Eberhard Karls Universität Tübingen
Mathematisch-Naturwissenschaftliche Fakultät
Wilhelm-Schickard-Institut für Informatik

Master Thesis Bioinformatics

Spatio-Temporal Patterns of European Heat Waves and Their Influence on Vegetation

Julia Pascale Hellmig

13.06.2022

Reviewers

Bedartha Goswami
(Informatik)
Wilhelm-Schickard-Institut für Informatik
Universität Tübingen

Heinz Köhler
(Biologie)
Institut für Evolution und Ökologie
Universität Tübingen

Abstract

In this thesis I identify spatio-temporal patterns of European heat waves between 1979 and 2020, cluster them into meaningful families and analyze their influence on vegetation. For this I introduce a novel definition of heat waves, which defines them as phenomena with a spatial and temporal extend. With two clustering algorithms I cluster the detected heat waves temporally by their occurence during the year and spatially by their spatial overlapping. After the detection of these spatio-temporal heat wave clusters I correlate different heat wave features with normalized differenced vegetation index data anomalies to assess the influence of heat waves on vegetation in different regions in Europe and during different seasons of the year. The results reveal interesting insights into the spatio-temporal distribution of heat waves over Europe and suggest an approach on how to detect influences of heat waves on vegetation.

Zusammenfassung

In dieser Arbeit identifiziere ich räumlich-zeitliche Muster europäischer Hitzewellen zwischen 1979 und 2020, gruppiere sie in aussagekräftige Familien und analysiere ihren Einfluss auf Vegetation. Dazu habe ich eine neue Definition für Hitzewellen entwickelt, die diese als Phänomene mit einer räumlichen und zeitlichen Ausdehnung definiert. Mit zwei aufeinanderfolgenden Clustering-Algorithmen gruppiere ich die detektierten Hitzewellen zeitlich nach ihrem Auftreten im Jahresverlauf und räumlich nach ihrer räumlichen Überlappung. Nach der Erkennung dieser räumlich-zeitlichen Hitzewellencluster korreliere ich verschiedene Hitzewellenmaße mit Anomalien eines Datensatz des normalisierten differenzierten Vegetationsindex, um den Einfluss von Hitzewellen auf die Vegetation in verschiedenen Regionen Europas und während verschiedener Jahreszeiten zu bewerten. Die Arbeit gibt interessante Einblicke in die räumlich-zeitliche Verteilung von Hitzewellen über Europa und schlägt einen Ansatz vor, wie man Einflüsse von Hitzewellen auf die Vegetation bestimmen kann.

Acknowledgements

I thank my two supervisors, Bedartha Goswami and Heinz Köhler for the support and feedback during the time I wrote my thesis. Also I want to thank the members of the working group Machine Learning in Climate Science for the great and enriching time. I learned a lot! Especially I want to thank Felix Strnad for interesting ideas on how to perform the clustering of the heat waves. I want to thank Susi and Sophia for proof-reading and all other people that accompanied me during my time in Tübingen.

iv

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Extreme heat events pose a multifaceted threat to humans, ecosystems and the earth in general. During the European heat wave in 2003 the mortality in France increased by 54 percent and the alpine glacier loss due to this heat wave was estimated at 5 to 10 percent [Fischer et al., 2007]. The Russian heat wave in 2010 caused a crop deficit of approximately 25 percent [Barriopedro et al., 2011] and the 2018 heat wave over Scandinavia and northern Europe was the main driver for several forest fires that caused severe damage to vegetation and animals [Yiou et al., 2020].

Driven by antrophogenic climate change, extreme heat events have become more frequent and intense during the last decades [Robinson et al., 2021, Chapman et al., 2019, IPCC, 2021]. Dosio et al. show that even with only 1.5°C global warming the magnitude and frequency of extreme heat events will increase in most regions of the earth, with 2°C global warming this frequency is likely to double compared to 1.5°C warming [Dosio et al., 2018].

To date it has been difficult to predict the occurrence of heat waves longer than two weeks in adavance. This is due to the various atmospheric and physical drivers and accelerators of extreme heat events. The ability to predict heat waves would allow governments to take preventive measures in order to decrease mortality, crop loss and to prevent ecosystem damages during heat waves [Lavaysse et al., 2019]. The IPCC defines a heat wave as a period of abnormally hot weather [IPCC, 2021]. Besides that, there is currently no more precise general definition of heat waves. This causes inconsistencies in research results, as often results are not comparable with each other when different definitions of heat waves are applied.

In this work I define and identify European heat waves over the last 40 years and use two clustering algorithms to identify meaningful spatio-temporal heat wave clusters. In a second step, I assess the influence of these heat wave clusters on vegetation. The identification and analysis of the heat waves is conducted with a promising new graph framework, DeepGraphs, developed by

Dominik Traxl [Traxl et al., 2016a].

This thesis is structured as follows: In chapter   2 an overview of the current research of Eupean heatwaves, their influence on vegetation and their physical drivers is provided. In chapter 3 the Deep Graphs framework, which is used to analyze Euopean heatwaves is presented, followed by a comprehensive description of the data used. The results are given in chapter 4. A discussion of the results and a short outlook conclude this thesis.

Chapter 2

Climatic and Computational Background

2.1 Climatic Background

2.1.1 Heat Wave Definition and Measurements

Up to date several different indices are used to define and detect heat waves and therefore make it impossible to compare results between different studies. The measure by which heat waves are defined differs, often depending on the underlying research questions. Common measures are maximum temperature, mean temperature and minimum nighttime temperature. Many developed heat wave indices are not applicable for all climates and seasons, as they are based on absolute temperatures or they only consider one characteristic of a heat wave and therefore can not represent the entire scope of the heat wave [Perkins and Alexander, 2013]. Newer approaches towards an unified definition of heat waves use percentile-based daily thresholds [Russo et al., 2014, Zschenderlein et al., 2019]. This makes it possible to detect heat waves in winter season and in colder regions as well.
Many definitions of heat waves are made at the grid cell level and therefore are not able to take the spatial extend of a heat wave into account [Sutanto et al., 2020]. Other approaches define geographical regions beforehand and then analyze heat waves within these clusters [Perkins-Kirkpatrick and Lewis, 2020, Zschenderlein et al., 2019]. Lo et al. published an approach in which heat waves are not defined at the grid cell level, but are considered as events with spatial extents that span multiple grid points. They define a heat wave magnitude scale that includes the temporal extend by taking the first and the last day of a heat wave and the spatial boundaries as the longitudes and latitudes of the most outwards voxels [Lo et al., 2021].

Russo et al. proposed a heat wave magnitude index (HWMI) in 2014
[Russo et al., 2014] and published an improved version, the heat wave mag-
nitude index daily (HWMId) in 2015, which makes it possible to measure the
magnitude of heat waves at different locations and compare them with each
other [Russo et al., 2015]. This index takes into account the spatial and tem-
poral size of the heat wave as well as the intensity of the extreme heat during
the heat wave and combines these aspects to define the magnitude of a heat
wave. The magnitude of a heat wave day d is defined as followed, where T_d is
the temperature at day d, T_{40y25p} and T_{40y75p} are the 25th and 75th percentile
of the full 40 year temperature datset, respectively.

$$M_d(T_d) = \begin{cases} \frac{T_d - T_{40y25p}}{T_{40y75p} - T_{40y25p}}, & \text{if } T_d > T_{40y25p} \\ 0, & \text{if } T_d \leq T_{40y25p} \end{cases} \qquad (2.1)$$

The magnitude of a heat wave is the sum of all magnitudes of the days belong-
ing to the heat wave. This measure takes into account the temporal extend as
well as the extremity of the temperature of the heat wave [Russo et al., 2015].
However with this definition of the heat wave magnitude, winter heat waves
will most likely all have magnitudes of zero, which might not be reflecting the
real impact they have on ecosystems and earth processes.

2.1.2 Physical Causes of Heat Waves in Europe

The increase in the frequency of heat events occuring over Europe can par-
tially be explained by the mean temperature increase over the last decades,
caused by greenhous gas emission. However, several extreme heat events
that occured in the last decades can not solely be explained by this phe-
nomenon. Mega heat waves such as the 2003, 2010 and 2018 heat waves, are
caused by atmospheric regimes. Heat waves over Europe are closely associated
with atmospheric blocking events, stationary fields of extremely high pressure
[Schaller et al., 2018]. Blocking events hinder the westerly wind flow and there-
fore cause a weather regime to stay at one location for an unusually long time.
Since blocking events are also associated with a cloudless sky, the blocks are
often accompanied by precipiation deficits. Pfahl and Wernli show that up to
80 percent of the summer heatwaves over northern Euope can be traced back
to a co-located blocking event [Pfahl and Wernli, 2012, Brunner et al., 2017].
However, a blocking event alone might not be able to cause extreme events like
the heat waves in 2003, 2010 and 2018. For an event of such length and mag-
nitude, land-atmospheric feedback is of great importance [Durre et al., 2000].
When soil moisture is low, the incoming radiation of the sun is mainly re-
leased as sensible heat instead of latent heat, causing the surrounding air to
heat up. Land-atmospheric feedback alone can not cause a heat wave, but it
is the driving force to influence length and intensity of an extreme heat event

[Miralles et al., 2019]. For the record breaking European heat wave of 2003 it has been shown that land-atmospheric feedback has had a great influence on the magnitude of the heat wave [Fischer et al., 2007].

The influence of large-scale teleconnections, like the North Atlantic Oscillation, on heat waves over Europe has been an ongoing field of research over the last couple of years, especially because understanding how large-scale teleconnections influence extreme heat events would improve the forecasting of such events tremendously [Pezza et al., 2012]. The weather over Europe is greatly influenced by North Atlantic anomalies. Atmospheric blocking events over Europe have been shown to be associated with positive phases of the North Atlantic Oscillation (NAO$^+$) and low pressures systems over the north atlantic can cause heat waves over Europe [Cassou et al., 2005]. Other causes for blocking patterns over Europe are quasi-stationary Rossby waves and the formation of a double Jet pattern [Kornhuber et al., 2017].

In trecent years, drivers and causes of heat waves have been studied extensively, revealing new insights on a constant basis. However, it has not yet been revealed which cause or driver of heatwaves contributes to which attribute of a heat wave and to what extend a heat wave is influenced by one driver.

2.1.3 Heat Waves and Vegetation

Temperature is one of the main drivers affecting plant growth and therefore vegetation in general. Higher temperatures, which often go hand in hand with increased solar radiation, are primarily beneficial for plant growth, as the rate of photosynthesis is enhanced. Rising mean temperatures in the last decades have caused prolonged growing seasons in mitladitudial zones. In addition to the greening effect that rising temperatures bring, extreme heat causes stress in plants, leading to changes in metabolism and the formation of reactive oxygen species [Baumbach et al., 2017]. A physiological process occuring in all green plants is photosynthesis. Within this process, photosystem II is especially heat sensitive and its activity is reduced when the plant is exposed to extreme heat [Hasanuzzaman et al., 2013]. Baumbach et al. have shown that whether vegetation does benefit from heat or whether it does not, depends largely on the type of vegetation and the stage of plant development at the time of the extreme heat events. In early spring high temperatures often seem to have a beneficial effect on plant growth. Later in the year, up until september, some vegetation types, especially cropland and grassland, show negative reactions to extremely high temperatures [Baumbach et al., 2017].

Besides temperature, soil moisture and water deficits are main drivers of plant growth [Liu et al., 2013]. The intensity of extreme heat events is often amplified by soil moisture deficits and as heat waves over Europe are mainly caused by atmospheric blocking events they often occur together with precipitation

deficits [Durre et al., 2000]. Therefore it has been suggested that a mean temperature increase is beneficial for plant growth but extreme heat events together with water scarcity causes damage to vegetation [Liu et al., 2013].

2.2   Computational Background

All of the below described algorithms, frameworks and statistical methods were used throughout the data-analysis of this thesis. The theoretical background of all methods used is described in this section.

2.2.1   Deep Graphs Framework

Deep Graphs is a novel graphs framework developed for analyzing large amounts of heterogenous data. A graph (G) consists of nodes (V) and edges (E) (see equation 2.2).

$$G = (V, E) \tag{2.2}$$

Nodes represent objects, which can contain several features and any edge between two nodes V_i and V_j represents the existence of a relation between these two nodes. The relation can be of any type and does not neccessarily need to be represented by a number. Again, a relation between two nodes can have several features such as a strength or a boolean value. With the extensive Python module deepgraphs, graphs can be easily filtered and partitioned, allowing to find potentially interesting intersection partitions of a graph [Traxl et al., 2016a]. For my analysis, nodes represent extreme heat days at one grid cell. They contain features such as a temperature, an integer-based location and time, a magnitude and latitude and longitude coordinates. Neighboring nodes are connected by edges and all nodes connected by edges are represented by a supernode which is a heat wave in this context. A supernode again has features like a HWMId value, day of year mean and longitude and latitude means. Those features are calculated from the features of the nodes belonging to a supernode.

The Deep Graphs framework has been applied to rainfall data to analyze spatio-temporal patterns and to determine the size distribution of rainfall clusters [Traxl et al., 2016a, Traxl et al., 2016b]. In 2021 the methodology has been applied to fire data to analyze spatio-temporal patterns of wildfires in amazonian forests [Cano-Crespo et al., 2021].

2.2.2 K-means Clustering

K-means clustering is based on Lloyd's algorithm and is an unsupervised partitioning clustering algorithm. Its goal is to assign every datapoint to the cluster where the cluster variance is kept minimal. As this problem is NP-hard, the algorithm follows a heuristic approach and initially sets the cluster medeoids randomly. The number of clusters K needs to be pre-given. After initialization the algorithm has two phases. The datapoints are assigned to clusters and afterwards the medoids of the clusters are recalculated. Based on the recalculated medoids, some datapoints might be assigned to different clusters in the next iteration of the algorithm. These two phases are repeated until the clusters do not change anymore [Lloyd, 1982]. The algorithm is very fast and therefore suited for large datasizes, but has some disadvantages. As the algorithm follows a heuristic approach its outcome is not neccessarily the optimal solution, but highly depends on the location and number of cluster medoids [Arora et al., 2016].

2.2.3 UPGMA Clustering

UPGMA stands for *unweight pair group method with arithmetic mean* and is a hierarchial agglomerative clustering method. The algorithm starts with all datapoints apart and consecutively merges the two objects with the smallest distance to each other into the same cluster. The distance between two datapoints or clusters is defined in a distance matrix which is updated in every iteration of the algorithm. In the end, a rooted dendrogram is built and the user can decide where to set the cut within the dendrogram revealing cluster memberships of the data points [Sukal and Miechener, 1958]. Since the problem of finding the optimal dendrogram for a set of datapoints is NP-hard, the algorithm follows a heuristic approach. Once connections between data points and clusters have been drawn, they cannot be undone later, even if the connections turn out to be less favorable than before [Omran et al., 2007].

2.2.4 Spearman's Rank Correlation Coefficient

The Spearman's rank correlation coefficient is a non-parametric rank coefficient. It is derived from the Pearson correlation coefficient, but has the advantages that it neither assumes that the underlying data is normally distributed nor that the relationship between two variables is neccessarily linear [Xiao et al., 2010]. Before calculating the correlation coefficient, the samples are converted into ranks, between which the correlation is then calculated. As for the Pearson's correlation coefficient, values of the Spearman correlation coefficient lie between -1 and +1, where 0 implies that there is no correlation existing between two datasets and -1 and +1 imply perfect anticorrelation

and correlation respectively. The Spearman's rank correlation coefficient is calculated as follows, where R(X) and R(Y) are the ranks of the samples [Spearman and Spearman, 1904]:

$$r_s = \frac{cov(R(X), R(Y))}{\sigma_{R(X)}\sigma_{R(Y)}} \tag{2.3}$$

2.2.5   Mann-Kendall Trend Test

The Mann-Kendall trend test tests statistically, whether a variable shows a monotonic upward or downward trend over time. One difference compared to linear regression is that the monotonic trend of a variable does not necessarily have to be linear. The initial assumption of the Mann-Kendall trend test, H_0 is, that there is no monotonic trend [Mann, 1945].

Chapter 3

Data and Methods

3.1 Data

During this thesis two datasets, temperature data and normalized difference vegetation index (NDVI) data, were used. Both datasets are introduced and described in the following section.

3.1.1 Temperature Data

Gridded Era5 reanalysis 2m-land-temperature data was downloaded from the Copernicus climate data store and used for the analysis [Hersbach et al., 2018]. Hourly data between 12 pm and 5 pm of the years 1979-2020 were retrieved and daily maximum temperatures were computed. The spatial resolution of the data is 0.25°x0.25° in a gridded coordinate system. The margins of the assessed region are 75°N/25°S and -50°W/44°E. The temperature values were converted from kelvin to degrees celcius by subtracting 273.15 from every value. For simplicity, the 366th day was removed from each leap year.

3.1.2 Normalized Difference Vegetation Index (NDVI) Data

The normalized difference vegetation index is calculated from the near-infrared and red light reflection ratio (see equation 3.1).

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \tag{3.1}$$

where NIR and RED stand for the amounts of near-infrared and red light, respectively, that are reflected from vegetation and captured by the sattelite. The equation is based on the phenomenum that clorophyll absorbs red light

and mesophyll scatters near-infrared light. Chlorophyll is a green pigment that gives plant leafs its green color and takes part in photosynthesis. Mesophyll is part of the leaf structure of plants. The NDVI can have values between -1 and 1. Negative values imply the absence of green vegetation. The higher the NDVI value is, the greener is the vegetation at a location. Photosynthesis is the main metabolic pathway through which plants gain energy. The NDVI can therefore be used as an indicator of plant productivity [Pettorelli et al., 2005].

The GIMMS NDVI3g dataset was downladed from the IRI/LDEO Climate Data Library. The data is retrieved from Advanced very high resolution ratiometer (AVHRR) instruments run by the US-American national oceanic and atmospheric administration (NOAA) [Pinzon and Tucker, 2014, Pinzon and Tucker, 2016]. The dataset contains biweekly datapoints with a 0.083°x0.083° resolution and spans a period from July 1981 to December 2015. For the analysis, the data was downsampled to a 0.25°x0.25° resolution by calculating the mean value. Monthly anomalies of the NDVI values were computed.

3.2   Methods

The python code for the complete methodology used in this thesis is openly accessible in the following Github repository: `https://github.com/JuHe0311/Heatwaves.git`

The following section describes the three methodological parts of the analysis of my Master's thesis. First I explain my definition of a heat wave and how they are detected. Secondly the clustering of the heat waves into families and clusters is described and finally, the process of vegetation correlation analysis is explained.

3.2.1   Definition of Heat Waves

To detect extreme heat days, a location site specific daily threshold was calculated according to Russo et al. For this purpose, the 95th percentile of a window of 31 days around the day of interest was calculated for all years within the dataset according to equation 3.2, where $A_{d,g}$ is the threshold at day d and grid cell g and $T_{y,i,g}$ is the maximum temperature in year y, at day i and grid cell g. If the maximum temperature on any day at a grid cell exceeds its daily threshold, it is considered to be an extreme heat event. When viewed in a 3D coordinate system with longitude, latitude and time as the axes, each of these extreme events can have up to 26 nearest neighbors, eight within the same time (t) dimension and nine each in t-1 and t+1 (see figure 3.1). A heat wave is defined as all nearest neigbors in this space-time coordinate system. As I was interested in heat waves rather than small local and short-term heat

events, I defined temporal and spatial thresholds: An event needs to last for at least three consecutive days and cover at least 100 unique grid cells to be considered a heat wave.

$$A_{d,g} = \cup_{y=1979}^{2020} \cup_{i=d-15}^{d+15} T_{y,i,g} \tag{3.2}$$
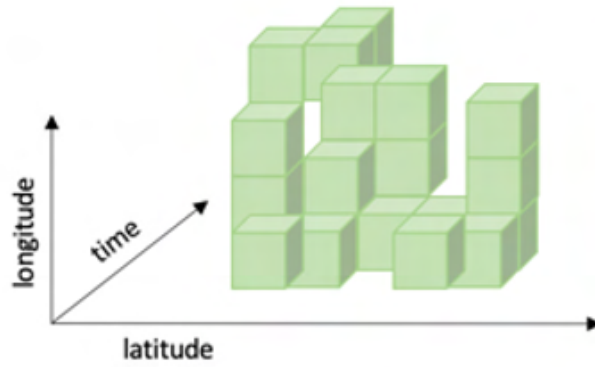
Figure 3.1: 3D representation of a heat wave. An example of a heat wave in a 3D coordinate system is given. Extreme heat events are represented as cubes and a heat wave is the union of all nearest neighbors of the extreme heat events. The three dimensions are the longitude, latitude and time of an extreme heat event.

3.2.2 Spatial and Temporal Clustering of Heatwaves

To cluster the heat waves spatially and temporally, two subsequent clusterings were applied. Spatio-temporal clustering was not possible in one clustering step as for both measures two different approaches were chosen. Temporal clustering was perfomed based on one feature of a heat wave and spatial clustering is based on the distances between each pair of heat waves. These two measures could not be combined in an easy way, therefore the choice was made to split the clustering into two separate steps. First the heat waves were clustered temporally with the K-means algorithm. For this the mean day of year

(doy) of every heat wave was mapped onto a circle. Every doy was expressed
by two values:

$$cos(doy) = cos(\frac{2doy * \pi}{365})$$ (3.3)

$$sin(doy) = sin(\frac{doy}{365 * 2\pi})$$ (3.4)

With this transformation it was achieved that the first and the last day of the
year are direct neighbors and are therefore considered similar in the clustering.
K-means clustering was performed with cos(doy) and sin(doy) as features. K
was set to four, as the four seasons have the strongest influence on European
weather. The clustering was repeated 100 times and the best result was re-
turned. The four clusters resulting from clustering will hereafter be referred
to as heat wave families.

After the first clustering step, UPGMA clustering was performed on every heat
wave family resulting from the K-means clustering separately. This clustering
step intended to capture the spatial distribution of the heat waves. A distance
matrix with distance measures between every pair of heat waves was created
based on the intersection strength of the heat waves, originally defined for
rainfall clusters by Traxl et al. [Traxl et al., 2016a].

The intersection strength between two heat waves i and j is defined as followed:

$$IS_{ij} = \frac{IC_{ij}}{min|L_i^{set}|, |L_j^{set}|}$$ (3.5)

where L_i^{set} and L_j^{set} are the sets of the grid cells that are contained in the heat
wave i and j. IC is the intersection cardinality between two heat waves, which
is defined as:

$$IC_{ij} = |L_i^{set} \cap L_j^{set}|$$ (3.6)

3.2.3   Vegetation Correlation Analysis

To analyze the influence of extreme heat events on vegetation, I calculcated
the spearman's rank correlation coefficient between two features of the heat
waves and NDVI monthly anomalies. The features of heat waves are summed
up over one season. A season is defined as the 10th and 90th percentile of
the monthly distribution of heat wave days within a heat wave family. One
feature is the HWMId sum calculated for one grid point and the other feature
is the number of extreme heat days for one grid point. To reduce noise in the
correlation analysis the heat wave clusters were edited. First all ocean grid
points were removed from the clusters. The Copernicus land-sea-mask dataset
was used for this purpose.All grid cells that have a value ≤ 0.5 are considered to

be predominantly ocean covered and are therefore removed from the clusters. Additionally, to obtain more centered clusters, all grid cells in which five or fewer heat waves occurred were removed from the cluster. By doing so, I hope to only analyze correlations of the grid cells that are significantly affected by heat waves in this cluster. The NDVI anomalies at the end of the respective season are retrieved and correlated with each of the two heat wave features. The p-value is calculated and only significant correlation values are used for analysis. The significance threshold for the correlation coefficient is set to $\alpha = 0.01$.

Chapter 4

Results

4.1 Heat Wave Detection

Using the definition of heat waves described above (3.2.1), the dataset identified 6,856 heat waves over Europe from 1979 to 2020. Among the 15 most intense heat waves, based on HWMId values, are several well known and studied heat waves of the past decades. All of the in figure 4.1 displayed heat waves were extensively studied in research papers [Russo et al., 2015, Yiou et al., 2020, Mecking et al., 2019, Bonne et al., 2015, Fischer et al., 2007, Barriopedro et al., 2011]. The 2003 heat wave affected mainly Western Europe and caused a massive increase in heat related mortality in France [Fischer et al., 2007]. 2010, a record breaking heat wave in Russia and Eastern Europe caused severe crop failure and forest fires [Barriopedro et al., 2011, Russo et al., 2015]. The summer heat wave in 2012 over Greenland lead to snow melt in an extend that has never been witnessed before [Bonne et al., 2015].
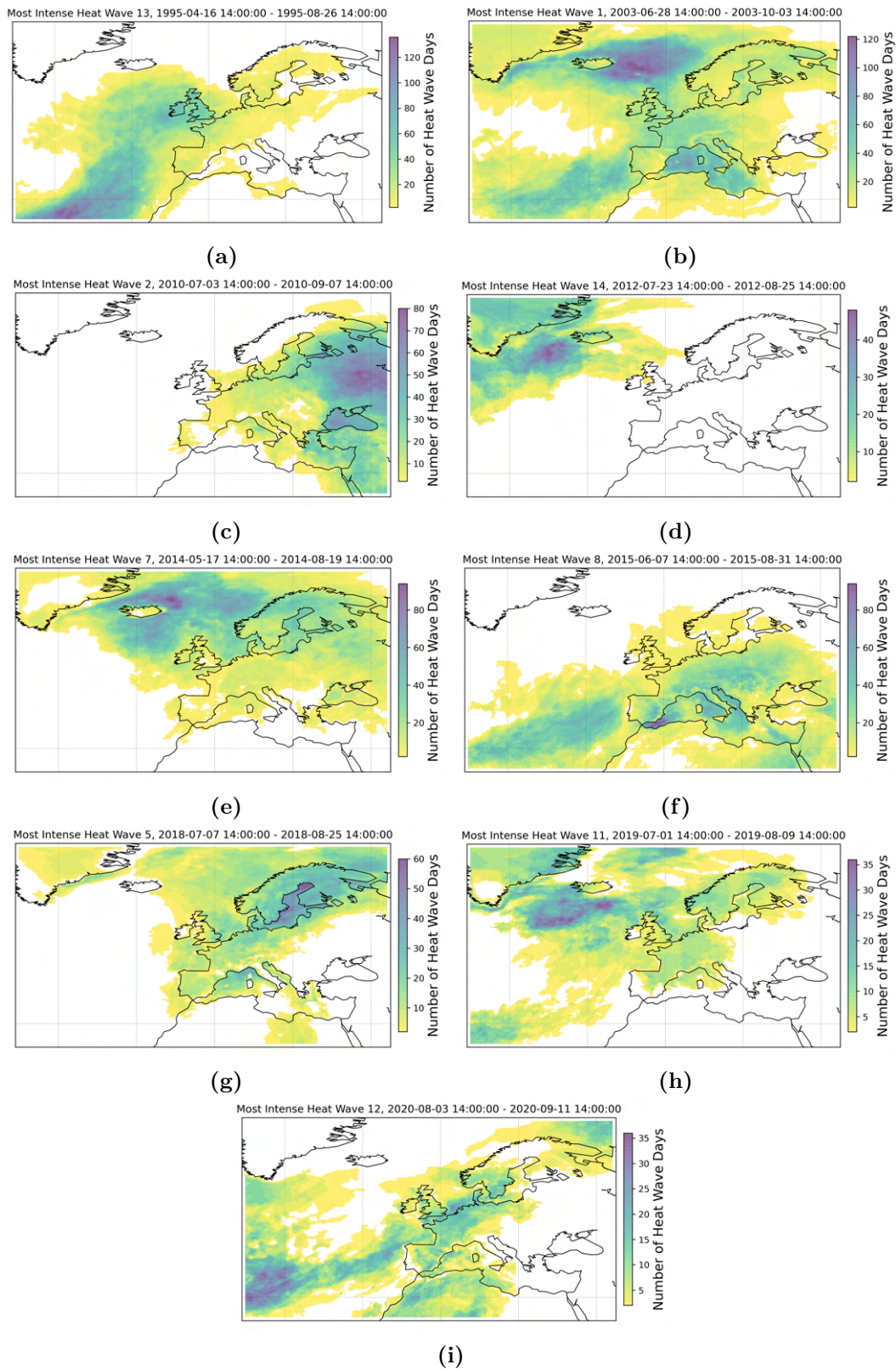
Figure 4.1: Spatial distribution of well known summer heat waves.
Nine summer heat waves that were found in my dataset but also reported in
other literature are shown here. The number of heat wave days represents the
number of extreme heat events at one grid cell: a) 1995 heat wave over Great
Britain, b) 2003 Western European heat wave, c) 2010 Russian heat wave, d)
2012 Greenland heat wave, e) 2014 Scandinavian heat wave, f) 2015 Central
European heat wave g) 2018 Scandinavian heat wave, h) 2019 Western European
heat wave i) 2020 Western European heat wave

Table 4.1 shows the basic characteristics of the five largest heat waves as measured by the number of heat events within a heat wave. The number of heat events is the number of nodes contained in one supernode. This means that the number of heat events is the sum of all individual heat events on a grid cell and day level that belong to one heat wave. Three of the five heat waves are also amongst the well known heat waves with the highest HWMId values depicted in figure 4.1. However, the dataset reveals other large heat waves that have not yet been in the focus of attention, such as the winter heat wave in 2020, which lasted 53 days and covered all of mainland Europe, but especially Eastern Europe.

Number	Start Time	End Time	Number of Heat Events	Timespan
1	28.06.2003	03.10.2003	1,362,710	98 days
2	03.04.2018	30.06.2018	870,226	89 days
2	16.04.1995	26.08.1995	785,314	133 days
3	17.05.2014	19.08.2014	764,550	95 days
4	22.07.2010	21.10.2010	734,818	92 days
5	25.01.2020	17.03.2020	721,998	53 days

Table 4.1: Characteristics of the five largest heat waves. The five largest heat waves detected in the dataset, measured on the number of heat events belonging to a heat wave are displayed here. The start and end time of the heat wave as well as the time span are given.

Figure 4.2 shows that the amount of heat waves happening throughout a year has increased over the years. Most heat waves that do not occur in summer have a magnitude of zero. As the magnitude of a heat wave is dependent on the spatial and temporal extend of a heat wave it was to be expected that the number of heat events belonging to a heat wave and the timespan of a heat wave slightly correlate with the HWMId. The distribution of heat waves over a year shows that there are slightly more heat waves in summer than in other seasons. However, there is a considerable number of heat waves in the other seasons. From the last column of figure 4.2 it can be seen that the intensity, timespan and spatial extend of heat waves appear to be increasing over the years.
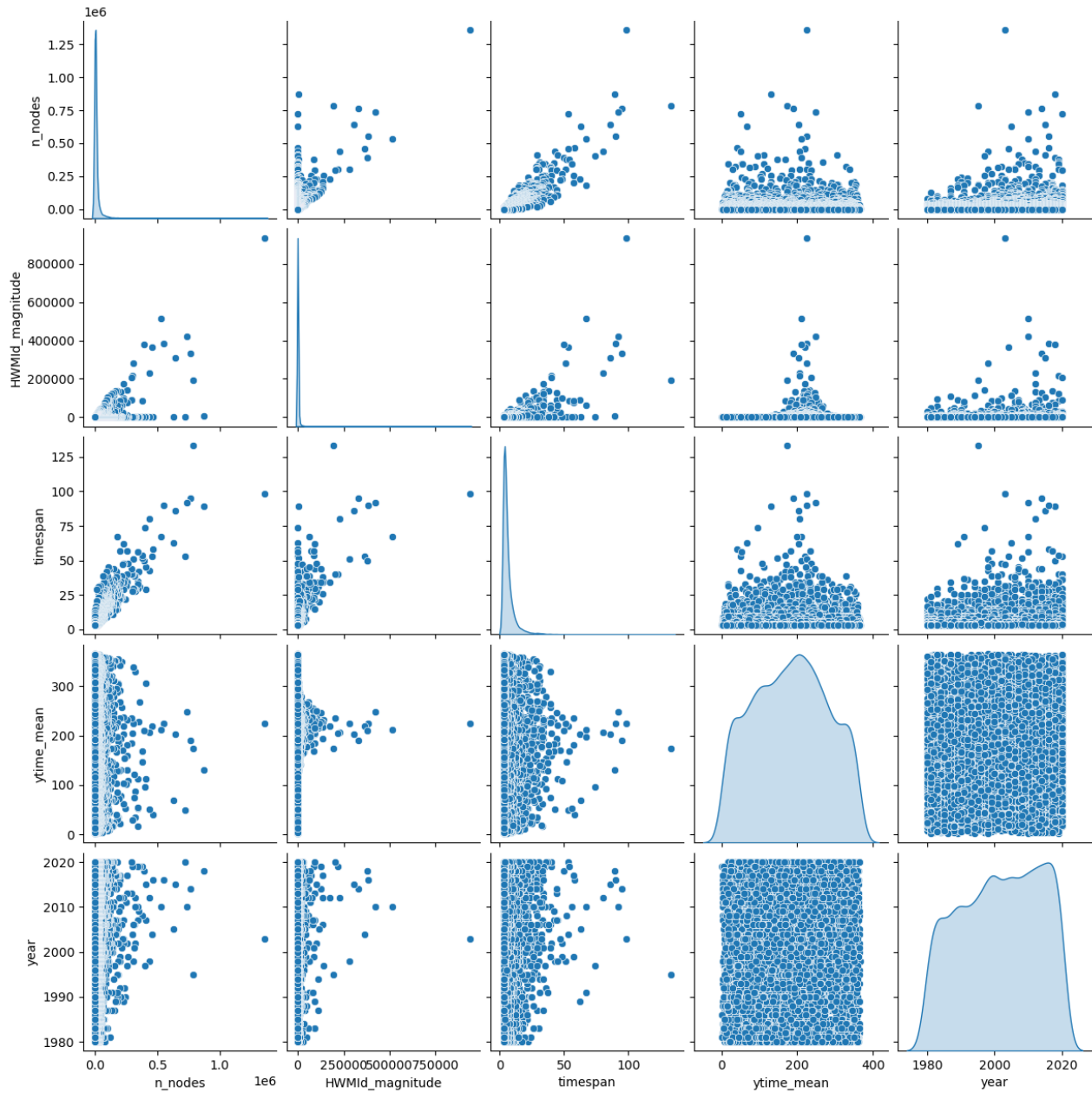
Figure 4.2: General statistics of the detected heat waves from 1979 to 2020. Features of the heat waves (number of heat events within a heat wave (n nodes), magnitude of a heat wave (HWMId magnitude), timespan of a heat wave, mean day of year of a heat wave (ytime mean) and year of occurence of a heat wave (year)) are plotted. The diagonal shows the distribution of the respective feature and the pairwise relationships between two features are plotted in the lower and upper triangle matrix.

4.2   Heat Wave Clustering
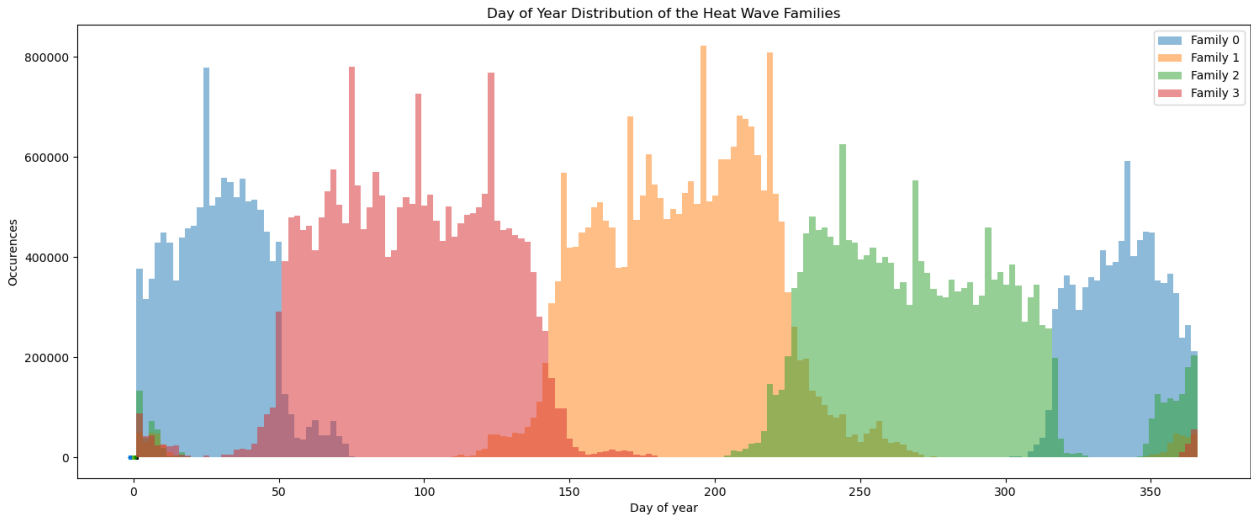
4.2.1   K-means Clustering

The heat waves were clustered by their mean day of year with the K-means clustering algorithm. K was set to four. In figure 4.3 the day of year distribution of the heat events belonging to the four heat wave families is shown. The four families represent the four seasons that mainly spelling central European weather. Family 0 contains winter heat waves and spans from December to January, family 1 is the summer heat wave family containing the most famous heat waves and spans from May to August. Family 2 and 3 contain fall and spring heat waves which span from August to November and from March to May respectively. The summarized characteristics for the four heat wave families can be found in table 4.2. The spatial distribution of the four heat wave families is depicted in the appendix in figure A.1.

Attempting to include more features into the K-means clustering in order to cluster heat waves by seasonality and length or intensity did not change the results of the clustering. This is due to the fact that the length of the heat waves is evenly distributed in the different seasons, which means that in all seasons there are few very long heat waves and many small heat waves. Including the magnitude of heat waves (HWMId sum) did not change the clustering either, since heat waves in winter, spring, and fall are mostly zero in magnitude.

When looking at the spatial distribution it is interesting to compare family 0 and 1 with each other. Familiy 0, the winter heat wave family, has a slight northwards tendency and tends to often center over the atlantic. In contrast, summer heat waves tend to occur over land and leave out the ocean. Also a southwards tendency in summer can be observed (see figure A.1 in the Appendix). However, there is no possibility to detect spatial clusters based solely on K-means clustering, therefore spatial clustering was performed with UPGMA clustering in a second step.

Family	Number of Heat Waves	Timespan
0	1513	December to January
1	1932	May to August
2	1708	August to November
3	1703	March to May

Table 4.2: Characteristics of heat wave families. Basic characteristics of the four heat wave families resulting from K-means clustering of the dataset from 1979 to 2020. The timespan is given by the months of the 10th and 90th percentile of the distribution of heat days belonging to one heat wave family.

Figure 4.3: Day of year distribution of heat wave families. The histogram shows the distribution of heat wave days of the whole dataset during the year. The four colors indicate the four heat wave families after K-means clustering.

4.2.2   UPGMA Clustering

After K-means clustering every heat wave family was clustered individually with UPGMA clustering to find meaningful spatial heat wave clusters. The threshold for the number of UPGMA clusters per heat wave family was chosen to be ten, based on previous research, in which Europe was divided into six regions [Zschenderlein et al., 2019, Stefanon et al., 2012]. As the margins of my dataset were broader, e.g. Greenland, North Africa and the Atlantic ocean are included in my dataset, I chose to add four additional regions.

Figures 4.4 to figures 4.7 show the spatial distribution of the heat wave clusters of the four heat wave families. The colors indicate the numbers of heat waves that occur in a grid cell over the whole time span. Only the clusters that are relevant for the subsequent vegetation correlation analysis are shown here. All other clusters that are ocean- or Greenland based can be found in the appendix in figures A.2 to A.5. Clusters over Greenland, Scandinavia, North Africa and Central Europe can be found in every heat wave family. Heat wave clusters over Great Britain, Central/Eastern Europe and Italy are present in more than one but not all heat wave families. Other heat wave clusters like the Sardegnan spring cluster or the Icelandic summer cluster are only present in one heat wave familiy, meaning that this heat wave cluster only exists in one specific season. A summary of the occurences of heat wave clusters can be found in table 4.3.

The heat wave clusters of every family are numbered by size, the smaller the number, the larger the cluster. It is noticeable that the Scandinavian, the North African cluster and the Greenland cluster are amongst the largest four clusters in all heat wave families.

Cluster Name	Family 0 (Winter)	Family 1 (Summer)	Family 2 (Fall)	Family 3 (Spring)
Scandinavia	Yes	Yes	Yes	Yes
Great Britain	No	Yes	No	Yes
Iceland	No	Yes	No	No
Eastern Europe	Yes	No	No	No
Central/East Europe	No	Yes	Yes	Yes
Central/West Europe	Yes	No	No	No
South France	No	No	Yes	No
Italy	Yes	Yes	No	Yes
Sardegna	No	No	No	Yes
Turkey/Middle East	No	No	Yes	No
North Africa	Yes	Yes	Yes	Yes
North Africa and Spain	Yes	Yes	Yes	Yes

Table 4.3: Summary of UPGMA clustering. The occurences of the different spatial heat wave clusters in the four heat wave families are summarized. The clusters are named based on their spatial center. Yes means the heat wave cluster occurs within a heat wave family, No means it does not.

(a) Cluster 0

(b) Cluster 2

(c) Cluster 4

(d) Cluster 5

(e) Cluster 6
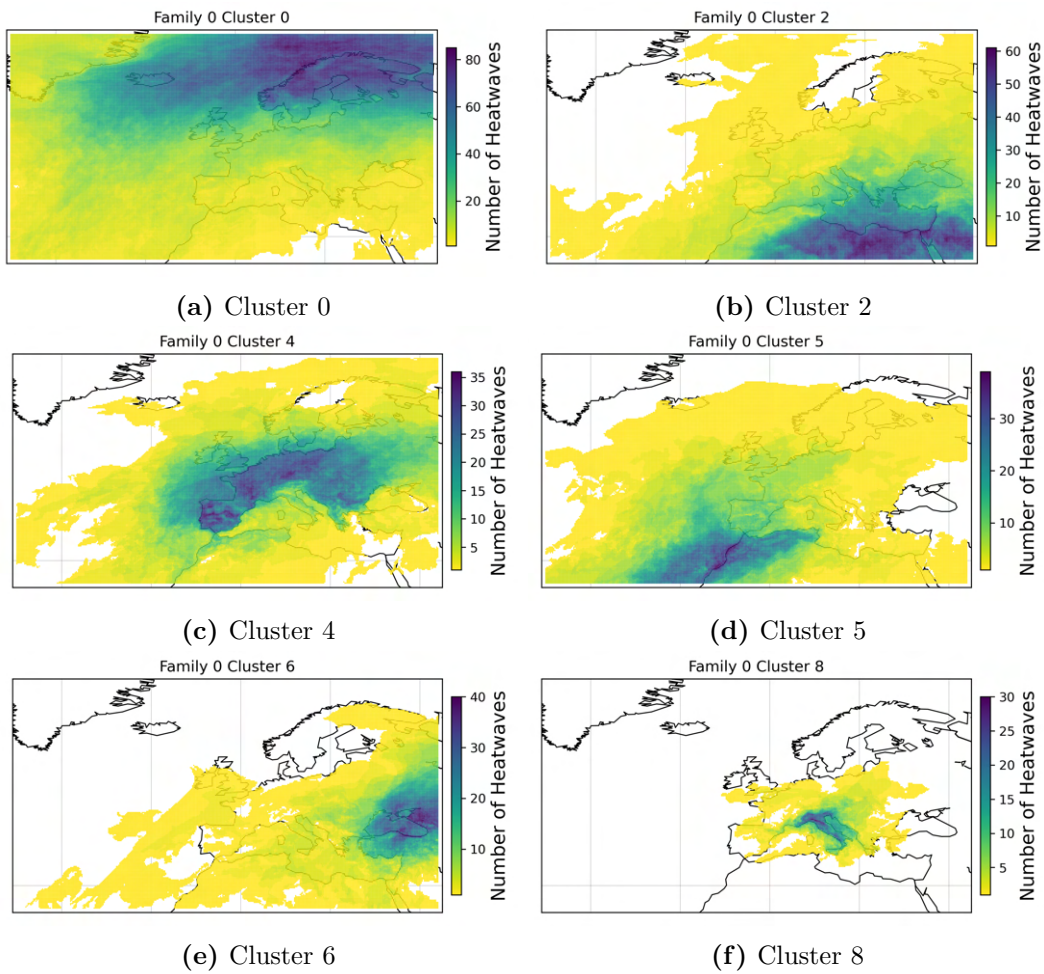
(f) Cluster 8

Figure 4.4: Spatial heat wave clusters of family 0. Spatial distribution of heat wave clusters of family 0 after UPGMA clustering. The number of heat waves gives the information on how often a grid cell is affected by different heat waves. Only the land clusters that are used for NDVI correlation analysis are shown here. All other clusters can be found in the Appendix in figure A.2.

(a) Cluster 0

(b) Cluster 2

(c) Cluster 3

(d) Cluster 4

(e) Cluster 5

(f) Cluster 7
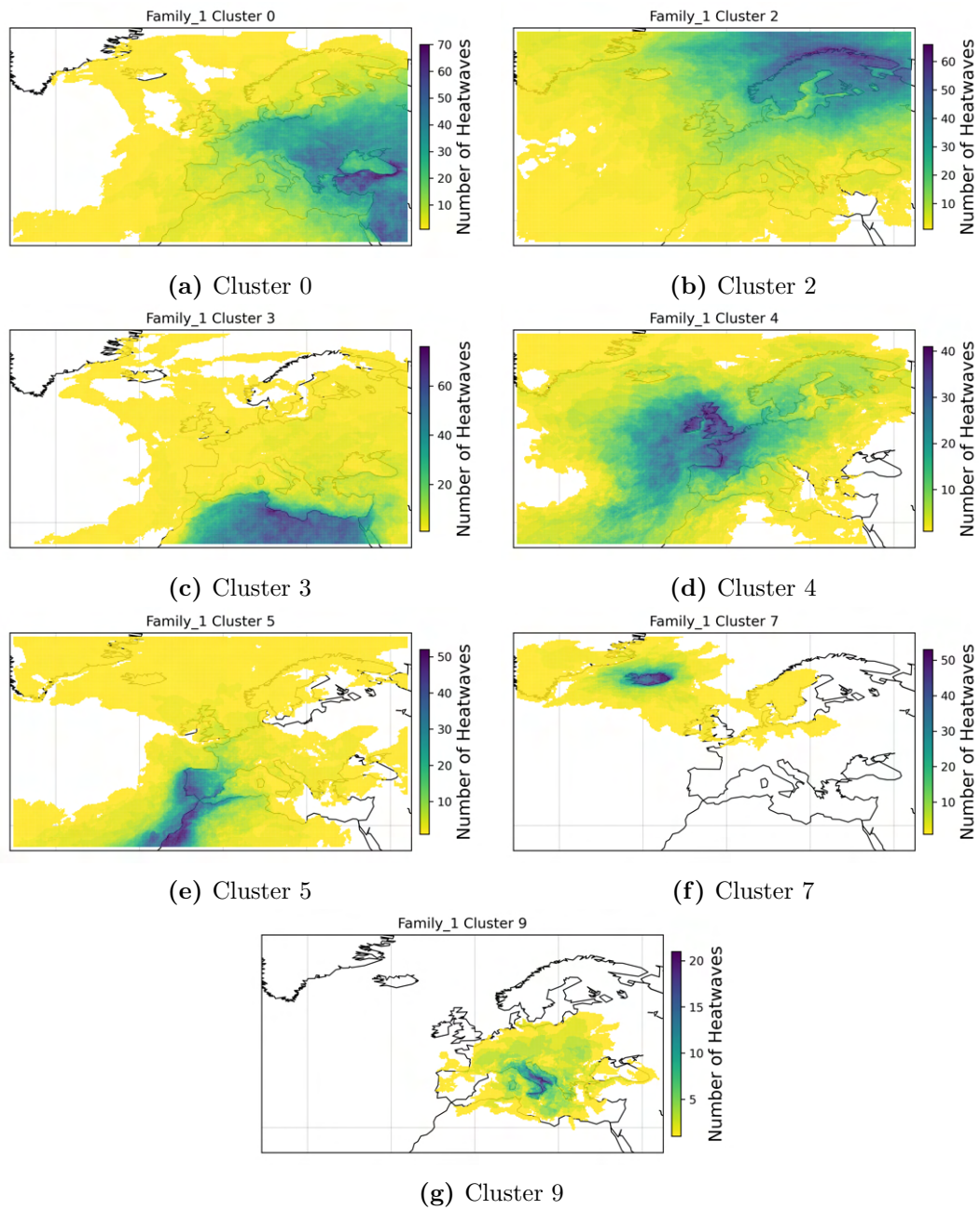
(g) Cluster 9

Figure 4.5: Spatial heat wave clusters of family 1. Spatial distribution of heat wave clusters of family 1 after UPGMA clustering. The number of heat waves gives the information on how often a grid cell is affected by different heat waves. Only the land clusters that are used for NDVI correlation analysis are shown here. All other clusters can be found in the Appendix in figure A.3.

(a) Cluster 0

(b) Cluster 1

(c) Cluster 4

(d) Cluster 6

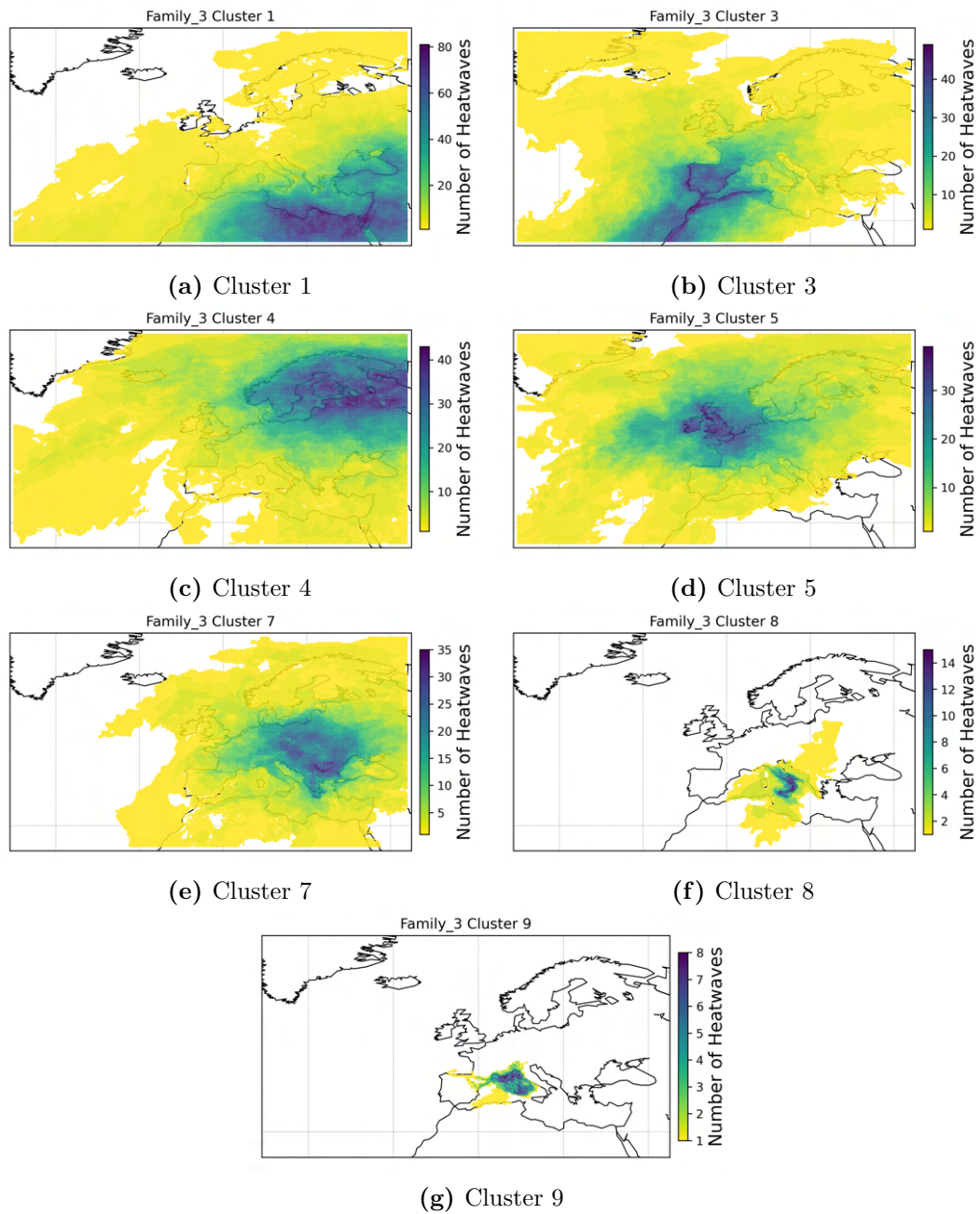**(e)** Cluster 7
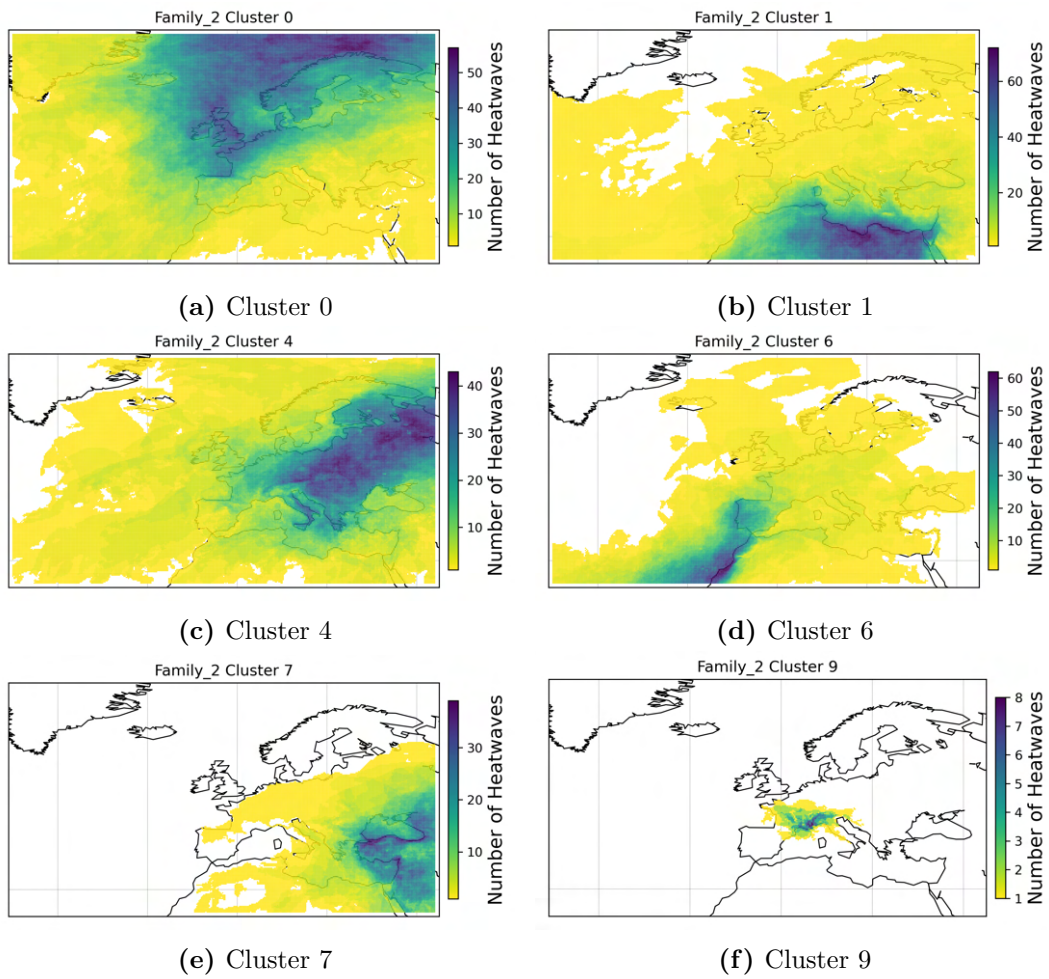
(f) Cluster 9

Figure 4.6: Spatial heat wave clusters of family 2. Spatial distribution of heat wave clusters of family 2 after UPGMA clustering. The number of heat waves gives the information on how often a grid cell is affected by different heat waves. Only the land clusters that are used for NDVI correlation analysis are shown here. All other clusters can be found in the Appendix in figure A.4.

(a) Cluster 1

(b) Cluster 3

(c) Cluster 4

(d) Cluster 5

(e) Cluster 7

(f) Cluster 8

(g) Cluster 9

Figure 4.7: Spatial heat wave clusters of family 3. Spatial distribution of heat wave clusters of family 3 after UPGMA clustering. The number of heat waves gives the information on how often a grid cell is affected by different heat waves. Only the land clusters that are used for NDVI correlation analysis are shown here. All other clusters can be found in the Appendix in figure A.5.

4.3   Vegetation Correlation Analysis

4.3.1   Mean Heat Wave Correlation Analysis

The mean of all significant Spearman's rank correlation coefficients over all years for each cluster from each heat wave familiy was calculated. Only mean correlations that were calculated from 30 or more individual correlation coefficients were considered in the analysis.

In heat wave family 1, which is the summer heat wave family, cluster 2 and 3 show significant correlations with NDVI data. However, this mean correlation is fairly weak. Compared to cluster 2, cluster 3 shows a higher positive mean correlation coefficient for the heat wave magnitude correlation (see table 4.4). Positive correlation in this context means that higher heat wave magnitude correlates with higher NDVI values, indicating greener vegetation. Negative correlation, as can be seen for cluster 1 in heat wave family 2, indicates that the more number of heat wave events at one grid cell occured the lower the NDVI values (see table 4.5). Both clusters with higher correlation coefficient values (cluster 3 from family 1 and cluster 1 from family 2) are clusters located over North Africa.

Cluster	NHWE Mean Correlation	HWMId Mean Correlation
2	-	-0.02
3	0.02	0.12

Table 4.4: Spearman's rank correlation coefficients of heat wave family 1. Mean spearman's rank correlation coefficient over all years for heat wave family 1. The mean is calculated from significant correlation coefficients of one cluster. Only clusters that have at least 30 significant correlation coefficients are shown here.Correlations between the features number of heat wave events per heat wave (NHWE) and the magnitude of a heat wave (HWMId) are calculated.

| Cluster	NHWE Mean Correlation	HWMId Mean Correlation
1	-0.1	-

Table 4.5: Spearman's rank correlation coefficients of heat wave family 2. Mean spearman's rank correlation coefficient over all years for the clusters of heat wave family 2. The mean is calculated from significant correlation coefficients of one cluster. Only clusters that have at least 30 significant correlation coefficients are shown here. Correlations between the features number of heat wave events per heat wave (NHWE) and the magnitude of a heat wave (HWMId) are calculated.

To answer the question whether the correlation between NDVI and heat wave measures within the different clusters changes over time, the time series progression of the correlation coefficient values together with a linear regression

line were plotted. Using the Mann-Kendall trend test I tested the significance of the slope of the regression line. The slope of the regression line was not significant for any of the correlation coefficient time series and is therefore not shown here.

4.3.2 Individual Heat Wave Correlation Analysis

In this section the individual correlation coefficient values of heat waves from all families with NDVI values are analyzed. Two different approaches were chosen to analyze NDVI and heat wave correlation more individually. First the dataset was split into three parts, to assess whether the correlation of NDVI values and heat wave features changes over time. For this I divided the dataset into three parts of ten to eleven years each (1981-1992, 1993-2003, 2004-2015) and calculated significant correlation coefficients. Boxplots of the correlation coefficient distributions for all three datasets were plotted and can be seen in figure 4.9 and figure 4.8. In heat wave family 1 an increasing mean correlation coefficient can be observed for correlation of NDVI and heat wave events (see figure 4.8b). For all other correlations there is no trend observable.
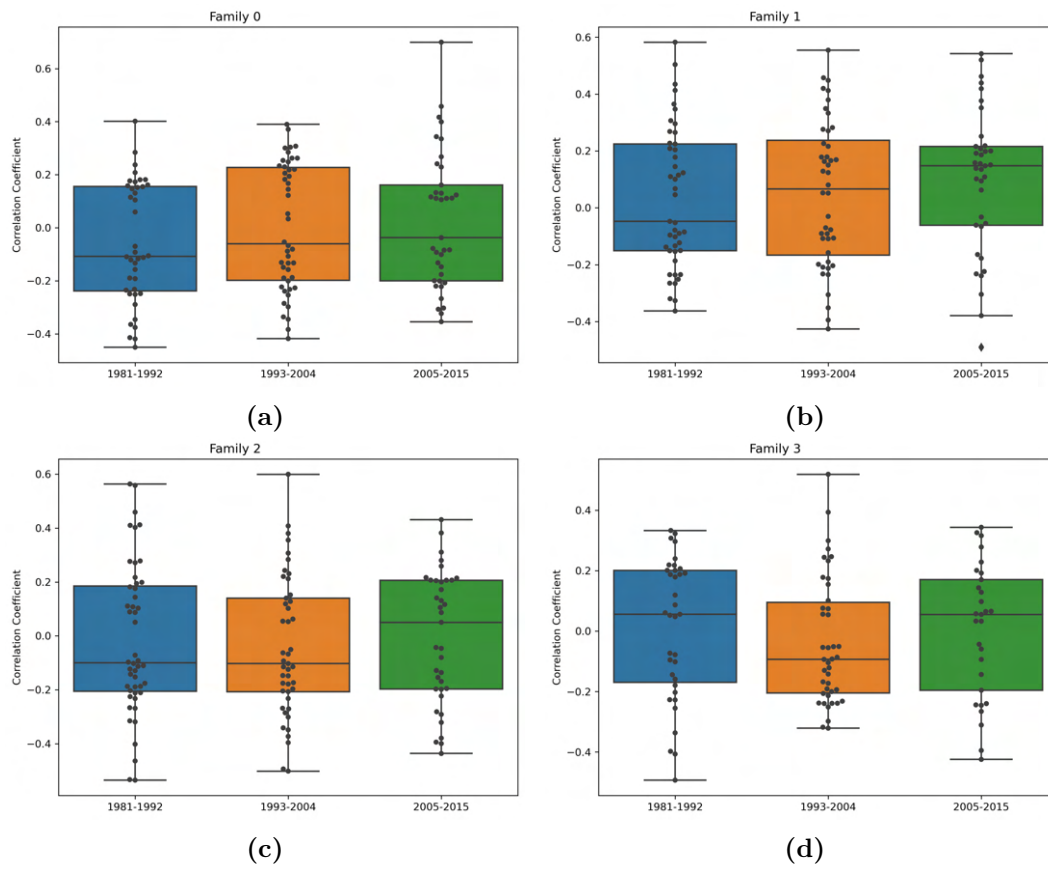
Figure 4.8: Boxpots of correlation coefficients (NHWE). Comparison of correlation coefficient value distributions of number of heat wave events correlation for: a) Family 0 (winter), b) Family 1 (summer), c) Family 2 (fall), d) Family 3 (spring). The dataset is split into three parts to compare the development of the correlation coefficients over time.

(a)

(b)

(c)
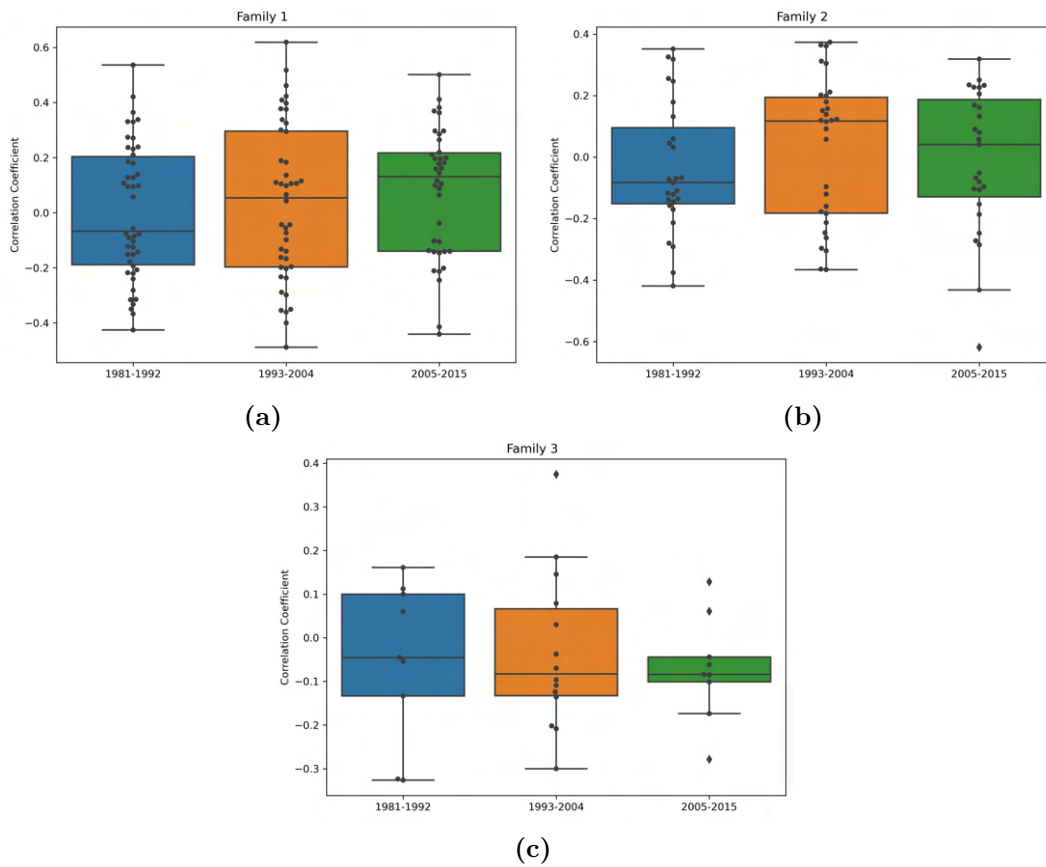
Figure 4.9: Boxpots of correlation coefficients (HWMId). Comparison of correlation coefficient value distributions of HWMId correlation for: a) Family 1 (summer), b) Family 2 (fall), c) Family 3 (spring). The dataset is split into three parts to compare the development of the correlation coefficients over time.

In a second approach the heat waves for the strongest correlating years and clusters were plotted individually. In family 0 and 3 (winter and spring) no high correlations between NDVI and HWMId could be found. The highest individual correlation found over all families is a winter heat wave cluster over Greenland in 2010 (see figure 4.10a). Across all families, it is noticeable that, with one exception, all highest correlating heat waves are either located in the south (North Africa, Spain) or in Scandinavia/Greenland. Independent of the family, correlation coefficients between heat wave features and NDVI values over Scandinavia are positive. In spring, the highest correlation coefficients are all negative (see figure 4.13) as well as the correlation coefficients for the southern heat waves in fall, as can be seen in figure 4.12.
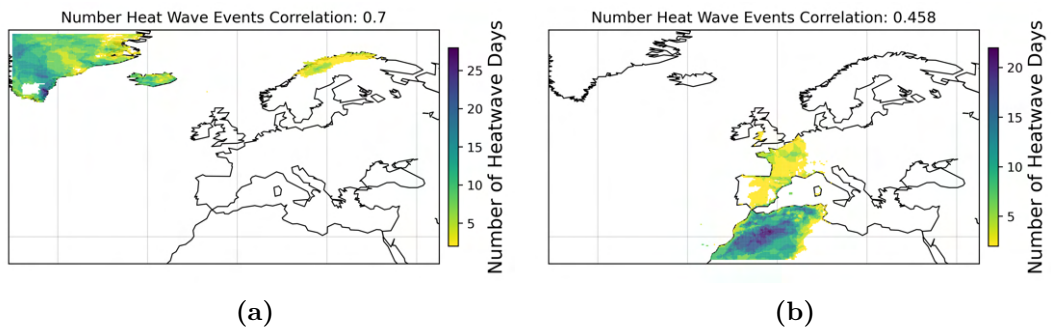
Figure 4.10: Highest individual correlating heat waves of family 0.
Highest correlating heat waves of family 0: a) 2010, heat wave cluster 0, b)
2007, heat wave cluster 5.
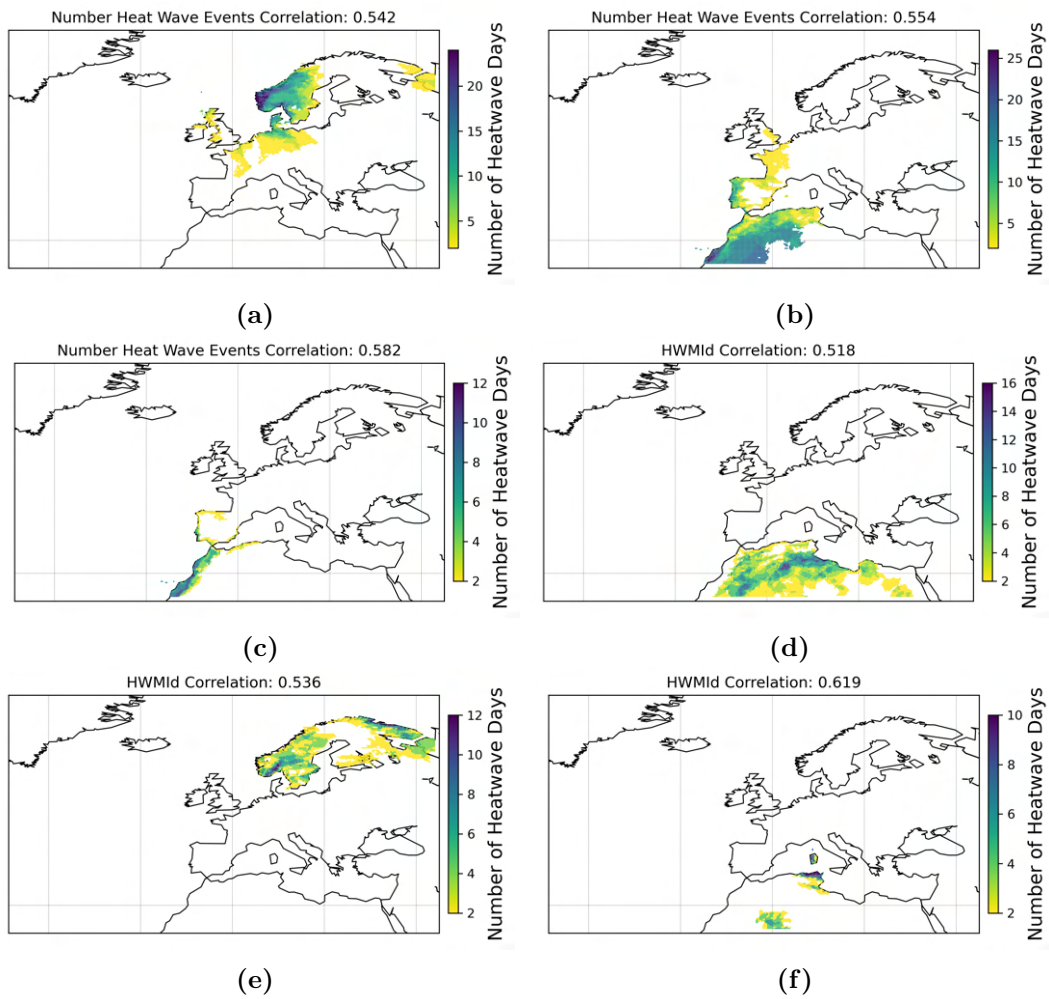
Figure 4.11: Highest individual correlating heat waves of family 1.
Highest correlating heat waves of family 1: a) 2008, heat wave cluster 2 b) 1996, heat wave cluster 5, c) 1982, heat wave cluster 5 , d) 2001, heat wave cluster 3, e) 1983, heat wave cluster 2, f) 1998, heat wave cluster 3.
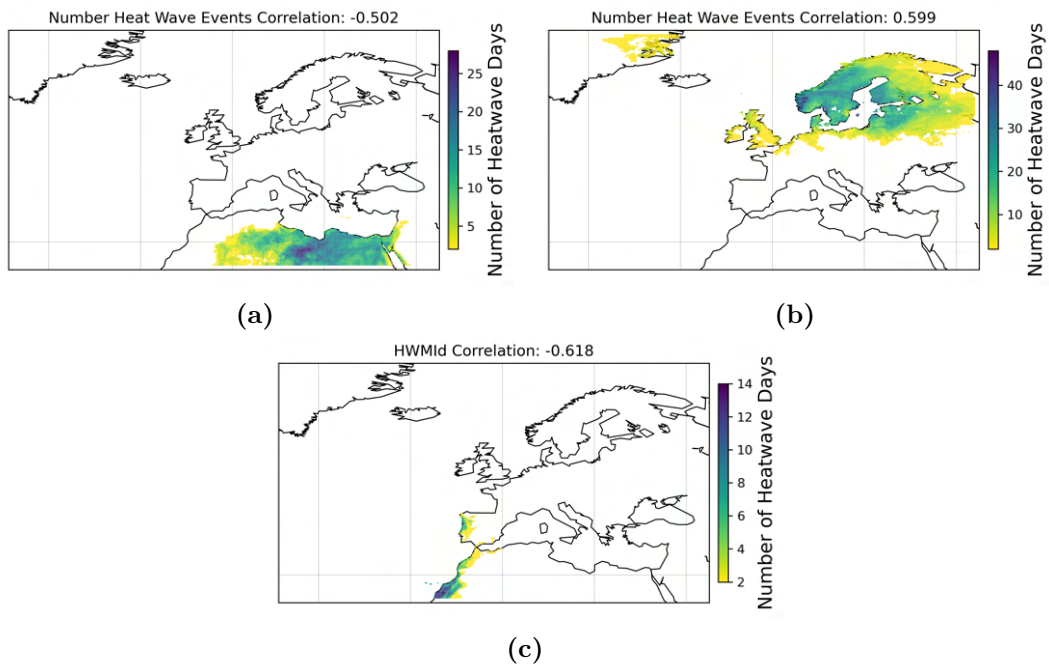
Figure 4.12: Highest individual correlating heat waves of family 2.
The spatial distributions of the highest correlating heat waves of family 2 are
shown: a) 1996, heat wave cluster 1, b) 2002, heat wave cluster 0, c) 2005, heat
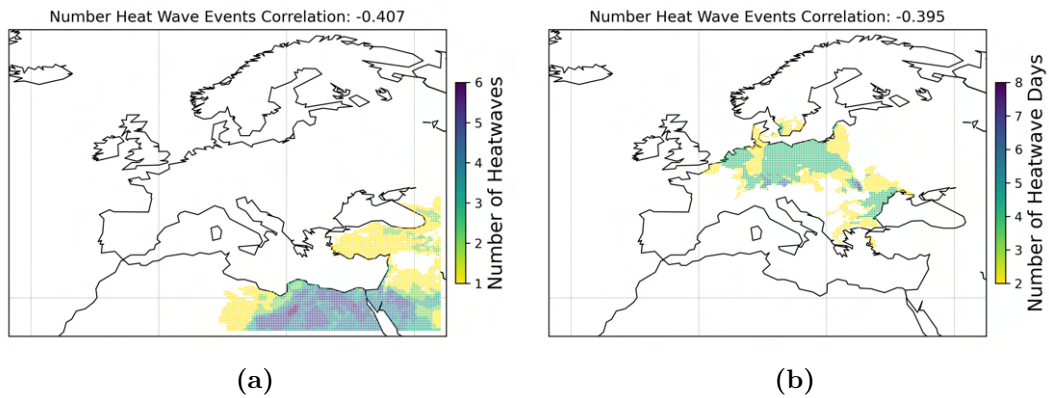wave cluster 6.

**Figure 4.13:   Highest individual correlating heat waves of family
3.**Highest correlating heat waves of family 3: a) 1991, heat wave cluster 1,
b) 2010, heat wave cluster 7.

Chapter 5

Discussion and Outlook

5.1 Definition and Identification of Heat Waves

In this work, I have developed a new approach to define and identify heat waves as phenomena that have spatial extents, rather than considering heat events only at the grid cell level. With this approach, I also do not rely on predefined fixed boundaries of regions, which makes spatio-temporal clustering of heat waves impossible. By detecting days of extreme heat with a quantile based daily and grid cell dependent threshold, I was able to detect heat waves in cold climates and heat waves throughout the whole year. Even though the distribution of the heat waves mean day of year peaks in summer, there are significant numbers of winter, spring and fall heat waves identified within the dataset (see figure 4.2). Amongst the 6,856 detected heat waves between 1979 and 2020 in Europe are all of the famous summer heat waves that were examined extensively in prior research. This shows that my approach to define and detect heat waves works. In addition to the well-known summer heat waves, my dataset revealed several other heat waves that have not been reported in literature. Among the five largest heat waves, in terms of spatial extend, there is one winter heat wave in 2020 that lasted from January to March (see table 4.1). This winter heat wave may not have been as severe when only considering heat-related deaths, wildfires, or crop failures, but these no less severe heat waves have other impacts that should not be discounted. Unusually warm temperatures in winter followed by colder temperatures cause damage to flora and fauna, as the growing season of plants starts too early and is then interrupted by another cold spell. Additionally winter and spring heat waves can cause ice sheet melting over Greenland. For future research it would be important to detect these heat waves and assess their impact on climate, ecosystems as well.

A striking difference between my approach of defining heat waves and previous

33

definitions is that the heat waves detected here are all significantly longer than the same heat waves reported in other literature. It is likely that a heat wave in my definition is seen as two or more individual heat waves in other definitions. The HWIMd defined by Russo et al. is based on a grid cell definition of a heat wave as well. Therefore it might have to be adapted in order to fit to my definition of heat waves. A weakness of the HWMId measure in general is that, winter and spring heat waves are rarely warm enough on an absolute scale to have a magnitude > 0. In my opinion, this undermines their severity and impact at different scales.

5.2   Spatio-Temporal Heat Wave Clusters

K-means clustering revealed the dependence of heat waves over Europe on the four seasons that mainly shape and influence European weather. The heat waves were divided into a winter, spring, summer and fall familiy. I tried to incorporate other attributes of heat waves as features into the K-means clustering in order to improve the clustering results. However, neither HWMId magnitude, nor timespan of the heat wave changed the outcome of the clustering in any way, implying that the mean day of year is the only important feature causing the variance in the data. Figure 4.2 shows that, with a few exceptions, the distribution of the timespan of the heat waves does not differ between the seasons. The HWMId is zero for most heat waves outside of summer, therefore this feature does not help to improve the clustering either. The number of clusters for UPGMA clustering was set to ten for every family. In previous studies Europe was divided into six regions, but since as my boundaries are broader and I include northern Africa and Greenland as well as the Atlantic, I decided to add four additional regions [Zschenderlein et al., 2019].

Spatial heat wave clusters over Scandinavia and the Iberian region (North/West Africa and Spain) were found. in all seasons Those clusters also always belong to the largest clusters, meaning they contain many heat waves. These findings fit to the results of Stefanon et al., as they identified the Iberian and Scandinavian heat wave clusters to be the most stable clusters in summer with a different clustering approach [Stefanon et al., 2012].

In my approach of spatial clustering, I found it difficult to define clear and stable margins between western, central and eastern Europe. A clear eastern European heat wave cluster can only be found in family 0, all other families, however contain a central/eastern European cluster. Nicely spatially separated heat wave clusters are the Italy clusters in winter spring and summer, the Sardegna cluster in spring and the Iceland cluster in Summer. The icelandic summer heat wave cluster is surprisingly large with 50 heat waves belonging to it. However, there is little to no prior research found on heat waves over Iceland and their impacts.

5.3 Influence of Heat Waves on Vegetation

Correlation analysis of NDVI data and heat wave attributes was used to evaluate the impact of heat waves on vegetation.To find differences between different spatio-temporal heat wave clusters, I assessed the different clusters of the families individually. The mean correlation coefficients of the clusters showed weak correlations, as the correlation coefficient values were close to zero. However seen in relation to the other correlation coefficient values, the highest mean correlation coefficient values were found for the North Africa clusters from family 1 and 2. The positive correlation between HWMId and NDVI values in summer suggests a beneficial effect of heat waves with high magnitude on vegetation greenness. This stands in contrast to other research, which suggests, that in regions with water scarcity and in general in summer, vegetation suffers from extreme heat and therefore, a negative correlation with NDVI values can be observed [Baumbach et al., 2017]. It is nevertheless noticeable, that most prior research did not specifically assess the influence of heat waves on vegetation in North Africa. In family 2 the North Africa cluster shows a weak negative correlation with the number of heat wave days. In fall, especially after hot summers with precipitation deficits, plants have used up their reserves they have had for surviving the summer and might be therefore more susceptible to heat waves.

The individual correlation analysis reveals that positive correlation should not autmatically be interpreted as something good. The highest correlation coefficient value for a single year and cluster was found for the 2010 winter heat waves over Greenland. Here the positive correlation however can be explained by extensive ice sheet melting in winter causing positive NDVI anomalies. The melting record over Greenland in 2010 has been analyzed in literature before [Tedesco et al., 2011]. Heat waves over colder climates like Scandinavia seem to have a positive effect on the greenness of vegetation. All individual correlation coefficients of Scandinavian clusters were positive. This suggests that the type of vegetation in Scandinavia, which is mostly coniferous and mixed forests, is less sucseptible to extreme heat than other landcover types [© European Union, 2022]. Also, sunlight is probably a limiting factor in this region rather than water, therefore more warm and clear sky weather conditions, which are often associated for summer heat waves, might be beneficial for plant growth. Similar results were reported by Baumbach et al. [Baumbach et al., 2017]. All highest individual correlation coefficient values in family 3, which is the spring heat wave family, are negative, this does not match prior research where high temperature extremes in spring were found to be beneficial for most vegetation types in Europe and resulted in high NDVI values [Baumbach et al., 2017]. However, the NDVI anomalies that are correlated with the heat wave measures are only a momentary picture of the vegetation greenness at the end of the season.

For summer heat waves a positive trend of the mean correlation coefficient of NDVI values and number of heat wave days can be observed 4.8b. Between 1981 and 2015 the greening effect of heat wave days on vegetation gets stronger. When taking into account that climate change is causing heat waves to be more extreme and frequent in the 21st century it seems like vegetation is currently still benefiting from the more frequent heat events. However, as mentioned above, one can not automatically conclude that greening of vegetation is a positive effect. This depends on the region under consideration. All other families do not show a trend of the correlation coefficients over the years. The effect of heat waves on vegetation in all other seasons than summer does not seem to change. However, there are some very extremely positively correlated data points in the 0 family from 2005 to 2015, suggesting that the correlation of individual events becomes more extreme in winter, but that most correlation values remain the same (see figure 4.8a).

5.4   Outlook

The issue of extreme heat events will become even more important in the future, because even in a world with a warming of only 1.5 degrees celsius, the number and magnitude of heat waves will increase. Therefore, it is of great interest to be able to predict heat waves in advance and to better understand their impacts at different scales. The current definition of the HWMId does not quite fit the definition of heat waves that I have proposed in this paper. Therefore, it would be beneficial to adapt the HWMId definition to apply it to my heat wave definition. For clustering heat waves, it would be interesting to develop an approach to cluster heat waves spatially and temporally in one step and see if similar results can be obtained. In order to better analyse the impact of heat waves on vegetation, it could be beneficial to include other variables such as precipitation anomalies, land cover data and drought indices in the analysis. This would allow more concrete statements to be made about the influence of heat waves on vegetation. It would also be interesting to investigate the influence of heat waves in other seasons on NDVI values. This could lead to insights into whether there are critical development phases for plants in which extreme heat causes long-term effects.
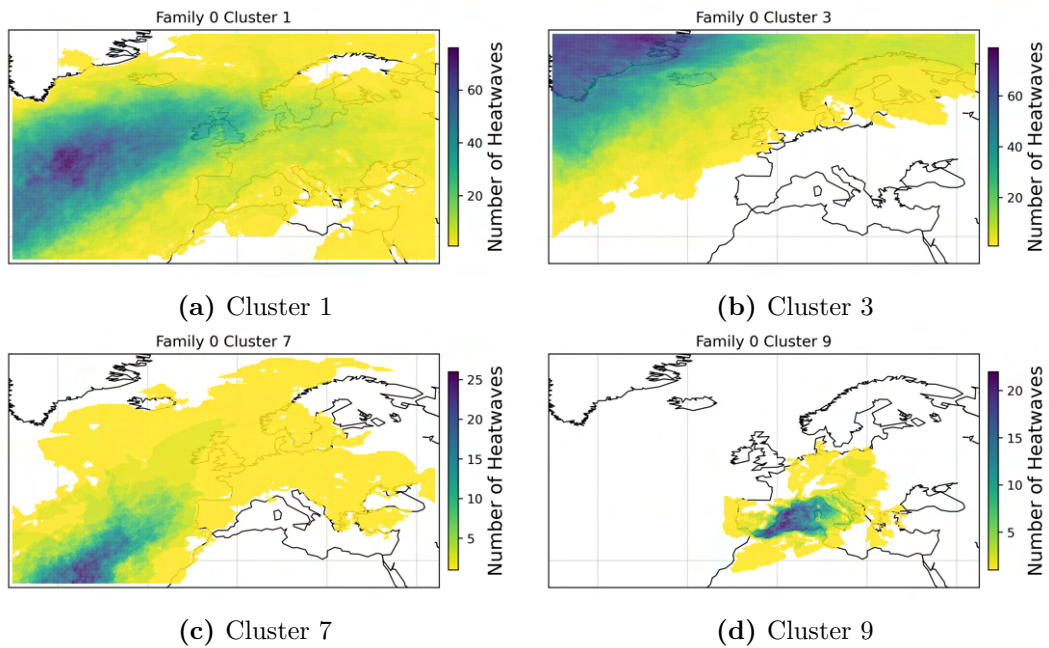
Appendix A

Further Tables and Figures
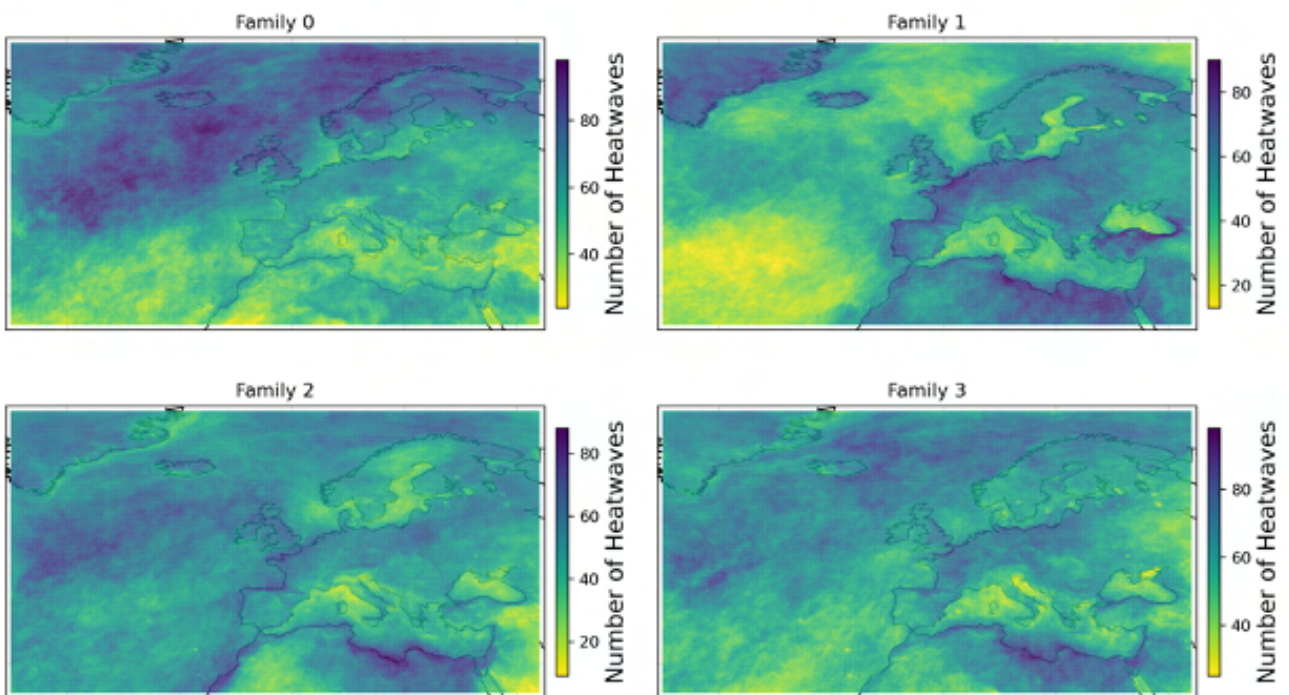
Figure A.1: Spatial distribution of the four heat wave families after K-means clustering. The spatial distribution of the four heat wave families is shown. The number of heat waves gives the information on how often a grid cell is affected by different heat waves.

(a) Cluster 1

(b) Cluster 3

(c) Cluster 7

(d) Cluster 9

Figure A.2: Ocean and Greenland clusters of heat wave family 0 after UPGMA clustering. The spatial distribution of all ocean and land clusters that are not included in the NDVI analysis of heat wave family 0 is shown. The number of heat waves gives the information on how often a grid cell is affected by different heat waves.

(a) Cluster 1

(b) Cluster 6
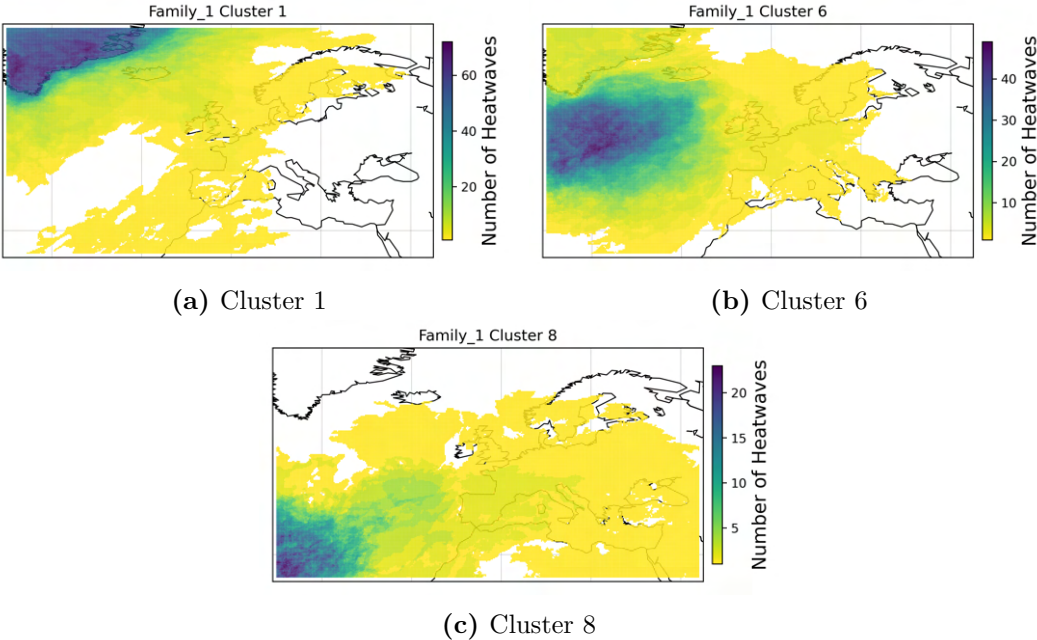
(c) Cluster 8

Figure A.3: Ocean and Greenland clusters of heat wave family 1 after UPGMA clustering. The spatial distribution of all ocean and land clusters that are not included in the NDVI analysis of heat wave family 1 is shown. The number of heat waves gives the information on how often a grid cell is affected by different heat waves.

(a) Cluster 2

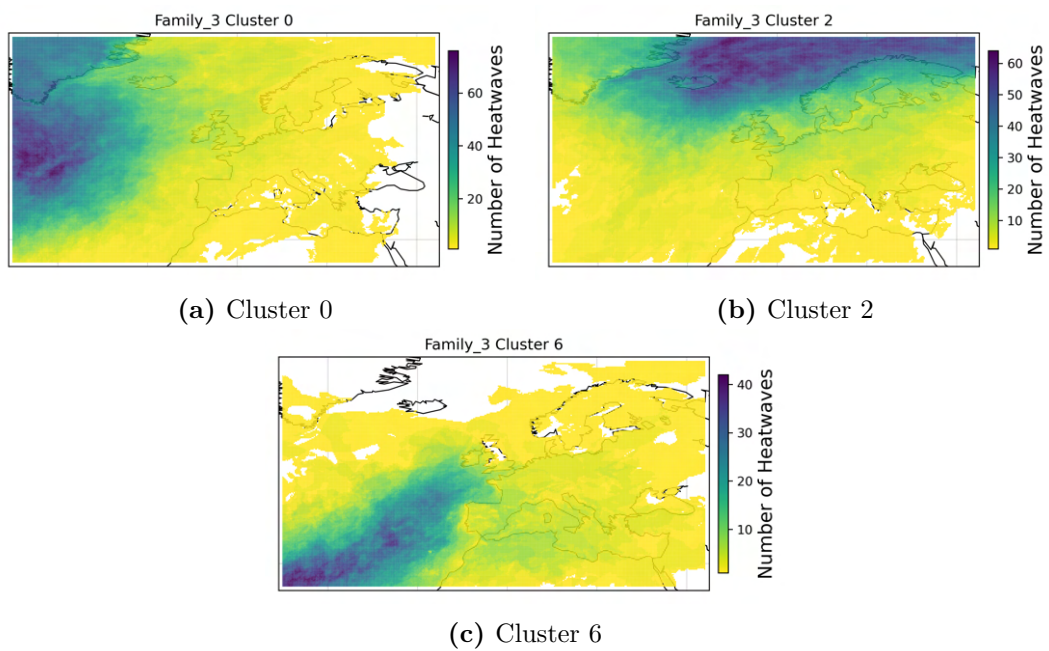**(b)** Cluster 3

(c) Cluster 5
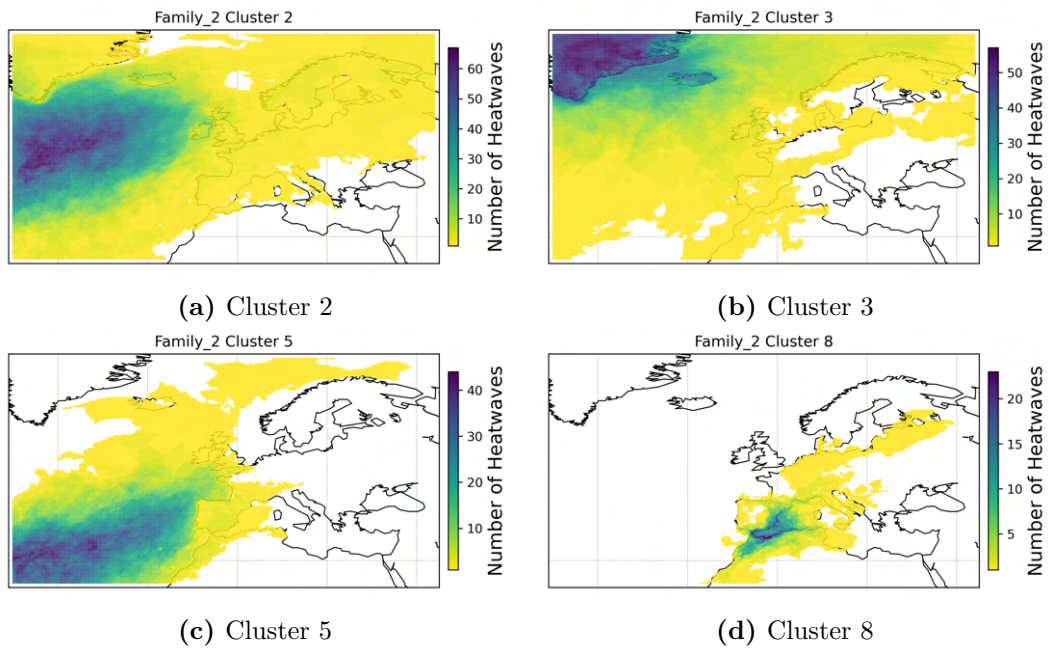
(d) Cluster 8

Figure A.4: Ocean and Greenland clusters of heat wave family 2 after UPGMA clustering. The spatial distribution of all ocean and land clusters that are not included in the NDVI analysis of heat wave family 2 is shown. The number of heat waves gives the information on how often a grid cell is affected by different heat waves.

(a) Cluster 0

(b) Cluster 2

(c) Cluster 6

Figure A.5: Ocean and Greenland clusters of heat wave family 3 after UPGMA clustering. The spatial distribution of all ocean and land clusters that are not included in the NDVI analysis of heat wave family 3 is shown. The number of heat waves gives the information on how often a grid cell is affected by different heat waves.

Bibliography

[Arora et al., 2016] Arora, P., Deepali, and Varshney, S. (2016). Analysis of K-Means and K-Medoids Algorithm for Big Data. *Physics Procedia*, 78(December 2015):507–512.

[Barriopedro et al., 2011] Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M., and Garcia-Herrera, R. (2011). The Hot Summer of 2010 : Map of Europe. *Science*, 332(April):220–224.

[Baumbach et al., 2017] Baumbach, L., Siegmund, J. F., Mittermeier, M., and Donner, R. V. (2017). Impacts of temperature extremes on European vegetation during the growing season. *Biogeosciences*, 14(21):4891–4903.

[Bonne et al., 2015] Bonne, J. L., Steen-Larsen, H. C., Risi, C., Werner, M., Sodemann, H., Lacour, J. L., Fettweis, X., Cesana, G., Delmotte, M., Cattani, O., Vallelonga, P., Kjær, H. A., Clerbaux, C., Sveinbjörnsdóttir, Á. E., and Masson-Delmotte, V. (2015). The summer 2012 Greenland heat wave: In situ and remote sensing observations of water vapor isotopic composition during an atmospheric river event. *Journal of Geophysical Research*, 120(7):2970–2989.

[Brunner et al., 2017] Brunner, L., Hegerl, G. C., and Steiner, A. K. (2017). Connecting atmospheric blocking to European temperature extremes in spring. *Journal of Climate*, 30(2):585–594.

[Cano-Crespo et al., 2021] Cano-Crespo, A., Traxl, D., and Thonicke, K. (2021). Spatio-temporal patterns of extreme fires in Amazonian forests. *European Physical Journal: Special Topics*, 230(14-15):3033–3044.

[Cassou et al., 2005] Cassou, C., Terray, L., and Phillips, A. S. (2005). Tropical Atlantic influence on European heat waves. *Journal of Climate*, 18(15):2805–2811.

[Chapman et al., 2019] Chapman, S. C., Watkins, N. W., and Stainforth, D. A. (2019). Warming Trends in Summer Heatwaves. *Geophysical Research Letters*, 46(3):1634–1640.

[Dosio et al., 2018] Dosio, A., Mentaschi, L., Fischer, E. M., and Wyser, K. (2018). Extreme heat waves under 1.5 °c and 2 °c global warming. *Environmental Research Letters*, 13(5).

[Durre et al., 2000] Durre, I., Wallace, J. M., and Lettenmaier, D. P. (2000). Dependence of extreme daily maximum temperatures on antecedent soil moisture in the contiguous United States during summer. *Journal of Climate*, 13(14):2641–2651.

[Fischer et al., 2007] Fischer, E. M., Seneviratne, S. I., Vidale, P. L., Lüthi, D., and Schär, C. (2007). Soil moisture-atmosphere interactions during the 2003 European summer heat wave. *Journal of Climate*, 20(20):5081–5099.

[Hasanuzzaman et al., 2013] Hasanuzzaman, M., Nahar, K., Alam, M. M., Roychowdhury, R., and Fujita, M. (2013). Physiological, biochemical, and molecular mechanisms of heat stress tolerance in plants. *International Journal of Molecular Sciences*, 14(5):9643–9684.

[Hersbach et al., 2018] Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Mu noz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N. (2018). Era5 hourly data on single levels from 1979 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*.

[IPCC, 2021] IPCC (2021). 2021: Climate change 2021: The physical science basis. *Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte*, page 1017/9781009157896.

[Kornhuber et al., 2017] Kornhuber, K., Petoukhov, V., Petri, S., Rahmstorf, S., and Coumou, D. (2017). Evidence for wave resonance as a key mechanism for generating high-amplitude quasi-stationary waves in boreal summer. *Climate Dynamics*, 49(5-6):1961–1979.

[Lavaysse et al., 2019] Lavaysse, C., Naumann, G., Alfieri, L., Salamon, P., and Vogt, J. (2019). Predictability of the European heat and cold waves. *Climate Dynamics*, 52(3-4):2481–2495.

[Liu et al., 2013] Liu, G., Liu, H., and Yin, Y. (2013). Global patterns of NDVI-indicated vegetation extremes and their sensitivity to climate extremes. *Environmental Research Letters*, 8(2).

[Lloyd, 1982] Lloyd, S. (1982). Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137.

[Lo et al., 2021] Lo, S. H., Chen, C. T., Russo, S., Huang, W. R., and Shih, M. F. (2021). Tracking heatwave extremes from an event perspective. *Weather and Climate Extremes*, 34:100371.

[Mann, 1945] Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica*, 13(3):245–259.

[Mecking et al., 2019] Mecking, J. V., Drijfhout, S. S., Hirschi, J. J., and Blaker, A. T. (2019). Ocean and atmosphere influence on the 2015 European heatwave. *Environmental Research Letters*, 14(11).

[Miralles et al., 2019] Miralles, D. G., Gentine, P., Seneviratne, S. I., and Teuling, A. J. (2019). Land–atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges. *Annals of the New York Academy of Sciences*, 1436(1):19–35.

[Omran et al., 2007] Omran, M. G., Engelbrecht, A. P., and Salman, A. (2007). An overview of clustering methods. *Intelligent Data Analysis*, 11(6):583–605.

[Perkins and Alexander, 2013] Perkins, S. E. and Alexander, L. V. (2013). On the measurement of heat waves. *Journal of Climate*, 26(13):4500–4517.

[Perkins-Kirkpatrick and Lewis, 2020] Perkins-Kirkpatrick, S. E. and Lewis, S. C. (2020). Increasing trends in regional heatwaves. *Nature Communications*, 11(1):1–8.

[Pettorelli et al., 2005] Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J. M., Tucker, C. J., and Stenseth, N. C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology and Evolution*, 20(9):503–510.

[Pezza et al., 2012] Pezza, A. B., van Rensch, P., and Cai, W. (2012). Severe heat waves in Southern Australia: Synoptic climatology and large scale connections. *Climate Dynamics*, 38(1-2):209–224.

[Pfahl and Wernli, 2012] Pfahl, S. and Wernli, H. (2012). Quantifying the relevance of atmospheric blocking for co-located temperature extremes in the Northern Hemisphere on (sub-)daily time scales. *Geophysical Research Letters*, 39(12):1–6.

[Pinzon and Tucker, 2014] Pinzon, J. E. and Tucker, C. J. A. (2014). Nonstationary 1981-2012 avhrr ndvi3g time series. *Remote Sens*, 6:6929–6960.

[Pinzon and Tucker, 2016] Pinzon, J. E. and Tucker, C. J. A. (2016). Nonstationary 1981-2015 avhrr ndvi3g.v1 time series: an update. *Remote Sens*.

[Robinson et al., 2021] Robinson, A., Lehmann, J., Barriopedro, D., Rahmstorf, S., and Coumou, D. (2021). Increasing heat and rainfall extremes now far outside the historical climate. *npj Climate and Atmospheric Science*, 4(1):3–6.

[Russo et al., 2014] Russo, S., Dosio, A., Graversen, R. G., Sillmann, J., Carrao, H., Dunbar, M. B., Singleton, A., Montagna, P., Barbola, P., and Vogt, J. V. (2014). Magnitude of extreme heat waves in present climate and their projection in a warming world. *Journal of Geophysical Research Atmospheres*, 119(22):12,500–12,512.

[Russo et al., 2015] Russo, S., Sillmann, J., and Fischer, E. M. (2015). Top ten European heatwaves since 1950 and their occurrence in the coming decades. *Environmental Research Letters*, 10(12).

[Schaller et al., 2018] Schaller, N., Sillmann, J., Anstey, J., Fischer, E. M., Grams, C. M., and Russo, S. (2018). Influence of blocking on Northern European and Western Russian heatwaves in large climate model ensembles. *Environmental Research Letters*, 13(5).

[Spearman and Spearman, 1904] Spearman and Spearman, C. (1904). The Proof and Measurement of Association between Two Things Author (s): C . Spearman Source : The American Journal of Psychology , Vol . 15 , No . 1 (Jan ., 1904), pp . 72-101 Published by : University of Illinois Press Stable URL : http://www.jstor.o. *The American journal of psychology*, 15(1):72–101.

[Stefanon et al., 2012] Stefanon, M., Dandrea, F., and Drobinski, P. (2012). Heatwave classification over Europe and the Mediterranean region. *Environmental Research Letters*, 7(1).

[Sukal and Miechener, 1958] Sukal, R. and Miechener, C. (1958). A statistical method for evaluating systematic relationships. *The University of Kansas Science Bulletin*, 1409(28).

[Sutanto et al., 2020] Sutanto, S. J., Vitolo, C., Di Napoli, C., D'Andrea, M., and Van Lanen, H. A. (2020). Heatwaves, droughts, and fires: Exploring compound and cascading dry hazards at the pan-European scale. *Environment International*, 134(October 2019):105276.

[Tedesco et al., 2011] Tedesco, M., Fettweis, X., Van Den Broeke, M. R., Van De Wal, R. S., Smeets, C. J., Van De Berg, W. J., Serreze, M. C., and Box, J. E. (2011). The role of albedo and accumulation in the 2010 melting record in Greenland. *Environmental Research Letters*, 6(1).

[Traxl et al., 2016a] Traxl, D., Boers, N., and Kurths, J. (2016a). Deep graphs-A general framework to represent and analyze heterogeneous complex systems across scales. *Chaos*, 26(6).

[Traxl et al., 2016b] Traxl, D., Boers, N., Rheinwalt, A., Goswami, B., and Kurths, J. (2016b). The size distribution of spatiotemporal extreme rainfall clusters around the globe. *Geophysical Research Letters*, 43(18):9939–9947.

[Xiao et al., 2010] Xiao, L., Ye, J., Esteves, R., and Rong, C. (2010). Using Spearman's correlation coefficients for exploratory data analysis on big dataset . *Concurrency and Computation: Practice and Experience*, 28(14):3866–3878.

[Yiou et al., 2020] Yiou, P., Cattiaux, J., Faranda, D., Kadygrov, N., Jézéquel, A., Naveau, P., Ribes, A., Robin, Y., Thao, S., van Oldenborgh, G. J., and Vrac, M. (2020). Analyses of the northern European summer heatwave of 2018. *Bulletin of the American Meteorological Society*, 101(1):S35–S40.

[Zschenderlein et al., 2019] Zschenderlein, P., Fink, A. H., Pfahl, S., and Wernli, H. (2019). Processes determining heat waves across different European climates. *Quarterly Journal of the Royal Meteorological Society*, 145(724):2973–2989.

[© European Union, 2022] © European Union, C. L. M. S. (2022). European environment agency (eea).

Selbständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Masterarbeit selbständig und nur mit den angegebenen Hilfsmitteln angefertigt habe und dass alle Stellen, die dem Wortlaut oder dem Sinne nach anderen Werken entnommen sind, durch Angaben von Quellen als Entlehnung kenntlich gemacht worden sind. Diese Masterarbeit wurde in gleicher oder ähnlicher Form in keinem anderen Studiengang als Prüfungsleistung vorgelegt.

Tübingen, 13.06.2022

Ort, Datum   Unterschrift