-stationarity, perform data smoothing, and expand the dataset. In our case, to enhance the identification of spatial trends and achieve a smoother time series, we divide the dataset from 1951 to 2020 into several sub-periods. We are using a rolling window of 30 years that shifts forward one year at a time. This approach results in a total of forty-one sub-periods, with each sub-period spanning 30 years. For instance, the first sub-period ranges from 1951 to 1980, the second from 1952 to 1981, and so on until the forty-first sub-period spans from 1991 to 2020. Our objective is to identify annual trend patterns in probabilities and intensities of extreme EFQ for 30-year overlapping sub-periods based on the daily dataset.

3.2.5 Threshold selection

The threshold selection is important in identifying and making statistical inferences about extreme events. To ensure consistency in studying the trend patterns of conditional probability and intensity of extreme EFQs and determine how they are affected by the chosen extreme value modeling methodology, we compute all threshold values from the baseline period and apply them consecutively to all sub-periods. The process of selecting thresholds for all methods and models described in Section 3.2 is explained below:

1. The STTC algorithm uses fixed temporal and θ_{h} thresholds, which are listed in Table 3.1;
2. The CLST method uses high thresholds that are fixed at the temporal quantile for each grid cell, as listed in the first column of Table 3.1;
3. The HT model is fitted by conditioning on PEFs. The simulated probabilities and return levels use the same θ_{h} thresholds derived from the STTC algorithm;
4. The GP model is fitted to the exceedances above these high thresholds for all

grid cells.

Temporal quantile (τ)	Temporal threshold	θ_{h}
0.95	4.97	10.52
0.98	6.32	10.18
0.99	7.49	9.16

Table 3.1: Fixed thresholds based on the baseline period for the PQ models.

3.2.6 Results

In the top panel of Figure 3.2, we display the empirical probability of extreme EFQ conditioned on PEFs using the STTC methodology. The pattern is akin to an inverted parabola, with the trend increasing until period twenty-six (1976-2005) and then gradually decreasing. The bottom panel shows the empirical probability of extreme EFQ conditioned on the high percentile thresholds using the CLST method. The pattern displays multi-decadal oscillations with no visible upward/downward trend.

The first panel of Figure 3.3 represents the empirical intensity of extreme EFQ, which is conditioned on PEFs using the STTC methodology. The pattern of this intensity is similar to the empirical STTC probability shown in Figure 3.2. The second panel displays the empirical intensity of extreme EFQ, which is conditioned on the high percentile thresholds using the CLST method. The 95th percentile of the CLST model exhibits multi-decadal oscillations with a weak upward trend. This trend becomes more intense in the 98th and 99th percentiles and appears as a straight line that monotonically increases with a few bumps.

The results for the HT model projections of 20-year conditional probabilities

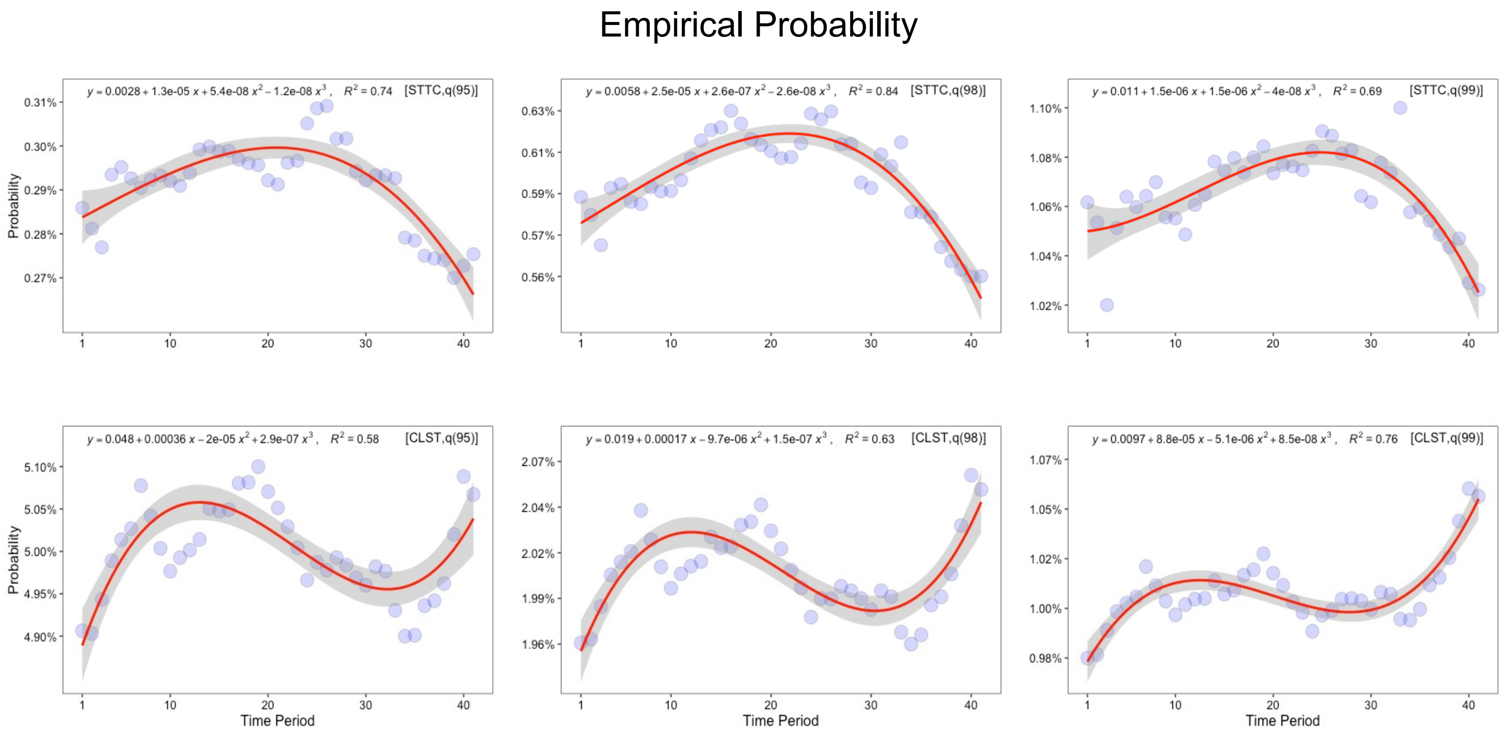

Figure 3.2: Top panel: the empirical probability of extreme EFQ conditioned on PEFs based on the STTC methodology. Bottom panel: the empirical probability calculated by the CLST method. The columns, from left to right, represent the 0.95, 0.98, and 0.99 temporal quantiles, respectively. In the top right corner, in square brackets, the model name and temporal percentile are shown. The red line is a 3rd-degree polynomial fitted, and the shaded area is a 95% confidence interval. The first time period is from 1951 to 1980, the second from 1952 to 1981, and so on, with the forty-first time period being from 1991 to 2020.

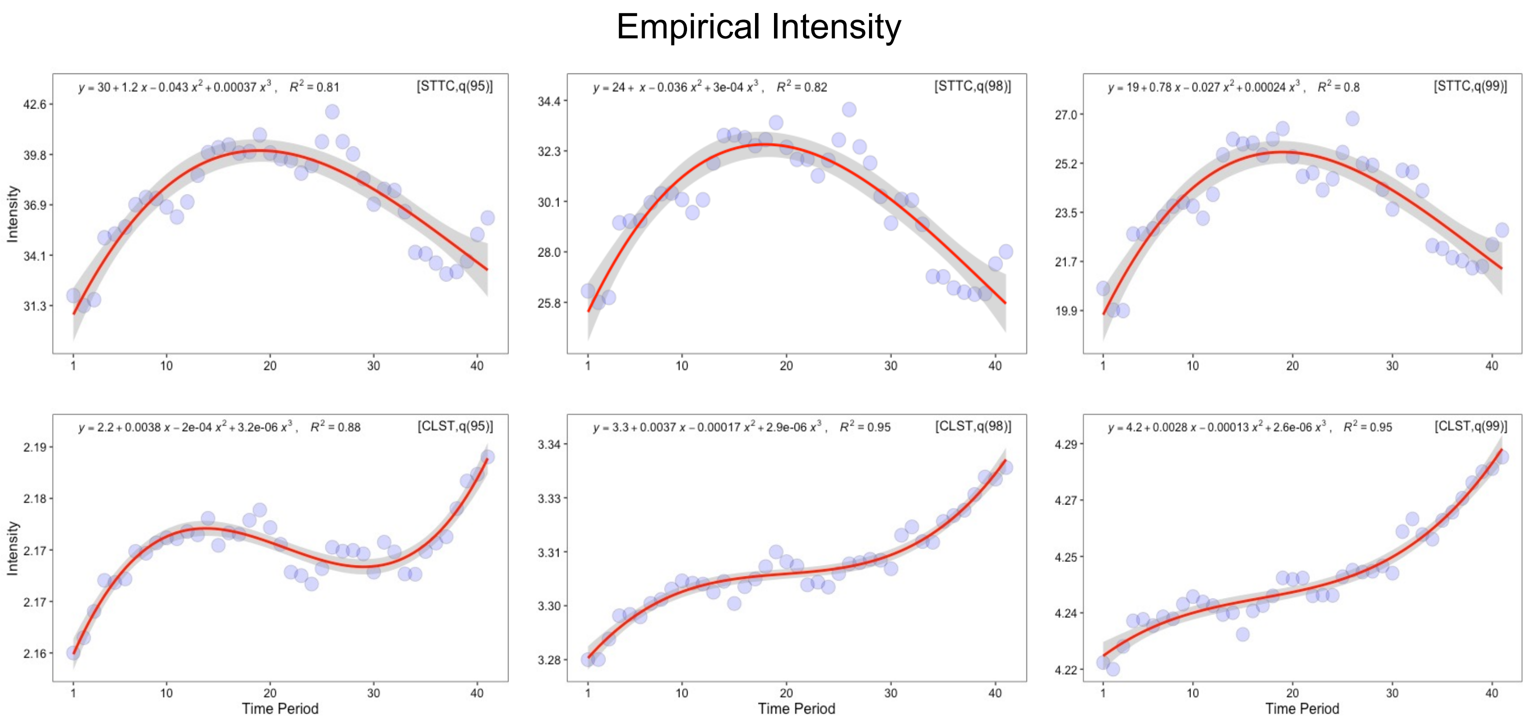

Figure 3.3: Top panel: the empirical intensity of extreme EFQ conditioned on PEFs based on the STTC methodology. Bottom panel: the empirical intensity calculated by the CLST method. By design, EFQ has no units. The columns, from left to right, represent the 0.95, 0.98, and 0.99 temporal quantiles, respectively. In the top right corner, in square brackets, the model name and temporal percentile are shown. The red line is a 3rd-degree polynomial fitted, and the shaded area is a 95% confidence interval. The first time period is from 1951 to 1980, the second from 1952 to 1981, and so on, with the forty-first time period being from 1991 to 2020.

are displayed in the top panel of Figure 3.4. In the bottom panel, 20-year probabilities conditioned on the high thresholds from the GP model are shown. The two trend patterns are different and are moving in opposite directions. The HT model projects an increase in the probability of extreme EFQ over the next 20 years, while the GP model projects a gradual decrease.

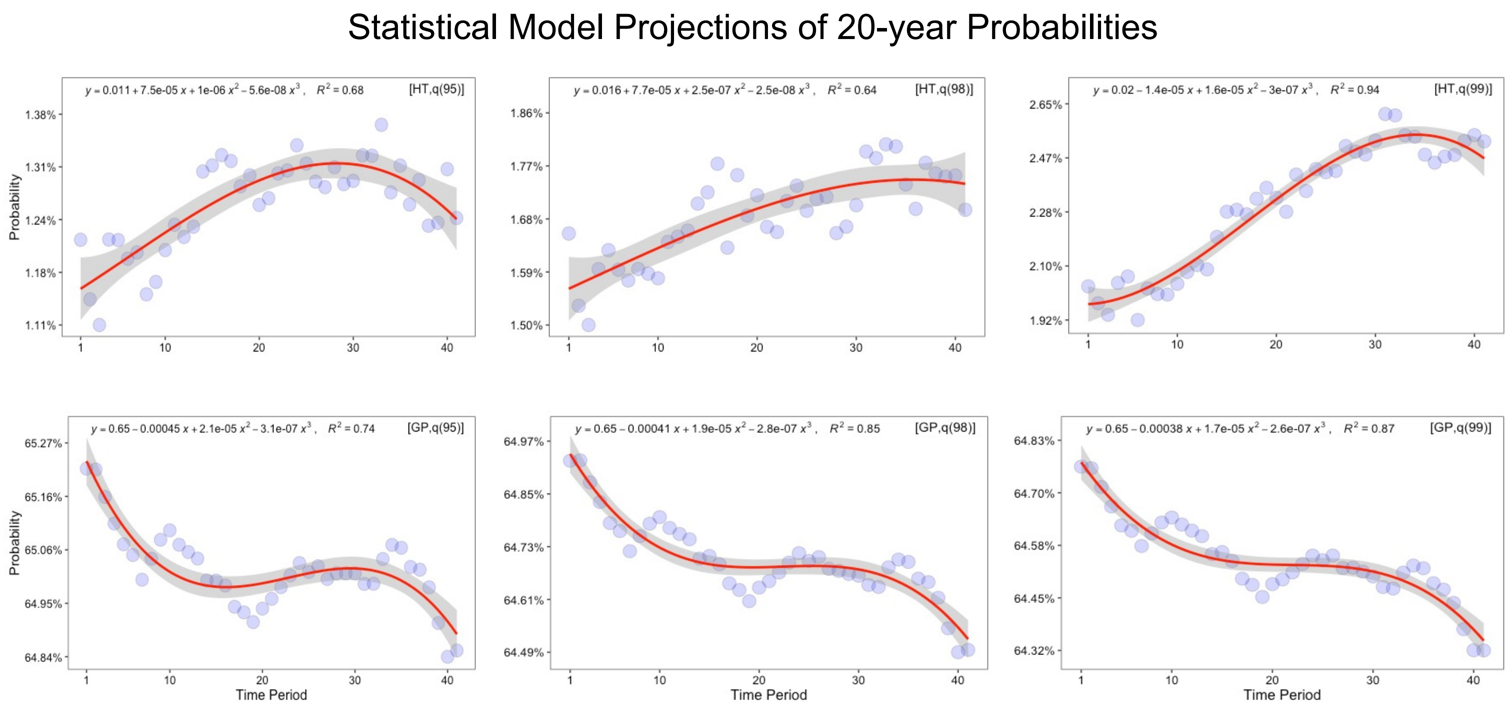

Figure 3.4: Top panel: the HT model projections of 20-year probability conditioned on PEFs. Bottom panel: the GP model projections of 20-year probability conditioned on the high threshold. The columns, from left to right, represent the 0.95, 0.98, and 0.99 temporal quantiles, respectively. In the top right corner, in square brackets, the model name and temporal percentile are shown. The red line is a 3rd-degree polynomial fitted, and the shaded area is a 95% confidence interval. The first time period is from 1951 to 1980, the second from 1952 to 1981, and so on, with the forty-first time period being from 1991 to 2020.

Figure 3.5 displays two distinct trend patterns between the HT model projections of 20-year return level conditioned on PEFs and the GP model projections of 20-year return level conditioned on the high threshold. However, both HT and GP models project a rise in 20-year return levels for extreme EFQ, though the HT model suggests it may dip back down in the later periods.

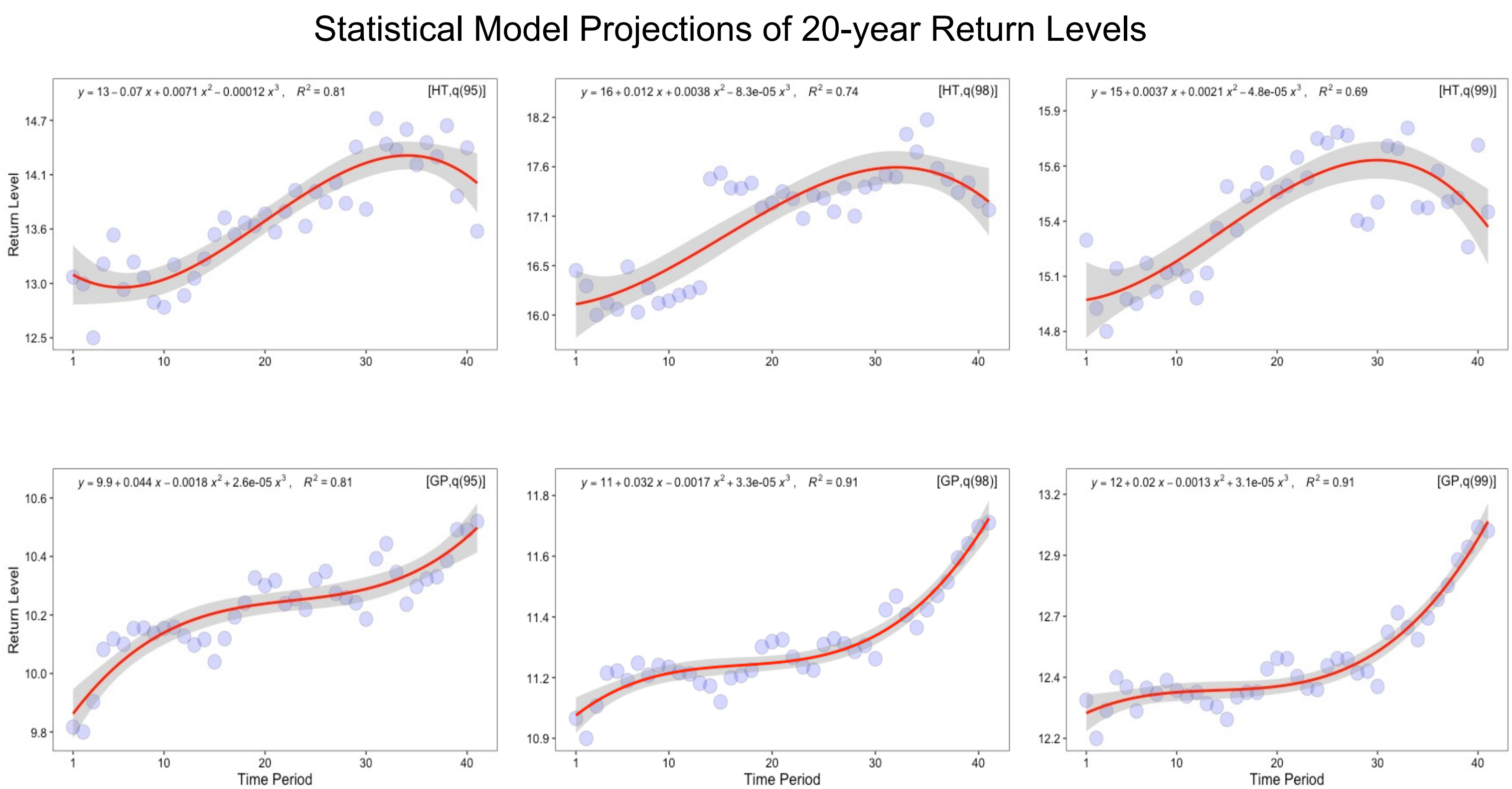

Figure 3.5: Top Panel: the HT model projections of 20-year return level conditioned on PEFs. Bottom panel: the GP model projections of 20-year return level conditioned on the high threshold. The columns, from left to right, represent the 0.95, 0.98, and 0.99 temporal quantiles, respectively. In the top right corner, in square brackets, the model name and temporal percentile are shown. The red line is a 3rd-degree polynomial fitted, and the shaded area is a 95% confidence interval. The first time period is from 1951 to 1980, the second from 1952 to 1981, and so on, with the forty-first time period being from 1991 to 2020.

3.2.7 Conclusions and discussion

Through spatial trend analysis, we show the significant contrasts between PQ and conventional extreme value models. Additionally, the process of spatially aggregating the model results for the trend construction can result in considerable differences in the overall trend pattern and direction, creating challenges for weather forecasting and climate adaptation initiatives. However, one could argue that the difference is due to the comparison of different physical quantities. The PQ models, for instance, are conditioned on persistent large-scale extreme spatial patterns that cover extensive geographical regions. On the other hand, univariate location-based methods are conditioned on high thresholds and lack the interconnectivity between individual grid cells. While we do agree with this viewpoint, we maintain that the current practice of aggregating the results of independent location-based methods without properly accounting for the connectivity between individual grid cells in atmospheric and oceanic science literature leads to a biased outcome for trend analysis. Therefore, it is important to avoid such practices and instead opt for the PQ modeling framework or other spatial EVA models that consider the interconnectivity between individual grid cells and factor in the dynamic and thermodynamic interactions between them. This will provide a more accurate depiction of trend direction and pattern, not only for synoptic scale processes but also for localized convective regions, especially when the former are becoming increasingly disconnected, as we previously indicated.

In the sensitivity analysis, the study examined the impact of different quantile-based temporal thresholds on the trend patterns of extreme EFQs. The results showed that the trend patterns were sensitive to the chosen temporal threshold in some experiments but not all.

3.3 Linking extreme EFQs to large-scale circulation patterns

The objective of this section is to gain a deeper understanding of the trends in extreme EFQ intensity patterns through the use of the PQ modeling framework with the STTC method. Additionally, we aim to identify the main drivers behind these trends. To accomplish this, we develop a Feature Finding Decomposition (FFD) algorithm which consists of the following three steps:

1. Feature Finding;
2. Image decomposition;
3. Large-scale circulation patterns connection.

3.3.1 Data

We follow the same process as in Section 1.3.1 to create EFQ without using a rolling accumulation window and focus on a high quantile (0.95).

For the large-scale circulation patterns, we use a monthly time series of climate indices retrieved from `https://psl.noaa.gov/data/climateindices/list/`, last visited 15 March 2023). Thirteen monthly climate indices were utilized from 1/1951 to 12/2020. The details regarding each index can be found in the Table 3.2.

Name	Description
AO	Arctic Oscillation: from NOAA Climate Prediction Center (CPC)
EA/WR	Eastern Atlantic/Western Russia: from CPC
EP/NP	East Pacific/North Pacific Oscillation: from CPC
NAO (Jones)	North Atlantic Oscillation: from Climatic Research Unit (CRU)
Niño 3	Eastern Tropical Pacific SST (5N-5S,150W-90W): from CPC
Niño 3.4	East Central Tropical Pacific SST (5N-5S)(170-120W): from CPC
NP	North Pacific Pattern: from Trenberth and Hurrell (1994)
ONI	Oceanic Niño Index: from CPC
PDO	Pacific Decadal Oscillation: from Joint Institute for the Study of the Atmosphere and Ocean (JISAO)
PNA	Pacific North American Index: from CPC
SOI	Southern Oscillation Index: from CPC
TNI	Trans-Niño Index: from Trenberth and Stepaniak (2001)
WP	Western Pacific Index: from CPC

Table 3.2: Climate Indices.

3.3.2 Feature Finding

Feature finding approaches have been widely used in forecast verification Gilleland et al. (2009b). Their primary objective is to assess how accurately numerical weather prediction models depict observed meteorological features. The main variations in these approaches primarily involve how they define features, handle spatially discontinuous features, match features between fields, and the types of output they generate. In this thesis, we use a feature-based approach that utilizes cylindrical convolution (smoothing) and thresholding for object identification, which was developed by Davis et al. (2006a,b) and implemented in `SpatialVx` Gilleland (2017b) package in R. This method has the following steps:

1. Cylindrical convolution for object smoothing;

2. Thresholding to detect object boundaries;

3. Mapping objects into simple geometric shapes;
4. Merging objects within the same field, if necessary.

We used this method to divide the extreme EFQ field into two feature classes based on their area: (i) the largest and (ii) the remaining.

3.3.3 Image decomposition

From the results of the feature finding step, we feed those features into the variational mode decomposition (VMD) module developed by Dragomiretskiy and Zosso (2013). The features are decomposed into a number of intrinsic mode functions (IMFs), each with specific properties, where an IMF is an amplitude-modulated and frequency-modulated (AM-FM) signal with narrow-band properties. The number of IMFs is determined by the algorithm, and in our case, it is equal to five. For example, the decomposition of the largest feature intensity from the FFD algorithm is displayed in Figure 3.6.

By excluding the high-frequency modes, specifically modes one through four, we can isolate and identify the trend pattern. We perform this process for the frequency and intensity of extreme EFQ conditioned on PEFs based on the STTC algorithm (Figures 3.7 and 3.8) as well as for the frequency and intensity of extreme EFQs based on the CLST method (Figure 3.9).

The STTC largest feature intensity has a shape resembling an inverted parabola, which is similar to Figure 3.2. The remaining feature intensity is gradually increasing, which is similar to the direction of CLST intensity in Figure 3.7 on the right

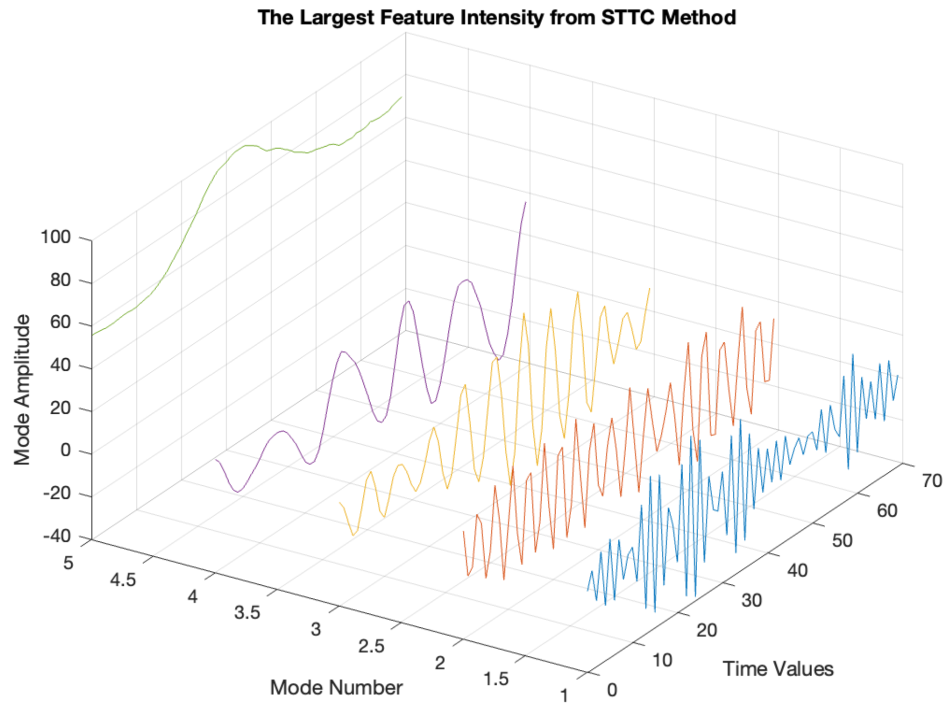

Figure 3.6: Signal decomposition for the largest feature intensity from the FFD algorithm.

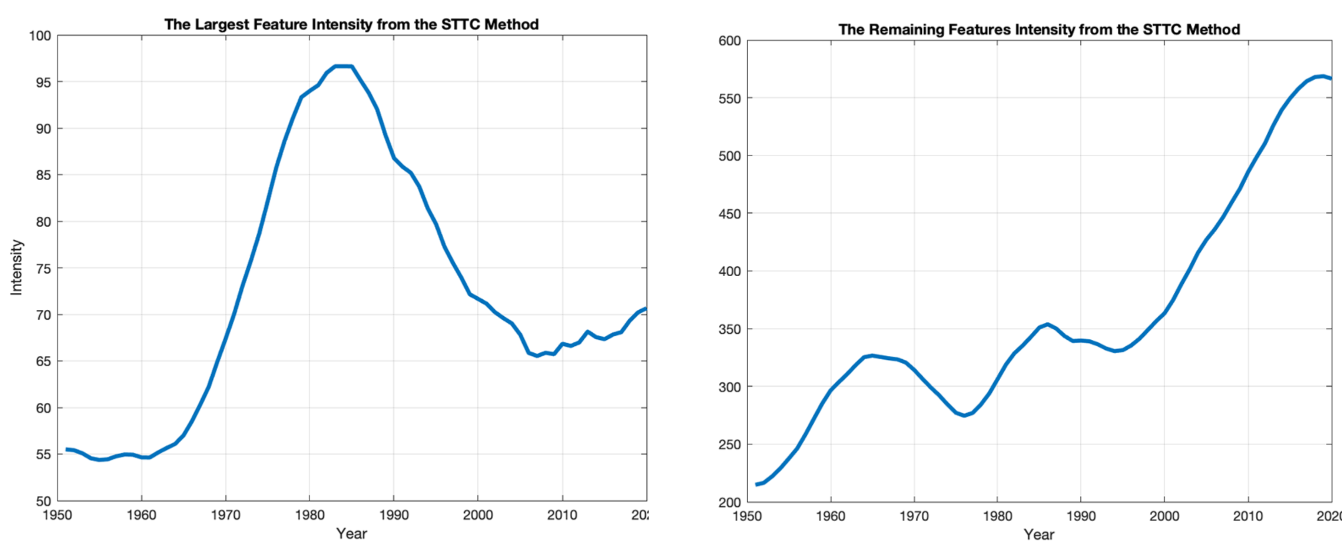

Figure 3.7: Left panel: the largest feature intensity of extreme EFQ conditioned on PEFs from the FFD algorithm; right panel: the remaining feature intensity of extreme EFQ conditioned on PEFs from the FFD algorithm. By design, EFQ has no units.

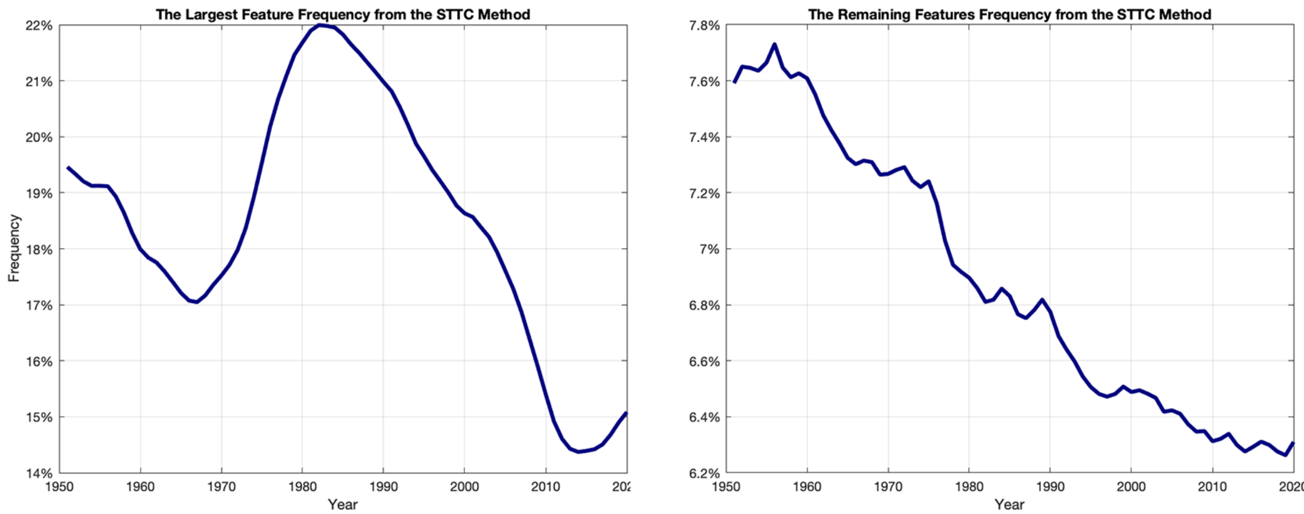

Figure 3.8: Left panel: the largest feature frequency of extreme EFQ conditioned on PEFs from the FFD algorithm; right panel: the remaining feature frequency of extreme EFQ conditioned on PEFs from the FFD algorithm.

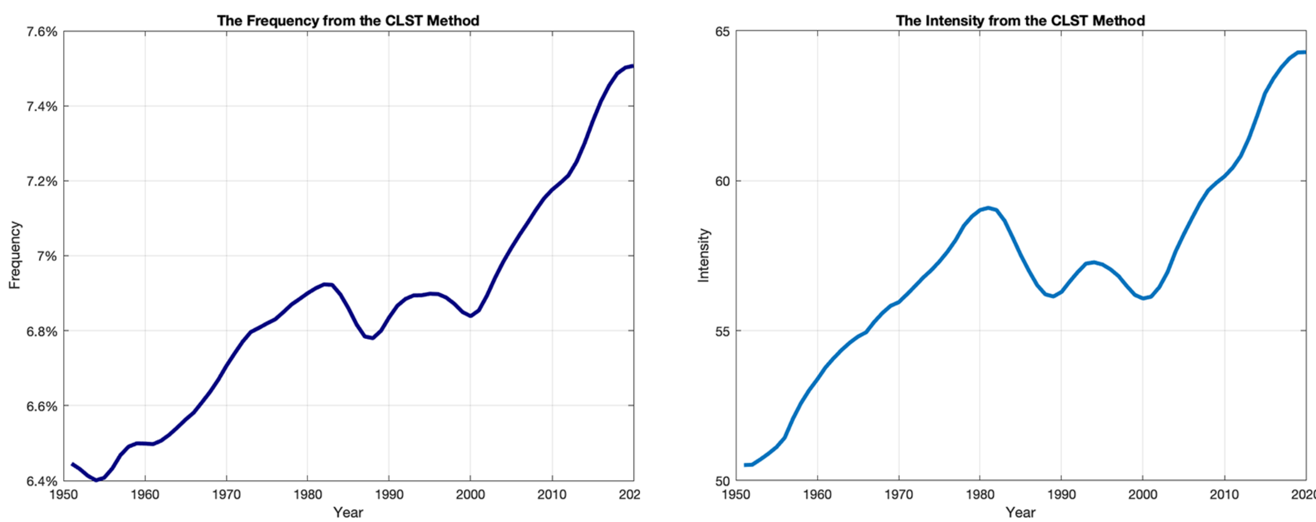

Figure 3.9: Left panel: the frequency of extreme EFQs from the CLST method; right panel: the intensity of extreme EFQs from the CLST method. By design, EFQ has no units.

panel. Both the largest and the remaining feature frequencies are decreasing, which is opposite to the increasing trend in CLST frequency.

3.3.4 Large-scale circulation patterns connection

In this study, we apply the Extreme Gradient Boosting algorithm (XGBoost; Chen and Guestrin, 2016) – a prevalent ensemble learning algorithm – implemented in the R package `xgboost` (Chen et al., 2019) to investigate the relationship between the trend pattern (IMF=5) in the largest feature intensity identified by the STTC algorithm and the teleconnection indices. XGBoost uses gradient boosting to create decision tree models. It has the ability to deal with missing values, has one of the best predictive accuracies, is flexible, and prevents overfitting by using regularization techniques to penalize complex models.

The data preparation for the XGBoost has the following steps:

1. Expanding the dataset with quantile selection;
2. Normalization of all data;
3. Expanding with IMFs;
4. Expanding with original indices.

Our process for creating the training dataset begins with using the original 13 teleconnection indices provided on a monthly basis from 1951 to 2020. To convert them to an annual basis, we take the 95th, 50th, and 5th percentiles from the original monthly data. The resulting time series are then normalized and decomposed into five IMFs. In addition, we include the original index time series (normalized but not decomposed). In total, our training set consists of 234 time series.

To train our algorithm, we use 75% of the data, reserving the remaining 25%

for validation. We conduct tuning by performing a grid search across multiple hyperparameters and utilizing 5,000 trees. The best tuning model based on the root mean square error (RMSE) is then used for making predictions. The best model explains 96% of the data variability and has an RMSE of 0.08.

The image depicted in Figure 3.10 showcases the importance plot of our final model. According to our findings, the Western Pacific Index, specifically its 5th percentile and 5th mode of decomposition (WP.05_5), is the most ideal teleconnection pattern that explains the variation in the trend pattern of the largest feature intensity from the FFD algorithm. The importance of WP.05_5 surpasses all other time series, indicating that its trend is strongly linked with the trend in the intensity of the largest feature of extreme EFQ derived from the FFD algorithm. Figure 3.11 shows the actual vs. predicted plot without residuals, and the actual vs. predicted plot with residuals is shown in Figure 3.12.

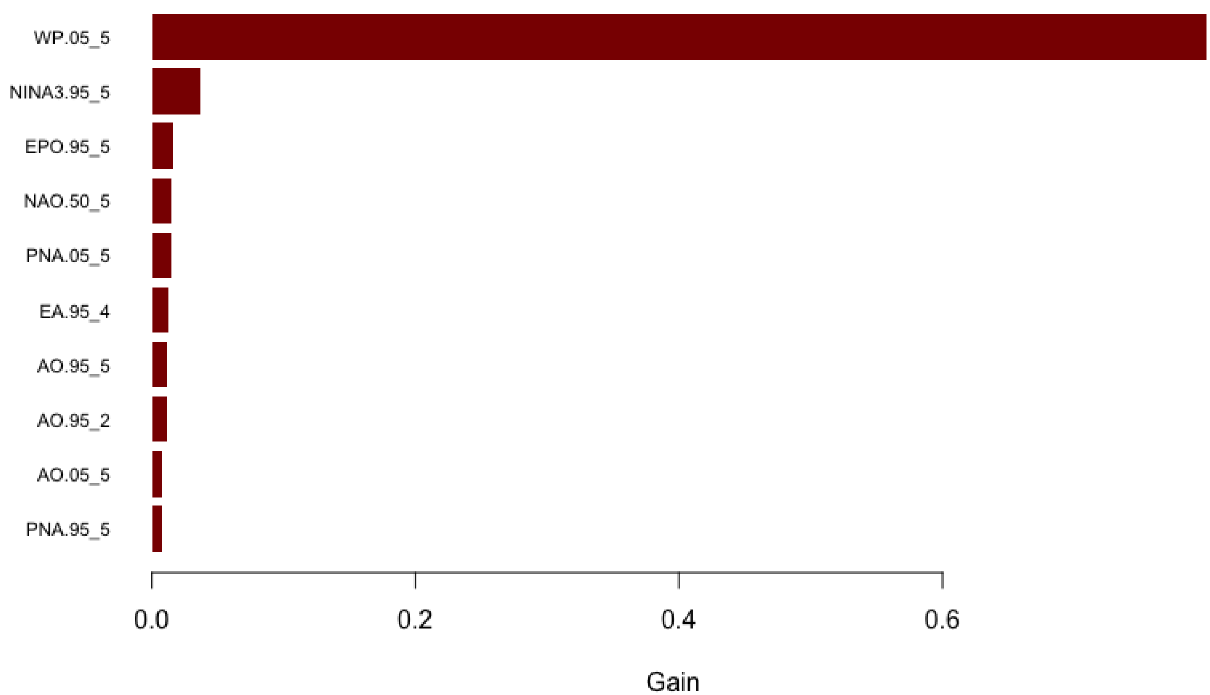

Figure 3.10: Importance plot of the final model.

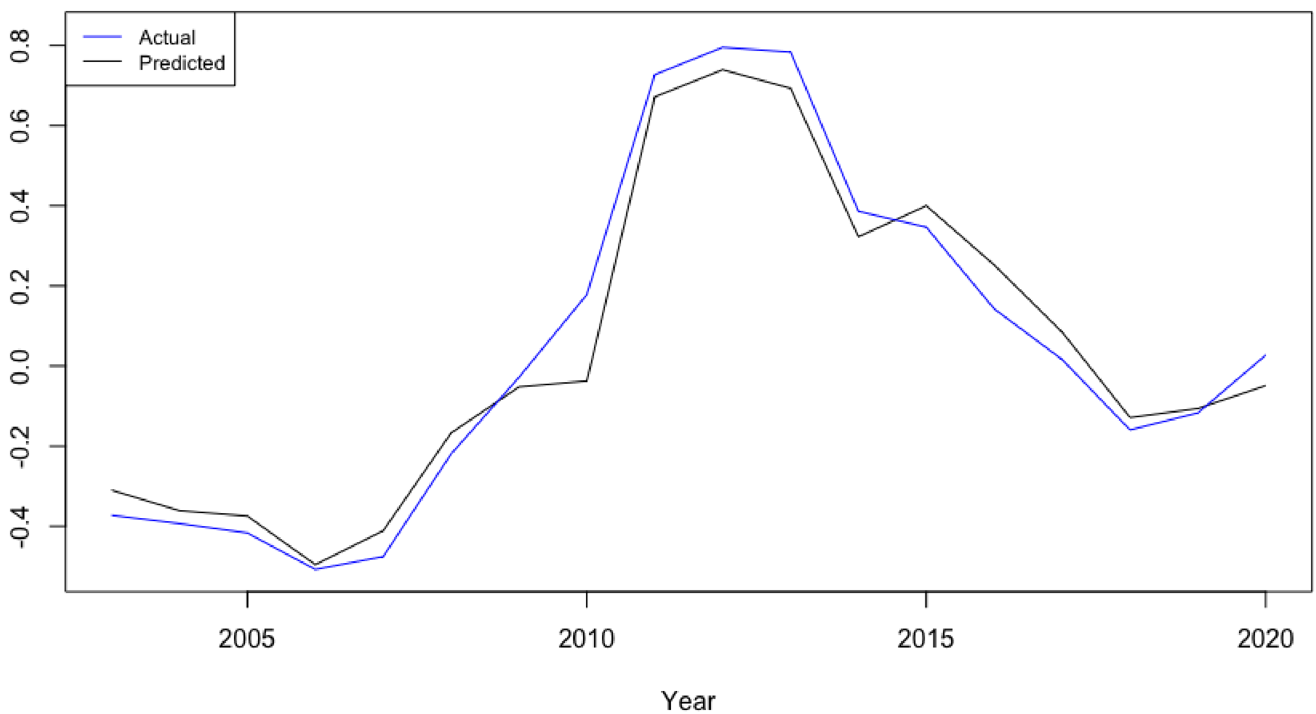

Figure 3.11: Actual vs. Predicted plot without residuals.

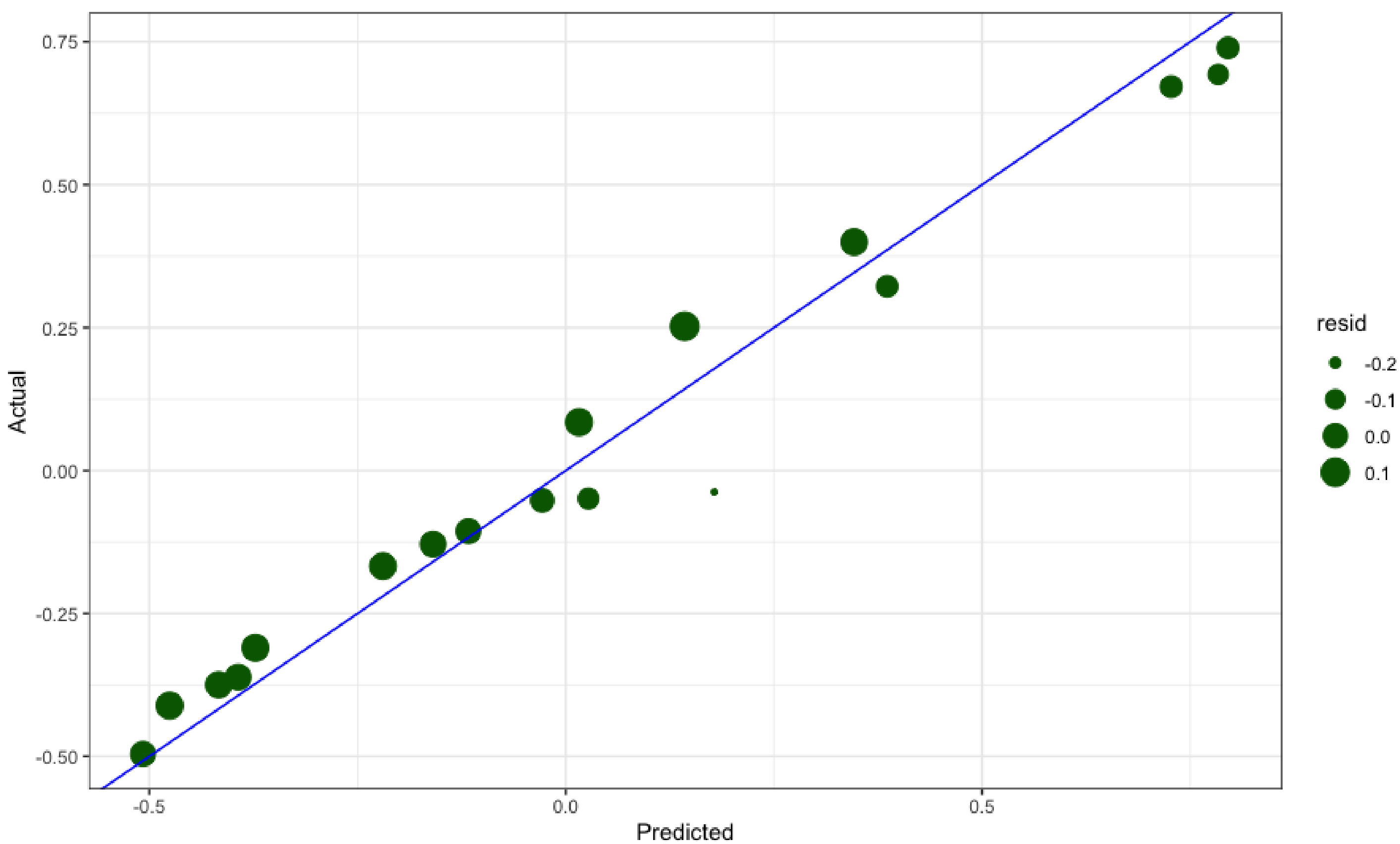

Figure 3.12: Actual vs. Predicted plot. The green circles show residual size.

3.3.5 Conclusions and discussion

The aim of this application was to establish a connection between extreme rainfall intensity trend patterns and large-scale circulations. We used the XGBoost algorithm to establish a link between STTC trend intensity and the WP.05_5 index, but we were unable to prove causation in the Granger causality test. During the image decomposition step of the FFD algorithm, we extracted four IMF modes to determine the trend component of the overall signal. One might wonder if those four IMF modes have a link with the large-circulation patterns as well. To answer this question, we performed cross-correlation between those modes and the remaining time series. Table 3.3 displays the results.

Mode Number	Xcorr	Index
IMF1	-0.65	WP_50
IMF2	0.63	NINA3_50
IMF3	-0.3	EPO_95
IMF4	0.85	TNI_95

Table 3.3: Cross correlation between WP.05 modes 1-4 and the rest of the time series.

We found that IMF1 displays a strong negative cross-correlation with the median of the WP index. Similarly, IMF2 exhibits a strong positive cross-correlation with the median of the NINA3 index. On the other hand, IMF3 shows a weak negative cross-correlation with the 95th percentile of the EPO index. Lastly, IMF4 displays a significant positive cross-correlation with the 95th quantile of the TNI index.

Finally, Figure 3.13 shows the trend pattern that is in the IMF5 of the 5th percentile of the WP index. This pattern is almost identical to the IMF5 of the STTC trend intensity in Figure 3.7, implying that the link we have established is strong but requires further investigation.

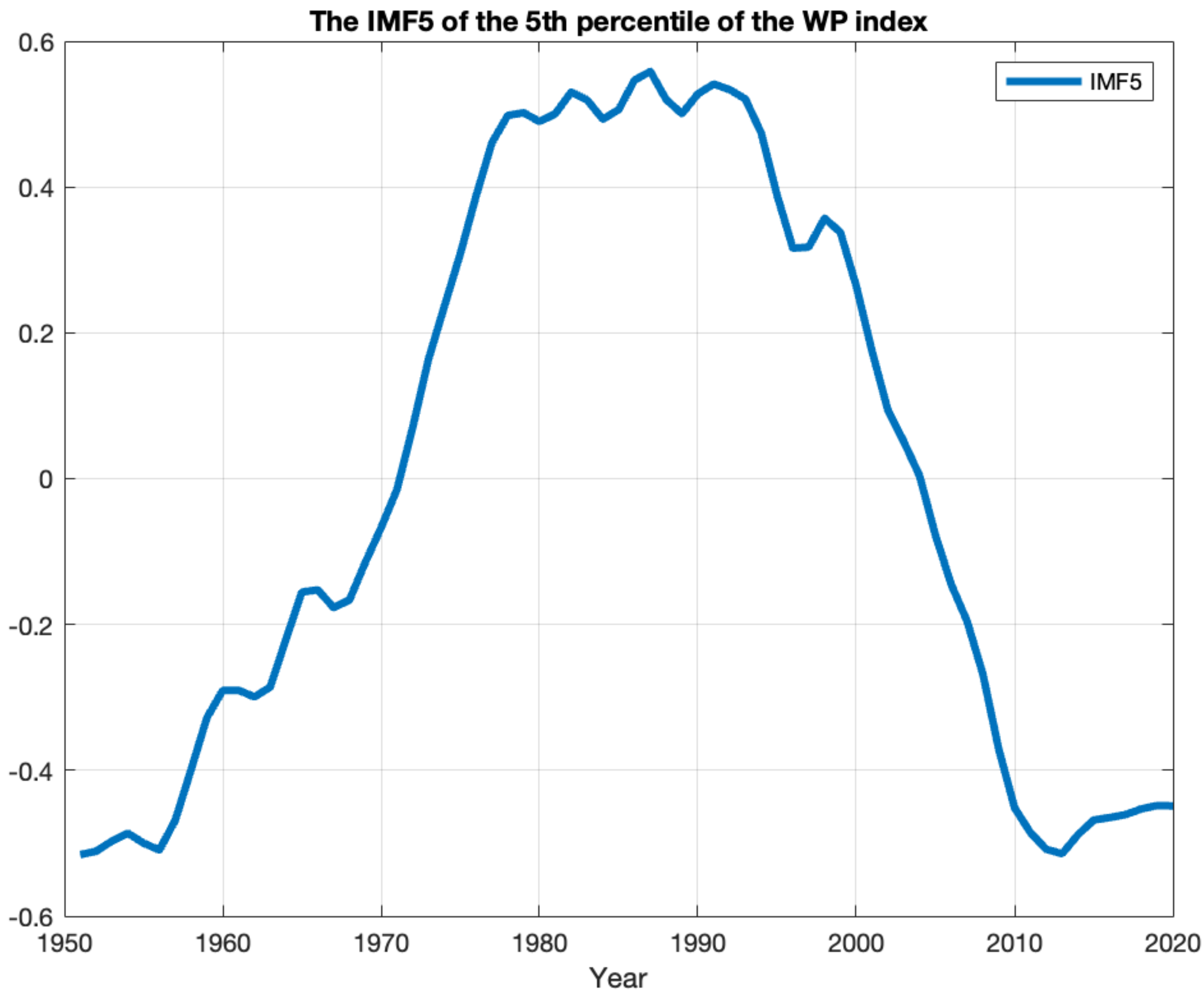

Figure 3.13: Trend pattern in the IMF5 of the 5th percentile of the WP index. The WP Index was normalized and has no units.

Chapter 4: Summary of major findings and future research

The lack of a consistent definition of an ”extreme” event presents a challenge in comparing various scientific studies. It hinders transparency and risk management processes, making it harder for decision-makers to make informed decisions. The dissertation thesis in Chapter 1 proposed a new generalized Spatio-Temporal Threshold Clustering (STTC) for extreme events using an integrative modeling approach. The algorithm can model extremes with high and low thresholds and is less sensitive to grid size for areal mean interpolation.

The STTC method automates the threshold selection process through the clustering algorithm. The clustering solution is directly linked to the threshold choice, allowing for user-defined spatial fields. The study also recommended addressing the unique connectivity/area trade-off to achieve desired geometrical properties. This involves finding the appropriate balance between the connectivity index and the area of non-zero grid points when determining a threshold for extreme events. As a result, the STTC algorithm can be utilized not only to determine thresholds for extreme events modeling but also to target required geometrical properties, making it a useful, practical tool for empirical EVA studies and spatial forecast verification. The algorithm can efficiently handle large datasets and has the potential to be integrated into climate modeling systems to detect spatial patterns at the synoptic scale that are associated with individual extreme events.

The STTC algorithm was employed to analyze the precipitation dataset, and essential field quantities (EFQs) were constructed to identify extreme EFQ patterns. These patterns could be asymmetric due to wind shear, making them difficult to identify using traditional location-specific methods. Further application of the STTC algorithm revealed a downward trend in the connectivity index of the conditional frequency of the extreme EFQ. If this trend continues, it could create enormous challenges for weather forecasting and climate adaptation efforts, particularly in the form of more localized droughts and heatwaves during the summer months. Therefore, the STTC algorithm offers a distinctive way to perform diagnostic analysis of the large-scale meteorological patterns specifically associated with extreme episodes, providing vital insights for climate researchers and policymakers.

In order to gain a better understanding of the STTC algorithm, we compared it with the Conventional Location Specific Threshold (CLST) method. The key contrasting features of each algorithm are outlined below:

1. The CLST method is a univariate approach that considers extreme values at individual locations and their temporal dependence;

2. The CLST method displays extremes at every non-empty grid cell and, in most cases, is deficient in geometric interconnectivity between spatial pixels;

3. The STTC algorithm is a multivariable method and applies to the overall spatial field that is conditioned on being extreme (i.e., conditioned on being

either positive (PEF) or negative extreme filed (NEF));

4. The STTC algorithm incorporates spatial connectivity among individual grid cells.

Both methods are empirical and objective.

In Chapter 2, we provided a detailed overview of the classical univariate and multivariate EVA models and highlighted their limitations. We also discussed the Heffernan and Tawn (HT) model, which has a number of advantages over classical multivariate EVA models. The HT model can handle complex multivariate extreme value distributions, account for a variety of dependence structures, estimate the conditional distribution of a multivariate variable given one of its components is large, and handle large datasets in space and time dimensions. However, it also has certain limitations, such as the need for large amounts of data to estimate the parameters accurately, sensitivity to the choice of conditioning variables, and the assumption of a stationary joint distribution of variables that may not hold in real-world scenarios.

Our work has several additional innovative aspects. Firstly, we applied the HT model with the rolling window approach to address potential non-stationarity in the dataset, which, to the best of our knowledge, has not been applied in this type of model previously. Secondly, we incorporated the HT model with the STTC algorithm to improve its processing time. This inclusion allowed us to only use PEF fields as input for the model instead of the entire dataset, which could have been a time-consuming process. Thirdly, we integrated the STTC algorithm and the conditional semi-parametric HT model into the new Propinquity (PQ) modeling framework, which includes various empirical and parametric models that calculate process level dependence of the large-scale extreme processes, primarily through the overall extreme spatial field. This approach differs from the CLST methods that are independent in time. Instead, it measures spatial dependence by using the positive or negative extreme field as the staircase to facilitate the modeling of spatial dependence at the process level of extreme EFQ.

We presented several EVA modeling applications in Chapter 3. The primary aim of the first applications was to estimate return levels and return periods for the severe flash flood that occurred in Ellicott City on July 30, 2016, and to investigate if these quantities are expected to change in the future. The study applied the univariate generalized Pareto (GP) model conditioned on a high threshold for this purpose. The rainfall data from the Ellicott City meteorological station was utilized to estimate return levels and return periods. The Climate-Weather Research and Forecasting (CWRF) model, coupled with RCP8.5 projections of NCAR's Community Earth System Model (CESM) model, was used for future projections.

The univariate GP model with a threshold of 30mm was fitted for both present and future model runs. The projections were made for a wide range of return periods from 2 to 800 years. The results showed that rainfall events of similar intensity are likely to be more severe and increase from 162 to 200mm, with return periods shortening from 450 to 150 years. These findings have important implications for weather forecasting and climate adaptation strategies, especially for areas prone to flash flooding.

In the second application, our focus was on investigating the impact of differences in geometrical properties between STTC and CLST methods on spatial trend analysis. Specifically, we compared empirical and statistical probabilities and intensities obtained through the PQ modeling framework with the STTC algorithm against those obtained through the univariate CLST method. Additionally, we compared outputs from the PQ modeling framework with the HT model coupled with the STTC algorithm against the GP model.

The results showed significant contrasts for both empirical and statistical methods between PQ and conventional extreme value models. Additionally, the process of spatially aggregating the model outputs that do not account for interconnectivity between individual grid cells for trend construction can result in considerable differences in the overall trend pattern and direction when compared with the models that do, creating challenges for weather forecasting and climate adaptation initiatives. The study suggested that it is important to avoid such practices and instead adopt the PQ modeling framework or other spatial EVA models that take into account the interconnectivity between individual grid cells. These models also consider the dynamic and thermodynamic interactions between the grid cells, which results in a more accurate depiction of trend direction and pattern for both synoptic scale processes and localized convective regions.

In the sensitivity analysis, the study investigated the impact of different quantile-based temporal thresholds on the trend patterns of extreme EFQ. The results indicated that some experiments were sensitive to the selected temporal threshold while others were not.

In the final application, the objective was to gain a deeper understanding of the trend in extreme EFQ intensity fields by linking them to large-scale circulation patterns. To achieve this, we developed a feature finding decomposition (FFD) algorithm consisting of three steps: feature finding, image decomposition, and large-scale circulation patterns connection. In the feature finding step, we utilized cylindrical convolution and thresholding techniques for object identification to divide the extreme EFQ intensity field into two feature classes based on their area. Our primary interest was to explain the trend pattern in the largest feature intensity obtained from this step. In the image decomposition step, we used the variational mode decomposition method to segregate the features into a number of intrinsic mode functions (IMFs), each with specific properties. The trend pattern corresponded to the fifth IMF, and that was one of the inputs to the last large-scale circulation patterns connection step of the FFD algorithm. In this step, we utilized the Extreme Gradient Boosting algorithm (XGBoost) to examine the relationship between the trend pattern in the largest feature intensity identified by the FFD algorithm and teleconnection indices. The findings indicated that the Western Pacific Index (WP), specifically its 5th percentile and 5th mode of decomposition (WP.05_5), is the most important teleconnection pattern that explains the variation in the trend pattern of the largest feature intensity from the FFD algorithm. Although the study was unable to prove causation through the Granger causality test, the importance of WP.05_5 surpasses all other time series, demonstrating a strong link between the trend in the intensity of the largest feature of extreme EFQ and the trend in WP.05_5.

In summary, the dissertation thesis presented a number of innovative contributions to the scientific literature on extreme events:

1. Developed a new generalized STTC algorithm for extreme events that identifies extreme episodes and links geometrical properties of those episodes with corresponding threshold choice;
2. Emphasized the importance of addressing the connectivity/area trade-off to choose a high threshold to ensure the feasibility of achieving desirable geometrical properties for determining large-scale spatial patterns;
3. Compared the STTC algorithm with the CLST method, showing that the former algorithm exhibited greater spatial connectivity among individual grid cells;
4. Discovered a downward trend in the connectivity index of the conditional frequency of the extreme EFQs derived from the precipitation dataset;
5. Applied the HT model with the rolling window approach to address potential non-stationarity in the dataset;
6. Integrated the STTC algorithm and the HT model into the new PQ modeling framework for spatio-temporal extreme events to segregate classes of models that can calculate process level dependence of large-scale extreme processes, primarily through the overall extreme spatial field;
7. Developed the FFD algorithm to link a trend in extreme essential field quantity intensity with large-scale circulation patterns;

In our future work, we plan to consider the empirical probability of extreme EFQs using the PQ modeling framework with the STTC algorithm. Furthermore, we would like to perform a similar FFD analysis to link major circulation patterns to it. We also aim to link empirical frequency and intensity fields to the physical processes to gain a better understanding of the dynamic and thermodynamic changes driving the observed patterns. Additionally, we plan to investigate whether an ensemble of climate models can replicate the observed spatio-temporal extreme event patterns and compare future projections of the climate models to the PQ modeling framework with the HT model coupled with the STTC algorithm to determine if the future is stationary. There are various other scientific inquiries to consider depending on the future needs, which are currently uncertain. However, one thing is certain: the integration of the spatio-temporal PQ modeling framework with the STTC method or HT model and the FFD algorithm is an invaluable tool for analyzing trend patterns for extreme events and associating them with large-scale circulations.

Appendix

Algorithm 1: Spatio-Temporal Threshold Clustering Method (STTC)

input: $Y(\mathbf{s}, t)$, with number of grid cells $\mathbf{s}$ and time points t, where $1 \leq \mathbf{s} \leq S$ and $1 \leq t \leq T$; $\theta = [\theta_1, \ldots, \theta_n]$ vector of initial thresholds, where $\theta_1 < \theta_2 <, \ldots, < \theta_n$;

output: $M = [m_1, m_2, m_3]$ vector of clusters medoids; $\Theta_M = [\theta_{m_1}, \theta_{m_2}, \theta_{m_3}]$ corresponding vector of thresholds;

`// Step 1` $\Leftarrow$ `EFQ construction`
$W \leftarrow 0$; `// user defined input for accumulation window`
Calculate EFQ over window W; `// see Section 1.3.1`

`// Step 2` $\Leftarrow$ `Dimension reduction`
$\eta \leftarrow 0.99$; `// user defined input`
$\tau \leftarrow 0.95$; `// user defined input`
Calculate g_t; `// see Section 1.3.2`
Calculate $\lambda[g_t]$; `// see Table 1.2 (temporal quantile definition)`
if $g_t > \lambda[g_t]$ **then** `// PEF (see Def. 1.3.2.2)`
    $\mathrm{H}(\mathbf{s}, t) \leftarrow$ `FindSubset(`Y`)`;

`// Step 3` $\Leftarrow$ `Spatial domain mapping`
foreach *threshold* θ **do**
    $f \leftarrow$ `CondFreq(`θ`)`; `// see Eq. (1.1)`
    $ci \leftarrow$ `Cindex` (f); `// see Eq. (1.3)`
    $si \leftarrow$ `Sindex` (f); `// see Eq. (1.4)`
    $ai \leftarrow$ `Aindex` (f);
    $\bar{ai} \leftarrow 1 - ai$; `// see Eq. (1.5)`

$ts \leftarrow$ `ColBind(`$ci, si, \bar{ai}$`)`; `// combine indices`

`// Step 4` $\Leftarrow$ `Time series clustering`
Calculate $d_{\mathrm{ACF}}(ts)$; `// Eq. (1.6)`
Calculate $d_{\mathrm{PACF}}(ts)$; `// Eq. (1.7)`
$M_{\mathrm{ACF}} \leftarrow$ `PAM` $(d_{\mathrm{ACF}}, 3)$; `// medoids for 3 clusters`
$\overline{Sil}_{\mathrm{ACF}} \leftarrow$ `Silhouette` (M_{ACF}); `// silhouette coefficient`
$M_{\mathrm{PACF}} \leftarrow$ `PAM` $(d_{\mathrm{ACF}}, 3)$;
$\overline{Sil}_{\mathrm{PACF}} \leftarrow$ `Silhouette` (M_{PACF});
$\overline{Sil}_{\mathrm{best}} \leftarrow$ `Max` $(\overline{Sil}_{\mathrm{ACF}}, \overline{Sil}_{\mathrm{PACF}})$;

`// Step 5` $\Leftarrow$ `Threshold selection`
$M_{\mathrm{best}} \leftarrow$ `MatchSilhouette2Medoid` $(\overline{Sil}_{\mathrm{best}})$;
$\Theta_{\mathrm{M}_{best}} \leftarrow$ `FindMedoidId` (M_{best}); `// choose thresholds`

Bibliography

Aberg, S. and Guttorp, P. (2007). Distribution of the maximum in air pollution fields. *Environmetrics*, 19(2):183–208, doi: 10.1002/env.866.

AghaKouchak, A., Nasrollahi, N., Li, J., Imam, B., and Sorooshian, S. (2010). Geometrical characterization of precipitation patterns. *J. Hydrometeor*, 12(2):274–285.

Ahijevych, D., Gilleland, E., Brown, B. G., and Ebert, E. E. (2009). Application of Spatial Verification Methods to Idealized and NWP-Gridded Precipitation Forecasts. *Weather and Forecasting*, 24(6):1485–1497.

Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A. M. G., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., Tagipour, A., Rupa Kumar, K., Revadekar, J., Griffiths, G., Vincent, L., Stephenson, D. B., Burn, J., Aguilar, E., Brunet, M., Taylor, M., New, M., Zhai, P., Rusticucci, M., and Vazquez-Aguirre, J. L. (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research*, 111(D5).

Ancona-Navarrete, M. A. and Tawn, J. A. (2002). Diagnostics for pairwise extremal dependence in spatial processes. *Extremes*, 5(3):271–285, doi: 10.1023/A:1024029128431.

Apputhurai, P. and Stephenson, A. G. (2013). Spatiotemporal hierarchical modelling of extreme precipitation in western australia using anisotropic gaussian random fields. *Environmental and Ecological Statistics*, 20(4):667–677, doi: 10.1007/s10651–013–0240–9.

Asadi, P., Davison, A. C., and Engelke, S. (2015). Extremes on river networks. *Ann. Appl. Statist.*, 9(4):2023–2050, doi: 10.1214/15–AOAS863.

Azais, J.-M. and Wschebor, M. (2005). On the distribution of the maximum of a gaussian field with d parameters. *Ann. Appl. Probab.*, 15(1A):254–278, doi: 10.1214/105051604000000602.

Bacro, J.-N., Bel, L., and Lantéjoul, C. (2010). Testing the independence of maxima: From bivariate vectors to spatial extreme fields. *Extremes*, 13(2):155–175, doi: 10.1007/s10687–010–0104–8.

Bader, B., Yan, J., and Zhang, X. (2017). Automated selection of r for the r largest order statistics approach with adjustment for sequential testing. *Statistics and Computing*, 27(6):1435–1451.

Bador, M., Naveau, P., Gilleland, E., Castellà, M., and Arivelo, T. (2015). Spatial clustering of summer temperature maxima from the cnrm-cm5 climate model ensembles & e-obs over europe. *Weather and Climate Extremes*, 9:17–24, doi: 10.1016/j.wace.2015.05.003.

Balkema, A. A. and de Haan, L. (1974). Residual Life Time at Great Age. *The Annals of Probability*, 2(5):792–804.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2015). *Hierarchical modeling and analysis for spatial data*. Number 135 in Monographs on statistics and applied probability. CRC Press, Taylor & Francis Group, Boca Raton, second edition edition.

Beguería, S. and Vicente-Serrano, S. M. (2006). Mapping the hazard of extreme rainfall by peaks over threshold extreme value analysis and spatial regression techniques. *Journal of Applied Meteorology and Climatology*, 45(1):108–124, doi: 10.1175/JAM2324.1.

Beirlant, J., Goegebeur, Y., Teugels, J., and Segers, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley Series in Probability and Statistics. Wiley, Chichester, West Sussex, England, U.K. 522 pp.

Bernard, E., Naveau, P., Vrac, M., and Mestre, O. (2013). Clustering of maxima: Spatial dependencies among heavy rainfall in France. *Journal of Climate*, 26(20):7929–7937, doi: 10.1175/JCLI–D–12–00836.1.

Blanchet, J. and Davison, A. C. (2011). Spatial modeling of extreme snow depth. *Ann. Appl. Stat.*, 5(3):1699–1725, doi: 10.1214/11–AOAS464.

Bocci, C., Caporali, E., and Petrucci, A. (2012). Geoadditive modeling for extreme rainfall data. *Advances in Statistical Analysis*, 97(2):181–193, doi: 10.1007/s10182–012–0192–7.

Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., and Zemp, M. (2014). The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy. *Bulletin of the American Meteorological Society*, 95(9):1431–1443.

Bonazzi, A., Cusack, S., Mitas, C., and Jewson, S. (2012). The spatial structure of European wind cyclones as characterized by bivariate exteme-value copulas. *Nat. Hazards Earth Syst. Sci.*, 12:1769–1782, doi: 10.5194/nhess–12–1769–2012.

Breiman, L. (2017). *Classification and regression trees*. Routledge.

Buhl, S. and Klüppelberg, C. (2015). Anisotropic Brown-Resnick space-time processes: estimation and model assessment. *Extremes*, 19(4):627–660, doi: 10.1007/s10687–016–0257–1.

Buishand, T. A., de Haan, L., and Zhou, C. (2008). On spatial extremes: With application to a rainfall problem. *Ann. Appl. Stat.*, 2(2):624–642, doi: 10.1214/08–AOAS159.

Bullock, R., Brown, B., and Fowler, T. (2016). Method for Object-Based Diagnostic Evaluation. Technical report, University Corporation For Atmospheric Research (UCAR):National Center for Atmospheric Research (NCAR):NCAR Library (NCARLIB). DOI: 10.5065/D61V5CBS.

Burauskaite-Harju, A. and Grimvall, A. (2013). Diagnostics for tail dependence in time-lagged random fields of precipitation. *Theoretical and Applied Climatology*, 112(3-4):629–636, doi: 10.1007/s00704–012–0748–1.

Butler, A., Heffernan, J. E., Tawn, J. A., and Flather, R. A. (2007). Trend estimation in extremes of synthetic North Sea surges. *J. R. Stat. Soc. C*, 56(4):395–414, doi: 10.1111/j.1467–9876.2007.00583.x.

Caiado, J., Crato, N., and Peña, D. (2006). A periodogram-based metric for time series classification. *Computational Statistics & Data Analysis*, 50(10):2668–2684.

Casson, E. and Coles, S. (1999). Spatial regression models for extremes. *Extremes*, 1(4):449–468, doi: 10.1023/A:1009931222386.

Cattiaux, J. and Ribes, A. (2018). Defining Single Extreme Weather Events in a Climate Perspective. *Bulletin of the American Meteorological Society*, 99(8):1557–1568.

Chen, C.-T. and Knutson, T. (2008). On the Verification and Comparison of Extreme Rainfall Indices from Climate Models. *Journal of Climate*, 21(7):1605–1621.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Chen, T., He, T., Benesty, M., and Khotilovich, V. (2019). Package 'xgboost'. *R version*, 90:1–66.

Cheng, D. and Schwartzman, A. (2015). Distribution of the height of local maxima of Gaussian random fields. *Extremes*, 18:213–240, doi: 10.1007/s10687–014–0211–z.

Coelho, C. A. S., Ferro, C. A. T., Stephenson, D. B., and Steinskog, D. J. (2008). Methods for exploring spatial and temporal variability of extreme events in climate data. *J. Climate*, 21(10):2072–2092, doi: 10.1175/2007JCLI1781.1.

Coles, S. (2001a). *An introduction to statistical modeling of extreme values*. Springer series in statistics. Springer, London ; New York.

Coles, S., Heffernan, J., and Tawn, J. (1999a). Dependence measures for extreme value analyses. *Extremes*, 2:339–365.

Coles, S. G. (1993). Regional modelling of extreme storms via max-stable processes. *J. Roy. Stat. Soc. B*, 55(4):797–816.

Coles, S. G. (2001b). *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, London, UK, 208 pp.

Coles, S. G. and Casson, E. (1998). Extreme value modelling of hurricane wind speeds. *Structural Safety*, 20(3):283–296, doi: 10.1016/S0167–4730(98)00015–0.

Coles, S. G., Heffernan, J. E., and Tawn, J. A. (1999b). Dependence measures for extreme value analysis. *Extremes*, 2:339–365, doi: 10.1023/A:1009963131610.

Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 53(2):377–92.

Coles, S. G. and Tawn, J. A. (1994). Statistical methods for multivariate extremes: an application to structural design (with discussion). *J. Roy. Stat. Soc. C (Applied Statistics)*, 43(1):1–48.

Coles, S. G. and Tawn, J. A. (1996). Modelling extremes of the areal rainfall process. *J. Roy. Stat. Soc. B*, 58(2):329–347.

Coles, S. G. and Walshaw, D. (1994). Directional modelling of extreme wind speeds. *J. Roy. Stat. Soc. C*, 43(1):139–157.

Cooley, D., Cisewski, J., Erhardt, R., Jeon, S., Mannshardt, E., Omolo, B., and Sun, Y. (2012a). A survey of spatial extremes: Measuring spatial dependence and modeling spatial effects. *REVSTAT*, 10:135–165.

Cooley, D., Cisewski, J., Erhardt, R. J., Jeon, S., Mannshardt, E., Omolo, B. O., and Sun, Y. (2012b). A survey of spatial extremes: Measuring spatial dependence and modeling spatial effects. *Revstat*, 10(1):135–165.

Cooley, D., Naveau, P., Jomelli, V., Rabatel, A., and Grancher, D. (2006a). A Bayesian hierarchical extreme value model for lichenometry. *Environmetrics*, 17(6):555–574, doi: 10.1002/env.764.

Cooley, D., Naveau, P., and Poncet, P. (2006b). *Variograms for spatial max-stable random fields*. In Dependence in Probability and Statistics. Ed. Bertail, P. and Doukhan, P. and Soulier, P. Springer Lecture Notes in Statistics no. 187.

Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102:824–840, doi: 10.1198/016214506000000780.

Cooley, D. and Sain, S. R. (2010). Spatial hierarchical modeling of precipitation extremes from a regional climate model. *JABES*, 15:381–402, doi: 10.1007/s13253–010–0023–9.

Cooley, D. and Thibaud, E. (2019). Decompositions of dependence for high-dimensional extremes. *Biometrika*, 106(3):587–604.

Craigmile, P. F., Cressie, N., Santner, T. J., and Rao, Y. (2005). A loss function approach to identifying environmental exceedances. *Extremes*, 8(3):143–159, doi: 10.1007/s10687–006–7964–y.

Cressman, G. P. (1959). AN OPERATIONAL OBJECTIVE ANALYSIS SYSTEM. *Monthly Weather Review*, 87(10):367–374.

Dai, A. (2001). Global Precipitation and Thunderstorm Frequencies. Part I: Seasonal and Interannual Variations. *Journal of Climate*, 14(6):1092–1111.

Dai, A., Rasmussen, R. M., Liu, C., Ikeda, K., and Prein, A. F. (2017). A new mechanism for warm-season precipitation response to global warming based on convection-permitting simulations. *Climate Dynamics*.

Daly, C., Neilson, R. P., and Phillips, D. L. (1994). A Statistical-Topographic Model for Mapping Climatological Precipitation over Mountainous Terrain. *Journal of Applied Meteorology*, 33(2):140–158.

Daly, C., Taylor, G., and Gibson, W. (1997). The PRISM approach to mapping precipitation and temperature. Preprints. 10th AMS Conf. on Applied Climatology, Reno, NV. *American Meteorological Society*, pages 10–12.

Davis, C., Brown, B., and Bullock, R. (2006a). Object-based verification of precipitation forecasts. part i: Methodology and application to mesoscale rain areas. *Monthly Weather Review*, 134(7):1772–1784.

Davis, C., Brown, B., and Bullock, R. (2006b). Object-based verification of precipitation forecasts. part ii: Application to convective rain systems. *Monthly Weather Review*, 134(7):1785–1795.

Davis, R. A., Klüppelberg, C., and Steinkohl, C. (2013a). Max-stable processes for modelling extremes observed in space and time. *Journal of Korean Statistical Society*, 42(3):399–414, doi: 10.1016/j.jkss.2013.01.002.

Davis, R. A., Klüppelberg, C., and Steinkohl, C. (2013b). Statistical inference for max-stable processes in space and time. *J. Roy. Stat. Soc. B*, 75(5):791–819, doi: 10.1111/rssb.12012.

Davis, R. A. and Mikosch, T. (2008). Extreme value theory for space-time processes with heavy-tailed distributions. *Stochastic Processes and their Applications*, 118(4):560–584, doi: 10.1016/j.spa.2007.06.001.

Davison, A. C. and Gholamrezaee, M. M. (2011). Geostatistics of extremes. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, pages 1–28, doi: 10.1098/rspa.2011.0412.

Davison, A. C., Huser, R., and Thibaud, E. (2013). Geostatistics of dependent and asymptotically independent extremes. *Mathematical Geosciences*, 45(5):511–529, doi: 10.1007/s11004–013–9469–y.

Davison, A. C., Padoan, S., and Ribatet, M. (2012). Statistical modelling of spatial extremes (with discussion). *Statistical Science*, 27(2):161–186, doi: 10.1214/11–STS376.

de Haan, L. (1984). A spectral representation for max-stable processes. *Ann. Probab.*, 12(4):1194–1204, doi: 10.1214/aop/1176993148.

de Haan, L. and de Ronde, J. (1998). Sea and wind: Multivariate extremes at work. *Extremes*, 1(1):1–7, doi: 10.1023/A:1009909800311.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer-Verlag, New York, N.Y., 288 pp.

de Haan, L. and Pereira, T. T. (2006). Spatial extremes: Models for the non-stationary case. *The Annals of Statistics*, 34(1):146–168, doi: 10.1214/009053605000000886.

de Haan, L. and Zhou, C. (2008). On extreme value analysis of a spatial process. *REVSTAT*, 6(1):71–81.

Dieker, A. B. and Mikosch, T. (2015). Exact simulation of Brown-Resnick random fields at a finite number of locations. *Extremes*, 18:301–314, doi: 10.1007/s10687–015–0214–4.

Diggle, P. and Ribeiro, P. J. (2007). *Model-based geostatistics*. Springer series in statistics. Springer, New York, NY.

Dombry, C., Éyi-Minko, F., and Ribatet, M. (2013). Conditional simulation of max-stable processes. *Biometrika*, 100(1):111–124, doi: 10.1093/biomet/ass067.

Donat, M., Alexander, L., Yang, H., Durre, I., Vose, R., and Caesar, J. (2013a). Global Land-Based Datasets for Monitoring Climatic Extremes. *Bulletin of the American Meteorological Society*, 94(7):997–1006.

Donat, M. G., Alexander, L. V., Yang, H., Durre, I., Vose, R., Dunn, R. J. H., Willett, K. M., Aguilar, E., Brunet, M., Caesar, J., Hewitson, B., Jack, C., Klein Tank, A. M. G., Kruger, A. C., Marengo, J., Peterson, T. C., Renom, M., Oria Rojas, C., Rusticucci, M., Salinger, J., Elrayah, A. S., Sekele, S. S., Srivastava, A. K., Trewin, B., Villarroel, C., Vincent, L. A., Zhai, P., Zhang, X., and Kitching, S. (2013b). Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset:

HADEX2-GLOBAL GRIDDED CLIMATE EXTREMES. *Journal of Geophysical Research: Atmospheres*, 118(5):2098–2118.

Dragomiretskiy, K. and Zosso, D. (2013). Variational mode decomposition. *IEEE transactions on signal processing*, 62(3):531–544.

Eastoe, E. F. (2009). A hierarchical model for non-stationary multivariate extremes: a case study of surface-level ozone and NO_x data in the UK. *Environmetrics*, 20(4):428–444, doi: 10.1002/env.938.

Economou, T., Stephenson, D. B., and Ferro, C. A. T. (2014). Spatio-temporal modelling of extreme storms. *Annals of Applied Statistics*, 8(4):2223–2246, doi: 10.1214/14–AOAS766.

Engelke, S., Malinowski, A., Kabluchko, Z., and Schlather, M. (2014). Estimation of hüsler–reiss distributions and brown–resnick processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):239–265, doi: 10.1111/rssb.12074.

Fawcett, L. and Walshaw, D. (2006). A hierarchical model for extreme wind speeds. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55(5):631–646, doi: 10.1111/j.1467–9876.2006.00557.x.

Föllmer, H. (2014). Spatial risk measures and their local specification: the locally law-invariant case. *Statistics and Risk Modeling*, 31(1):79–101, doi: 10.1515/strm–2013–5001.

Fowler, H. J., Ekström, M., Kilsby, C. G., and Jones, P. D. (2005). New estimates of future changes in extreme rainfall across the UK using regional climate model integrations. *J. Hydrol. Amsterdam*, 300:212–233, doi: 10.1016/j.jhydrol.2004.06.017.

French, J. P. and Davis, R. A. (2013). The asymptotic distribution of the maxima of a Gaussian random field on a lattice. *Extremes*, 16:1–26, doi: 10.1007/s10687–012–0149–y.

Friederichs, P. and Thorarinsdottir, T. L. (2012). Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction: VERIFICATION FOR EXTREME VALUE DISTRIBUTIONS. *Environmetrics*, 23(7):579–594.

Fuentes, M., Henry, J., and Reich, B. (2013). Nonparametric spatial models for extremes: application to extreme temperature data. *Extremes*, 16(1):75–101, doi: 10.1007/s10687–012–0154–1.

Fukutome, S., Liniger, M. A., and Süveges, M. (2015). Automatic threshold and run parameter selection: a climatology for extreme hourly precipitation in switzerland. *Theoretical and Applied Climatology*, 120(3):403–416.

Gaetan, C. and Grigoletto, M. (2007). A hierarchical model for the analysis of spatial rainfall extremes. *J. Agricult., Biolog. and Environ. Stat.*, 12(4):434–449, doi: 10.1198/108571107X250193.

Galeano, P. and Peña, D. P. (2000). Multivariate analysis in vector time series. *Resenhas do Instituto de Matemática e Estatística da Universidade de São Paulo*, 4(4):383–403.

Genton, M. G., Ma, Y., and Sang, H. (2011). On the likelihood function of Gaussian max-stable processes. *Biometrika*, 98:481–488, doi: 10.1093/biomet/asr020.

Genton, M. G., Padoan, S. A., and Sang, H. (2015). Multivariate max-stable spatial processes. *Biometrika*, 102(1):215–230, doi: 10.1093/biomet/asu066.

Ghosh, S. and Mallick, B. K. (2008). A hierarchical bayesian spatio-temporal model for extreme precipitation events. *Environmetrics*, 22(2):192–204, doi: 10.1002/env.1043.

Gilleland, E. (2017a). A New Characterization within the Spatial Verification Framework for False Alarms, Misses, and Overall Patterns. *Weather and Forecasting*, 32(1):187–198.

Gilleland, E. (2017b). Spatialvx: Spatial forecast verification. R package version 0.6-1, doi: 10.5065/4px3-5a05.

Gilleland, E., Ahijevych, D., Brown, B. G., Casati, B., and Ebert, E. E. (2009a). Intercomparison of Spatial Forecast Verification Methods. *Weather and Forecasting*, 24(5):1416–1430.

Gilleland, E., Ahijevych, D., Brown, B. G., Casati, B., and Ebert, E. E. (2009b). Intercomparison of spatial forecast verification methods. *Weather and forecasting*, 24(5):1416–1430.

Gilleland, E., Ahijevych, D. A., Brown, B. G., and Ebert, E. E. (2010). Verifying Forecasts Spatially. *Bulletin of the American Meteorological Society*, 91(10):1365–1373.

Gilleland, E., Brown, B. G., and Ammann, C. M. (2013). Spatial extreme value analysis to project extremes of large-scale indicators for severe weather. *Environmetrics*, 24(6):418–432, doi: 10.1002/env.2234.

Gilleland, E., Bukovsky, M., Williams, C. L., McGinnis, S., Ammann, C. M., Brown, B. G., and Mearns, L. O. (2016a). Evaluating NARCCAP model performance for frequencies of severe-storm environments. *Advances in Statistical Climatology, Meteorology and Oceanography*, 2(2):137–153.

Gilleland, E., Bukovsky, M., Williams, C. L., McGinnis, S., Ammann, C. M., Brown, B. G., and Mearns, L. O. (2016b). Evaluating NARCCAP model performance for frequencies of severe-storm environments. *Advances in Statistical Climatology, Meteorology and Oceanography*, 2(2):137–153, doi: 10.5194/ascmo–2–137–2016.

Gilleland, E. and Katz, R. W. (2016). extRemes 2.0: An Extreme Value Analysis Package in R. *Journal of Statistical Software, Articles*, 72(8):1–39, doi: 10.18637/jss.v072.i08.

Gilleland, E. and Nychka, D. (2005). Statistical models for monitoring and regulating ground-level ozone. *Environmetrics*, 16(5):535–546, doi: 10.1002/env.720.

Gilleland, E., Nychka, D., and Schneider, U. (2006). Spatial models for the distribution of extremes. *Hierarchical Modelling for the Environmental Sciences: Statistical Methods and Applications*, 216:170–183.

Gregersen, I. B., Madsen, H., Rosbjerg, D., and Arnbjerg-Nielsen, K. (2013). A spatial and nonstationary model for the frequency of extreme rainfall events. *Water Resources Research*, 49(1):127–136, doi: 10.1029/2012WR012570.

Groisman, P. Y., Knight, R. W., Easterling, D. R., Karl, T. R., Hegerl, G. C., and Razuvaev, V. N. (2005). Trends in Intense Precipitation in the Climate Record. *Journal of Climate*, 18(9):1326–1350.

Hainy, M., Müller, W. G., and Wagner, H. (2015). Likelihood-free simulation-based optimal design with an application to spatial extremes. *Stochastic Environmental Research and Risk Assessment*, 30(2):481–492, doi: 10.1007/s00477–015–1067–8.

Hanel, M., Buishand, T. A., and Ferro, C. A. T. (2009). A nonstationary index flood model for precipitation extremes in transient regional climate model simulations. *J. Geophys. Res.-Atmos.*, 114(D15107):1–16, doi: 10.1029/2009JD011712.

Harris, D., Foufoula-Georgiou, E., Droegemeier, K. K., and Levit, J. J. (2001). Multiscale statistical properties of a high-resolution precipitation forecast. *Journal of Hydrometeorology*, 2(4):406–418, doi: 10.1175/1525–7541(2001)002¡0406:MSPOAH¿2.0.CO;2.

Heaton, M. J., Katzfuss, M., Ramachandar, S., Pedings, K., Gilleland, E., Mannshardt-Shamseldin, E., and Smith, R. L. (2011). Spatio-temporal models for large-scale indicators of extreme weather. *Environmetrics*, 22(3):294–303, doi: 10.1002/env.1050.

Heffernan, J. E. and Resnick, S. I. (2007). Limit laws for random vectors with an extreme component. *Ann. Appl. Probab.*, 17(2):537–571, doi:10.1214/105051606000000835.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):497–546, doi: 10.1111/j.1467–9868.2004.02050.x.

Hrafnkelsson, B., Morris, J. S., and Baladanadayuthapani, V. (2012). Spatial modeling of annual minimum and maximum temperatures in iceland. *Meteorol. Atmos. Phys.*, 116(1–2):43–61, doi: 10.1007/s00703–010–0101–0.

Hsing, T., Hüsler, J., and Reiss, R. (1996). The extremes of a triangular array of normal random variables. *Ann. Appl. Probab.*, 6:671–686, doi: 10.1214/aoap/1034968149.

Huerta, G. and Sansó, B. (2007). Time-varying models for extreme values. *Environmental and Ecological Statistics*, 14(3):285–299, doi: 10.1007/s10651–007–0014–3.

Huser, R. and Davison, A. C. (2013). Composite likelihood estimation for the Brown–Resnick process. *Biometrika*, 100(2):511–518, doi: 10.1093/biomet/ass089.

Huser, R. and Davison, A. C. (2014). Space-time modelling of extreme events. *J. Roy. Stat. Soc. B*, 76(2):439–461, doi: 10.1111/rssb.12035.

Huser, R. and Genton, M. G. (2016). Non-stationary dependence structures for spatial extremes. *Journal of Agricultural, Biological, and Environmental Statistics*, 21(3):470–491, doi: 10.1007/s13253–016–0247–4.

Huser, R., Opitz, T., and Thibaud, E. (2017). Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures. *Spatial Statistics*, 21:166–186, doi: 10.1016/j.spasta.2017.06.004.

Huser, R. and Wadsworth, J. L. (2022). Advances in statistical modeling of spatial extremes. *Wiley Interdisciplinary Reviews: Computational Statistics*, 14(1):e1537.

IPCC (2012). *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, UK, and New York, NY, USA. 582 pp.

Jiang, Y., Cooley, D., and Wehner, M. F. (2020). Principal component analysis for extremes and application to us precipitation. *Journal of Climate*, 33(15):6441–6451.

Joe, H. (1997). *Multivariate models and dependence concepts*, volume 73. Chapman & Hall, London, 424 pp., London.

Jolliffe, I. T. and Stephenson, D. B., editors (2003). *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. J. Wiley, Chichester, West Sussex, England ; Hoboken, NJ.

Jonathan, P., Ewans, K., and Flynn, J. (2014a). On the estimation of ocean engineering design contours. *Journal of Offshore Mechanics and Arctic Engineering*, 136(4):041101–041101–8, doi: 10.1115/1.4027645.

Jonathan, P. and Ewans, K. C. (2011). A spatiodirectional model for extreme waves in the Gulf of Mexico. *Journal of Offshore Mechanics and Arctic Engineering*, 133(1):1–9, doi: 10.1115/1.4001949.

Jonathan, P., Ewans, K. C., and Randell, D. (2014b). Non-stationary conditional extremes of northern North Sea storm characteristics. *Environmetrics*, 25(3):172–188.

Kabluchko, Z., Schlather, M., and de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann. Probab.*, 37(5):2042–2065, doi: 10.1214/09–AOP455.

Katz, R. W., Parlange, M. B., and Naveau, P. (2002). Statistics of extremes in hydrology. *Water Resources Research*, 25:1287–1304.

Katz, R. W., Parlange, M. B., and Tebaldi, C. (2003). Stochastic modeling of the effects of large-scale circulation on daily weather in the southeastern U.S. *Climatic Change*, 60:189–216, doi: 10.1023/A:1026054330406.

Kaufman, L. and Rousseeuw, P. (1987). Clustering by means of medoids, *Statistical data analysis based on the L_1-norm and related methods*. North-Holland; Sole distributors for the U.S.A. and Canada, Elsevier Science Pub. Co, Amsterdam ; New York : New York, N.Y., U.S.A.

Kaufman, L. and Rousseeuw, P. J., editors (1990). *Partitioning Around Medoids (Program PAM).* John Wiley & Sons, Inc., Hoboken, NJ, USA. DOI: 10.1002/9780470316801.ch2.

Kaur, D. and Kaur, Y. (2014). Various image segmentation techniques: A review. *International Journal of Computer Science and Mobile Computing*, 3(5):809–814.

Keef, C., Papastathopoulos, I., and Tawn, J. A. (2013a). Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model. *Journal of Multivariate Analysis*, 115:396–404, doi: 10.1016/j.jmva.2012.10.012.

Keef, C., Svensson, C., and Tawn, J. A. (2009a). Spatial dependence in extreme river flows and precipitation for Great Britain. *Journal of Hydrology*, 378(3–4):240–252, doi: 10.1016/j.jhydrol.2009.09.026.

Keef, C., Tawn, J. A., and Lamb, R. (2013b). Estimating the probability of widespread flood events. *Environmetrics*, 24(1):13–21, doi: 10.1002/env.2190.

Keef, C., Tawn, J. A., and Svensson, C. (2009b). Spatial risk assessment for extreme river flows. *J. Roy. Stat. Soc. C (Applied Statistics)*, 58(5):601–618, doi: 10.1111/j.1467–9876.2009.00672.x.

Knutson, T., Kossin, J., Mears, C., Perlwitz, J., Wehner, M., Wuebbles, D., Fahey, D., Hibbard, K., Dokken, D., Stewart, B., and Maycock, T. (2017). Ch. 3: Detection and Attribution of Climate Change. Climate Science Special Report: Fourth National Climate Assessment, Volume I. Technical report, U.S. Global Change Research Program. DOI: 10.7930/J01834ND.

Koch, E. (2017). Spatial risk measures and applications to max-stable processes. *Extremes*, 20(3):635–670, doi: 10.1007/s10687–016–0274–0.

Kottas, A. and Sansó, B. (2007). Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis. *Journal of Statistical Planning and Inference*, 137:3151–3163, doi: 10.1016/j.jspi.2006.05.022.

Kunkel, K. E., Easterling, D. R., Kristovich, D. A., Gleason, B., Stoecker, L., and Smith, R. (2010). Recent increases in U.S. heavy precipitation associated with tropical cyclones: HEAVY PRECIPITATION AND TROPICAL CYCLONES. *Geophysical Research Letters*, 37(24):n/a–n/a.

Kunkel, K. E., Easterling, D. R., Kristovich, D. A. R., Gleason, B., Stoecker, L., and Smith, R. (2012). Meteorological Causes of the Secular Variations in Observed Extreme Precipitation Events for the Conterminous United States. *Journal of Hydrometeorology*, 13(3):1131–1141.

Kunkel, K. E., Easterling, D. R., Redmond, K., and Hubbard, K. (2003). Temporal variations of extreme precipitation events in the United States: 1895-2000: TEMPORAL VARIATIONS IN THE U.S. *Geophysical Research Letters*, 30(17):n/a–n/a.

Kunkel, K. E., Karl, T. R., and Easterling, D. R. (2007). A Monte Carlo Assessment of Uncertainties in Heavy Precipitation Frequency Variations. *Journal of Hydrometeorology*, 8(5):1152–1160.

Lau, W. K.-M., Wu, H.-T., and Kim, K.-M. (2013). A canonical response of precipitation characteristics to global warming from CMIP5 models: PRECIPITATION AND GLOBAL WARMING. *Geophysical Research Letters*, 40(12):3163–3169.

Liang, X.-Z., Li, L., Kunkel, K. E., Ting, M., and Wang, J. X. (2004). Regional climate model simulation of US precipitation during 1982–2002. Part I: Annual cycle. *Journal of Climate*, 17(18):3510–3529.

Liang, X.-Z., Xu, M., Yuan, X., Ling, T., Choi, H. I., Zhang, F., Chen, L., Liu, S., Su, S., Qiao, F., et al. (2012a). Regional climate–weather research and forecasting model. *Bulletin of the American Meteorological Society*, 93(9):1363–1387.

Liang, X.-Z., Xu, M., Yuan, X., Ling, T., Choi, H. I., Zhang, F., Chen, L., Liu, S., Su, S., Qiao, F., He, Y., Wang, J. X. L., Kunkel, K. E., Gao, W., Joseph, E., Morris, V., Yu, T.-W., Dudhia, J., and Michalakes, J. (2012b). Regional Climate–Weather Research and Forecasting Model. *Bulletin of the American Meteorological Society*, 93(9):1363–1387.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2017). *cluster: Cluster Analysis Basics and Extensions*. R package version 2.0.6 — For new features, see the 'Changelog' file (in the package source).

Mannshardt-Shamseldin, E. C., Smith, R. L., Sain, S. R., Mearns, L. O., and Cooley, D. (2010). Downscaling extremes: A comparison of extreme value distributions in point-source and gridded precipitation data. *Ann. Appl. Statist.*, 4:484–502, doi: 10.1214/09–AOAS287.

Mendes, B. and Pericchi, L. (2008). Assessing conditional extremal risk of flooding in Puerto Rico. *Stoch. Environ. Res. Risk. Assess.*, 23(3):399–410, doi: 10.1007/s00477–008–0220–z.

Mendes, J. M., Cortés de Zea Bermudez, P., Pereira, J., Turkman, K. F., and Vasconcelos, M. J. P. (2010). Spatial extremes of wildfire sizes: Bayesian hierarchical models for extremes. *Environ. Ecol. Stat.*, 17:1–28, doi: 10.1007/s10651–008–0099–3.

Menne, M. J., Durre, I., Korzeniewski, B., McNeill, S., Thomas, K., Yin, X., Anthony, S., Ray, R., Vose, R. S., Gleason, B. E., and Houston, T. G. (2012a). Global Historical Climatology Network - Daily (GHCN-Daily), Version 3. DOI: 10.7289/v5d21vhz.

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., and Houston, T. G. (2012b). An Overview of the Global Historical Climatology Network-Daily Database. *Journal of Atmospheric and Oceanic Technology*, 29(7):897–910.

Mercadier, C. (2006). Numerical bounds for the distributions of the maxima of some one- and two-parameter gaussian processes. *Advances in Applied Probability*, 38(1):149–170, doi: 10.1239/aap/1143936145.

Min, S.-K., Zhang, X., Zwiers, F. W., and Hegerl, G. C. (2011). Human contribution to more-intense precipitation extremes. *Nature*, 470:378 EP.

Montero, P. and Vilar, J. A. (2014). TSclust: An R package for time series clustering. *Journal of Statistical Software*, 62(1):1–43.

Naveau, P., Guillou, A., Cooley, D., and Diebolt, J. (2009). Modelling pairwise dependence of maxima in space. *Biometrika*, 96(1):1–17, doi: 10.1093/biomet/asp001.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer series in statistics. Springer, New York, 2nd edition.

Northrop, P. J. and Jonathan, P. (2010). Threshold modeling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics*, 22(7):799–809, doi: 10.1002/env.1106.

Oesting, M., Schlather, M., and Friederichs, P. (2017). Statistical post-processing of forecasts for extremes using bivariate brown-resnick processes with an application to wind gusts. *Extremes*, 20(2):309–332, doi: 10.1007/s10687–016–0277–x.

Olinda, R. A., Blanchet, J., dos Santos, C. A. C., Ozaki, V. A., and Ribeiro, P. J. (2014). Spatial extremes modeling applied to extreme precipitation data in the state of Paraná. *Hydrol. Earth Syst. Sci. Discuss.*, 11:12731–12764, doi: 10.5194/hessd–11–12731–2014.

Oliver, E. C. J., Wotherspoon, S. J., and Holbrook, N. J. (2014). Estimating extremes from global ocean and climate models: A Bayesian hiearchical model approach. *Progress in Oceanography*, 122:77–91, doi: 10.1016/j.pocean.2013.12.004.

Padoan, S. A., Ribatet, M., and Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489):263–277, doi: 10.1198/jasa.2009.tm08577.

Palutikof, J. P., Brabson, B. B., Lister, D. H., and Adcock, S. T. (1999). A review of methods to calculate extreme wind speeds. *Meteorological Applications*, 6(2):119–132, doi: 10.1017/S1350482799001103.

Peli, T. and Malah, D. (1982). A study of edge detection algorithms. *Computer Graphics and Image Processing*, 20(1):1–21.

Pendergrass, A. G. and Hartmann, D. L. (2014). Changes in the Distribution of Rain Frequency and Intensity in Response to Global Warming*. *Journal of Climate*, 27(22):8372–8383.

Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, 3(1):119–131.

Poon, S.-H., Rockinger, M., and Tawn, J. (2004). Extreme value dependence in financial markets: Diagnostics, models, and financial implications. *The Review of Financial Studies*, 17(2):581–610, doi: 10.1093/rfs/hhg058.

Reich, B. J., Cooley, D. S., Foley, K. M., Napelenok, S. L., and Shaby, B. A. (2013). Extreme value analysis for evaluating ozone control strategies. *Annals of Applied Statistics*, 7:739–762, doi: 10.1214/13–AOAS628.

Reich, B. J., Fuentes, M., and Dunson, D. B. (2011). Bayesian spatial quantile regression. *Journal of the American Statistical Association*, 106:6–20, doi: 10.1198/jasa.2010.ap09237.

Reich, B. J. and Shaby, B. A. (2012). A hierarchical max-stable spatial model for extreme precipitation. *The Annals of Applied Statistics*, 6(4):1430–1451, doi: 10.1214/12–aoas591.

Reiss, R. and Thomas, M. (2007). *Statistical Analysis of Extreme Values: with applications to insurance, finance, hydrology and other fields*. Birkhäuser, 530pp., 3rd edition.

Ribatet, M. (2013). Spatial extremes: max-stable processes at work. *Journal de la Société Françise de Statistique*, 154(2):156–177.

Ribatet, M., Cooley, D. S., and Davison, A. C. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, 22:813–845, doi: 10.5705/ss.2009.248.

Ribes, A., Zwiers, F. W., Azaïs, J.-M., and Naveau, P. (2017). A new statistical approach to climate change detection and attribution. *Climate Dynamics*, 48(1-2):367–386.

Rietsch, T., Naveau, P., Gilardi, N., and Guillou, A. (2013). Network design for heavy rainfall analysis. *J. Geophys. Res.*, 118(23):13075–13086, doi: 10.1002/2013JD020867.

Rootzén, H., Segers, J., and L. Wadsworth, J. (2018a). Multivariate peaks over thresholds models. *Extremes*, 21(1):115–145.

Rootzén, H., Segers, J., and Wadsworth, J. L. (2018b). Multivariate generalized pareto distributions: Parametrizations, representations, and properties. *Journal of Multivariate Analysis*, 165:117–131.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.

Sahoo, P. K., Soltani, S., and Wong, A. K. (1988). A survey of thresholding techniques. *Computer vision, graphics, and image processing*, 41(2):233–260.

Sang, H. (2008). *Extreme value modeling for space-time data with meteorological applications*. Ph.D. thesis, Duke University, 150 pp.

Sang, H. and Gelfand, A. (2010). Continuous spatial process models for spatial extreme values. *Journal of Agricultural, Biological, and Environmental Statistics*, 15:49–65, doi: 10.1007/s13253–009–0010–1. 10.1007/s13253-009-0010-1.

Sang, H. and Gelfand, A. E. (2009). Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics*, 16(3):407–426, doi: 10.1007/s10651–007–0078–0.

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT–Statistical Journal*, 10(1):33–60.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, 5(1):33–44, doi: 10.1023/A:1020977924878.

Schlather, M. and Tawn, J. A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, 90(1):139–156, doi: 10.1093/biomet/90.1.139.

Schliep, E. M., Cooley, D., Sain, S. R., and Hoeting, J. A. (2010). A comparison study of extreme precipitation from six different regional climate models via spatial hierarchical modeling. *Extremes*, 13:219–239, doi: 10.1007/s10687–009–0098–2.

Sezgin, M. et al. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic imaging*, 13(1):146–168.

Shaby, B. A. (2014). The open-faced sandwich adjustment for MCMC using estimating functions. *Journal of Computational and Graphical Statistics*, 23(3):853–876, doi: 10.1080/10618600.2013.842174.

Shaby, B. A. and Reich, B. J. (2012). Bayesian spatial extreme value analysis to assess the changing risk of concurrent high temperatures across large portions of European cropland. *Environmetrics*, 23:638–648, doi: 10.1002/env.2178.

Shang, H., Yan, J., and Zhang, X. (2011). El Niño–Southern oscillation influence on winter maximum daily precipitation in California in a spatial model. *Water Resources Research*, 47(11):1–9, doi: 10.1029/2011WR010415.

Sibson, R. (1981). Interpolating multivariate data. In *A Brief Description of Natural Neighbor Interpolation*, chapter 2, pages 21–36. John Wiley & Sons, New York.

Smith, E. L. and Stephenson, A. G. (2009). An extended gaussian max-stable process model for spatial extremes. *Journal of Statistical Planning and Inference*, 139(4):1266–1275, doi: 10.1016/j.jspi.2008.08.003.

Smith, R. L. (1987). Estimating Tails of Probability Distributions. *The Annals of Statistics*, 15(3):1174–1207.

Smith, R. L. (1990). Max-stable processes and spatial extremes. *Unpublished Manuscript*, Available at: `http://www.stat.unc.edu/postscript/rs/spatex.pdf` (downloaded 21 June 2018):1–32.

Smith, R. L. (2003). Statistics of extremes, with applications in environment, insurance and finance. *Extreme values in finance, telecommunications and the environment*, 1:78.

Staid, A., Pinson, P., and Guikema, S. D. (2014). Probabilistic maximum-valued wind prediction for offshore environments. *Wind Energy*, 18(10):1725–1738, doi: 10.1002/we.1787.

Stephenson, A. G. (2009). High-dimensional parametric modelling of multivariate extreme events. *Australian & New Zealand Journal of Statistics*, 51(1):77–88, doi: 10.1111/j.1467–842X.2008.00528.x.

Stephenson, A. G., Shaby, B. A., Reich, B. J., and Sullivan, A. L. (2014). Estimating spatially varying severity thresholds of a forest fire danger rating system using max-stable extreme-event modeling. *Journal of Applied Meteorology and Climatology*, 54(2):395–407.

Stephenson, A. G. and Tawn, J. A. (2005). Exploiting occurrence times in likelihood inference for componentwise maxima. *Biometrika*, 92(1):213–227, doi: 10.1093/biomet/92.1.213.

Stephenson, D. (2008). Definition, diagnosis, and origin of extreme weather and climate events. In *Climate Extremes and Society*, Diaz, Henry F. and Murnane, Richard J. (eds.). Cambridge University Press, Cambridge, New York. 348.

Strokorb, K., Ballani, F., and Schlather, M. (2015). Tail correlation functions of max-stable processes: Construction principles, recovery and diversity of some mixing max-stable processes with identical TCF. *Extremes*, 18:241–271, doi: 10.1007/s10687–014–0212–y.

Sun, C. and Liang, X.-Z. (2020). Improving US extreme precipitation simulation: sensitivity to physics parameterizations. *Climate Dynamics*, 54(11-12):4891–4918.

Sun, Y., Solomon, S., Dai, A., and Portmann, R. W. (2007). How Often Will It Rain? *Journal of Climate*, 20(19):4801–4818.

Therneau, T., Atkinson, B., and Ripley, B. (2017). *rpart: Recursive Partitioning and Regression Trees*. R package version 4.1-11.

Thibaud, E., Aalto, J., Cooley, D. S., Davison, A. C., and Heikkinen, J. (2016). Bayesian inference for the Brown-Resnick process, with an application to extreme low temperatures. *Ann. Appl. Stat.*, 10(4):2303–2324, doi: 10.1214/16–AOAS980.

Thibaud, E., Mutzner, R., and Davison, A. C. (2013). Threshold modeling of extreme spatial rainfall. *Water Resources Research*, 49(8):4633–4644, doi: 10.1002/wrcr.20329.

Thibaud, E. and Opitz, T. (2015). Efficient inference and simulation for elliptical Pareto processes. *Biometrika*, 102(4):855–870, doi: 10.1093/biomet/asv045.

Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423.

Tippett, M. K. and Cohen, J. E. (2016). Tornado outbreak variability follows Taylor's power law of fluctuation scaling and increases dramatically with severity. *Nature Communications*, 7:10668 EP 1–7, doi: 10.1038/ncomms10668.

Trapp, R. J., Diffenbaugh, N. S., Brooks, H. E., Baldwin, M. E., Robinson, E. D., and Pal, J. S. (2007). Changes in severe thunderstorm environment frequency during the 21st century caused by anthropogenically enhanced global radiative forcing. *Proceedings of the National Academy of Sciences*, 104(50):19719–19723, doi: 10.1073/pnas.0705494104.

Trenberth, K. E. and Hurrell, J. W. (1994). Decadal atmosphere-ocean variations in the pacific. *Climate Dynamics*, 9:303–319.

Trenberth, K. E. and Stepaniak, D. P. (2001). Indices of el niño evolution. *Journal of climate*, 14(8):1697–1701.

Turkman, K. F., Turkman, M. A. A., and Pereira, J. M. (2010). Asymptotic models and inference for extremes of spatio-temporal data. *Extremes*, 13(4):375–397, doi: 10.1007/s10687–009–0092–8.

Tye, M. R. and Cooley, D. S. (2015). A spatial model to examine rainfall extremes in Colorado's Front Range. *Journal of Hydrology*, 530:15–23, doi: 10.1016/j.jhydrol.2015.09.023.

Vannitsem, S. and Naveau, P. (2007). Spatial dependences among precipitation maxima over belgium. *Nonlin. Processes Geophys.*, 14:621–630.

Varin, C., Reid, N. M., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42.

Wadsworth, J. L. and Tawn, J. (2022). Higher-dimensional spatial extremes via single-site conditioning. *Spatial Statistics*, 51:100677.

Wadsworth, J. L. and Tawn, J. A. (2012). Dependence modelling for spatial extremes. *Biometrika*, 99(2):253–272, doi: 10.1093/biomet/asr080.

Wadsworth, J. L. and Tawn, J. A. (2013). A new representation for multivariate tail probabilities. *Bernoulli*, 19(5B):2689–2714, doi:10.3150/12–BEJ471.

Wadsworth, J. L. and Tawn, J. A. (2014). Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika*, 101(1):1–15, doi: 10.1093/biomet/ast042.

Wang, Y. and Stoev, S. (2010). On the structure and representations of max-stable processes. *Adv. in Appl. Probab.*, 42:855–877, doi: 10.1239/aap/1308662488.

Westra, S. and Sisson, S. A. (2011). Detection of non-stationarity in precipitation extremes using a max-stable process model. *J. Hydrology*, 406:119–128, doi: 10.1016/j.jhydrol.2011.06.014.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences.* Number v. 100 in International geophysics series. Elsevier/Academic Press, Amsterdam ; Boston, 3rd ed edition.

Yaglom, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions. Volume 1: Basic Results*. Springer-Verlag, New York, N.Y., 526 pp.

Yee, T. W. and Stephenson, A. G. (2007). Vector generalized linear and additive extreme value models. *Extremes*, 10(1):1–19, doi: 0.1007/s10687–007–0032–4.

Yee, T. W. and Wild, C. J. (1996). Vector generalized additive models. *J. Roy. Stat. Soc. B*, 58(3):481–493.

Yogamangalam, R. and Karthikeyan, B. (2013). Segmentation techniques comparison in image processing. *International Journal of Engineering and Technology (IJET)*, 5(1):307–313.

Zhang, J., Craigmile, P. F., and Cressie, N. (2008). Loss function approaches to predict a spatial quantile and its exceedance region. *Technometrics*, 50(2):216–227, doi: 10.1198/004017008000000226.

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., Trewin, B., and Zwiers, F. W. (2011). Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdisciplinary Reviews: Climate Change*, 2(6):851–870.

Zheng, F., Thibaud, E., Leonard, M., and Westra, S. (2015). Assessing the performance of the independence method in modeling spatial extreme rainfall. *Water Resources Research*, 51:7744–7758, doi: 10.1002/2015WR016893.

Zhu, M., Lakshmanan, V., Zhang, P., Hong, Y., Cheng, K., and Chen, S. (2011). Spatial verification using a true metric. *Atmospheric Research*, 102(4):408–419.

Zivot, E. and Wang, J. (2006). *Modeling financial time series with S-PLUS*, volume 2. Springer.