BONNER METEOROLOGISCHE ABHANDLUNGEN

Heft 98 (2025) (ISSN 0006-7156)
Herausgeber: Andreas Hense

Svenja Szemkus

COMPACT DESCRIPTION FOR HIGH-RESOLUTION SPATIAL WEATHER EXTREMES

BONNER METEOROLOGISCHE ABHANDLUNGEN

Heft 98 (2025) (ISSN 0006-7156)
Herausgeber: Andreas Hense

Svenja Szemkus

COMPACT DESCRIPTION FOR HIGH-RESOLUTION SPATIAL WEATHER EXTREMES

Compact Description for High-Resolution Spatial Weather Extremes

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
M.Sc
Svenja Szemkus
aus
Köln

Bonn, Juli, 2024

Diese Arbeit ist die ungekürzte Fassung einer der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn im Jahr 2024 vorgelegten Dissertation von Svenja Szemkus aus Köln.

This paper is the unabridged version of a dissertation thesis submitted by Svenja Szemkus born in Cologne to the Faculty of Mathematical and Natural Sciences of the Rheinische Friedrich-Wilhelms-Universität Bonn in 2024.

Anschrift des Verfassers: Address of the author:

Svenja Szemkus
Institut für Geowissenschaften, Abt. Meteorologie
Universität Bonn
Auf dem Hügel 20
D-53121 Bonn

Gutachterin/Betreuerin: PD Dr. Petra Friederichs,
Rheinische Friedrich-Wilhelms-Universität Bonn
Gutachter: Prof. Dr. Andreas Hense,
Rheinische Friedrich-Wilhelms-Universität Bonn

Tag der Promotion: 17. Oktober 2024

Abstract

Detecting local climate change signals, particularly within the context of extreme weather events, is challenging due to the significant internal variability of the climate system. This thesis, part of the BMBF-funded ClimXtreme Phase I project, aims to improve the signal-to-noise ratio of climate change signals during extreme weather events through advanced data compression methods. Our approach is twofold, emphasizing data-adaptive techniques and spectral decomposition. These methods are evaluated for their effectiveness in describing heat waves, droughts, and precipitation extremes.

We first analyze Principal Component Analysis (PCA). The focus on extremes is achieved using the tail pairwise dependence matrix (TPDM) proposed by Cooley and Thibaud (2019). Applying this method to daily temperature maxima and meteorological droughts of varying durations, we identify an effective technique for analyzing and compactly describing large-scale multivariate weather extremes. Additionally, we introduce the cross-TPDM to identify patterns of concurrent extremes across two variables. We propose an extreme pattern index (EPI) that provides a pattern-based spatial aggregation of extremes and demonstrate that a heat wave definition based on EPI effectively detects major heat waves across Europe. For addressing simultaneous extremes in two variables, we extend this approach by introducing the threshold-based EPI (TEPI). Using the European heat waves of 2003 and 2010 as examples, we show that TEPI describes the large-scale compound character of heat waves and droughts.

Next, we examine wavelet transformation, specifically utilizing the complex dual-tree wavelet, which has proven efficient for precipitation fields (see Brune et al., 2021). Focusing on extreme precipitation, we find that the wavelet transform accurately represents these extremes without requiring specific adaptations to method or data. Comparing two European reanalyses (COSMO-REA6, CERRA) for their representation of hourly precipitation, we discover that CERRA cannot accurately resolve small-scale convective events. A scale-aware detection study for three regions in northern, southern and western Germany reveals consistent trends of increasing intensity in small-scale events during summer and in large-scale events during winter.

Kurzzusammenfassung

Die Identifizierung lokaler Signale des Klimawandels, insbesondere im Zusammenhang mit extremen Wetterereignissen, ist aufgrund der hohen internen Variabilität des Klimasystems eine Herausforderung. Diese Dissertation ist Teil des vom BMBF geförderten Projektes ClimXtreme Phase I, mit dem Ziel das Signal-Rausch-Verhältnis von Klimasignalen bei Extremwetterereignissen durch den Einsatz fortgeschrittener Datenkompressionsverfahren zu verbessern. Unser Ansatz ist zweigeteilt und konzentriert sich auf datenadaptive Techniken und spektrale Zerlegung. Diese Methoden werden auf ihre Wirksamkeit bei der Beschreibung von Hitzewellen, Dürren und Niederschlagsextremen untersucht.

Zunächst analysieren wir die Hauptkomponentenanalyse (PCA). Der Fokus auf Extreme wird durch die von Cooley und Thibaud (2019) vorgeschlagene Tail Pairwise Dependence Matrix (TPDM) erreicht. Durch die Anwendung dieser Methode auf tägliche Temperaturmaxima und meteorologische Dürren unterschiedlicher Dauer identifizieren wir eine effektive Technik zur Analyse und kompakten Beschreibung großräumiger multivariater Wetterextreme. Zusätzlich führen wir die Cross-TPDM ein, um Muster von gleichzeitigen Extremen in zwei Variablen

zu identifizieren. Wir schlagen einen Extreme Pattern Index (EPI) vor, eine musterbasierte räumliche Aggregation von Extremen. Um gemeinsame Extreme in zwei Variablen zu erfassen, erweitern wir diesen Ansatz durch die Einführung eines Schwellwert-basierten EPI (TEPI). Am Beispiel der europäischen Hitzewellen von 2003 und 2010 zeigen wir, dass EPI und TEPI eine geeignete Beschreibung des großräumig vernetzten Charakters von Hitzewellen und Dürren geben.

Als nächstes untersuchen wir die Wavelet-Transformation, insbesondere das komplexe Dual-Tree-Wavelet, das sich bereits als effizient für Niederschlagsfelder erwiesen hat (siehe Brune et al., 2021). Bei der Untersuchung extremer Niederschläge stellen wir fest, dass die Wavelet-Transformation diese Extreme gut abbildet, ohne dass spezifische Anpassungen der Methode oder der Daten erforderlich sind. Wir vergleichen zwei europäische Reanalysen (COSMO-REA6, CERRA) hinsichtlich ihrer Darstellung von stündlichem Niederschlag und zeigen, dass die Darstellung von kleinskaligen konvektiven Ereignisse in CERRA unszureichend ist. Eine skalen basierte Detektionsstudie für drei Regionen in Nord-, Süd- und Westdeutschland zeigt konsistente Trends zunehmender Intensität für kleinskalige Ereignisse im Sommer und für großskalige Ereignisse im Winter.

Contents

Abstract **III**

1 Introduction **1**
1.1 Motivation . . . 1
1.2 Findings & Outline . . . 3

2 Extremal Dependence **5**
2.1 Motivation & Previous Work . . . 5
2.2 Tail-Pairwise Dependencies . . . 7
2.2.1 Regular variation . . . 7
2.2.2 Extremal Dependence Measure . . . 8
2.2.3 Tail Pairwise Dependence Matrix . . . 10
2.3 Principal Component Analysis . . . 11
2.3.1 ... in Gaussian World . . . 12
2.3.2 ... in Extreme Setting . . . 12
2.4 New Developments . . . 17
2.4.1 Extremal Pattern Index . . . 17
2.4.2 Common extremes in two variables . . . 18
2.4.3 Extremal Pattern Index for two variables . . . 20
2.5 Conclusion & Outlook . . . 21

3 Spectral Filter Approach **23**
3.1 Motivation & Previous Work . . . 23
3.2 Wavelet Theory . . . 26
3.2.1 1 Dimension: Discrete Wavelet Transform . . . 27
3.2.2 2 Dimensions: Discrete Wavelet Transform . . . 32
3.2.3 Further Considerations . . . 35
3.3 Data and preprocessing . . . 38
3.3.1 Wavelet transform & boundary treatment . . . 39
3.3.2 Analysis of hourly precipitation totals . . . 40
3.3.3 Analysis of wavelet representation . . . 42
3.4 Changes in the Localized Wavelet Spectrum . . . 47
3.4.1 Changes in Intensity . . . 47
3.4.2 Changes in Timing . . . 52
3.5 Conclusion & Outlook . . . 57

A Appendix Extremal Dependence **61**
A.1 Paper: Szemkus et al. 2024 . . . 61

A.2 R-Package: "ExtrPatt" 83

B Appendix Spectral Filter **95**
B.1 Clausius-Clapeyron Relation 95
B.2 Additional Figures 96
B.2.1 Density plots from hourly precipitation by season 96
B.2.2 Scaling estimates for CERRA & COSMO-REA6, JJA & DJF 97
B.3 Changes in Hourly Precipitation 102
B.3.1 Changes in Intensity 102
B.3.2 Changes in Timing 104

Bibliography **107**

List of Abbreviations **115**

1. Introduction

1.1. Motivation

In recent years, there has been an undoubted increase in attention to the impact of human-induced climate change on global weather and climate extremes. The latest report from Intergovernmental Panel on Climate Change (IPCC) clearly underlines this fact, revealing that human-induced global warming has led to a 1.1°C warming of global surface temperature between 2011 and 2020 compared to the 1850-1900 pre-industrial levels (Masson-Delmotte et al., 2021). According to a press release from Copernicus services of the European Centre for Medium-Range Weather Forecasts (ECMWF)[1], the year 2023 has been confirmed as the warmest calendar year in global temperature data records going back to 1850 with global temperatures on average 1.48°C warmer than the pre-industrial level. This ongoing warming trend is already having a significant impact on various weather and climate extremes in all regions of the world (Masson-Delmotte et al., 2021).

Europe has recently experienced a series of extreme weather events, including heat waves and devastating floods, which have attracted global attention due to their significant socio-economic impact. According to the current European State of the Climate report (ESOTC 2023[2]), heat-related mortality has increased by around 30% over the last 20 years, and heat-related deaths are estimated to have increased in 94% of the European regions monitored. In July 2021, Germany was hit by devastating heavy rainfall, which led to widespread flooding and flash floods, causing severe damage, particularly in the Ahr valley in western Germany. The German Committee for Disaster Reduction (DKKV) reported at least 189 deaths and an estimated total loss of around 33 billion euros (DKKV, 2022). In light of these catastrophic impacts, the Federal Ministry for the Environment, Nature Conservation, and Nuclear Safety (BMUV) published a press release in 2022 stating that the costs of damage caused by extreme weather in Germany totaled over 80 billion euros in recent years[3].

In view of these climatic challenges, there is an urgent need for comprehensive and differentiated analyses to understand the complex interactions between climate change and extreme weather events. The ClimXtreme project, funded by the Federal Ministry of Education and Research (BMBF), is a collaborative network of various research institutions in Germany that is advancing research on extreme weather events in the context of climate change[4]. This thesis was produced as part of the ClimXtreme project. At the time of submission of this thesis, ClimXtreme is in its second phase. However, this work summarises findings from the first phase, which focused on addressing the following research questions:

1. Has past climate change caused more extreme weather events?

[1] `https://climate.copernicus.eu/copernicus-2023-hottest-year-record` (last accessed: 24.06.2024)
[2] `https://climate.copernicus.eu/esotc/2023` (last accessed: 24.06.2024)
[3] `https://www.bmuv.de/pressemitteilung/` (last accessed: 14.05.2024).
[4] `https://climxtreme.net` (last accessed: 24.06.2024).

2. Will future climate change modify the occurrence of extreme weather events?

This thesis addresses these research questions by analyzing statistical data compression methods and their applicability in the context of climate change. Specifically, two methods with different approaches to data compression are investigated in this regard, as detailed below. A further focus is placed on investigating how these methods can improve our understanding of extreme weather events.

Statistical data compression is a technique for reducing the size of the data by exploiting its statistical properties. Raw data often contains redundant information that can be tightened and irrelevant information (e.g., background noise) that can be reduced. As a result, data compression methods can achieve a more efficient visualization, whereby essential features of the raw data are retained, and only a minimum of information is lost. As a result, statistical compression methods offer great potential for statistical hypothesis testing, as trends that are superimposed by noise in the raw data can be recognised. In the high-dimensional space of the climate system, natural variability and a portion of internal variability overlay climate change signals (Fischer and Knutti, 2016). Data compression can improve the signal-to-noise ratio, increasing the ability to detect weaker signals in highly variable local weather systems.

There are various statistical methods for data compression, ranging in complexity from simple data aggregation (e.g., using mean or maximum values) to advanced deep learning algorithms. Both approaches can be considered forms of data compression, as they extract and encode key features from raw data rather than storing every detail. However, simple data aggregation often results in significant information loss, especially in highly variable datasets or those with outliers. Conversely, deep learning decomposes raw data by utilizing numerous underlying features, which can make the results difficult to interpret.

This thesis focuses on methods that lie between deep learning and data aggregation in terms of complexity, decomposing raw data based on 1) their underlying dependence structure and 2) their underlying frequency components. Specifically, this thesis examines Principal Component Analysis (PCA) as a data-adaptive method and wavelet transform as a spectral decomposition method for their applicability to extreme weather events. Both methods are widely used in climate science and benefit from extensive experience in their application and interpretation.

PCA is probably the most frequently used multivariate statistical method in climate science (see Wilks, 2006b). It dates back to Lorenz, 1956, who used the term empirical orthogonal function (EOF). To this day, both terms are used almost interchangeably (Wilks, 2006b). In traditional PCA, data is decomposed based on a covariance matrix that captures the second-order moments. The strength of this method lies in its ability to identify patterns where these moments are particularly pronounced. However, as the covariance matrix effectively describes the bulk of the underlying distribution, it provides little to no information about extreme features. A promising approach for applying PCA to extremes has recently been developed by Dan Cooley and Emeric Thibaud (see Cooley and Thibaud, 2019). They utilize statistical extreme value theory to define the Tail Pairwise Dependence Matrix (TPDM). The TPDM captures extreme dependence and can be applied analogously to the covariance matrix in PCA. The first part of this thesis examines the applicability of this approach to extreme weather events, specifically its usefulness in addressing the ClimXtreme research questions listed above.

Spectral decomposition, on the other hand, decomposes the data on the basis of its underlying frequency components. These methods differ essentially in the choice of their a priori basis functions, which are used to decompose the raw data. The best known method of spec-

tral decomposition is the Fourier transform, which goes back to Fourier, 1888 and uses sinusoidal basis functions. The wavelet transform, a more recent development in statistical data analysis, introduces basis functions called wavelets that are localised within the signal, thus extending the concept of the Fourier transform to analyse localised variations. The concept of wavelets was developed by (Haar, 1909) and experienced a significant boost towards the end of the 20th century through several contributions that provide efficient algorithms and a framework for the systematic development of new wavelets (see e.g. Mallat, 1989; Meyer, 1997; Daubechies, 1988). Since then, the wavelet transform has become an invaluable tool in various research fields, including signal processing, image compression, and time-frequency analysis. Therefore, wavelets are widely used in the analysis and description of precipitation, including spatial forecast verification (see, e.g. Buschow et al., 2019; Brune et al., 2020) and further analysis (Benestad et al., 2022; Buschow, 2023b). The second part of this thesis analyses the applicability of wavelets to extreme weather events. In particular, it is analysed how wavelets can contribute to answering the ClimXtreme research questions listed above.

1.2. Findings & Outline

This thesis examines the PCA for extremes proposed by Cooley and Thibaud, 2019 and the wavelet transform for their applicability and benefits of an improved understanding of extreme weather events. The most important finding is that both methods are useful in this regard. However, since these methods are based on different approaches to data compression and thus extract different key features, it is not surprising that they can be used to answer different research questions, as detailed below. Therefore, This thesis is divided into two parts, each of which can be read individually. Each part offers a detailed motivation, a comprehensive description of the respective data compression method, and a discussion of the results.

Chapter 2 is dedicated to the PCA for extremes as proposed in Cooley and Thibaud, 2019. This method offers a pattern-based analysis of spatial weather extremes, where common extreme patterns are identified first, and their temporal development is analyzed both individually and jointly. The joint intensity of these identified patterns can be summarized in an Extremal Pattern Index (EPI), which describes the occurrence of extremes within the area under consideration. Meanwhile, examining individual patterns provides insights into the development of extreme events in specific regions. When applied to daily maximum temperatures, the EPI provides a suitable heat wave definition, minimizing dependence on predefined thresholds in grid point space. The framework is then extended to investigate the co-occurrence of extremes in two variables, providing a suitable description of compound drought and heat waves. For this purpose, the cross-tail pairwise dependence matrix (cross-TPDM) is introduced, which serves as an analog to the cross-covariance matrix but for extremes and can be derived from the TPDM. The EPI is also extended to capture common extremes in two variables by introducing the threshold-based EPI (TEPI). This extension is shown to be suitable for describing compound heat and drought extremes. The findings of this work have been published in Szemkus and Friederichs, 2024 and contributed to Xoplaki et al., 2023. Box 1.1 provides a summary of the key findings obtained in this study.

Box 1.1: Key findings chapter 2

- Identification of large-scale patterns related to heat waves and droughts.
- Definition of pattern-based spatial aggregation index: EPI.
- Extention of the approach to common extremes in two variables.
- Extention the EPI to capture simultaneous & compound extremes in two variables.

Chapter 3 explores the wavelet transformation for its application on extreme precipitation events. The dual-tree wavelet is selected due to its proven accuracy in representing spatial precipitation fields (Brune et al., 2021). For the analysis, three areas in Germany are selected: Southern Germany (SG), Northern Germany (NG), and Western Germany (WG), each measuring around 275×275 km. Hourly precipitation data from two reanalysis datasets, COSMO-REA6 and CERRA, is utilized. By applying the wavelet transform, the precipitation fields are decomposed for each time step into spectral energies at different spatial resolutions. Each resolution contains a proportion of the total energy from the input field. It can be shown that the resolutions can be approximated to a spatial scale in km: 5.5 km, 11 km, 22 km, 44 km, 88 km, and 176 km. Initially, the representation of extreme precipitation in both datasets, COSMO-REA6 and CERRA, is analyzed. The wavelet transform allows for scale-aware analyses, revealing good agreement between both datasets, except for the smallest scales (approximately 5.5 km and 11 km), indicating that both datasets resolve convective processes differently. Subsequently, both datasets are analyzed for changes in extreme precipitation with time, including changes in intensity and seasonality. Consistent trends are found towards more small-scale extremes in the summer months and more large-scale extremes in the winter months. In the COSMO-REA6 data, consistent trends are observed in WG towards 1) more intense precipitation extremes across all scales in the summer months and 2) changes in seasonality, with events occurring more concentrated in the northern hemisphere summer months. Although further analyses and the use of additional datasets are needed to support these results, it is concluded that the wavelet transformation is a useful tool for the scale-based analysis of precipitation extremes. Box 1.2 provides a summary of the key findings obtained in this study.

Box 1.1: Key findings chapter 3

- The dual-tree wavelet transform provides a suitable tool for scale-aware analysis of precipitation extremes.
- COSMO-REA6 and CERRA show consistent trends towards small-scale extremes in the summer months and more large-scale extremes in the Winter months.

2. Extremal Dependence

2.1. Motivation & Previous Work

According to the latest IPCC report (Masson-Delmotte et al., 2021), there is strong evidence of an increase in the frequency and severity of heat waves, droughts, and concurrent events. However, these changes are primarily attributed to thermodynamic processes. While significant changes in precipitation deficits are only found in isolated regions, strong signals are found in agricultural and ecological droughts that are highly related to changes in evapotranspiration (see, e.g. Masson-Delmotte et al., 2021; Knutti and Sedláček, 2013; Cook et al., 2018; Dai et al., 2018). Temperature and precipitation, on the other hand, are strongly negatively correlated during summer. Thus, warmer temperatures naturally lead to an increase in the co-occurrence of hot temperatures and meteorological droughts, even if precipitation does not change (Zscheischler and Seneviratne, 2017).

When analyzing heat waves and droughts, it is common practice to use grid point-based definitions and to evaluate the identified events based on various criteria such as magnitude, spatial extent, and duration (see, e.g. Russo et al., 2015; Lavaysse et al., 2018; Griggs and Noguer, 2002). A heatwave is usually defined as a consecutive series of days in which temperatures exceed a predefined threshold (Robinson, 2001). For the investigation of meteorological droughts, the Standartized Precipitation Index (SPI) (McKee et al., 1993) is commonly applied to monthly, seasonal, or even annual precipitation totals (see e.g. Sienz et al., 2012; Keyantash and Dracup, 2002). This index indicates the position of the precipitation values transformed into the normal distribution on their density curve. In the context of the SPI, the term "drought" is usually applied when the precipitation value falls below a threshold of minus one standard deviation. However, as both types of events have a natural, certain spatial extent, it seems reasonable to include this in the respective definitions. Consequently, some heatwave definitions use an additional threshold value for their spatial extent (see, e.g. Perkins and Alexander, 2013).

When summarizing heat waves and droughts within a specific region, conventional methods typically assess their magnitude, spatial extent, or duration based on predefined events as per established definitions. E.g., the German weather service offers a comprehensive comparison of European heat waves from 1950 to 2023[1]. However, relying solely on threshold-based definitions may result in significant information loss, prompting the need for a more inclusive approach.

Climate indices, such as, e.g., those for the North Atlantic Oscillation (NAO) and El Niño (Trenberth and Stepaniak, 2001; Hurrell et al., 2003), provide a broader perspective on large-scale climate patterns. These indices have become invaluable tools in climate monitoring and analysis, offering simplified insights into complex climate conditions. By reducing complex climate conditions to more manageable information, they provide a deeper understanding of regional climate dynamics. Some climate indices, e.g., for NAO index (Barnston and Livezey,

[1] https://www.dwd.de/EN/ourservices/rcccm/int/rcccm_int_hwkltr.html (last accessed: 26.06.2024)

1987), rely on PCA to effectively summarize and characterize large-scale climate phenomena. However, classical PCA primarily focuses on describing the bulk of the underlying distribution, as represented by first and second-order moments, and may not capture information about the tail of the distribution.

To maintain a focus on weather extremes, alternatives to traditional PCA, such as the PCA framework proposed by Cooley and Thibaud, 2019, explicitly designed for investigating extreme events, should be utilized. This approach ensures that the analysis remains sensitive to extreme weather occurrences, providing a more comprehensive understanding of their impacts on regional climate dynamics.

Cooley and Thibaud, 2019 propose the TPDM that gives an effective dependence measure for extremes. The TPDM is specifically designed to have all the necessary properties to perform a PCA and thus can be used as an analog to the covariance matrix for extremes in PCA studies. Jiang et al., 2020 and Rohrbeck and Cooley, 2021 propose a PCA-study using the TPDM with applications to daily precipitation over the United States (Jiang et al., 2020) as well as historical river flow levels across northern England and southern Scotland (Rohrbeck and Cooley, 2021).

Notably, there are further approaches for the mathematically consistent development of PCA for extremes (see Schlather and Reinbott, 2021; Schlather, 2024). Following Schlather and Reinbott, 2021, extreme distributions (here max-stable distributions) cannot be fully described by bivariate characteristics such as covariance. Therefore, the decomposition of such derived matrices would be insufficient. However, these approaches are not yet applicable in practice. Another interesting perspective arises from the logistic PCA (Landgraf and Lee, 2020), which can be applied to binary data. This method was recently applied to heat waves by Buschow, 2023a.

To retain the focus on the consideration of extreme events, the study explores how the PCA framework from Cooley and Thibaud, 2019 can be utilized to obtain a compressed description of extreme weather events.

The results of this chapter have been previously published in the following peer-reviewed paper:

- Svenja Szemkus and Petra Friederichs (Jan. 2024). “Spatial patterns and indices for heat waves and droughts over Europe using a decomposition of extremal dependency.” en. In: *Advances in Statistical Climatology, Meteorology and Oceanography* 10.1, pp. 29–49. DOI: 10.5194/ascmo-10-29-2024

The remainder of this chapter is therefore as follows: Section 2.2 introduces the theoretical framework of regular variation underlying the definition of TPDM. Subsequently, the definition of TPDM is derived, and a discussion on the estimators as outlined in Cooley and Thibaud, 2019; Jiang et al., 2020 is provided. Section 2.3 introduces the PCA framework for TPDM, along with a discussion on its similarities and differences compared to traditional PCA. Section 2.4 summarizes the results obtained in Szemkus and Friederichs, 2024. Finally, Section 2.5 presents conclusions and outlines future directions.

2.2. Tail-Pairwise Dependencies

2.2.1. Regular variation

This section introduces the theory of regular variations, first in the univariate and second in the bivariate setting. For more detailed information and proofs, the reader is referred to Resnick, 2007, from which the main information in this section is taken.

Regularly varying variables are defined by the tail behavior of their distribution function $F(x)$ as x tends to infinity. Thus, the definition does not characterize the full distribution, which provides a more general approach than common statistical models for extremes.

Box 2.1: Definition Regular Variation

Let X be a non-negative random variable with positive and measurable distribution function $F(x)$ on $\mathbb{R}_+$. We call F a regularly varying function with *exponent of variation* $\rho \in \mathbb{R}$, denoted $F \in \mathrm{RV}_\rho$, if for all $t > 0$:

$$\lim_{x\to\infty} \frac{F(xt)}{F(x)} = t^\rho. \tag{2.1}$$

In applications, one is particularly interested in functions whose tail is regularly varying. This can be expressed through the tail function $\bar{F}(x) = 1 - F(x)$. We say that X is regularly varying with index $\alpha > 0$ if its tail function is regularly varying with exponent of variation $\rho = -\alpha$ (Resnick, 2007, Theorem 3.6).

The Generalized Extreme Value Distribution (GEV) gives a good example of a regular varying function. From the well-known formulation of the GEV (see e.g. Coles, 2001), given a random variable x: $G(x) = \exp(-x^{-\alpha})$, it can be shown for $x \to \infty$ that

$$1 - G(x) \sim x^{-\alpha}. \tag{2.2}$$

From this, it follows that that a non-negative random vector following a GEV, is regularly varying with tail-index α in the limit $x \to \infty$.

An interpretation from vague convergence theory is commonly applied to analyze the asymptotic properties of regularly varying functions. For the definition, see, e.g., Billingsley, 1968. The concept of vague convergence states that for large x, the tail function $\bar{F}(x)$ converges to a limiting function. In the context of regularly varying variables, this limiting function often takes a specific form, reflecting the asymptotic behavior of heavy-tailed distributions. Formally, a random variable X is said to be regularly varying with an index $\alpha > 0$, if the following vague convergence holds (Resnick, 2007, Theorem 3.6):

$$n\mathbb{P}\left[\frac{X}{b_n} > x\right] \xrightarrow{v} \nu(x) \tag{2.3}$$

as $n \to \infty$, where b_n is a sequence of positive numbers such that $b_n \to \infty$, and $\nu(x)$ is a non-degenerate probability function with the following asymptotic behavior:

$$\nu(x) \sim x^{-\alpha} \quad \text{as} \quad x \to \infty. \tag{2.4}$$

In other words, as the threshold x increases, the tail probabilities of X behave like a power law with exponent $-\alpha$, and this behavior is reflected in the limiting distribution $\nu(x)$.

We now turn to regular variation on a 2-dimensional non-negative space $\mathbb{R}^2 = [0,\infty)\backslash\{0\} \times [0,\infty)\backslash\{0\}$. An extension to negative values is possible, although not considered in this work (see Resnick, 2007). Let be given a 2-dimensional random vector $X \in \mathbb{R}^2$. The multivariate analogue to equation 2.3 is:

$$n\mathbb{P}\left[\frac{X}{b_n} \in \cdot\right] \xrightarrow{v} \nu(\cdot) \tag{2.5}$$

For the multivariate case, it is convenient to transform the state space using a generalized polar coordinate transformation. The unit-sphere N_+ defines on any norm $\|\cdot\|$ as:

$$N_+ = \left\{x \in \mathbb{R}^2 : \|x\| = 1\right\}. \tag{2.6}$$

With polar coordinate transform $T : [0,\infty)\backslash\{0\} \times [0,\infty)\backslash\{0\} \to [0,\infty)\backslash\{0\} \times N_+$ defined as:

$$T(x) = \left(\|x\|, \frac{x}{\|x\|}\right) =: (r, \boldsymbol{\omega}) \qquad \text{for } \|x\| > 0, \tag{2.7}$$

where $r = \|x\| \in [0,\infty)\backslash\{0\}$ is the radial component and $\boldsymbol{\omega} = x/\|x\| \in N_+$ is the angular component. To rewrite equation 2.5 in the form of polar coordinates, an estimate $\hat{T}(\nu)$ is needed. Following Resnick, 2007, Proposition 5.5, consider an independent pair of random variables (R,Ω) on $[0,\infty)\backslash\{0\} \times N_+$. It is further assumed that R is regular varying with index $\alpha > 0$: $R \in \mathrm{RV}_{-\alpha}$ and that $\Omega \in N_+$ follows a distribution function S, that is often referred to as spectral- or angular measure. This gives the following formulation:

$$n\mathbb{P}[\frac{R}{b_n} > r] \xrightarrow{v} r^{-\alpha}, \quad r > 1 \qquad \mathbb{P}[\Omega \in \boldsymbol{\omega}] \xrightarrow{v} S(\boldsymbol{\omega}) \tag{2.8}$$

as $r \to \infty$. The 2-dimensional random vector $T(X) = (R,\Omega)$ is then regularly varying (Resnick, 2007, page 198), which is formulated as

$$n\mathbb{P}\left[\left(\frac{R}{b_n}, \Omega\right)\right] = n\mathbb{P}\left[\frac{R}{b_n} > r\right] n\mathbb{P}\left[\Omega \in \boldsymbol{\omega}\right] \xrightarrow{v} \nu(\cdot) \times S(\boldsymbol{\omega}). \tag{2.9}$$

Hence, the tail behavior of the radial component R follows the power-law distribution as described in equation 2.4. Conversely, the distribution of angular components converges towards the distribution S for sufficiently large R and characterizes the behavior of angular components. As pointed out in the following, S holds information about extremal dependencies (Larsson and Resnick, 2012) and can be used to derive a measurement for this.

2.2.2. Extremal Dependence Measure

Consider a bivariate random vector $X = (X_1, X_2)$, that is regularly varying with index α: $X \in \mathrm{RV}_{-\alpha}$. Further, choose any norm $\|\cdot\|$ and calculate the polar transform (eq. 2.7) $T(X) = (R,\Omega)$ to obtain the angular component $\Omega = (\Omega_1, \Omega_2)$ and radial component R. The Extremal Dependence Measure (EDM) (Resnick, 2002) then defines on the spectral measure S (equation 2.9) as:

$$\mathrm{EDM}(X_1, X_2) = \int_{N_+} \Omega_1 \Omega_2 S(d\boldsymbol{\omega}). \tag{2.10}$$

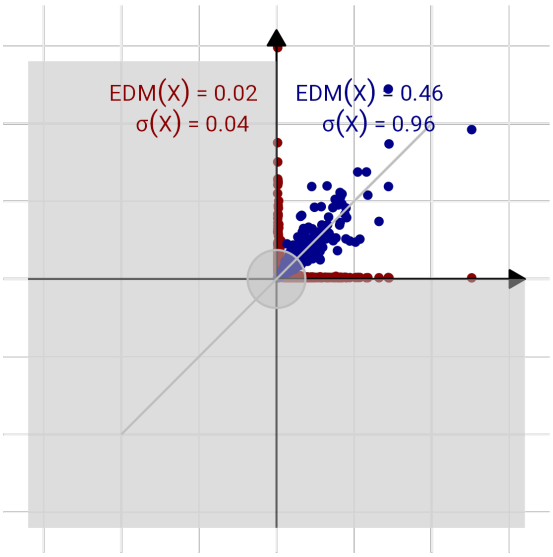


Figure 2.1.: Scatterplot of daily temperature measurements at three observational stations $T = (T_C, T_D, -T_B)$ in Bremen (T_B), Cologne (T_C) and Düsseldorf (T_D). The data has been pre-processed to be regularly varying with index $\alpha = 2$. The gray circle indicates r_{th}, which is the 98% quantile of radial components r. EDM is estimated according to equation 2.11 and σ is estimated from equation 2.13.

If the measure S concentrates on the respective coordinate axis in $\mathbb{R}^2$, then $\text{EDM}(X_1, X_2) = 0$. If S concentrates on the line where $\Omega_1 = \Omega_2$ then $\text{EDM}(X_1, X_2)$ is maximal. Larsson and Resnick, 2012 propose an estimator of $\text{EDM}(X_1, X_2)$ for independent copies $x_t = (x_{t,1}, x_{t,2})$ with $t, \ldots, n$. Set $r_t = ||x_t||$ and choose a high threshold r_{th}. The index function $\mathbb{I}$ is introduced to ensure that only observations larger than r_{th} are considered for the estimation. $n_{exc} = \sum_{t=1}^{n} \mathbb{I}_{[r_t > r_{th}]}$ is the number of values r_t exceeding the threshold r_{th}, resulting in the estimate:

$$\widehat{\text{EDM}}(X_1, X_2) = \frac{1}{n_{exc}} \sum_{t=1}^{n} \frac{x_{t,1} x_{t,2}}{r_t^2} \mathbb{I}_{[r_t > r_{th}]}. \tag{2.11}$$

Figure 2.1 illustrates the concept of EDM using the example of two 2-dimensional variables. Both variables are regularly varying with index $\alpha = 2$ and have been chosen as an illustrative example of 1) near independence, i.e., variables arranged close to the coordinate axes (red), and 2) near-perfect dependence, i.e., variables distributed around the center line between the two axes (blue). The threshold r_{th} is set to the 0.98-quantile of radial components, visually represented by a gray circle. When data points cluster near the respective coord. axis, the EDM approximates zero, indicating minimal statistical dependence in extreme values. If EDM$= 0$, the variables are asymptotic independent in mathematical terminology (Larsson and Resnick, 2012). Conversely, EDM increases when the data points move toward the center line, indicating extremal dependence. The maximum EDM value is reached when the data coincides with the grey line, which means perfect dependence.

Thus, the EDM is a suitable measurement for extreme dependencies in a bivariate setting. It may be interpreted as analogous to variance in the non-extreme setting (Larsson and Resnick, 2012). A few years ago, Cooley and Thibaud, 2019 reformulated the EDM in p-dimensional

space. They further propose a standardization of the EDM between zero and one, resulting in a matrix (the TPDM), that can be interpreted as an analog to the correlation matrix for extremes (Cooley and Thibaud, 2019).

2.2.3. Tail Pairwise Dependence Matrix

Cooley and Thibaud, 2019 reformulate the EDM as proposed in Larsson and Resnick, 2012. They restrict to the case where:

1. $\mathbb{R}^p = [0,\infty)\backslash\{0\}^p$ and thus extend the EDM from bivariate case to p dimensions.
2. $X \in \mathrm{RV}_{-2}$, the regular varying case with index $\alpha = 2$.
3. Usage of the L_2-norm $||\cdot||_2$ for transformation into polar coordinates.

Let $x_t = (x_{t,1},\dots,x_{t,p})$, with $t = 1,\dots,n$ be independent copies of X. Cooley and Thibaud, 2019 define the $p \times p$ dimensional TPDM Σ_X as:

$$\sigma_{X_{i,j}} = \int_{N_+} \Omega_1\Omega_2 S(d\boldsymbol{\omega}) \qquad \Sigma_X = (\sigma_{X_{i,j}})_{i,j=1,\dots,p}. \tag{2.12}$$

Where $\sigma_{X_{i,j}}$ equals $\mathrm{EDM}(X_i, X_j)$, the i'th, j'th element of Σ_X (Cooley and Thibaud, 2019). Due to te restrictions above, Σ_X is positive-semidefinite, i.e. it can be shown for an $Y \in \mathbb{R}^p$ that $Y^T\Sigma_X Y \geq 0$. As in equation 2.10, we fix a high threshold r_{th} and write $n_{exc} = \sum_{t=1}^n \mathbb{I}_{[r_t > r_{th}]}$. We restrict to the L_2-norm when estimating radial component $r_t = ||x_t||_2$ and further introduce a normalization constant $m = r_{th}^2 n^{-1} \sum_{t=1}^n \mathbb{I}(r_t > r_{th})$. If X is preprocessed to have common unit scale, we set $m = p$ (Cooley and Thibaud, 2019). The estimator of each element of the TPDM $\hat{\sigma}_{X_{i,j}}$ proposed by Cooley and Thibaud, 2019, is given as:

$$\hat{\sigma}_{X_{i,j}} = \frac{m}{n_{exc}} \sum_{t=1}^{n} \frac{x_{t,i}x_{t,j}}{r_t^2} \mathbb{I}_{[r_t > r_{th}]}. \tag{2.13}$$

The radial component r_t is estimated from all data points within the study area at time t. The estimator is guaranteed to be positive definite. For applications involving large areas and/or small-scale events, Jiang et al., 2020 propose an alternative estimator where the radial component is estimated on pairwise estimators:

$$\hat{\sigma}_{X_{i,j}} = \frac{2}{n_{exc}} \sum_{t=1}^{n} \frac{x_{t,i}x_{t,j}}{r_{t,ij}^2} \mathbb{I}_{[r_{t,ij} > r_{th,ij}]}. \tag{2.14}$$

Here $r_{t,ij} = \sqrt{x_{t,i}^2 + x_{t,j}^2}$. The threshold $r_{th,ij}$ can be set individually for each pairwise element of TPDM. If X is preprocessed to have common unit scale, we now set $m = 2$ (Jiang et al., 2020). This alternative estimator introduces increased computational complexity compared to the original estimator, as it requires the individual calculation of the radial component and threshold value for each pair. Unlike the original estimator (see equation 2.14), it is not guaranteed to be positive definite.

2.3. Principal Component Analysis

Let $x_t, t = 1,\dots,n$ be independent copies of a p-dimensional random vector $X = (X_1,\dots,X_p)^\top$ Simply put, Principal Component Analysis (PCA) reduces the matrix $X = (x_1,\dots,x_n)$ into $V \in \mathbb{R}^{k\times p}$, containing fewer, i.e. $k < n$ new rows called the eigenvectors (Wilks, 2006b). The eigenvectors are linear combinations of the original columns and form an orthogonal basis on $\mathbb{R}^{n\times p}$. Let $V = (v_1,\dots,v_k)$ be the matrix whose columns are the k eigenvectors. The projection of X onto the subspace spanned by V is XV. The reconstruction of X from this projection is given by:

$$\hat{X} = (XV)V^\top. \tag{2.15}$$

To ensure V forms an orthogonal basis, one further assumes $VV^\top = I_k$, where I_k is the $k \times k$ identity matrix.

Formally, the PCA can be derived by minimizing the reconstruction error using the L2-norm (see Pearson, 1901) with:

$$\min_V ||X - \hat{X}||_2^2 \tag{2.16}$$

$$= \min_V ||X - XVV^\top||_2^2. \tag{2.17}$$

It can be shown (see e.g. Abdi and Williams, 2010), that this problem reduces to finding V that maximizes $V^\top SV$, where $S = X^\top X$.

When X is normally distributed, i.e., $X \sim N(E[X], S_X)$ with zero mean, $S = S_X$, the covariance matrix. Thus, classical PCA is typically employed in the Gaussian setting, which gives several benefits for interpreting the results, where each eigenvector explains a specific amount of total variance within the data.

However, in principle, every matrix, that is *symetric*, *non-negative* and *positive-semidefinite* (for definition, see box 2.2) can be applied for PCA.

Box 2.2: Definition positive-semidefinite matrix

Let M be a symmetric matrix in $\mathbb{R}^{n\times n}$, and let z be a non-zero column vector in $\mathbb{R}^{n\times 1}$. M is positive-semidefinite if the quadratic form $z^\top Mz$ is non-negative for all $z \neq 0$. In other words, M is positive semidefinite if and only if $z^\top Mz \geq 0$ for all non-zero vectors z.

Outside the Gaussian setting, however, it is essential to investigate which properties are maximized by the eigenvectors in order to interpret them correctly. Cooley and Thibaud, 2019 take this into account. They define a matrix W whose rows are the random vectors $\boldsymbol{\omega}_t = x_t/r_t$, for which $r_t > r_{thr}$, and show that the estimate of TPDM (equation 2.13) can be rewritten in the form $\hat{\Sigma}_X = m/n_{exc}\hat{W}^\top\hat{W}$. This allows a straightforward interpretation of the eigenvectors of TPDM as being close to those of the covariance matrix, representing the extremal dependence structure. However, a significant difference is that the matrix W does not have a mean value of zero, resulting in the first eigenvector representing the mean value.

Notably, there is ongoing research in mathematical sciences to generalize the classical PCA (equation 2.16) to minimize properties of distribution functions that do not have second moments, e.g., extreme value distributions. Some first approaches have been published to date (Schlather and Reinbott, 2021). However, these methods are not yet applicable in practice and need further development.

A naturally arising question is how the eigenvectors of the TPDM actually differ from those of the covariance matrix. The following, therefore, uses the illustrative example of daily maximum temperature in the northern hemispheric Summer (JJA) months from the ERA5 reanalysis for the period from 2000 to 2023 to provide an insight into the decomposition of the TPDM. First, the classical PCA is illustrated. Secondly, the decomposition of TPDM is discussed, and the necessary assumptions and additional steps compared to classical PCA are introduced.

The calculations were performed within the ClimXtreme central evaluation system (XCES) at the German Climate Computing Center (DKRZ[2]). Sebastian Buschow & Etor Lucio-Eceiza implemented the code that we apply for PCA in the Gaussian setting. Svenja Szemkus and Etor Lucio-Eceiza implemented the code that calculates the PCA using the TPDM.

2.3.1. ... in Gaussian World

Assume X to be normally distributed, with zero mean and unit standard deviation. The latter is not necessary for the classical PCA, but results in the covariance matrix matching the correlation matrix, which is standardised between zero and one, just like the TPDM.

The estimate of the covariance matrix $\hat{S}_X = \frac{1}{n}X^\top X$ decomposes into its eigenvectors V and eigenvalues Λ, such that $S_X = V\Lambda V^\top$. The principal components η are then derived by projecting the data onto the eigenvectors: $\eta = V^\top X$. In the context of standardized, normally distributed X, it can be demonstrated that each principal component η_k follows a normal distribution $\eta_k \sim N(0, \lambda_k)$ with zero mean and standardisation according to its respective eigenvalue λ_k. The eigenvectors are normalized to unit length $||V_k|| = 1$, and the sum of eigenvalues equals the dimensionality of the data $\sum_{i=1}^{p} \lambda_i = p$.

Figures 2.2 and 2.3 show the eigenvalues and principal components, as computed from daily maximum temperature data as described above. Daily temperature data can generally be assumed to follow a normal distribution. The seasonal cycle is removed by subtracting the monthly long-term mean from the data at each grid point. Subsequently, the data is standardized. This results in the dataset being approximately normally distributed with mean zero and variance one. The first eigenvector mostly represents the long-term temperature trend in the northern hemisphere (Fig. 2.3). For the higher eigenvectors, typical patterns for large-scale climate variations such as El Niño and Tropical Atlantic variability can be found. The principal components are normally distributed with zero mean. Some of them (see, e.g., $k = 2$: Nino3 and $k = 3$: Northern Tropical Atlantic index) closely follow the time series of climate indices associated with each pattern (see, NOAA Climate Prediction Centre, available at `https://psl.noaa.gov/data/climateindices/`; last accessed: 01.07.2024).

2.3.2. ... in Extreme Setting

Now, assume X to be regular, varying with index $\alpha = 2$, and estimate the TPDM $\hat{\Sigma}_X$ according to equation 2.14, the estimator given by Jiang et al., 2020 for larger areas. The TPDM satisfies all the necessary properties to perform a PCA. However, as previously noted, the positive definiteness of this specific estimator is not assured. In line with the approach adopted by the authors, we address this issue by determining the nearest positive-semidefinite matrix utilizing the methodology proposed by Higham, 2002. Similar to the decomposition in the classical setting, the decomposition $\hat{\Sigma}_X = V\Lambda V^\top$ provides the eigenvector matrix V with eigenvectors V_i,

[2] `https://www.xces.dkrz.de/` (last ac- cessed: 01.07.2024)

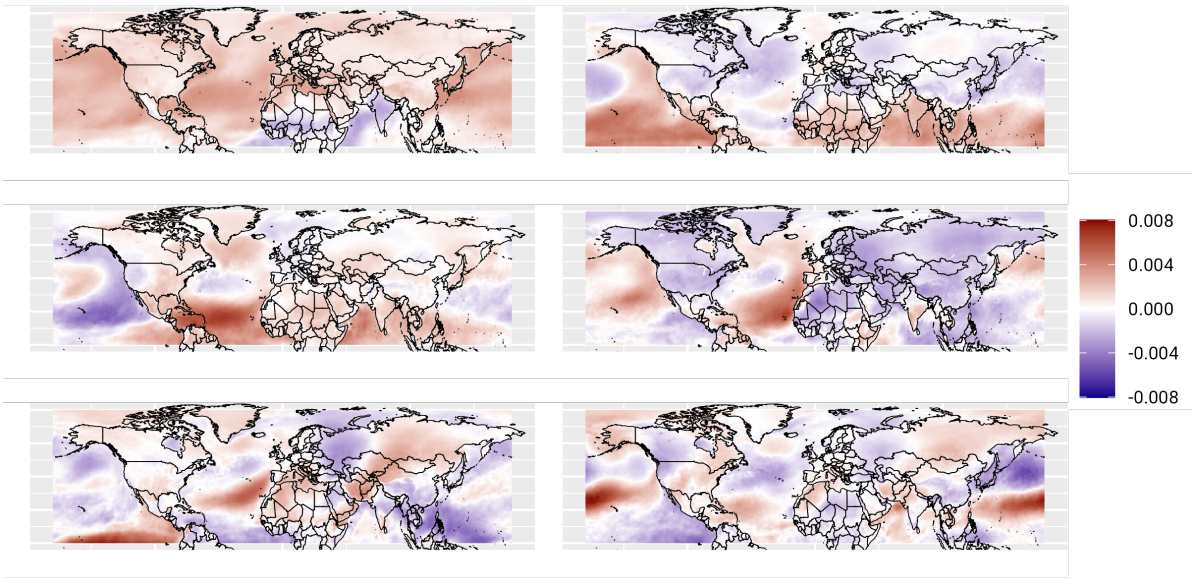


Figure 2.2.: First six (from left to right, top to bottom) eigenvectors of daily maximum temperature data, obtained in JJA months in the northern hemisphere. The data has been standardized, and the seasonal circle has been removed by subtracting the seasonal climatological cycle.

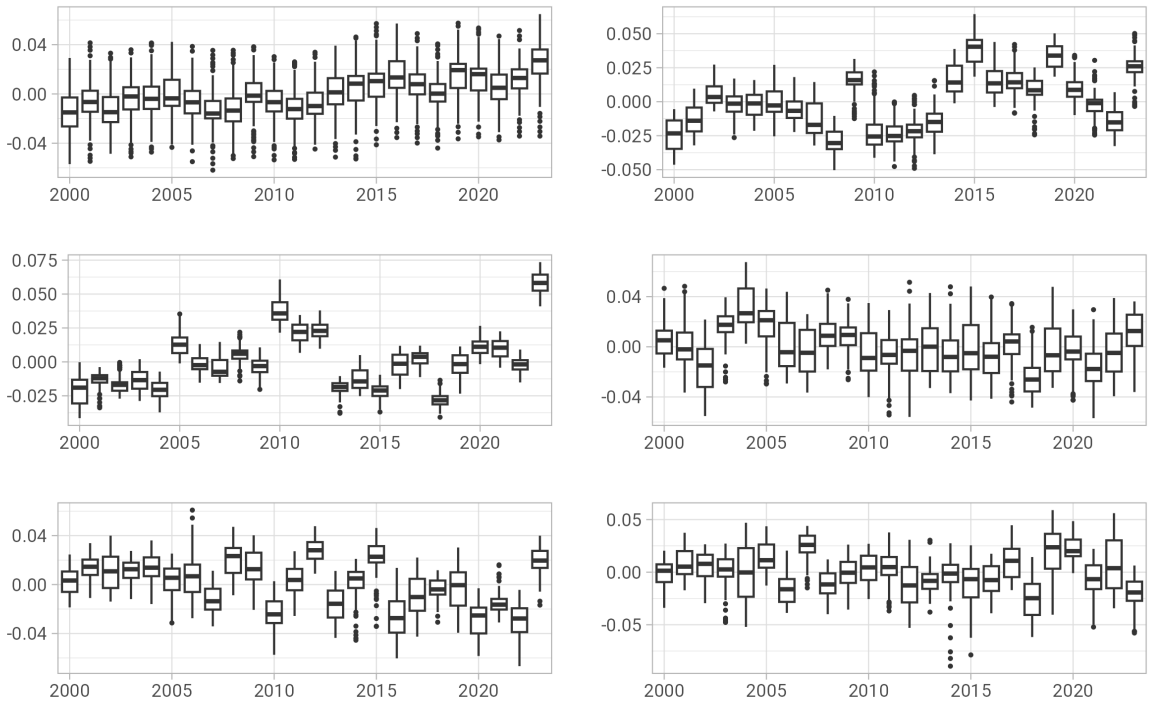


Figure 2.3.: Yearly boxplot-representation of Principal Components according to eigenvectors from Fig. 2.2

$i = 1, \dots, k$ as well as the diagonal-matrix Λ, containing the non-negative eigenvalues $\lambda_1, \dots, \lambda_k \geq 0$.

However, some adjustments need to be made for the results to be mathematically consistent. Since the TPDM is defined based on positive real numbers (see, e.g., Fig. 2.1), the eigenvectors of the TPDM should, by definition, form an orthogonal basis in $\mathbb{R}_+^p$, the positive real numbers. However, the eigenvectors of a PCA live by construction in $\mathbb{R}^p = (-\infty, \infty)^p$. Cooley and Thibaud, 2019 overcome this issue by defining a new vector space $\mathbb{X}^p = [0, \infty)^p$, along with a transformation function $\tau : \mathbb{R}^p \to \mathbb{X}^p$ and it's inverse τ^{-1}. They define mathematical operations on $\mathbb{X}^p$, via transformation, calculation, and back-transformation as summarised in Tab. 2.1, along with the transformation function $\tau(y)$ and it's inverse:

$$\tau(y) = \log(1 + \exp(y)) \tag{2.18}$$

$$\tau^{-1}(y) = \log(\exp(y) - 1). \tag{2.19}$$

It can be shown that $\mathbb{X}$ satisfies all mathematical requirements for a vector space (see Cooley and Thibaud, 2019). τ is known as the softplus function in neuronal networks (Cooley and Thibaud, 2019). It is continuous, infinitely differentiable, and, importantly, has a negligible effect on large values. Thus, the Fréchet distribution, as well as the orthogonality, is preserved under this transformation.

Table 2.1.: Definition of mathematical opperations in $\mathbb{X}^p$ as provided in Cooley and Thibaud, 2019.

vector addition	$x_1 \oplus x_2 = \tau\big(\tau^{-1}(x_1) + \tau^{-1}(x_2)\big)$	For $x_1, x_2 \in \mathbb{X}^p$
scalar multiplication	$a \circ x_i = \tau\big(a\tau^{-1}(x_i)\big)$	For $a \in \mathbb{R}$, $x_i \in \mathbb{X}^p$
additive identity	$0_{\mathbb{X}^p} = \tau(0)$	
additive inverse	$-x_i = \tau\big(-\tau^{-1}(x_i)\big)$	For $x_i \in \mathbb{X}^p$
matrix multiplication	$A \circ x_i = \tau\big(A\tau^{-1}(x_i)\big)$	For $A \in \mathbb{R}^{p \times q}$, $x \in \mathbb{X}^q$

To ensure that the eigenvectors $V_i \in \mathbb{R}^k$ form an orthogonal basis in $\mathbb{R}^+$, Cooley and Thibaud, 2019 simply transform them componentwise in $\mathbb{X}$ such that $E_i = \tau(V_i)$, for each of the $i = 1, \dots, k$ components. Since X is assumed to follow the GEV, $X \in \mathbb{X}^p$ such that, in order to calculate the principal components, we back-transform $X \to \mathbb{R}^p$ to apply the matrix multiplication:

$$\eta = V^\top \tau^{-1}(X). \tag{2.20}$$

$\eta \in \mathbb{R}$ are not transformed to $\mathbb{X}$. This is unnecessary as they, unlike the eigenvectors, per definition, do not have to be in $\mathbb{R}^+$.

Figures 2.4 and 2.5 illustrate the first six eigenvectors and principal components of the TPDM, derived from daily maximum temperature data as detailed above. The eigenvectors have been transformed to positive real space $\mathbb{R}^+$ using equation 2.19. Hence values $> \tau(0)$ are considered as positive and values $< \tau(0)$ as negative. The data undergoes standardization at each grid point, utilizing the climatological mean and standard deviation to eliminate the seasonal variation. Although daily temperature data generally do not follow the GEV, a common practice involves applying marginal transformations to fit the data (Cooley and Thibaud, 2019). Thus, following Szemkus, 2023, the empirical distribution $\hat{F}_i(x)$ is estimated at each grid point $i = 1, \dots, p$, and subsequently, the marginal transformation $g_i(x) = (-\log(\hat{F}_i(x)))^{-1/2}$ is applied for data preprocessing. The first eigenvector is positive with minimal spatial variabil-

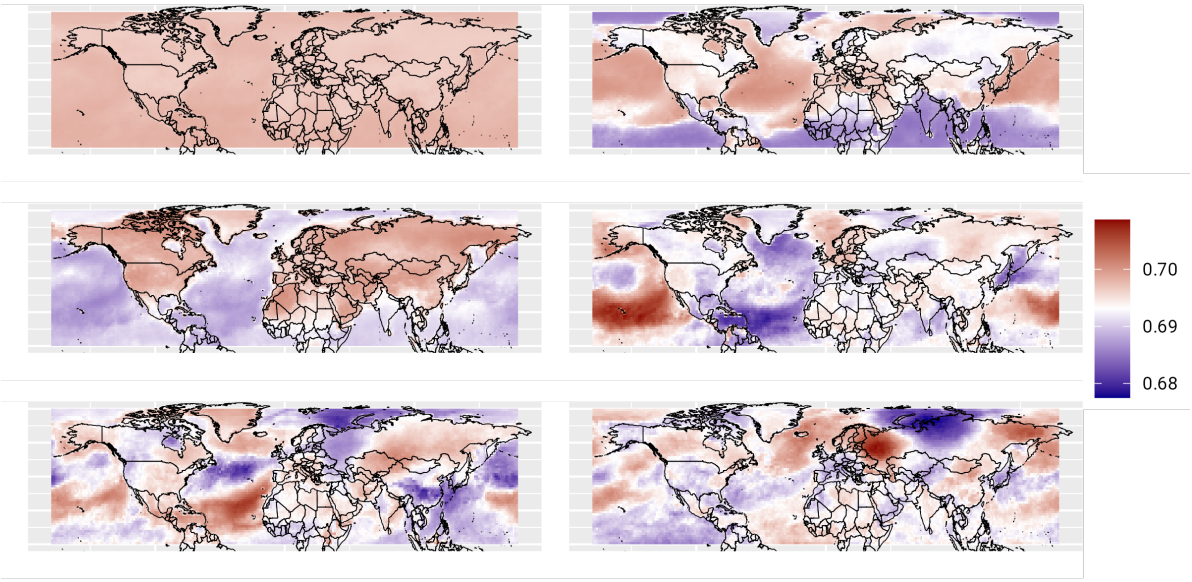


Figure 2.4.: First six (from left to right, top to bottom) eigenvectors of daily maximum temperature data, obtained in JJA months in the northern hemisphere. Each vector is transformed to the positive reals $\mathbb{R}^p \to \mathbb{X}$ such that $\tau(0) = \log(2) \approx 0.693$ corresponds to zero.

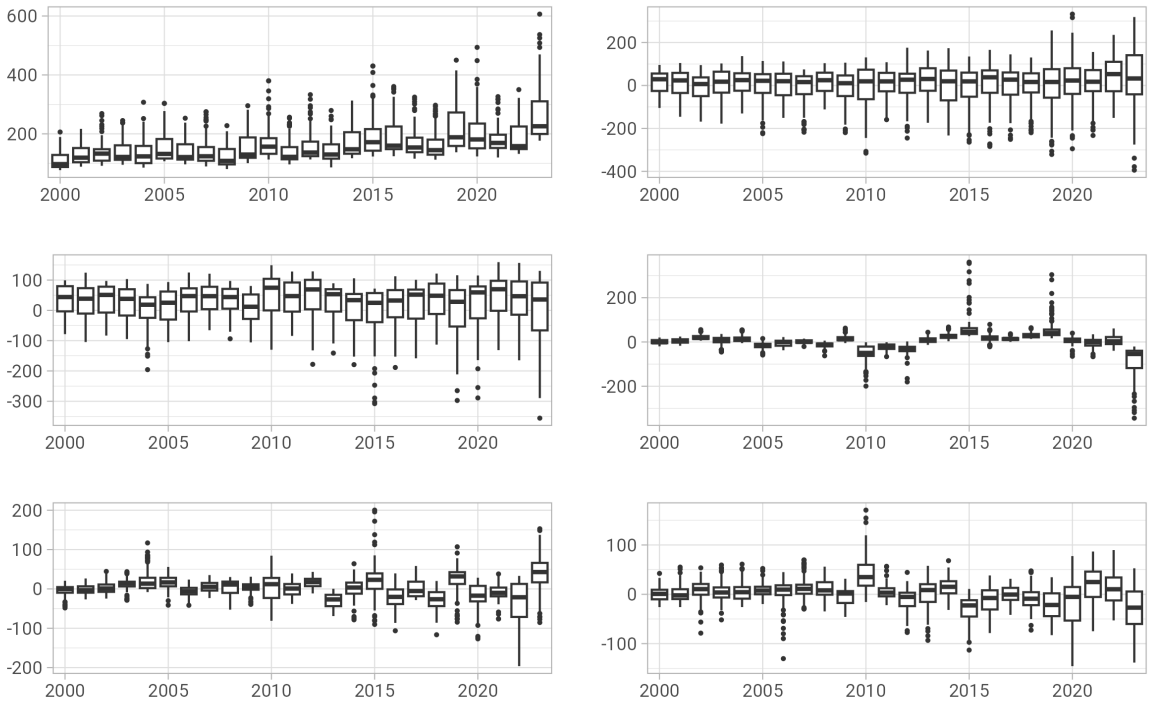


Figure 2.5.: Principal components of TPDM, according to Fig. 2.4.

ity, a characteristic previously noted and discussed in Szemkus and Friederichs, 2024; Jiang et al., 2020; Cooley and Thibaud, 2019. The second and third eigenvectors capture seasonal variability and land-sea variability. Despite its removal during standardization, the unexpected persistence of seasonality suggests that the process may not entirely eliminate all seasonal variations, particularly evident in the tail behavior of temperature data. The higher-order eigenvectors describe more localized patterns for temperature extremes. For instance, we identify the tripolar North Atlantic pattern, previously associated with heat waves (Beobide-Arsuaga et al., 2023) in the fifth eigenvector. We also identify the double-jet structure over Eurasia in the sixth eigenvector that has been recognized as a driving factor behind the increase in European heat waves (Rousi et al., 2022).

The analysis shows that employing the TPDM instead of the correlation matrix produces distinct spatial patterns. While the patterns derived from the correlation matrix predominantly reflect large-scale climate variations, those derived from the TPDM correspond to known patterns associated with heat waves. They provide information about periods in which the corresponding pattern is very pronounced and thus provide a compact representation of heat wave activity in a particular region.

2.4. New Developments

So far, this chapter has been dedicated to summarising previous studies' theoretical foundations and findings. This section now summarises the results obtained in this thesis. The main findings have already been published in the following peer-reviewed paper:

- Svenja Szemkus and Petra Friederichs (Jan. 2024). "Spatial patterns and indices for heat waves and droughts over Europe using a decomposition of extremal dependency." en. In: *Advances in Statistical Climatology, Meteorology and Oceanography* 10.1, pp. 29–49. DOI: 10.5194/ascmo-10-29-2024

My supervisor, Petra Friederichs, and I jointly developed the idea for this study. I took on the tasks of data preparation, analysis, and visualization, as well as the initial drafting of the manuscript. Petra Friederichs contributed to the improvement and revision of the manuscript and the discussion and interpretation of the results.

The complete software needed for the reproduction of Szemkus and Friederichs, 2024 has been published as CRAN-Package for the statistic software R:

- Svenja Szemkus (2023). *ExtrPatt: Spatial Dependencies and Indices for Extremes*. URL: https://CRAN.R-project.org/package=ExtrPatt (visited on 12/20/2023)

The full version of our publication is attached in Appendix A.1. This section, therefore, summarizes the achievements and developments of the most important results.

2.4.1. Extremal Pattern Index

In Szemkus, 2023, we introduce the EPI defined by the leading principal components of TPDM. As shown in section 2.3, the leading patterns of TPDM capture the dynamics of large-scale weather extremes, and the principal components provide information about their temporal occurrence. Thus, a compressed representation of the leading principal components captures strong features in one or more components and can summarise the occurrence of heat waves over the entire area. We introduce the EPI to provide a spatial aggregation that is high when individual patterns or a linear combination of patterns are particularly strongly pronounced. This allows us to identify spatially extended extreme events with minimal reliance on predefined thresholds and regions of influence in the grid point space. We can show that this results in little information being lost with respect to heat wave description.

Let be given m leading principal components η_t at time t as derived by equation 2.20 with corresponding eigenvalues $\lambda_{1,\dots,m}$. The EPI is defined as:

$$\mathrm{EPI}_t^{\eta} = \frac{\sqrt{\sum_{k=1}^{m} \eta_{t,k}^2}}{\sqrt{\sum_{j=1}^{m} \lambda_j}}. \tag{2.21}$$

Figure 2.6 outlines the workflow for calculating the EPI. Pre-processing is crucial as the atmospheric data usually does not follow a Fréchet distribution. Moreover, although the treatment of high-resolution datasets tends to be computationally expensive, they do not increase the information content to the same extent. Therefore, we propose a three-stage preprocessing strategy: 1) Standardization at the grid point level, 2) reduction of the number of grid points, and 3) transformation of the data into a Fréchet distribution with a tail index

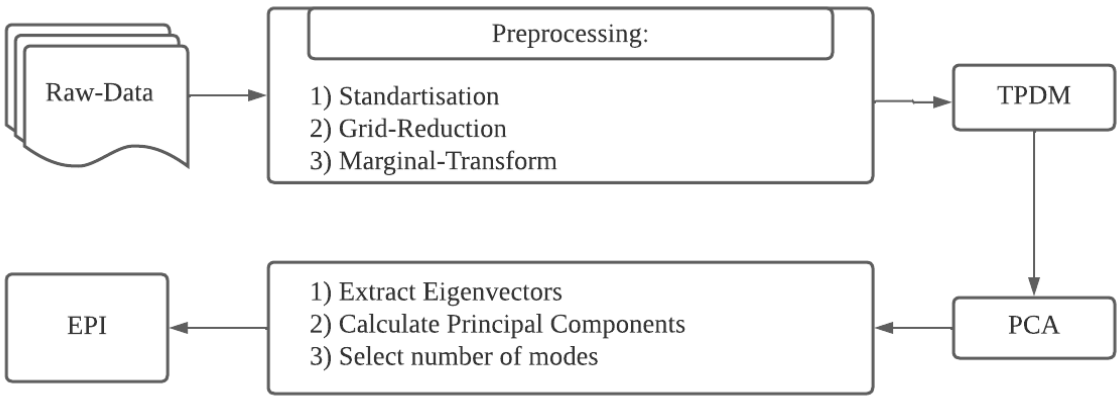


Figure 2.6.: Workflow for calculation of EPI as proposed in Szemkus and Friederichs, 2024.

$\alpha = 2$. The TPDM is then estimated using the pre-processed data, followed by a PCA. Although these steps are mathematically complicated, their application is relatively simple due to the availability of existing functions and packages for the PCA in various programming languages. Currently, the code for estimating the TPDM is only available in R (Szemkus, 2023), but this may change in the future. After performing the PCA, extracting the leading patterns, and calculating their principal components, the number of patterns to be included in the EPI is selected. Alternatively, one can decide to use all patterns. Our analysis reveals that the first ten temperature patterns adequately describe typical heat waves. However, when analyzing further variables or weather extremes, the number of principal components may need to be adjusted.

In Szemkus and Friederichs, 2024, our methodology is applied to the COSMO-REA6 reanalysis data. We demonstrate its effectiveness in capturing the basic characteristics of heat waves. Introducing a novel approach, we define heat waves based on the EPI. We say a heat wave is present in the study area when the EPI exceeds a certain threshold for a minimum of five consecutive days. We find that our EPI-based definition closely matches conventional, grid-point-based definitions. This underlines the suitability of the EPI as a robust index for spatially related extreme event activities within a predefined region. We also succeed in minimizing the reliance on subjective user-defined thresholds.

We conclude that the EPI is a valuable tool for diverse applications such as attribution studies, climate monitoring, and characterization of extreme weather events.

2.4.2. Common extremes in two variables

After initially focusing on describing heat waves, we extend our approach and examine co-occurring extremes in two variables. We propose a cross-TPDM (as an analog to the cross-covariance matrix) that describes pairwise extremal dependencies between the two variables. The cross TPDM of two random vectors $X, Y \in \mathrm{RV}_{-\alpha}$ is, by definition, a part of the TPDM of the joint random vector $Z = (X, Y)$ (see Fig. 2.7). We provide a definition and estimator thereof (see Box 2.3).

When addressing pairwise dependencies between two variables, Singular Value Decomposition (SVD) is commonly employed. SVD can be seen as an extension of PCA to rectangular ma-

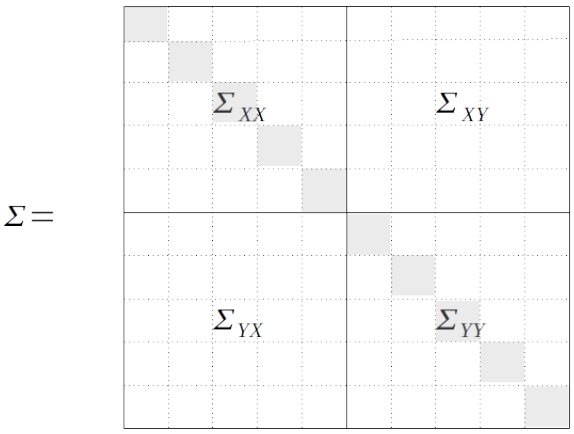


Figure 2.7.: Illustration how the TPDM of joint random-vector $Z = \{X, Y\}$ breaks down in individual TPDM Σ_{XX} and Σ_{YY} and cross-TPDM Σ_{XY} and Σ_{YX}.

trices (Björnsson and Venegas, 1997). The decomposition of the cross-TPDM with $\Sigma_{XY} = U\Lambda V^\top$ yields left- and right-singular vectors U and V, along with the diagonal matrix Λ containing singular values. Analogous to PCA, these singular vectors represent pairs of spatial patterns that account for the maximum possible proportion of extremal dependence. Consequently, each leading mode provides a set of spatial patterns in which extremes in both variables are likely to occur simultaneously.

We apply this methodology to analyze compound meteorological drought and heat waves using COSMO-REA6 reanalysis data. Specifically, we examine daily maximum temperature associated with meteorological drought conditions. Meteorological drought is considered as anomaly of accumulated precipitation, taking into account different duration times from one week to several months. To effectively capture the different dynamics of short- and long-term drought conditions, we introduce a time series shift and analyse their interaction with heat waves.

Our results show a significant influence of preceeding meteorological drought conditions on the occurrence and intensification of heat waves. In addition, we could identify characteristic time intervals between the occurrence of drought and subsequent heat waves on different time scales.

Box 2.3: Estimation of Cross-TPDM

Let x_t, y_t, $t = 1, \ldots, n$ be samples of two Fréchet random vectors X and Y, with components $x_{t,i}$, $i = 1, \ldots, p$ and $y_{t,j}$, $j = 1, \ldots, q$. The estimate of each element of the cross-TPDM Σ_{XY}, analog to the estimate of TPDM in equation 2.14, is given as:

$$\hat{\sigma}_{ij} = 2n_{ij,exc}^{-1} \sum_{t=1}^{n} w_{t,i}^{x} w_{t,j}^{y} \mathbb{I}_{(r_{t,ij} > r_{0,ij})}. \tag{2.22}$$

With radial component $r_{t,ij} = \sqrt{x_{t,i}^2 + y_{t,j}^2}$, threshold $r_{0,ij}$, angular components of X and Y respectively: $w_{t,i}^{x} = \frac{x_{t,i}}{r_{t,ij}}; w_{t,j}^{y} = \frac{y_{t,j}}{r_{t,ij}}$ and $n_{ij,exc} = \sum_{t=1}^{n} \mathbb{I}_{(r_{t,ij} > r_{0,ij})}$, the number of threshold exceedances.

2.4.3. Extremal Pattern Index for two variables

Finally, we extend the EPI to capture simultaneous extremes in two variables. In contrast to utilizing principal components, our new estimator relies on expansion coefficients derived from the projections of the Fréchet distributed data X and Y onto the left and right singular vectors: $\eta = U^T \tau^{-1}(X)$ and $\nu = V^T \tau^{-1}(Y)$. We here employ the transformation function $\tau()$ as described in section 2.3. Similar to PCA, the expansion coefficients provide evidence of extreme weather events in the corresponding pattern described by the singular vector. We define the EPI for the cross-TPDM analogously to equation 2.21 as:

$$\mathrm{EPI}_t^{\eta,\nu} = \frac{\sqrt{\sum_{k=1}^{m}(\eta_{t,k}^2 + \nu_{t,k}^2)}}{\sqrt{\sum_{j=1}^{m}\lambda_j}}. \tag{2.23}$$

This definition of EPI characterizes scenarios where either one or both of the corresponding expansion coefficients are high. To account for simultaneous occurrences of extremes, we introduce a threshold-based approach that considers both the left and right expansion coefficients of each mode k to be prominently expressed. We refer to this as the Threshold-based Extremal Pattern Index (TEPI):

$$\mathrm{TEPI}_t^{\eta,\nu} = \frac{\sqrt{\sum_{k=1}^{m}(\eta_{t,k}^2 + \nu_{t,k}^2)\mathbb{I}_{(|\eta_{t,k}|>q_\eta, |\nu_{t,k}|>q_\nu)}}}{\sqrt{\sum_{j=1}^{m}\lambda_j}}, \tag{2.24}$$

with threshold values q_η and q_ν. The threshold ensures that the associated patterns are simultaneously highly pronounced.

Figure 2.8 shows the changed workflow compared to using the original TPDM and a single variable. While the pre-processing steps remain the same, they are now applied to both datasets individually. This approach ensures that the resulting expansion coefficients have the same magnitude, which is essential for calculating EPI and TEPI with the cross-TPDM.

Once again, we apply this methodology to the COSMO-REA6 reanalysis data. To comprehensively capture the different dynamics of compound heat waves and droughts, we estimate the TEPI for various durations. Additionally, we refer to the previously identified characteristic time shift between the occurrence of meteorological droughts and heat waves. To showcase the utility of TEPI, we focus our analysis on the 2003 and 2010 heat waves. We examine their distinct dynamics in terms of compound extremes as described by the TEPI. Our analysis demonstrates that the TEPI reconstructs the known characteristics of these events, thus highlighting its efficacy as a valuable tool for analyzing compound extremes.

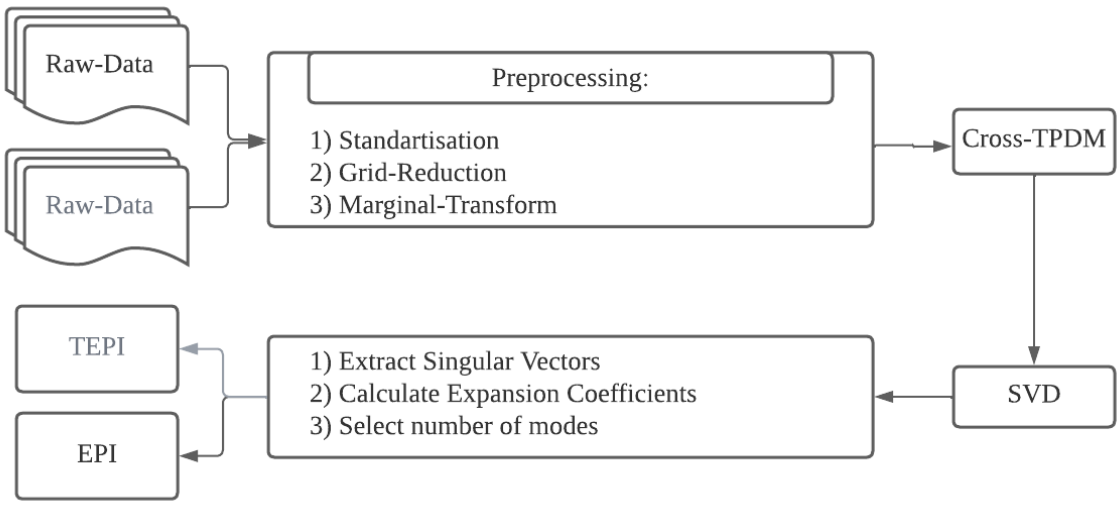


Figure 2.8.: Workflow for calculation of EPI and TEPI from cross-TPDM as proposed in Szemkus and Friederichs, 2024.

2.5. Conclusion & Outlook

The novel pattern-based analysis proposed in this study has proven to be a valuable tool for describing spatially extended extreme weather events. In Xoplaki et al., 2023, we analysed the complex dynamics of the 2018 European heatwave, also taking into account the associated meteorological drought conditions. We identified and investigated the spatial patterns that contributed to this event and analysed and assessed the temporal occurrence of drought and heatwave using the EPI.

Other studies that apply our pattern-based analysis are 1) the ClimXtreme assessment of continuous precipitation in winter 2023/24 over Germany (Grieger et al., 2024), in which the underlying spatial patterns are identified and their properties assessed with respect to historical events and 2) a study on the attribution of heat waves and droughts using the EPI by Meurer, 2023. This study showed that heat waves have been significantly influenced by anthropogenic climate change since around 1984. It also highlighted that the European heatwaves of 2003 and 2010 were significantly impacted by climate change. An increase in the joint occurrence of heatwaves and droughts could also be attributed to climate change.

Further work could include extending the compact description framework using the spatial patterns and the EPI to include a temporal component. Currently, the framework summarises two main characteristics of weather extremes: 1) intensity and 2) spatial extent. These characteristics are crucial for analysing spatially extended extreme weather events, such as heat waves and droughts. However, a third crucial property is the duration of these events. Therefore, the integration of a temporal component into the description framework would be of great benefit. The addition of a temporal component would allow for a more comprehensive analysis of extreme weather events by capturing their persistence over time.

3. Spectral Filter Approach

3.1. Motivation & Previous Work

The latest IPCC report (Masson-Delmotte et al., 2021) summarises findings of changes in daily precipitation. It concludes that 1) for 1.5°C of global warming, longer warm seasons and shorter cold seasons can be expected and 2) the magnitude and intensity of daily precipitation extremes have "very likely" increased across Europe since 1950, particularly in northern hemispheric Summer and Winter months (see, e.g. Helama et al., 2018; Madsen et al., 2014; Field et al., 2012). For Germany, studies reveal weak (decreasing) trends in intensity during the Summer months and more pronounced increasing changes during Winter (Zolina et al., 2008; Zeder and Fischer, 2020). Studies on changes in the seasonality of precipitation extremes are rare. A recent publication suggests weak signals in precipitation extremes towards occurrences later in the years for isolated regions over Germany (Peter et al., 2024). Brönnimann et al., 2018 analyses daily precipitation extremes over the Alpine region and finds a trend towards less frequent events in Summer and more frequent events in Spring and Autumn.

Changes in daily precipitation extremes are usually found to closely follow the Clausius-Clapeyron (CC) relation. A detailed description of this relation is given in the appendix section B.1. The interested reader is further referred to Trenberth et al., 2003. According to the CC relation, the intensity of extreme precipitation is expected to increase by about 7% per degree of global warming as a consequence of the increased water-holding capacity of a warmer atmosphere (see, e.g. Guerreiro et al., 2018; Fischer and Knutti, 2016). However, recent studies of diurnal precipitation extremes indicate that, for certain locations, the intensifications exceed CC-scaling. An increase with a ratio about two times higher than the CC ratio is observed at several locations in Western Europe (Panthou et al., 2014; Park and Min, 2017) and the Netherlands (Lenderink and Van Meijgaard, 2010; Lenderink et al., 2011).

Current research suggests that this unexpectedly strong climate change signal is related to physical processes associated with convective clouds (Lenderink and Van Meijgaard, 2010; Field et al., 2012). Potential mechanisms include dynamical feedback within cloud core updrafts, in-cloud interactions driven by cold pools, and quasi-geostrophic large-scale vertical uplifting (Fowler et al., 2021).

As stratiform and convective precipitation events occur under distinct atmospheric conditions, both types are likely to react individually to climate change and can, therefore, strengthen or weaken each other. Analyzing these events separately has the potential to significantly improve the signal-to-noise ratio, thereby revealing changes that might otherwise remain obscured in the raw data.

Convective precipitation's essential vertical movement can be triggered by frontal passages, orography, or ground warming. In Germany, the latter or a combination is commonly observed, explaining why convective precipitation intensifies during the second half of Summer months. Convective events often manifest as thunderstorms with individual cells exhibiting a typical diameter of less than 10 km and a relatively short lifespan of 30 to 60 minutes (Bott, 2016). In contrast, stratiform precipitation is caused by large-scale uplift processes, often associated

with fronts. Characterized by a slow, large-scale uplift (on average about 10cm/s^{-1} on warm fronts), stratiform events are of a much larger spatial and temporal extent in comparison. Their spatial extent can be up to 200-300 km (Bott, 2016). The differentiation of both event types becomes more complicated when individual thunderstorm cells combine to form multicells. Under favorable atmospheric conditions, such as vertical shear, multicells can continuously initiate new convective single cells and exceed their typical size and lifetime. These systems form a continuous precipitation area of at least 100 km in one direction, with a typical lifetime of about 10 hours (Bott, 2016).

Various approaches have been carried out to discriminate between stratiform and convective precipitation. They typically rely on remote-sensing data (cloud drop size, ice, graupel presence, etc.) (e.g. Wang et al., 2021; Ahmed and Schumacher, 2015) or routine meteorological observations, as these typically include cloud-type descriptions (Chernokulsky et al., 2019; Ye et al., 2017). Alternative approaches include the investigation of the spatio-temporal characteristics of stratiform and convective precipitation.

Promising results have been obtained from studies that take into account the spatiotemporal structure of precipitation events (see also Brune et al., 2018). Lengfeld et al., 2019 use hourly and daily radar data over Germany and show that the spatial extent of precipitation is significantly reduced in Summer, compared to the winter months, which is due to the increased amount of convective precipitation events in Summer. Han et al., 2016 distinguish between convective and stratiform events using daily precipitation data from multiple weather stations. They do so by incorporating precipitation amounts from neighboring weather stations to account for the spatial structure of the events.

Another perspective to describe extreme precipitation events on various scales comes from methods that accumulate precipitation to different duration levels. Such methods include, e.g., intensity–duration–frequency (IDF)-curves as applied in Fauer et al., 2021 or the cross-scale weather extremity index (xWEI) as proposed in Voit and Heistermann, 2022. These methods differ from spectral decomposition as they do not apply a low-pass filter. This means they consider all events up to a certain duration instead of separating events into varying scales. As a result, they offer an alternative perspective on precipitation events.

In a recent publication, Benestad et al., 2022 employed an approach using a wavelet transform to investigate evolving precipitation patterns within the global ERA5 reanalysis. Their methodology involved the application of a two-dimensional Haar wavelet transform. The study highlighted a significant shift in precipitation variability towards smaller spatial scales in tropical regions. However, it has been subsequently demonstrated that this shift is primarily induced by changes in the observation system rather than actual climate change (Buschow, 2023b). Nevertheless, the study demonstrated the suitability of spectral filtering approaches for scale-dependent climate data investigations.

There have been several studies that decompose hourly precipitation fields using wavelet transformation (Weniger et al., 2017; Kapp et al., 2018). For hourly precipitation fields, the complex dual-tree wavelet transform is particularly suited, as demonstrated in Buschow et al., 2019; Buschow and Friederichs, 2020; Brune et al., 2020.

This study explores how the complex dual-tree wavelet transform can be utilized to investigate changes in extreme precipitation events over Germany. To do so, the two European reanalyses, COSMO-REA6 and CERRA, are utilized, focussing on three areas in Germany.

The remainder of this chapter is as follows: Section 3.2 introduces the wavelet transform, emphasizing key achievements and useful properties. Section 3.3 describes the data and the preprocessing steps. It also provides a comprehensive analysis of how precipitation events are

represented in the two datasets, COSMO-REA6 and CERRA. In section 3.4, both datasets are examined for non-stationarities. The analysis is divided into the examination of changes in intensity and seasonality. Finally, Section 3.5 presents conclusions and outlines future directions.

3.2. Wavelet Theory

This section provides an overview of the wavelet transform. The main information is taken from Daubechies, 1988; Mallat, 1989; Meyer, 1997; Mallat, 1998. Starting with fundamental definitions and properties in the continuous case, the focus shifts to the discrete wavelet transform. This section then systematically moves from one-dimensional to two-dimensional cases and extends the theoretical framework to account for spatial complexities. This section is intended to give the reader an understanding of wavelet theory and to provide a foundation for the subsequent analyses and applications in this study.

A Wavelet Transform (WT) is a mathematical technique for decomposing data, functions, or operators into their frequency spaces, making it a powerful tool in spectral analysis. In WT, signals are analyzed using a set of basis functions called wavelets, denoted as $\psi(t)$. The term 'wavelet' originates from 'little wave' and reflects the property that wavelets integrate to zero, known as vanishing mean (Daubechies, 1992):

$$\int \psi(t) = 0. \tag{3.1}$$

Unlike in traditional Fourier Transform (FT), which uses sinusoids as basis functions, wavelet basis functions are localized both in time and frequency domains. Given a continous signal $f(t)$ the decomposition into frequency space using the FT (3.2) and WT (3.3) are as follows:

$$f(\omega) = \frac{1}{\sqrt{2\pi}} \int \exp^{-i\omega t} f(t) dt \tag{3.2}$$

$$f(a,b) = \frac{1}{\sqrt{a}} \int \psi\Big(\frac{t-b}{a}\Big) f(t) dt, \tag{3.3}$$

where ω is the frequency. The function $\psi(t)$, is known as mother-wavelet, the function $\psi^{a,b}(t) = \frac{1}{\sqrt{a}}\psi\big(\frac{t-b}{a}\big)$ is called a daughter wavelet, with continuous parameters a and b for dilating and translating (a is also known as scale parameter). Thus, in comparison with the FT that decomposes a signal using fixed basis functions, the WT offers a more flexible approach by utilizing a diverse set of basis functions that can be adapted to the characteristics of the signal. The time-frequency localization enables the WT not only to identify the frequency content of a particular signal but also to determine when these frequencies occur.

Figure 3.1 demonstrates this advantage of time-frequency localization using three simple examples: 1) a signal that abruptly changes its frequency to a fixed value, 2) a signal with linear decreasing frequency, and 3) the reverse version of 2). The FT provides information about the frequency content of each signal but does not convey any information about the time of occurrence. Consequently, the FT of examples 2) and 3) are identical. The WT, in contrast, offers insights into the temporal occurrence of events such as the frequency jump in example 1) and can differentiate between the linear increase and decrease of frequencies in examples 2) and 3).

Figure 3.2 illustrates how scaling parameter a and parameter b adapt the wavelet $\psi^{a,b}(t)$ with respect to its respective frequency. The wavelet function decomposes the frequency axis in dyadic intervals with exponential growth (Mallat, 1998). At low frequencies, the wavelet stretches, and at fast fluctuations, the wavelet compresses. This allows a good temporal resolution at higher frequencies, allowing the wavelet to 'zoom in' on short-lived, highly-frequent phenomena.

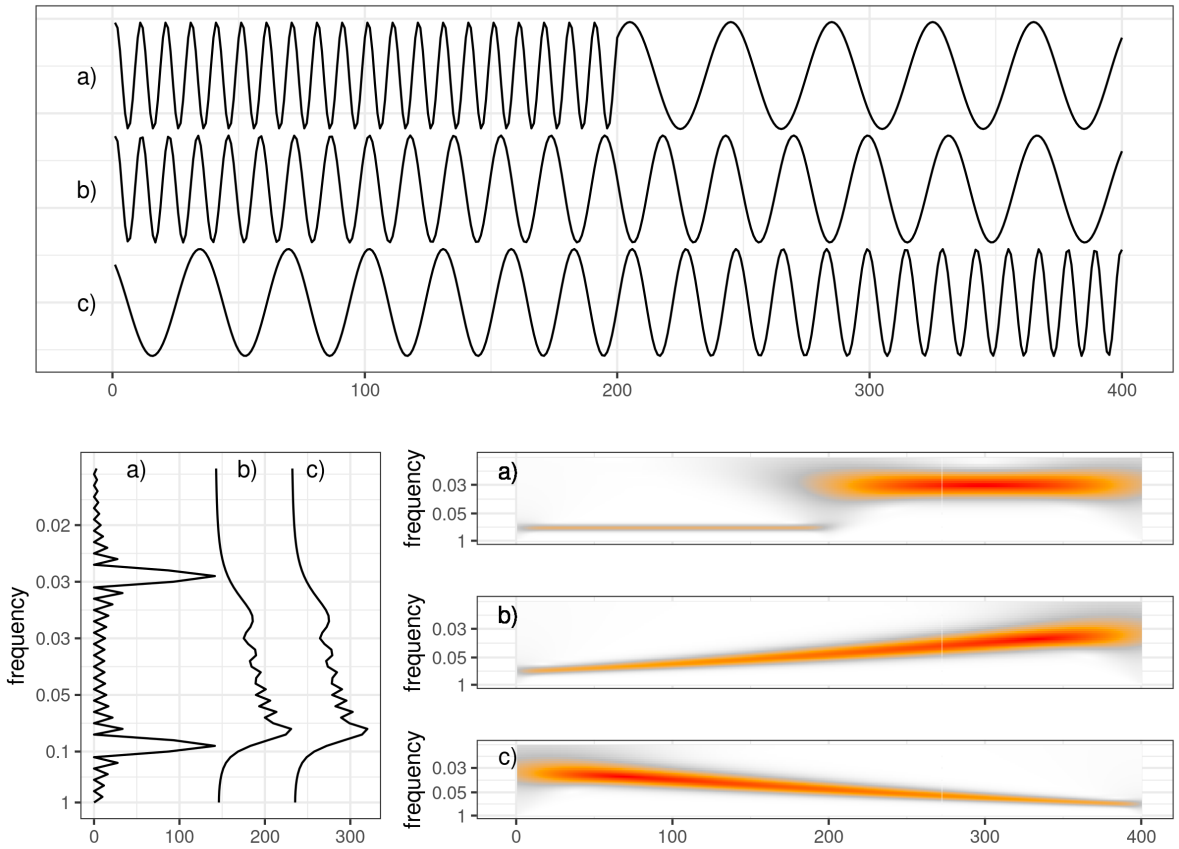


Figure 3.1.: Top: Timeseries of three signals $f(t) = \cos(2\pi * F(t)t) + \sin(2\pi F(t)t)$. a) with apprupt frequency change $F(t)$ from 0.1 to 0.025 at $t = 200$, b) linear decreasing frequency with $F(t) = 0.1$ at $t = 1$ and $F(t) = 0.025$ at $t = 400$ and c) linear increasing frequency with $F(t) = 0.025$ at $t = 1$ and $F(t) = 0.1$ at $t = 400$.
Bottom: Frequency content obtained from FT (left) Frequency content obtained from WT (right). he frequency axes are logarithmic.

3.2.1. 1 Dimension: Discrete Wavelet Transform

Consider a discrete signal $f(t)$. The Discrete Wavelet Transform (DWT) can be calculated from equation 3.3. Given t and ω are discrete and equally spaced with $\omega = m\omega_0$, $t = nt_0$ and $m, n \in \mathbb{Z}$, $\omega_0, t_0 > 0$. The parameters a and b, assumed to be continuous beforehand in equation 3.3 are now also considered to be discrete with $a = a_0^m$ and $b = nb_0a_0^m$ where $a_0, b_0 > 0$. The relation of a and b ensures a time-frequency resolution as given in Fig. 3.2. The discretized wavelet function $\psi^{m,n}(t)$ and DWT $f^{m,n}$ are then given as follows:

$$\psi^{m,n}(t) = \frac{1}{\sqrt{a_0^m}}\psi\left(\frac{t - nb_0a_0^m}{a_0^m}\right) \tag{3.4}$$

$$f^{m,n} = \langle f(t), \psi^{m,n}(t)\rangle = \int \psi^{m,n}(t)f(t)dt. \tag{3.5}$$

If a_0 and b_0 are chosen near 1 and 0, respectively, the resulting wavelet turns out to be redundant and very close to the continuous wavelets (Daubechies, 1988). To find orthogonal wavelets, i.e. $\langle \psi^{m,n}(t), \psi^{m',n'}(t) \rangle = \delta(m,m')\delta(n,n')$ where $\delta(m,m') = 1$ if $m = m'$ and 0 otherwise, a typically choice is $a_0 = 2$ and $b_0 = 1$ (Daubechies, 1988; Meyer, 1997; Mallat, 1989, see e.g.). Orthogonal wavelets establish an orthogonal basis within $L^2(\mathbb{R})$, the space of signals f with measurable and integrable energy. Consequently, they offer a non-redundant representation and ensure perfect reconstruction of signals from their decomposition. However, orthogonal wavelets are not shift-invariant, meaning small shifts in the input signal can result in significant changes in the transformation.

This section continues with an overview of Multiresolution Approximation (MRA) for DWT that was introduced by Meyer and Mallat in 1986. The MRA approach marked a significant advance and revolutionized DWT as it provides 1) an efficient algorithm for its computation and 2) a framework for the systematic development of new orthonormal wavelet families. To visualize the application of the MRA approach, an example is given utilizing the Haar wavelet (Haar, 1909), which is characterized by its simplicity. Although the Haar wavelet was invented in 1909, it can be shown that the principles of MRA can be applied. We then end by introducing complex wavelets, specifically the Dualtree Complex Wavelet Transform (DT-CWT), which is a more recent development in DWT (Kingsbury, 1998).

Multiresolution Approximation

The MRA framework, introduced by Mallat and Meyer in 1986 (Meyer, 1997; Mallat, 1989), offers a hierarchical structure for understanding signal information across different resolutions. Within this framework, a signal f is decomposed into components at various resolutions, enabling representations ranging from coarse approximations to fine details.

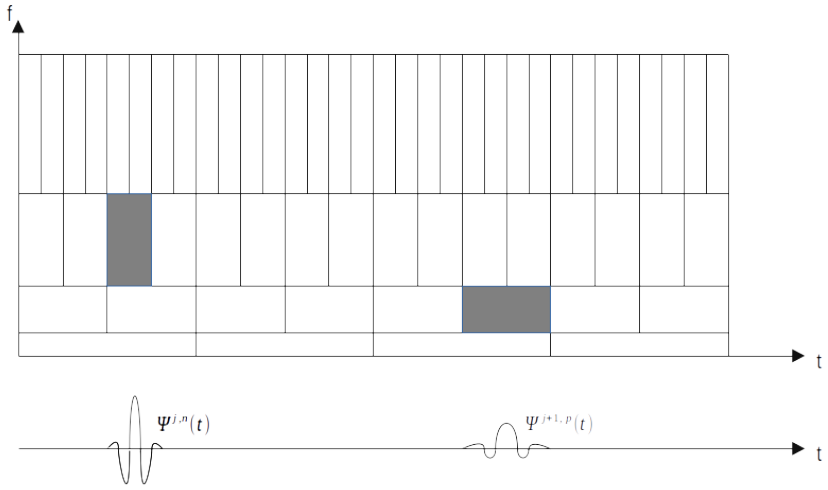


Figure 3.2.: Illustration of time-frequency resolution for the WT (figure adapted from Mallat, 1998).

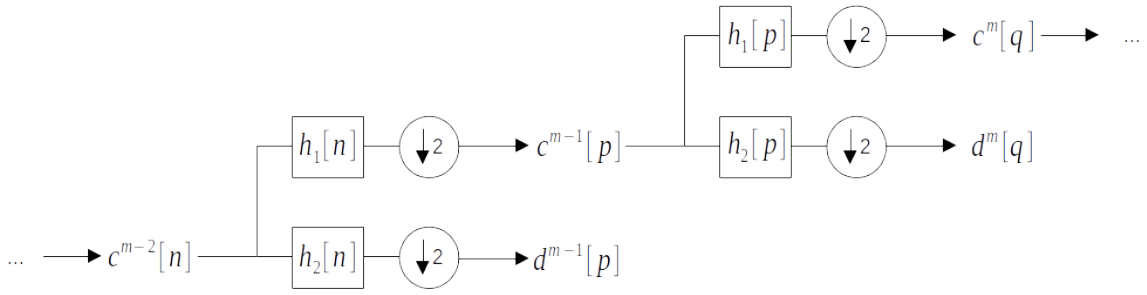


Figure 3.3.: Scematic illustration of the one-dimensional pyramid algorithm for the DWT. h_1 and h_2 denote applications of the high- or low-pass filter, respectively. $2\downarrow$ represents a down-sampling step.

Box 2.1: Definition MRA

Let $L^2(\mathbb{R})$ denote the vector space of signals f with measurable and integrable energy. Within this space, $V_{2^m} \subset L^2(\mathbb{R})$ represents the set of all possible signal approximations at a resolution of 2^m where $m \in \mathbb{Z}$. m denotes different levels of detail. Thus, V_{2^m} serves as a multiresolution approximation of $L^2(\mathbb{R})$.

There exists a unique function $\phi(t)$, referred to as the scaling function or 'father wavelet', whose dilation and translations form an orthonormal basis in V_{2^m}:

$$\phi^{m,n}(t) = \frac{1}{\sqrt{2^m}}\phi\left(\frac{t-n}{2^m}\right). \tag{3.6}$$

Similarly, there exists a unique function $\psi^{m,n}(t)$ forming an orthonormal basis in the vector space W^{2^m}, which is the orthogonal complement of V_{2^m}, such that $V_{2^m} \perp W_{2^m}$, where $W_{2^m} \subset L^2(\mathbb{R})$. This function $\psi^{m,n}(t)$ is the wavelet function:

$$\psi^{m,n}(t) = \frac{1}{\sqrt{2^m}}\psi\left(\frac{t-2^m n}{2^m}\right). \tag{3.7}$$

In summary, Mallat and Meyer demonstrated that one can derive wavelet functions forming an orthonormal basis of W_{2^m} and thereby also an orthonormal basis of $L^2(\mathbb{R})$ through dilation the scaling function $\phi(t)$. They further utilize a corresponding low-pass & high-pass filter pair h_0 h_1 such that:

$$h_1[n] = (-1)^{1-n}h_0[1-n] \tag{3.8}$$

$$h_0[n] = \frac{1}{\sqrt{2}}\sum_t \phi\left(\frac{t}{2}\right)\phi(t-n) \tag{3.9}$$

$$\frac{1}{\sqrt{2}}\psi(\frac{t}{2}) = \sum_n h_1[n]\phi(t-n) = \sum_n (-1)^{1-n}h_0[1-n]\phi(t-n). \tag{3.10}$$

For detailed information and proof, the reader is referred to Mallat, 1998.

To provide an efficient algorithm for DWT-computation that gives perfect reconstruction

Mallat, 1989 introduces the scaling coefficient $c^m[p] = \langle f, \phi^{m,n} \rangle$ and the wavelet coefficient $d^m[p] = \langle f, \psi^{m,n} \rangle$. $c^m[p]$ serves as an approximation of the original signal at resolution step m, while $d^m[p]$ gives some difference measure between $c^m[p]$ and $c^{m-1}[p]$. Since $\phi^{m,n}$ and $\psi^{m,n}$ form orthonormal bases of V_{2^m} and W_{2^m} respectively, the combination of $c^m[n]$ and $d^m[n]$ provides all the information needed to reconstruct the signal. It can be further shown, that c^{m+1} and d^{m+1} can be directly calculated from $c^m[n]$:

$$c^{m+1}[p] = \sum_n h[n-2p]c^m[n] \tag{3.11}$$

$$d^{m+1}[p] = \sum_n g[n-2p]c^m[n]. \tag{3.12}$$

Therefore, the DWT can be computed using a pyramid algorithm that initially calculates $c^m[n]$ at its finest resolution and then proceeds step-wise through each resolution level until reaching the coarsest scale. A schematic illustration of this algorithm is provided in Fig. 3.3.

Haar Wavelet

Consider a simple, step-wise scaling function with a corresponding mirror filter $h_0[n]$, as computed from equation 3.9:

$$\phi(x) = \begin{cases} 1 \text{ if } 0 \le x < 1 \\ 0 \text{ else} \end{cases} \qquad h_0[n] = \begin{cases} 1/\sqrt{2} \text{ if } n = 0,1 \\ 0 \text{ else} \end{cases} \tag{3.13}$$

The corresponding mother wavelet and with high-pass filter function $h_1[n]$, can be calculated from equations 3.10 and 3.8:

$$\psi(x) = \begin{cases} 1 \text{ if } 0 \le x \le \frac{1}{2} \\ -1 \text{ if } \frac{1}{2} \le x \le 1 \\ 0 \text{ else} \end{cases} \qquad h_1[n] = \begin{cases} -1 \text{ if } n = 0 \\ 1 \text{ if } n = 1 \\ 0 \text{ else} \end{cases} \tag{3.14}$$

Figure 3.4 shows $\phi(x)$ and $\psi(x)$. $\psi(x)$ is the Haar wavelet, the earliest wavelet, dating back to 1909 (Haar, 1909). The Haar wavelet is recognized for its simplicity and efficiency in signal analysis. Importantly, the functions $\phi^{m,n}(t)$ and $\psi^{m,n}(t)$ (equations 3.7 and 3.6), derived from $\phi(x)$ and $\psi(x)$, respectively, form an orthonormal basis in $L^2(\mathbb{R})$. However, the Haar wavelet has poor frequency localization and, therefore, does not accurately capture the frequency content of the signals. Due to its discontinuity, it is not differentiable. It also suffers from a common disadvantage of discrete wavelets, namely that it is shift-variant, which means that small shifts in the input signal can lead to significant differences in its scale- and wavelet coefficients. Given any signal $f(t)$ of length 2^l with $l \in \mathbb{Z}$, we construct the signal approximation (scaling coefficients) $c^m[p]$ and wavelet coefficients $c^m[p]$ with $m = 0, \ldots, l$ at each resolution step from equations 3.11 and 3.12:

$$c^{m+1}[p] = \frac{1}{\sqrt{2}}(c^m[2p] + c^m[1+2p]) \tag{3.15}$$

$$d^{m+1}[p] = c^m[1+2p] - c^m[2p]. \tag{3.16}$$

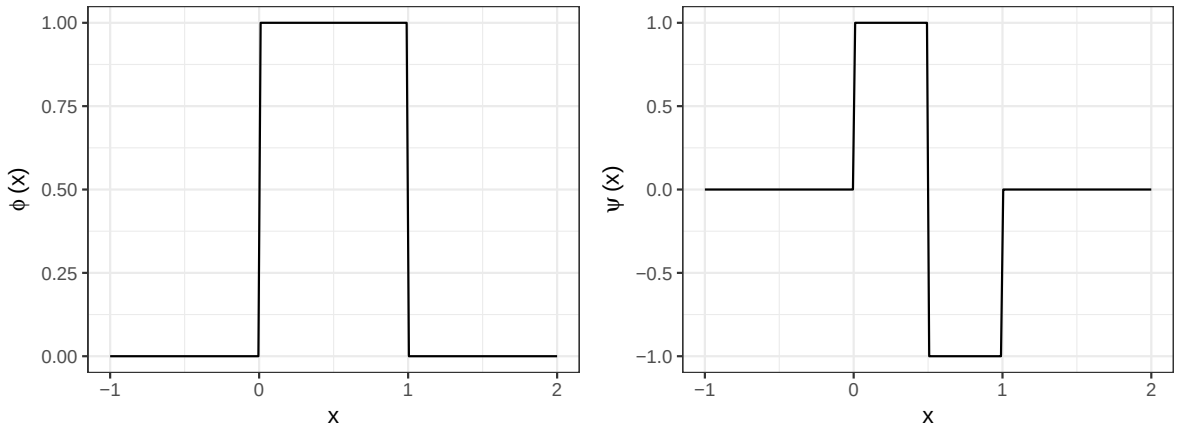


Figure 3.4.: Illustration of Haar wavelet $\psi(x)$ and it's scaling function $\phi(x)$.

Complex Dual-tree Wavelet

This section considers the Complex Wavelet Transform (CWT) that relies on complex but discrete wavelets:

$$\psi_c(t) = \psi_r(t) + i\psi_i(t), \tag{3.17}$$

where $\psi_r(t)$ and $\psi_i(t)$ represent the real and imaginary parts, respectively.

Generally, complex wavelets do not form an orthogonal basis in $L^2(\mathbb{R})$. However, in comparison to real-valued DWT, they are (approximately) shift-invariant and provide both magnitude and phase information. However, complex wavelets tend to be more computationally expensive compared to orthonormal wavelets.

First introduced in 1998 by Kingsbury (Kingsbury, 1998), the DT-CWT utilizes two real-valued DWTs, one for the real part and one for the imaginary part of the signal. Given two discrete wavelets $\psi_h(t)$ and $\psi_g(t)$, associated with a real-valued DWT each, the complex wavelet is given as $\psi(t) := \psi_h(t) + i\psi_g(t)$. The real-valued wavelet coefficients d_h and d_h, equivalently form the complex coefficient of DT-CWT: $d_c = d_h + id_g$. By splitting the wavelet in two, knowledge and methods from real-valued DWT can be applied. This, e.g., includes criteria for wavelet selection and established algorithms for their computation. The algorithm for the computation of DT-CWT is shown in Fig. 3.5. The name dual-tree comes from the two branches for real and imaginary parts, respectively, each consisting of a regular MRA algorithm, as already described in section 3.2.1.

However, the complex wavelet $\psi(t)_c$ requires to be approximately analytic (for definition, see Box 2.2). Thus, the DT-CWT cannot utilize common discrete wavelets as, e.g., the Haar wavelet. We hence write $\psi_g(t) \sim H\{\psi_h(t)\}$. With $H\{\ldots\}$, the Hilbert transform.

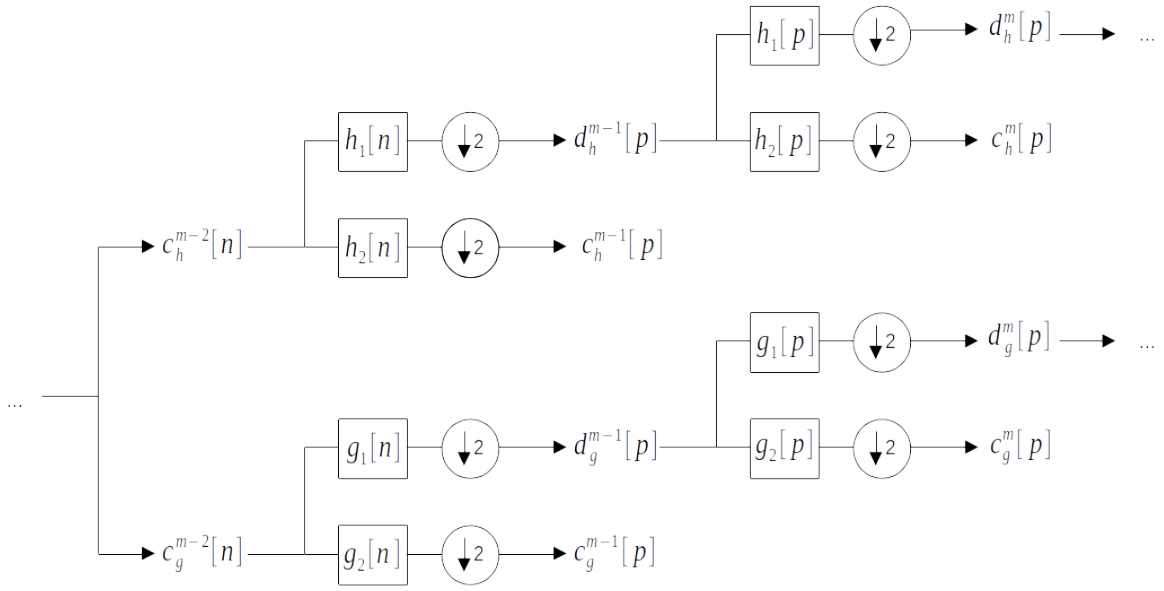


Figure 3.5.: Scematic illustration of the one-dimensional pyramid algorithm for the DT-CWT. h_1, g_t and h_2, g_2 denote the high- or low-pass filter, respectively. $2\downarrow$ represents a down-sampling step.

Box 2.2: Definition: analytic signal

An analytic signal is defined as a complex-valued function characterized by the absence of negative frequency components. The real and imaginary components of an analytic signal are related through the application of the Hilbert transformation.
The Hilbert transform creates a 90-degree phase shift within the frequency domain. The result of this transformative process is that negative frequency components are effectively eliminated, leaving only the positive frequency components within the resulting signal.

3.2.2. 2 Dimensions: Discrete Wavelet Transform

This section considers a 2-dimensional signal $f(x,y) \in L^2(\mathbb{R}^2)$. The DWT counterpart of equation 3.19 for the 2-dimensional case can be written as follows:

$$\psi_k^{m,n}(x,y) = \frac{1}{2^m}\psi^k\left(\frac{x-2^m n_x}{2^m}, \frac{y-2^m n_y}{2^m}\right) \tag{3.18}$$

$$f_k^{m,n} = \langle f(x,y), \psi_k^{m,n}(x,y)\rangle = \int\int \psi_k^{m,n}(x,y) f(x,y) dxdy. \tag{3.19}$$

with $k = 1,2,3$. The following section provides a generalization of 1) the MRA framework and 2) the DT-CWT for the 2-dimensional case.

Multiresolution Approximation

To extend the MAR approximation to two dimensions Mallat, 1989; Meyer, 1997 define a wavelet family of three directional wavelets that, each construct a separable orthonormal basis

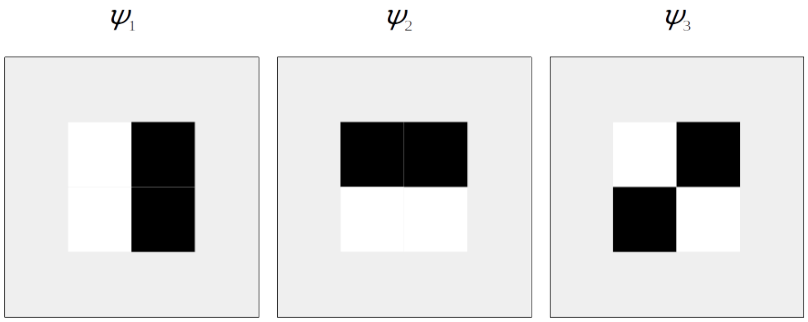


Figure 3.6.: Schematic illustration of Haar-wavelets in two dimensions according to equations 3.21-3.23. The white and black areas indicate positive and negative values, respectively.

in $L^2(\mathbb{R}^2)$: $\{\psi_1^{m,n}(x,y), \psi_2^{m,n}(x,y), \psi_3^{m,n}(x,y)\}$, the daughter wavelets. Their definition relies on the scaling function ϕ and wavelet ψ as introduced in section 3.2.1:

$$\phi(x,y) = \phi(x)\phi(y) \tag{3.20}$$
$$\psi_1(x,y) = \phi(x)\psi(y) \tag{3.21}$$
$$\psi_2(x,y) = \psi(x)\phi(y) \tag{3.22}$$
$$\psi_3(x,y) = \psi(x)\psi(y). \tag{3.23}$$
$$\tag{3.24}$$

Figure 3.6 illustrates the three wavelets associated with the Haar wavelet. These wavelets, denoted as $\psi_1(x,y)$ (vertical), $\psi_2(x,y)$ (horizontal), and $\psi_3(x,y)$ (diagonal), each capture signals oriented in distinct directions.

Consequently, when employing the MRA algorithm in two dimensions, three distinct wavelet coefficients are obtained for each level of resolution. This decomposition allows for a comprehensive analysis of the signal's structure and content across different orientations:

$$c^m[p,q] = \langle f, \phi^{m,n}(x,y)\rangle \tag{3.25}$$
$$d_k^m[p,q] = \langle f, \psi_k^{m,n}(x,y)\rangle \tag{3.26}$$

Figure 3.7 presents a two-dimensional decomposition, using the Uni-bonn logo as an illustrative example. The Figure showcases wavelet coefficients corresponding to resolution steps one, two, and three. This visualization effectively demonstrates the separation of horizontal, vertical, and diagonal structures into the three distinct wavelet coefficients.

Complex Dual-tree Wavelet

The 2-dimensional DT-CWT builds on the two real-valued discrete wavelets, $\psi_h(t)$ and $\psi_g(t)$ with scaling functions $\phi_h(t)$ and $\phi_g(t)$, as introduced in section 3.2.1. In two dimensions, the DT-CWT is oriented, approximately analytic, and shift-invariant. For further detail, see Selesnick, 2002. Analogous to the real-valued DWT, this approach defines a wavelet family of directional daughter wavelets that construct a separable orthonormal basis in $L^2(\mathbb{R}^2)$. However, the dual-tree approach expands the wavelet family to include not three, but six complex wavelets $\{\psi_1^{m,n}(x,y), \ldots, \psi_6^{m,n}(x,y)\}$ with real- and imaginary part respectively.

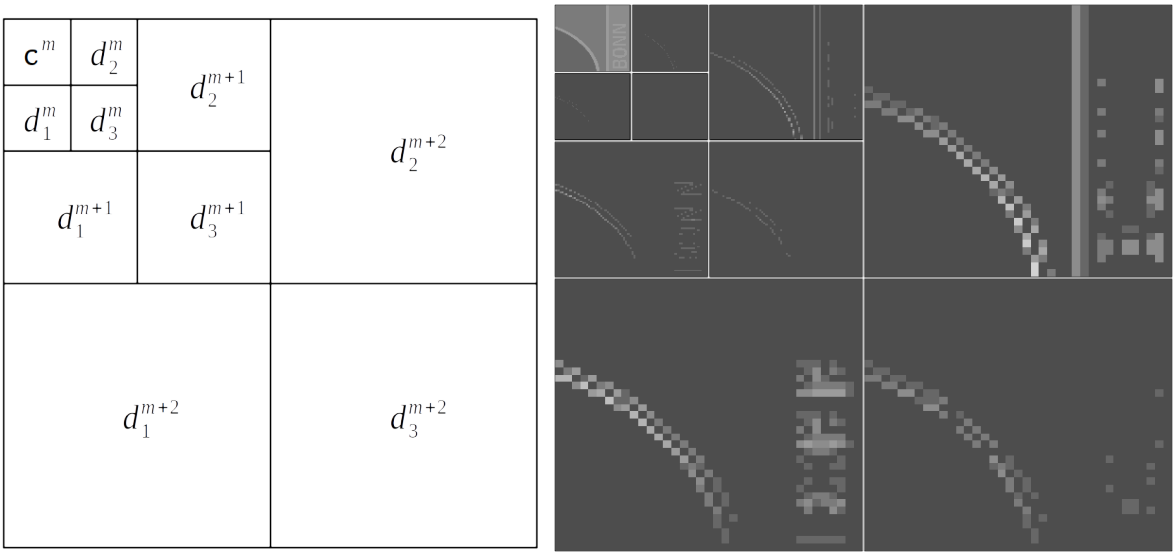


Figure 3.7.: left: structure of wavelet coefficients $d_k^m[p]$; right: 2-Dimensional Haar-Wavlet transform of Uni-Bonn logo for three resolution steps. Shown are the absolute values of $d_k^m[p]$.

Given the following separable 2-D wavelet bases:

$$\psi_{1,1}(x,y) = \phi_h(x)\psi_h(y) \qquad \psi_{2,1}(x,y) = \phi_g(x)\psi_g(y) \tag{3.27}$$
$$\psi_{1,2}(x,y) = \psi_h(x)\phi_h(y) \qquad \psi_{2,2}(x,y) = \psi_g(x)\phi_g(y) \tag{3.28}$$
$$\psi_{1,3}(x,y) = \psi_h(x)\psi_h(y) \qquad \psi_{2,3}(x,y) = \psi_g(x)\psi_g(y) \tag{3.29}$$

$$\psi_{3,1}(x,y) = \phi_g(x)\psi_h(y) \qquad \psi_{4,1}(x,y) = \phi_h(x)\psi_g(y) \tag{3.30}$$
$$\psi_{3,2}(x,y) = \psi_g(x)\phi_h(y) \qquad \psi_{4,2}(x,y) = \psi_h(x)\phi_g(y) \tag{3.31}$$
$$\psi_{3,3}(x,y) = \psi_g(x)\psi_h(y) \qquad \psi_{4,3}(x,y) = \psi_h(x)\psi_g(y). \tag{3.32}$$

The real- and imaginary parts of the complex daughter wavelets are calculated as follows:

$$R(\psi_i(x,y)) = \frac{1}{\sqrt{2}}(\psi_{3,i}(x,y) + \psi_{4,i}(x,y)) \tag{3.33}$$
$$R(\psi_{i+3}(x,y)) = \frac{1}{\sqrt{2}}(\psi_{3,i}(x,y) - \psi_{4,i}(x,y)) \tag{3.34}$$

$$I(\psi_i(x,y)) = \frac{1}{\sqrt{2}}(\psi_{1,i}(x,y) - \psi_{2,i}(x,y)) \tag{3.35}$$
$$I(\psi_{i+3}(x,y)) = \frac{1}{\sqrt{2}}(\psi_{1,i}(x,y) + \psi_{2,i}(x,y)) \tag{3.36}$$

for $i = 1,2,3$.

Figure 3.8 illustrates the six orientated and complex wavelets as described above. Each of these (daughter) wavelets isolates one well-defined direction: 15°, 45°, 75°, 105°, 135°, 165°. In direct comparison with the Haar wavelet (Fig. 3.6), it can be easily seen that the DT-CWT

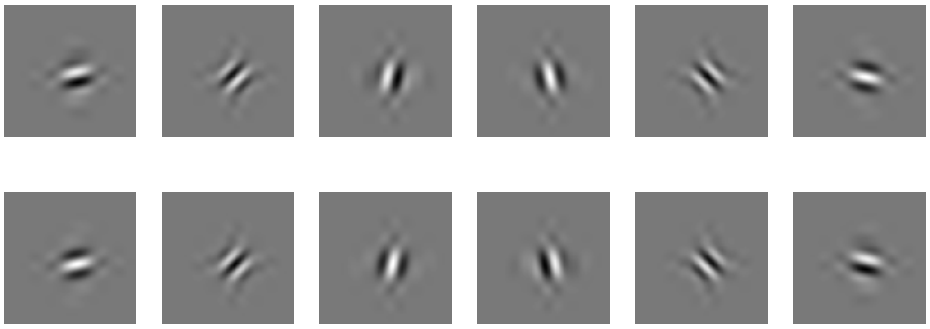


Figure 3.8.: Six directed daughter wavelets for the DT-CWT as defined by Kingsbury, 1998. Top row shows the real part and bottom row the imaginary part.

gives a more specific extraction of directional information. Notably, the real and imaginary components of each wavelet align in orientation, essentially representing phase-shifted versions of one another. Consequently, the magnitude of each wavelet is an approximately bell-shaped function. This property ensures unambiguous localization of image features within the support of each daughter wavelet.

As for the one-dimensional case, the DT-CWT can be performed by running two separate DWTs for real- and imaginary parts, respectively. The resulting algorithm is about four times more computationally expensive than for real-valued DWT.

3.2.3. Further Considerations

Up to this point, this section has provided a broad overview of DWT. This subsection now addresses additional considerations relevant to the application of DWT to spatial precipitation fields described.

Redundant Discrete Wavelet Transform

Let's consider a spatial field on $2^8 \times 2^8 = 512 \times 512$ grid points. The 2-dimensional DT-CWT, employing the MRA algorithm as described above, decomposes the precipitation field into $m = 0, \ldots, 8$ resolution steps where resolution step m has dimension $2^{8-m} \times 2^{8-m}$. The estimator of the wavelet coefficient at each grid point, for each m, could, however, provide precise information about the location of the respective events in space. The redundant discrete wavelet transform (RDWT), initially introduced by Nason et al., 2000 and adapted to DT-CWT by Nelson et al., 2018, provides precisely this grid point-wise resolution of spatial wavelet coefficients. Simply said, the wavelet is shifted to every possible position and then averaged over the results, which is analogous to dropping the constraint $b = nb_0a_m$ in equation 3.4.

Mathematically speaking, the signal $f(x,y)$ is divided into overlapping or non-overlapping windows $w^i(x,y)$, where i indexes the window. Within each window, the 2-dimensional DT-CWT is performed independently, resulting in localized wavelet coefficients $d_k^{m,i}[p,q]$ where m and k represent scale and orientation indices, respectively. The counterpart to equation 3.26 for the

RDWT algorithm is given as:

$$d_k^{m,i}[p,q] = \langle w^i(x,y), \psi_k^{m,n}(x,y) \rangle. \tag{3.37}$$

The respective field of wavelet coefficients on the original grid $d_k^m(x,y)$ is

$$d_k^m(x,y) = \frac{1}{N(x,y)} \sum_i d_k^{m,i}[x - x_0, y - y_0]. \tag{3.38}$$

$N(x,y)$ is the number of windows $w^i(x,y)$ contributing to position (x,y) and (x_0,y_0) are the coordinates of the top-left corner of window $w^i(x,y)$. In practice, these windows are not directly implemented in the RDWT. Instead, one adds zeros in the filters and drops the down-sampling steps. The RDWT offers shift invariance, improved time-frequency localisation and extended display options, disadvantages are increased complexity and tendency to overfitting.

Wavelet Spectrum

As detailed in section 3.2.1, the complex wavelet coefficients $d_k^m(x,y) = d_k^m(x,y)_R + i d_k^m(x,y)_I$ provide magnitude and phase information. In this study, the magnitude is considered as it contains the energy information of the underlying signal at a particular scale and orientation. It is calculated as follows:

$$|d_k^m(x,y)| = \sqrt{(d_k^m(x,y)_R)^2 + (d_k^m(x,y)_I)^2}. \tag{3.39}$$

The *wavelet spectrum*, i.e., the wavelet analog to the Fourier spectrum, is defined as the squared magnitude $|d_k^m(x,y)|^2$. In the case of employing the RDWT, it is usually referred to as the local wavelet spectrum (LWS) (see e.g. Nelson et al., 2018; Eckley et al., 2010). Following Nelson et al., 2018, the LWS can be used to obtain an estimate of spatial covariances. However, a bias correction of the local wavelet spectrum is required, as this estimator overemphasizes larger scales. This bias can be removed by multiplying the LWS by a matrix that depends only on the choice of wavelet and the dimensions of the input field.

After removing the bias, the orientation information of LWS, as provided by DT-CWT, is summarised as follows:

$$|d^m(x,y)|^2 = \sum_{k=1}^{6} |d_k^m(x,y)|^2 \tag{3.40}$$

with $k = 1,\ldots,6$. The remainder of this chapter refers to $|d^m(x,y)|^2$ as the LWS where m is the resolution step. The following section explores the spatial scale represented by each resolution step.

As Selesnick, 2002 state, a suitable interpretation for LWS can be derived, as it follows the principles of a Parseval's Energy theorem. This theorem is known from FT and states that the total energy of a signal in the time domain equals the total energy of its FT in the frequency domain. This principle extends to the DT-CWT as follows:

$$\sum_m \sum_{x,y} (|d^m(x,y)|^2) = \sum_{x,y} |f(x,y)|^2 \tag{3.41}$$

When applied to RDWT, this theorem is subject to inaccuracies due to the redundancy introduced and the tendency to overestimate larger scales, which cannot be completely eliminated

by the bias correction. However, despite this limitation, it provides a point of reference for analyzing the contribution of the different resolution steps to the total wavelet energy.

Quantification of the spatial scale

This section aims to define an equivalent scale in km for the dual-tree daughter wavelets ψ_m with resolution steps $m = 0,\dots,6$. Buschow, 2023b draws an analogy to spherical harmonics to establish a relationship between the resolution levels and the equivalent wavelength. He finds an approximately exponential relationship. Brune et al., 2018 draws a connection from an interpretation rooted in the MRA framework and quantifies the spatial scale given by each resolution step m as $gs \cdot 2^m$, where gs is the grid point spacing in km.

Let be given a $2^9 \times 2^9 = 512 \times 512$ grid that contains only zero values, except for a certain domain of $k \times k$ grid points in the center that are set to one. Figure 3.9 shows the normalised LWS for each $k = 1,\dots 50$. The lines show the approximations as suggested in Buschow, 2023b (dark gray) and Brune et al., 2018 (light gray). The 2^m approximation provides a rough approximation, but underestimates the spatial scale for large k. This can be improved through exponential fitting. However, this is not surprising since the parameters of the exponential distribution were estimated based on the illustrated curve. Both estimators provide suitable results.

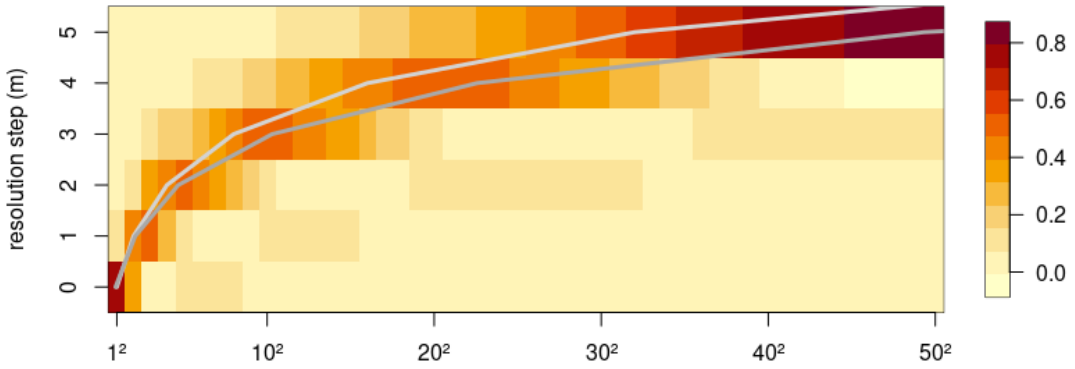


Figure 3.9.: Normalized LWS (equation 3.40) from a 512×512 field containing zeros, where a $k \times k$ domain of ones is included. The x-axis shows k^2, the y-axis $m = 0,\dots,5$. The light gray line shows $k = 2^m$ approximation as proposed in Brune et al., 2018. Darkgray line gives exponential fit between m and maximum k for each m: $k = -0.8\exp(0.78 \cdot m)$.

3.3. Data and preprocessing

This study investigates hourly precipitation from regional reanalysis data, as gridded data is a fundamental requirement for the DT ℂWT. At the time of writing, there are two European reanalyses available. Copernicus European Regional ReAnalyses (CERRA) (Ridal et al., 2024) and COSMO-REA6 (Bollmeyer et al., 2015). Both datasets provide hourly precipitation data and are open accessible for download (CERRA: Copernicus Data Storage[1]; COSMO-REA6: Climate Data Center of german weather service (DWD[2])). This study uses both datasets to assess the uncertainty from the choice of reanalysis. Observation data is used to compare and quantify the results for hourly precipitation totals. Therefore, data from DWD weather stations is used, which is also open accessible for download[3].

The CERRA dataset (Schimanke et al., 2021) provides consistent historical reconstructions for Europe and is available from 1984 to 2021 at the time of writing, with this period being updated regularly. The boundary conditions come from ERA5 global reanalysis data, and the core of the reanalysis is the hydrostatic numerical weather prediction and data assimilation system HARMONIE-ALADIN. It has a horizontal resolution of 5.5 km on a Lambert conformal conical grid and roughly covers the CORDEX domain for Europe (EURO-11)[4]. The model domain is shown in Fig. 3.12 (right). Hourly precipitation data are available in CERRA with lead times/forecast steps between one and six hours, and this study uses the first three hours.

COSMO-REA6 was developed within the Hans-Ertel Centre for Weather Research (HErZ) at the University of Bonn. It is based on the non-hydrostatic numerical weather prediction model COSMO with boundary conditions from ERA-Interim, which was discontinued after 2019. The COSMO-REA6 data is therefore available from 1995 to end of Summer 2019. The model domain meets the specifications of CORDEX EURO-11 but exceeds them by a finer spatial resolution of 6 km (compared to 12 km) (see left part of Fig. 3.12; Bollmeyer et al., 2015).

As CERRA and COSMO-REA6 are operating on different grids, pre-processing is essential to

[1] `https://cds.climate.copernicus.eu` (last accessed: 04.07.2024)
[2] `https://opendata.dwd.de/climate_environment/REA/` (last accessed: 04.07.2024)
[3] `https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/hourly/` (last accessed: 04.07.2024)
[4] `https://euro-cordex.net` (last accessed: 04.07.2024)

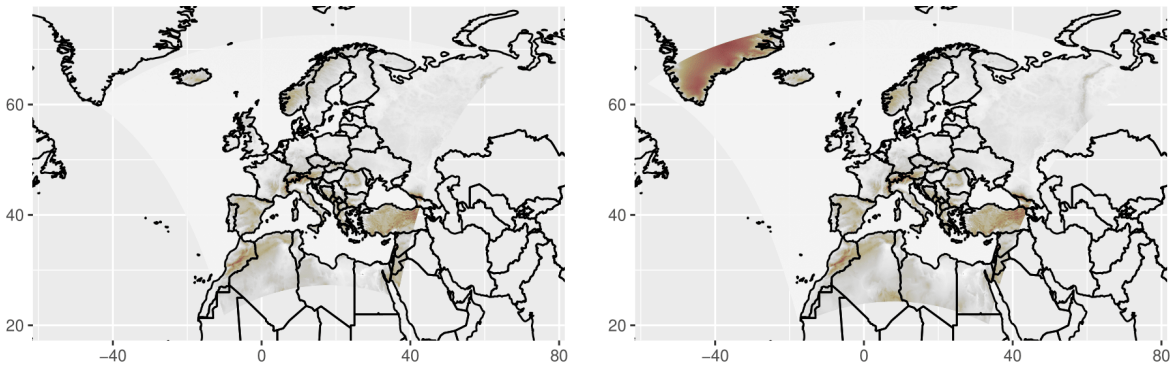
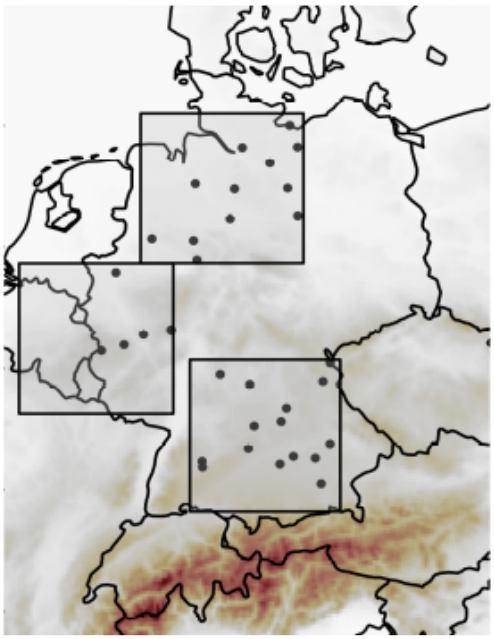


Figure 3.10.: Original mapping from COSMO-REA6 (left) and CERRA (right). The color shows the geographical height.



	Latitude	Longitude
SG	47.00 – 49.45	8.57 – 11.80
NG	51.00 – 53.45	7.48 – 10.80
WG	48.55 – 51.00	5.00 – 8.25

Figure 3.11.: Left: Graphical illustration of the three designated regions over Germany (SG, NG, WG). The color indicates geographical height. Right: Geographical coordinates.

ensure consistent results. To subsequently interpolate the rotated coordinates from COSMO-REA6 to Lambert coordinates as provided by CERRA, the nearest-neighbor approach implemented in the Climate Data Operator (CDO; Schulzweida, 2023) is applied. Mass-preserving interpolation methods are usually preferred for interpolating precipitation amounts. In this particular application, however, the original grids are already quite close, so the much simpler nearest-neighbor approach is sufficient. This study is limited to the period from 1995 to Summer 2019, which is covered by both CERRA and COSMO-REA6 data. A further focus is set by the selection of three areas in Germany, which will be examined in more detail. This enables a detailed and computationally less intensive evaluation. The selected areas include parts of SG, WG, and NG as illustrated in Fig. 3.11. The observation data has also been adapted to the time period and the regions. A total of 30 weather stations that provide a continuous time series of hourly precipitation in the selected regions were identified (black dots).

3.3.1. Wavelet transform & boundary treatment

In wavelet analysis, careful consideration of the boundary conditions is crucial to avoid spurious signals and aliasing in the resulting image. Our spatial wavelet transformation uses 400×400 grid points with the center over Germany as illustrated in Fig. 3.12). The MRA algorithm, which is implemented in the R-package *dualtrees* (Buschow, 2023a) requires that the number of grid points n is a power of two such that $n \sim 2^k$ with $k \in \mathbb{Z}$.

The predefined area 400×400 is thus expanded to a grid size of 512×512 by filling it with zeros. To subsequently eliminate edges at the borders, the outer 50 grid points are linearly smoothed to zero. After estimating the LWS at each of the 512×512 grid points, the predefined areas SG, NG, and WG are extracted, each containing 50 grid points. These areas are far

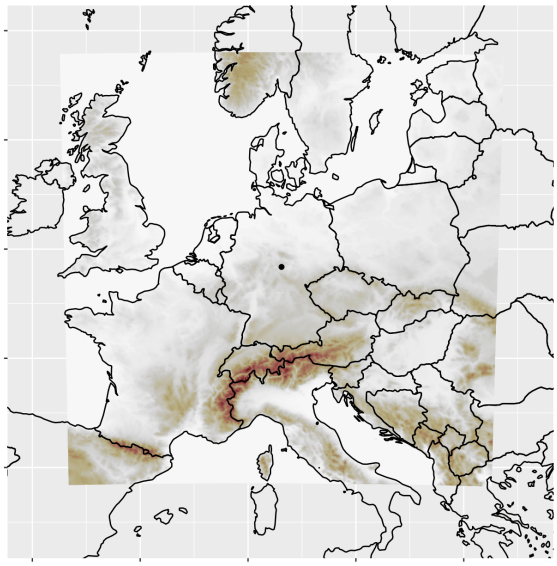


Figure 3.12.: Selected area for wavelet transform: a 400×400 grid point region centered over Germany (black dot). Geographical height is represented by color.

enough away from the border not to be affected by boundary issues.

Following section 3.2.3, Tab. 3.1 finally summarizes the equivalent spatial scale in km for the dual-tree wavelets with resolution steps $m = 0, \ldots, 5$. We use the logarithmic approximation $k = 2^m$ as suggested in Buschow, 2023b.

Table 3.1.: Approximate resolution of wavelet spectra for resolution steps $0, \ldots, 5$ as derived in section 3.2.3.

m	0	1	2	3	4	5	6	7
nx/ny	1	2	4	8	16	32	64	128
resolution [km]	5.5	11	22	44	88	176	352	704

3.3.2. Analysis of hourly precipitation totals

We first analyze hourly precipitation totals from CERRA and COSMO-REA6 datasets within the regions NG, SG, and WG. To avoid spurious signals, values below 0.01mm are set to 0.

Figure 3.13 shows kernel density estimates from the time series of the spatial mean (left) and maximum (right) precipitation totals. For CERRA and COSMO-REA6, we analyze data for each region separately. Due to the limited number of stations available, we aggregate all 30 weather stations for observational data. For the spatial mean, we find a general agreement between CERRA and COSMO-REA6 but discrepancies compared to the observational data. We ascribe this to uncertainties due to the limited number of stations considered. For the spatial maximum, we find clear discrepancies between COSMO-REA6 and CERRA, indicating that COSMO-REA6 contains systematically higher precipitation totals. This is also reflected in the maximum value of both models (COSMO-REA6: 91mm; CERRA: 49mm). The observational data lies between the two models. However, we expect that they will not be able to adequately

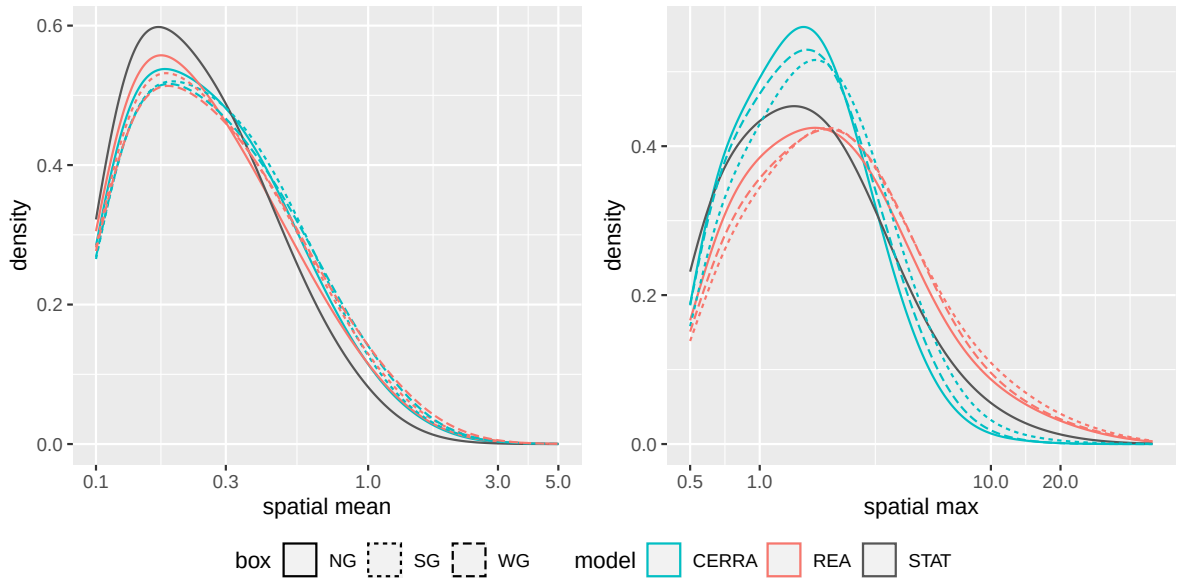


Figure 3.13.: Kernel density estimates of aggregated precipitation totals. Left: time series of spatial mean; Right: time series of spatial maximum. The color indicates the respective data set (black: observations, blue: CERRA, red: REA6). The line type indicates the area (SG, NG, WG). For observational data, the aggregation is done over all stations.

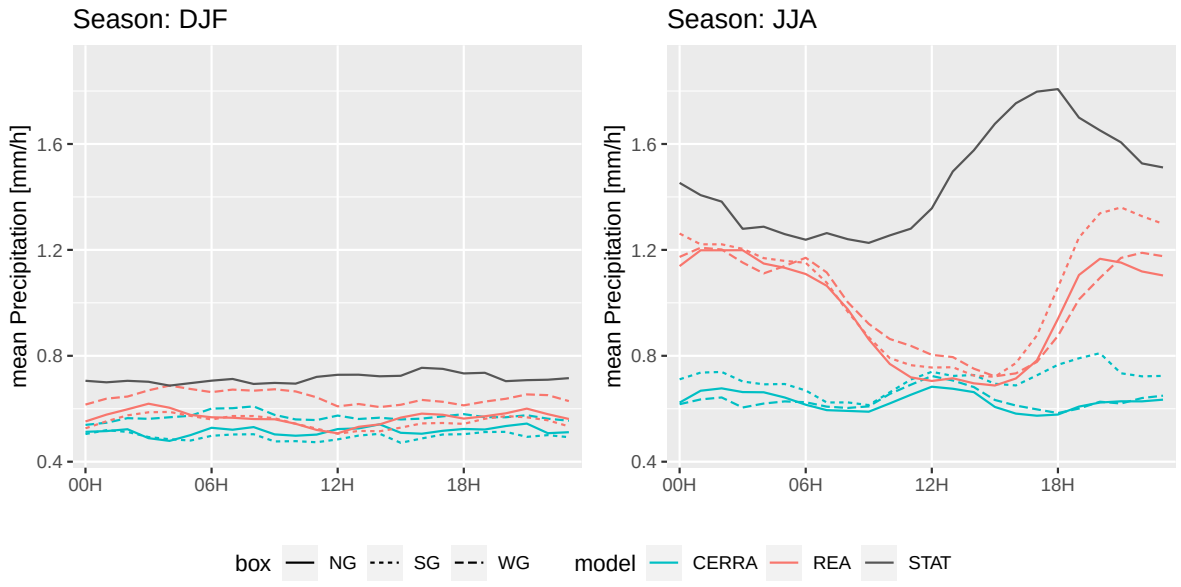


Figure 3.14.: Mean diurnal cycle of hourly precipitation totals [mm] for winter months (DJF, left) and summer months (JJA, right). The calculation is done on values exceeding 0.1 mm/h. The color indicates the respective data set (black: observations, blue: CERRA, red: REA6). The line type indicates the area (SG, NG, WG). For observational data, the calculation is done over all stations

map precipitation maxima due to the low spatial coverage of the weather stations used. In Appendix Fig. B.2, we analyze the density plots for each season. The highest discrepancies are in the northern hemispheric summer (JJA) months. We find only minor differences in the northern hemispheric winter (DJF) months.

As detailed in section 3.1, small-scale convective precipitation events are dominant during the JJA months in the mid-latitudes. They tend to occur in the evening, with a typical peak around 18 UTC over Germany, which is mainly attributed to ground warming during the day. Figure 3.14 shows the mean diurnal variation of hourly precipitation in the months JJA and DJF. The calculation is done for values > 0.1 mm. These are then averaged by area, time of day and season.

The diurnal cycle obtained from COSMO-REA6 is comparable to the results from previous studies as, e.g., Bollmeyer et al., 2015; Wahl et al., 2017. Minor differences result from our focus on limited areas. As, e.g., Bollmeyer et al., 2015 discuss, COSMO-REA6 gives a good representation of the diurnal cycle from observation data (black line), but with a time delay of around three hours. As a result, the evening peak occurs at 21 UTC instead of 18 UTC. This is likely due to a delayed initiation of convective precipitation in the COSMO model and can be improved by assimilating precipitation data. This has been done in the COSMO-REA2 dataset, which significantly improved the representation of diurnal circle (Bollmeyer et al., 2015; Wahl et al., 2017).

With CERRA, the diurnal circle is less evident, and the precipitation intensities are significantly lower than for COSMO-REA6 data. This is especially true for the JJA months when the diurnal circle is especially characterized by convective precipitation events. The CERRA dataset is relatively new, and, to the best of our knowledge, no publications have explained this result to date. The HARMONIE-ALADIN system is based on the hydrostatic approximation, which provides a possible explanation. In contrast, the COSMO model does not use this approximation, which also gives a better representation of convective precipitation events from the theoretical perspective.

3.3.3. Analysis of wavelet representation

To evaluate the spectral representation of precipitation events, we utilize the LWS $|d^m(x,y)|^2$ as derived in section 3.2.3. $m = 0,\ldots,5$ is the resolution step that has an equivalent spatial scale in km as given in Tab. 3.1.

We start our analysis of precipitation events in the wavelet space with an illustrative example of a heavy precipitation event on 29 June 2017. This event was particularly pronounced over the Berlin metropolitan region and was classified as exceptionally severe (see, e.g. Caldas-Alvarez et al., 2022). As Caldas-Alvarez et al., 2022 shows, over $11,000$ convective cells were triggered in connection with this event. The triggering mechanism was a convergence line driven by counter-circulations over the Atlantic and Central Europe. Although the event was particularly pronounced over Berlin, we find accompanying precipitation events in SG, which began in the early morning of 29 June and slowly moved northwards. Figure 3.15 shows 1) the accumulated precipitation totals at 17-06-29 (right) and 2) hourly precipitation at 17-06-29 02:00 UTC from radar-based RADOLAN-RW dataset (Winterrath et al., 2012).

Figure 3.16 illustrates hourly precipitation amount at 02:00 UTC from CERRA and COSMO-REA6 for SG. Both datasets roughly agree on mean and maximum precipitation amounts (mean: 1.98 max: 18.12 for CERRA; mean: 1.91 max: 22.17 for COSMO-REA6) and also agree on the precipitation area between Nuremberg, Stuttgart, and Munich. The spatial patterns

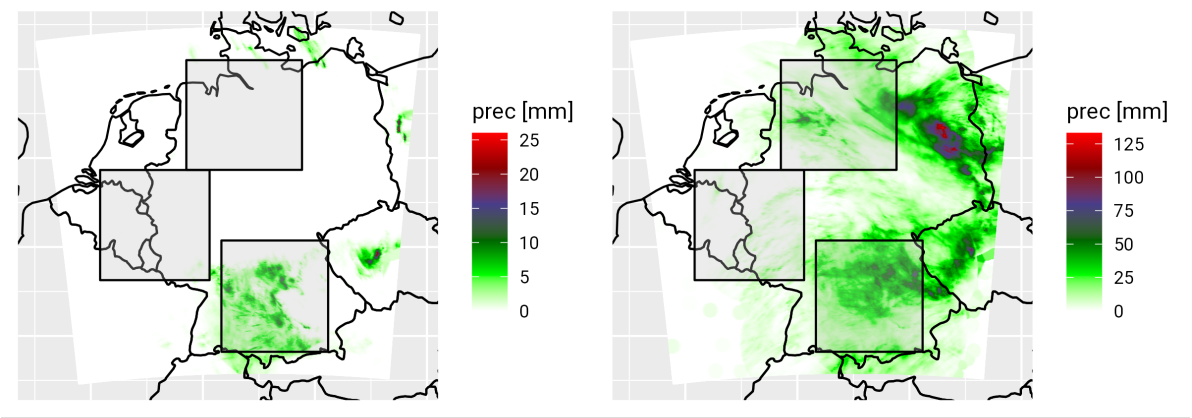


Figure 3.15.: Left: Precipitation totals on 17-06-29 02:00 UTC; Right: daily precipitation totals on 17-06-29, given by RADOLAN-RW product from DWD.

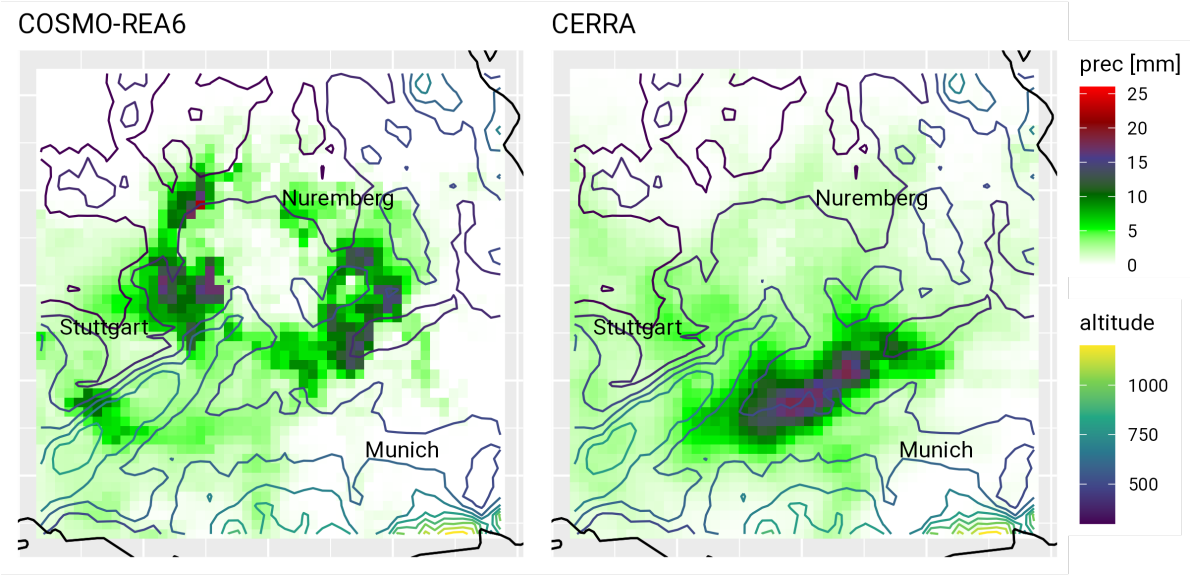


Figure 3.16.: Hourly precipitation totals at 2017-06-29 02:00 UTC in Southern Germany. The raw data from CERRA (left) and COSMO-REA6 (right) are shown.

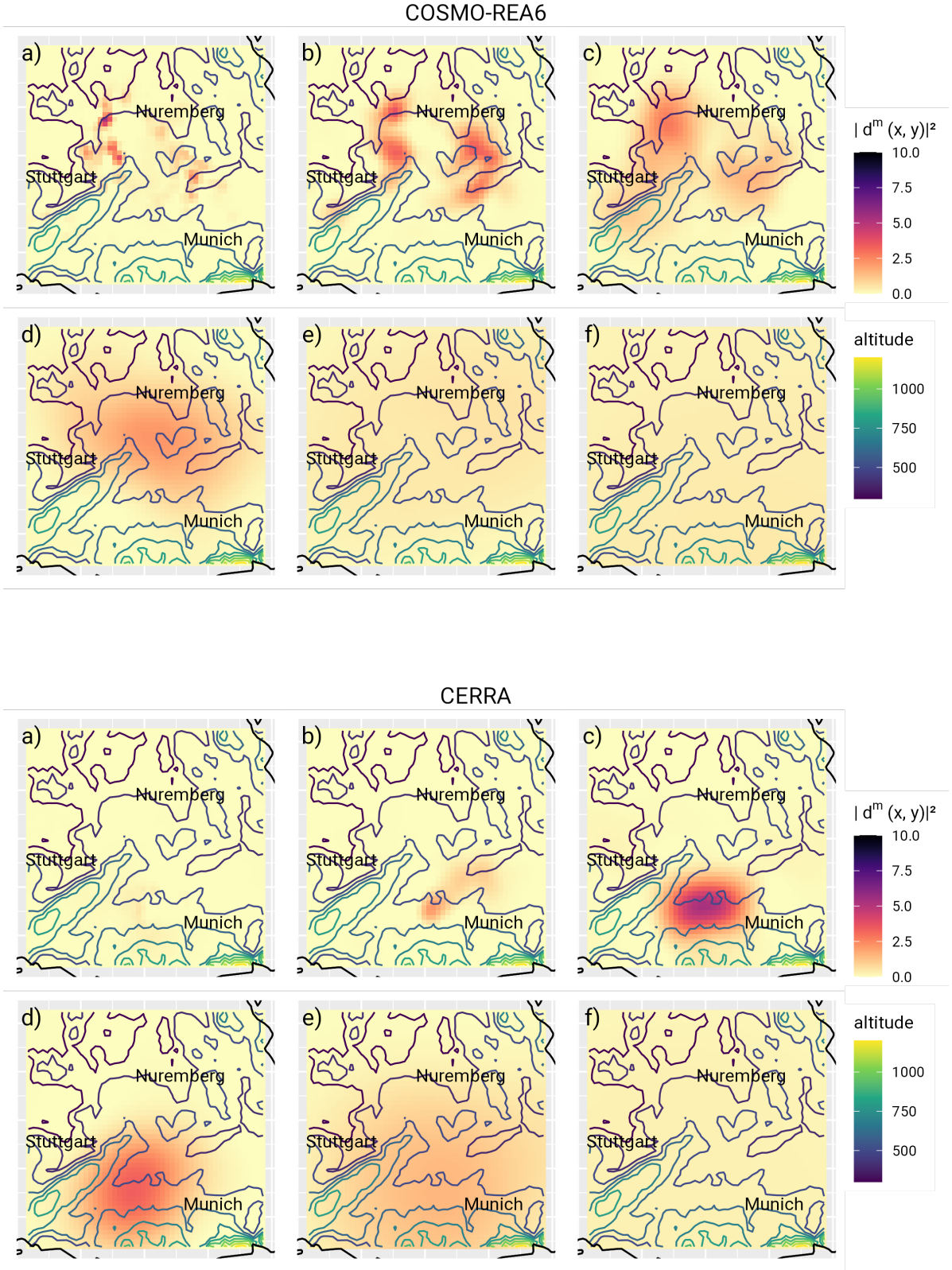


Figure 3.17.: Localized wavelet spectrum $|d^m(x,y)|^2$ for $m = 0,\ldots,5$ (a), …, f)) on 2007-06-15 17:00 UTC in SG from COSMO-REA6 (top) and CERRA (bottom).

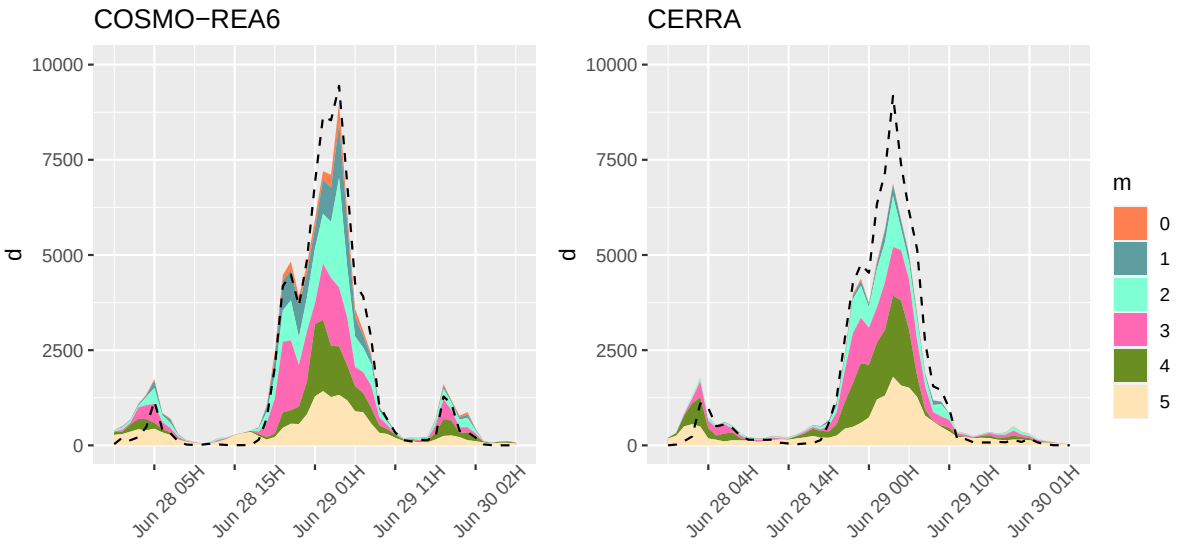


Figure 3.18.: Total wavelet energy $d = \sum_{m=0}^{5}(d^m)$ from June-28 00:00 UTC to June-30 03:00 UTC. Colors indicate $d^m = \sum_{x,y}(|d^m(x,y)|^2)$ for resolution steps $m = 0,\dots 5$. The black line shows the total squared precipitation.

differ. In comparison, the precipitation pattern in CERRA is more extensive, while in COSMO-REA6, it is characterized by individual cells.

Figure 3.17 shows the LWS for resolution steps $m = 0,\dots,5$. The resolution steps $m = (0,1)$ (approximately 5.5 km and 11 km) show two separate precipitation cells within COSMO-REA6. In CERRA, the smaller-scale events are less evident, no such separated cells are visible, and the peak values are at $m = 2,3$ (about 22 km and 44 km). $m = 5.6$ are less pronounced in both models. We conclude that the wavelet method provides meaningful decomposition of precipitation patterns. It reflects what would be visually readable and, therefore, provides meaningful decomposition into patterns of different spatial scales.

For a more detailed analysis, we define the total wavelet energy for each m as $d^m = \sum_{x,y}|d^m(x,y)|^2$ and the total energy amount as $d = \sum_m d^m$. As detailed in section 3.2.3, the ideal scenario imposes that d and squared precipitation are identical at each timestep (see Parseval's Energy theorem; equation 3.41). Our application, however, introduces several uncertainties. These include the utilization of redundant wavelet transform and the restriction to smaller study areas after applying the wavelet transform. We also do not consider resolution steps $m > 5$. Figure 3.18 shows the time series of total squared precipitation from 28 June to 30 June (dotted line). The contribution of each resolution step m to the total wavelet energy d (d^m) is indicated by color. Despite the differences in $m = 0,1$ identified before, total precipitation and total wavelet energy d^m are roughly identical between CERRA and COSMO-REA6. For COSMO-REA6, squared precipitation and d mostly coincide. In CERRA, missing the energy content from the smallest scales, we observe higher differences. This could be because the precipitation event is larger in CERRA. Therefore, the energies recorded by $m = 1,\dots,5$ may not capture all the amounts resolved in the data. We further observe that d^m at larger m tends to overshadow smaller m, despite their relatively low expression in the spatial field (Fig. 3.18). This is primarily due to the larger spatial extent of higher m, which is more evident

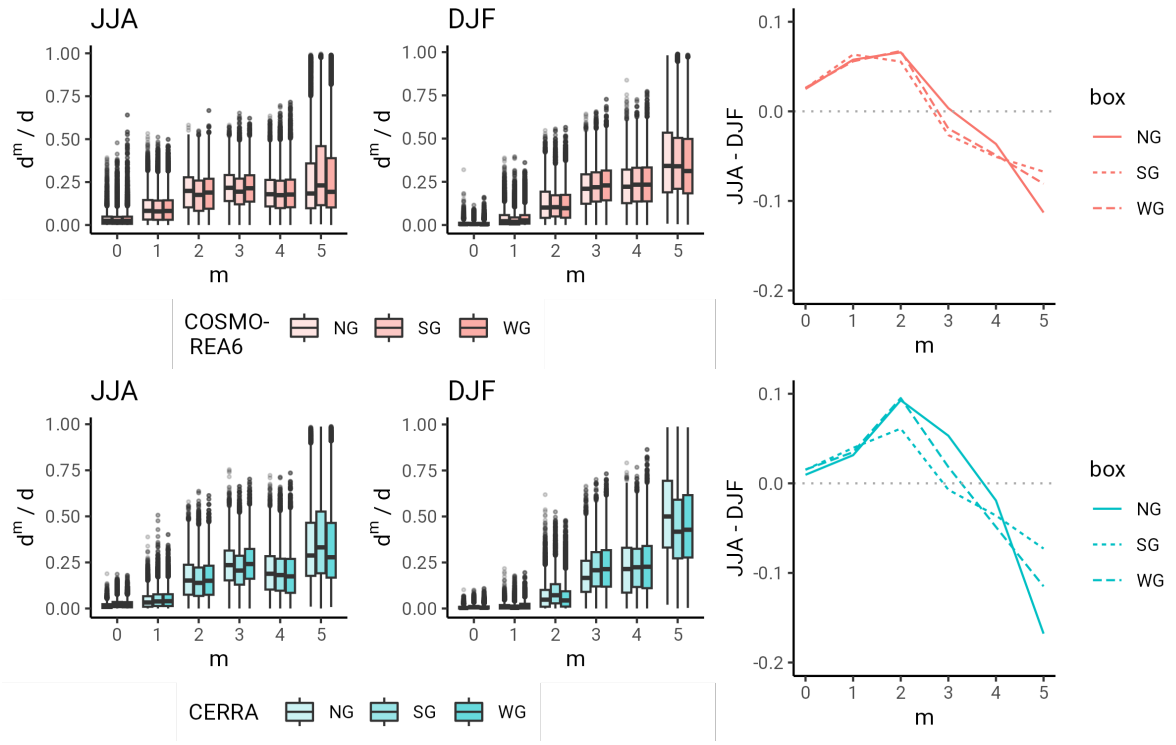


Figure 3.19.: Left: boxplot representation from hourly d^m/d values for JJA and DJF months Right: Difference in mean d^m/d between JJA and DJF. Linetype/shading represents the respective area (SG, NG, WG). The top row is COSMO-REA6, and the bottom row is CERRA. The calculation is done on timesteps where d exceeds the 0.8-quantile.

in the aggregation. Similar investigations to RDWT typically only consider grid points where precipitation is actually present (see, e.g. Buschow and Friederichs, 2020; Brune et al., 2020). However, as we analyze the time series of LWS at each grid point in the following section, this approach is quite suitable for our study.

We conclude this section with an investigation of seasonality in LWS. Figure 3.19 shows a boxplot representation of hourly time series of normalized wavelet energies d^m/d for JJA and DJF months by area and dataset (Top: COSMO-REA6; Bottom: CERRA). Only timesteps where d exceeds the 0.8-quantile are considered. The plot on the right shows the difference in mean d^m/d between JJA and DJF. The fraction d^m/d of the total wavelet energy with resolution step $m = 0, 1, 2$ is larger during JJA than DJF, whereas total wavelet energy is more pronounced on $m = 3, 4, 5$ DJF months. This indicates that JJA precipitation generally has a larger fraction of the total wavelet energy on smaller scales than DJF months. There are regional differences, such as a tendency in NG towards larger scales in the JJA months, which is consistent in both CERRA and COSMO-REA6.

3.4. Changes in the Localized Wavelet Spectrum

The non-stationarity of extreme weather events may include three aspects: changes over time in their intensity, frequency, and timing (see, e.g., Slater et al., 2021). This study utilizes the LWS to investigate changes in 1) their intensity and 2) their timing, i.e., their typical occurrence days throughout the year.

In appendix section B.3, we show results from similar trend analysis with hourly precipitation data and find isolated trends. In COSMO-REA6, we find changes in WG towards more intense events during JJA months and a decrease in the variability of precipitation extremes throughout the year. In CERRA, we find evidence for increased intensity in SG in DJF months.

We here analyze the LWS of hourly precipitation fields at resolution steps $m = 0,\dots,5$ separately to determine the suitability of the wavelet decomposition for two purposes: 1) assessing whether the trends observed in the precipitation data can be attributed to specific spatial scales, and 2) analysing whether the separation into different spatial scales reveals additional changes compared to hourly precipitation data.

3.4.1. Changes in Intensity

Various studies have investigated changes in the intensity of extreme precipitation. We follow Zeder and Fischer, 2020; Westra et al., 2013 who find trends to more intense precipitation extremes over Germany in DJF daily observation data. We first derive seasonal maxima of LWS for each grid point and resolution step m. We then fit a GEV distribution and apply a log-ratio test to access evidence for a non-stationary location parameter as, e.g., suggested by Coles, 2001. The log-ratio test evaluates whether one of two maximum-likelihood estimates offers a better fit. Specifically, we test a stationary location parameter against a linear, time-dependent location parameter:

$$\begin{aligned} H_0: &\quad \mu = \mu_0 \\ H_1: &\quad \mu = \mu_0 + t\mu_1 \end{aligned}$$

Where t is the year $1995,\dots,2018$. To quantify changes, we estimate the scaling estimate $\beta = \mu_1/\mu_0$ at each location and resolution step $m = 0,\dots,5$.

Although we originally performed the significance testing for all seasons, the remainder of this section is limited to the description of JJA and DJF, as these are the seasons where we find consistent and significant changes. In Fig. 3.20, we exemplary show the spatial β field for SG in JJA months. The dots denote grid points exceeding the significance level of $\alpha = 0.025$. For the remaining β-fields for NG and WG, as well as JJA and DJF months, here we refer to appendix Section B.2.2.

For the JJA months, we generally observe positive trends in SG. In NG, the pattern is more scattered, while in WG, we find consistent positive trends in the Rhineland, parts of the Pfalz, Hocheifel, and Ahr Valley. Significant trends are mostly on small scales ($m = 0,1,2$). This is consistent in COSMO-REA6 and CERRA, except for WG, where COSMO-REA6 produces significant patterns across all resolution steps.

For the DJF months, we find a tendency toward negative trends on small scales ($m = 0,1,2$) and positive trends on large scales ($m = 4,5$) across all regions. Significant trends in SG are mostly on large scales ($m = 4,5$). In NG and WG, significant trends are mainly evident on small scales ($m = 0,1,2$). Again, COSMO-REA6 and CERRA produce roughly consistent results.

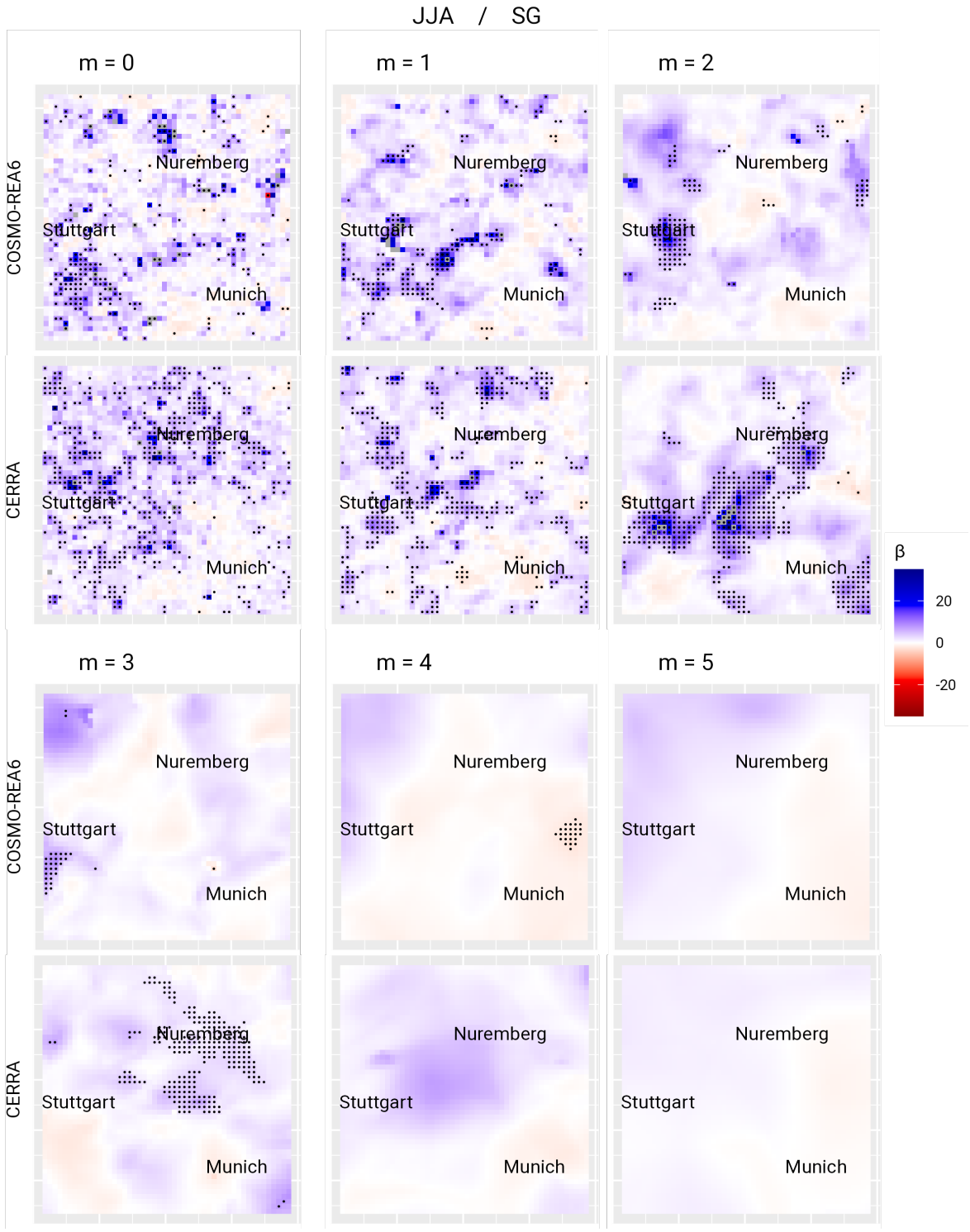


Figure 3.20.: Spatial fields of $\beta = \mu_1/\mu_0$ derived from GEV-fit with a non-stationary location parameter $\mu = \mu_0 + \mu_1 t$, estimated from seasonal LWS maxima in JJA months in SG. Shown are resolution steps $m = 0,\ldots,5$ for COSMO-REA6 and CERRA. Black dots indicate gridpoints where a log-ratio test against a stationary location parameter $\mu = \mu_0$ exceeds the 0.025 significance level.

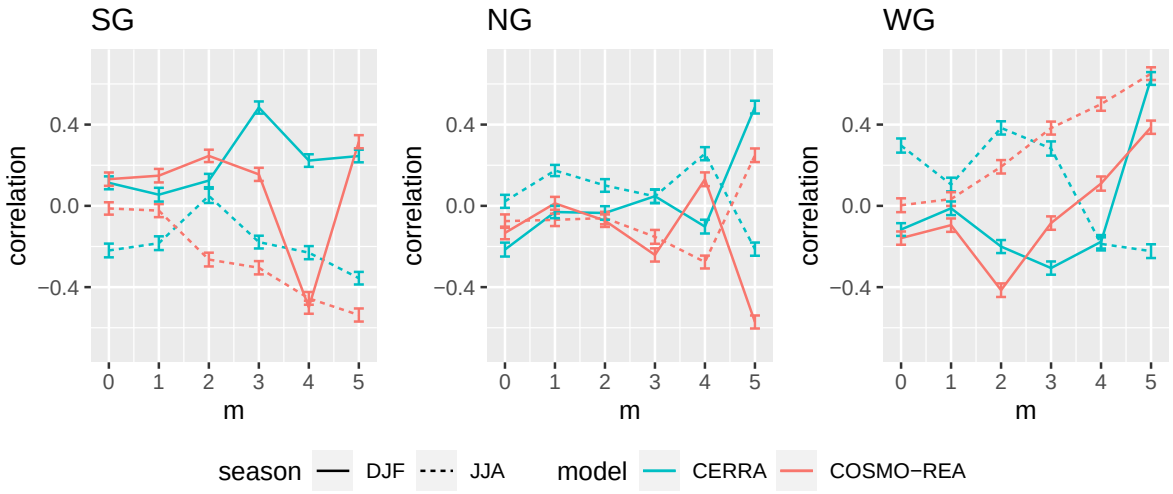


Figure 3.21.: Correlation between scaling estimates β and topographical height. Error bars indicate the 95% confidence interval from 1000 bootstrap-samples where spatial correlation is destroid.

Figure 3.21 shows the spatial correlation between the β fields and topographical height obtained from COSMO-REA6 and CERRA data. In SG, we observe a negative correlation during the JJA months and a positive correlation during the DJF months at larger scales ($m = 3,4,5$). Conversely, in NG, the correlation is minimal. This is consistent in COSMO-REA6 and CERRA. In WG, we observe a positive correlation during the JJA months at small scales ($m = 0,1,2,3$) in CERRA and at large scales ($m = 3,4,5$) in COSMO-REA6. During the DJF months, there is a tendency towards negative correlation at small scales and positive correlation at large scales, although less pronounced. The complex topography and associated processes in WG could explain why COSMO-REA6 and CERRA produce such different correlations. WG includes large parts of the Eifel and the Cologne/Bonn area, where the topography strongly influences precipitation patterns. These include, e.g., lee-effects, which have been shown to affect regional climatology west of the Eifel (Szemkus, 2016). COSMO can, in principle, resolve these effects (Werner, 2014), we are not aware of any studies for CERRA.

We now investigate whether significant trends are present in the spatial β fields. Techniques from field-significance provide evidence regarding whether trends observed at specific grid points are representative of the broader field (Zeder and Fischer, 2020; Benjamini and Hochberg, 1995). For example, the False Detection Rate (FDR) counts the number of falsely rejected null hypotheses in the entire field, which is used to estimate the field significance (Wilks, 2006a). If the β fields are correlated, the FDR estimator must be adjusted (Benjamini and Heller, 2007). We do so by generating bootstrap samples of the β field, where the temporal relation is destroyed. Field significance is declared when the number of grid points with a significance level α exceeds the $1-\alpha$ quantile of estimated FDR. Computing the appropriate number of bootstrap estimates for 50×50 fields is computationally intensive, as estimating a single field requires several hours on a standard computer, even with parallel computing. Our estimate of the significance level $\alpha = 0.025$ relies on a total of 600 bootstrap samples for each resolution step. Since LWS lives on well-defined spatial scales, it is reasonable to assume that the probability distribution of FDR does not depend on parameters such as season or area, which saves computation time. We obtain 20×20 random samples rather than estimating the

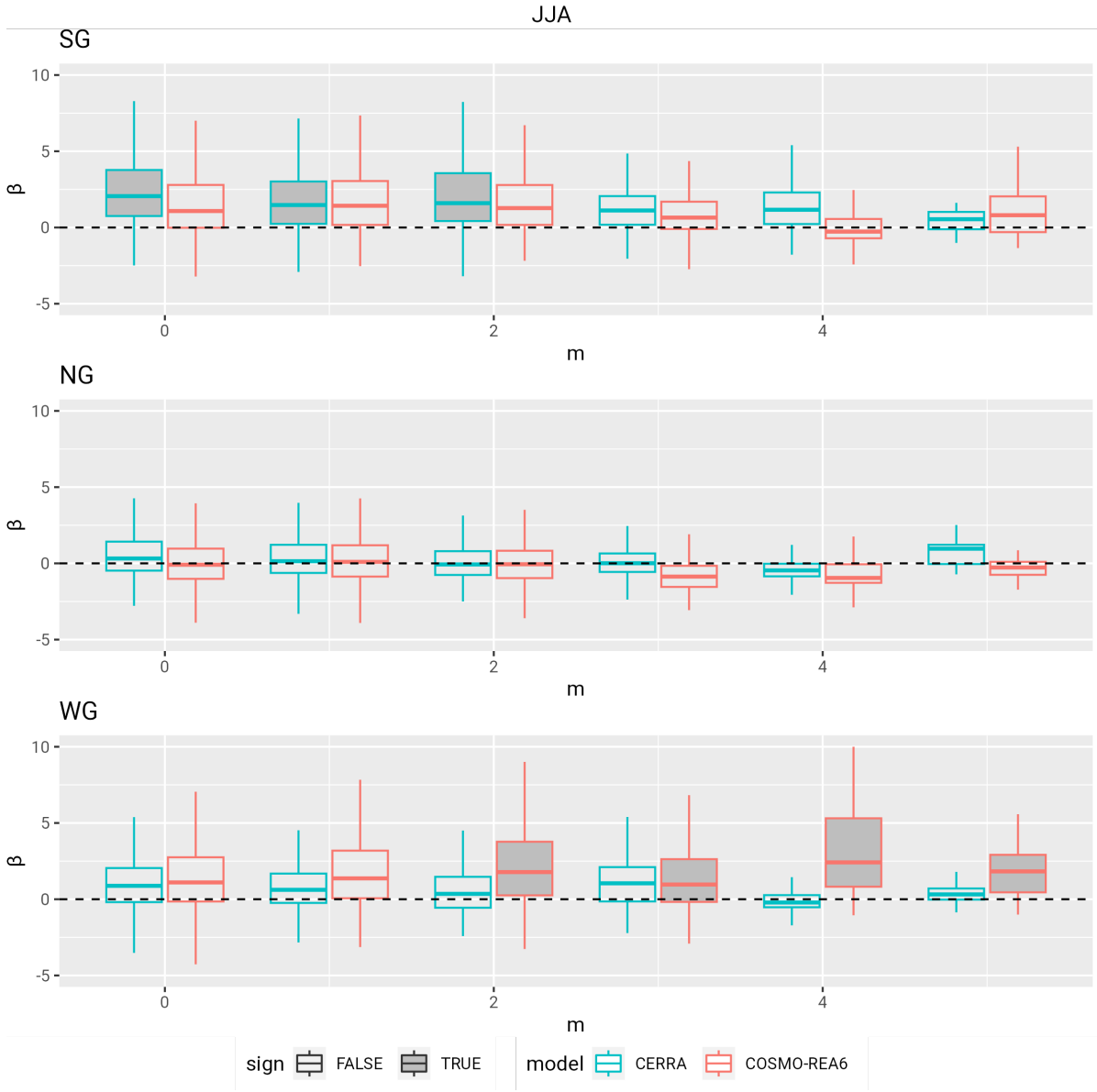


Figure 3.22.: Boxplot representation of scaling estimate β as shown in Fig. 3.20 and appendix figures B.4, B.6 for JJA months. The boxplots are seperated by area, season, and resolution step m for CERRA (blue) and COSMO-REA6 (red). Areas exceeding the significance level of 0.025 for a specific resolution step are marked gray.

entire 50×50 grid.

Figures 3.22 and 3.23 show boxplots of the β fields for JJA and DJF months. Areas that exceed field significance at $\alpha = 0.025$ are shaded in gray. The corresponding mean values of β are provided in Table 3.2. For the JJA months, we find significant increasing trends in SG on small scales in CERRA. COSMO-REA6 shows similar trends but does not provide significance. In NG, no trends are apparent. In WG, CERRA exhibits the same structure as in SG, with positive trends at small scales, although these are not significant. In contrast, COSMO-REA6 shows increasing trends across all scales, exceeding the significance level for $m = 2,3,4,5$. For

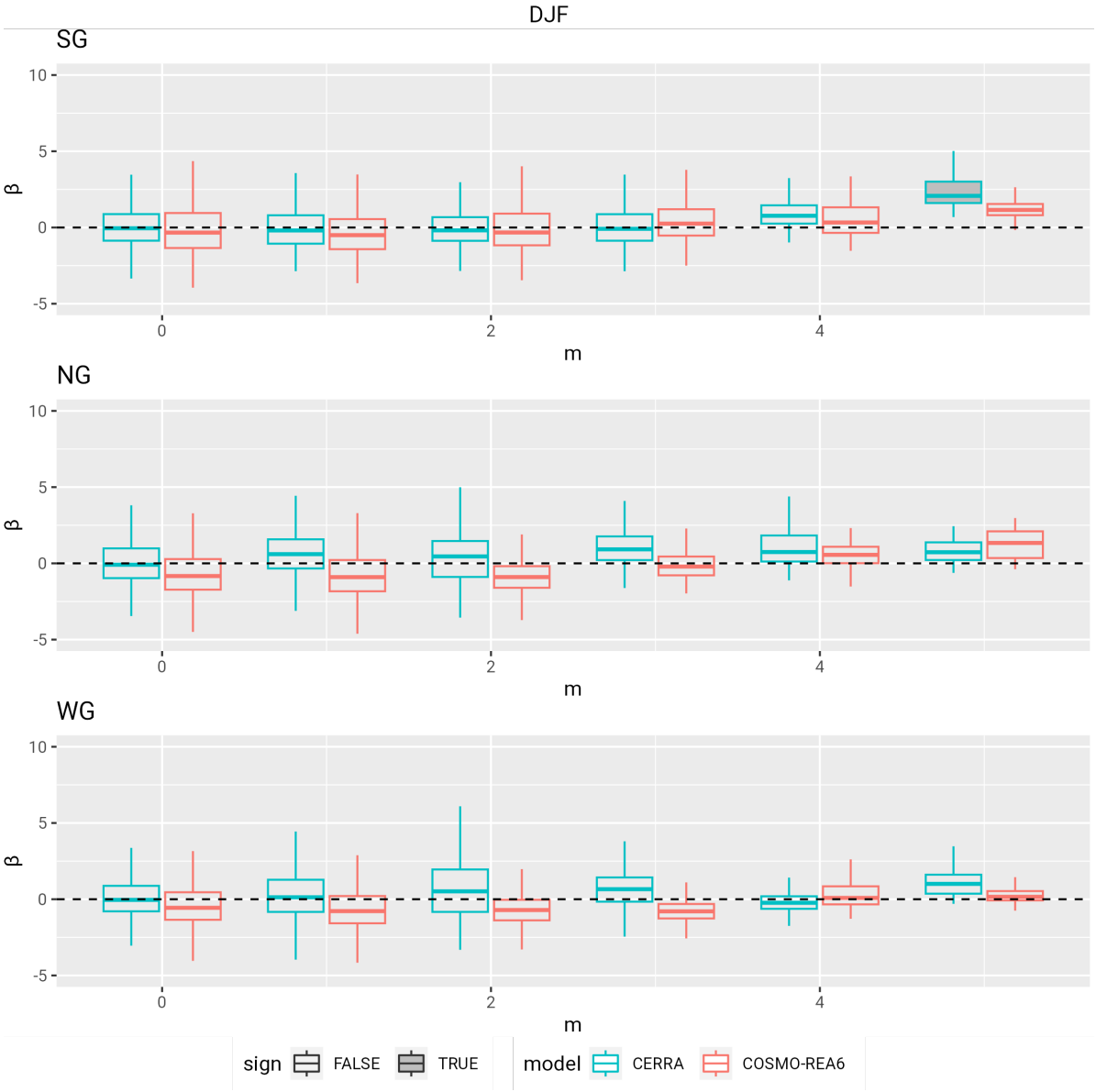


Figure 3.23.: Boxplot representation of scaling estimate β as shown in Figs. B.3, B.5 and B.7 for DJF months. The boxplots are seperated by area, season, and resolution step m for CERRA (blue) and COSMO-REA6 (red). Areas exceeding the significance level of 0.025 for a specific resolution step are marked gray.

the DJF months, we find a tendency towards negative changes at small scales ($m = 0, 1, 2$) in COSMO-REA6. For large scales ($m = 4.5$), there are increasing tendencies in both COSMO-REA6 and CERRA, which is consistent across all regions. CERRA shows significance in SG on the largest scale ($m = 5$).

Table 3.2.: Mean values of scaling estimate β as shown in Figs. 3.22 and 3.23 for each area, resolution step, model, and season (JJA, DJF). Shown are the values for COSMO-REA6/CERRA. Values exceeding the significance level $\alpha = 0.025$ are marked with a star.

	JJA			DJF		
	SG	NG	WG	SG	NG	WG
m=0	2.6/2.9*	0.3/0.7	0.0/1.2	0.0/0.1	-0.5/0.2	-0.3/0.2
m=1	2.6/2.1*	0.4/0.5	2.3/1.1	-0.2/0.0	-0.6/0.7	-0.5/0.4
m=2	1.9/2.7*	-0.0/0.2	4.4*/1.1	0.0/0.0	-0.8/0.5	-0.6/1.6
m=3	1.1/1.2	-0.8/0.2	3.3*/1.1	0.4/0.1	-0.0/1.1	-0.7/0.7
m=4	0.1/1.5	-0.6/-0.4	3.4*/0.0	0.4/0.9	0.5/1.1	0.3/-0.2
m=5	1.0/0.4	-0.3/0.7	1.7*/0.0	1.2/3.1*	1.3/0.8	0.2/1.1

3.4.2. Changes in Timing

This section explores changes in the seasonality of LWS. In principle, the same method as in section 3.4.1 could be applied. However, we find the grid point-wise tests to be quite sensitive to threshold values. Added to this is the high computational effort required to compute the grid point-wise tests. Therefore, we use a pooling approach to increase the robustness and reliability of our results. Similar approaches have been used to analyze changes in precipitation intensity, e.g., in Tradowsky et al., 2023; Vautard et al., 2015.

For each year, grid point and resolution step m, we identify the day with the largest value in the LWS. The values above each grid point are summarised into a sample, whereby we decluster by removing duplicate events that occurred within 48 hours. In doing so, we identify

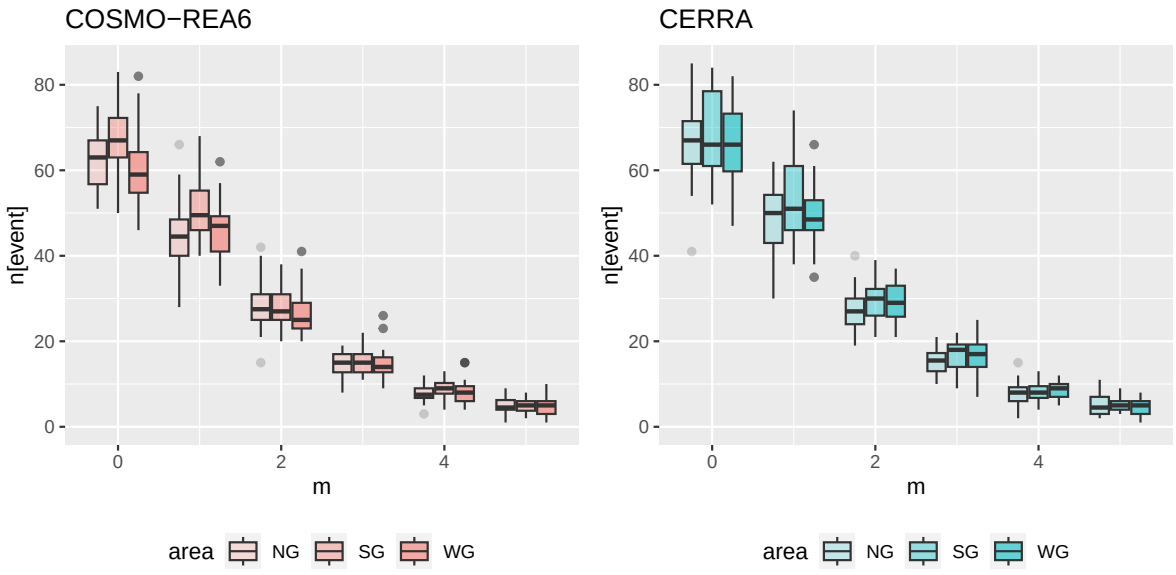


Figure 3.24.: Boxplot representation of the number of events identified as the most extreme event of the year at least one grid point of the respective area, model, and spatial resolution. Shown are results from the period 1995-2018 for COSMO-REA6 (left) and CERRA (right). The shading indicates the respective area (SG,NG,WG)

events recognized as the strongest of the year in at least one grid point. This is done for SG, NG, and WG separately, assuming similar seasonal cycles within the areas.

Figure 3.24 shows the number of events per year identified for each resolution step and area. On average, we identify between 50 and 60 events per year at the lowest resolution, whereas at the highest resolution, just about five events per year are identified. This outcome is not unexpected, given the diverse types of events with varying spatial extents identified at each resolution step. A sample size of less than 20 events may be too small for statistical conclusions. Therefore, we increase the sample size by considering not only the highest event at each grid point but also the second highest event and so on until a minimum sample of 20 events is reached. As a downside, we may loosen our focus on extremes.

Figure 3.25 shows histograms of occurrence dates for the highest events identified in COSMO-REA6 and CERRA. All resolution steps have an almost unimodal and symmetrical distribution. Thus, the mean and variance are suitable for describing their characteristics. As previous studies suggest, when analyzing changes in seasonality using precipitation data, we rely on angular measures (see box 2.3) to analyze changes in seasonality as the day of the year is a cyclic variable (see, e.g., Dhakal et al., 2015). However, our estimates, based on circular measurements, are roughly consistent with the standard mean and variance. Most events occur during the warm season, May to September. COSMO-REA6 and CERRA show similar behavior on small scales ($m = 0, 1$). For larger scales, CERRA has a less clear concentration of extremes in the warm season. As a result, the variance for these scales is higher in CERRA. The mean values tend to be higher for larger scales (COSMO-REA6: $m = 1$: 202, $m = 5$: 215; CERRA: $m = 1$: 197, $m = 5$: 224). This implies a tendency for events on large scales to occur later in the year. This is consistent between COSMO-REA6 and CERRA. However, CERRA shows a more pronounced shift (around four weeks from smallest to largest scale) compared to COSMO-REA6 (around 2 weeks from smallest to largest scale).

Box 2.3: Circular Mean & Variance

Die Berechnung des kreisförmigen Mittelwerts und der Varianz erfolgt über die radiale Komponente. The transformation between the day of the year and the radial component is done with:

$$r(d) = (d-1) \times 2 \times \frac{\pi}{365.25} \qquad d(r) = \frac{r \times 365.25}{2 \times \pi} + 1. \tag{3.42}$$

$r(d)$ is the radial component and $d(r)$ is the day of the year. Subsequently, circular mean and variance are calculated as follows:

$$\text{mean}(r) = \text{atan2}\left(\overline{\sin(r)}, \overline{\cos(r)}\right) \tag{3.43}$$

$$\text{var}(r) = 1 - \sqrt{\left(\overline{\sin(r)}\right)^2 + \left(\overline{\cos(r)}\right)^2}. \tag{3.44}$$

$\overline{(\ldots)}$ is the standard mean measure.

In Figs. 3.26 and 3.27, we show the time series of the yearly mean and variance of occurrence days for each scale, area, and model. The mean values show more pronounced variability in CERRA compared to COSMO-REA6, which is particularly evident for larger scales ($m = 4.5$). The overall mean (dotted lines) is comparable between COSMO-REA6 and CERRA, hovering around ~ 200, corresponding to the middle to end of July. The variance in CERRA is gener-

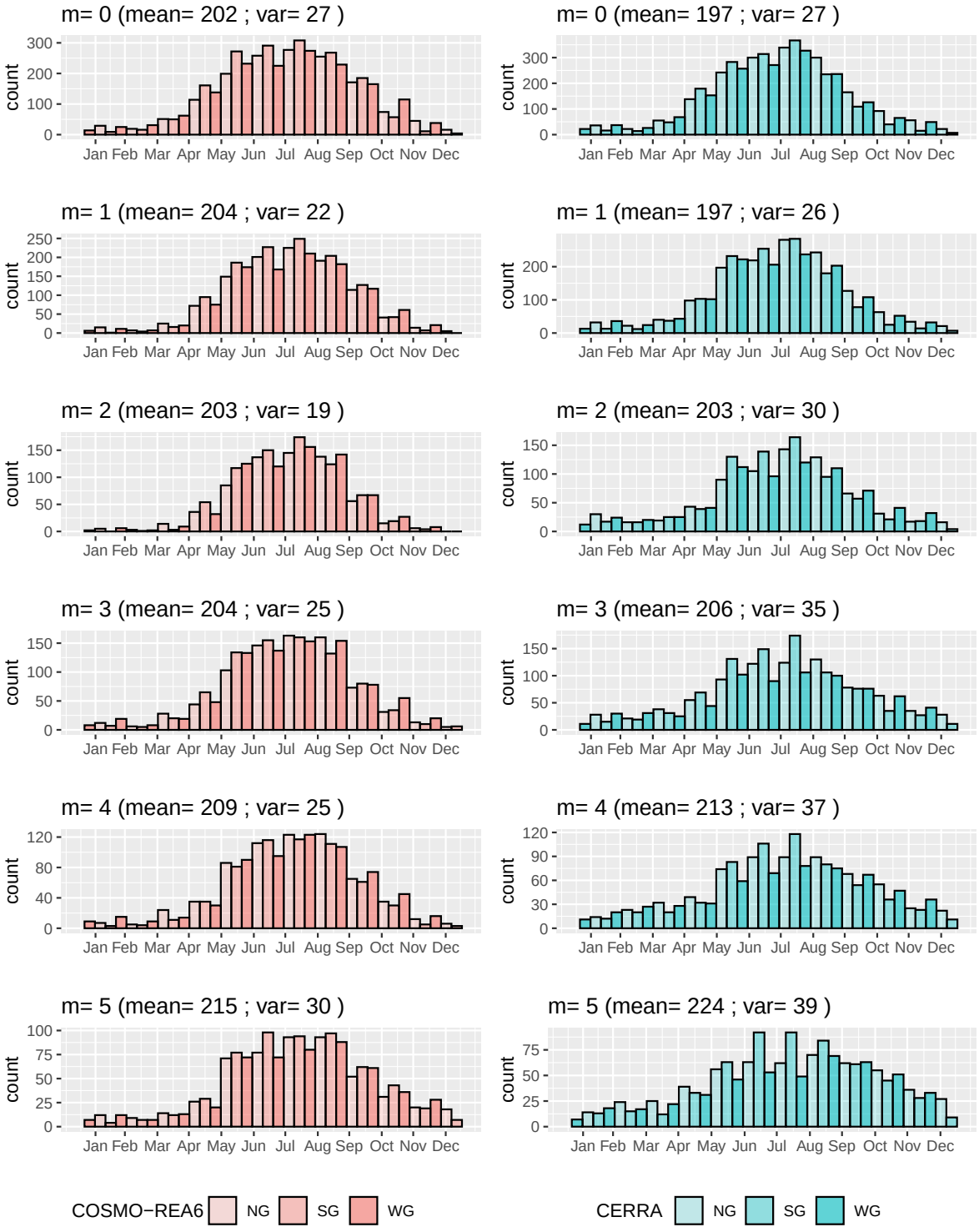


Figure 3.25.: Histogramm of occurrence dates from events identified as the most extreme event of the year at at least one grid point of the respective area (SG,NG,WG), model (left:COSMO-REA6; right:CERRA) and spatial resolution (top: $m = 0$ to bottom $m = 5$). The results for the period 1995-2018 are shown.

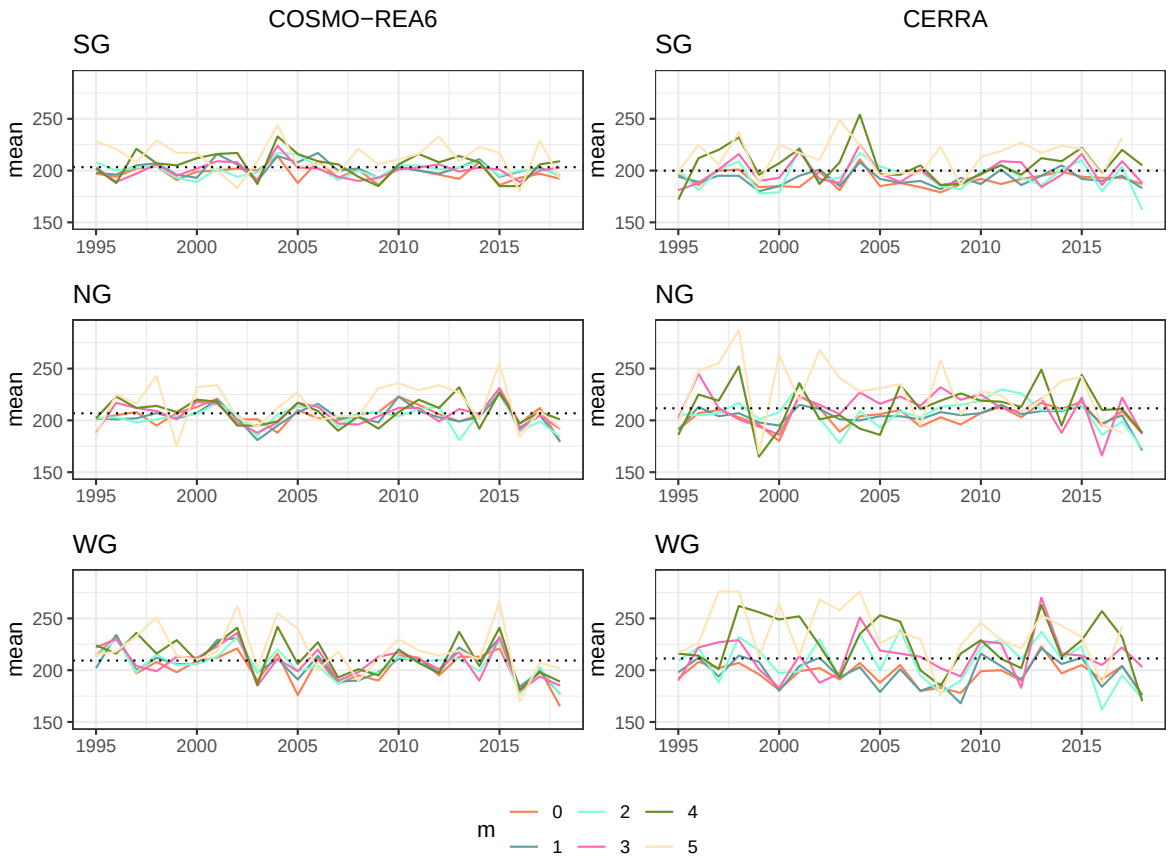


Figure 3.26.: Timeseries of yearly circular mean (see box 2.3) of occurrence dates from events identified as the most extreme event of the year at at least one grid point of the respective area (top to bottom: SG, NG, WG), model (left: COSMO-REA6; right: CERRA) and spatial resolution (color). The results from the period 1995-2018 are shown.

ally higher than in COSMO-REA6, as previously noted in Fig. 3.25. However, both models consistently indicate that SG exhibits the lowest variance. There are slight trends suggesting a decrease in both variance and mean in WG, particularly in COSMO-REA6. The subsequent analysis in this section will determine if these trends are statistically significant at a certain significance level.

We employ the non-parametric Mann-Kendall test (Kendall, 1948; Mann, 1945) to find evidence for non-stationarities within the yearly time series of mean and variance of occurrence days. Separate tests are conducted using the mean values and variances from Figs. 3.26 and 3.27, as well as for the respective regions and resolution steps m. To further quantify the respective change in days/year, we fit a linear trend $m(t) = m_0 + m_1 t$ where t is the year. The parameter m_0 describes the respective change in days per year t. Table 3.3 summarizes the m_1 values.

Values exceeding the $\alpha = 0.025$ level of significance are indicated by a star. Overall, there is a tendency for both the mean value and the variance to decrease. This indicates a tendency for extreme precipitation events to 1) occur earlier in the year and 2) less spread out over the year. However, the signal is weak and not significant for SG and NG. For WG, a significant trend is present in COSMO-REA6 at smaller resolution steps $m = 0, \ldots, 2$, indicating a decrease

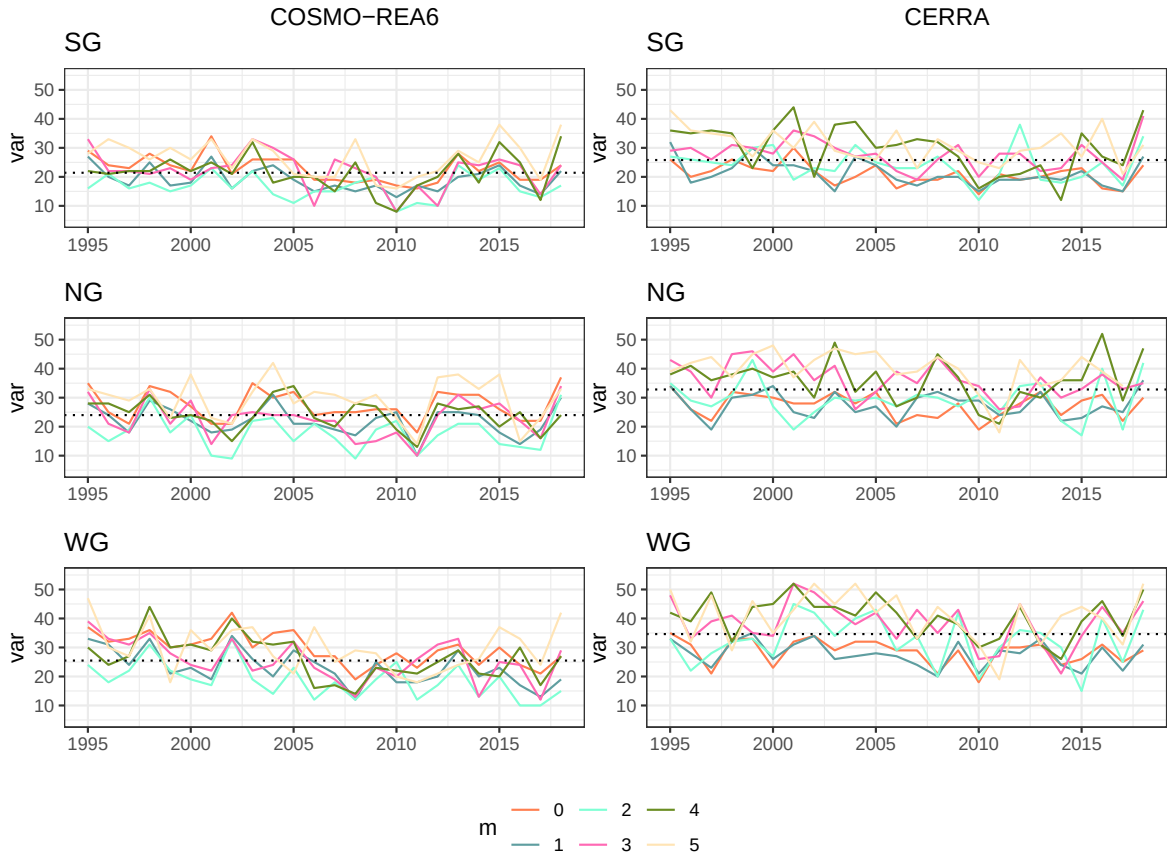


Figure 3.27.: Same as Fig. 3.26 but for timeseries of yearly circular variance.

in the variance of about 0.5 days per year. CERRA, on the other hand, also shows weaker and non-significant trends for WG. Although not shown, this trend strengthens when examining the moving average and moving variance over periods of two or more years. In these cases, significant, decreasing trends in the annual mean values in WG for COSMO-REA6 and resolution steps $m = 4, 5$ are also observed. To the best of our knowledge, no similar trends have been identified in or around the WG area so far. A connection with the trends discussed in section 3.4.1 is possible and even probable. However, the physical drivers behind these trends remain unknown. In order to assess the uncertainties and verify the observed trends, it is advisable to analyze the underlying factors and dynamics in the COSMO-REA6 dataset that are driving this trend. However, this is not part of this work; it provides an outlook for future research.

Table 3.3.: Parameter d_1 according to linear fit $d(t) = d_0 + d_1 t$ for time series of yearly circular mean (left) and variance (right) according to Figures 3.26 and 3.27. The values for COSMO-REA6/CERRA are shown. Values exceeding the significance level $\alpha = 0.025$ are marked with an star.

	Mean			Variance		
	SG	NG	WG	SG	NG	WG
m=0	-0.2/0.0	0.0/0.2	-0.1/-0.2	-0.3/-0.2	0.0/-0.1	-0.5*/-0.2
m=1	-0.1/0.0	-0.1/-0.1	-0.7/-0.3	-0.2/-0.3	-0.1/0.0	-0.5*/-0.2
m=2	-0.0/-0.3	-0.1/0.3	-0.8/-0.7	-0.1/-0.1	-0.1/0.0	-0.4*/0.0
m=3	0.1/0.1	-0.1/-0.4	-0.8/0.4	-0.2/-0.2	-0.0/-0.4	-0.5/-0.3
m=4	-0.3/0.0	-0.1/-0.4	-1.1/-0.6	0.0/-0.4	-0.2/-0.1	-0.5/-0.3
m=5	-0.4/1.0	0.1/-0.1	-0.9/-3	-0.1/-0.3	-0.1/0.1	-0.2/-0.2

3.5. Conclusion & Outlook

In this study, we analyzed the DT-CWT for its applicability in a scale-aware analysis of precipitation extremes over Germany. Our focus was twofold: first, on the representation of extreme precipitation events across different spatial scales, and second, on investigating potential changes in the context of climate change.

We utilized hourly precipitation data from reanalysis datasets, comparing the two European regional reanalyses, COSMO-REA6 and CERRA, and decompose the precipitation for each time step into spectral energies on different spatial resolutions. Each resolution step contains a proportion of the total energy from the input field. We determine an equivalent spatial scale for each resolution step and estimate the spatial scale in km for each resolution step $m = 0, \ldots, 5$ as $5.5 \cdot 2^m$ km, where 5.5 is the grid spacing. Our results are similar to Brune et al., 2018, but underestimate large scales. We recommend the exponential approximation proposed in Buschow, 2023b for a more accurate determination of the spatial scales.

We compare the representation of extreme precipitation in COSMO-REA6 and CERRA on different spatial scales and find good agreement for larger scales (> 22 km). On smaller scales (< 22 km), COSMO-REA6 is more active compared to CERRA. To analyze this more specifically, we look at the representation of the diurnal cycle in both models and find that it is not adequately reproduced in CERRA. Since the diurnal cycle of precipitation is particularly characterized by convective events (e.g., triggered by ground warming), we conclude that CERRA cannot adequately represent these events. The validation of our results is challenging due to the lack of comparative literature. In personal communication, the authors of Ridal et al., 2024 agreed that the diurnal cycle in CERRA is unrealistic. They suggest the CERRA-Land product (Verrelle et al., 2022) for a more accurate representation. However, this product is not suitable for our wavelet approach due to problems with the boundary conditions at the land-sea transitions. The differences between COSMO-REA6 and CERRA could possibly be explained by the use of the hydrostatic approximation in the HARMONIE-ALADIN system, the core of CERRA. COSMO-REA6 does not use this approximation (Bollmeyer et al., 2015). As indicated in section 3.3, CERRA data is available for lead times between one and six hours. We use the first three hours here. It is possible that using other, longer lead times will improve the representation of convective events in CERRA. However, since we already clarified that the diurnal cycle in CERRA is unrealistic, we did not include a further analysis in this regard.

The local wavelet energy, i.e., the spatial sum of LWS: $d^m = \sum_{x,y} d^m(x,y)$, is suitable as a measure of the energy content for each resolution step $m = 0,\ldots,5$. A useful interpretation results from the Parseval energy theorem, which states that the total energy of the system, i.e., the squared amount of precipitation, is equivalent to $d = \sum_m d^m$. We thus conclude that the DT-CWT is suitable for a compact description of precipitation extremes on different spatial scales. In our application, however, we face several uncertainties resulting from the RDWT and the pre-processing of our data. For RDWT, the grid point resolution of the high scales leads to an overrepresentation after spatial aggregation. As Brune et al., 2020; Buschow et al., 2019 show, this bias could be reduced by excluding grid points without precipitation. In case the user is interested in a description of the energy content of the spatial scales and less in the spatial localization, we recommend a non-redundant DWT as this reduces the bias and is less computationally expensive.

In a final step, we examine the LWS for possible non-stationarities. Our focus is on changes in intensity and seasonality. To assess possible changes in intensity, we calculate seasonal block maxima for each grid point and apply the log-ratio test to check for non-stationarity in the location parameter of GEV. We obtain spatial fields of scaling estimates that we compare for COSMO-REA6 and CERRA. We find rough agreement between the two datasets, except for WG. In order to summarize a significance level for changes within the areas SG, NG and WG, we apply methods from field significance. In COSMO-REA6 and CERRA, we find consistent trends towards more intense small-scale extremes in the JJA months and more intense large-scale extremes in the DJF months. The COSMO-REA6 data shows consistent trends in WG towards more intense precipitation extremes across all scales in JJA months. To assess possible changes in seasonality, we identify events in LWS that were the highest event of the year at least one grid point. In doing so, we identify at least 20 events per year in each area and resolution step m, and we analyze changes in the variance and mean of their temporal occurrence over the years. We find evidence for changes in the variance in WG for COSMO-REA6.

We compare our scale-aware analysis results with similar trend analysis on hourly precipitation and find similar trends in WG in COSMO-REA6 for JJA months. The scale-dependent trends that we identify from both data sets are not evident from hourly precipitation data. Therefore, we can conclude that our wavelet approach detects trends in precipitation that are not apparent from the raw data. Therefore, our wavelet method is suitable for the scale-based trend analysis.

Categorizing the identified trends is challenging as precipitation characteristics are known to be influenced by large-scale atmospheric circulations (Peter et al., 2024). We, therefore, only detect these trends and do not attribute them to contributing factors such as climate change. This is the first step of a so-called detection & attribution study (Hegerl et al., 2010). The uncertainties in our results are manifold. They result from the limited time series (1995-2019) and uncertainties from the data sets. Added to this is the lack of relevant literature on CERRA, which makes it difficult to assess these uncertainties. It has further been shown that results obtained from studies like this are quite sensitive to the data sets used (see, e.g. Lengfeld et al., 2023). We thus recommend applying the methodology to further data sets in order to validate our results. In order to attribute our identified trends, an application to climate models will be necessary. However, the regional and global models available to date cannot adequately resolve precipitation events (Ritzhaupt and Maraun, 2023; Ritzhaupt and Maraun, 2024). Convection-permitting models (CPMs), in which convection is explicitly resolved, give a promising approach. However, these data sets are relatively new and are not yet available in sufficient numbers.

There is a lack of comparative literature. Although there is existing literature on changes in extreme precipitation over Germany Zeder and Fischer, 2020; Masson-Delmotte et al., 2021; Zolina et al., 2008, it generally focuses on daily precipitation data, as the available time series of hourly data are often considered too short to identify robust trends (Fowler et al., 2021). Our results for large scales ($m = 4.5$) are generally consistent with the analyses based on daily precipitation data. In Germany, such analyses show on average a positive trend in the DJF months and a slightly negative trend in the JJA months. We find positive trends in DJF and no negative trends in JJA months (except for WG). Following the flood situation in July 2021, the WG region was particularly analyzed for extreme precipitation events (see e.g. Kreienkamp et al., 2021; Tradowsky et al., 2023; Lengfeld et al., 2023; Ludwig et al., 2023; Mohr et al., 2023). Kreienkamp et al., 2021; Tradowsky et al., 2023 detect an increase in block maxima of daily precipitation from E-OBS data over western Germany. This trend is, however, not evident in HYRAS-dataset[5] provided by German Weather Service (DWD), which highlights the difficulties encountered when analyzing trends in extreme precipitation, as well as their sensitivity to the datasets used.

Further work could include extending the wavelet approach to the consideration of the underlying space-time scales of precipitation events. This would allow for a more comprehensive analysis of extreme precipitation by capturing their persistence over time. One possible approach to take this into account is the use of the 3-dimensional DT-CWT (Selesnick and Li, 2003).

[5] `https://www.dwd.de/DE/leistungen/hyras/hyras.html` (last accessed: 09.07.2024)

A. Appendix Extremal Dependence

A.1. Paper: Szemkus et al. 2024

Adv. Stat. Clim. Meteorol. Oceanogr., 10, 29–49, 2024
https://doi.org/10.5194/ascmo-10-29-2024
© Author(s) 2024. This work is distributed under
the Creative Commons Attribution 4.0 License.

ASCMO
Open Access

Spatial patterns and indices for heat waves and droughts over Europe using a decomposition of extremal dependency

Svenja Szemkus and Petra Friederichs

Institute of Geosciences, University of Bonn, Bonn, Germany

Correspondence: Svenja Szemkus (sszemkus@uni-bonn.de)

Received: 31 January 2023 – Revised: 3 November 2023 – Accepted: 9 November 2023 – Published: 22 January 2024

Abstract. We present a method for the analysis and compact description of large-scale multivariate weather extremes. Spatial patterns of extreme events are identified using the tail pairwise dependence matrix (TPDM) proposed by Cooley and Thibaud (2019). We also introduce the cross-TPDM to identify patterns of common extremes in two variables. An extremal pattern index (EPI) is developed to provide a pattern-based aggregation of temperature. A heat wave definition based on EPI is able to detect the most important heat waves over Europe. As an extension for considering simultaneous extremes in two variables, we propose the threshold-based EPI (TEPI) that captures the compound character of spatial extremes. We investigate daily temperature maxima and precipitation deficits at different accumulation times and find evidence that preceding precipitation deficits have a significant influence on the development of heat waves and that heat waves often co-occur with short-term drought conditions. We exemplarily show for the European heat waves of 2003 and 2010 that TEPI is suitable for describing the large-scale compound character of heat waves.

1 Introduction

Extreme weather events over Europe, such as heat waves and droughts, but also heavy rainfall events have repeatedly attracted attention in recent years due to their dramatically high impact on socioeconomic systems. The most prominent European heat waves in the 21st century occurred in 2003, 2010, and 2018. The high temperatures, which lasted for a period of several weeks to months, in combination with compound drought periods, led to disastrous effects on socioeconomic and local natural systems, such as an increased number of wildfires, crop failure, and increased health-related deaths (see, e.g., Liu et al., 2020; Fischer and Schär, 2010). For instance, according to WHO, more than 70 000 additional deaths in 12 European countries can be attributed to the European heat wave in 2003 (Anderson et al., 2004), which was very likely the hottest summer in 500 years (Luterbacher et al., 2004).

Several studies indicate that an increase in the global mean temperature and its variability leads to more frequent heat wave and drought events that are characterized by longer durations and greater severity (Gershunov and Douville, 2009; Dai et al., 2018; Guerreiro et al., 2018; Della-Marta et al., 2007). This is also confirmed in the recent IPCC report (IPCC, 2021), saying that European hot extremes (including heat wave) have significantly increased in intensity and number since 1950. Particularly severe heat waves in Europe occurred in recent years in 2018, 2019, and 2022, often in connection with drought conditions. The 2022 heat wave resulted in record-breaking temperatures in France, Spain, the United Kingdom, and Germany (Rousi et al., 2023). Our understanding of the underlying dynamics of heat waves and, consequently, the changes that can be expected under climate change have improved considerably in recent years. In contrast, the dynamics and underlying structure of co-occurring temperature and drought events are much less understood. However, studies show that concurrent drought events can strongly influence heat extremes (Zscheischler and Seneviratne, 2017). For example, for the heat wave of 2003, it has been demonstrated that the co-occurring drought was a key contributor to the extreme temperature anomalies (e.g., Fischer et al., 2007; Black et al., 2004; Ferranti and Viterbo,

2006). Therefore, when considering potential changes under climate change, joint consideration of heat and drought extremes is necessary.

Several definitions and indices for drought and heat waves can be found in the literature. Heat wave indices usually include daily maximum temperature exceeding a certain threshold for several consecutive days (Robinson, 2001). Standard indices for heat waves are, e.g., the heat and cold wave index (HCWI) (Lavaysse et al., 2018), which is implemented in the Copernicus European Drought Observatory or the heat wave duration index (HWDI) and has been recommended by the IPCC (Griggs and Noguer, 2002). Standard meteorological drought indices are, e.g., the Palmer drought severity index (PDSI) (Palmer, 1965) and the standardized precipitation index (PSI) (McKee et al., 1993). A recent, promising approach to identifying heat waves and droughts is presented by Schädler and Breil (2021). They base their evaluation on so-called regional climate networks (RCNs) and identify extremes over Germany. Indices for compound heat waves and drought are less common in the literature. Recent approaches to index definition include the dry-hot magnitude index (Wu et al., 2019) or the standardized compound event indicator (Hao et al., 2019).

Detection and attribution (D&A) provides a powerful statistical tool for analyzing climate change's complex causes and effects. Reviews of the application of D&A to meteorological data are given in Slater et al. (2021) and Ummenhofer and Meehl (2017). In the high-dimensional space of the climate system, both the natural variability of the system and some internal variability overshadow the signals of climate change (Fischer and Knutti, 2016). Therefore, local changes, especially on smaller scales, only become visible if the data are analyzed in a compressed form. Data compression is therefore of great importance for a successful and reliable D&A study.

There exists a variety of statistical methods of different complexity for information compression. For example, a relatively simple approach is to aggregate the data spatially, but this usually neglects the spatial dependence structure of the underlying extreme values. Other techniques may include filtering using Fourier or wavelet transforms or adaptive data methods such as principal component analysis (PCA). However, most data compression methods focus on the description of the bulk of the underlying distribution, as described by first- and second-order moments, and provide little to no information about the tail of the distribution. Adaptive data methods have been widely used to study the spatial structure of heat waves. For example, Stefanon et al. (2012) use a clustering approach to identify six common heat wave patterns over Europe. Schubert et al. (2014) identify and analyze spatial patterns related to precipitation and temperature based on rotated empirical orthogonal functions (EOFs) over Russia. The studies agree that the patterns identified are consistent with large-scale blocking anticyclones associated with Rossby waves.

For the D&A of extreme events, mathematical methods are needed that achieve a targeted description of extremes. The framework in which these methods operate is the multivariate extreme value theory (MEVT) with broad applications in mathematical and meteorological research (e.g., Beirlant et al., 2006; Coles et al., 2001). New approaches combine MEVT with existing information compression methods. The rich experience with these methods in atmospheric research allows their targeted application to extreme events. For example, Morris et al. (2019) use an approach based on empirical basis functions to model the spatial dependence of low-rank max-stable processes. Janßen and Wan (2020) and Saunders et al. (2021) propose clustering approaches combined with an extremal framework, and Goix et al. (2017) also use a clustering approach to design a statistical learning procedure for anomaly detection based on a dataset with standard Pareto margins. Boulaguiem et al. (2022) combine generative adversarial networks with extreme value theory to model spatial dependencies in temperature and precipitation extremes. This shows that some of these methods are suitable for the description of composite extrema in two variables. For example, Vignotto et al. (2021) use an extreme value dependence clustering approach to identify spatial patterns of situations where precipitation and wind extremes are very likely to occur simultaneously.

Another promising approach comes from Cooley and Thibaud (2019). They propose the tail pairwise dependence matrix (TPDM), which summarizes extremal dependence using a regular varying framework and a threshold-based estimator. They build on the work of Larsson and Resnick (2012), who first defined this matrix for the bivariate case and show that the TPDM has some favorable properties that it shares with the covariance matrix so that the TPDM can be used as an analog to the covariance matrix for extremes. Jiang et al. (2020) and Rohrbeck and Cooley (2023) propose a PCA study using the TPDM with applications to daily precipitation over the United States (Jiang et al., 2020) as well as historical river flow levels across northern England and southern Scotland (Rohrbeck and Cooley, 2023). Drees and Sabourin (2019) also investigate a PCA based on the TPDM but focus more on describing the mathematical background. Finally, Fix et al. (2021) use the TPDM to propose a new model for spatial dependencies inspired by the simultaneous autoregressive (SAR) model with application to a gridded dataset of daily precipitation over Colorado.

In this paper, we introduce the extremal pattern index (EPI) for spatial weather extremes. The EPI is constructed based on the leading patterns derived from a PCA for extremes, as proposed by Jiang et al. (2020). Climate indices based on PCA, e.g., for the North Atlantic Oscillation (Barnston and Livezey, 1987), are already well-established tools in climate monitoring and analysis. By replacing the covariance matrix with the TPDM, we shift our focus to identifying extreme weather patterns, such as heat waves and droughts, while leaving the PCA procedure basically unchanged.

We use daily maximum 2 m temperature (maxT2m) from COSMO-REA6 reanalysis to derive the TPDM and identify spatial patterns associated with temperature extremes. Its EPI describes the intensity and spatial extent of heat waves in the area under consideration. To test the suitability of the EPI for describing heat waves, we focus on the two European heat waves in 2003 and 2010, about which many publications already exist. We compare our findings with existing literature.

We further explore the potential use and benefits of the TPDM in describing co-occurring heat waves and meteorological droughts. We again use maxT2m and define precipitation deficit (PD) as the additive reciprocal of daily precipitation accumulated over periods ranging from 11 to 47 d. To identify the pairwise extremal dependence between both variables, we introduce the cross-TPDM and provide an estimator thereof. We decompose the estimator of the cross-TPDM using a singular value decomposition (SVD), providing insight into the co-occurrence of extreme spatial patterns in two variables. Finally, we apply an analogue definition of the EPI using the SVD of the cross-TPDM. In order to target the index to co-occurring events, a modulated threshold-based estimator of EPI (TEPI) is defined.

The remainder of the paper is organized as follows. First, the theoretical framework of regular variation underlying the definition of the TPDM is presented in Sect. 2. Here we also discuss the estimation of the TPDM and its decomposition. Accordingly, we introduce the cross-TPDM in Sect. 3. We then discuss their evaluation (Sect. 4) and the definition of the various extremal pattern indices (Sect. 5). Section 6 discusses the data and some pre-processing steps. The results are presented in Sect. 7 with a discussion first on the temperature extremes and the EPI to define heat waves and then the compound heat wave and drought events. The paper concludes with a summary and discussion in Sect. 8.

2 The tail pairwise dependence matrix (TPDM)

2.1 Definition of the TPDM

The definition of the TPDM is based on regularly varying random variables. Regularly varying variables are defined by the tail behavior of their distribution function $F(x)$ in the limit $x \to \infty$. A summary on regular variation is given by Resnick (2007) saying "regularly varying functions are those functions which behave asymptotically like power functions". Let X be a non-negative random variable with positive and measurable distribution function $F(x)$ on $\mathbb{R}_+$. We call F to be a regularly varying function with tail index $\rho \in \mathbb{R}$ (denoted by $F(X) \in RV_\rho$), if for all $t > 0$

$$\lim_{x\to\infty} \frac{F(xt)}{F(x)} = t^\rho. \tag{1}$$

If the tail function, $\overline{F}(X) = 1 - F(X)$, is a regularly varying function, i.e. $\overline{F}(X) \in RV_{-\rho}$ $(\rho > 0)$, then X is a regularly varying random variable with tail index ρ (Resnick, 2007, Theorem 3.6), and we write $X \in RV_\rho$. If X is Fréchet-distributed with $F(x) = \exp(-x^{-\alpha})$ and $x \geq 0$, then X is a regularly varying random variable with tail index α (Resnick, 2007).

Now let X be a positive multivariate random variable of dimension p with Fréchet-distributed margins and tail index α. In the following, we denote $X \in RV_\alpha^+$. For multivariate regularly varying random variables, it is convenient to transform the random vector to generalized polar coordinates with the angular component ω and the radial component r. With any norm $||\cdot||$ we define the unit sphere $\mathbb{S}_+^{p-1}$ as

$$\mathbb{S}_+^{p-1} = \left\{x \in \mathbb{R}_+^p : ||x|| = 1\right\} \tag{2}$$

and the polar coordinate transform as

$$x \to \left(||x||, \frac{x}{||x||}\right) =: (r, \omega) \quad \text{for} \quad ||x|| > 0, \tag{3}$$

with $r = ||x||$ and $\omega = x/||x||$. If X is regularly varying, then its radial and angular components become conditionally independent for $r \to \infty$.

To understand the derivation and definition of the TPDM, we explore the properties of a regularly varying random variable. This exploration is done through the lens of weak convergence (Resnick, 2007), restricted to the special case of vague convergence, denoting convergence of Radon measures. Given a multivariate, Fréchet-distributed random vector $X \in RV_\alpha^+$, the definition of regular variation says that there is a sequence b_n with $n \to \infty$ and a limit measure $\nu(\cdot)$, such that we can write $n\mathbb{P}\left[X/b_n \in \cdot\right] \xrightarrow{v} \nu(\cdot)$ as $n \to \infty$. Here, $\xrightarrow{v}$ denotes vague convergence in $M_+([0,\infty]^p \setminus \{0\})$, the space of Radon measures on $(0,\infty] \setminus \{0\}$. After the polar transform, one can specify $\nu(\mathrm{d}r \times \mathrm{d}\omega) = \rho r^{-\rho-1}\mathrm{d}r\mathrm{d}H(\omega)$ where ρ is the tail index of X, and H is some Radon measure, referred to as the angular measure on $\mathbb{S}_+^{p-1}$. It can be interpreted as a probability measure for ω (for a detailed description see, e.g., Larsson and Resnick, 2012; Cooley and Thibaud, 2019; Resnick, 2007).

Larsson and Resnick (2012) found that the measure H, on the unit sphere $\mathbb{S}_+^{p-1}$, holds information about extremal dependencies. They define the extremal dependence measure of any regularly varying, bivariate random vector based on its angular components after polar transformation.

Cooley and Thibaud (2019) expand the original approach to higher dimensions. They concentrate on Fréchet-distributed variables with tail index $\alpha = 2$ and choose the L_2 norm when transforming into polar coordinates. They define the tailed pairwise dependence between the ith and jth element of X, using its angular component ω as

$$\sigma_{ij} = \int_{\mathbb{S}_+^{p-1}} \omega_i \omega_j \mathrm{d}H(\omega). \tag{4}$$

The TPDM (Σ_X) is then a $p \times p$ matrix, which summarizes the pairwise extremal dependencies of X for all mar-

gins $i, j = 1, \ldots, p$: $\Sigma_X = (\sigma_{ij})_{i,j=1\ldots p}$. Cooley and Thibaud (2019) show that by their choice of parameters, the properties of the TPDM are very close to those of the covariance matrix. The most important property for our application is that the TPDM is symmetric and positive definite with real and positive eigenvalues. For rigorous derivations and proofs as well as additional literature, the reader is referred to Cooley and Thibaud (2019).

2.2 Estimation of the TPDM

Let x_t, $t = 1, \ldots, n$ be a sample of a p-dimensional random vector X with Fréchet margins and sample components $x_{t,i}, i = 1, \ldots, p$. The estimator of the individual elements of the TPDM is given by Jiang et al. (2020) as

$$\hat{\sigma}_{ij} = 2n_{ij,\mathrm{exc}}^{-1} \sum_{t=1}^{n} w_{t,i} w_{t,j} \mathbb{I}_{(r_{t,ij} > r_{0,ij})}. \tag{5}$$

with $r_{t,ij} = \sqrt{x_{t,i}^2 + x_{t,j}^2}$ as well as $w_{t,i} = x_{t,i}/r_{t,ij}$ and $w_{t,j} = x_{t,j}/r_{t,ij}$. $r_{0,ij}$ is a threshold that can be set individually, with the index function $\mathbb{I}$ ensuring that only observations larger than $r_{0,ij}$ are considered for the estimation, and $n_{ij,\mathrm{exc}} = \sum_{t=1}^{n} \mathbb{I}_{(r_{t,ij} > r_{0,ij})}$ is the number of elements which are greater than $r_{0,ij}$. The positive definiteness of the TPDM is not guaranteed by the proposed estimator in Eq. (5). If the estimator is not positive definite, we follow Cooley and Thibaud (2019) by computing the nearest correlation matrix (Higham, 2002).

Pre-processing is required when using meteorological data. Again we follow Jiang et al. (2020) and transform the marginal distribution of the original data to Fréchet margins with tail index $\alpha = 2$. Let X^{orig} be the original random vector with marginal distribution $F_i^{\mathrm{orig}}(X), i = 1, \ldots, p$ and X the transformed random vector with Fréchet margins $F_i(X) = \exp(-x_i^{-\alpha})$. Due to the conservation of probability, we have to ensure that $F_i(X) = F_i^{\mathrm{orig}}(X^{\mathrm{orig}})$, which leads us to the transformation

$$X_i = \left(-\log F_i^{\mathrm{orig}}(X^{\mathrm{orig}}) \right)^{-1/\alpha}. \tag{6}$$

As an estimator for F_i^{orig}, we may either use the empirical distribution function or fit a parametric distribution (e.g. normal distribution).

2.3 Decomposition of TPDM

The eigenvalue decomposition (EVD) of a covariance matrix is a standard tool in climate data analysis to reduce the dimensionality of high-dimensional weather and climate data (see, e.g., Wallace et al., 1992). The eigenvectors (EVs) form an orthogonal, complete base system. Here we use an estimator based on extreme dependencies instead of covariances, which leads to a targeted consideration of extremal dependencies. For an introduction to EVD and its applications, see, for example, Wilks (2011), Jolliffe (1986), and Björnsson and Venegas (1997). The decomposition of the TPDM, with $\Sigma_{XX} = U \Lambda U^T$, provides the eigenvector matrix U with eigenvectors $u^{(k)}, k = 1, \ldots, p$ and components $u_1^{(k)}, \ldots u_p^{(k)}$, as well as the diagonal-matrix Λ, containing the eigenvalues $\lambda_1, \ldots \lambda_p$.

The TPDM is defined on the basis of positive real numbers. The transformation to Fréchet margins leads to the input variable X being consistent with this assumption. However, the EVs of the TPDM live by construction in $\mathbb{R}^p$. For the EVs to form an orthogonal basis in $\mathbb{R}_+^p$, we have to transform the EVs from $\mathbb{R}^p$ to $\mathbb{R}_+^p$. We again follow Cooley and Thibaud (2019) and use $\tau(y) = \log(1 + \exp(y))$ with its back-transform $\tau^{-1}(y) = \log(\exp(y) - 1)$. The Fréchet distribution, as well as the orthogonality of the EVs, is preserved under this transformation. We transform the EVs componentwise with $e_j^{(k)} = \tau(u_j^{(k)})$, for each of the $j = 1, \ldots, p$ components and $k = 1, \ldots, p$ vectors. The time series of principal components (PCs), in contrast, are calculated by a projection of the back-transformed data X on the respective EV as

$$\boldsymbol{\eta} = U^T \tau^{-1}(X). \tag{7}$$

We can reconstruct a filtered data set for each sample x_t ($t = 1, \ldots, n$) of the p-dimensional input variable X using $\tilde{x}_t = \tau(\eta_t U)$.

3 The cross-TPDM

3.1 Definition of the cross-TPDM

The TPDM presented above provides information on the pairwise dependencies between the margins of a random vector. Here we introduce the cross-TPDM as a measure of extremal dependency between two random vectors, such as spatial fields of maxT2m and PD.

Let X and Y be Fréchet random vectors with tail index α and dimensions p and q, respectively. For the following definitions, we consider a joint random vector $Z = (X_1, \ldots, X_p, Y_1, \ldots, Y_q)^T$ of dimension $(p+q)$. As in Sect. 2, we use the L_2 norm $||z||$ to define the unit sphere:

$$\mathbb{S}_+^{(p+q)-1} = \{ z \in \mathbb{R}_+^{p+q} : ||z|| = 1 \}. \tag{8}$$

Along with the transformation of z into polar coordinates,

$$z \to \left(||z||, \frac{z}{||z||} \right) =: (r, \omega), \tag{9}$$

with $r = ||z||$ and $\omega = z/||z||$. Following Eq. (4) we define the joint TPDM, which has the dimensions $(p+q) \times (p+q)$. It separates into Σ_{XX}, Σ_{YY}, and the cross-TPDM Σ_{XY}, which is a $p \times q$ matrix. If we break down ω into its x and y components $\omega^x = (x_1/||z||, \ldots, x_p/||z||)^T$ and $\omega^y = (y_1/||z||, \ldots, y_q/||z||)^T$, the cross-TPDM Σ_{XY} has the en-

tries

$$\sigma_{ij} = \int_{\mathbb{S}_+^{(p+q)-1}} \omega_i^x \omega_j^y \mathrm{d}H(\omega). \tag{10}$$

The cross-TPDM is the counterpart of the cross-covariance matrix and represents an extension of the TPDM to two variables. By definition, it is a component of the joint TPDM, obtained from the joint random vector Z. While it is conceivable that the joint TPDM could also be used to extract information about the extremal dependence between X and Y, this approach is relatively rare and involves higher computational costs.

3.2 Estimation of the cross-TPDM

Let x_t, y_t, $t = 1, \ldots, n$ be samples of two Fréchet random vectors X and Y, with components $x_{t,i}$, $i = 1, \ldots, p$ and $y_{t,j}$, $j = 1, \ldots, q$. To estimate the cross-TPDM, we compute, for the radial component,

$$r_{t,ij} = \sqrt{x_{t,i}^2 + y_{t,j}^2} \tag{11}$$

and, for the angular components,

$$w_{t,i}^x = \frac{x_{t,i}}{r_{t,ij}}; w_{t,j}^y = \frac{y_{t,j}}{r_{t,ij}}. \tag{12}$$

We then estimate each element of the cross-TPDM Σ_{XY} as

$$\hat{\sigma}_{ij} = 2n_{ij,\mathrm{exc}}^{-1} \sum_{t=1}^{n} w_{t,i}^x w_{t,j}^y \mathbb{I}_{(r_{t,ij} > r_{0,ij})}. \tag{13}$$

with threshold $r_{0,ij}$ and $n_{ij,\mathrm{exc}} = \sum_{t=1}^{n} \mathbb{I}_{(r_{t,ij} > r_{0,ij})}$, the number of exceedances.

3.3 Decomposition of the cross-TPDM

The SVD of a cross-covariance matrix is a standard tool in climate data analysis to reduce the dimensionality of high-dimensional weather and climate data (see, e.g., Wallace et al., 1992). The singular vectors (SVs) form an orthogonal basis, which can be used for information compression. Here we use an estimator based on extreme dependencies instead of cross-covariances, which leads to a targeted consideration of extremal dependencies. For an introduction to SVD and its applications, see, for example, Wilks (2011), Jolliffe (1986), and Björnsson and Venegas (1997).

The decomposition of the cross-TPDM with $\Sigma_{XY} = U \Lambda V^T$ provides left and right singular vectors (SVs), as well as the matrix Λ, which contains the singular values on its diagonal. The SVs can be interpreted as pairs of spatial patterns that explain the maximum possible proportion of extremal dependence, analogous to SVD in Newman and Sardeshmukh (1995). As they point out, one should be cautious in the physical interpretation of the patterns, since they are first of all statistical patterns. However, data reduction based on SVs does not require the physical interpretability of the patterns. The singular values provide information about the extent to which the dependencies are explained by the respective SV patterns.

The cross-TPDM is defined on the basis of positive real numbers. The transformation to Fréchet margins leads to the input variables X and Y being consistent with this assumption. However, the SVs live by construction in $\mathbb{R}^p$. For the SVs to form an orthogonal basis in $\mathbb{R}_+^p$, we have to transform the SVs from $\mathbb{R}^p$ to $\mathbb{R}_+^p$. We again follow Cooley and Thibaud (2019) and use $\tau(y) = \log(1 + \exp(y))$ with its back-transform $\tau^{-1}(y) = \log(\exp(y) - 1)$. The Fréchet distribution, as well as the orthogonality of the SVs, is preserved under this transformation. Again, we transform the SVs component-wise. The time series of expansion coefficients (ECs) are calculated by a projection of the back-transformed data X and Y on the respective SVs as

$$\boldsymbol{\eta} = U^T \tau^{-1}(X) \tag{14}$$

$$\boldsymbol{\nu} = V^T \tau^{-1}(Y). \tag{15}$$

Reconstructions of each sample x_t and y_t ($t = 1, \ldots, n$) are obtained using $\tilde{x}_t = \tau(\eta_t U)$ and $\tilde{y}_t = \tau(\nu_t V)$.

4 Evaluation

To assess how many EVs or SVs are needed to represent the main features of the data, we use the fraction of explained covariance, which is

$$\mathrm{ECF}_k = \frac{\lambda_k}{\sum_{i=1}^{p} \lambda_i}, \tag{16}$$

with λ_k the kth largest eigenvalue or singular value in Λ. The term ECF (explained covariance function) in the context of (cross-)TPDM may be misleading because, strictly speaking, SVs no longer provide information about (cross-)covariance but about extremal dependence. Nevertheless, we use the term here because it is a fixed term for measuring the explained dependencies in PCA or SVD.

To examine the extent to which each pattern is dominated by individual events that occurred during the training period and find an appropriate number of modes that yields a representative subset of patterns, we perform 2-fold cross-validation. There are several measurements that aim to find the "optimal" number of modes after PCA (see, e.g., Abdi and Williams, 2010). We use Q_m^2 as a rather simple but effective measure to assess overfitting (i.e., the point at which the quality of the prediction decreases with an increasing number of modes). To derive Q_m^2 we divide the time series into two periods of equal length and calculate a (cross-)TPDM for each period, respectively.

The reconstructions follow Sect. 2.3 and 3.3. The within-sample reconstruction $\tilde{x}_t^m = \tau(\tilde{u}^T \tau^{-1}(x_t)\tilde{u})$ uses the first

m EVs/SVs $\tilde{u}$ derived from the period including t, while the out-of-sample reconstruction $\hat{x}_t^m$ uses the corresponding EVs/SVs $\hat{u}$ of the period not including t. The reconstructions of the two periods are then merged to calculate Q_m^2 defined as

$$Q_m^2 = 1 - \frac{\sum_{t=1}^{n} ||x_t - \hat{x}_t^m||_2}{\sum_{t=1}^{n} ||x_t - \tilde{x}_t^{m-1}||_2}, \tag{17}$$

with $||\cdot||_2$ the L_2 norm. Q_m^2 is a measure of information gain relative to the increase of noise. A positive Q_m^2 means that the mth EV/SV mode significantly improves an out-of-sample reconstruction relative to the in-sample reconstruction using $m-1$ EV/SV modes. The number of modes m at which Q_m^2 becomes negative indicates the point at which overfitting starts to occur. The standard error of Q_m^2 is assessed by resampling using a block bootstrap method of annual data with 100 samples.

5 Extremal pattern index

We define an extremal pattern index (EPI) based on the PCs using the L_2 norm. The L_2 norm is chosen since both PCA and SVD may be defined as an L_2-minimization problem (see, e.g., Abdi and Williams, 2010). Let be given m PCs η_t at time t as derived by Eq. (7) using the m leading EV of a TPDM. We further incorporate the leading eigenvalues/singular values $\lambda_{1,\dots,m}$ to normalize the EPI, which we define at time t as

$$\mathrm{EPI}_t^{\eta} = \frac{\sqrt{\sum_{k=1}^{m} \eta_{t,k}^2}}{\sqrt{\sum_{j=1}^{m} \lambda_j}}. \tag{18}$$

The EPI represents a spatial aggregation and shows when individual patterns or a linear combination of patterns are particularly pronounced. The first EV/SVs of the (cross-)TPDM show little spatial variation (see, e.g., Cooley and Thibaud, 2019), so $\eta_{t,1}$ and $\nu_{t,1}$ essentially represent the spatial mean of the Fréchet-distributed variables. The higher modes are those representing the spatial variation patterns. Thus, the EPI can become extreme even if the first principal component is not outstanding.

The SVD of the cross-TPDM provides two ECs, for the right and the left singular vectors. We define the EPI for the cross-TPDM analogously as

$$\mathrm{EPI}_t^{\eta,\nu} = \frac{\sqrt{\sum_{k=1}^{m} (\eta_{t,k}^2 + \nu_{t,k}^2)}}{\sqrt{\sum_{j=1}^{m} \lambda_j}}. \tag{19}$$

This definition of EPI describes scenarios where either one or both of the corresponding ECs are high. To account for the simultaneous occurrence of extremes, we introduce a threshold-based approach that considers both the left and right ECs of each mode k to be prominently expressed. We refer to this as the threshold-based extremal pattern index (TEPI).

$$\mathrm{TEPI}_t^{\eta,\nu} = \frac{\sqrt{\sum_{k=1}^{m} (\eta_{t,k}^2 + \nu_{t,k}^2) \mathbb{I}_{(|\eta_{t,k}|>q_\eta, |\nu_{t,k}|>q_\nu)}}}{\sqrt{\sum_{j=1}^{m} \lambda_j}}, \tag{20}$$

with threshold values q_η and q_ν. The threshold ensures that the associated patterns are simultaneously highly pronounced.

6 Application

6.1 Data description

This study relies on the COSMO-REA6 regional reanalysis (Bollmeyer et al., 2015). COSMO-REA6 was developed within the Hans Ertel Centre for Weather Research (HErZ) at the University of Bonn. COSMO-REA6 is currently available for the period from 1995 to 2019 at the opendata FTP server[1] of the German meteorological service (DWD).

We use accumulated PD, as well as daily maxT2m. PD is defined as the additive inverse of precipitation accumulated over various periods between 11 and 47 d using a centered moving-average approach (e.g., for an accumulation period of 11 d, PD is estimated at the center including the 5 d preceding and following the center date, respectively). The use of the moving average for precipitation is less common in the literature, but for our application, we consider it appropriate. Otherwise, we would rather focus on the precipitation deficit before a heat wave. (e.g., PD from 1 to 30 June in conjunction with maxT2m on 30 June). The spatial domain of COSMO-REA6 is according to the CORDEX EUR-11 specifications[2] but with a higher horizontal resolution of about 6 km horizontal grid space. We limit our analysis to land areas and also exclude countries that have little to no precipitation during northern hemispheric summer months. We further restrict our analysis to the northern hemispheric summer months from June to August, when heat waves have particularly strong impacts on social and environmental systems through associated effects such as forest fires and heat stress.

6.2 Data pre-processing

Our pre-processing requires three steps: (1) in order to reduce spatial non-stationarity we choose to standardize the data at grid point level. Along with standardization, we also remove seasonal trends, which we identify in maxT2m and PD. To this end, we estimate mean and standard deviation at each grid point and day of the year, which are then

[1] https://opendata.dwd.de/climate_environment/REA/COSMO_REA6/ (last access: 20 December 2023)

[2] https://euro-cordex.net/060374/ (last access: 20 December 2023)

used for standardization. (2) The spatial fields in COSMO-REA6 contain 824×848 grid points, resulting in a matrix of $698\,752 \times 698\,752$ pairwise dependence measures. To limit the computational cost, the number of grid points is reduced to a 121×117 grid. For consistency with extreme value theory, we reduce the grid size using spatial block maxima instead of standard interpolation techniques. To this end, we compute the maximum in a 7×7 neighborhood of grid points for spatial reduction. (3) To transform maxT2m and PD to a Fréchet distribution with tail index α we use the empirical distribution function with

$$\hat{F}^{\text{orig}}_{Y_i}(y) = \frac{1}{n} \sum_{t=1}^{n} \mathbf{I}(y^{\text{orig}}_{t,i} \leq y). \tag{21}$$

6.2.1 Choice of threshold

The choice of the threshold value for estimating the TPDM is always critical. If the threshold is too low, there is a risk that the asymptotic limit of the extreme value model has not yet been reached, which leads to biases in the estimation. A threshold that leaves too few data points to which the model can be fitted leads to large uncertainties in the estimates. To determine a suitable threshold, we consider the rank of the TPDM and determine its stability as a function of the threshold value in Fig. 1. A stable linear downward trend is given for thresholds below the 98 % quantile (red line), but a rapid decline is seen for higher thresholds. The 98 % quantile, therefore, seems suitable as a threshold for further calculations.

Figure 2 shows maxT2m and the 15d PD at two grid points in COSMO-REA6 near Paris and Bonn. Only the data outside the gray shaded area (i.e., above the 98 % quantile) are used to estimate the cross-TPDM. The dependence between maxT2m at grid points with a distance between Paris and Bonn is strong even for the extreme values since the temperature variations are usually dominated by large-scale patterns and therefore still show dependencies, even at large distances. In contrast, the dependencies in the PD and between PD and maxT2m are much lower, which is due to the small-scale structures of the extreme values in the PD.

Figure 2 may be used to illustrate the asymptotic dependence/independence between Fréchet-distributed time series. The measure $H(\omega)$ of the angular component ω describes the tail dependence on the unit sphere $\mathbb{S}_+^{p-1}$ (Sect. 2). With asymptotic independence, we expect the mass of $H(\omega)$ to be at the edges, so we do not expect an extreme event to occur in both variables simultaneously. With asymptotic dependence, corresponding probability mass exists on the unit circle, and we observe events that are extreme in both components.

As Jiang et al. (2020) point out, the assumption of asymptotic dependence, which is the basis of the concept of regular variations, may not be fulfilled. Jiang et al. (2020) focus on precipitation data, which are clearly not asymptotically dependent at the continental scale. This is less clear

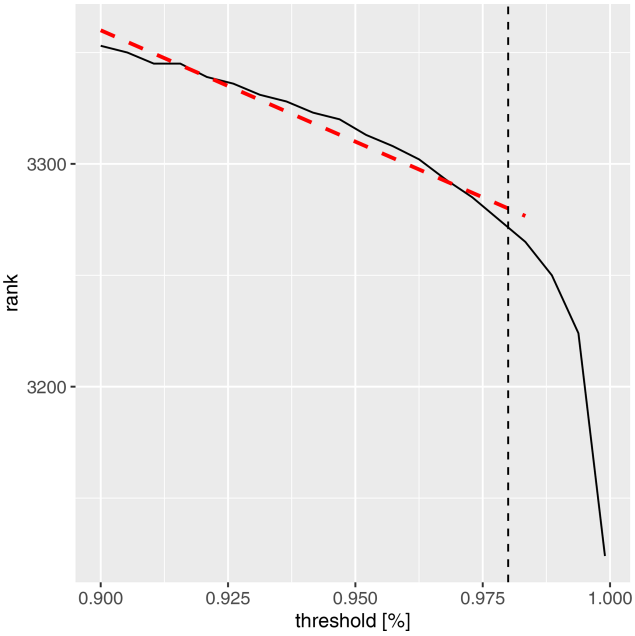


Figure 1. Rank of the TPDM, calculated for maxT2m and varying thresholds r_0 (Eq. 5). Vertical line: 98 % quantile. Red line: linear least-squares fit of the rank using a threshold between 90 % and 98 % quantile.

for regional-scale maxT2m and PD and cross-dependence. The TPDM method is nevertheless well suited to decompose (cross-)dependencies in our case since we typically do not reach the asymptotic limit for real data. For real data, we are more likely to be in the pre-asymptotic limit, which is characterized by significant dependencies, as seen in Fig. 2.

7 Results

7.1 Spatial patterns of extreme daily maximum 2m temperature

We start our analysis with the estimation of the TPDM for maxT2m. The estimation is based on maxT2m values exceeding the local 98 % quantile on the reduced grid (see Sect. 6.1). Figure 3 shows the first 12 EVs of the TPDM. The first mode represents the nonzero spatial mean and consists of nearly constant values with only small variations of order 10^{-2}. Since the data are defined in $\mathbb{R}_+$ (see Sect. 2.3), the mean value is not removed unlike in standard PCA (cf. Jiang et al., 2020; Cooley and Thibaud, 2019).

Higher-order EVs show large-scale patterns associated with the typical dipole and multipole structures also known from PCA. Since the EVs are transformed on $\mathbb{R}_+$, values $> \log(2)$ refer to positive and values $< \log(2)$ to negative anomalies (see Sect. 2.3). The wave-like structures are strongly reminiscent of westward-propagating Rossby waves, which are the main driving force for large-scale circu-

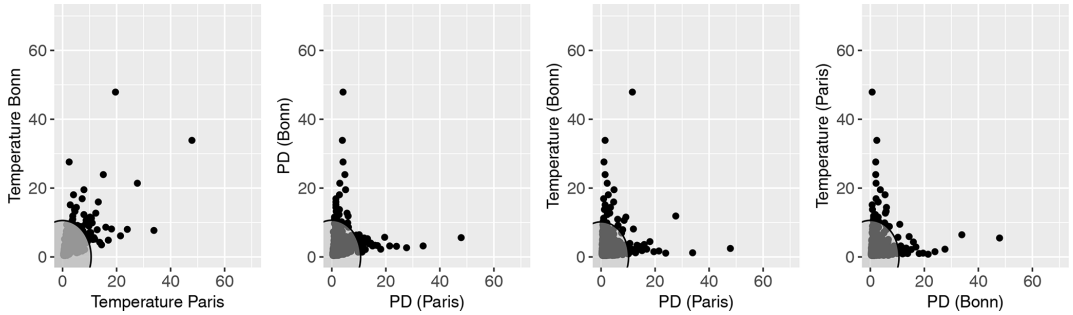


Figure 2. Scatter plots of Fréchet-transformed maxT2m and PD (black dots) at two grid points in COSMO-REA6 at Paris and Bonn. PD has been calculated from 15 d accumulated precipitation. The circle indicates the respective 98 % quantile, and the values inside the circle (gray shading) are not used for the estimation of the (cross-)TPDM.

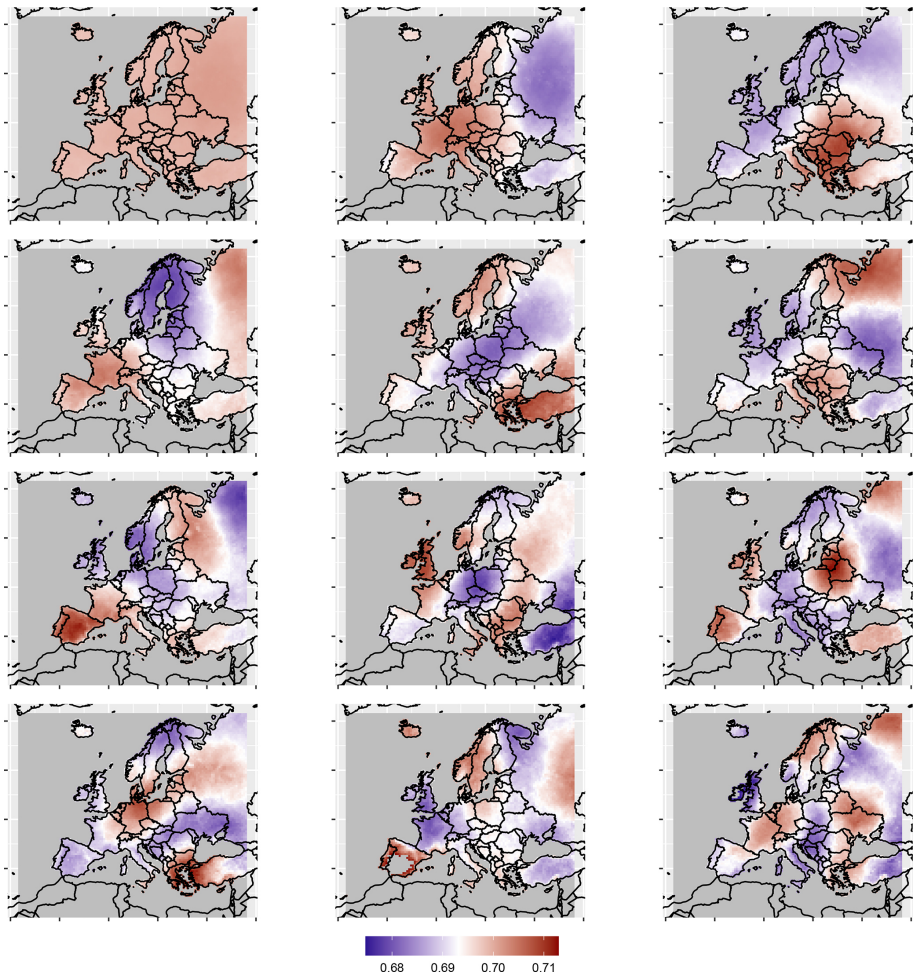


Figure 3. First 12 (from top to bottom, left to right) EVs of the TPDM for maxT2m in northern hemispheric summer months (JJA). Each vector is transformed to the positive real numbers $\mathbb{R}^p \to \mathbb{R}_+^p$ such that $\tau(0) = \log(2)$ corresponds to zero.

lation patterns in northern mid-latitudes (see, e.g., Liu et al., 2020; Schubert et al., 2011; Gershunov and Douville, 2009).

An important question is how representative the EVs are and at what number of modes significant information is no longer added by further modes but overfitting prevails. In Figs. 4 and 5 we show explained variance and Q^2 as introduced in Sect. 4. The first EV contributes to about 25 % of the total variance, followed by less than 8 % for the second EV. As in Jiang et al. (2020) and Cooley and Thibaud (2019), the first EV (i.e. changes in the spatial mean maxT2m) dom-

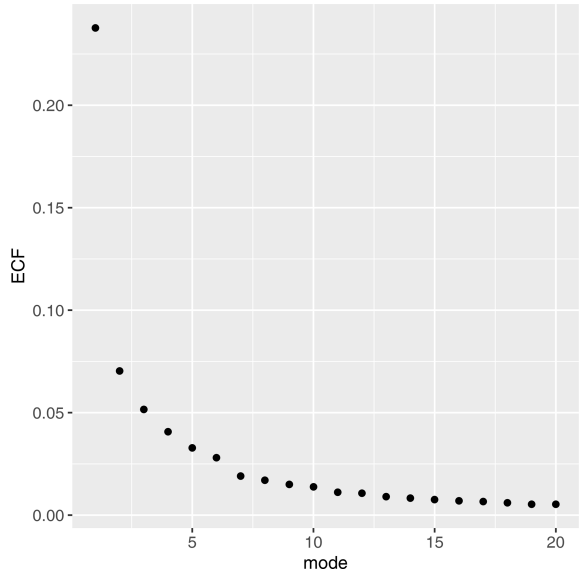


Figure 4. Explained variance ECF_k for each mode k (Eq. 16). Calculated from TPDM using maxT2m in northern hemispheric summer months (JJA).

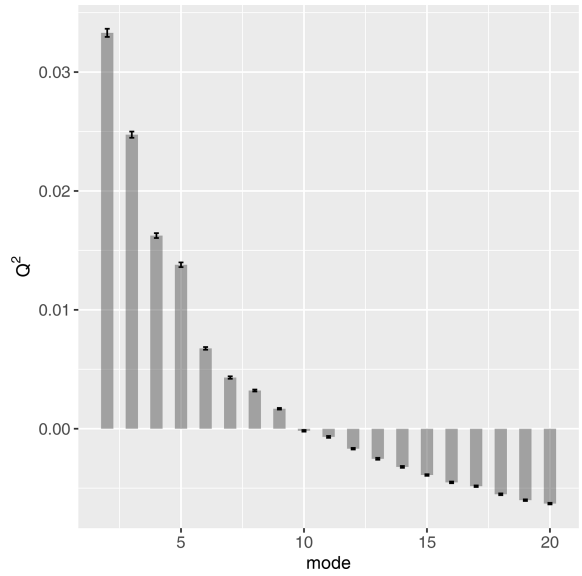


Figure 5. Q_m^2 for the first $m = 20$ modes (Eq. 17). Calculated from TPDM for maxT2m in northern hemispheric summer months (JJA). The error bars indicate the standard error determined using 100 bootstrap samples.

inates the variance. About 52 % (75 %) of the variance is explained by the first 10 (50) EVs. The out-of-sample information gain as measured by Q_m^2 is positive for the leading 10 EVs. Both measures indicate that the first 10 EVs provide a representative subset for the spatial variation patterns of extreme maxT2m.

7.2 Extreme pattern index for temperature extremes

Our calculation of the EPI^η (Eq. 18) uses the first 10 PCs η_k with $k = 1, \ldots, 10$. The time series of the PCs for the first six EVs are displayed in Fig. A1 in the Appendix. Figure 6 shows EPI^η for the June to September months from 1995 to 2019 with values varying between 0.5 and 13.5. Peaks in EPI^η indicate intense heat waves within the studied region. Although the 2010 summer heat wave prominently stands out in EPI^η, we also identify other peaks associated with typical heat waves. The large outliers in the PCs as well as in EPI^η indicate a heavy tail behavior, as already shown by Cooley and Thibaud (2019) for the PCs.

We will now examine the two heat waves of 2003 and 2010 in more detail using our framework. Both heat waves have been extensively studied and analyzed in the existing literature. The 2010 heat wave is characterized by the highest peak EPI in the period considered. This indicates that the 2010 heat wave was the most extreme event in terms of its combination of intensity and spatial extent over the period considered. The 2003 heat wave, on the other hand, did not reach the highest peak but is characterized by its long, highly pronounced duration. Finally, both heat waves had significant impacts on socio-economic systems, underlining their importance for understanding extreme heat events (see Beniston, 2004; Barriopedro et al., 2011; Kueh and Lin, 2020; Schubert et al., 2011; Gershunov and Douville, 2009). Figure 7 (bottom panel) shows EPI^η for the June to September months in 2003. The respective averaged maxT2m anomalies are shown in the upper panels for the periods during periods as marked with (a) to (c). We observe a first heat wave in mid-July (a) in western Europe, which is most pronounced over the United Kingdom, Iceland, and Norway. This short-term event is followed by a few non-extreme days (b), after which we observe a very severe heat wave over central and southern Europe from late July to mid-August (c) (see also García-Herrera et al., 2010; Liu et al., 2020). In Fig. 7 we also display the reconstruction of the maxT2m fields using only the leading 10 EVs. The spatial structure of the heat wave pattern in maxT2m is well represented within the 10-dimensional subspace. The reconstructed patterns are somewhat smoother and thus more clearly highlight the areas affected by the heat wave. Thus, the subspace spanned by the 10 eigenvectors of the TPDM is sufficient to describe the 2003 heat wave and fulfills the goal of a compact description of large-scale heat waves.

In Fig. 8, we analyze the Russian heat wave in 2010. High values of EPI^η consistently capture their evolution from the second week of July to mid-August, as analyzed, e.g., in Barriopedro et al. (2011) or Liu et al. (2020). Since we observe the strongest expressions of EPI^η in the first half of August, we focus on this period in the following. The beginning of August is characterized by the most intensive phase of the heat wave according to EPI^η (a), with its center located over Moscow, western Russia. In the second week, we observe a

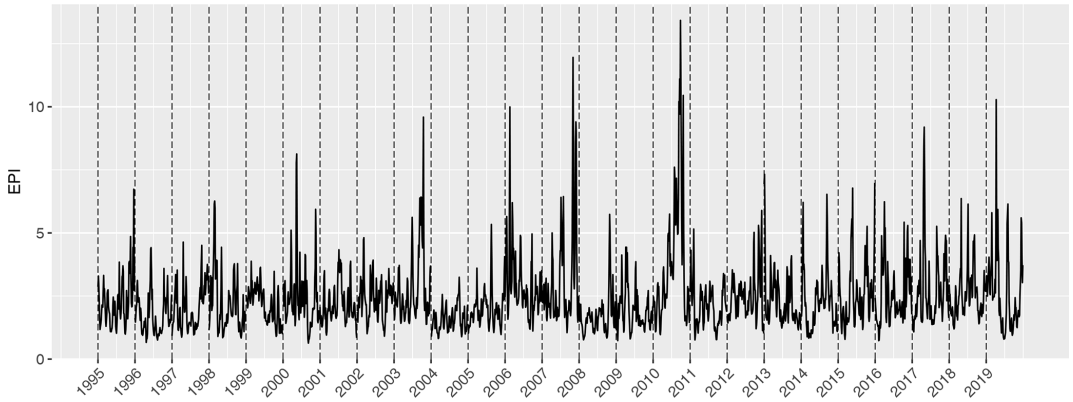


Figure 6. EPI$^{\eta}$ as calculated from first 10 PCs after a PCA of the TPDM for maxT2m. Shown are the northern hemispheric summer months (JJA) from 1995 to 2019.

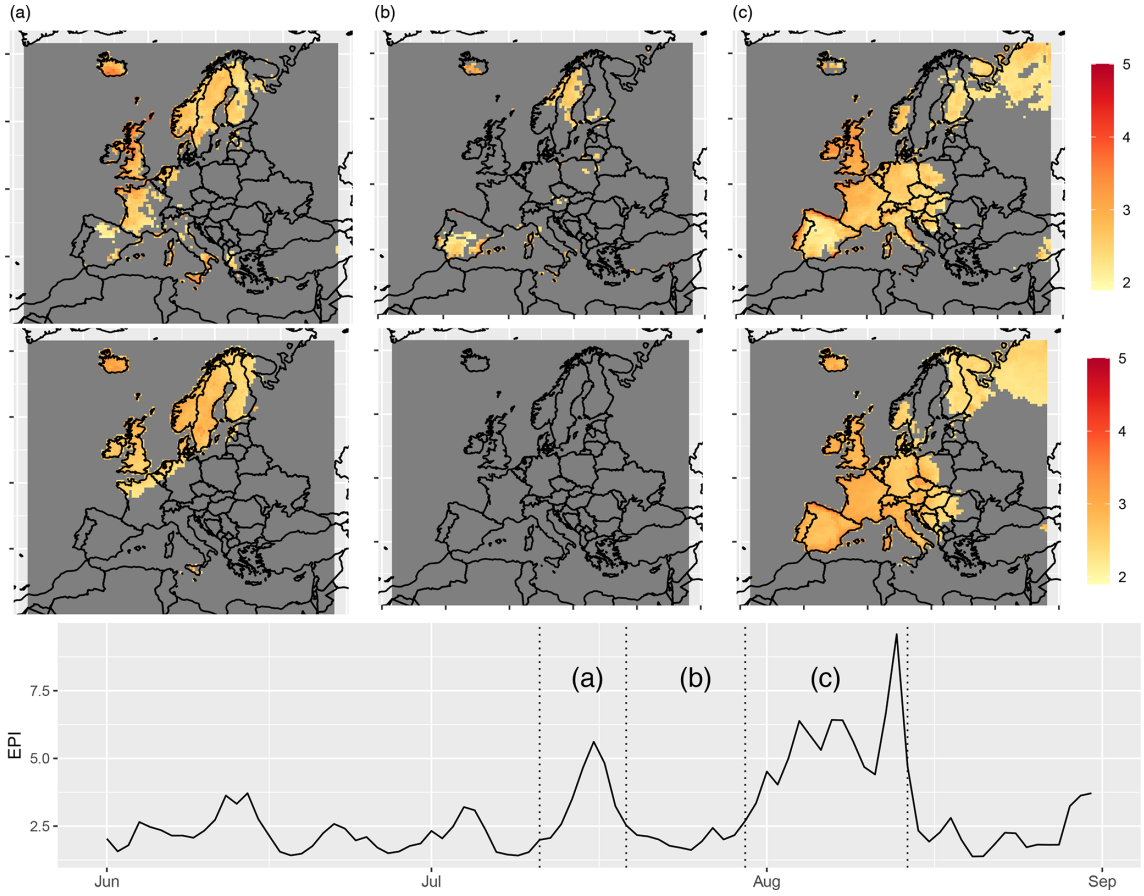


Figure 7. Overview of the temporal evolution of the 2003 heat wave. Top: pattern of mean anomalies within the periods **(a)** 11–19 July 2003, **(b)** 19–30 July 2003, and **(c)** 30 July–14 August 2003. We show the maxT2m anomalies (upper pattern) and the reconstruction, using the first 10 modes of PCs (lower pattern). All patterns use the same color scale and only show values exceeding the 98 % quantile, respectively. Bottom: EPI$^{\eta}$ as in Fig. 6 but for JJA months in 2003.

westward shift of the maxT2m anomalies, accompanied by a short drop in EPI$^{\eta}$ (b). Then the heat wave intensifies again for 2 d, with a new center over Belarus and Lithuania (c), before it collapses in mid-August. This particular feature of the 2010 heat wave has received little attention in the literature, although it is clearly visible in station measurements in, e.g., Moscow (Grumm, 2011). Consequentially, the Russian heat wave can also be well described by our method.

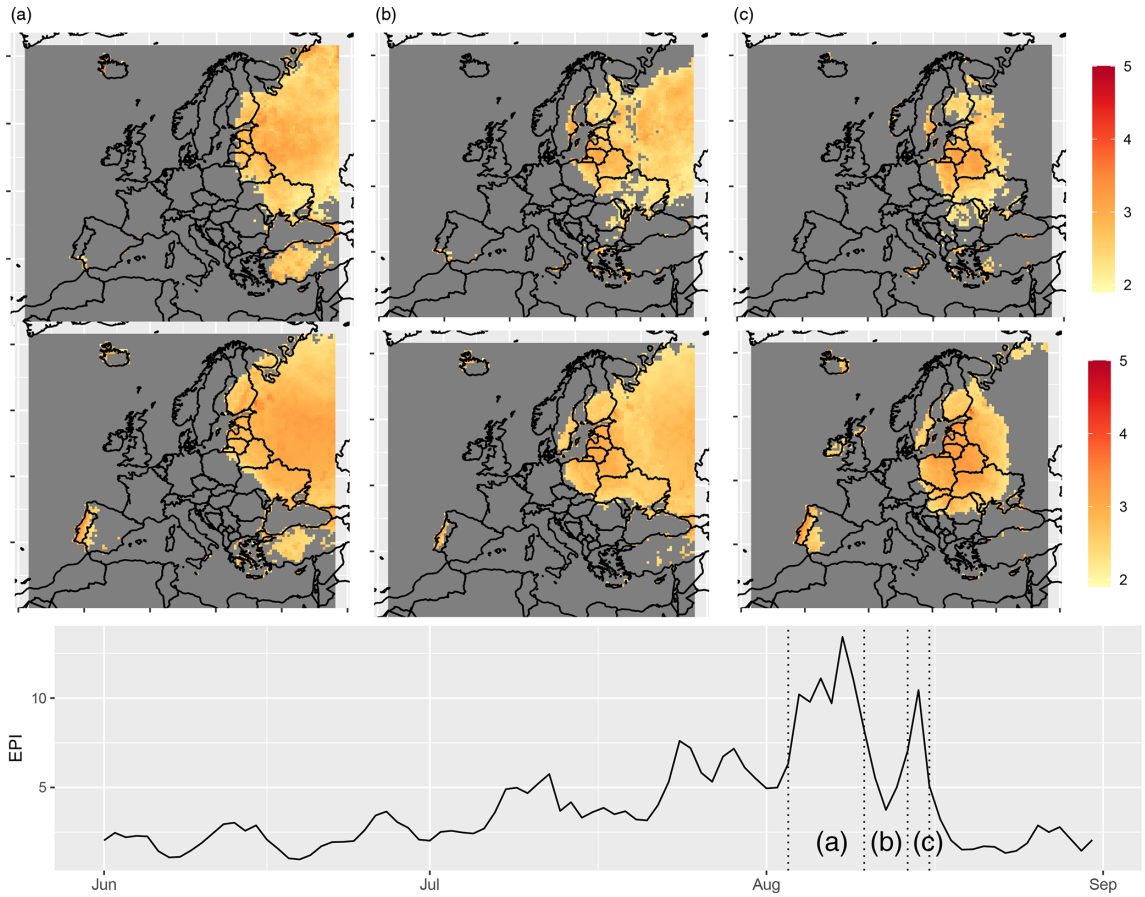


Figure 8. Overview of the temporal evolution of the 2010 heat wave. Top: pattern of mean anomalies within the periods **(a)** 3–10 August 2010, **(b)** 10–14 August 2010, and **(c)** 14–16 August 2010. We show the maxT2m anomalies (upper pattern) and the reconstruction, using the first 10 modes of PCs (lower pattern). All patterns use the same color scale and only show values exceeding the 98 % quantile, respectively. Bottom: EPI^{η} as in Fig. 6 but for JJA months in 2010.

To further evaluate the suitability of EPI^{η} for heat wave detection and representation, we compare event definitions using EPI^{η} with other heat wave definitions. A heat wave is usually defined on the basis of grid points (hereafter GRD event), e.g., a 5 d continuous exceedance of the 98 % quantile of maxT2m, with seasonal trends removed. A GRD event exists when a 5 d continuous exceedance occurs in at least 3 % of the grid points. Using maxT2m of COSMO-REA6, we obtain a number of 32 GRD events during the summer months in 1995–2019 in Europe. To rank heat wave events in terms of their "extremeness", DWD compares predefined heat waves (i.e. GRD events) based on the properties of duration, maximum number of grid points affected (GRDsize), and average magnitude of maxT2m anomalies (GRDmean)[3]. The time period of the 15 strongest GRD events with respect to GRDsize and GRDmean is displayed in black in Table 1.

We define an EPI event as a period when EPI^{η} exceeds the 80 % quantile for at least 5 consecutive days. The threshold is chosen so that we get a similar number of heat waves for the EPI events as for the GRD events, yielding a total of 28 heat waves. In Table 1, we compare the top 15 EPI events according to their average EPI^{η} (EPImean), with the top 15 GRD events according to their GRDsize and GRDmean. Of the top 15 GRD events, 3 (June 2002, August 2008, and August 2012) are not classified as EPI events. For each of these events, the maxT2m fields exceeding the 98 % quantile and the EPI^{η} are shown in the Appendix Fig. A2. We find EPI^{η} exceeds the 80 % quantile in all three GRD events, but the duration is less than 5 d in each case. Of the top 15 EPI events, 13 are identified as being among the top 15 GRD events in terms of GRDsize as well, suggesting that EPI^{η} is particularly successful in the identification of large-scale extremes. This is largely due to the property of PCA, which detects large-scale patterns mainly in the first few modes. Large-scale extremes are thus weighted more strongly than small-scale events.

The Cooley and Thibaud framework requires the assumption of independent and identically distributed samples.

[3]https://www.dwd.de/EN/ourservices/rcccm/int/rcccm_int_hwkltr.html (last access: 7 July 2022)

Table 1. Periods of strongest heat waves defined by EPI and GRD events. Periods indicated in black are within the top 15 strongest events using the respective index (EPImean, GRDsize, and GRDmean). Periods in bold are heat waves with a lower rank than 15, and underlined periods show no correspondence in the respective counterpart (EPI or GRD).

EPImean	GRDsize	GRDmean
1995-08-17–1995-08-22	**1995-08-17–1995-08-22**	1995-08-17 1995-08-22
1998-06-11–1998-06-21	1998-06-11–1998-06-21	**1998-06-11–1998-06-21**
	2002-06-08–2002-06-12	2002-06-08 2002-06-12
2003-07-14–2003-07-18	2003-07-14–2003-07-19	2003-07-14 2003-07-19
2003-07-31–2003-08-14	2003-07-31–2003-08-14	2003-07-31 2003-08-14
2006-06-02–2006-06-06	2006-06-02–2006-06-06	2006-06-02 2006-06-06
2006-06-18–2006-06-23	**2006-06-18–2006-06-23**	**2006-06-18–2006-06-23**
2006-07-06–2006-07-11	**2006-07-06 2006-07-10**	2006-07-06 2006-07-10
2007-07-15–2007-07-25	2007-07-15–2007-07-25	2007-07-15 2007-07-25
2007-08-14–2007-08-19	2007-08-13–2007-08-20	**2007-08-13–2007-08-20**
2007-08-21–2007-08-27	2007-08-22–2007-08-27	2007-08-22 2007-08-27
	2008-08-15 2008-08-19	**2008-08-15 2008-08-19**
2009-06-25–2009-06-29	**2009-06-25–2009-06-29**	2009-06-25 2009-06-29
2010-07-07–2010-08-17	2010-07-07–2010-07-14	**2010-07-07–2010-07-14**
2010-07-07–2010-08-17	2010-07-16–2010-08-17	**2010-07-16–2010-08-17**
	2012-08-17 –2012-08-21	2012-08-17 2012-08-21
2014-08-03–2014-08-07	**2014-08-03–2014-08-07**	**2014-08-03–2014-08-07**
2015-06-30–2015-07-08	2015-06-30–2015-07-08	2015-06-30 2015-07-08
2016-06-17–2016-06-27	**2016-06-17–2016-06-27**	**2016-06-17–2016-06-27**
2017-06-29–2017-07-04	2017-06-29–2017-07-04	2017-06-29 2017-07-04
2017-08-01–2017-08-06	**2017-08-01–2017-08-06**	2017-08-01 2017-08-06
2019-06-12–2019-06-16	2019-06-12–2019-06-16	**2017-08-01–2017-08-06**
2019-06-25–2019-07-02	2019-06-25–2019-07-02	2019-06-25 2019-07-02
2019-07-23–2019-07-27	2019-07-23–2019-07-27	2019-07-23 2019-07-27

This assumption is generally not given for daily values of maxT2m and PD. However, it turns out that our method is quite robust in this respect. We evaluate this in Appendix Fig. A3 using the EPI calculated on the one hand from the patterns of daily maxT2m data and on the other hand from the patterns of weekly maxima of maxT2m. It can be seen that the EPI is robust to small deviations in individual patterns. The range of EPI values changes slightly as the first 10 modes explain a higher proportion of the variance.

7.3 Decomposition of the cross-TPDM for maxT2m and PD

We have shown that the decomposition of the TPDM provides a set of spatial patterns to describe extremal dependencies in maxT2m and leads to a suitable definition of a heat wave index. We now turn to the cross-TPDM, which encodes the extremal covariability of maxT2m and PD. Commonly used indices, e.g., the standard precipitation index (SPI), are calculated on relatively long accumulation times from 3 to 48 months (McKee et al., 1993). However, so-called "flash droughts", e.g., occur instantaneously at sub-seasonal scales. They are characterized by rapid development and intensification of drought, usually accompanied by exceptionally high temperatures (see, e.g., Peters et al., 2002; Hunt et al., 2009; Cook et al., 2018). To capture this structure, Mo and Lettenmaier (2016), e.g., calculate PD using 5 d precipitation averages. Here, we want to capture both perspectives on the dynamics and the evolution of compound heat wave and drought events and therefore assess extremal covariability for different accumulation times from 11 to 47 d.

To allow for a computation of the cross-TPDM for many different accumulations times, we need to lower the computational costs by further reducing the grid size. For this purpose, we increase the neighborhood in the COSMO-REA6 grid to 21×21 grid points and take the respective maximum of PD and maxT2m as the value (in Sect. 7.2 we used a neighborhood of 7×7 grid points). With this approach, we obtain a grid of size 40×39. To ensure that no important information gets lost in this new reduction, in Appendix Fig. A4 we exemplarily show the EPI^{η} as calculated in Sect. 7.2 against the EPI^{η} on the highly reduced grid. The differences in the temporal behavior are small, and the EPI^{η} is robust to this further data reduction. Although not shown here, we have produced similar diagrams for the other indices we have defined, with similar results.

Not only the duration of a dry spell but also when it occurs in relation to the heat wave can be important for dynamic interpretation. For example, there is evidence that a dry phase can promote the development of an intense heat period (see,

e.g., Sousa et al., 2020). To account for different feedback mechanisms between PD and maxT2m, we introduce a time shift between PD and maxT2m, in the sense that PD leads maxT2m in time when there is a negative shift. The cross-TPDM and its decomposition using SVD are calculated for each time shift and accumulation time, respectively.

We assess statistical significance based on the singular values. To this end, we generate a 100-head bootstrap sample where we draw maxT2m and PD of a season independently so that the temporal relationship is destroyed while seasonality is preserved in the data. In Fig. 9, we show the first and the sum of the first and second singular values of the cross-TPDM, normalized by the respective average over the bootstrap samples. This is shown as a function of time shift and accumulation level. For values > 1 the measure of the extremal covariability between maxT2m and PD is larger than on average in the bootstrap sample. White dots indicate values that exceeded the 95 % confidence interval of the values in the bootstrap sample with "no statistical association".

For short accumulation times, maximum association is obtained for instantaneous anomalies. For higher accumulation times, the association is maximal for negative shifts when anomalies in PD precede the maxT2m anomalies. This indicates that a heat wave is preferentially formed when it is preceded by a drought period. Considering the sum of more singular values (see Appendix Fig. A5 for the first 10 singular values), this time-shifted dependence structure weakens but remains significant. We conclude that the statistical association between the large-scale patterns in matT2m and PD is significant and that the accumulation time of PD must be evaluated together with a temporal shift.

We now turn to the large-scale patterns of extremes in maxT2m and PD. In Fig. 10, we show the patterns for an accumulation time of 11 d and a time shift of 0 d. The large-scale patterns again show the characteristic wave-like structure previously discussed in Sect. 7.1. The corresponding anomalies in maxT2m and PD have a very similar structure, with the regions of the strongest gradients in PD slightly shifted compared to the gradients in maxT2m (second pattern – eastwards, third pattern – southwards). Although a physical interpretation of the SVs should always be approached with caution (see Sect. 3.3), it is possible that eastward-propagating Rossby waves (Sect. 7.1) could explain this pattern if, for example, strong gradients in the temperature fields are associated with frontal systems typically associated with precipitation. The leading patterns are also very similar for other accumulation times and time shifts.

7.4 Extremal pattern indices for compound maxT2m and PD extremes

As in Sect. 5 we base our calculation of an extremal pattern index on the expansion coefficients, where η_k represents the respective expansion coefficient of maxT2m and ν_k of PD. We calculate the $\mathrm{EPI}^{\eta,\nu}$ (Eq. 19) and the $\mathrm{TEPI}^{\eta,\nu}$ (Eq. 20) using η_k and ν_k of the leading 10 SVD modes of maxT2m and PC, respectively. As a threshold for the $\mathrm{TEPI}^{\eta,\nu}$, we use the 90 % quantile, respectively. This gives us a sufficient number of events without losing focus on the extremes. In this way, we obtain a flexible tool for the analysis of different event characteristics of compound extremes in maxT2m and PD.

Our investigation focuses on three cases with varying accumulation periods and lead times for PD anomalies: one with 11 d accumulations and immediate anomalies, another with 35 d accumulations and a PD anomaly lead time of 8 d, and a third case with 93 d accumulations and a PD anomaly lead time of 38 d. The first case represents short-term, coincident events, while the second and third cases represent monthly and seasonal PD driving a heat wave. In all cases, we find significant statistical dependence between maxT2m and PD (Fig. A5). Although not shown, we obtain a significant value of 1.13 for the third case 93 d accumulation and 38 d shift. Figure 11 displays the $\mathrm{TEPI}^{\eta,\nu}$ for all cases. The heat waves in 2003 and 2010 are indicated with colors.

Both the 2003 and 2010 heat waves show a strong signal in instantaneous TEPI. $\mathrm{TEPI}^{\eta,\nu}$ events coincide with heat wave events as indicated by the EPI^{η}, suggesting that both heat waves were accompanied by short-term PD. This indicates typical weather conditions during heat waves with clear skies, strong solar radiation, and little to no precipitation (Fischer et al., 2007).

Furthermore, the 2003 heat wave was preceded by a PD that was mainly evident on a seasonal timescale, whereas the 2010 heat wave was preceded by a more pronounced PD on a monthly timescale. In both cases, the high agreement of $\mathrm{TEPI}^{\eta,\nu}$ and EPI^{η} suggests that PD contributed to the extreme maxT2m during 2003 and 2010 (see also Fischer et al., 2007; Fink et al., 2004; Zaitchik et al., 2006; Christian et al., 2020).

Thus, the heat waves in 2003 and 2010 display different dynamics concerning common extremes of PD and maxT2m. Our TEPI definition effectively captures these characteristics, demonstrating that both events meet the IPCC definition of compound events as "two or more extreme events occurring simultaneously or sequentially" (Seneviratne et al., 2012).

We thus demonstrate that our definition of TEPI for different accumulation times and temporal shifts summarizes many characteristics of compound extremes and represents a suitable index.

8 Discussion and conclusion

In this study, we apply the promising approach of Cooley and Thibaud (2019) to describe extremal dependencies using the TPDM. The TPDM has comparable properties to the covariance matrix and can be decomposed similar to PCA while focusing on the tail behavior of the variables. To describe the dynamics of large-scale temperature extremes, we

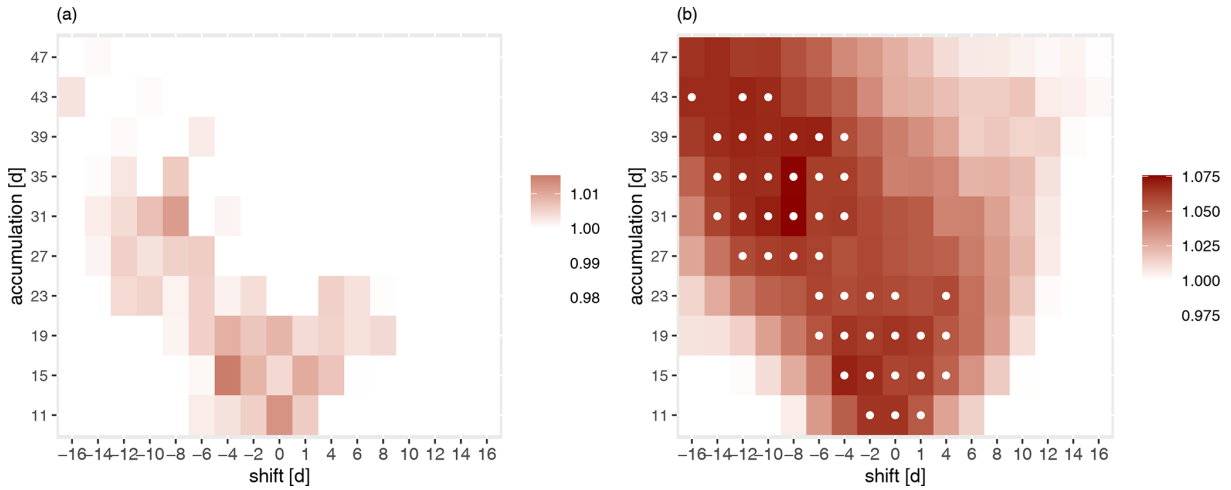


Figure 9. First and first + second singular values of the cross-TPDM calculated from maxT2m and PD in northern hemispheric summer months for varying shift and accumulation times. Each value has been normalized by the respective mean of 100 bootstrap samples where the temporal relation is destroyed. White dots indicate values exceeding the 95 % confidence interval of "no statistical relation".

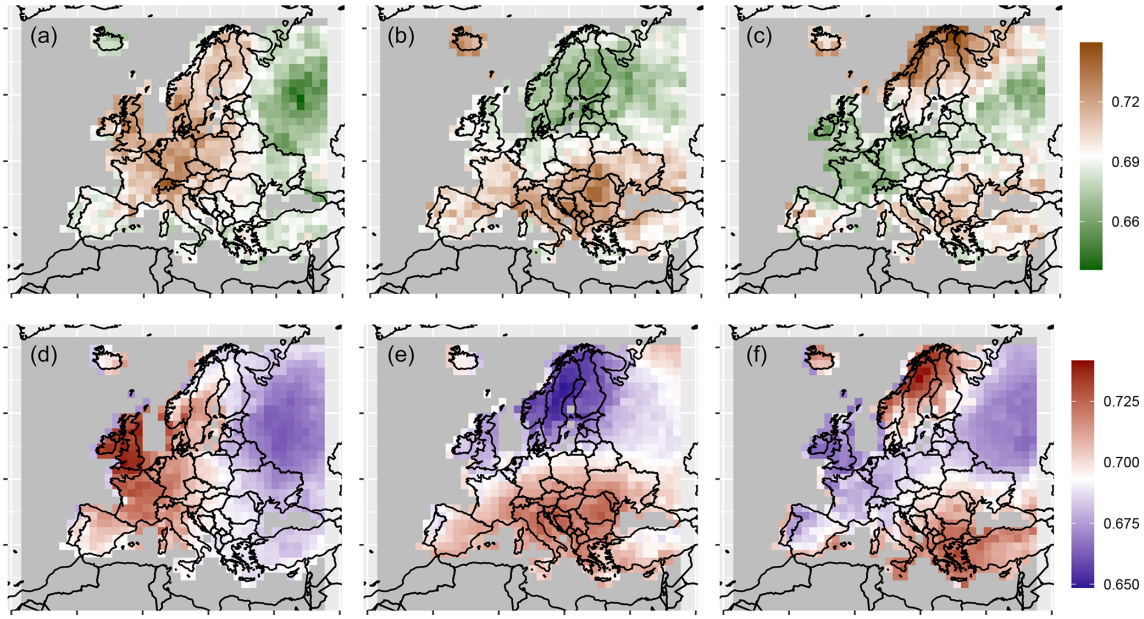


Figure 10. Second, third, and fourth left and right SVs of the cross-TPDM, associated with anomalies in maxT2m **(d, e, f)** and 11 d PD **(a, b, c)**. As in Fig. 3, we transformed each vector to the positive real numbers $\mathbb{R}^p \to \mathbb{R}^p_+$ such that $\tau(0) = \log(2)$ corresponds to zero.

calculated the TPDM based on maxT2m and identified typical temperature patterns of European heat waves. Our results are in line with previous studies that identify large-scale blocking events as the primary mechanism behind European heat waves (e.g. Schubert et al., 2014; Stefanon et al., 2012). We determine the optimal number of patterns to describe the main features of European heat waves and show exemplarily for the 2003 and 2010 heat waves that the description of the dynamics of large-scale temperature extremes is suitable with only the 10 leading eigenmodes.

We define the extremal pattern index (EPI) based on the time series of PCs and obtain a pattern-based spatial aggregation index. Thus, we can specifically identify spatially related extreme events with minimal reliance on pre-defined thresholds and regions of influence in grid point space. Although we cannot completely eliminate the use of a threshold, as it is a fundamental factor in estimating extreme dependencies based on the TPDM, all data, even those below the threshold, are included in the EPI estimator. We can show that this results in little information being lost with respect to the heat wave description.

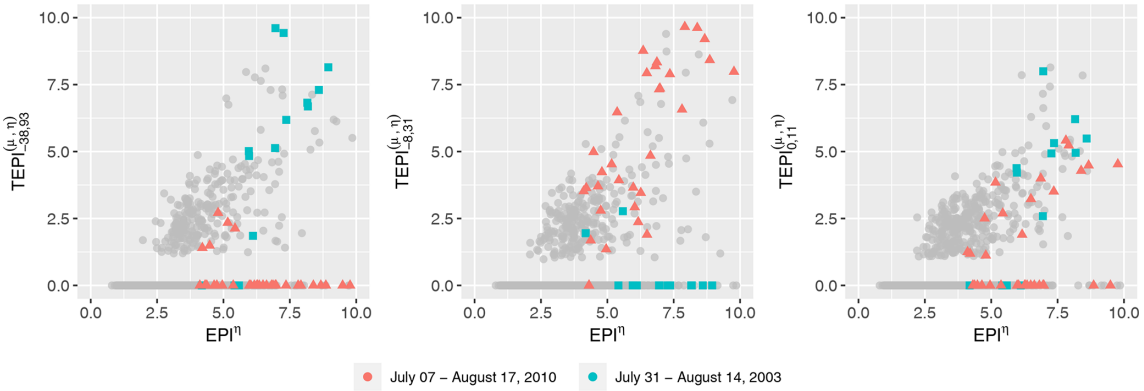


Figure 11. Scatter plot of EPI^{η} and $\mathrm{TEPI}^{\mu,\eta}$. $\mathrm{TEPI}^{\mu,\eta}$ is shown for calculation from PD using three different combinations of shift and accumulation (11 d accumulation, 0 d shift; 35 d accumulations, 8 d shift; and 93 d accumulation, 46 d shift). All indices were calculated using the first 10 PCs of the cross-TPDM from maxT2m and PD for northern hemispheric summer months. The 2010 and 2003 heat waves (31 July–14 August 2003; 7 July–17 August 2010), as identified in Table 1, are highlighted in blue and red, respectively.

Using the 2003 and 2010 European heat waves as examples, we examine how well the EPI captures the essential characteristics of heat waves. The temporal progression of both heat waves can be effectively illustrated using the EPI and aligns with the findings of previous analyses. Some characteristics, such as the collapse of the 2010 heat wave between 10 and 14 August, can be clearly identified using the EPI yet have received little exploration in the existing literature. An additional condition of heat waves besides spatial extent and intensity is usually temporal persistence. This condition can also be applied to EPI by defining heat waves. In Sect. 7, e.g., we define a heat wave (i.e., an EPI event) such that the EPI exceeds a threshold for at least 5 consecutive days. We propose an EPI-based definition of heat waves that, compared to a conventional definition, gives very consistent results. This shows that the EPI is suitable as an alternative to conventional heat wave indices with minimum reliance on user-defined thresholds. The EPI is thus a powerful tool for, e.g., attribution studies, climate monitoring, and characterization of extreme weather conditions.

Since pronounced heat waves over Europe cover a large part of the area studied, the spatial mean describes the course of the heat waves relatively well. Nevertheless, the higher modes of the PCs contribute significantly to the description of heat waves, as we discuss in Sect. 7.1. It can be assumed that the dominant influence of the first PC on the EPI would decrease when analyzing a larger geographical area.

To include the temporal persistence of heat waves as another condition, one could include only those patterns that exhibit a certain persistence and then aggregate them over time. However, too hard a condition could be detrimental.

After initially focusing on the description of heat waves using maxT2m, we extend our approach and go beyond the examination of individual variables. We deal with common extremes in maxT2m and PD. To this end, we propose a cross-TPDM model as an analog to the cross-covariance matrix that describes pairwise extremal dependencies between the two variables. Previous studies for compound events have usually referred to monthly or even seasonal variables (Wu et al., 2019; Hao et al., 2019). We calculate PD from daily precipitation using different duration scales. To generalize our approach and apply it to compound events as well, we shift the time series against each other. This allows us to investigate the dynamics of compound extremes in maxT2m and PD, even at relatively short duration levels. Using the singular values following an SVD, we analyze the dynamics of compound extremes in PD and maxT2m and show that preceding extremes in PD exert a significant influence on the evolution of heat waves (see, e.g., Christian et al., 2020). The investigation of additional variables such as hydrological drought indices and boundary layer heights would undoubtedly be interesting and may provide new insights into physical relationships.

Last but not least, we extend the EPI from considering a single variable to describing simultaneous extremes in two variables. We introduce a threshold-based estimator, the TEPI, which ensures simultaneous extremes of the underlying patterns. To demonstrate the utility of the TEPI, we again focus on the 2003 and 2010 heat waves and analyze their different dynamics in terms of compound extremes in maxT2m and PD using the TEPI. We analyze both heat waves using two event types: short-term coincident events and long-term precipitation deficit triggering a heat wave. In both cases, we can use TEPI to reconstruct the known characteristics of these events and demonstrate its effectiveness as a valuable tool for analyzing compound extremes. TEPI is a way to focus on specific events. A disadvantage of our formulation of TEPI is the hard condition of exceeding a threshold (see Fig. 11). We need to investigate whether a less strict selection or weighting would be beneficial.

There are several indicators for assessing climate indices (see, e.g., Keyantash and Dracup, 2002; Redmond, 1991).

Both EPI and TEPI are robust to application to different data sources, although we believe they are most successful in identifying large-scale extremes. In principle, they are applicable to point (e.g., measured data) and gridded data sets. The mathematical theory and derivation are quite complex, but the application is comparatively simple since existing functions for PCA and SVD, etc., can be used in various programming languages. Jiang et al. (2020) also make their R code for calculating TPDM, data pre-processing, and performing PCA available.

The cost-intensiveness of calculating TPDM based on high-resolution climate models is a universal problem that affects methods for dealing with spatial dependencies. This can possibly be solved by clever mathematical reformulation or extension of the working memory. As shown in Fig. A4, the data reduction approach we propose, using the spatial block maximum, provides a robust estimate of the EPI.

The TPDM assumes asymptotic dependence. However, we recognize that this assumption may not hold for maxT2m, PD, and their interdependencies. As Jiang et al. (2020), we nevertheless choose to decompose using the TPDM, as it opens up a new way to examine extreme patterns in maxT2m and PD. The resulting patterns are consistent with previous findings in the field and provide a meaningful dynamic context.

Appendix A

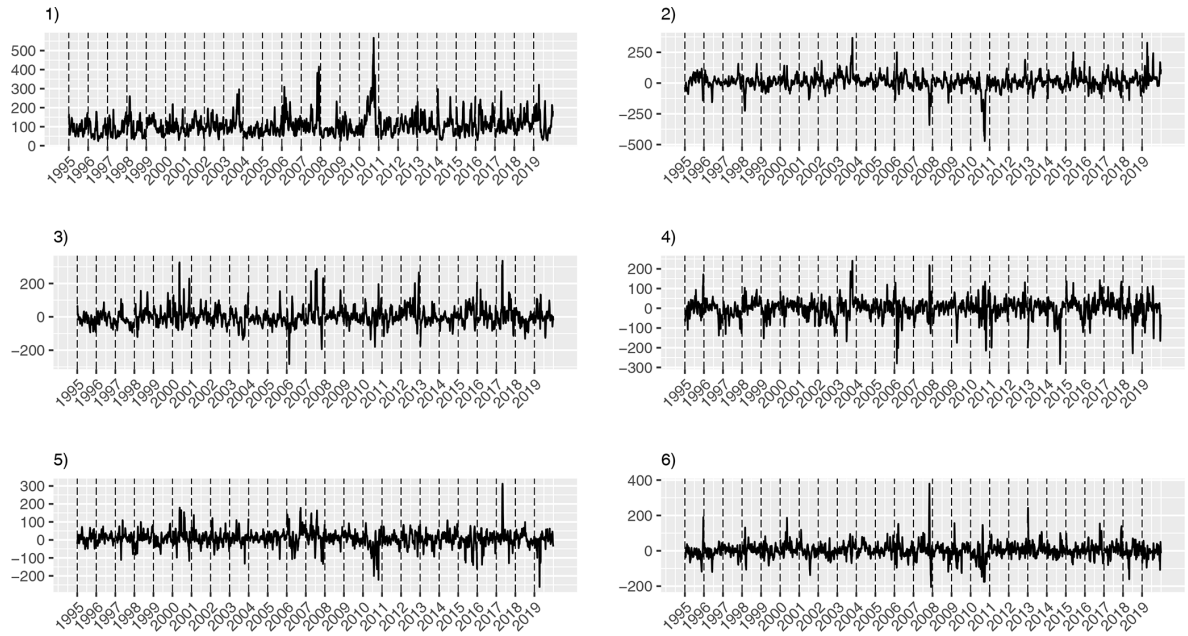


Figure A1. First six (from left to right, top to bottom) PCs of the TPDM as calculated from maxT2m in northern hemispheric summer months (JJA).

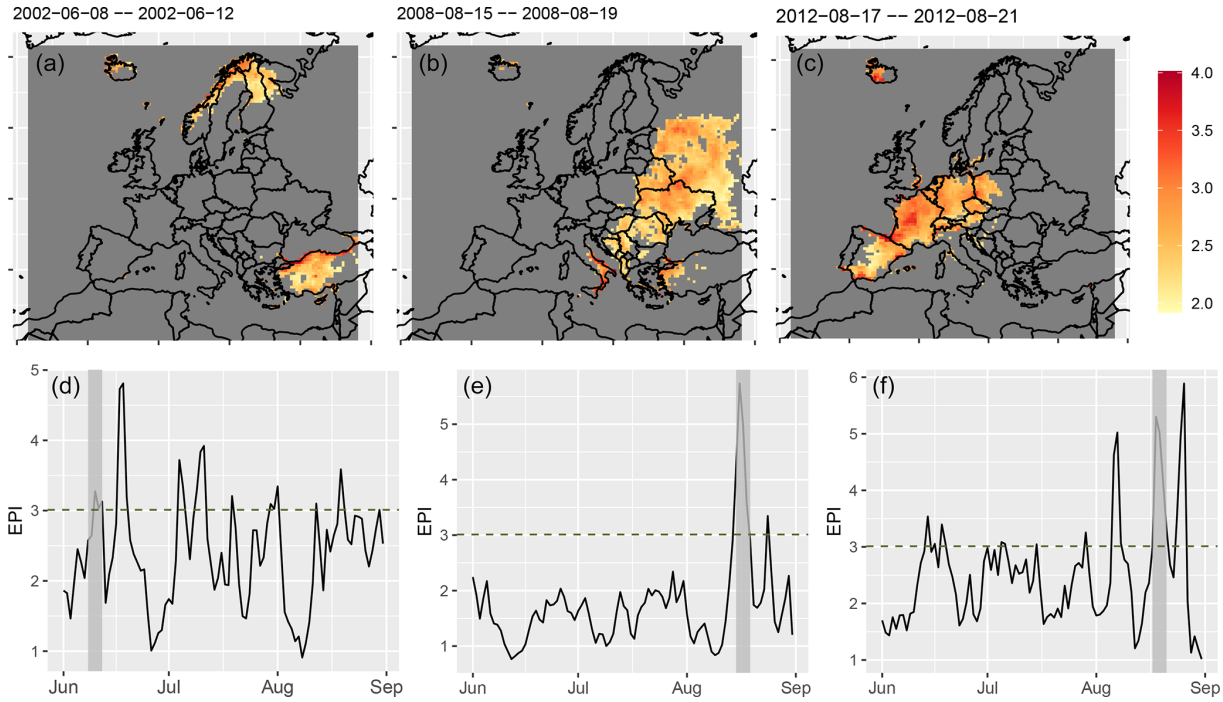


Figure A2. EPI$^{\eta}$ and maximum maxT2m anomaly of top 15 EPI events (according to EPImean) which are not identified as GRD events (Sect. 7.2). **(a, b, c)** Maximum maxT2m anomaly. Shown are only grid points exceeding the 98 % quantile. **(d, e, f)** EPI$^{\eta}$ as shown in Fig. 6 but for single years 2002, 2008, and 2012. Time periods of corresponding EPI events are marked in gray.

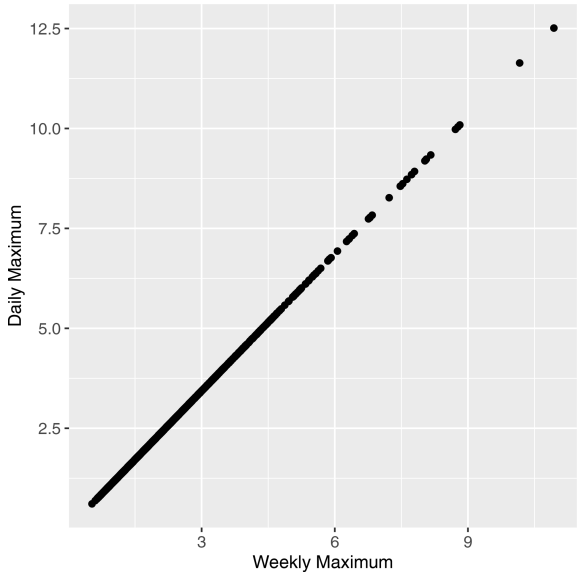


Figure A3. Scatter plot representation of EPI$^{\eta}$ calculated on EVs and eigenvalues of daily maxT2m and weekly maximum of maxT2m.

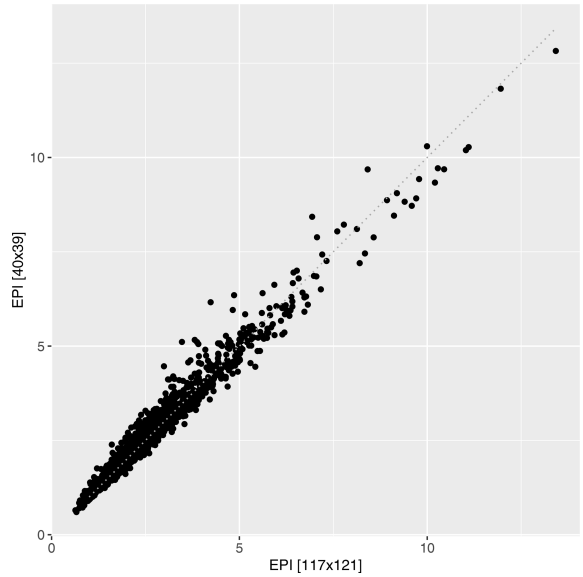


Figure A4. Scatter plot representation of EPI^{η} calculated on $[117 \times 121]$ grid (see Sect. 7.2) vs. EPI^{η} calculated on $[40 \times 39]$ grid (Sect. 3).

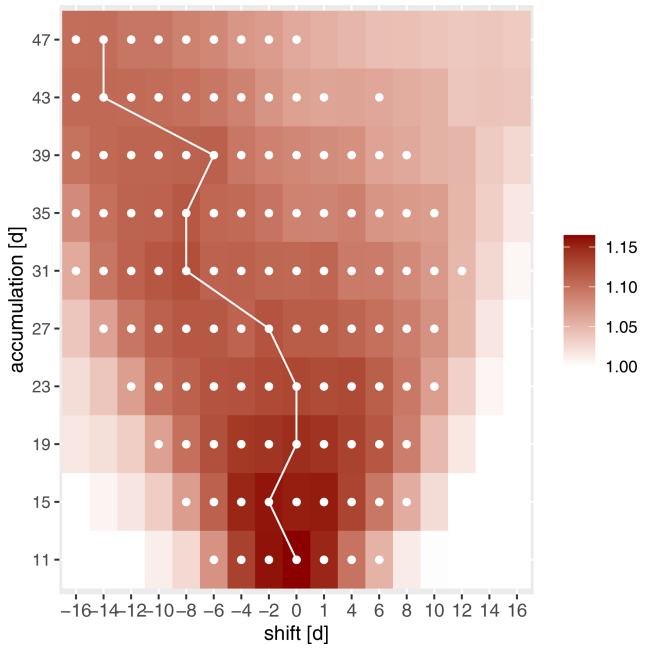


Figure A5. As in Fig. 9 but for the sum of first 10 singular values. The white line connects the maximum values over each accumulation stage.

Code availability. Code for data pre-processing and estimation of (cross-)TPDM and (T)EPI is provided as the R package "ExtrPatt". It is available at https://CRAN.R-project.org/package=ExtrPatt (Szemkus et al., 2023).

Data availability. The COSMO REA6 dataset is available via the DWD Climate Data Centre at https://opendata.dwd.de/climate_environment/REA/COSMO_REA6/ (DWD/HErZ, 2023).

Author contributions. SvS and PF developed the idea for this study. Data processing, analysis, and visualization were performed by SvS, who also led the writing, with contributions from PF. Both authors contributed to the discussion and interpretation of the results.

Competing interests. The contact author has declared that neither of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Special issue statement. This article is part of the special issue "Past and future European atmospheric extreme events under climate change". It is not associated with a conference.

Acknowledgements. This work was conducted as part of the ClimXtreme project Module B/CoDEx. We are grateful to our project partners Marco Oesting and Carolin Forster from the University of Stuttgart and Sebastian Buschow, University of Bonn, for valuable suggestions and fruitful discussions.

Financial support. This research has been supported by the Bundesministerium für Bildung und Forschung (ClimXtreme Module B/CoDEx) (grant no. 01LP1902A).

Review statement. This paper was edited by Dan Cooley and reviewed by three anonymous referees.

References

Abdi, H. and Williams, L. J.: Principal component analysis, Wires Comput. Stat., 2, 433–459, https://doi.org/10.1002/wics.101, 2010.

Anderson, H. R., Atkinson, R. W., Peacock, J., Marston, L., and Konstantinou, K.: Meta-analysis of time-series studies and panel studies of particulate matter (PM) and ozone (O3): report of a WHO task group, Tech. rep., Copenhagen: WHO Regional Office for Europe, https://iris.who.int/handle/10665/107557, 2004.

Barnston, A. G. and Livezey, R. E.: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns, Mon. Weather Rev., 115, 1083–1126, 1987.

Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M., and García-Herrera, R.: The hot summer of 2010: redrawing the temperature record map of Europe, Science, 332, 220–224, 2011.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J. L., De Waal, D. and Ferro, C.: Statistics of extremes: theory and applications, John Wiley & Sons, 2006.

Beniston, M.: The 2003 heat wave in Europe: A shape of things to come? An analysis based on Swiss climatological data and model simulations, Geophys. Res. Lett., 31, L02202, https://doi.org/10.1029/2003GL018857, 2004.

Björnsson, H. and Venegas, S.: A manual for EOF and SVD analyses of climatic data, CCGCR Report, 97, 112–134, 1997.

Black, E., Blackburn, M., Harrison, G., Hoskins, B., and Methven, J.: Factors contributing to the summer 2003 European heatwave, Weather, 59, 217–223, 2004.

Bollmeyer, C., Keller, J., Ohlwein, C., Wahl, S., Crewell, S., Friederichs, P., Hense, A., Keune, J., Kneifel, S., Pscheidt, I., Redl, S., and Steinke, S.: Towards a high-resolution regional reanalysis for the European CORDEX domain, Q. J. Roy. Meteor. Soc., 141, 1–15, 2015.

Boulaguiem, Y., Zscheischler, J., Vignotto, E., van der Wiel, K., and Engelke, S.: Modeling and simulating spatial extremes by combining extreme value theory with generative adversarial networks, Environmental Data Science, 1, 4, https://doi.org/10.1017/eds.2022.4, 2022.

Christian, J. I., Basara, J. B., Hunt, E. D., Otkin, J. A., and Xiao, X.: Flash drought development and cascading impacts associated with the 2010 Russian heatwave, Environ. Res. Lett., 15, 094078, https://doi.org/10.1088/1748-9326/ab9faf, 2020.

Coles, S., Bawa, J., Trenner, L., and Dorazio, P.: An introduction to statistical modeling of extreme values, vol. 208, Springer, https://doi.org/10.1007/978-1-4471-3675-0, 2001.

Cook, B. I., Mankin, J. S., and Anchukaitis, K. J.: Climate change and drought: From past to future, Current Climate Change Reports, 4, 164–179, 2018.

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

Dai, A., Zhao, T., and Chen, J.: Climate change and drought: a precipitation and evaporation perspective, Current Climate Change Reports, 4, 301–312, 2018.

Della-Marta, P. M., Haylock, M. R., Luterbacher, J., and Wanner, H.: Doubled length of western European summer heat waves since 1880, J. Geophys. Res.-Atmos., 112, D15103, https://doi.org/10.1029/2007JD008510, 2007.

Drees, H. and Sabourin, A.: Principal component analysis for multivariate extremes, arXiv [preprint], https://doi.org/10.48550/arXiv.1906.11043, 26 June 2019.

DWD/HErZ – Deutscher Wetterdienst – Climate Data Center/Hans Ertel Centre for Weather Research: COSMO-REA6 regional reanalysis, https://opendata.dwd.de/climate_environment/REA/COSMO_REA6/, last access: 20 December 2023.

Ferranti, L. and Viterbo, P.: The European summer of 2003: Sensitivity to soil water initial conditions, J. Climate, 19, 3659–3680, 2006.

Fink, A. H., Brücher, T., Krüger, A., Leckebusch, G. C., Pinto, J. G., and Ulbrich, U.: The 2003 European summer heatwaves and drought-synoptic diagnosis and impacts, Weather, 59, 209–216, 2004.

Fischer, E. M. and Knutti, R.: Observed heavy precipitation increase confirms theory and early models, Nat. Clim. Change, 6, 986–991, 2016.

Fischer, E. M. and Schär, C.: Consistent geographical patterns of changes in high-impact European heatwaves, Nat. Geosci., 3, 398–403, 2010.

Fischer, E. M., Seneviratne, S. I., Vidale, P. L., Lüthi, D., and Schär, C.: Soil moisture–atmosphere interactions during the 2003 European summer heat wave, J. Climate, 20, 5081–5099, 2007.

Fix, M. J., Cooley, D. S., and Thibaud, E.: Simultaneous autoregressive models for spatial extremes, Environmetrics, 32, e2656, https://doi.org/10.1002/env.2656, 2021.

García-Herrera, R., Díaz, J., Trigo, R. M., Luterbacher, J., and Fischer, E. M.: A review of the European summer heat wave of 2003, Crit. Rev. Env. Sci. Tec., 40, 267–306, 2010.

Gershunov, A. and Douville, H.: Extensive summer hot and cold spells under current and possible future climatic conditions: Europe and North America, Climate Extremes and Society, 74–98, https://doi.org/10.1017/Cbo9780511535840.008, 2009.

Goix, N., Sabourin, A., and Clémençon, S.: Sparse representation of multivariate extremes with applications to anomaly detection, J. Multivariate Anal., 161, 12–31, https://doi.org/10.1016/j.jmva.2017.06.010, 2017.

Griggs, D. J. and Noguer, M.: Climate change 2001: the scientific basis. Contribution of working group I to the third assessment report of the intergovernmental panel on climate change, Weather, 57, 267–269, 2002.

Grumm, R. H.: The central European and Russian heat event of July–August 2010, B. Am. Meteor. Soc., 92, 1285–1296, 2011.

Guerreiro, S. B., Dawson, R. J., Kilsby, C., Lewis, E., and Ford, A.: Future heat-waves, droughts and floods in 571 European cities, Environ. Res. Lett., 13, 034009, https://doi.org/10.1088/1748-9326/aaaad3, 2018.

Hao, Z., Hao, F., Singh, V. P., and Zhang, X.: Statistical prediction of the severity of compound dry-hot events based on El Niño-Southern Oscillation, J. Hydrol., 572, 243–250, 2019.

Higham, N. J.: Computing the nearest correlation matrix – a problem from finance, IMA J. Numer. Anal., 22, 329–343, 2002.

Hunt, E. D., Hubbard, K. G., Wilhite, D. A., Arkebauer, T. J., and Dutcher, A. L.: The development and evaluation of a soil moisture index, Int. J. Climatol., 29, 747–759, 2009.

IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/9781009157896, 2021.

Janßen, A. and Wan, P.: k-means clustering of extremes, Electron. J. Stat., 14, 1211–1233, 2020.

Jiang, Y., Cooley, D., and Wehner, M. F.: Principal Component Analysis for Extremes and Application to US Precipitation, J. Climate, 33, 6441–6451, 2020.

Jolliffe, I. T.: Principal components in regression analysis, in: Principal component analysis, 129–155, Springer, https://doi.org/10.1007/978-1-4757-1904-8_8, 1986.

Keyantash, J. and Dracup, J. A.: The quantification of drought: an evaluation of drought indices, B. Am. Meteor. Soc., 83, 1167–1180, 2002.

Kueh, M.-T. and Lin, C.-Y.: The 2018 summer heatwaves over northwestern Europe and its extended-range prediction, Sci. Rep.-UK, 10, 1–18, 2020.

Larsson, M. and Resnick, S. I.: Extremal dependence measure and extremogram: the regularly varying case, Extremes, 15, 231–256, 2012.

Lavaysse, C., Cammalleri, C., Dosio, A., van der Schrier, G., Toreti, A., and Vogt, J.: Towards a monitoring system of temperature extremes in Europe, Nat. Hazards Earth Syst. Sci., 18, 91–104, https://doi.org/10.5194/nhess-18-91-2018, 2018.

Liu, X., He, B., Guo, L., Huang, L., and Chen, D.: Similarities and differences in the mechanisms causing the European summer heatwaves in 2003, 2010, and 2018, Earth's Future, 8, e2019EF001386, https://doi.org/10.1029/2019EF001386, 2020.

Luterbacher, J., Dietrich, D., Xoplaki, E., Grosjean, M., and Wanner, H.: European seasonal and annual temperature variability, trends, and extremes since 1500, Science, 303, 1499–1503, 2004.

McKee, T. B., Doesken, N. J., and Kleist, J.: The relationship of drought frequency and duration to time scales, in: 8th Conf. Appl. Climatol., Anaheim, CA, https://climate.colostate.edu/pdfs/relationshipofdroughtfrequency.pdf (last access: 23 December 2023), 17–22 January 1993.

Mo, K. C. and Lettenmaier, D. P.: Precipitation deficit flash droughts over the United States, J. Hydrometeorol., 17, 1169–1184, 2016.

Morris, S. A., Reich, B. J., and Thibaud, E.: Exploration and inference in spatial extremes using empirical basis functions, J. Agr. Biol. Envir. St., 24, 555–572, 2019.

Newman, M. and Sardeshmukh, P. D.: A caveat concerning singular value decomposition, J. Climate, 8, 352–360, 1995.

Palmer, W. C.: Meteorological drought, vol. 30, US Department of Commerce, Weather Bureau, https://books.google.de/books?id=kyYZgnEk-L8C (last access: 23 December 2023), 1965.

Peters, A. J., Walter-Shea, E. A., Ji, L., Vina, A., Hayes, M., and Svoboda, M. D.: Drought monitoring with NDVI-based standardized vegetation index, Photogramm. Eng. Rem. S., 68, 71–75, 2002.

Redmond, K.: Climate monitoring and indices, Drought management and planning: international drought information center, Department of Agricultural Meteorology, Am. Metorol. Soc., 9, 29–33, 1991.

Resnick, S. I.: Heavy-tail phenomena: probabilistic and statistical modeling, Springer Science & Business Media, ISBN 0387450246, 9780387450247, 2007.

Robinson, P. J.: On the definition of a heat wave, J. Appl. Meteorol. Climatol., 40, 762–775, 2001.

Rohrbeck, C. and Cooley, D.: Simulating flood event sets using extremal principal components, Ann. Appl. Stat., 17, 1333–1352, 2023.

Rousi, E., Fink, A. H., Andersen, L. S., Becker, F. N., Beobide-Arsuaga, G., Breil, M., Cozzi, G., Heinke, J., Jach, L., Niermann, D., Petrovic, D., Richling, A., Riebold, J., Steidl, S., Suarez-Gutierrez, L., Tradowsky, J. S., Coumou, D., Düsterhus, A., Ellsäßer, F., Fragkoulidis, G., Gliksman, D., Handorf, D., Haustein, K., Kornhuber, K., Kunstmann, H., Pinto, J. G., Warrach-Sagi, K., and Xoplaki, E.: The extremely hot and dry 2018 summer in central and northern Europe from a multi-faceted weather and climate perspective, Nat. Hazards Earth Syst. Sci., 23, 1699–1718, https://doi.org/10.5194/nhess-23-1699-2023, 2023.

Saunders, K., Stephenson, A., and Karoly, D.: A regionalisation approach for rainfall based on extremal dependence, Extremes, 24, 215–240, 2021.

Schädler, G. and Breil, M.: Identification of droughts and heatwaves in Germany with regional climate networks, Nonlin. Processes Geophys., 28, 231–245, https://doi.org/10.5194/npg-28-231-2021, 2021.

Schubert, S., Wang, H., and Suarez, M.: Warm season subseasonal variability and climate extremes in the Northern Hemisphere: The role of stationary Rossby waves, J. Climate, 24, 4773–4792, 2011.

Schubert, S. D., Wang, H., Koster, R. D., Suarez, M. J., and Groisman, P. Y.: Northern Eurasian heat waves and droughts, J. Climate, 27, 3169–3207, 2014.

Seneviratne, S., Nicholls, N., Easterling, D., Goodess, C., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., Reichstein, M., Sorteberg, A., Vera, C., and Zhang, X.: Changes in climate extremes and their impacts on the natural physical environment, in: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation, edited by: Field, C. B., Barros, V., Stocker, T. F., Qin, D., Dokken, D. J., Ebi, K. L., Mastrandrea, M. D., Mach, K. J., Plattner, G.-K., Allen, S. K., Tignor, M., and Midgley, P. M., A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC), Cambridge University Press, Cambridge, UK, and New York, NY, USA, 109–230, 2012.

Slater, L. J., Anderson, B., Buechel, M., Dadson, S., Han, S., Harrigan, S., Kelder, T., Kowal, K., Lees, T., Matthews, T., Murphy, C., and Wilby, R. L.: Nonstationary weather and water extremes: a review of methods for their detection, attribution, and management, Hydrol. Earth Syst. Sci., 25, 3897–3935, https://doi.org/10.5194/hess-25-3897-2021, 2021.

Szemkus, S., Cooley, D., and Jiang, Y.: ExtrPatt: Spatial Dependencies and Indices for Extremes, https://CRAN.R-project.org/package=ExtrPatt (last access: 20 December 2023), 2023.

Sousa, P. M., Barriopedro, D., García-Herrera, R., Ordóñez, C., Soares, P. M., and Trigo, R. M.: Distinct influences of large-scale circulation and regional feedbacks in two exceptional 2019 European heatwaves, Commun. Earth & Environ., 1, 1–13, 2020.

Stefanon, M., D'Andrea, F., and Drobinski, P.: Heatwave classification over Europe and the Mediterranean region, Environ. Res. Lett., 7, 014023, https://doi.org/10.1088/1748-9326/7/1/014023, 2012.

Ummenhofer, C. C. and Meehl, G. A.: Extreme weather and climate events with ecological relevance: a review, Philos. T. Roy. Soc. B, 372, 20160135, https://doi.org/10.1098/rstb.2016.0135, 2017.

Vignotto, E., Engelke, S., and Zscheischler, J.: Clustering bivariate dependencies of compound precipitation and wind extremes over Great Britain and Ireland, Weather and Climate Extremes, 32, 100318, https://doi.org/10.1016/j.wace.2021.100318, 2021.

Wallace, J. M., Smith, C., and Bretherton, C. S.: Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies, J. Climate, 5, 561–576, 1992.

Wilks, D. S.: Statistical methods in the atmospheric sciences, vol. 100, Academic press, ISBN 0123850223, 9780123850225, 2011.

Wu, X., Hao, Z., Hao, F., Singh, V. P., and Zhang, X.: Dry-hot magnitude index: A joint indicator for compound event analysis,

Environ. Res. Lett., 14, 064017, https://doi.org/10.1088/1748-9326/ab1ec7, 2019.

Zaitchik, B. F., Macalady, A. K., Bonneau, L. R., and Smith, R. B.: Europe's 2003 heat wave: a satellite view of impacts and land–atmosphere feedbacks, Int. J. Climatol., 26, 743–769, 2006.

Zscheischler, J. and Seneviratne, S. I.: Dependence of drivers affects risks associated with compound events, Sci. Adv., 3, e1700263, https://doi.org/10.1126/sciadv.1700263, 2017.

A.2. R-Package: "ExtrPatt"

Package 'ExtrPatt'

November 7, 2023

Type Package

Title Spatial Dependencies and Indices for Extremes

Version 0.1-4

Author Svenja Szemkus [aut, cre],
Dan Cooley [ctb],
Yuing Jiang [ctb]

Description An implementation of
1) the tail pairwise dependence matrix (TPDM) as described in Jiang & Cooley (2020) <doi:10.1175/JCLI-D-19-0413.1>
2) the extremal pattern index (EPI) as described in Szemkus & Friederichs ('Spatial patterns and indices for heatwave and droughts over Europe using a decomposition of extremal dependency'; submitted to ASCMO 2023).

Depends R (>= 3.5.0)

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Imports Matrix, doParallel, stats, foreach, MASS, parallel, utils

RoxygenNote 7.2.3

NeedsCompilation no

Maintainer Svenja Szemkus <sszemkus@uni-bonn.de>

Repository CRAN

Date/Publication 2023-11-07 18:50:05 UTC

R topics documented:

compute.EPI 2
decls 3
est.tpdm 4
invTrans 5
pca.tpdm 6
precipGER 7

svd.tpdm 7
to.alpha.2 8
trans 8
wrapper.EPI 9

Index **11**

`compute.EPI` *Estimation of EPI*

Description

Estimates the extremal pattern index (EPI) from either the 'm' principle components after a PCA or left- and right expansion coefficients after an SVD. In case of a SVD, the threshold-based EPI (TEPI) can optionally be calculated.

Usage

```
compute.EPI(coeff, m = 1:10, q = 0.98)
```

Arguments

`coeff`	A list, containing the t x n dimensional principle components/expansion coefficients of TPDM. Can also be output of function 'est.tpdm'.
`m`	numeric vector: Containing the Principle Components from which EPI shall be computed (e.g. with modes = c(1:10), the EPI is calculated on first ten principle components)
`q`	Optional: A threshold for computation of TEPI

Details

Given the first 'm' modes of principle components u and eigenvalues after a PCA, the EPI is given as:

$$EPI_t^u = \sqrt{\sum_{k=1}^{m}(u_{t,k}^2)/\sum_{j=1}^{m} e_j}.$$

Given the first 'm' modes of expansion coefficients u and v and singular values e after a SVD, the EPI and TEPI are given as:

$$EPI_t^{u,v} = \sqrt{\sum_{k=1}^{m}(u_{t,k}^2 + v_{t,k}^2)/\sum_{j=1}^{m} e_j}.$$

$$TEPI_t^{u,v} = \sqrt{(\sum_{k=1}^{m}(u_{t,k}^2 + v_{t,k}^2)/\sum_{j=1}^{m} e_j)|_{(|u_{t,k}|>q_u, |vt,k|>q_v)}}.$$

Value

An array of length t, containing EPI. TEPI is computed if if q > 0.

References

Szemkus & Friederichs (2023)

Examples

```
data    <- precipGER

data.alpha2  <- to.alpha.2(data$pr)
Sigma   <- est.tpdm(data.alpha2,anz_cores =1)
res.pca <- pca.tpdm(Sigma, data.alpha2)
EPI <- compute.EPI(res.pca, m = 1:10)

plot(data$date, EPI, type='l')
```

---

decls *Declustering*

Description

Declustering routine, which will can be applied on radial component r in estimation of the TPDM. Subroutine of est.tpdm.

Usage

```
decls(x, th, k)
```

Arguments

x	Real vector
th	Threshold
k	Cluster length

Value

numeric vector of declustered threshold exceedances

Author(s)

Yuing Jiang, Dan Cooley

References

Jiang & Cooley (2020) <doi:10.1175/JCLI-D-19-0413.1>

See Also

est.tpdm

`est.tpdm` *Estimation of TPDM*

Description

Estimation of tail pairwise dependence matrix (TPDM)

Sub-Routine of est.row.tpdm. Calculates one element of the TPDM

Usage

```
est.tpdm(X, Y = NULL, anz_cores = 1, clust = NULL, q = 0.98)

est.row.tpdm(x, Y, clust = NULL, q = 0.98)

est.element.tpdm(x, y, clust = NULL, q = 0.98)
```

Arguments

`X`	A t x n dimensional, numeric data-matrix with t: Number of time steps and n: Number of grid points/stations
`Y`	A t x n dimensional, numeric Data-matrix with t: Number of time steps and n: Number of grid points/stations
`anz_cores`	Number of cores for parallel computing (default:1); Be careful not to overload your computer!
`clust`	Optional: If clust = NULL, no declustering is performed. Else, declustering according to cluster-length 'clust'.
`q`	Threshold for computation of TPDM. Only data above the 'q'-quantile will be used for estimation. Choose such that 0<q<1.
`x`	Array of length t, where t is the number of time steps
`y`	Same as x

Details

Given a random vector X with components $x_{t,i}, x_{t,j}$ with $i, j = 1, \ldots, n$ and it's radial component $r_{t,ij} = \sqrt{x_{t,i}^2 + x_{t,j}^2}$ and angular components $w_{t,i} = x_{t,i}/r_{t,ij}$ and $w_{t,j} = x_{t,j}/r_{t,ij}$, the i'th,j'th element of the TPDM is estimated as:

$$\hat{\sigma}_{ij} = 2n_{ij,exc}^{-1} \sum_{t=1}^{n} w_{t,i} w_{t,j}|_{(r_{t,ij} > r_{0,ij})}$$

. Given two random vectors X and Y with components $x_{t,i}, y_{t,j}$ with $i, j = 1, \ldots, n$, and it's radial component $r_{t,ij} = \sqrt{x_{t,i}^2 + y_{t,j}^2}$ and angular components $w_{t,i}^x = \frac{x_{t,i}}{r_{t,ij}}; w_{t,j}^y = \frac{y_{t,j}}{r_{t,ij}}$, the i'th,j'th element of the cross-TPDM is estimated as:

$$\hat{\sigma}_{ij} = 2n_{exc}^{-1} \sum_{t=1}^{n} w_{t,i}^x w_{t,j}^y |_{(r_{t,ij} > r_{0,ij})}$$

.

Value

An n x n matrix, containing the estimate of the TPDM

Array containing the estimate of one row of the TPDM.

Value containing the estimate of one element of the TPDM.

References

Jiang & Cooley (2020) <doi:10.1175/JCLI-D-19-0413.1>; Szemkus & Friederichs (2023)

Examples

```
data    <- precipGER

data.alpha2       <- to.alpha.2(data$pr)
Sigma   <- est.tpdm(data.alpha2,anz_cores =1)
```

`invTrans` *Transformation function*

Description

Applies the inverse transformation $t^{-1}(v) = \log(\exp(v) - 1)$

Usage

```
invTrans(v)
```

Arguments

v	Real, positive vector

Details

Transformation from real, positive vector in real vector under preservation of frechet-distribution.

Value

Real vector, containing the result of inverse transformation function.

Author(s)

Yuing Jiang, Dan Cooley

References

Cooley & Thibaud (2019) <doi:10.1093/biomet/asz028>

See Also

svd.tpdm, pca.tpdm

`pca.tpdm` *Principal Component Analysis for TPDM*

Description

Calculates principal component analysis (PCA) of given TPDM

Usage

```
pca.tpdm(Sigma, data)
```

Arguments

`Sigma`	A n x n data array, containing the TPDM, can be output of est.tpdm.
`data`	A t x n dimensional, numeric Data-matrix with t: Number of time steps and n: Number of grid points/stations.

Value

list containing

- pc: The Principal Components of TPDM
- basis: The Eigenvectors of TPDM
- extremal.basis: The Eigenvectors of TPDM but transformed in positive reals with trans

Author(s)

Yuing Jiang, Dan Cooley

References

Jiang & Cooley (2020) <doi:10.1175/JCLI-D-19-0413.1>

precipGER *daily Precipitation over Southern Germany*

Description

Daily Precipitation at several stations in Germany

Usage

```
data(precipGER)
```

Format

A list containing containing

- pr: data-array
- date: time-information
- lon,lat: longitude & latitude information

Details

Daily Precipitation Data

Daily precipitation data from several wather station in southern Germany (longitude <50) over the years 2000-2019. The data has been downloaded from opendata server of german weather service (https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/daily/kl/historical/).

Source

Quelle: Deutscher Wetterdienst

svd.tpdm *Singular Value decomposition for cross-TPDM*

Description

Calculates singular value decomposition (SVD) of given cross-TPDM

Usage

```
svd.tpdm(Sigma, X, Y)
```

Arguments

Sigma	A n x n data array, containing the cross-TPDM, can be output of est.tpdm.
X	A t x n dimensional, numeric Data-matrix with t: Number of time steps and n: Number of grid points/stations.
Y	Same as X but for second variable.

Value

List containing

- pcU, pcV: The left- and right expansion coefficients of cross-TPDM
- U, V: The left- and right singular Vectors of cross-TPDM
- extr.U, extr.V: The left- and right singular vectors of cross-TPDM, but transformed in positive reals with trans

---

`to.alpha.2` *Probability integral transformation*

Description

Performs transformation to make all of the margins follow a Frechet distribution with tail-index alpha = 2.

Usage

```
to.alpha.2(data, orig = NULL)
```

Arguments

`data`	A t x n dimensional, numeric Data-matrix with t: Number of time steps and n: Number of grid points/stations
`orig`	If known: original distribution of data (currently implemented: 'normal' or 'gamma'), else: NULL

Value

Data-matrix of same dimension as 'data', but in Frechet-margins with tail-index 2

`trans` *transformation function*

Description

Applies the transformation $t(x) = \log(1 + \exp(x))$

Usage

```
trans(x)
```

Arguments

`x`	Real vector

Details

Transformation from real vector in real, positive vector under preservation of Frechet-distribution.

Value

Real, positive vector, containing the result of transformation function.

Author(s)

Yuing Jiang, Dan Cooley

References

Cooley & Thibaud (2019) <doi:10.1093/biomet/asz028>

See Also

svd.tpdm, pca.tpdm

`wrapper.EPI` *Wrapper function*

Description

Handles all steps for estimation of EPI from raw-data: 1) Preprocessing into Frechet-Margins 2) Estimation of TPDM 3) Calculation of Principal Components 4) Estimation of EPI

Usage

```
wrapper.EPI(
  X,
  Y = NULL,
  q = 0.98,
  anz_cores = 1,
  clust = NULL,
  m = 1:10,
  thr_EPI = NULL
)
```

Arguments

`X`	A t x n dimensional Data-matrix with t: Number of time steps and n: Number of grid points/stations
`Y`	Optional: Sames as X but for second variable: If Y!=NULL, cross-TPDM instead of TPDM and SVD instead of PCA is computed
`q`	Threshold for computation of TPDM. Only data above the 'q'-quantile will be used for estimation. Choose such that $0 < q < 1$.

anz_cores	Number of cores for parallel computing (default: 5)
clust	Optional_ Uf clust = NULL, no declustering is performed. Else, declustering according to cluster-length 'clust'
m	Numeric vector: Containing the principal components/expansion coefficients (in case of Y!=NULL) from which the EPI shall be computed (default: modes = c(1:10), calculates the EPI on first ten principle Components)
thr_EPI	Only if Y!=NULL: Threshold for computation of TEPI. Expansion-coefficients that exceed the 'q'-quantile will be used for estimation. Choose such that 0 < q < 1.

Value

In case of Y =NULL: A list containing:

- basis: The Eigenvectors of TPDM
- pc: The principal components of TPDM
- extremal.basis: The Eigenvectors of TPDM but transformed in positive reals with trans
- EPI: Extremal pattern index

In case of Y !=NULL: A list containing:

- U, V: The left- and right singular Vectors of cross-TPDM
- extr.U, extr.V: The left- and right singular vectors of cross-TPDM, but transformed in positive reals with trans
- pcU, pcV: The left- and right expansion coefficients of cross-TPDM
- EPI: Extremal pattern index
- TEPI: Threshold-based extremal pattern index

References

Szemkus & Friederichs 2023

Examples

```
data    <- precipGER

result  <- wrapper.EPI(data$pr, m = 1:50)

rbPal <- colorRampPalette(c('blue', 'white','red'))
Col <- rbPal(10)[as.numeric(cut(result$basis[,2],breaks = 10))]
plot(data$lat, data$lon,col=Col)
plot(data$date, result$EPI, type='l')
```

Index

∗ **datasets**
precipGER, 7

compute.EPI, 2

decls, 3

est.element.tpdm *(*est.tpdm*)*, 4
est.row.tpdm, *4*
est.row.tpdm *(*est.tpdm*)*, 4
est.tpdm, *3, 4*, 4, *6, 7*

invTrans, 5

pca.tpdm, *6*, 6, *9*
precipGER, 7

svd.tpdm, *6*, 7, *9*

to.alpha.2, 8
trans, *6*, *8*, 8, *10*

wrapper.EPI, 9

B. Appendix Spectral Filter

B.1. Clausius-Clapeyron Relation

This appendix section provides a short derivation of the Clausius-Clapeyron (CC) relation, which was originally introduced by Trenberth et al., 2003. It states that for every degree of warming, the atmosphere can retain about 7% more water vapor, which is proportional to an upper bound on total precipitable water. The CC relation goes back to the famous Clausius-Clapeyron equation formulated by Émile Clapeyron and Rudolf Clausius in 1834. The derivation, uses the liquid-gas approximation of the CC equation assuming that:

1. the density of water vapour is much lower than the density of water.
2. water vapour behaves like an ideal gas.
3. latent heat vaporization L is a constant and i.e. does not depend on temperature.

Where the CC equation takes the form:

$$\frac{de_s}{dT} = \frac{Le_s}{RT^2} \tag{B.1}$$

where e_s is saturation vapor pressure, L is latent heat vaporization and R, the ideal gas constant. To derive the relation between saturation vapour pressure and temperature, equation B.1 is integrated from T_0 to T which can be expressed as:

$$\frac{1}{T} \sim \frac{1}{T_0} - \frac{\Delta T}{T_0^2} \tag{B.2}$$

Thus, for $\Delta T = 1°\text{C}$, the percentage increase is approximately $\frac{L}{RT_0^2}$. Inserting typical values as follows: $T_0 = 273\text{K}$, $L \approx 2.5 \cdot 10^6 \frac{J}{kg}$ and $R \sim 461 \frac{J}{kgK}$ results in:

$$\frac{L}{RT_0^2} \sim 0.07 = 7\% \tag{B.3}$$

B.2. Additional Figures

B.2.1. Density plots from hourly precipitation by season

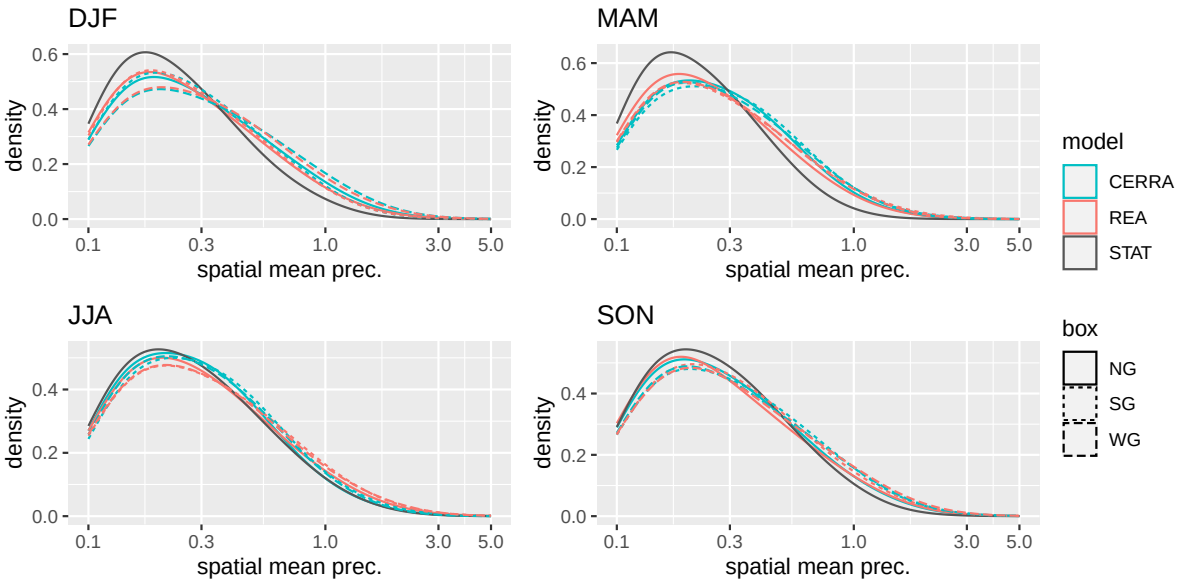


Figure B.1.: Same as Fig. 3.13, but for the spatial mean and divided by season.

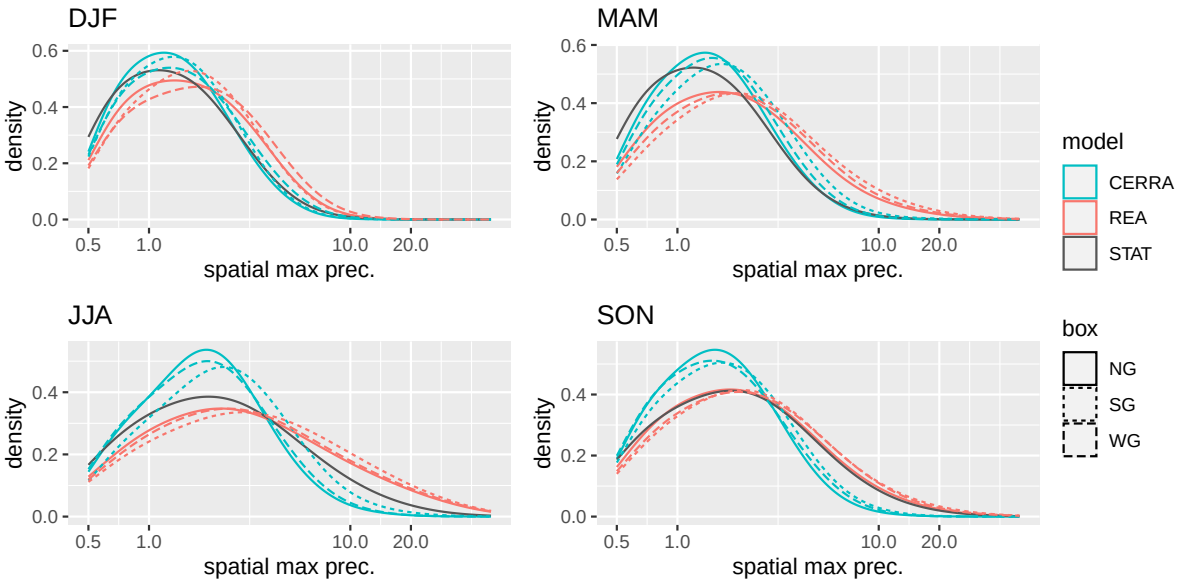


Figure B.2.: Same as Fig. 3.13, but for the spatial maximum and divided by season.

B.2.2. Scaling estimates for CERRA & COSMO-REA6, JJA & DJF

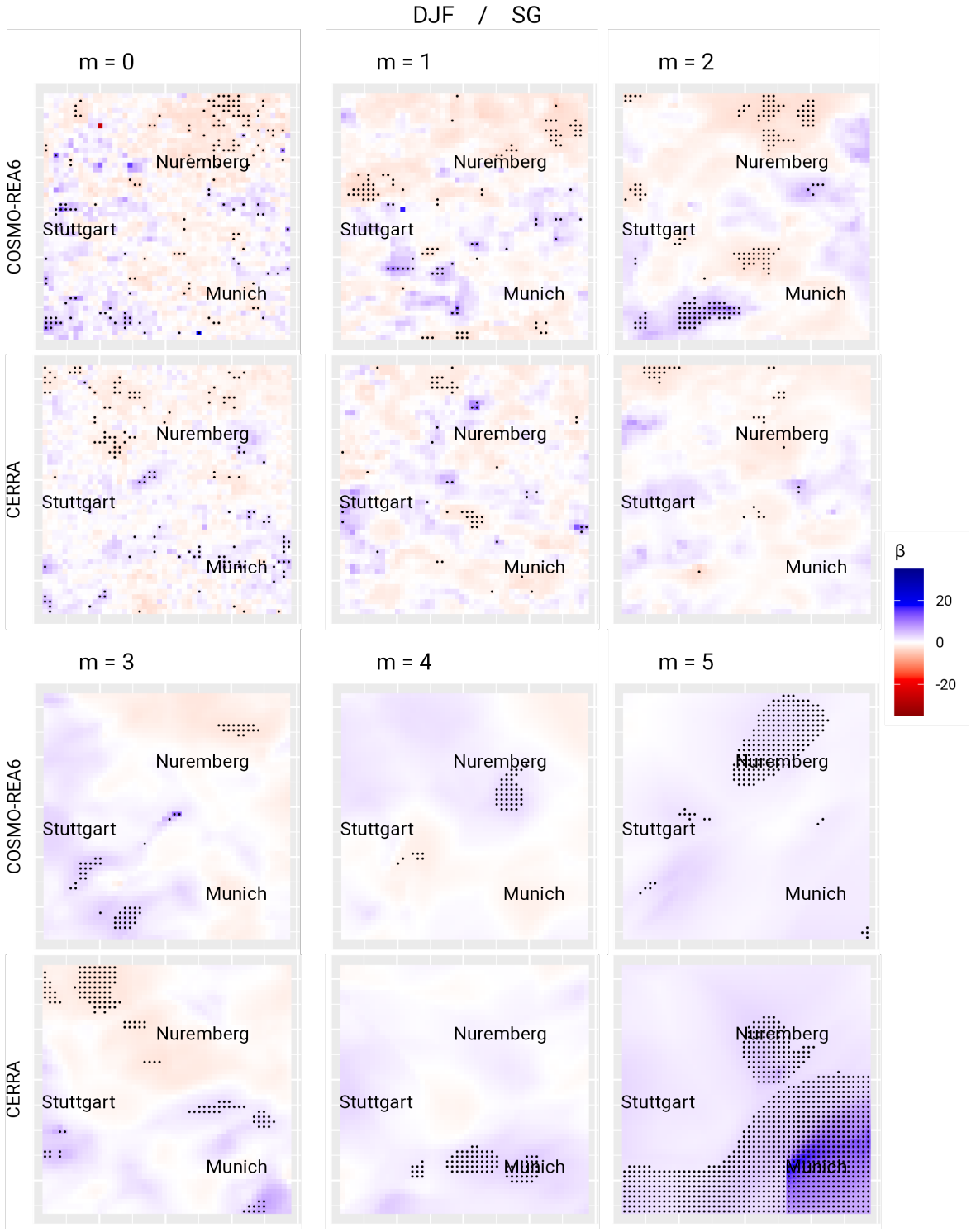


Figure B.3.: Same as Fig. 3.20 but for DJF months.

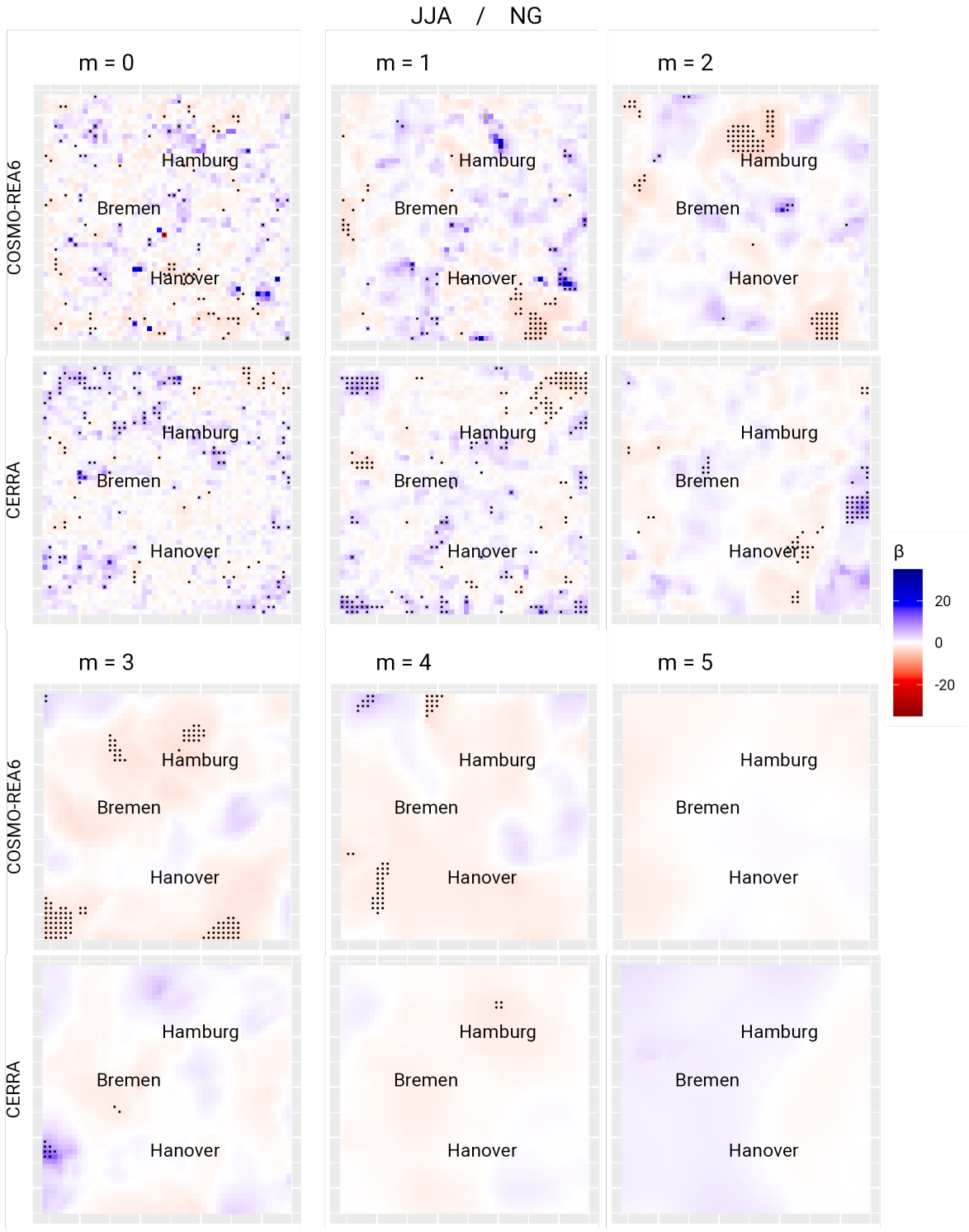


Figure B.4.: Same as Fig. 3.20 but for JJA months in NG.

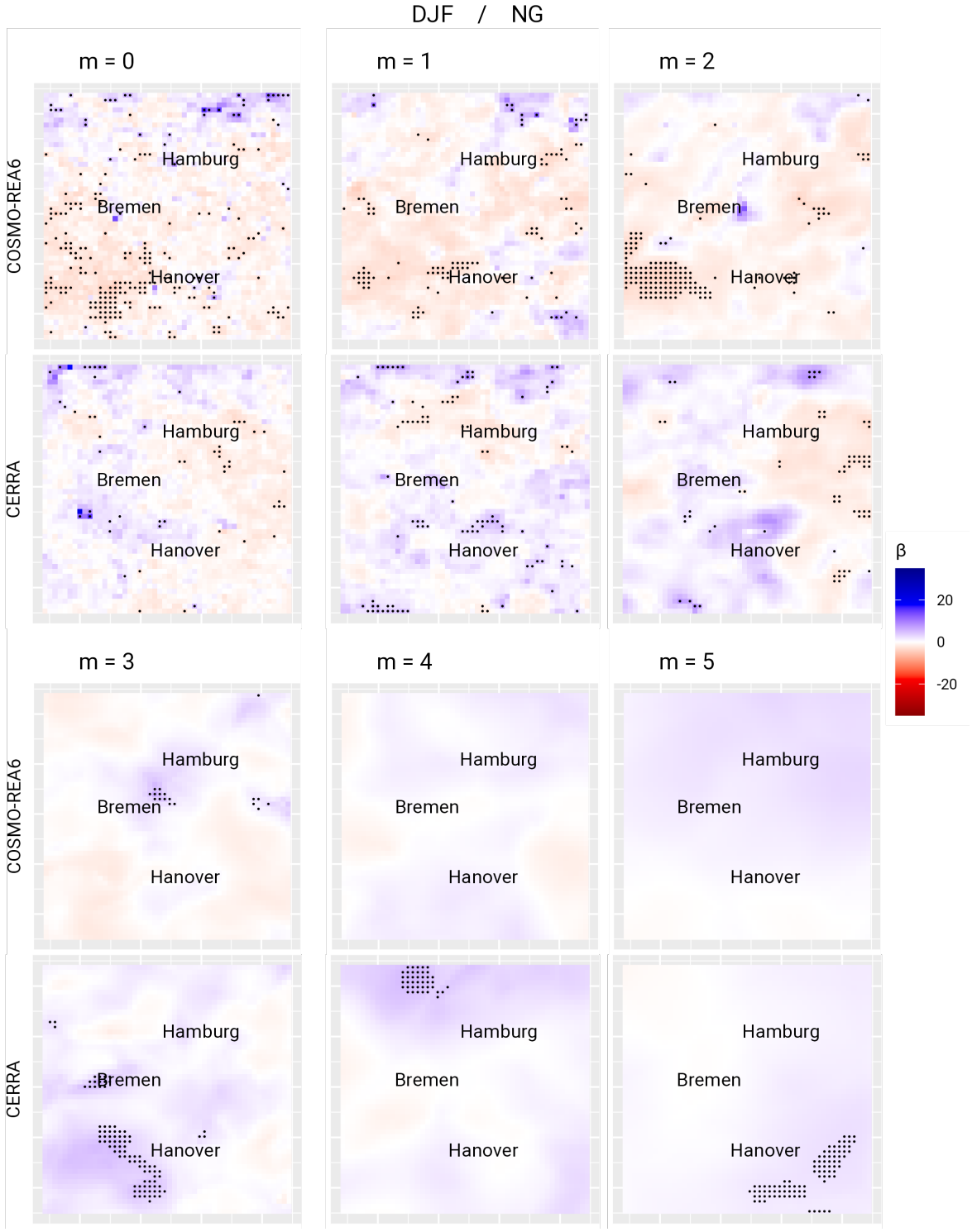


Figure B.5.: Same as Fig. 3.20 but for DJF months in NG.

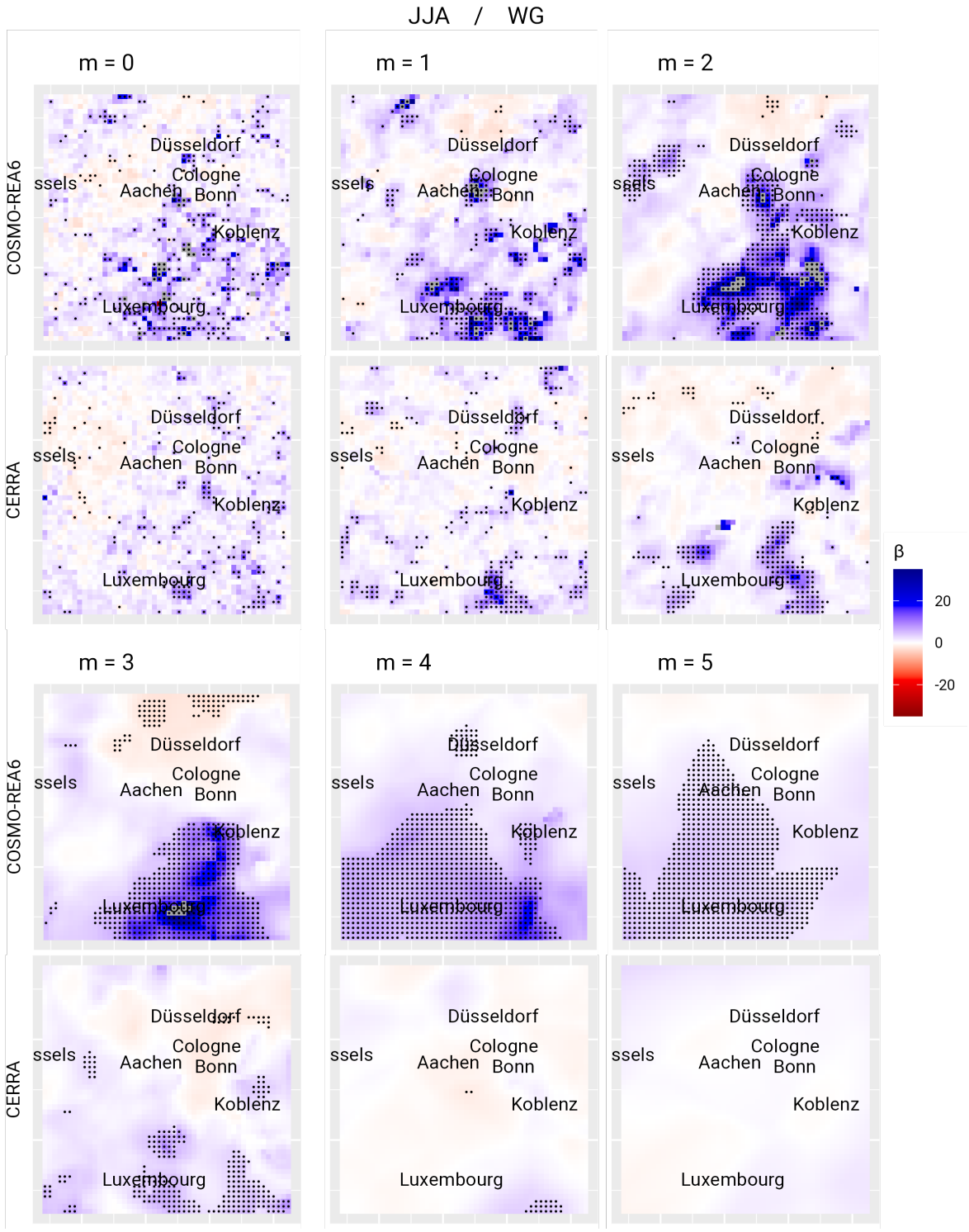


Figure B.6.: Same as Fig. 3.20 but for JJA months in WG.

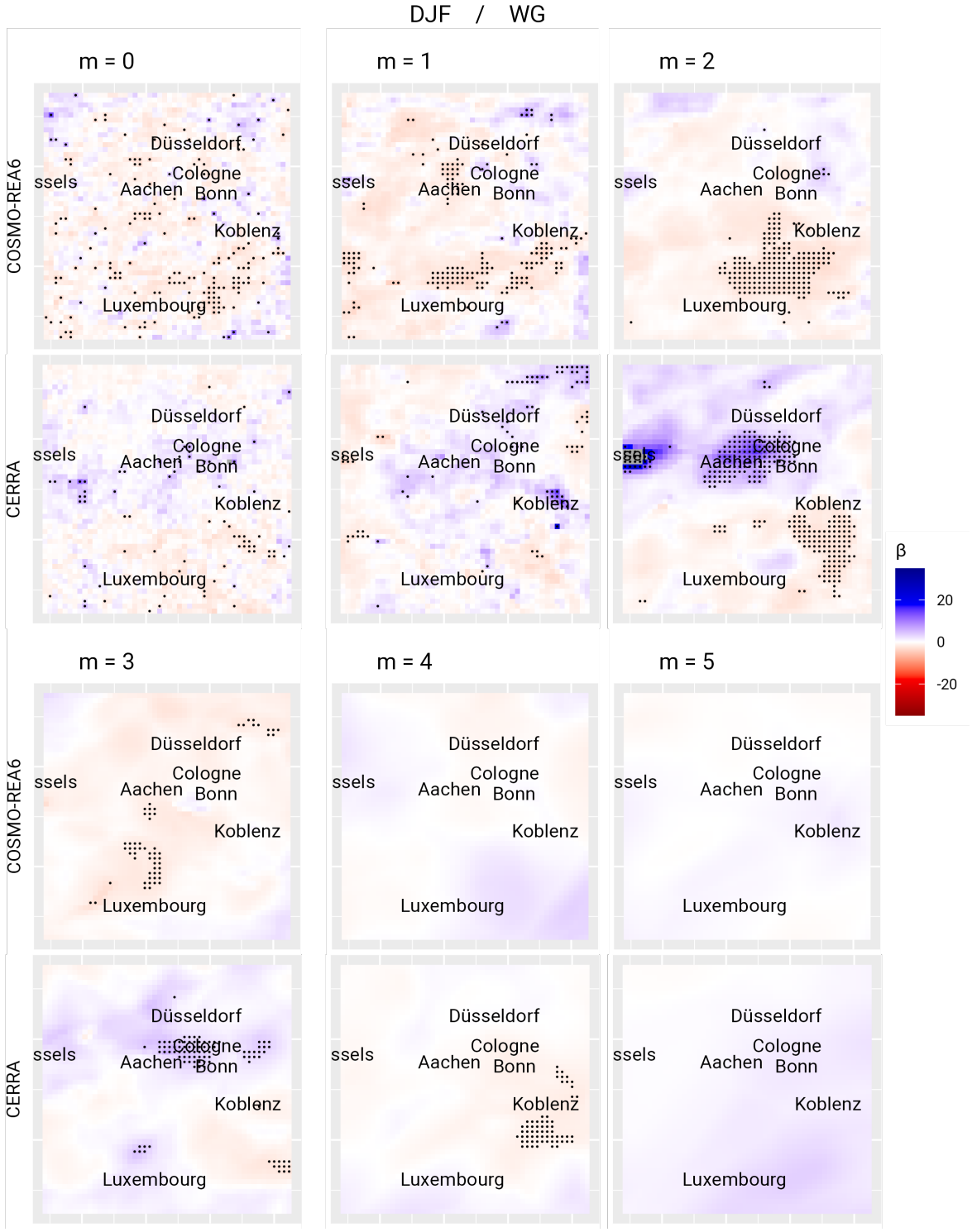


Figure B.7.: Same as Fig. 3.20 but for DJF months in WG.

B.3. Changes in Hourly Precipitation

We here provide an investigation of changes in timing and intensity of precipitation totals from COSMO-REA6 and CERRA. Methods and Figures are analogue to section 3.4.

B.3.1. Changes in Intensity

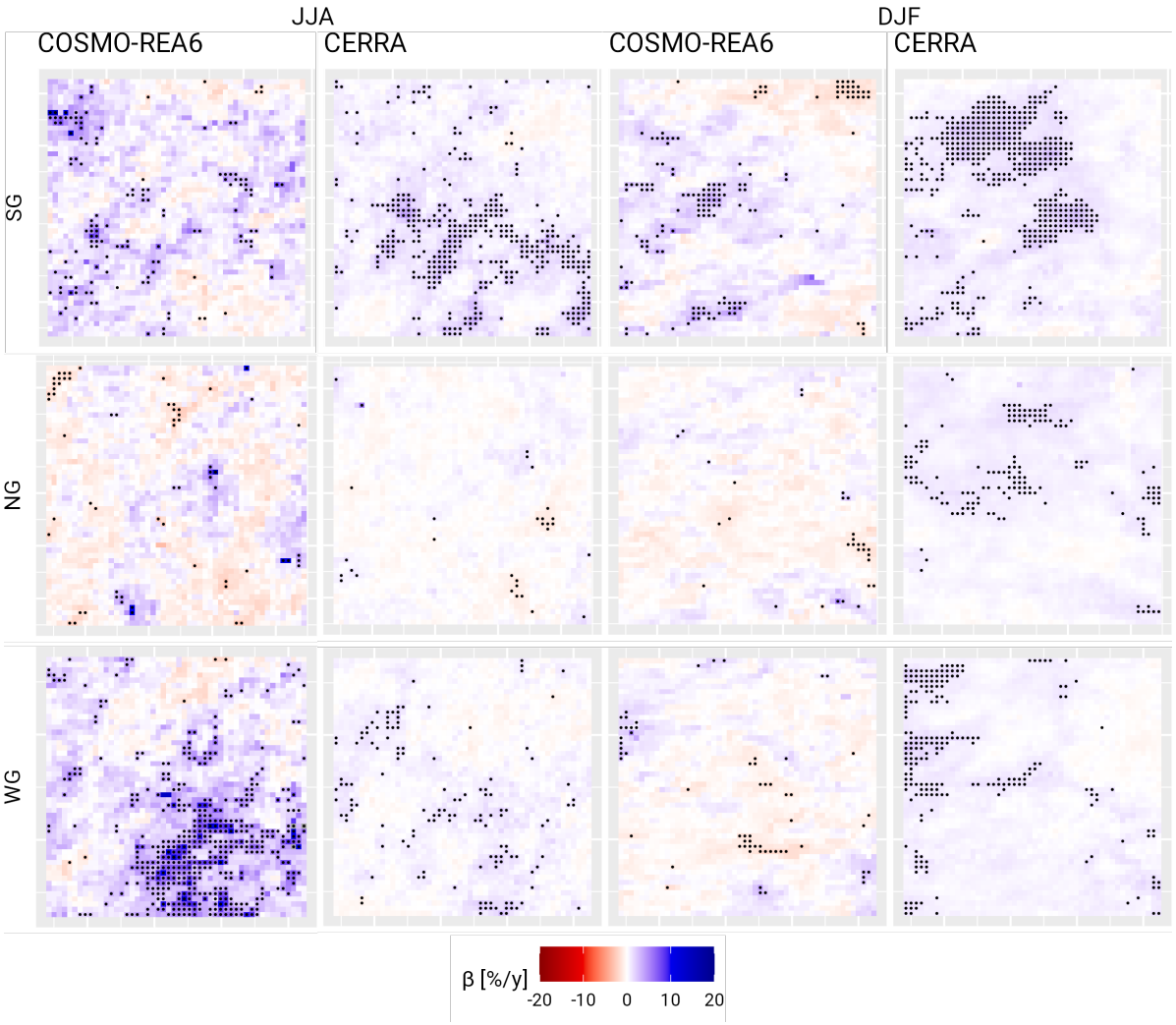


Figure B.8.: Same as Fig. **??**, but for hourly precipitation data from COSMO-REA6 (left) and CERRA (right). The results are shown for the months JJA (left) and DJF (right). For SG, NG and WG (from top to bottom).

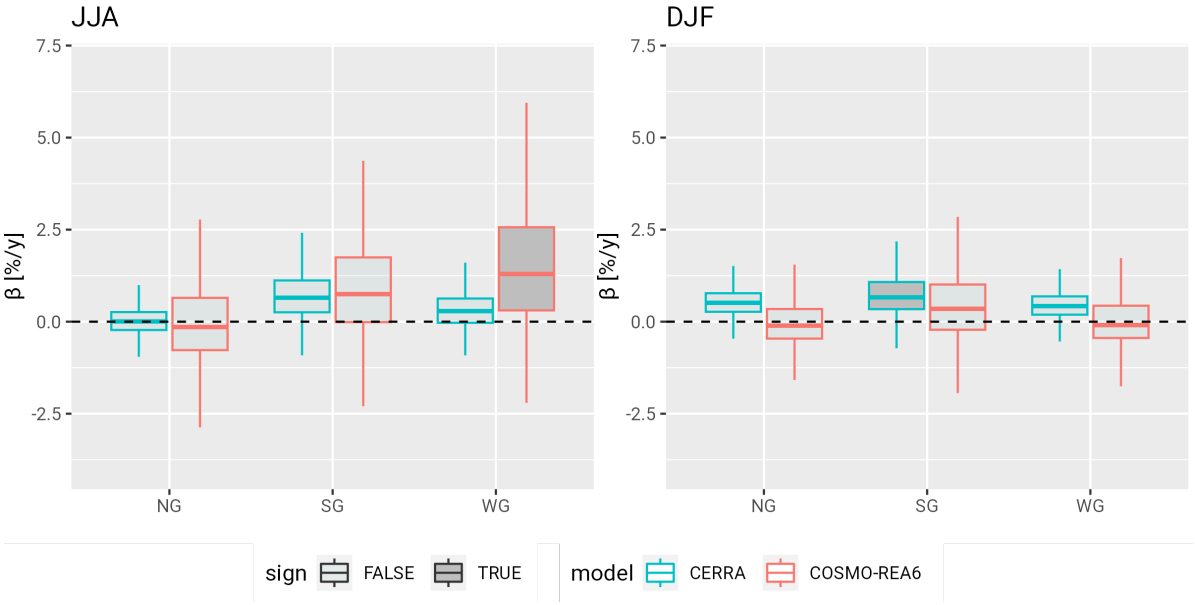


Figure B.9.: Same as Fig. 3.22 but for hourly precipitation data from COSMO-REA6 (red) and CERRA (blue). The results are shown for the months JJA (left) and DJF (right).

Table B.1.: Same as 3.2 but for hourly precipitation data from COSMO-REA6 and CERRA.

	JJA			DJF		
	SG	NG	WG	SG	NG	WG
COSMO-REA6	0.99	0.03	1.76*	0.42	-0.04	0.02
CERRA	0.73	0.03	0.033	0.73*	0.51	0.45

B.3.2. Changes in Timing

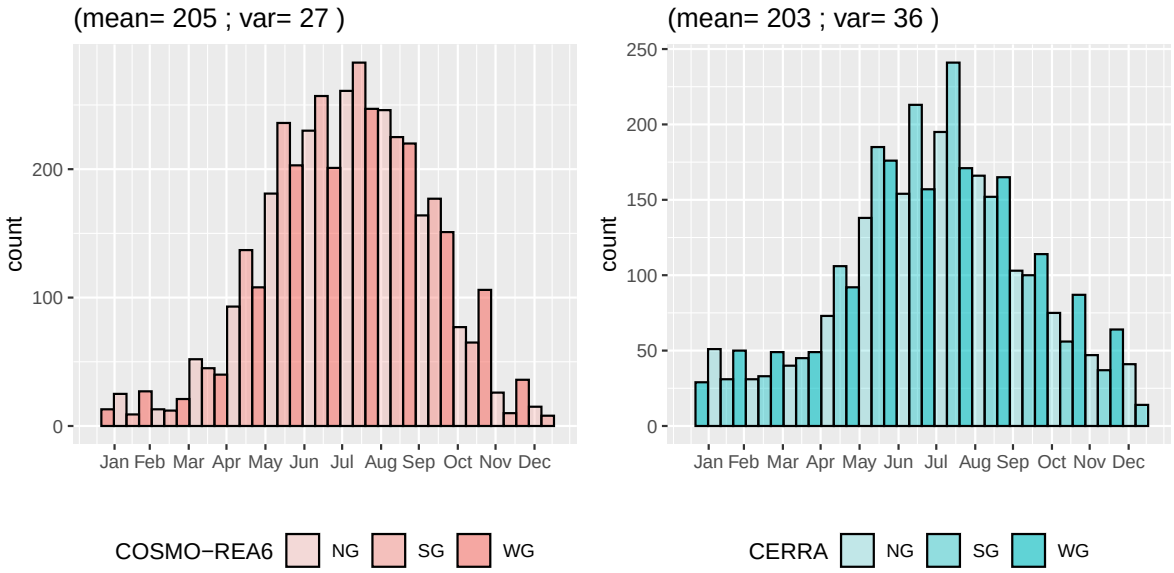


Figure B.10.: Same as Fig. 3.25, but for hourly precipitation data from COSMO-REA6 (left) and CERRA (right). Shown are results from the time period 1995-2018.

Table B.2.: Same as 3.3 but for hourly precipitation data from COSMO-REA6 and CERRA.

	mean			var		
	SG	NG	WG	SG	NG	WG
COSMO-REA6	-0.16	-0.30	-0.75	-0.25	-0.07	-0.43*
CERRA	-0.08	-0.23	0.33	-0.22	0.13	0.30

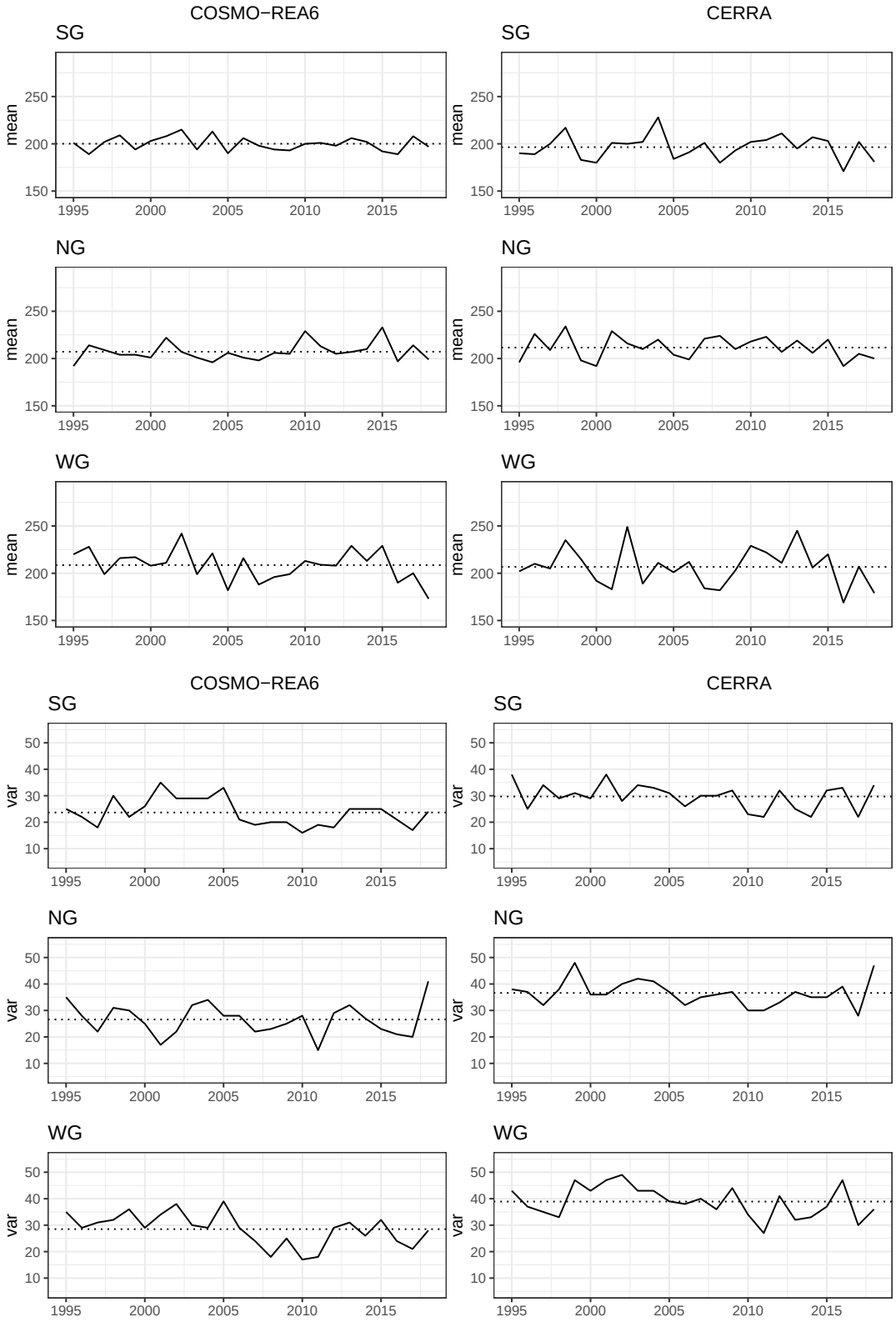


Figure B.11.: Same as Figs. 3.26 and 3.27, but for hourly precipitation data from COSMO-REA6 (left) and CERRA (right). Shown are results from the time period 1995-2018.

Bibliography

Abdi, Hervé and Lynne J Williams (2010). "Principal component analysis." In: *Wiley interdisciplinary reviews: computational statistics* 2.4. Publisher: Wiley Online Library, pp. 433–459.

Ahmed, Fiaz and Courtney Schumacher (Dec. 2015). "Convective and stratiform components of the precipitation-moisture relationship." en. In: *Geophysical Research Letters* 42.23. DOI: 10.1002/2015GL066957. (Visited on 11/20/2023).

Barnston, Anthony G. and Robert E. Livezey (1987). "Classification, Seasonality and Persistence of Low-Frequency Atmospheric Circulation Patterns." In: *Monthly Weather Review* 115. DOI: 10.1175/1520-0493(1987)11.

Benestad, Rasmus E. et al. (May 2022). "Global hydro-climatological indicators and changes in the global hydrological cycle and rainfall patterns." en. In: *PLOS Climate* 1.5. Ed. by Xander Wang. DOI: 10.1371/journal.pclm.0000029.

Benjamini, Yoav and Ruth Heller (Dec. 2007). "False Discovery Rates for Spatial Signals." en. In: *Journal of the American Statistical Association* 102.480, pp. 1272–1281. DOI: 10.1198/016214507000000941.

Benjamini, Yoav and Yosef Hochberg (Jan. 1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." en. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 57.1, pp. 289–300. DOI: 10.1111/j.2517-6161.1995.tb02031.x.

Beobide-Arsuaga, Goratz et al. (Jan. 2023). "Spring Regional Sea Surface Temperatures as a Precursor of European Summer Heatwaves." en. In: *Geophysical Research Letters* 50.2, e2022GL100727. DOI: 10.1029/2022GL100727.

Billingsley, Patrick (1968). *Convergence of Probability Measures*. Wiley & Sons. ISBN: 978-471-07242-7.

Björnsson, H and SA Venegas (1997). "A manual for EOF and SVD analyses of climatic data." In: *CCGCR Report* 97.1, pp. 112–134.

Bollmeyer, C. et al. (Jan. 2015). "Towards a high-resolution regional reanalysis for the European CORDEX domain." en. In: *Quarterly Journal of the Royal Meteorological Society* 141.686, pp. 1–15. DOI: 10.1002/qj.2486.

Bott, Andreas (2016). *Synoptische Meteorologie: Methoden der Wetteranalyse und-prognose*. Springer-Verlag. ISBN: 978-3-662-48194-3.

Brune, Sebastian, Sebastian Buschow, and Petra Friederichs (Apr. 2020). "Observations and high-resolution simulations of convective precipitation organization over the tropical Atlantic." en. In: *Quarterly Journal of the Royal Meteorological Society* 146.729, pp. 1545–1563. DOI: 10.1002/qj.3751.

– (Apr. 2021). "The local wavelet-based organization index – Quantification, localization and classification of convective organization from radar and satellite data." en. In: *Quarterly Journal of the Royal Meteorological Society* 147.736, pp. 1853–1872. DOI: 10.1002/qj.3998.

Brune, Sebastian, Florian Kapp, and Petra Friederichs (Oct. 2018). "A wavelet-based analysis of convective organization in ICON large-eddy simulations." en. In: *Quarterly Journal of the Royal Meteorological Society* 144.717, pp. 2812–2829. DOI: 10.1002/qj.3409.

Brönnimann, Stefan et al. (July 2018). "Changing seasonality of moderate and extreme precipitation events in the Alps." en. In: *Natural Hazards and Earth System Sciences* 18.7, pp. 2047–2056. DOI: 10.5194/nhess-18-2047-2018.

Buschow, Sebastian (2023a). *dualtrees: Decimated and Undecimated 2D Complex Dual-Tree Wavelet Transform*. URL: https://CRAN.R-project.org/package=dualtrees (visited on 12/20/2023).

– (Nov. 2023b). "Tropical convection in ERA5 has partly shifted from parameterized to resolved." en. In: *Quarterly Journal of the Royal Meteorological Society*, qj.4604. DOI: 10.1002/qj.4604.

Buschow, Sebastian and Petra Friederichs (Mar. 2020). "Using wavelets to verify the scale structure of precipitation forecasts." en. In: *Advances in Statistical Climatology, Meteorology and Oceanography* 6.1, pp. 13–30. DOI: 10.5194/ascmo-6-13-2020.

Buschow, Sebastian, Jakiw Pidstrigach, and Petra Friederichs (Aug. 2019). "Assessment of wavelet-based spatial verification by means of a stochastic precipitation model (wv_verif v0.1.0)." en. In: *Geoscientific Model Development* 12.8, pp. 3401–3418. DOI: 10.5194/gmd-12-3401-2019.

Caldas-Alvarez, Alberto et al. (Nov. 2022). "Meteorological, impact and climate perspectives of the 29 June 2017 heavy precipitation event in the Berlin metropolitan area." en. In: *Natural Hazards and Earth System Sciences* 22.11, pp. 3701–3724. DOI: 10.5194/nhess-22-3701-2022.

Chernokulsky, Alexander et al. (Mar. 2019). "Observed changes in convective and stratiform precipitation in Northern Eurasia over the last five decades." en. In: *Environmental Research Letters* 14.4. DOI: 10.1088/1748-9326/aafb82.

Coles, Stuart (2001). *An introduction to statistical modeling of extreme values*. en. Springer series in statistics. London ; New York: Springer. ISBN: 978-1-85233-459-8.

Cook, Benjamin I., Justin S. Mankin, and Kevin J. Anchukaitis (June 2018). "Climate Change and Drought: From Past to Future." en. In: *Current Climate Change Reports* 4.2, pp. 164–179. DOI: 10.1007/s40641-018-0093-2.

Cooley, D and E Thibaud (Sept. 2019). "Decompositions of dependence for high-dimensional extremes." en. In: *Biometrika* 106.3, pp. 587–604. DOI: 10.1093/biomet/asz028.

Dai, Aiguo, Tianbao Zhao, and Jiao Chen (Sept. 2018). "Climate Change and Drought: a Precipitation and Evaporation Perspective." en. In: *Current Climate Change Reports* 4.3, pp. 301–312. DOI: 10.1007/s40641-018-0101-6.

Daubechies, Ingrid (Oct. 1988). "Orthonormal bases of compactly supported wavelets." en. In: *Communications on Pure and Applied Mathematics* 41.7, pp. 909–996. DOI: 10.1002/cpa.3160410705.

– (1992). *Ten lectures on wavelets*. en. CBMS-NSF regional conference series in applied mathematics 61. Philadelphia, Pa: Society for Industrial and Applied Mathematics. ISBN: 978-0-89871-274-2.

Dhakal, Nirajan et al. (June 2015). "Nonstationarity in seasonality of extreme precipitation: A nonparametric circular statistical approach and its application." en. In: *Water Resources Research* 51.6, pp. 4499–4515. DOI: 10.1002/2014WR016399.

DKKV (2022). *Die Flutkatastrophe im Juli 2021 in Deutschland. Ein Jahr danach: Aufarbeitung und erste Lehren für die Zukunft*. Nr. 62. Bonn: DKKV-Schriftenreihe. ISBN: 978-3-933181-72-5.

Eckley, Idris A., Guy P. Nason, and Robert L. Treloar (June 2010). "Locally stationary wavelet fields with application to the modelling and analysis of image texture: Modelling and Analy-

sis of Image Texture." en. In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, no–no. DOI: 10.1111/j.1467-9876.2009.00721.x.

Fauer, Felix S. et al. (Dec. 2021). "Flexible and consistent quantile estimation for intensity–duration–frequency curves." en. In: *Hydrology and Earth System Sciences* 25.12, pp. 6479–6494. DOI: 10.5194/hess-25-6479-2021.

Field, Christopher B. et al., eds. (May 2012). *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change*. en. 1st ed. Cambridge University Press. URL: https://www.cambridge.org/core/product/identifier/9781139177245/type/book (visited on 11/14/2023).

Fischer, E. M. and R. Knutti (Nov. 2016). "Observed heavy precipitation increase confirms theory and early models." en. In: *Nature Climate Change* 6.11, pp. 986–991. DOI: 10.1038/nclimate3110.

Fourier, Jean Baptiste Joseph (1888). *Théorie analytique de la chaleur*. Vol. 1. Gauthier-Villars. ISBN: 978-2-01-302758-8.

Fowler, Hayley J. et al. (Jan. 2021). "Anthropogenic intensification of short-duration rainfall extremes." en. In: *Nature Reviews Earth & Environment* 2.2, pp. 107–122. DOI: 10.1038/s43017-020-00128-6.

Grieger, Jens et al. (2024). "Dauerniederschläge und Weihnachtshochwasser im Winter 2023/24." In: DOI: 10.17169/refubium-43523.

Griggs, David J. and Maria Noguer (Aug. 2002). "Climate change 2001: The scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change." en. In: *Weather* 57.8, pp. 267–269. DOI: 10.1256/004316502320517344.

Guerreiro, Selma B. et al. (Sept. 2018). "Detection of continental-scale intensification of hourly rainfall extremes." en. In: *Nature Climate Change* 8.9, pp. 803–807. DOI: 10.1038/s41558-018-0245-3.

Haar, Alfred (1909). "Zur Theorie der orthogonalen Funktionensysteme: Erste Mitteilung." de. In: *Mathematische Annalen* 69.3, pp. 331–371. DOI: 10.1007/BF01456326.

Han, Xuqing et al. (Mar. 2016). "The roles of convective and stratiform precipitation in the observed precipitation trends in Northwest China during 1961–2000." en. In: *Atmospheric Research* 169, pp. 139–146. DOI: 10.1016/j.atmosres.2015.10.001.

Hegerl, Gabriele C et al. (2010). "Good practice guidance paper on detection and attribution related to anthropogenic climate change." en. In: *Intergovernmental Panel on Climate Change (IPCC) Working Group*. DOI: 10.1029/2004GL020876.

Helama, Samuli et al. (June 2018). "Reconstruction of precipitation variability in Estonia since the eighteenth century, inferred from oak and spruce tree rings." en. In: *Climate Dynamics* 50.11-12, pp. 4083–4101. DOI: 10.1007/s00382-017-3862-z.

Higham, N. J. (July 2002). "Computing the nearest correlation matrix–a problem from finance." en. In: *IMA Journal of Numerical Analysis* 22.3, pp. 329–343. DOI: 10.1093/imanum/22.3.329.

Hurrell, James W. et al. (2003). "An overview of the North Atlantic Oscillation." en. In: *Geophysical Monograph Series*. Ed. by James W. Hurrell et al. Vol. 134. Washington, D. C.: American Geophysical Union, pp. 1–35. ISBN: 978-0-87590-994-3.

Jiang, Yujing, Daniel Cooley, and Michael F Wehner (2020). "Principal Component Analysis for Extremes and Application to U.S. Precipitation." en. In: *Jurnal of Climate* 33, p. 11. DOI: 10.1175/JCLI-D-19-0413.1.

Kapp, Florian et al. (Aug. 2018). "Spatial verification of high-resolution ensemble precipitation forecasts using local wavelet spectra." en. In: *Climatic Change* 176.90. arXiv: 1808.07667. DOI: 10.1007/s10584-023-03502-7.

Kendall, Maurice George (1948). *Rank correlation methods*. Oxford University Pres. ISBN: 978-0-19-520837-5.

Keyantash, John and John A. Dracup (2002). "The quantification of drought: An evaluation of drought indices." In: *American Meteorological Society*. DOI: 10.1175/1520-0477-83.8.1167.

Kingsbury, Nick (1998). "The dual-tree complex wavelet transform: A new efficient tool for image restoration and enhancement." en. In: *9th European Signal Processing Conference*. ISSN: 978-960-7620-06-4.

Knutti, Reto and Jan Sedláček (Apr. 2013). "Robustness and uncertainties in the new CMIP5 climate model projections." en. In: *Nature Climate Change* 3.4, pp. 369–373. DOI: 10.1038/nclimate1716.

Kreienkamp, Frank et al. (2021). *Rapid attribution of heavy rainfall events leading to the severe flooding in Western Europe during July 2021*. URL: https://www.worldweatherattribution.org (visited on 05/01/2024).

Landgraf, Andrew J. and Yoonkyung Lee (Nov. 2020). "Dimensionality reduction for binary data through the projection of natural parameters." en. In: *Journal of Multivariate Analysis* 180, p. 104668. DOI: 10.1016/j.jmva.2020.104668.

Larsson, Martin and Sidney I. Resnick (June 2012). "Extremal dependence measure and extremogram: the regularly varying case." en. In: *Extremes* 15.2, pp. 231–256. DOI: 10.1007/s10687-011-0135-9.

Lavaysse, Christophe et al. (Jan. 2018). "Towards a monitoring system of temperature extremes in Europe." en. In: *Natural Hazards and Earth System Sciences* 18.1, pp. 91–104. DOI: 10.5194/nhess-18-91-2018.

Lenderink, G. et al. (Sept. 2011). "Scaling and trends of hourly precipitation extremes in two different climate zones – Hong Kong and the Netherlands." en. In: *Hydrology and Earth System Sciences* 15.9, pp. 3033–3041. DOI: 10.5194/hess-15-3033-2011.

Lenderink, Geert and Erik Van Meijgaard (Apr. 2010). "Linking increases in hourly precipitation extremes to atmospheric temperature and moisture changes." en. In: *Environmental Research Letters* 5.2, p. 025208. DOI: 10.1088/1748-9326/5/2/025208.

Lengfeld, Katharina et al. (Dec. 2019). "Characteristic spatial extent of hourly and daily precipitation events in Germany derived from 16 years of radar data." en. In: *Meteorologische Zeitschrift* 28.5, pp. 363–378. DOI: 10.1127/metz/2019/0964.

Lengfeld, Katharina et al. (Mar. 2023). "Brief communication: On the extremeness of the July 2021 precipitation event in western Germany." en. In: *Natural Hazards and Earth System Sciences* 23.3, pp. 1227–1232. DOI: 10.5194/nhess-23-1227-2023.

Lorenz, Edward N. (1956). *Empirical orthogonal functions and statistical weather prediction*. 1. Massachusetts Institute of Technology, Department of Meteorology Cambridge.

Ludwig, Patrick et al. (Apr. 2023). "A multi-disciplinary analysis of the exceptional flood event of July 2021 in central Europe – Part 2: Historical context and relation to climate change." en. In: *Natural Hazards and Earth System Sciences* 23.4, pp. 1287–1311. DOI: 10.5194/nhess-23-1287-2023.

Madsen, H. et al. (Nov. 2014). "Review of trend analysis and climate change projections of extreme precipitation and floods in Europe." en. In: *Journal of Hydrology* 519, pp. 3634–3650. DOI: 10.1016/j.jhydrol.2014.11.003.

Mallat, Stephane (1998). *a Wavelet tour of signal processing*. 2nd ed.
Mallat, Stephane G. (Sept. 1989). “Multiresolution Approximations and Wavelet Orthonormal Bases of L 2 (R).” en. In: *Transactions of the American Mathematical Society* 315.1, p. 69. DOI: 10.2307/2001373.
Mann, Henry B. (July 1945). “Nonparametric Tests Against Trend.” en. In: *Econometrica* 13.3, p. 245. DOI: 10.2307/1907187.
Masson-Delmotte, V. et al. (2021). “IPCC, 2021.” en. In: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* Cambridge University Press.2. DOI: doi:10.1017/9781009157896..
McKee, Thomas B, Nolan J Doesken, and John Kleist (1993). “The relationship of drought frequency and duration to time scales.” en. In: *Proceedings of the 8th Conference on Applied Climatology* 17.22.
Meurer, Pascal (2023). “Attribution von Hitzewellen und Dürren in Europa.” Bachelorarbeit Meteorologie. Bonn: Rheinische Friedrich-Wilhelms-Universität Bonn.
Meyer, Yves (1997). *Ondelettes et opérateurs Volume 1: Ondelettes*. fre. OCLC: 1038041932. Paris, France: Editions Hermann. ISBN: 978-2-7056-6125-0.
Mohr, Susanna et al. (Feb. 2023). “A multi-disciplinary analysis of the exceptional flood event of July 2021 in central Europe – Part 1: Event description and analysis.” en. In: *Natural Hazards and Earth System Sciences* 23.2, pp. 525–551. DOI: 10.5194/nhess-23-525-2023.
Nason, Guy P., R. Von Sachs, and Gerald Kroisandt (July 2000). “Wavelet Processes and Adaptive Estimation of the Evolutionary Wavelet Spectrum.” en. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 62.2, pp. 271–292. DOI: 10.1111/1467-9868.00231.
Nelson, J. D. B. et al. (Nov. 2018). “The locally stationary dual-tree complex wavelet model.” en. In: *Statistics and Computing* 28.6, pp. 1139–1154. DOI: 10.1007/s11222-017-9784-0.
Panthou, Gérémy et al. (Oct. 2014). “Relationship between Surface Temperature and Extreme Rainfalls: A Multi-Time-Scale and Event-Based Analysis*.” en. In: *Journal of Hydrometeorology* 15.5, pp. 1999–2011. DOI: 10.1175/JHM-D-14-0020.1.
Park, In-Hong and Seung-Ki Min (Dec. 2017). “Role of Convective Precipitation in the Relationship between Subdaily Extreme Precipitation and Temperature.” en. In: *Journal of Climate* 30.23, pp. 9527–9537. DOI: 10.1175/JCLI-D-17-0075.1.
Pearson, Karl (Nov. 1901). “LIII. *On lines and planes of closest fit to systems of points in space*.” en. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2.11, pp. 559–572. DOI: 10.1080/14786440109462720.
Perkins, S. E. and L. V. Alexander (July 2013). “On the Measurement of Heat Waves.” en. In: *Journal of Climate* 26.13, pp. 4500–4517. DOI: 10.1175/JCLI-D-12-00383.1.
Peter, Madlen, Henning W Rust, and Uwe Ulbrich (2024). “Interannual variations in the seasonal cyle of extreme precipitation in Germany and the response to climate change.” en. In: DOI: 10.5194/nhess-24-1261-2024.
Resnick, Sidney (2002). “The extremal dependence measure and asymptotic independence.” In: *Taylor \& Francis*. DOI: 10.1081/STM-120034129.
Resnick, Sidney I (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media. ISBN: 987-1-4419-2024-9.

Ridal, Martin et al. (2024). "CERRA, the Copernicus European Regional Reanalysis system." en. In: *Quarterly Journal of the Royal Meteorological Society*. DOI: 10.1002/qj.4764.

Ritzhaupt, N. and D. Maraun (Jan. 2023). "Consistency of Seasonal Mean and Extreme Precipitation Projections Over Europe Across a Range of Climate Model Ensembles." en. In: *Journal of Geophysical Research: Atmospheres* 128.1, e2022JD037845. DOI: 10.1029/2022JD037845.

Ritzhaupt, Nicole and Douglas Maraun (June 2024). "State-of-the-art climate models reduce dominant dynamical uncertainty in projections of extreme precipitation." en. In: *Environmental Research: Climate* 3.2, p. 021001. DOI: 10.1088/2752-5295/ad2eb2.

Robinson, Peter J (2001). "On the definition of a heat wave." In: *Journal of Applied Meteorology and Climatology* 40.4. Publisher: American Meteorological Society, pp. 762–775. DOI: 10.1175/1520-0450.

Rohrbeck, Christian and Daniel Cooley (2021). "Simulating flood event sets using extremal principal components." In: *The Annals of Applied Statistics*. Publisher: arXiv Version Number: 2. DOI: 10.1214/22-AOAS1672.

Rousi, Efi et al. (July 2022). "Accelerated western European heatwave trends linked to more-persistent double jets over Eurasia." en. In: *Nature Communications* 13.1, p. 3851. DOI: 10.1038/s41467-022-31432-y.

Russo, Simone, Jana Sillmann, and Erich M Fischer (2015). "Top ten European heatwaves since 1950 and their occurrence in the coming decades." In: *Environmental Research Letters* 10.12. Publisher: IOP Publishing, p. 124003. DOI: 10.1088/1748-9326/10/12/124003.

Schimanke, S. et al. (2021). *CERRA sub-daily regional reanalysis data for Europe on model levels from 1984 to present*. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). DOI: 10.24381/cds.7c27fd20. (Visited on 03/14/2023).

Schlather, Martin (Apr. 2024). *An algebraic generalization of the entropy and its application to statistics*. en. arXiv:2404.05854 [math]. URL: http://arxiv.org/abs/2404.05854 (visited on 06/26/2024).

Schlather, Martin and Felix Reinbott (Dec. 2021). *A semi-group approach to Principal Component Analysis*. en. arXiv:2112.04026 [math, stat]. URL: http://arxiv.org/abs/2112.04026 (visited on 06/26/2024).

Schulzweida, Uwe (2023). *CDO User Guide*. URL: https://doi.org/10.5281/zenodo.10020800.

Selesnick, Ivan W. and Ke Yong Li (Nov. 2003). "Video denoising using 2D and 3D dual-tree complex wavelet transforms." en. In: ed. by Michael A. Unser, Akram Aldroubi, and Andrew F. Laine. San Diego, California, USA, p. 607. DOI: 10.1117/12.504896.

Selesnick, I.W. (May 2002). "The design of approximate Hilbert transform pairs of wavelet bases." en. In: *IEEE Transactions on Signal Processing* 50.5, pp. 1144–1152. DOI: 10.1109/78.995070.

Sienz, F., O. Bothe, and K. Fraedrich (July 2012). "Monitoring and quantifying future climate projections of dryness and wetness extremes: SPI bias." en. In: *Hydrology and Earth System Sciences* 16.7, pp. 2143–2157. DOI: 10.5194/hess-16-2143-2012.

Slater, Louise J et al. (2021). "Nonstationary weather and water extremes: a review of methods for their detection, attribution, and management." In: *Hydrology and Earth System Sciences* 25.7. Publisher: Copernicus GmbH, pp. 3897–3935. DOI: 10.5194/hess-25-3897-2021.

Szemkus, Svenja (2016). "Regionale Klimatologie: Niederschlagsföhneffekt Bonn." Bachelorthesis. Bonn: University Bonn.

– (2023). *ExtrPatt: Spatial Dependencies and Indices for Extremes*. URL: https://CRAN.R-project.org/package=ExtrPatt (visited on 12/20/2023).
Szemkus, Svenja and Petra Friederichs (Jan. 2024). "Spatial patterns and indices for heat waves and droughts over Europe using a decomposition of extremal dependency." en. In: *Advances in Statistical Climatology, Meteorology and Oceanography* 10.1, pp. 29–49. DOI: 10.5194/ascmo-10-29-2024.
Tradowsky, Jordis S. et al. (July 2023). "Attribution of the heavy rainfall events leading to severe flooding in Western Europe during July 2021." en. In: *Climatic Change* 176.7, p. 90. DOI: 10.1007/s10584-023-03502-7.
Trenberth, Kevin E. and David P. Stepaniak (Apr. 2001). "Indices of El Niño Evolution." en. In: *Journal of Climate* 14.8, pp. 1697–1701. DOI: 10.1175/1520-0442.
Trenberth, Kevin E. et al. (Sept. 2003). "The Changing Character of Precipitation." en. In: *Bulletin of the American Meteorological Society* 84.9, pp. 1205–1218. DOI: 10.1175/BAMS-84-9-1205.
Vautard, R. et al. (Dec. 2015). "Extreme Fall 2014 Precipitation in the Cévennes Mountains." en. In: *Bulletin of the American Meteorological Society* 96.12, S56–S60. DOI: 10.1175/BAMS-D-15-00088.1.
Verrelle, A. et al. (2022). *CERRA-Land sub-daily regional reanalysis data for Europe from 1984 to present*. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). DOI: 10.24381/cds.a7f3cd0b.
Voit, Paul and Maik Heistermann (Aug. 2022). "A new index to quantify the extremeness of precipitation across scales." en. In: *Natural Hazards and Earth System Sciences* 22.8, pp. 2791–2805. DOI: 10.5194/nhess-22-2791-2022.
Wahl, Sabrina et al. (Oct. 2017). "A novel convective-scale regional reanalysis COSMO-REA2: Improving the representation of precipitation." en. In: *Meteorologische Zeitschrift* 26.4, pp. 345–361. DOI: 10.1127/metz/2017/0824.
Wang, Yadong et al. (Jan. 2021). "Separation of convective and stratiform precipitation using polarimetric radar data with a support vector machine method." en. In: *Atmospheric Measurement Techniques* 14.1, pp. 185–197. DOI: 10.5194/amt-14-185-2021.
Weniger, Michael, Florian Kapp, and Petra Friederichs (Jan. 2017). "Spatial verification using wavelet transforms: a review." en. In: *Quarterly Journal of the Royal Meteorological Society* 143.702, pp. 120–136. DOI: 10.1002/qj.2881.
Werner, Mike (2014). "Eifelföhn: Leeeffekte im Einflussbereich der Eifel." Bachelorthesis. Bonn: University Bonn.
Westra, Seth, Lisa V. Alexander, and Francis W. Zwiers (June 2013). "Global Increasing Trends in Annual Maximum Daily Precipitation." en. In: *Journal of Climate* 26.11, pp. 3904–3918. DOI: 10.1175/JCLI-D-12-00502.1.
Wilks, D. S. (Sept. 2006a). "On "Field Significance" and the False Discovery Rate." en. In: *Journal of Applied Meteorology and Climatology* 45.9, pp. 1181–1189. DOI: 10.1175/JAM2404.1.
Wilks, Daniel S. (2006b). *Statistical methods in the atmospheric sciences*. en. 2nd ed. International geophysics series v. 91. Amsterdam ; Boston: Academic Press. ISBN: 978-0-12-751966-1.
Winterrath, Tanja, Wolfgang Rosenow, and Elmar Weigl (2012). *Winterrath, Tanja, Wolfgang Rosenow, and Elmar Weigl. "On the DWD quantitative precipitation analysis and nowcasting system for real-time application in German flood risk management.*

Xoplaki, Elena et al. (Aug. 2023). *Compound events in Germany in 2018: drivers and case studies*. en. DOI: 10.5194/egusphere-2023-1460.

Ye, Hengchun et al. (Sept. 2017). "More frequent showers and thunderstorm days under a warming climate: evidence observed over Northern Eurasia from 1966 to 2000." en. In: *Climate Dynamics* 49.5-6, pp. 1933–1944. DOI: 10.1007/s00382-016-3412-0.

Zeder, Joel and Erich M. Fischer (Sept. 2020). "Observed extreme precipitation trends and scaling in Central Europe." en. In: *Weather and Climate Extremes* 29, p. 100266. DOI: 10.1016/j.wace.2020.100266.

Zolina, Olga et al. (Mar. 2008). "Seasonally dependent changes of precipitation extremes over Germany since 1950 from a very dense observational network." en. In: *Journal of Geophysical Research: Atmospheres* 113.D6, 2007JD008393. DOI: 10.1029/2007JD008393.

Zscheischler, Jakob and Sonia I. Seneviratne (June 2017). "Dependence of drivers affects risks associated with compound events." en. In: *Science Advances* 3.6, e1700263. DOI: 10.1126/sciadv.1700263.

Further Resources

Statistical analysis, data handling, and visualization:

- R Core Team (2021). *A Language and Environment for Statistical Computing*. Vienna, Austria. URL: https://www.R-project.org/

Text processing and citation:

- Corporation for Digital Scholarship (2024). *Zotero*. Accessed: 2021-07-01. URL: https://www.zotero.org
- Overleaf (2024). *Overleaf: Online LaTeX Editor*. URL: https://www.overleaf.com

Formulation and spelling correction:

- DeepL (2024). *DeepL Translator*. Cologne, Germany. URL: https://www.deepl.com
- Grammarly (2024). *Grammarly*. San Francisco, United States. URL: https://www.grammarly.com
- OpenAI (2024). *ChatGPT: OpenAI's GPT-4 Model*. URL: https://www.openai.com/chatgpt

List of Names and Abbreviations

GEV Generalized Extreme Value Distribution 7
PCA Principal Component Analysis 2
IPCC Intergovernmental Panel on Climate Change 1
SPI Standartized Precipitation Index 5
TPDM Tail Pairwise Dependence Matrix 2
NAO North Atlantic Oscillation 5
EDM Extremal Dependence Measure 8
EPI Extremal Pattern Index 3
TEPI Threshold-based Extremal Pattern Index 20
SVD Singular Value Decomposition 18
CC Clausius-Clapeyron 23
WT Wavelet Transform 26
DWT Discrete Wavelet Transform 27
FT Fourier Transform 26
RDWT redundant discrete wavelet transform 35
MRA Multiresolution Approximation 28
CWT Complex Wavelet Transform 31
DT-CWT Dualtree Complex Wavelet Transform 28
WG Western Germany 4
NG Northern Germany 4
SG Southern Germany 4
IDF intensity–duration–frequency 24
LWS local wavelet spectrum 36
CERRA Copernicus European Regional ReAnalyses 38
HErZ Hans-Ertel Centre for Weather Research 38
FDR False Detection Rate 49
DWD German Weather Service 59
JJA northern hemispheric Summer 12

Acknowledgements

Wen die Dankbarkeit geniert,
Der ist übel dran;
Denke, wer dich erst geführt,
Wer für dich getan!

Johann Wolfgang von Goethe
(1749 - 1832)

In diesem Sinne sei allen Menschen ganz herzlich gedankt, die mich beim Erstellen dieser Dissertation unterstützt haben.

Zunächst ist hier natürlich meine Betreuerin, Petra Friederichs, zu nennen. Dafür, dass ich nach meinem "Seitensprung" in die Softwareentwicklung wieder herzlich "aufgenommen" wurde und meine Dissertation zu diesem spannenden Thema anfertigen durfte. Ebenso möchte ich mich für die zahlreichen Gespräche, fruchtbaren Diskussionen und die hervorragende fachliche Beteuung bedanken.

Ausdrücklich danken, möchte ich ebenfalls allen, mit denen ich im Laufe meiner Dissertation zusammen arbeiten durfte. Meinen Kollegen, sowie auch ehemaligen Mitgliedern unseres Instituts. Euer Zuspruch und unsere Diskussionen haben wesentlich zum Gelingen dieser Arbeit beigetragen. Besonders möchte ich an dieser Stelle Sebastian Buschow für seine fachliche Unterstützung, insbesondere in Bezug auf Kapitel 3 meiner Dissertation danken.

Mein herzlicher Dank gilt auch den Mitgliedern meiner Promotionskommission. Neben meiner Betreuerin Petra Friederichs sind hier Andreas Hense, Andreas Bott und Thomas Heckelei zu nennen. Vielen Dank für die Bereitschaft.

Zu guter Letzt, möchte ich meiner Familie danken. Ohne die Unterstützung meiner Eltern, wäre dieser Weg nicht möglich gewesen.

BONNER METEOROLOGISCHE ABHANDLUNGEN

Herausgegeben vom Institut für Geowissenschaften der Universität Bonn, Abteilung Meteorologie, durch Prof. Dr. H. FLOHN (Hefte 1-25), Prof. Dr. M. HANTEL (Hefte 26-35), Prof. Dr. H.-D. SCHILLING (Hefte 36-39), Prof. Dr. H. KRAUS (Hefte 40-49), ab Heft 50 durch Prof. Dr. A. HENSE.

Heft 1-79: siehe `https://www.ifgeo.uni-bonn.de/de/abteilungen/meteorologie/bibliothek/bibliothek-meteorologie#BMA`

80-96: open access, verfügbar unter `https://bonndoc.ulb.uni-bonn.de/xmlui/handle/20.500.11811/1627`

Heft 80: ***Tanja Zerenner***: Atmospheric downscaling using multi-objective genetic programming, 2016, [erschienen] 2017, X, 191 S.

Heft 81: ***Sophie Stolzenberger***: On the probabilistic evaluation of decadal and paleoclimate model predictions, 2017, IV, 122 S.

Heft 82: ***Insa Thiele-Eich***: Flooding in Dhaka, Bangladesh, and the challenge of climate change, 2017, V, 158 S.

Heft 83: ***Liselotte Bach***: Towards a probabilistic regional reanalysis for Europe, 2017 [erschienen] 2018, VI, 114 S.

Heft 84: ***Yen-Sen Lu***: Propagation of land surface model uncertainties in terrestrial system states, 2017, [erschienen] 2018, X, 120 S.

Heft 85: ***Rüdiger Hewer***: Stochastic physical models for wind fields and precipitation extremes, 2018, 99 S.

Heft 86: ***Sebastian Knist***: Land-atmosphere interactions in multiscale regional climate change simulations over Europe, 2018, VIII, 147 S.

Heft 87: ***Jessica Keune***: Integrated terrestrial simulations at the continental scale: Impact of groundwater dynamics and human water use on groundwater-to-atmosphere feedbacks during the European heatwave in 2003, 2019, IX, 172 S.

Heft 88: ***Christoph Beekmans***: 3-D Cloud Morphology and Evolution Derived from Hemispheric Stereo Cameras, 2019, [erschienen] 2020, VIII, 118 S.

Heft 89: ***Nils Weitzel***: Climate field reconstructions from pollen and macrofossil syntheses using Bayesian hierarchical models, 2019, [erschienen] 2020, XII, 153 S.

Heft 90: ***Alexander Kelbch***: Investigations to quantify individual exposure to solar ultraviolet erythemal radiation including cloud meteorological impact, 2020, III, 107 S.

Heft 91: ***Mari L. Schmidt***: Improvement of hail detection and nowcasting by synergistic combination of information from polarimetric radar, model predictions, and in-situ observations, 2020, VI, 136 S.

Heft 92: ***Sebastian Brune***: Der Wavelet-basierte Organisationsindex als Maß der konvektiven Organisation über Deutschland und dem tropischen Atlantik, 2021, IV, 121 S.

Heft 93: ***Sebastian Buschow***: Spatial Verification with Wavelets, 2022, V, 195 S.

Heft 94: ***Michael Langguth***: Representation of deep convection at gray-zone resolutions - Implementing and testing the HYbrid MAss flux Convection Scheme (HYMACS) in the ICON model, 2022, VI, 173 S.

Heft 95: ***Timon Netzel***: Quantitative paleoclimate reconstructions in the European region based on multiple proxies, 2023, VI, 179 S.

Heft 96: ***Lukas Hubert Leufen***: Time Filter Assisted Deep Learning to Predict Air Pollution, 2023, IV, 156 S.

Heft 97: ***Maike Hacker***: Modelling Fog and Low Stratiform Clouds in the Namib Desert with COSMO-FOG, 2024, VIII, 176 S.

Heft 98: ***Svenja Szemkus***: Compact Description for High-Resolution Spatial Weather Extremes, 2025, VI, 117 S.

Heft 99: ***Lijie Zhang***: Modelling Secondary Circulation in Convective Boundary Layer Using Large Eddy Simulation, 2024, VI, 117 S.



Institut für Geowissenschaften
Abteilung Meteorologie
Mathematisch Naturwissenschaftliche Fakultät
Universität Bonn

institut für geowissenschaften
METEOROLOGIE