# INVESTIGATING DRIVERS AND DYNAMICS OF HOT AND DRY EXTREMES IN EUROPE BY APPLYING MACHINE LEARNING APPROACHES

**Dissertation zur Erlangung des Doktorgrades**
**an der Fakultät für Geowissenschaften der**
**Ludwig-Maximilians-Universität München**

vorgelegt von

Elizaveta Felsche

eingereicht am

12.02.2024, München

Supervisor: Prof. Ralf Ludwig
Department of Geography, Ludwig-Maximilians-Universität München
Munich, Bavaria, Germany
Co-Supervisor: Dr. Martin Leduc
Ouranos, Montréal, Canada
Centre pour l'étude et la simulation du climat à l'échelle régionale (ESCER),
Université du Québec à Montréal, Montréal, Canada

Day of the oral exam: 28.05.24

*"Never doubt that a small group of thoughtful, committed citizens can change the world; indeed, it is the only thing that ever has."*

Margaret Mead

# Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Prof. Ralf Ludwig, for his support throughout this time. From the master's thesis project onward, you provided me with the opportunity to follow my scientific interests, guided and supported me in addressing academic questions and decisions, and trusted in the very special PhD setup in which I was involved.

I would like to thank Dr. Martin Leduc who sparked my interest in climate sciences many years ago by offering a great opportunity to visit Montreal to work in Climate Science and supported me at the end of my dissertation. I am incredibly honoured to have you as my second supervisor.

My sincere appreciation goes to my colleagues from the LMU working group for the memorable experiences and fruitful scientific discussions we shared, from Weihnachtsgrillen to trips to Vienna for the EGU. My special thanks goes out to Andrea Böhnisch for the immensely productive and enjoyable collaboration that enriched my PhD years and made me feel less like a lonely warrior.

A huge thank you goes out to the wonderful team at CDTM. It's been a thrilling ride, and I've enjoyed growing with you, laughing with you, and extinguishing thousands of fires that burn out there at the Center every day, continually strengthened by the knowledge that we will face all of the challenges as a team. I extend my thanks to Gesa Biermann, Tom Schelo, Aaron Defort, Philipp Hofsommer, Philipp Hulm, Theresa Doppstadt, Michael Fröhlich, Amelie Pahl, Franz Waltenberger, Jose Vega, Carla Pregel Hoderlein, Charlotte Kobiella, Felix Dörpmund, Vera Eger, Ferran Pla Cardona, Julia Balowski, Martin Wessel, Cyrine Chaabani, Samuel Valenzuela, and Florian Wiethof.

I am incredibly grateful to my wonderful parents, Olga and Vladimir Pechenov, who supported me throughout my study years and in my professional career, giving me everything I needed to grow. To my grandparents for being fantastic role models and supporters from the heart. Last but not least, I have to say a huge thank you to my husband and life partner, Frido, for believing in me, offering unwavering support, and holding my hand when I needed it the most. Your encouragement has been a source of strength throughout this dissertation process.

# Summary

Heatwaves and droughts are increasingly impacting our ecosystem, infrastructure and society through their rising intensity and frequency due to the present effect of global warming. A series of recent extreme heatwaves and droughts in Europe, like in 2003, 2006, 2018 and 2022, have revealed the vulnerability by causing various impacts from increased heat stress on humans and the ecosystem to missing cooling water for thermal power plants and increased incidence of climate-sensitive diseases. The frequency and intensity of heatwaves and droughts are expected to increase in most regions of Europe in the upcoming years.

In the past decade, artificial intelligence has shown its impressive capacity to solve complex, non-linear problems and increase pattern recognition capabilities. In 2016, artificial intelligence beat the world champion in Go, a challenge previously regarded as unsolvable by computers. Recent ChatGPT-3 and GPT-4 launches have highlighted AI's potential in text generation. Much of it lies in the application of machine learning in climate sciences due to the complexity of the earth system and the huge need for predictability and mitigation of various climate extremes. Increased volume of observational and model data facilitates the development of machine learning applications in climate sciences. Explainable AI approaches allow an interpretation of the results and an examination of the physical consistency of the results. Pattern recognition and clustering approaches allow to derive consistent groupings of data based on inherent features.

This thesis aims to investigate the predictability and interrelation of heatwaves and droughts in Europe using machine learning approaches in a large ensemble. Physical interpretation of the results is achieved via explainable AI methods. This dissertation employs data from a Single Model Initial Condition Large Ensemble (SMILE). A SMILE consists of multiple model runs with the same external forcing but slightly varied initial conditions. The differences in the initial conditions allow to estimate the uncertainty that arises due to nonlinear interactions in the climate system. A large ensemble covers the natural variability of the climate and provides robust statistical estimations of extreme events due to the large quantity of available events. This thesis uses a Regional Climate Model (RCM), which, in contrast to General Circulation Models (GCMs), has a finer spatial reso-

lution and allows the resolution of fine-scale processes, such that the effects of local geographies like mountains become visible.

This thesis includes three publications which use different machine learning methods and which have applications in drought prediction, identification of heatwave patterns, and compound hot and dry events. This interdisciplinary research draws relevance from meteorology, climatology, statistics, and data science and contributes to scientific progress in geosciences.

In the first publication (Felsche and Ludwig, 2021), an Artificial Neural Network (ANN) is trained to predict drought occurrence in two European regions, Munich and Lisbon, using data from a large ensemble of climate simulations. Atmospheric and soil variables from the Canadian Regional Climate Model 5 Large Ensemble (CRCM5-LE) and teleconnection indices such as the North Atlantic Oscillation (NAO) are used as input data for the prediction. Standardised Precipitation Index with a cumulative time of one month (SPI1) is used as the predictor variable. The best-performing algorithms can correctly classify drought or no drought for about $55-57\%$ of the events in both domains. The models are then analysed for variables important for prediction using explainable AI methods like shapely values. Shapely values assign each variable a relative importance for the prediction, allowing the comparison of the contribution of each variable. The North Atlantic Oscillation (NAO) index and air pressure one month before the event were found to be the most important variables for prediction for both domains. Additionally, for the Lisbon domain, northward near-surface wind is found to have an important contribution. For the Munich domain, in contrast, teleconnection indices like the Scandinavian Oscillation Index (SCA) and East Atlantic/Western Russia Oscillation (EAWR) five months before the event are found among the strongest predictors. The study concludes that the Lisbon domain prediction is more reliable regarding the accuracy and contribution of individual variables.

In the second publication (Felsche et al., 2023), machine-learning-based clustering is applied to summer heatwaves in Europe to identify typical heatwave patterns in Europe in the current climate. Therefore, the data from the CRCM5-LE model is used. The patterns are identified via an unsupervised machine learning approach of hierarchical agglomerative clustering with cosine similarity as a distance measure. A total of nine significant spatial patterns are found, ranging from the Iberian Peninsula to North-Eastern Europe. The patterns are consistent with previous studies, observations and historical extreme heatwave events. Additionally, the publication investigated the influence of a soil moisture deficit in spring on a prolonged heatwave in summer in the identified region and

the effect of a prolonged heatwave in summer on the soil moisture deficit in the following fall via quantile regression slopes. The results of Felsche et al. (2023) indicate a predictive power of soil moisture in the preceding winter/spring for a heatwave occurrence in summer in South European regions like Greece and South Italy, Southeast Europe, Iberian Peninsula and Western Europe. Moreover, the study finds a correlation between a high number of heatwave days in summer and a soil moisture deficit in the following fall for Northern European regions like Scandinavia, Northeast Europe and Southeast Europe.

The third publication (Felsche et al., 2024) investigates the probability of occurrence of historical extreme compound hot and dry events in the future. The study identifies the most extreme past event occurrences from reanalysis data ERA5 in the past two decades, 2001 to 2022, for Europe's nine typical regions of compound hot and dry events. It quantifies the event probabilities with CRCM5-LE for three Global Warming Levels (GWL; +1.2K, +2K, and +3K). GWL is defined via the difference in mean surface air temperature with regards to the historical period 1850-1900. The study uses copula statistics for bivariate probability quantification and uses a Survival Kendall probability definition. The nine sub-regions are identified using the same approach as in publication II. The results show that 2003 was the most extreme historical hot and dry event in four European sub-regions located in Southwest Europe. The findings show a significant increase in the frequency of extreme past events under higher warming levels. For some events, the probability increases up to five to six times when comparing GWL2 and GWL3 and up to 46 percent (every second summer) under the latter scenario. Additionally, the study performs a hot and dry compound event climatology analysis by evaluating changes between the present and GWL3 periods. It shows a northward shift in the climatology of hot and dry events. Eastern Europe's current climate is extending into parts of the Baltic Sea Coast and Scandinavia, the area experiencing a wet and moist climate in the Alps is shrinking, and the hot and dry climate currently observed in the Balkan Peninsula is expected to extend into substantial parts of Eastern Europe under a future GWL of +3K.

This thesis studies the predictability and interrelationship of heatwaves and droughts and their evolution with climate change by applying different machine learning techniques. With the help of artificial intelligence methods such as ANN and hierarchical agglomerative clustering, the study advances the scientific knowledge on the predictability of droughts. The thesis shows that air pressure and teleconnection indices are essential predictors for droughts in Munich and Lisbon, as well as that for Northern Europe number of heatwave days in summer, can serve as a predictor for

an agricultural drought in fall. For summer heatwaves in South Europe, the soil moisture deficit in the previous spring can serve as a predictor. The thesis shows that compound hot and dry events will become more probable with rising Global Warming Levels in most European regions. At the same time, there is a difference of up to five to six times in event occurrence when comparing GWL +2K to GWL +3K, underlining the benefits of sticking to a two-degree target. This scientific knowledge is valuable for further studies on the predictability of hot and dry events. The evolution in the future is valuable for decision-makers for implementation of mitigation measures to address the impacts of these events on human health, agriculture, and the environment.

# Zusammenfassung

Hitzewellen und Dürren beeinträchtigen zunehmend unser Ökosystem, unsere Infrastruktur und unsere Gesellschaft, da sie aufgrund der derzeitigen Auswirkungen der globalen Erwärmung an Intensität und Häufigkeit zunehmen. Eine Reihe von extremen Hitzewellen und Dürren in Europa in den Jahren 2003, 2006, 2018 und 2022 haben die Vulnerabilität deutlich gemacht, indem sie verschiedene Auswirkungen, von erhöhtem Hitzestress auf Menschen und das Ökosystem bis hin zu fehlendem Kühlwasser für Wärmekraftwerke und dem vermehrten Auftreten von klimasensitiven Krankheiten, verursacht haben. Es ist zu erwarten, dass die Häufigkeit und Intensität von Hitzewellen und Dürren in den meisten Regionen Europas in den kommenden Jahren zunehmen wird.

In den letzten zehn Jahren hat die künstliche Intelligenz (KI) beeindruckende Fähigkeiten gezeigt komplexe, nicht lineare Probleme zu lösen und Mustererkennung zu verbessern. Im Jahr 2016 besiegte die künstliche Intelligenz den Weltmeister in Go, eine Herausforderung, die zuvor als unlösbar für Computer galt. Die kürzlichen Veröffentlichungen von ChatGPT-3 und GPT-4 haben das Potenzial der KI bei der Texterstellung aufgezeigt. Die Anwendung von KI in den Klimawissenschaften birgt ein großes Zukunftspotenzial aufgrund der Komplexität des Erdsystems und des enormen Bedarfs an Vorhersagbarkeit und Mitigation der verschiedenen Klimaextreme. Die wachsende Menge an Beobachtungs- und Modelldaten erleichtert die Entwicklung von Anwendungen für KI.

Ziel dieser Arbeit ist es, die Vorhersagbarkeit und den Zusammenhang von Hitzewellen und Dürren in Europa mithilfe von Ansätzen des maschinellen Lernens in einem Large Ensemble zu untersuchen. Physikalische Interpretation der Ergebnisse erfolgt durch die Anwendung statistischer Methoden, sowie Methoden as dem Bereich der Explainable AI. In dieser Dissertation werden Daten aus einem Single Model Initial Condition Large Ensemble (SMILE) verwendet. Ein SMILE besteht aus mehreren Modellläufen mit dem gleichen externen Antrieb, aber leicht veränderten Initialbedingungen. Die Unterschiede in den Anfangsbedingungen ermöglichen es, die Unsicherheiten abzuschätzen, die durch nichtlineare Wechselwirkungen im Klimasystem entstehen. Ein

Large Ensemble deckt die natürliche Variabilität des Klimas ab und liefert aufgrund der großen Menge an verfügbaren Ereignissen statistisch robuste Abschätzungen von Extremereignissen. In dieser Arbeit wird ein regionales Klimamodell (RCM) verwendet, das im Gegensatz zu allgemeinen Zirkulationsmodellen (GCM) eine feinere räumliche Auflösung hat und die Darstellung feiner Prozesse ermöglicht, sodass die Auswirkungen lokaler Geografien wie Gebirge sichtbar werden.

Diese Dissertation umfasst drei Publikationen, in denen verschiedene Methoden des maschinellen Lernens angewandt werden und die sich auf die Vorhersage von Dürren, die Identifizierung von Hitzewellenmustern und die Wahrscheinlichkeitsquantifizierung von gleichzeitigen Hitze- und Dürreereignissen beziehen. Diese interdisziplinäre Forschung ist im Bereich der Meteorologie, Klimatologie, Statistik und Data Science zu verorten und trägt zum wissenschaftlichen Fortschritt in den Geowissenschaften bei.

In der ersten Publikation (Felsche and Ludwig, 2021) wird ein künstliches neuronales Netzwerk (ANN) trainiert, um das Auftreten von Dürren in zwei europäischen Regionen, München und Lissabon, anhand von Daten aus einem Large Ensemble von Klimasimulationen vorherzusagen. Atmosphären- und Bodenvariablen aus dem Canadian Regional Climate Model 5 Large Ensemble (CRCM5-LE) und Telekonnektionsindizes wie die Nordatlantische Oszillation (NAO) werden als Input für die Vorhersage verwendet. Als Vorhersagevariable wird der standardisierte Niederschlagsindex mit einer kumulativen Zeit von einem Monat (SPI1) verwendet. Die leistungsstärksten Algorithmen können für etwa $55-57\%$ der Ereignisse in beiden Domänen korrekt klassifizieren, ob eine Dürre zu erwarten ist oder nicht. Die Modelle werden dann auf Variablen untersucht, die für die Vorhersage wichtig sind, indem erklärbare KI-Methoden wie Shapely-Werte verwendet werden. Shapely-Werte weisen jeder Variable eine relative Bedeutung für die Vorhersage zu und ermöglichen so den Vergleich des Beitrags der jeweiligen Variablen. NAO und der Luftdruck einen Monat vor dem Ereignis erwiesen sich für beide Domänen als die wichtigsten Variablen für die Vorhersage. Darüber hinaus ist der oberflächennahe Wind für Lissabon eine der wichtisten Variablen. Für München hingegen gehören Telekonnektionsindizes wie der Skandinavische Oszillationsindex (SCA) und die Ostatlantik/Westrussland-Oszillation (EAWR) fünf Monate vor dem Ereignis zu den stärksten Prädiktoren. Die Studie kommt zu dem Schluss, dass die Vorhersage für Lissabon zuverlässiger ist, was die Genauigkeit und den Beitrag der einzelnen Variablen angeht.

In der zweiten Publikation (Felsche et al., 2023) wird ein KI-basiertes Clustering auf sommerliche

Hitzewellen in Europa angewandt, um typische Hitzewellenmuster in Europa unter den derzeitigen klimatischen Bedingungen zu identifizieren. Dazu werden die Daten des CRCM5-LE-Modells verwendet. Die Studie wendet hierarchisches agglomeratives Clustering mit Cosinusähnlichkeit als Distanzmaß an um die Hitzewellenmuster zu identifiziert. Es wurden insgesamt neun signifikante räumliche Muster gefunden, die von der Iberischen Halbinsel bis nach Nordosteuropa reichen. Die Muster stimmen mit früheren Studien, Beobachtungen und historischen extremen Hitzewellenereignissen überein. Darüber hinaus wurde in der Publikation der Einfluss eines Bodenfeuchtedefizits im Frühjahr auf eine lang anhaltende Hitzewelle im Sommer und die Auswirkung einer lang anhaltenden Hitzewelle im Sommer auf das Bodenfeuchtedefizit im darauffolgenden Herbst in den jeweiligen Regionen mittels Quantilsregressionskurven untersucht. Die Ergebnisse von Felsche et al. (2023) zeigen eine Vorhersagekraft der Bodenfeuchte im vorangegangenen Winter/Frühjahr für das Auftreten einer Hitzewelle im Sommer in südeuropäischen Regionen wie Griechenland und Süditalien, Südosteuropa, der Iberischen Halbinsel und Westeuropa an. Darüber hinaus wird in der Studie für nordeuropäische Regionen wie Skandinavien, Nordosteuropa und Südosteuropa ein Zusammenhang zwischen einer hohen Anzahl von Hitzewellentagen im Sommer und einem Bodenfeuchtedefizit im folgenden Herbst aufgezeigt.

Die dritte Publikation (Felsche et al., 2024) untersucht die Wahrscheinlichkeit des Auftretens historischer gleichzeitiger Hitze- und Dürreereignisse (*compound events*) in der Zukunft. Die Studie identifiziert die extremsten Ereignisse der Vergangenheit über den Reanalysedatensatz ERA5 in den letzten zwei Jahrzehnten, 2001 bis 2022, für die neun typischen Regionen Europas. Sie quantifiziert die Ereignis-wahrscheinlichkeiten mit CRCM5-LE für drei Global Warming Levels (GWL; +1,2K, +2K und +3K). GWLs sind über die Differenz der mittleren Oberflächenlufttemperatur zum Zeitpunkt der Betrachtung zum historischen Zeitraum 1850-1900 definiert. Die Studie verwendet Copula-Statistiken zur Quantifizierung bivariater Wahrscheinlichkeiten und verwendet eine Survival -Kendall -Wahrscheinlichkeitsdefinition. Die neun Unterregionen werden nach dem gleichen Ansatz wie in Publikation II ermittelt. Die Ergebnisse zeigen, dass 2003 in vier von neun europäischen Teilregionen das extremste historische Hitze- und Dürreereignis war. Die Analysen zeigen eine erhebliche Zunahme der Wahrscheinlichkeit von *compound events* bei höheren Erwärmungsniveaus. Bei einigen Ereignissen steigt die Wahrscheinlichkeit zwischen GWL2 und GWL3 um das Fünf- bis Sachsfache, und um bis zu 46 Prozent (jeden zweiten Sommer) unter dem letzteren Szenario. Darüber hinaus führt die Studie eine Analyse der Klimatologie heißer und trockener *compound events* durch, indem sie die Veränderungen zwischen dem heutigen Zeitraum

und GWL3 bewertet. Das derzeitige Klima in Osteuropa dehnt sich auf Teile der Ostseeküste und Skandinavien aus, die Region mit feuchtem und nassem Klima in den Alpen schrumpft, und das heiße und trockene Klima, das derzeit auf der Balkanhalbinsel zu beobachten ist, wird sich bei einem künftigen GWL von +3 K voraussichtlich auf wesentliche Teile Osteuropas ausdehnen.

In dieser Arbeit werden die Vorhersagbarkeit und die Wechselbeziehung von Hitzewellen und Dürren sowie deren Veränderung im Zusammenhang mit dem Klimawandel durch die Anwendung verschiedener KI-Verfahren untersucht. Mit Hilfe von Methoden der künstlichen Intelligenz wie ANN und hierarchischem agglomerativen Clustering bringt die Studie die wissenschaftlichen Erkenntnisse über die Vorhersagbarkeit von Dürren voran. Die Arbeit zeigt, dass für Dürren in München und Lissabon Luftdruck- und Telekonnektionsindizes wesentliche Prädiktoren sind, und, dass für Nordeuropa die Anzahl der Hitzewellentage im Sommer als Prädiktor für landwirtschaftliche Dürren im Herbst dienen kann. Für sommerliche Hitzewellen in Südeuropa kann das Bodenfeuchtedefizit im vorangegangenen Frühjahr als Prädiktor dienen. Die Dissertation zeigt, dass heiße und trockene *compound events* mit dem Anstieg des globalen Erwärmungsniveaus in den meisten europäischen Regionen wahrscheinlicher werden. Gleichzeitig ergibt sich ein Unterschied in der Häufigkeit der Ereignisse um das Fünf- bis Sechsfache, wenn man GWL +2K mit GWL +3K vergleicht. Dies unterstreicht die Schwere der Konsequenzen, wenn die globale Erderwärmung nicht auf Zwei-Grad reduziert wird. Diese wissenschaftlichen Erkenntnisse sind wertvoll für weitere Studien über die Vorhersagbarkeit von Hitzewellen und Dürren. Ebenso, können die Ergebnisse als Entscheidungsgrundlage dienen um Maßnahmen zur Mitigation der Auswirkungen von Hitzewellen und Dürren auf menschliche Gesundheit, Landwirtschaft und Umwelt durchzusetzen.

# Contents

# List of Figures

# Acronyms

**ANN** Artificial Neural Network

**CanESM2-LE** Canadian Earth System Model version 2 Large Ensemble

**CRCM5** Canadian Regional Climate Model version 5

**CRCM5-LE** Canadian Regional Climate Model version 5 Large Ensemble

**EA** East Atlantic Oscillation

**EA/WR** East Atlantic/Western Russia Oscillation

**ECMWF** European Centre for Medium-Range Weather Forecasts

**GCM** General Circulation Model

**GWL** Global Warming Level

**IPCC** Intergovernmental Panel on Climate Change

**NAO** North Atlantic Oscillation

**RCP** Representative Concentration Pathway

**SCA** Scandinavian Oscillation

**SMILE** Single-Model Initial-Condition Large Ensemble

# 1. Introduction

Our environment, our infrastructure and ecosystem are accustomed to deal with calculated extremes: there is a certain expectation value for which we adjust. Our homes are built to withstand the average wind and weathers, the dams are built to hold the 100 year flood, the public transport system can only deal with a certain amount of snow. Nevertheless we again and again experience extremes where our society and ecosystem fail to deal with: be it the catastrophic flash flood in the Ahr valley in Germany in 2021 (Truedinger et al., 2023) leading to deaths and massive distractions, the strongest agricultural drought in 2018 in central and northern Europe leading to huge crop failures (Li et al., 2020) or the strong snowfall in Munich in 2023 (Livingston, 2023), where no trains, busses, or airplanes were working for multiple days. Having the ability to predict the occurrence of such kind of events becomes crucial, especially if we can mitigate the excess mortality, economic damages, supply shortages and infrastructure failure. Prediction of the events needs an advanced understanding of the underlying drivers and teleconnections that lead to the formation of extreme events.

Europe experienced a rising trend in the occurrences of heatwaves and droughts in the past two decades, with record-breaking events like 2003, 2018, 2022. Heatwaves and droughts pose an immense challenge for research for a multitude of reasons. First of all extreme events happen rarely by definition. Comprehensive historical weather observations are available for roughly 140 years only, since 1880 (Kaspar et al., 2017). If an event happens only every 20 years, each of the observed locations will have a record of 7 events only to be researched. This a number way too low for statistically robust analyses. Secondly, the events of heatwaves and droughts happen on a vast scale with a multitude of variables that influence their formation, ranging from atmospheric conditions to land surface processes. Due to the internal chaotic nature of the weather system we cannot expects that an interconnection found once will hold true for other event occurrences. Moreover, the events of heatwaves and droughts are highly interrelated. A heatwave can trigger a drought and vice versa also across seasons or even across different regions, thereby further complicating the investigations.

Reports by the Intergovernmental Panel on Climate Change (IPCC) that are summarizing the cur-

rent state of research underline the rising occurrence of extreme weather events due to climate change, including heatwaves and droughts, with varying regional impacts. It is scientific consensus that Europe is experiencing currently and will face in the future more frequent, prolonged, and intense heatwaves and droughts (Bednar-Friedl et al., 2022). As our climate undergoes significant changes, understanding the dynamics of heatwaves and droughts becomes not only a scientific imperative but also a critical foundation for developing effective strategies for adaptation and mitigation.

The research presented here employs state-of-the-art machine learning techniques to enhance our comprehension of spatial and temporal patterns associated with heatwaves and droughts. The thesis uses Artificial Neural Networks (ANNs) as a supervised method for prediction and clustering as an unsupervised approach for identifying dominant spatial patterns.

This dissertation contributes to advancing the understanding of heatwaves and droughts by making use of a Single Model Initial Condition Large Ensemble (SMILE). The employed SMILEs consist of a regional nested climate model, that is run several times with slightly iterated initial conditions, thereby producing data not only from one model run, but 50. This method ensures correct estimation of the natural climate variability in contrast to climate-change-induced signal. One other benefit of SMILEs is that in contrast to conventional simulations those produce a high number of extremes available for investigation.

In following the state of the research on heatwaves and droughts, data and machine learning methods are presented in **section 2**. Research questions are introduced in **section 3**, followed by the three publications. The first publication within this dissertation focuses on monthly predictability of droughts in Lisbon and Munich, utilizing ANNs (**Paper I; section 4.1**). In the second publication, we apply clustering techniques to discern typical patterns of heatwaves across Europe and investigate the seasonal connection of heatwaves occurring in those regions with a soil moisture deficit in the previous or following season (**Paper II; section 4.2**). Lastly, this dissertation explores occurrences of past compound events—simultaneous occurrences of heatwaves and droughts. By leveraging the SMILE approach, we quantify the probability of past extreme compound events and project their occurrence in the future (**Paper III; section 4.3**). The key findings of the three publications are discussed in **section 5**.

# 2. Scientific Basis

## 2.1 Heatwaves and droughts

Prolonged periods of above-average temperatures are commonly known as heatwaves (Miralles et al., 2019). Meteorological drought is commonly defined as a period of reduced precipitation (Miralles et al., 2019). However, other definitions exist that focus more on soil moisture (agricultural drought) or streamflow and reservoir levels (hydrological drought) (Sheffield and Wood, 2011).

The weather in the Northern Hemisphere is determined by the constant alternation of cyclones and anticyclones arriving from the West (Roedel, 2013; Domeisen et al., 2023). Rossby waves are the high latitude wind meanders between those cyclones and anticyclones (Bravar and Kavvas, 1991), see Figure 2.1. Heatwaves and droughts occur when a large-scale circulation anomaly, namely an anticyclone, remains over the same geographical area for a prolonged period (Bravar and Kavvas, 1991). Due to the weather system's chaotical nature, no single mechanism causes this prolonged stopping of the alternation of cyclones and anticyclones and contributes to the formation of those events. Generally speaking, two main factors influence the formation of the events: dynamic and thermodynamical drivers (Suarez-Gutierrez et al., 2020). Dynamical drivers are large-scale circulation changes and atmospheric conditions influencing the events' formation. Thermodynamic drivers account for local effects as land-atmosphere interactions that are introducing feedbacks that could weaken or strengthen the events. A study found that for heatwaves dynamical drivers and thermodynamic conditions contribute each between 20 and 70% to the events anomalies (Wehrli et al., 2019). Already now Climate Change is playing a significant role, contributing between 10% to 40% of the heatwave event anomaly.

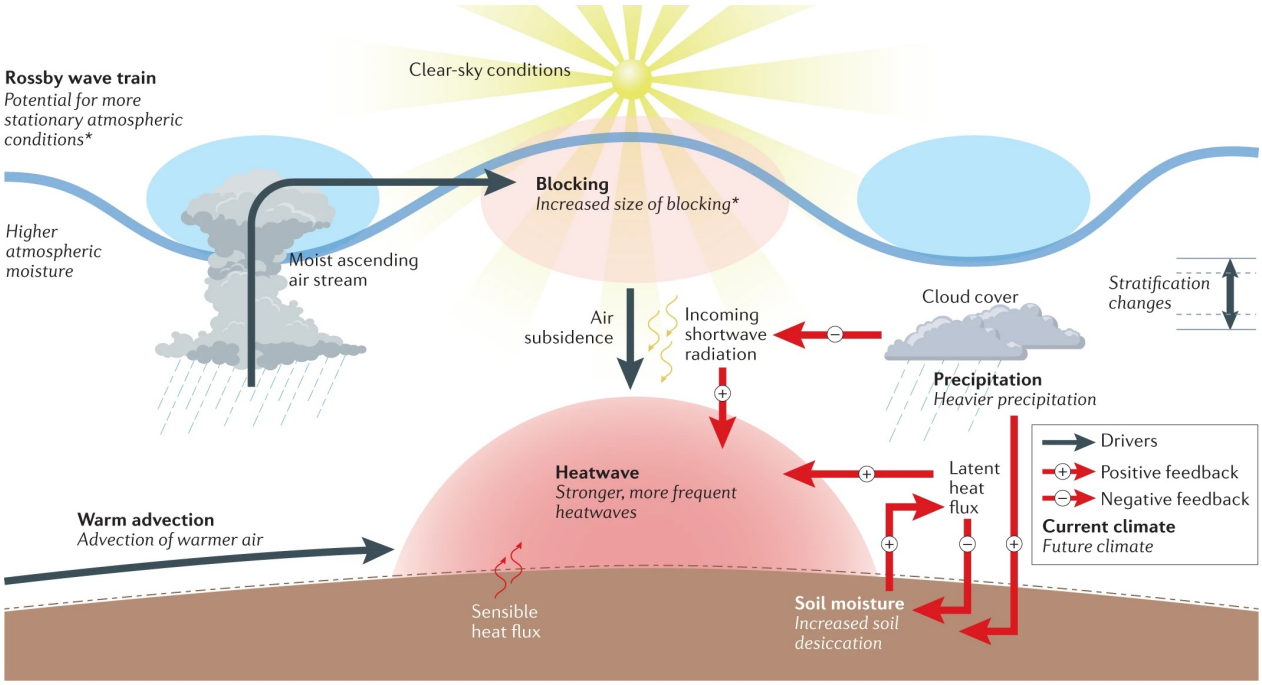


Figure 2.1: Schematical overview of drivers contributing to the formation of heatwaves and meteorological droughts in the mid-latitudes. Boldface text indicates processes in current climate and italic text projected changes under anthropogenic warming. Asterisks indicate projected changes with high uncertainty. Adapted from Domeisen et al. (2023).

### 2.1.1 Large-scale circulations influencing the formation

Large-scale atmospheric circulation patterns strongly determine European climate variability (Lim, 2015). The pressure system over the Atlantic Ocean and Europe can be divided into four main modes of variability. The four leading modes are displayed in Figure 2.2 along with the percentage of variance they explain (Felsche and Ludwig, 2021). Those modes are connected to temperature and precipitation anomalies in certain regions. The most dominant pattern, North Atlantic Oscillation (**NAO**), is characterised by the atmospheric pressure anomaly between the Azores High and the Icelandic Low. It accounts for roughly 30% of variability. A positive NAO index in summer is associated with dry and warm conditions in north-west Europe, whereas southern Europe and the Mediterranean experience cooler and wetter conditions (Folland et al., 2009). The Scandinavian Oscillation (**SCA**) (Figure 2.2) has the primary centre of action around the Scandinavian Peninsula. A positive SCA index is usually associated with increased precipitation in the midlatitude north-eastern Atlantic and decreased precipitation around north-eastern Europe (Bueh and Nakamura, 2007). The East Atlantic Oscillation (**EA**) (Figure 2.2 (C)) consists of a dipole: a low-pressure sys-

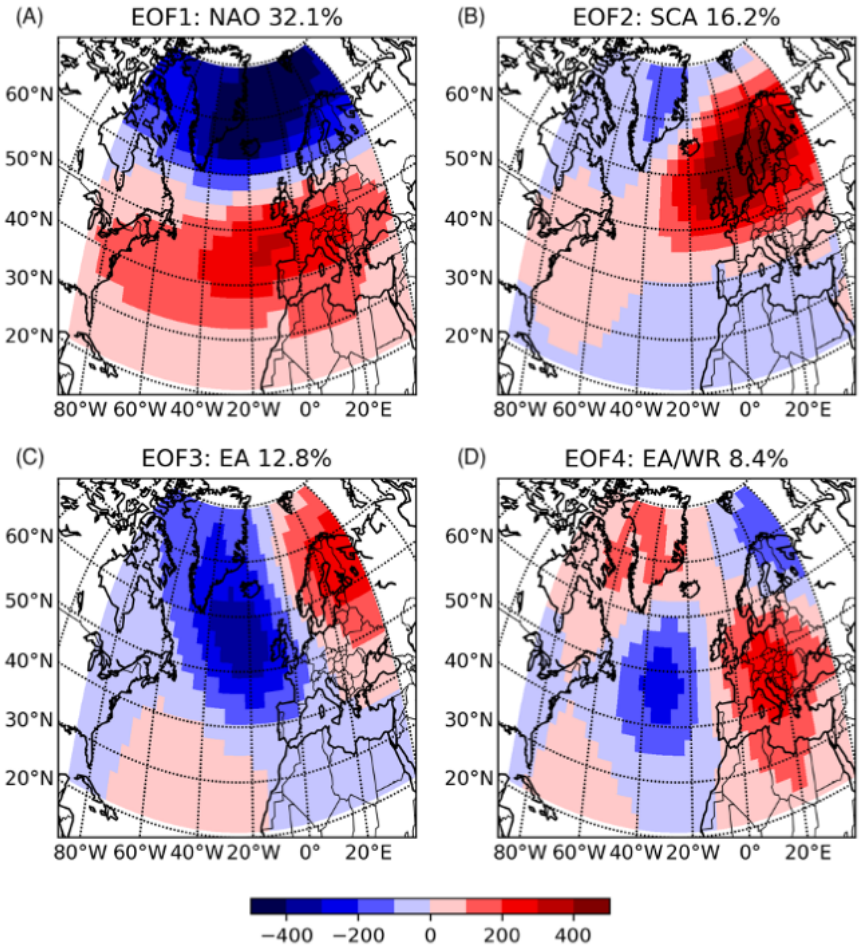


Figure 2.2: First four leading eigenfunctions of the mean sea level pressure in CanESM2. The percentage of variance the mode explains is given on top of the figures, from Felsche and Ludwig (2021)

tem situated in the North Atlantic basin and a high-pressure system over South Europe and South Atlantic (Barnston and Livezey, 1987). A study by Mikhailova and Yurovsky (2016) depicted a close positive relationship between the EA index and the surface air temperature in Europe in winter. A positive correlation between EA and precipitation was found in Spain, France and West Germany (Barnston and Livezey, 1987). The East Atlantic/Western Russia Oscillation (**EA/WR**) (Figure 2.2 (D)) is characterised by two main large-scale anomalies north from the Caspian Sea and western Europe (Lim, 2015). The positive EA/WR phase is defined by a negative pressure anomaly over the Atlantic near $40°W$ and $40-45°N$ and a positive pressure anomaly over central Europe. A study by Lim (2015) showed that positive precipitation anomalies over midlatitude Atlantic and negative anomalies over the European region are associated with the positive EA/WR.

Blocking is another key mechanism contributing to the formation of heatwaves and droughts in

Europe (Figure 2.1). Blocking occurs when a strong anticyclonic persistent weather pattern forms, so strong that it blocks the incoming storms (Sheffield and Wood, 2011) and is self-sustaining (Kautz et al., 2022). This leads to the area under the blocking experiencing hot and dry conditions for an extended period (Sheffield and Wood, 2011). Only a strong incoming anticyclone is able to break it. It has been shown that blocking conditions in Europe can be induced by a positive phase of NAO (Li et al., 2020). An atmospheric blocking caused historical events like the 2003 and 2018 heatwaves and droughts (Kautz et al., 2022; Kueh and Lin, 2020).

### 2.1.2 Land-atmosphere interactions influencing the formation

The influence of land surface on weather is mostly moduled through soil moisture (Hsu and Dirmeyer, 2023). Soil moisture has a direct effect through the partitioning of land heat fluxes, as portrayed in Figure 2.3.

This leads to two distinct evaporation regimes: moisture-limited and energy-limited (Seneviratne et al., 2010; Teuling et al., 2009). In the first case, enough energy is present for evaporation; however, due to dry soil, only little energy can be transmitted to the latent heat flux, which usually contributes to the formation of convective clouds. Therefore, sensible heat flux prevails (Figure 2.3 left). This is the case in the Southern regions of Europe, such as the Mediterranean. The energy-limited regime refers to a case where enough soil moisture is present. However, not enough energy is present. The incoming solar radiation leads to a high evaporative fraction and high latent heat flux (Figure 2.3 right). This regime prevails in Northern Europe in summer (Teuling et al., 2009). The region in between is defined as transitional, where time-wise, both regimes might occur and, therefore, strongly influence the large-scale atmospheric dynamics.

Moreover, heatwaves and droughts can have a mutually reinforcing relationship (Osman et al., 2022; Felsche et al., 2023). With the start of a heatwave, there is a drying effect on soil and vegetation, causing a reduction in evaporation. This, in turn, diminishes the probability of rainfall, creating conditions for the emergence of droughts. On the other hand, with the onset of drought, evaporation decreases. The resulting reduction in cloud cover allows more solar radiation to reach the land surface, increasing the probability of heatwave formation (Figure 2.1). This bidirectional interaction underlines the complex interdependency between heatwaves and droughts, highlighting that through the feedback loops, the events can trigger each other and self-intensify (Miralles et al.,

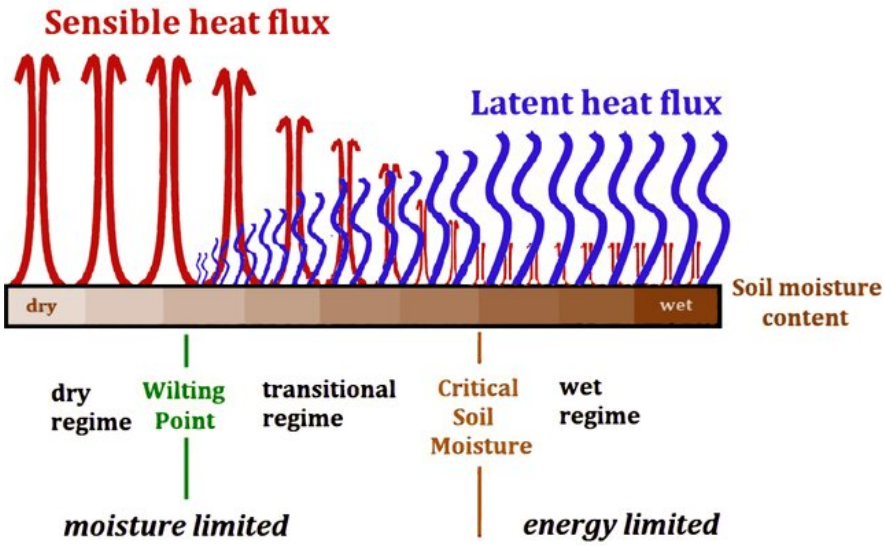


Figure 2.3: Schematical overview of moisture-limited and energy-limited regimes; adapted from Hsu and Dirmey-er (2023)

2019).

The described feedbacks are non-linear and dependent on the observed regions. Quesada et al. (2012) confirm that dry conditions in winter/spring seasons prevail before hot summers over Southern Europe. Moreover, a rainfall deficit in the Mediterranean in spring favours the formation of heatwaves in Northern Europe as the rainfall deficit propagates northward throughout the summer (Vautard et al., 2007; Stefanon et al., 2012). For summers it is confirmed that anomalously dry Western and Northern European (Della-Marta et al., 2007), as well as in South-Eastern European (Hirschi et al., 2011; Whan et al., 2015) summers significantly correlate with the occurrence of heatwaves in those regions (Della-Marta et al., 2007).

### 2.1.3 Observed and projected changes in heatwave and drought occurrence

Since 1950, most regions worldwide have observed a significant increase in heatwave days, maximum duration, and cumulative heat (Perkins-Kirkpatrick and Lewis, 2020). A major heat and/or drought event occurred in Europe every couple of years in the past two decades: 2003, 2010, 2012, 2015, 2018, 2022 (Russo et al., 2015; Hanel et al., 2018; Felsche et al., 2024). In most European capitals, the number of heatwave days and long heatwaves increased when comparing 1998-2015 to 1980-1997 (Bednar-Friedl et al., 2022). The number of droughts has increased in Mediterranean

regions and Central and Eastern Europe (Spinoni et al., 2017). Fewer droughts have been observed in North Europe (Spinoni et al., 2017). Compounding heatwaves and droughts have larger impacts through the increased stress of two extremes and the self-intensifying nature due to the interplay between extreme events. There is *medium confidence*, that compound hazards of warming and precipitation have become more frequent in the past years in Europe (Bednar-Friedl et al., 2022; Mukherjee and Mishra, 2021) (definition of *medium confidence* according to the IPCC).

Under Global Warming conditions, an increase in the median, as well as extreme temperatures, is to be expected. However, due to local drivers and interactions, the effect on precipitation and the related drought phenomenon has to be modelled via climate projections.

Climate model projections estimate that the described trends will continue throughout the 21st century (Garcia-Leon et al., 2021; Ballester et al., 2010). Heatwaves will likely become a major threat for Europe and European cities (Bednar-Friedl et al., 2022). A doubling of heatwave frequency is expected until the middle of the century (2050), with an even larger increase in severe heatwaves (Lhotka et al., 2018). Climate change projections show an increased likelihood of drought occurrence under moderate and extreme greenhouse emission scenarios of RCP 2.6, RCP 6.0 and RCP 8.5 (Grillakis, 2019; Böhnisch et al., 2021). In all scenarios, an increase in spatial extent and duration is expected. The drying trend is expected to be stronger in Southern regions of Europe like the Iberian Peninsula and the Mediterranean (Tuel and Eltahir, 2021; Böhnisch et al., 2021). Regarding future compound heatwave and drought events, an increase is expected in most European regions (Boehnisch et al., 2023). Northern France, Southern Germany, Switzerland, Southern Ireland, and the Western coasts of the Black Sea are expected to experience a tenfold increase in occurrence under the business-as-usual scenario.

If Global Warming is limited to $2°C$, the number of compound hot and dry events would be at least halved (Boehnisch et al., 2023). Limiting the increase below $2°C$ would reduce the share of the European population exposed to hydro-meteorological events by 40% in Europe compared to a $3°C$ scenario (Farinosi et al., 2020).

## 2.2 Observational and modelled climate data

Different kinds of climate data exist, each providing own constraints and advantages, such that the data source has to be intentionally chosen to fit the research question. There is observational data, reanalysis data and model data.

Observational data is measured data via a variety of tools, from weather stations to airplanes and satellites (Saltikoff et al., 2019). The main benefit of observational data is that it is the closest representation we can use to study historical weather events. However, many limitations persist: there is only a limited set of variables that can be measured, the spatial coverage is never complete and is dependent on the density of measuring stations, and finally, the quality of the data can vary, e.g. due to the device uncertainty, dependence of the measurement on certain weather conditions or due to device maintenance, power outages, etc. (Tippett et al., 2015). Statistical tools exist that help to overcome these limitations by extrapolating the data to a spatially consistent grid; however, the physical relationships between the variables might be violated (Notz, 2015).

Reanalysis data is physically consistent modelling data that is merged with historical observations by using data assimilation (Baatz et al., 2021). The assimilation technique ensures that the numerical results reproduce the observations as well as possible while maintaining a physically consistent solution for regions where observational data is missing (Baatz et al., 2021). Moreover, this method has the benefit of producing many variables that cannot be measured. Limitations include that reanalysis is limited to the range of observational data, which is available only from the 1950s. Moreover, models come with known uncertainties, such as the dependence on the number of vertical levels in the model, as well as the limited accuracy in the representation of the upper boundary level (Lilensten et al., 2016). Last but not least, reanalyses require high computational power due to model complexity (Lilensten et al., 2016).

Last but not least, conventional modelling data could be used for various purposes, e.g., to better understand the past and the future, where we do not have historical data, or to assess rare meteorological events. Single-Model Initial-Condition Large Ensembles (SMILEs) are an essential recent climate research and modelling development. The approach enlarges the amount of produced data immensely. A SMILE is produced by repeatedly running a model with slightly iterated initial conditions under the same greenhouse gas forcing scenario (Maher et al., 2021). This methodology allows to distinguish between the climate-change-induced signal and the natural climate variability,

which refers to the internal chaotic nature of the weather system, where slight changes might lead to very different results (Maher et al., 2021).

## Regional Single-Model Initial-Condition Large Ensembles (SMILEs)

There are global as well as regional SMILEs. Global climate models (GCMs) cover the whole Earth's surface and have a spatial resolution of 100 to 450 km due to high computational costs (Liang-Liang et al., 2022). The resolution is often too coarse for investigating extreme hydrometeorological events (Leduc et al., 2019). This problem is solved by nested Regional Climate Models that have a dynamical code and parametrisation; however, they are taking the boundary conditions from a GCM (Leduc et al., 2019). Leduc et al. (2019) find a better representation of local extremes in the regional model when compared to the global model. Another study confirms that RCMs can better represent temperature and precipitation in high mountain regions like the Alps, as well as Central Europe in summer and north-eastern Europe in winter (Kendon et al., 2010).

Figure 2.4 displays the climate change projections of December precipitation for the first 24 members of the CRCM5-LE, the dataset that has been used in this dissertation. The difference in the change signal between the members is evident. The added value of the ensemble lies in the proper estimation of the climate variability, which is only observable when all members are considered. Moreover, regional SMILEs produce a huge statistical basis of extreme events available for investigation, as they produce thousands of model years (Maher et al., 2021). This makes them particularly useful for investigating events like droughts and heatwaves.

Limitations of SMILEs concern the same issues as with conventional modelling. Each model has structural differences from one another in modelling the physical mechanisms, in the choices during the setup and tuning of the model (Lupo et al., 2013; Leduc et al., 2019). This model uncertainty can only be assessed through a comparison to other models under the same greenhouse gas forcing scenario (von Trentini et al., 2019).

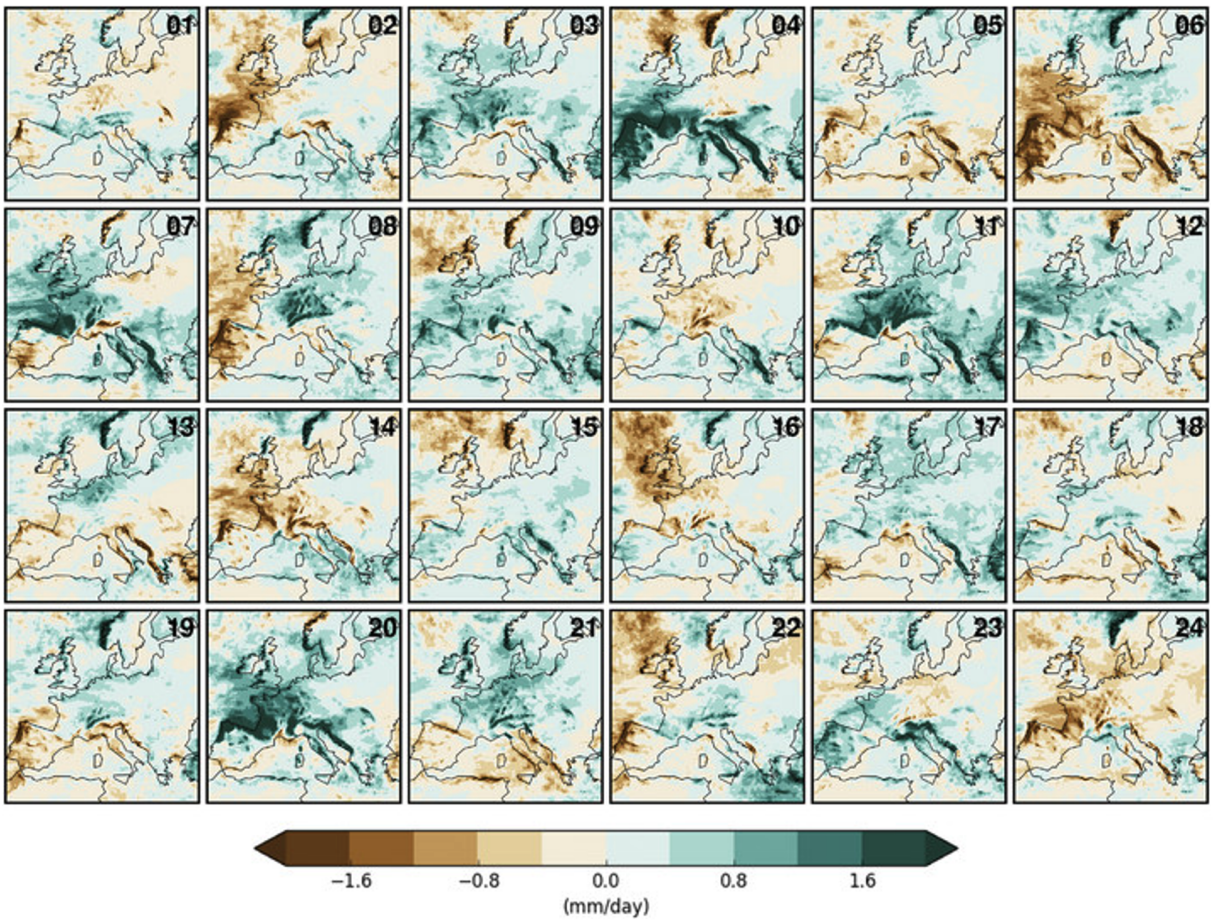


Figure 2.4: Short-term climate change projections (2020–39 vs 2000–19) for mean December precipitation from the ensemble members 1–24 over the EU domain, adapted from Leduc et al. (2019)

## 2.3 Machine Learning

Machine Learning is a field of study in computational informatics that aims to train machines to learn from the provided data without explicitly providing the rules for the generalisation (Shinde and Shah, 2018). Machine Learning has a wide application in various fields, where it is too costly to develop deductive algorithms to fulfil the task. There are many applications ranging from speech recognition, computer vision, and text production to weather forecasts and tumour identification (Shinde and Shah, 2018).

A general differentiation exists between a *supervised* and an *unsupervised* learning approaches. The first refers to learning problems in which target data is given (Bishop, 2006, p. 3). If the

target data is discrete, e.g. 0 or 1, then the problem is called a *classification* problem. In case the target variable is continuous, the problem is called a *regression* problem (Bishop, 2006, p. 3). In *unsupervised* learning, the goal is to discover patterns within the input data without a pre-defined true result. Explicit examples are *clustering* to group the data, *density estimation* of a dataset or the projection of the data to a lower-dimensional space for the purpose of *visualisation* or *compression* (Bishop, 2006, p. 3).

This thesis applies Artificial Neural Networks (ANNs) and a clustering approach in the publications. In the following, those approaches will be introduced.

### 2.3.1 Foundations of ANNs

ANNs are a supervised learning approach. The design of ANNs is inspired by the architecture of the brain (Russell and Norvig, 2009, p. 728). Like connected brain neurons, an ANN consists of interconnected nodes. These nodes have the capacity to learn as they encounter information. They are connected with each other by directed links (Russell and Norvig, 2009). The nodes are organised in layers (Zhang, 2010; Russell and Norvig, 2009). Each node in the layer is connected to all neurons in the preceding and following layer if it is a fully connected network. In an ANN, the learning process of nodes is facilitated through the adjustment of weights associated with the connections between nodes. Each node represents a computational unit that receives input signals and computes an output. The input layer interacts with input data X. The output layer produces the aimed result $y_{predicted}$. A multitude of hidden layers can be located between $input$ and $output$. An exemplary architecture is shown in Figure 2.4 on the left.

A defined activation function, $g$, is used for each neural network to calculate the node output (Russell and Norvig, 2009). Activation functions play a crucial role in the functioning of neural networks as they introduce non-linearity, complexity and expressive power to the model. Without them, the neural network would be a linear transformation of the data, regardless of the number of layers or nodes. A link between the node $i$ and the node $j$ serves to propagate the activation $a_i$ from $i$ to $j$. A numeric weight $w_{i,j}$ is assigned to each connection. The output of the node is computed by:

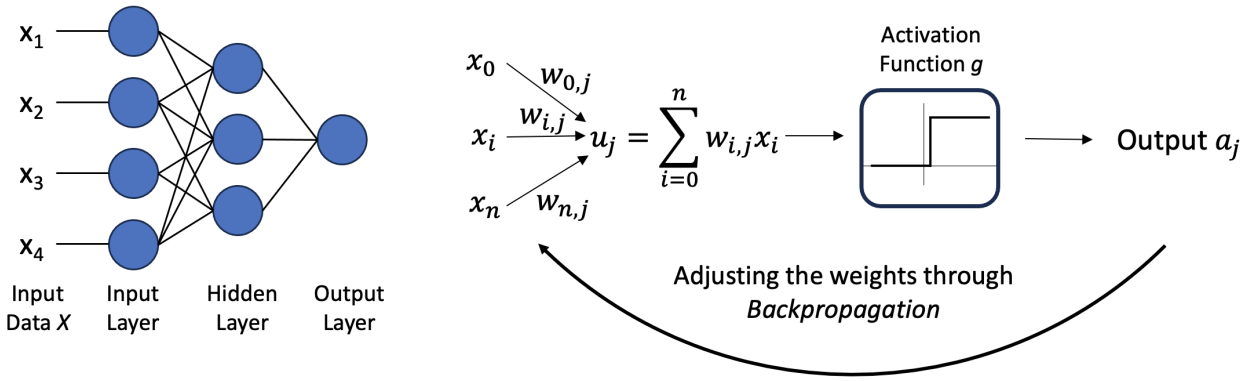


Figure 2.5: Schematical overview of an artificial neural network

$$a_j = g(in_j) = g\left(\sum_{i=0}^{n} w_{i,j} a_i\right) \tag{2.1}$$

where $g$ is an activation function (Russell and Norvig, 2009, p. 728). For example, the activation function could be a hard threshold modelled with a step function or a logistic function called a sigmoid. A schematical overview is displayed in Figure 2.5 on the right.

The optimisation process in artificial neural networks involves adjusting the weights, denoted as $w_{i,j}$, to minimise the error, or loss, between the $y_{true}$ and $y_{predicted}$. Two fundamental algorithms are employed for this purpose: Gradient Descent and Backpropagation (Rumelhart et al., 2013).

The adjustment of weights is guided by the gradient of the loss function with respect to the weights (Janocha and Czarnecki, 2017) (gradient descent). This gradient indicates the direction and magnitude of the steepest ascent of the error. This is first calculated only for the final layer. Backpropagation is then applied to transmit this information from the final layer in the opposite direction of the gradient, effectively descending the error landscape (Russell and Norvig, 2009; Rumelhart et al., 2013). The weight update rule is governed by the learning rate, a hyperparameter that influences the size of each step in the weight adjustment process. The algorithm iteratively applies this process until a satisfactory level of convergence is achieved, and the network produces accurate predictions (Russell and Norvig, 2009; Rumelhart et al., 2013).

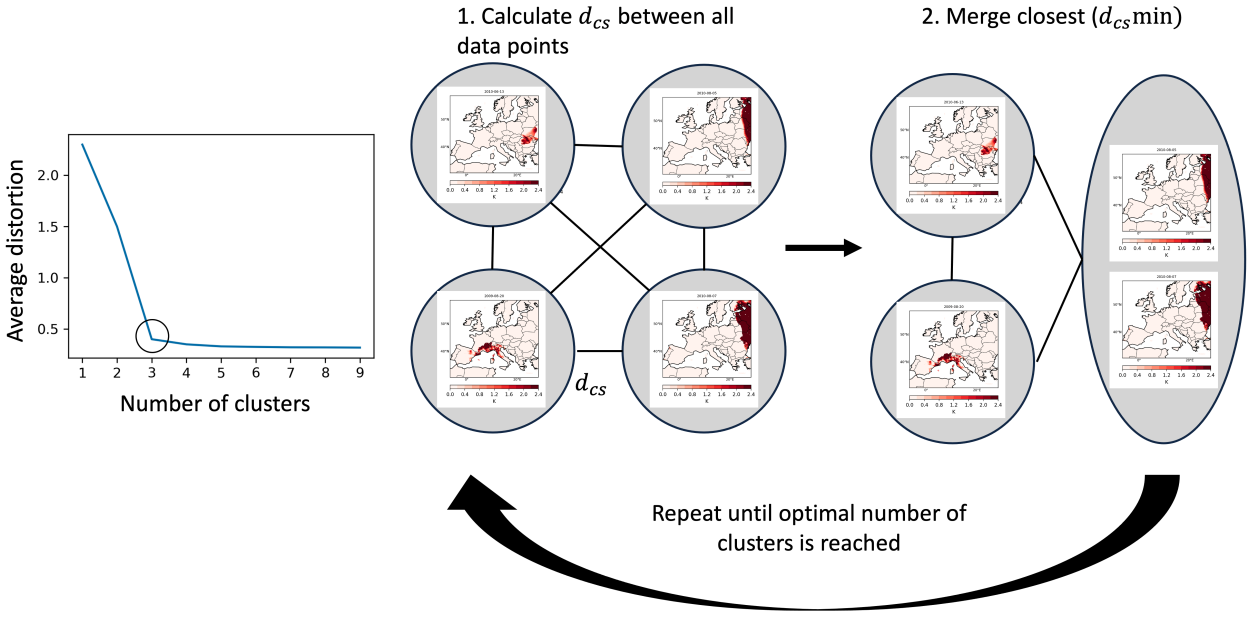


Figure 2.6: Schematical overview of clustering. Left: exemplary distortion plot with the elbow at 3. Right: Steps in the Hierarchical Agglomerative Clustering Algorithm

### 2.3.2 Foundations of Clustering

Clustering is a fundamental method in Data Mining (Rokach and Maimon, 2005), which can be applied as a pattern recognition method. Clustering aims to discover knowledge, hence a new set of categories or groups that share similar characteristics without prior knowledge of the right outcome. The created subsets should be organised in such a way that each member belongs to one and only one subset. Many algorithms use differing approaches for clustering, such as taking the distance between data points, the density, or particular statistical distributions. Two basic types of clustering algorithms exist: hierarchical and relocation methods (Fraley and Raftery, 1998).

Hierarchical methods produce a sequence of subgroups, each corresponding to a different overall number of clusters. Thereby, we can start with each datapoint belonging to their own cluster and merge them - *'agglomerative'*, or we start with all datapoints in the same cluster and split the groups at each stage - *'divisive'*(Fraley and Raftery, 1998; Murtagh and Contreras, 2017). The agglomerative algorithm is schematically shown in Figure 2.6. An optimisation criterion has to be introduced for the splitting or merging to happen. Broadly used criteria for hierarchical agglomerative clustering include a single link (nearest neighbour), complete link (farthest neighbour) or sum of squares (Fraley and Raftery, 1998; Rokach and Maimon, 2005).

Relocation methods have the core idea of moving data points from one cluster to another so that a criterion, like the within-group sum of squares, is minimised. A predefined number of clusters has to be given for this method. The most common and broadly used relocation method is k-means (Fraley and Raftery, 1998).

In both approaches, the problem of determining the correct number of clusters is imminent. The elbow method is one of the most common and easiest approaches to do that. The elbow method measures the distortion for different numbers of clusters in order to identify the optimal number that achieves a balance between minimizing distortion and avoiding excessive complexity. The distortion score is calculated via the squared distance of every point of the cluster to the assigned center. The number of clusters corresponding to the distortion curve's knee is then chosen as the final number of clusters (Jung et al., 2003; Kodinariya and Makwana, 2013). Figure 2.6 on the left shows an exemplary plot of the distortion score curve.

Every clustering approach needs a pre-defined similarity or distance measure, which can tell how similar or dissimilar two data points are. A valid distance measure should be symmetric and obtain a zero value for identical vectors. Cosine Similarity distance $d(r,q)$ is a frequently used measure which gives the angle between two vectors. The smaller the angle between two data points, the more similar they are. The distance between two vectors $r$ and $q$ is therefore defined as follows:

$$d(r,q) = 1 - cs(r,q) \tag{2.2}$$

$$cs(r,q) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M} r_{i,j}q_{i,j}}{(\sum_{i=1}^{N}\sum_{j=1}^{M} r_{i,j})^{1/2}(\sum_{i=1}^{N}\sum_{j=1}^{M} p_{i,j})^{1/2}} \tag{2.3}$$

$cs(r,q)$ refers to the cosine similarity measure between two vectors (Cheng and Wallace, 1993). It is defined as 1 for parallel vectors and as 0 for orthogonal.

### Explainable AI methods

The increased complexity of Machine Learning models turns the algorithms in "black box" approaches, due to the missing transparency in how the models come to their decisions. Explainable Artificial Intelligence is a field aiming to develop methods that would allow interpretability and in-

sight into decision criteria. (Linardatos et al., 2021; Holzinger et al., 2022). In the thesis the SHAP (SHapley Additive exPlanations) method is a applied. The methodology is inspired by game theory (Lundberg and Lee, 2017). It estimates an average marginal contribution of each input feature on the prediction of the result and therefore allows to calculate a relative importance score to each input variable.

### 2.3.3 State of the art usage of AI in Climate Sciences

Although many machine learning methods have been known since the 1960s, the methods initially received little development due to missing computational power needed to train the models (Bochenek and Ustrnul, 2022). Only with a massive increase in computational power and data volumes over the past two decades are machine learning and artificial intelligence enjoying their golden era. The 2012 Deep Learning Revolution marks the point where Deep Learning approaches outperformed other machine learning approaches, demonstrating the capacity to solve more complex problems and increasing the pattern recognition capabilities (Willingham, 2023). 2016 marked another significant milestone, where, for the first time, artificial intelligence was able to beat the world champion in Go. This ancient game before was considered an unsolvable challenge for AI (Sharifani and Amini, 2023; Willingham, 2023). Most recent Chat GPT-3 and GPT-4 launches in 2020 and 2023, respectively, have yet again underlined the immense potential of artificial intelligence (Liu et al., 2023). Those models showcase advanced text and content understanding and demonstrate astonishing capabilities to interact with humans and generate texts from prose to poems and programming code (Liu et al., 2023).

Machine Learning has broad applications in climate sciences. One application is in numerical weather prediction research, where machine learning models are predicting specific variables like wind (Heinermann and Kramer, 2016) or precipitation (Ahmed et al., 2020), the occurrence of extreme events like droughts (Jiang and Luo, 2022), hurricanes (Asthana et al., 2021) or heatwaves (Jacques-Dumas et al., 2022), or even deliver continuous forecasts for the whole atmospheric system (Huntingford et al., 2019). A new hurdle was recently breached in this space: Weather Prediction based on Deep Learning delivered more accurate forecasts than the most accurate physics-based model developed by the European Centre for Medium-Range Weather Forecasts (ECMWF). Huawei's model "Pangu -Weather" and Google's Model "GraphCast" deliver global weather predictions for over ten days at a high spatial resolution (Lam et al., 2023; Bi et al., 2023). Both

models were trained on available reanalysis data. The big advantage of those models is that once trained, they can be run on conventional computers and deliver results within a minute. In contrast, conventional weather prediction models require giant supercomputers (Lam et al., 2023). Nevertheless, both models still have several weaknesses, like the underestimation of extreme events and the inability to deliver ensemble forecasts. Also, doubts persist whether the models can predict global warming-triggered unseen weather events since machine learning models learn from known data (Lam et al., 2023; Bi et al., 2023). Disregarding those challenges, it becomes clear that the potential for further development is enormous.

Another popular application of machine learning lies within synoptic meteorology and climatology, which deals with finding generalised weather patterns. In many cases, no strict and precise definitions exist, such that machine learning can help divide the dataset based on common characteristics. Both applications, supervised when training to label already defined synoptic classes and unsupervised when discovering patterns, find broad usage in research (Stryhal and Plavcová, 2023; Lee and Sheridan, 2012; Mittermeier et al., 2022).

# 3. Research Questions

This thesis aims to investigate the drivers, dynamics and interrelation of hot and dry extremes in Europe. Heatwaves and droughts are both events that share the same large-scale atmospheric drivers and that will intensify under Global Warming conditions. Due to the internal chaotic nature of the weather system, we are facing limited understanding and predictability of the generalised patterns of those events. This thesis thrives on improving this understanding to allow for timely event mitigation and portray the consequences of climate change. This thesis used two fundamental concepts that made the investigation possible: the SMILE framework and machine learning approaches. The SMILE framework allowed for statistically reliable investigation of extreme events with a return period of 20 years or more. Due to the 50-members, 50 times more extreme heat and drought events were available for classification and driver investigation. Machine learning and artificial intelligence are one of the biggest technological revolutions and potentials of our time. In the thesis, basic machine learning concepts were applied to research the physical features of heatwaves and droughts. Explainable AI approaches supported the investigation and allowed to determine the variables that allowed the prediction.

The scientific research projects that are part of this thesis provide answers to the following research question:

> *Can machine learning approaches facilitate the research on prediction and interrelation of heatwaves and droughts in Europe?*
> *If yes, what can we learn from it about the physical nature of the events and the expected changes in the future?*

All three publications address the overarching research question. A schematical overview of the publications in Figure 3.1 portrays the investigated event type, the timeframe and the applied machine learning method. In the following, the specific research questions of each publication are introduced:

1. The first publication investigates drought prediction for two European locations with differing climates: Munich and Lisbon. The research questions are the following:

    - *Q1.1:* What is the potential of machine learning in drought prediction? What opportunities for research does it offer?

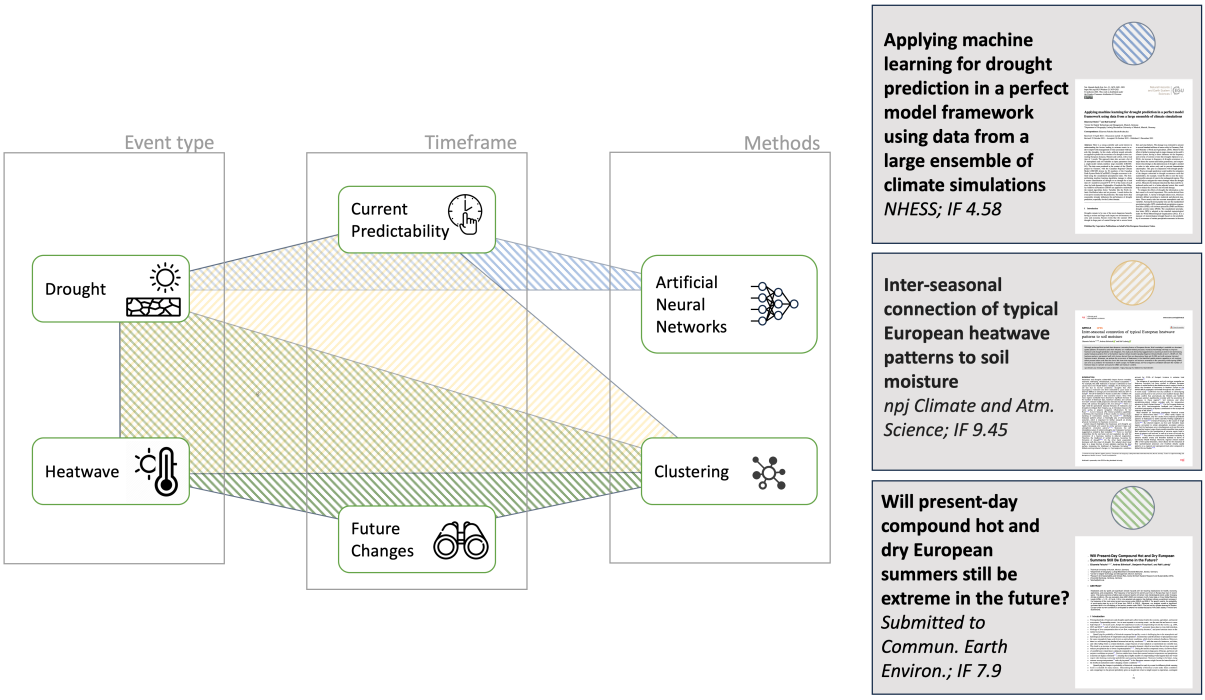


Figure 3.1: Schematic overview of the publications for the investigated event type, timeframe and machine learning method

- *Q1.2:* What are the relevant variables for drought prediction in Lisbon? What are the relevant variables for drought prediction in Munich?
- *Q1.3:* What role does seasonality play in the prediction? Are there particular seasons where droughts are better predictable?

2. In the second research project, the phenomenon of European summer heatwaves and their interrelation with the seasonal soil moisture anomaly is investigated. For any study dealing with spatio-temporal events, the question of meaningful dimensionality reduction becomes eminent. Therefore, the study aims to find typical heatwave patterns that can be used to study the interrelation with soil moisture.

   - *Q2.1:* How can we identify typical patterns of heatwaves in Europe with the help of machine-learning-powered clustering?
   - *Q2.2:* How does a soil moisture deficit in spring influence the occurrence of a summer heatwave in the identified regions?
   - *Q2.3:* What is the influence of a heatwave in summer on the soil moisture deficit in the following fall for each of the identified regions?

3. The third research project brings the first two together. It investigates compound hot and dry events by first identifying the typical patterns with the same methodology as in Paper II and

then quantifying the probabilities of occurrence of historical compound hot and dry events in Europe under current conditions, a Global Warming Level (GWL) of $+2°C$ and a $+3°C$ degree.

- *Q3.1:* How can we identify typical patterns of compound hot and dry events in Europe with the help of machine-learning-powered clustering? What are the most extreme historical compound hot and dry events in those regions?
- *Q3.2:* How will the probability of historical compound hot and dry events change under the influence of climate change under two possible scenarios of GWL2 and GWL3?
- *Q3.3:* How will the climatology of compound hot and dry events change under future climate?

Two of the three papers have been published in peer-reviewed journals: *Natural Hazards and Earth System Sciences* (NHESS) and *npj Climate and Atmospheric Science*. The last paper has been submitted to *Communications Earth & Environment*. All the publications are introduced by an overview page, which consists of the most relevant information on the publication, including a plain language summary, information on the authors, authors' contributions, the journal, and the publication status. The papers are presented chronologically, as the last publication combines concepts introduced in the first two.

The thesis is based on interdisciplinary research in meteorology, climatology, statistics and data science and presents state-of-the-art geoscience research. Modern geosciences apply computer sciences to model and analyse the physical processes of the Earth System in contrast to older and traditional methods of observations via pen and paper, measurements via physical devices and traditional statistics. The enhanced tools for analysis via machine learning are of high potential for the future to discover physical knowledge by learning from the data and comparing to existing physical theories. The presented methods are part of interpretable machine learning and pattern recognition approaches. Future applications of machine learning in geosciences will boost the performance both in modelling and analysis.

In addition to the three publications that are part of the thesis, the author contributed to three projects in collaboration with co-authors. The research targets related topics of dynamics of heatwaves, the expected changes of compound dry and hot events in agricultural areas, as well as the expected impact of climate change on crop yields.

Böhnisch, A., **Felsche, E.**, Ludwig, R. (2023). European heatwave tracks: using causal discovery to detect recurring pathways in a single-regional climate model large ensemble. Environmental Research Letters, 18(1), 014038.

Boehnisch, A., **Felsche, E.**, Mittermeier, M., Poschlod, B., Ludwig, R. (2023). Future hotspots of compound dry and hot summers emerge in European agricultural areas. Authorea Preprints.

Schmidt, M., **Felsche, E.** (2023). The effect of climate change on crop yield anomaly in Europe. Climate Resilience and Sustainability, 3(1), e261.

# 4. Scientific Publications

## 4.1 Paper I: Applying machine learning for drought prediction in a perfect model framework using data from a large ensemble of climate simulations

**Reference:** Felsche, E., Ludwig, R. (2021). Applying machine learning for drought prediction in a perfect model framework using data from a large ensemble of climate simulations. Natural Hazards and Earth System Sciences, 21(12), 3679-3691.

**Plain language summary:** There is a strong scientific and social interest in understanding the factors leading to extreme events like droughts in order to improve risk management. This study applies an artificial neural network (ANN) to predict drought occurrence in two European regions with differing climatology: Munich and Lisbon, with a lead time of 1 month. The approach takes into account a list of 28 atmospheric and soil variables from a single-model initial-condition large ensemble (CRCM5-LE) as input parameters to the ANN model. Drought occurrence is defined using the Standardized Precipitation Index (SPI). The best-performing machine learning algorithms manage to obtain a correct classification of drought or no drought for a lead time of 1 month for around $55\% - 57\%$ of the events of each class for both domains. An analysis of the variables that have the highest impact on the prediction is performed. The study shows that the North Atlantic Oscillation index and air pressure 1 month before the event have the highest importance for the prediction for both domains. For the model trained for the Lisbon domain, the variables of northward near-surface wind (vas) and evaporation (evspsbl) are the next strongest predictors. For the Munich domain, teleconnection indices EAWR and SCA 5 months before the event are found among the strongest predictors. In general the prediction for the Lisbon domain is more reliable in terms of absolute accuracy and contribution of individual variables.

**Author's contribution:** This study was conceptualised by EF under the supervision of RL. Formal analysis, visualization of results and writing of the original draft were performed by EF. All authors contributed to the interpretation of the findings and revision of the paper.

**Scope of the journal:** "Natural Hazards and Earth System Sciences (NHESS) is a not-for-profit interdisciplinary and international journal dedicated to the public discussion and open-access publication of high-quality studies and original research on natural hazards and their consequences.

Embracing a holistic Earth system science approach, NHESS serves a wide and diverse community of research scientists, practitioners, and decision makers concerned with detection of natural hazards, monitoring and modelling, vulnerability and risk assessment, and the design and implementation of mitigation and adaptation strategies, including economical, societal, and educational aspects." (EGU, 2024)

**Status:** published

**Journal:** Natural Hazards and Earth System Sciences (NHESS)

**Impact Factor (2-Year):** 4.58

Nat. Hazards Earth Syst. Sci., 21, 3679–3691, 2021
https://doi.org/10.5194/nhess-21-3679-2021


Natural Hazards and Earth System Sciences
Open Access
EGU

# Applying machine learning for drought prediction in a perfect model framework using data from a large ensemble of climate simulations

**Elizaveta Felsche**[1,2] **and Ralf Ludwig**[2]

[1]Center for Digital Technology and Management, Munich, Germany
[2]Department of Geography, Ludwig Maximilian University of Munich, Munich, Germany

**Correspondence:** Elizaveta Felsche (felsche@cdtm.de)

Received: 10 April 2021 – Discussion started: 15 April 2021
Revised: 25 October 2021 – Accepted: 26 October 2021 – Published: 3 December 2021

**Abstract.** There is a strong scientific and social interest in understanding the factors leading to extreme events in order to improve the management of risks associated with hazards like droughts. In this study, artificial neural networks are applied to predict the occurrence of a drought in two contrasting European domains, Munich and Lisbon, with a lead time of 1 month. The approach takes into account a list of 28 atmospheric and soil variables as input parameters from a single-model initial-condition large ensemble (CRCM5-LE). The data were produced in the context of the ClimEx project by Ouranos, with the Canadian Regional Climate Model (CRCM5) driven by 50 members of the Canadian Earth System Model (CanESM2). Drought occurrence is defined using the standardized precipitation index. The best-performing machine learning algorithms manage to obtain a correct classification of drought or no drought for a lead time of 1 month for around 55 %–57 % of the events of each class for both domains. Explainable AI methods like SHapley Additive exPlanations (SHAP) are applied to understand the trained algorithms better. Variables like the North Atlantic Oscillation index and air pressure 1 month before the event prove essential for the prediction. The study shows that seasonality strongly influences the performance of drought prediction, especially for the Lisbon domain.

## 1 Introduction

Droughts remain to be one of the most dangerous hazards, having a serious and large-scale impact on environment, society and economy. Recent events like the summer 2018 drought in huge parts of central Europe led to severe forest fires and crop failures. The damage was estimated to amount to several hundred millions of euros solely in Germany (Federal Ministry of Food and Agriculture, 2018). Moreover the effect of global warming leads to major changes in the earth's climate system, having a direct influence on the frequency and severity of extreme events like droughts (Spinoni et al., 2016). An increase in frequency of drought occurrence is a major threat for current and future generations, and comprehensive knowledge on the phenomenon of drought is needed in order to take action early and to prevent humanitarian catastrophes. This goes in conjunction with drought prediction. Precise drought prediction would enable the mitigation of the dangers connected to drought occurrences such that stakeholders, for example, would be able to store the maximal possible amount of water in the endangered regions. This would help to mitigate the water shortage when the drought arrives. Measures for demand reduction like that could be introduced earlier and to a better-adjusted extent; this would help to reduce the economic and societal damage.

To mitigate the effects of droughts the information on the their onset is of crucial importance. This can be derived from a drought index. A variety of drought indices exist, which are typically defined according to statistical and physical measures. These mostly take into account atmospheric and soil variables. Among the most popular ones are the standardized precipitation index (SPI), standardized precipitation evaporation index (SPEI), soil moisture percentile (SMP) and Palmer drought severity index (PDSI). The standardized precipitation index (SPI) is adopted as the standard meteorological index by World Meteorological Organization (2012). It is a measure of meteorological drought based on the probability of occurrence of certain precipitation amounts in the area

of interest (Sheffield and Wood, 2011). Studies on drought prediction by Belayneh et al. (2016) and Bonaccorso et al. (2015) use SPI as a prediction variable for the forecast.

Forecasting of any physical phenomenon can be done by either a physical, conceptual or data-driven model. The latter ones are widely used due to their rapid development times and the flexibility in input parameters. McGovern et al. (2017) argue that AI methods have a high potential for prediction of extremes due to the ability of machine learning methods to learn from past data, to handle large numbers of input variables, to integrate physical understanding into the models and to discover additional knowledge from the data.

A review of seasonal drought prediction given by Hao et al. (2018) identifies two typical predictor groups of variables: large-scale climate indices that reflect the atmosphere–ocean circulation patterns and local climate variables. The first ones are known to correlate with precipitation patterns in special regions and therefore are naturally correlated with the occurrence of drought. The teleconnection indices important for European precipitation include North Atlantic Oscillation (NAO), Scandinavian Oscillation (SCA), East Atlantic–Western Russia Oscillation (EAWR), East Atlantic Oscillation (EA) and Atlantic Multidecadal Oscillation (AMO) (Hao et al., 2018). As shown by Folland et al. (2009) a positive NAO index in summer is associated with dry and warm conditions in the northwest of Europe, whereas southern Europe and the Mediterranean experience cooler and wetter conditions. More information on the influence of the NAO, SCA, EA and EAWR on the European climate can be found in Folland et al. (2009), Bueh and Nakamura (2007), Mikhailova and Yurovsky (2016), Lim (2015); Barnston and Livezey (1987) and Sheffield et al. (2009). A positive phase of AMO is associated with humid conditions over Great Britain and parts of Scandinavia and with dry conditions in the Mediterranean (Sheffield and Wood, 2011, p. 26); the negative phase is associated with a reversed pattern: dry conditions in Great Britain and wet conditions in the Mediterranean. A study by Sheffield et al. (2009) showed a correlation between the amount of droughts and AMO of 62 % with a significance at the 90 % level. A recent study by Bonaccorso et al. (2015) uses NAO for prediction of probability of drought occurrence for Sicily. The local climate variables like precipitation, temperature and soil moisture were also used as inputs to reflect the conditions at the time the prediction occurs. Belayneh et al. (2016) and Bonaccorso et al. (2015) used SPI for the past months as input variable to the algorithm. A study by Morid et al. (2007) used precipitation as an input parameter.

This paper examines the possibilities of meteorological drought prediction with the lead time of 1 month, applying artificial neural networks (ANNs) for two domains with different climate: one with Mediterranean (Lisbon) and one with continental climate (Munich) (Ceglar et al., 2019). Both sites experienced an increase in drought frequency when comparing 2015 and 1950 and are projected to keep rising under RCP4.5 as well as RCP8.5. (Spinoni et al., 2017). Observational data offer only a limited field for drought investigation, as can be seen from the following approximation. Systematical weather observations started in 1781 by the Societas Meteorologica Palatina (Kington, 1980). In this study $\mathrm{SPI1} < -1$ is used as a threshold for drought occurrence. It corresponds to the 15 % driest months (Keyantash and National Center for Atmospheric Research Staff, 2018) and can be estimated by a total number of 430 observed events until the year 2020 (Eq. 1).

$$(2020 - 1781)\,\text{years} \cdot 12\,\text{months}\,\text{yr}^{-1} \cdot 15\,\% = 430\ \text{events} \quad (1)$$

Compared to that, CRCM5-LE offers a total number of roughly 4500 events when using the first 50 years from the climate simulation data (1955–2005) (see Eq. 2).

$$50\ \text{years per member} \cdot 50\ \text{members} \cdot 12\,\text{months}\,\text{yr}^{-1} \cdot 15\,\% = 4500\ \text{events} \quad (2)$$

This is a difference of an order of magnitude. The more data are available the better the predictions that can be derived by a drought-predicting machine learning model and the more can be learned about drought formation. According to von Trentini et al. (2020) precipitation in summer and winter derived from the European gridded dataset (E-OBS) does fall to a high percentage into the range produced by CRCM5-LE for the historic period. Therefore, the CRCM5-LE proves applicable to this study, and its larger number of extreme events can be used as input to the machine learning algorithms. In this study a variety of ANNs are trained. The best-performing models are investigated, using explainable AI methods to understand the results.

While no comparable study exists for the Munich domain, Santos et al. (2014) performed a drought prediction based on SPI6 for Portugal for the months April, May and June using the following input variables: sea surface temperatures (JFM), NAO (DJFM) and cumulative precipitation (NDJFM for $\mathrm{SPI6_{April}}$, DJFM for $\mathrm{SPI6_{May}}$, JFM for $\mathrm{SPI6_{June}}$). The best results were achieved for the prediction of SPI6 for April, with a correlation coefficient of 0.98. SPI6 for May and June referred to a correlation coefficient of 0.78 and 0.77, respectively.

## 2 Data and methods

### 2.1 Datasets

To investigate the predictability of drought data from the single-model initial-condition large ensemble (SMILE) consisting of 50 members, the Canadian Regional Climate Model 5 Large Ensemble (CRCM5-LE) is used. The data were produced within the scope of the ClimEx project (Leduc et al., 2019, http://www.climex-project.org, last access: 28 November 2021). The CRCM5-LE was generated

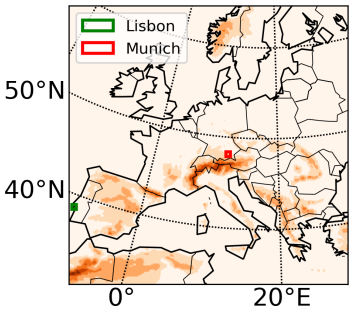


**Figure 1.** CRCM5 topography.

by dynamical downscaling of the data provided by the 50-member initial-condition Canadian Earth System Model 2 using the Canadian Regional Climate Model 5 (Martynov et al., 2013). The data have a resolution of 0.11° (12 km) and are produced for the years 1950–2099 for a European and an eastern North America domain. For the years 1950–2005 the historical greenhouse gas concentrations and aerosol emissions are being used. Starting from 2005, the model introduces the RCP8.5 (IPCC, 2013) forcing scenario. A total of 42 atmospheric variables are available at a temporal resolution of 1 to 3 h. They are used on a monthly basis as input to the machine learning algorithms. The list of variables is provided in Table 1.

In the study we use monthly sea level pressure (*psl*) from the driving model CanESM2-LE (Kushner et al., 2018; Kirchmeier-Young et al., 2016) for the calculation of North Atlantic Oscillation (NAO), Scandinavian Oscillation (SCA), East Atlantic Oscillation (EA) and East Atlantic–Western Russia Oscillation (EAWR) over the whole Atlantic basin (20–80° N, 90° W–40° E). The Atlantic Multidecadal Oscillation (AMO) is calculated using the sea surface temperature (SST) over 0–60° N, 0–80° W from the CanESM2. Only the period 1955–2005 is considered in order to stay within the scope of historical climate. The CRCM5 domain is displayed in Fig. 1. For the machine learning training a grid point situated at 48.11° N, 11.91° E is referenced as Munich, and 38.67° N, 9.17° W is referenced as Lisbon.

### 2.2 Input variables for drought prediction

In order to calculate NAO, SCA, EA and EAWR, the method introduced by Hurrell et al. (2003) is used: a principal component analysis (PCA) of the monthly psl is performed over the 20–80° N, 90° W–40° E domain. The leading eigenvectors, scaled by the amount of variance they explain, represent the leading circulation patterns of the atmospheric system. The first eigenvector corresponds to NAO, the second one to SCA, the third one to EA and the fourth one to EAWR. To calculate the teleconnection indices (NAO, SCA, EA, EAWR) the eofs package described in Dawson (2016) is used. It is an implementation of the technique of empirical orthogonal functions (EOFs) (Dawson, 2016). The leading modes of the PCA corresponding to NAO, SCA, EA and EAWR derived from the CanEsm2 dataset are shown in Fig. 2.

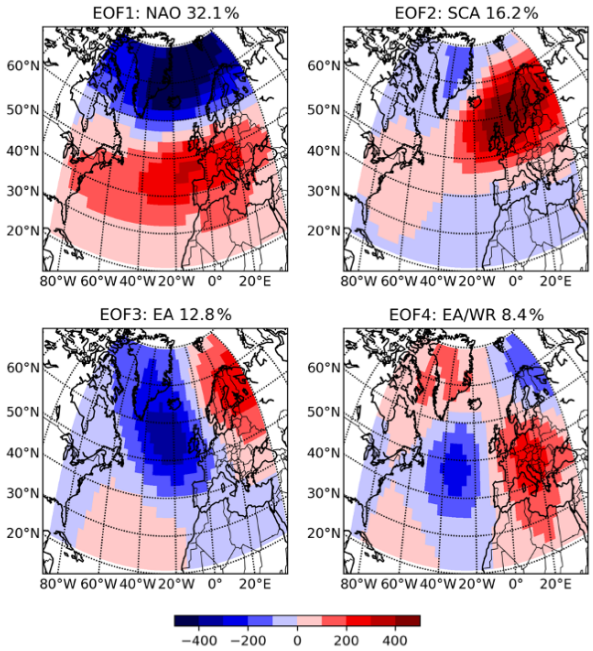


**Figure 2.** First four leading eigenfunctions of the mean sea level pressure in CanESM2. Percentage of variance the mode explains is given at the top of the panels.

AMO is calculated by spatial averaging over the 0–60° N, 0–80° W area of the anomaly of sea surface temperature (Trenberth, 2011). Additionally the 10-year running mean of AMO is calculated as an input variable as it is widely used in various studies and was shown to be correlated with precipitation (Enfield et al., 2001).

Variable subset selection helps to limit the computational time and to improve predictive accuracy (Kumar and Minz, 2014). In order to eliminate redundant variables, Pearson's $R$ between all the CRCM5 variables for the chosen domains is calculated. Pearson's $R$ ($\rho_{X,Y}$) is a measure of linear correlation between two variables $X$ and $Y$; $\rho_{X,Y}$ equals 1 if the correlation is totally positive, 0 if there is no linear correlation and $-1$ if the correlation is total negative (Guyon and Elisseeff, 2003). For two samples $x$ and $y$, Pearson's $R$ is defined in the following way:

$$\rho_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}. \quad (3)$$

The bar refers to the average over the index $i$ (Guyon and Elisseeff, 2003). Pearson's $R$ is a popular and easy method for feature selection of continuous variables as introduced in

**Table 1.** The 42 monthly atmospheric and soil variables from CRCM5-LE. TOA refers to top of atmosphere.

| | | | | | |
|---|---|---|---|---|---|
| *clt* | Total cloud fraction | % | *prw* | Water vapor path | $\mathrm{kg\,m^{-2}}$ |
| *dds* | Near-surface dew point depression | K | *ps* | Surface air pressure | Pa |
| *evspsbl* | Evaporation | $\mathrm{kg\,m^{-2}\,s^{-1}}$ | *psl* | Sea level pressure | Pa |
| *evspsblland* | Water evaporation from land | $\mathrm{kg\,m^{-2}\,s^{-1}}$ | rlds | Surface downwelling longwave radiation | $\mathrm{Wm^{-2}}$ |
| *hfls* | Surface upward latent heat flux | $\mathrm{Wm^{-2}}$ | *rlus* | Surface upwelling longwave radiation | $\mathrm{Wm^{-2}}$ |
| *hfss* | Surface upward sensible heat flux | $\mathrm{Wm^{-2}}$ | *rlut* | TOA Outgoing longwave radiation | $\mathrm{Wm^{-2}}$ |
| *hurs* | Near-surface relative humidity | % | *rsaa* | Shortwave radiation absorbed by atmosphere | $\mathrm{Wm^{-2}}$ |
| *huss* | Near-surface specific humidity | 1 | *rsds* | Surface downwelling shortwave radiation | $\mathrm{Wm^{-2}}$ |
| *mrfso* | Soil frozen water content | $\mathrm{kg\,m^{-2}}$ | *rsdt* | TOA incident shortwave radiation | $\mathrm{Wm^{-2}}$ |
| *mrlso* | Soil liquid water content | $\mathrm{kg\,m^{-2}}$ | *rsus* | Surface upwelling shortwave radiation | $\mathrm{Wm^{-2}}$ |
| *mrro* | Total runoff | $\mathrm{kg\,m^{-2}s-1}$ | *rsut* | TOA outgoing shortwave radiation | $\mathrm{Wm^{-2}}$ |
| *mrros* | Surface runoff | $\mathrm{kg\,m^{-2}\,s^{-1}}$ | *sfcWindmax* | Daily maximum near-surface wind speed | $\mathrm{m\,s^{-1}}$ |
| *mrso* | Total soil moisture content | $\mathrm{kg\,m^{-2}}$ | *snc* | Snow area fraction | % |
| *mrsos* | Moisture in upper portion of soil column | $\mathrm{kg\,m^{-2}}$ | *snd* | Snow depth | m |
| *prc* | Convective precipitation | $\mathrm{kg\,m^{-2}\,s^{-1}}$ | *snw* | Surface snow amount | $\mathrm{kg\,m^{-2}}$ |
| *prdc* | Deep convective precipitation | $\mathrm{kg\,m^{-2}\,s^{-1}}$ | *tas* | Near-surface air temperature | K |
| *prfr* | Freezing rain | $\mathrm{kg\,m^{-2}\,s^{-1}}$ | *tasmax* | Daily maximum near-surface temperature | K |
| *pr* | Precipitation | $\mathrm{kg\,m^{-2}\,s^{-1}}$ | *tasmin* | Daily minimum near-surface temperature | K |
| *prlp* | Liquid precipitation | $\mathrm{kg\,m^{-2}\,s^{-1}}$ | *ts* | Surface temperature | K |
| *prrp* | Refrozen rain | $\mathrm{kg\,m^{-2}\,s^{-1}}$ | *uas* | Eastward near-surface wind | $\mathrm{m\,s^{-1}}$ |
| *prsn* | Snowfall flux | $\mathrm{kg\,m^{-2}\,s^{-1}}$ | *vas* | Northward near-surface wind | $\mathrm{m\,s^{-1}}$ |

**Table 2.** List of sorted-out variables.

| Kept variable | Sorted-out variable | Pearson's $R$ |
|---|---|---|
| *hurs* | *dds* | −0.9879 |
| *evspsbl* | *evspsblland* | 0.9994 |
| *evspsbl* | *hfls* | 0.9988 |
| *mrso* | *mrlso* | 0.9991 |
| *rlut* | *rlaa* | −0.9549 |
| *tas* | *rlds* | 0.9550 |
| *tas* | *rlus* | 0.9960 |
| *rnt* | *rns* | 0.9954 |
| *rnt* | *rsaa* | 0.9831 |
| *rnt* | *rsdt* | 0.9970 |
| *rnt* | *rss* | 0.9872 |
| *rnt* | *rst* | 0.9926 |
| *tas* | *tasmax* | 0.9932 |
| *tas* | *tasmin* | 0.9864 |

Biesiada and Duch (2007); $\rho$ is calculated for all possible permutations of the 41 input variables. The ones correlating to a high degree are examined, and a threshold of 0.95 is chosen. In Table 2 a list of sorted-out variables and the corresponding values of Pearson's $R$ is given. The high correlation values can be explained by a physical relationship between the variables: e.g., the total evaporation (*evspsbl*) is almost the same as evaporation from land (*evspsblland*) as there are no relevant water bodies in the chosen domains. Out of the full list of 42 variables, 14 are sorted out as being redundant.

### 2.3 Standardized precipitation index

The standardized precipitation index (SPI) is a precipitation-based index introduced by McKee et al. (1993). For the calculation of SPI a continuous monthly precipitation dataset is used. The index can be calculated on different timescales: typically, it is 1, 3, 6, 12 or 24 months. As a first step the precipitation values are accumulated for the needed timescale. The resulting dataset is fitted to a gamma distribution for each month separately and then transformed to a normal distribution such that the mean SPI is zero. The SPI value for a given precipitation is then the number of standard deviations from normal. Because of the normalization, the SPI is especially useful to represent wetter and drier climates as well as to account for differences among seasons. As the two study sites are having different meteorological conditions, the SPI provides a convenient and comparable measure (Zargar et al., 2011). As noted in Yoon et al. (2012) the accumulation period of the SPI value needs to be chosen equal to or less than the prediction lead time as otherwise the precipitation values needed for the mathematical calculation of the SPI would be given as input to the machine learning algorithm. Therefore the accumulation period of 1 month is chosen. SPI1 is calculated for Lisbon and Munich each using the data from 1955–2005 from all members as reference.

### 2.4 Machine learning

This study investigates drought predictability applying the technique of supervised machine learning for this purpose. Machine learning is a promising tool for the analysis of complex and data-rich phenomena as droughts (McGovern et al., 2017). The Python package Keras, a high-level neural net-

work package, is used for the design of the machine learning models (Chollet et al., 2015) as it allows the design of neural networks in an easy way by adding layers. Three crucial elements are needed to perform drought prediction by supervised machine learning: input data; a target variable to be predicted; and a computation pipeline, which includes the machine learning algorithm.

The data from the years 1957–1999 are used as training data; the years 2000–2005 are used for the testing purpose. Each of the time periods is available 50 times as we are dealing with an ensemble of 50 members. This results in 2150 model years for training and 250 years for testing. A small fraction of the training data are used for the validation of the machine learning algorithms. The target variable chosen for the prediction of droughts is SPI1. Two classes for the prediction are identified in the following way: SPI1 $< -1$ is defined as an event and is initialized with 1; SPI1 $> -1$ is initialized with 0 and corresponds to a non-drought event. The lead time of 1 month is chosen for the prediction as it has been used in previous studies by Yoon et al. (2012) and Deo et al. (2017). Moreover shorter prediction lead times usually obtain better results when compared to longer periods, as seen in Bonaccorso et al. (2015). After the feature selection 28 variables originating directly from the CRCM5-LE dataset are used as input. In addition to those the teleconnection indices NAO, SCA, EA, EAWR, AMO and AMO10 are used as input.

To predict, for example, a drought or non-drought in April of 1980, the data for 12 months before the event are used as input. This is NAO and other teleconnection and atmospheric variables for the period April 1989–March 1980; for a prediction of an event in May 1980, May 1989–April 1980 is used as input. The 12 months before the event are chosen in accordance with the study by Morid et al. (2007), who found that the best-performing drought prediction model was the one including the value up to 12 months before the predicted one. We perform a time series prediction with no limitation on special months or seasons to be inspected.

For this analysis we use a supervised machine learning algorithm, an artificial neural network (ANN). ANNs are algorithms whose design is inspired by the architecture of the human brain with its neurons (Russell and Norvig, 2009); they both consists of connected nodes. A link between the node $i$ and the node $j$ serves to propagate the activation $a_i$ from $i$ to $j$. To each connection a numeric weight $w_{i,j}$ is assigned. The output of the node is computed by

$$a_i = g(in_j) = g\left(\sum_{i=0}^{n} w_{i,j} a_i\right) \tag{4}$$

(Russell and Norvig, 2009, p. 728). The activation function defines the output of the node. In order to have stable learners with confident predictions a function with a soft threshold is recommended (Russell and Norvig, 2009). In this study the following three activation functions are used: sigmoid, rectified linear unit (ReLU), exponential linear unit (ELU). Sigmoid activation is especially useful for the output layer (Russell and Norvig, 2009), while ReLU and ELU both have the property of allowing very fast optimization (Maas, 2013).

The *sigmoid* function, also called the logistic function, is defined in the following way:

$$\text{Logistic}(x) = \frac{1}{1+e^{-x}} \tag{5}$$

(Russell and Norvig, 2009). This function has an output between 0 and 1. This can be interpreted as a probability of belonging to the class 1. One of the main disadvantages of the sigmoid activation function is the vanishing gradient problem: at higher, almost saturated layers with values of 1 or $-1$, the gradients become nearly 0, resulting in a slow optimization convergence (Russell and Norvig, 2009, p. 726).

*ReLU* refers to rectified linear unit and shows better performance when dealing with the vanishing gradient problem (Maas, 2013). ReLU is defined in the following way:

$$f(x) = \max(0, x). \tag{6}$$

*ELU* refers to the exponential linear unit and was introduced by Clevert et al. (2016). Clevert et al. (2016) claim that in experiments the ELU activation led to faster learning and significantly better generalization performance than ReLU and sigmoid activation. The function is defined as

$$f(x) = \begin{cases} x \text{ if } x > 0 \\ \alpha(\exp(x) - 1) \text{ if } x \leq 0; \end{cases} \tag{7}$$

$\alpha$ controls the value to which an ELU saturates for negative inputs. Per default the value is set to 1 such that the function saturates at $-1$.

Two kinds of layers are used in this study: dense and dropout. *Dense* refers to a regular fully connected neural network layer. *Dropout* refers to a layer which is randomly setting a fraction of inputs to zero at each update. This technique is used to prevent overfitting and therefore to improve the performance of the algorithm (Chollet et al., 2015). The first part of the study concentrates on the methodological search for the best-performing algorithms. A pipeline to search for the best-performing architecture, value for L2 regularization and loss function is built up.

The model performance is evaluated using accuracy and F1 score (Sasaki, 2007). The latter one is especially useful when training on datasets with an imbalanced class distribution as it is in the case of our dataset. Accuracy is defined in the following way:

$$\text{Accuracy} = \frac{\text{Number of right predictions}}{\text{Total number of samples}}. \tag{8}$$

F1 score is a harmonic measure between precision and recall. Precision is the amount of true positives with respect to the amount of positively classified data. Recall is the amount

of true positives with respect to the total number of positives in the data. The F1 score is defined in the following way:

$$\text{F1 score} = 2\frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (9)$$

Due to the class imbalance within the dataset we require that the accuracy of each class is at least 50 %. In that case given the distribution of the test dataset of 1803 non-drought events to 387 droughts for Lisbon and 1848 non-drought events to 352 drought events for Munich a marginal F1 score of 0.26 for Lisbon and 0.24 for Munich is given.

The best-performing models are additionally evaluated using the Heidke skill score (HSS). The range of the HSS is $-\infty$ to 1. Values below zero indicate that the random forecast (a forecast which randomly assigns the labels) has a better performance than the trained model. HSS of 1 indicates a perfect forecast. HSS is defined in the following way:

$$\text{HSS} = 2\frac{ad - bc}{(a+c)(c+d) + (a+b)(b+d)}, \quad (10)$$

where $a$ is the number of true positives, $b$ the number of false positives, $c$ number of false negatives and $d$ number of true negatives.

The second part of the study analyzes the best-performing algorithms (one for the Lisbon domain, one for the Munich domain) by applying explainable AI methods. SHAP (SHapley Additive exPlanations) is a state-of-the-art method for interpretation of machine learning models, which was inspired by game theory (Lundberg and Lee, 2017). It estimates for each input feature an average marginal contribution to the prediction of the result and therefore allows a comparison of the contributions among different features. In addition to that the difference in predictability among the seasons is calculated and compared to gain a better understanding on the influence of seasonal weather patterns.

An overview of the proposed methodology can be found in Fig. 3.

## 3 Results

This study consists of two parts: the first part deals with a systematical search for the best-performing setup of the ANN model for the two domains of interest: Munich and Lisbon. A repeated training is conducted by varying the values of parameters like the architecture of the hidden layers, L2 regularization and the loss function. In the second part of the analysis the best-performing models for the two domains are analyzed using explainable AI methods.

### 3.1 Model training results

For the design of the ANN it is crucial to perform fine-tuning of the model parameters to find the optimal setup. An architecture has to have enough layers and neurons to capture the complexity of the dataset (Goodfellow et al., 2016). In order to find the best architecture the learning curve of the algorithm is inspected. The learning curve shows the loss of the training and validation datasets on the weights during the training (Goodfellow et al., 2016). Two examples are shown in Fig. 4. The plot shown at the top refers to an architecture which is not able to capture the complexity of the dataset: the loss is hardly decreasing in the training or validation data. The bottom figure refers to an architecture which overfits: in the last epochs the loss of the validation dataset is rising, while it decreases in the training dataset.

In this way a network is searched which captures the given complexity of the dataset. This is reached with an algorithm consisting of at least five layers. Additionally two dropout layers, which set a specified number of nodes to zero in a random way, are introduced in order to fight overfitting.

#### 3.1.1 L2 regularization

L2 regularization is a broadly applied method to prevent overfitting in the training data (Bishop, 2007). The main idea behind regularization is to add a penalty term to the loss function, which will punish the classifier for complexity and force some of the weights to zero (Russell and Norvig, 2009). In case of L2 regularization the punishing term is proportional to the L2 norm of the weight vector. The weight of the punishing term $\lambda$ determines the relative importance of the regularization.

The results of the training with different values of $\lambda$ for L2 regularization are shown in Table 3. Training results are displayed in this particular case as the regularization is introduced to prevent overfitting. Generally the performance for the test dataset is more important and will be inspected in following experiments. If $\lambda$ is set to zero the regularization term vanishes. Especially in those cases the overfitting is high. For Lisbon overall higher performance could be seen for $\lambda$ values around 0.01, 0.001 or 0.0001. Models that are trained on the Munich dataset perform better with the $\lambda$ value of 0.001. Since the performance of the model with regards to the F1 score has a higher importance for an imbalanced dataset than the pure accuracy, the value of 0.001 is chosen for the following ANN model training.

#### 3.1.2 Loss function

As a next step the influence of the different loss functions on the model performance is investigated. Loss function is a function to evaluate how well a specific algorithm manages to fit the training data (Janocha and Czarnecki, 2016). It is an important part of the optimization function which has a direct influence on the updating of the weights of the ANN (Russell and Norvig, 2009). In addition to overall accuracy and F1 metric, the accuracies of the non-drought and drought classes in the test dataset are displayed. The results are shown in Table 4. Binary cross-entropy, mean absolute error and hinge

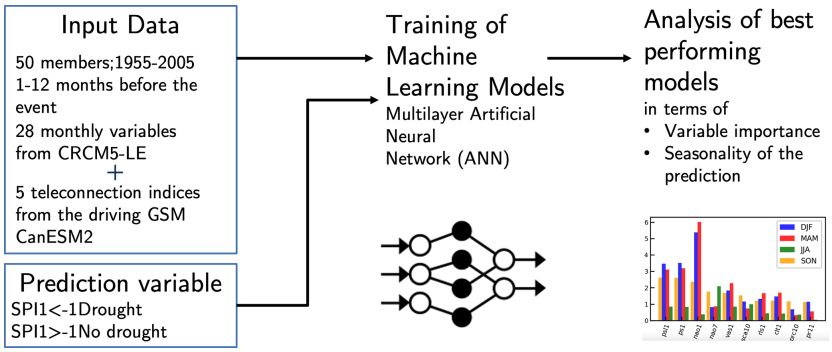


**Figure 3.** Overview of the proposed methodology.

**Table 3.** Results of ANN training for different values for λ for L2 regularization. λ of 0.001 (bold) is chosen for both domains for subsequent training since the performance of the model with regards to the F1 score has a higher importance for an imbalanced dataset than the pure accuracy.

| λ | Lisbon | | | | Munich | | | |
|---|---|---|---|---|---|---|---|---|
| | Train | | Test | | Train | | Test | |
| | Acc | F1 | Acc | F1 | Acc | F1 | Acc | F1 |
| 0 | 0.961 | 0.861 | 0.733 | 0.206 | 0.959 | 0.865 | 0.787 | 0.176 |
| 0.1 | 0.495 | 0.233 | 0.373 | 0.294 | 0.506 | 0.241 | 0.536 | 0.215 |
| 0.01 | 0.517 | 0.245 | 0.460 | 0.269 | 0.519 | 0.268 | 0.431 | 0.275 |
| **0.001** | **0.572** | **0.261** | **0.540** | **0.288** | **0.490** | **0.288** | **0.563** | **0.266** |
| 0.0001 | 0.765 | 0.472 | 0.627 | 0.259 | 0.823 | 0.557 | 0.719 | 0.189 |

loss functions show the best performance for the Munich domain. In contrast to that, for the Lisbon domain only the mean absolute error loss function has an accuracy of higher than 0.5. Also in the case of the Munich domain mean absolute error shows a higher performance with regards to the F1 score. Therefore mean absolute error is used for further analysis.

### 3.1.3 Model architecture

Lastly the models are trained on both domains using different architectures. Table 5 displays the model training results for the test dataset. The column "architecture" refers to the number of neurons in each dense (De) layer separated by the *-sign. For dropout (Dr) layers the fraction of weights which are randomly set to zero is given. The model architecture consists overall of seven layers. For example the architecture for the model in the first line of Table 5 is the following:

1. dense layer with 4000 neurons
2. dropout layer randomly setting 50 % of weights to zero
3. dense layer with 1000 neurons
4. dropout layer randomly setting 50 % of weights to zero
5. dense layer with 500 neurons
6. dense layer with 100 neurons
7. dense layer with 5 neurons.

We require the accuracy of both classes individually to be higher than 0.5 and search for an F1 score as high as possible. In the case of the Lisbon domain, three trained models satisfy the criterion of at least 50 % accuracy of each class: the model in the first, in the fourth and in the last row. The best performance in terms of F1 score is obtained for the last model with the following architecture: 5000*0.5*4000*0.5*1000*500*100. For the Munich domain only the first and the fourth models satisfy the criterion of at least 50 % accuracy for each class. For further analyses the first model is chosen as it shows the highest F1 score. The following model architecture is used for the Munich domain: 4000*0.5*1000*0.5*500*100*5. For the best-performing models HSS equals 0.06 for Lisbon and 0.04 for Munich. These results confirm that the obtained prediction is better than the one obtained by a random forecast and therefore does show a weak prediction skill. In the next step those models are analyzed using explainable AI methods.

**Table 4.** Performance of the model for different loss functions for the test dataset. Acc nd refers to the accuracy on the non-drought class and Acc d to the accuracy of the drought class. Mean absolute error (bold) is chosen for subsequent analysis since for Munich and Lisbon it shows an accuracy of at least 0.5 for both classes and a higher performance with regards to the F1 score.

| Loss function | Lisbon | | | | Munich | | | |
|---|---|---|---|---|---|---|---|---|
| | Acc nd | Acc d | Acc | F1 | Acc nd | Acc d | Acc | F1 |
| **Mean absolute error** | **0.511** | **0.516** | **0.540** | **0.288** | **0.500** | **0.582** | **0.512** | **0.276** |
| Mean squared error | 0.440 | 0.655 | 0.479 | 0.312 | 0.562 | 0.509 | 0.553 | 0.267 |
| Binary cross-entropy | 0.436 | 0.610 | 0.467 | 0.292 | 0.589 | 0.440 | 0.565 | 0.245 |
| Hinge | 0.229 | 0.753 | 0.323 | 0.287 | 0.568 | 0.486 | 0.555 | 0.259 |
| Squared hinge | 0.486 | 0.501 | 0.489 | 0.261 | 1.000 | 0.000 | 0.840 | 0.000 |

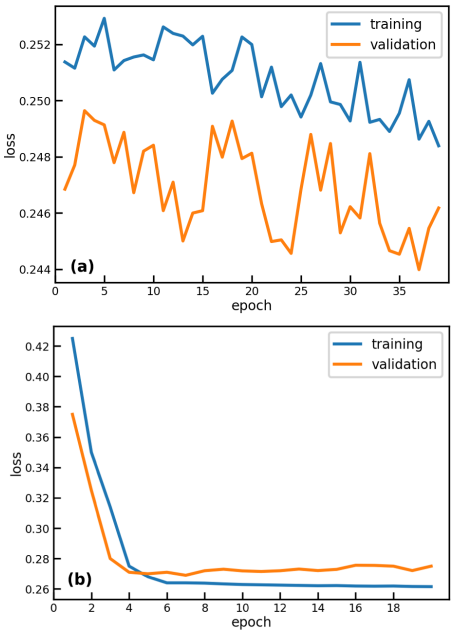


**Figure 4.** Learning curve for two chosen fitting examples: algorithm complexity insufficient **(a)** and overfitting **(b)**.

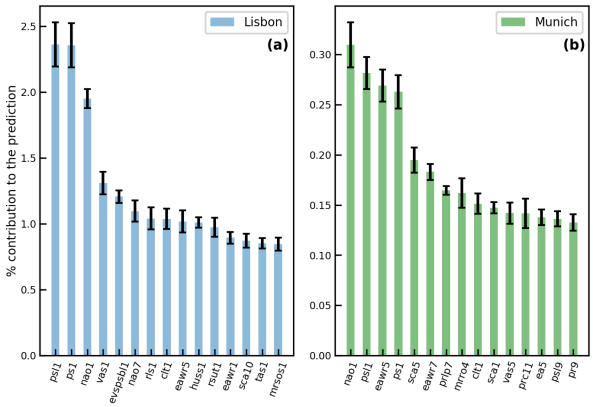


**Figure 5.** Mean Shapely values normalized to the contribution to the prediction for the top 15 variables, with the highest importance for Lisbon **(a)** and Munich **(b)** in the test dataset. The number behind the variable name refers to the number of months before the event (NAO1–NAO value 1 month before the predicted event). The results indicate that for the Lisbon domain *psl1* and *ps1* are the most influential drought predictors; for Munich this is NAO1.

## 3.2 Explainable AI methods for the analysis of the best-performing algorithms

### 3.2.1 Shapely values

For the Munich and Lisbon domain Shapely values are calculated using the results of the best-performing models for the test dataset. For the calculation each of the 12 months used as input to the predicting algorithm for each variable is considered individually, resulting in 28 atmospheric variables $\times 12 + 6$ teleconnection indices $\times 12 = 408$ variables. The number behind the variable name refers to the number of months before the event (NAO1–NAO value 1 month before the predicted event). The results are shown in Fig. 5. Since the calculation of Shapely values is computationally expensive, they are calculated five times on a subset of 500 data points. The error bars displayed in black on the plot indicate that the uncertainties are smaller than the nominal values of the variable contributions. The nominal Shapely values are normed and recalculated to a percentage of contribution to the prediction; e.g., the NAO1 value explains roughly 2.3 % of the prediction for the Lisbon domain.

We see that for both domains the contribution to the prediction is broadly distributed among the many input variables. Between Lisbon and Munich, Shapely values show a distinct difference in the nominal values of the feature contributions: values for Lisbon are about 6 times higher than those for Munich (e.g., the contribution of NAO1 for Munich is around 0.3 % and for Lisbon around 1.9 %).

For the Lisbon domain, the variables with a higher-impact are sea level pressure (*psl*), surface pressure (*ps*) and NAO 1 month before the event. The first two variables are strongly autocorrelated for the Lisbon domain due to its location at

**Table 5.** Performance of models for the Lisbon and Munich domains for different variations in architecture on the test dataset. Acc nd refers to the accuracy of the non-drought class and Acc d to the accuracy of the drought class. The sixth model architecture is chosen for subsequent analysis for Lisbon and first for Munich (bold) due to an accuracy of at least 0.5 of both classes and a higher performance with regards to the F1 score.

| Neurons | Architecture | Lisbon | | | | Munich | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Acc nd | Acc d | Acc | F1 | Acc nd | Acc d | Acc | F1 |
| De*Dr*De*Dr*De*De*De | 4000*0.5*1000*0.5*500*100*5 | 0.511 | 0.516 | 0.540 | 0.288 | **0.562** | **0.509** | **0.553** | **0.267** |
| De*Dr*De*Dr*De*De*De | 5000*0.5*1000*0.5*500*100*5 | 0.581 | 0.496 | 0.566 | 0.292 | 0.378 | 0.693 | 0.428 | 0.279 |
| De*Dr*De*Dr*De*De*De | 5000*0.5*4000*0.5*500*100*5 | 0.457 | 0.602 | 0.483 | 0.296 | 0.725 | 0.338 | 0.663 | 0.243 |
| De*Dr*De*Dr*De*De*De | 5000*0.5*4000*0.5*1000*100*5 | 0.570 | 0.501 | 0.558 | 0.290 | 0.527 | 0.514 | 0.525 | 0.257 |
| De*Dr*De*Dr*De*De*De | 5000*0.5*4000*0.5*1000*500*5 | 0.402 | 0.635 | 0.444 | 0.292 | 0.683 | 0.409 | 0.640 | 0.266 |
| De*Dr*De*Dr*De*De*De | 5000*0.5*4000*0.5*1000*500*100 | **0.575** | **0.526** | **0.566** | **0.305** | 0.420 | 0.619 | 0.452 | 0.266 |

the sea. The strong influence of *ps* and *psl* and NAO shows the influence of the atmospheric pressure system on drought formation in Lisbon. It is also striking that the influence of the local pressure seems to be higher than the influence of NAO. The next two variables for the Lisbon domain with the strongest contribution to the prediction are northward near-surface wind (*vas*) and evaporation (*evspsbl*). The latter variable has a very direct influence on the formation of drought given that if evaporation is getting lower, the probability of formation of rain clouds also decreases (Sheffield and Wood, 2011). The contribution of *vas* to drought formation in Lisbon needs to be further studied. For the Munich domain the highest influence is found for NAO1, *psl1*, EAWR5 and *ps1*. The results indicate that NAO is the most influential drought predictor for Munich. Additionally the contribution of EAWR5 and SCA5 on the Munich domain cannot be neglected as they are found within the top five predictors. A further investigation of this relationship is of interest for the understanding of drought formation in Munich.

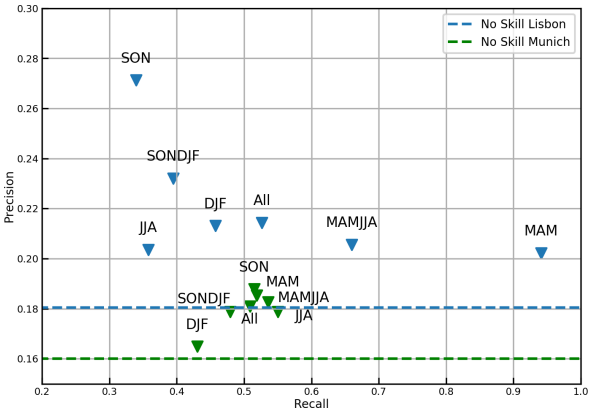


**Figure 6.** The effect of seasonality on precision and recall for Lisbon (blue) and Munich (green). The results indicate that for the Munich and Lisbon domain better drought predictability is possible in spring, fall and summer.

### 3.2.2 Seasonality

In order to evaluate the influence of seasonality on the prediction the performance of the model is calculated separately for the four seasons. Since the distribution between the drought and non-drought classes is different among the seasons (e.g., range of 17 % to 19 % of drought events for the Lisbon domain) a rescaling of the number of drought and non-drought events is performed to ensure comparability among the results. To compare the performance a precision recall plot is used (Saito and Rehmsmeier, 2015). Recall and precision are calculated for each of the four seasons (MAM, JJA, SON, DJF) and for the 2 half-years (MAMJJA and SONDJF) using the estimated scaling factors. Results of the calculation are shown in Fig. 6. The dotted line marks the line under which the classifier shows no skill. The line is defined as a proportion of drought events against overall number of events (Saito and Rehmsmeier, 2015). For the Lisbon domain it becomes evident that the model performance is very different across seasons: higher precision of around 0.23 can be found during the winter half-year. However for the spring season and summer half-year the recall rises, while precision goes down. For the Munich classifier the results for the different seasons are closer together in terms of recall. It shows a worse performance for the winter months (DJF), while fall, spring and summer show an overall better model performance. This is an indication that for the Munich domain, better drought predictability is possible in spring, fall and summer.

An additional analysis is conducted to calculate the Shapely values separately for the four season and the two domains in order to understand the influence of the different variables on the prediction. The results of the analysis can be seen in Figs. 7 and 8. The results for the Lisbon domain show that NAO1 is the strongest predictor in winter and spring, while the contribution of pressure to drought predictability is higher in fall, followed by NAO1. In contrast, for the summer season NAO1 is not among the top 10 predictors but rather other teleconnection indices like EAWR5, NAO7 and SCA7. Those teleconnection indices originate from winter months, when NAO was shown to have the highest impact on the pre-

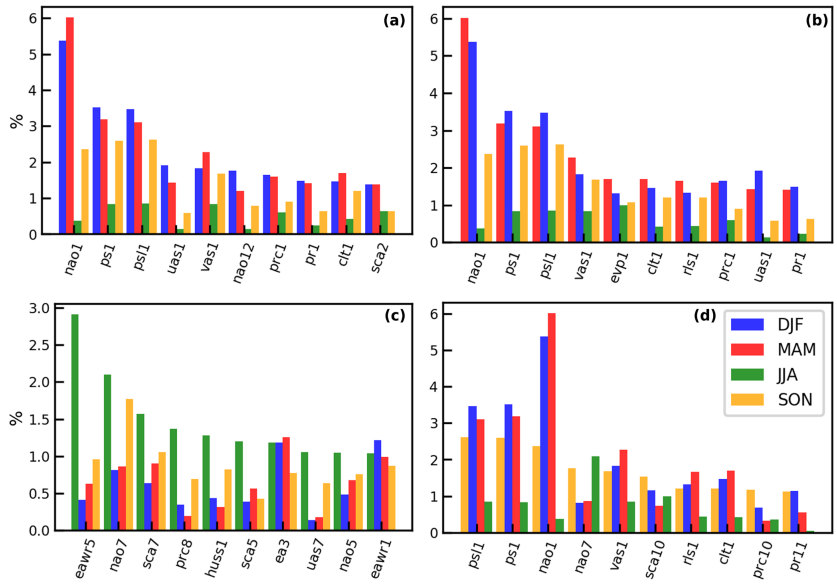


**Figure 7.** Shapely values for Lisbon calculated separately for the four seasons and sorted by the maximum contribution in DJF **(a)**, MAM **(b)**, JJA **(c)** and SON **(d)** for the test dataset; evspsbl abbreviated as evp.

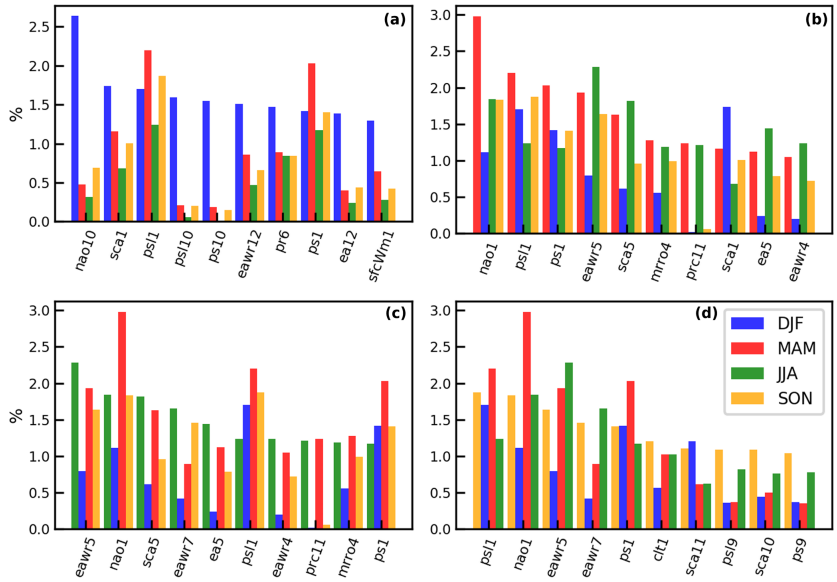


**Figure 8.** Shapely values for Munich calculated separately for the four seasons and sorted by the maximum contribution in DJF **(a)**, MAM **(b)**, JJA **(c)** and SON **(d)** for the test dataset; sfcWindmax abbreviated as sfcWm.

diction. However, given the low performance of the model in the summer season, further investigation is needed. For the Munich domain NAO1 has one of the highest contributions for spring, summer and fall, while it cannot be found among the strongest predictors for winter. EAWR5 is one of the strongest predictors for summer, spring and fall. The feature contributions for predictions in the winter season in Munich indicate that atmospheric variables 10 or 12 months before the event might be drought indicators.

## 4 Discussion and conclusion

Drought is a multiscale phenomenon, and its formation and evolution are different for every climatology and season. In this study, we (i) explored the possibilities of using the data provided by CRCM5-LE to predict droughts using ANNs and (ii) applied explainable AI methods to gain a better understanding of the results. A drought event is defined as an SPI1 less than $-1$ at the given site. The first half of the study deals with the systematic search for the best-performing models. For the Lisbon domain the best results are obtained by the model with L2 regularization of 0.001; mean absolute error as a loss function; and the architecture 5000*0.5*4000*0.5*1000*500*100, where five layers are fully connected, and two layers are dropout layers. For the Munich domain, the best results are obtained by the model with L2 regularization of 0.001; mean absolute error as a loss function; and the architecture 4000*0.5*1000*0.5*500*100*5, where five layers are fully connected, and two layers are dropout layers. The best-performing models obtain accuracies of 57 % for the Lisbon domain and 55 % for the Munich domain.

The precision of the prediction in both cases is rather moderate as a high percentage of data are misclassified. For Lisbon, classifier precision remains at around 22 %. This means that one out of four predicted drought events is an actual drought. For the Munich case, this ratio is even lower and amounts to 18 %. However, the models provide an important basis for the development of future drought-predicting models and offer a fruitful ground for the investigation of influence of single input variables during different seasons on drought formation.

Compared to the study by Santos et al. (2014), who investigated drought predictability in Portugal, the weak prediction accuracies of our study are not surprising. In Santos et al. (2014), SPI6 for April, May and June is predicted; however precipitation amounts for the months until March were also given as input. As SPI6 is calculated using the sum of 6 months precipitation, the model receives over half of the information it needs for the calculation of the value. As no similar studies exist for the Munich domain, no comparison can be performed.

The second half of the study concentrates on the analysis of the obtained algorithms using explainable AI methods. Among the strongest predictors for the domains are NAO, *psl* and *ps* 1 month before the event. This underlines the importance of the atmospheric system on the drought formation. For the model trained for the Lisbon domain, the variables of northward near-surface wind (*vas*) and evaporation (*evspsbl*) followed. For the Munich domain, EAWR and SCA 5 months before the event are found among the strongest predictors. In general the percentages of the contribution of the strongest predictors for the Munich domain are around 6 times lower than those for the Lisbon domain.

This study indicates that seasonality is a crucial factor for drought predictions. Precision and recall of the prediction are lower in summer for the Lisbon domain and in winter for the Munich domain. Moreover, while for the Munich domain the spread of precision and recall across the seasons is rather low, huge differences are found for the Lisbon domain: the trained model obtained higher recall and lower precision for spring and higher precision and lower recall for fall when comparing to the baseline of all data. The results show that for the Lisbon domain, NAO1 is the strongest predictor in winter and spring, while the contribution of pressure to drought predictability is higher in fall, followed by the contribution of NAO1. For the Munich domain, NAO1 is found to have one of the highest contributions for spring, summer and fall, while it could not be found among the 10 strongest predictors for winter.

Further investigations are of interest for scientific research on both objectives. In terms of drought prediction, further research is possible within the same setting. The field of AI is evolving rapidly, showing new algorithms, methods and frameworks, such that there is a high potential for finding better-suited algorithms (Hao, 2019). One of the main limitations of this study remains that an application of the obtained framework on observation data is not possible due to the fact that observational data lack a multitude of variables which are used as input in this study, e.g., heat fluxes and radiation. However the results obtained by Shapely value calculation are of high importance for the choice of variables for the development of a future model which potentially could be applied to observational data. Given the high Shapely importance of NAO for drought prediction, other large-scale variables, such as atmospheric blocking, can be added to the input variables. Moreover, the application to new domains is of interest to investigate the regionality of drought prediction possibilities. Explainable AI methods offer an important approach to improve the current limitations of machine learning models; their application is of high importance in the field of physical geography since it enables a physical interpretation of statistical results to be provided.

*Data availability.* Ensemble model data used in this study may be retrieved from the following sources: CanESM2-LE data are available via https://open.canada.ca/data/en/dataset/aa7b6823-fd1e-49ff-a6fb-68076a4a477c (Environment and Climate Change Canada, 2020). CRCM5-LE data can be retrieved at https://climex-data.srv.lrz.de/Public/ (ClimEx project, 2020).

*Author contributions.* This study was conceptualized by EF under the supervision of RL. Formal analysis, visualization of results and writing of the original draft were performed by EF. All authors contributed to the interpretation of the findings and revision of the paper.

*Competing interests.* The contact author has declared that neither they nor their co-author has any competing interests.

*Disclaimer.* Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

*Special issue statement.* This article is part of the special issue "Recent advances in drought and water scarcity monitoring, modeling, and forecasting (EGU2019, session HS4.1.1/NH1.31)". It is a result of the European Geosciences Union General Assembly 2019, Vienna, Austria, 7–12 April 2019.

*Acknowledgements.* The CRCM5-LE was created within the ClimEx project, which was funded by the Bavarian State Ministry for the Environment and Consumer Protection. Computations of the CRCM5-LE were made on the SuperMUC supercomputer at the Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities. We acknowledge Environment and Climate Change Canada for providing the CanESM2-LE driving data.

*Review statement.* This paper was edited by Athanasios Loukas and reviewed by three anonymous referees.

## References

Barnston, A. G. and Livezey, R. E.: Classification, Seasonality and Persistence of Low-Frequency Atmospheric Circulation Patterns, Mon. Weather Rev., 115, 1083–1126, https://doi.org/10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2, 1987.

Belayneh, A., Adamowski, J., Khalil, B., and Quilty, J.: Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction, Atmos. Res., 172–173, 37–47, https://doi.org/10.1016/j.atmosres.2015.12.017, 2016.

Biesiada, J. and Duch, W.: Feature Selection for High-Dimensional Data – A Pearson Redundancy Based Filter, in: Computer Recognition Systems 2, edited by: Kurzynski, M., Puchala, E., Wozniak, M., and Zolnierek, A., Springer, Berlin, Heidelberg, 242–249, 2007.

Bishop, C. M.: Pattern Recognition and Machine Learning (Information Science and Statistics), 1st edn., Springer, Berlin, Heidelberg, 2007.

Bonaccorso, B., Cancelliere, A., and Rossi, G.: Probabilistic forecasting of drought class transitions in Sicily (Italy) using Standardized Precipitation Index and North Atlantic Oscillation Index, J. Hydrol., 526, 136–150, https://doi.org/10.1016/j.jhydrol.2015.01.070, 2015.

Bueh, C. and Nakamura, H.: Scandinavian pattern and its climatic impact, Q. J. Roy. Meteor. Soc., 133, 2117–2131, https://doi.org/10.1002/qj.173, 2007.

Ceglar, A., Zampieri, M., Toreti, A., and Dentener, F.: Observed Northward Migration of Agro-Climate Zones in Europe Will Further Accelerate Under Climate Change, Earths Future, 7, 1088–1101, https://doi.org/10.1029/2019EF001178, 2019.

Chollet, F., et al.: Keras, available at: https://keras.io (last access: 29 November 2021), 2015.

Clevert, D. A., Unterthiner, T., and Hochreiter, S.: Fast and accurate deep network learning by exponential linear units (ELUs), arXiv [preprint], arXiv:1511.07289, 23 November 2015.

ClimEx project: CRCM5-LE, available at: https://climex-data.srv.lrz.de/Public/ (last access: 30 November 2021), 2020.

Dawson, A.: eofs: A Library for EOF Analysis of Meteorological, Oceanographic, and Climate Data, Journal of Open Research Software, 4, e14, https://doi.org/10.5334/jors.122, 2016.

Deo, R. C., Kisi, O., and Singh, V. P.: Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model, Atmos. Res., 184, 149–175, https://doi.org/10.1016/j.atmosres.2016.10.004, 2017.

Enfield, D. B., Mestas-Nuñez, A. M., and Trimble, P. J.: The Atlantic Multidecadal Oscillation and its relation to rainfall and river flows in the continental U.S., Geophys. Res. Lett., 28, 2077–2080, https://doi.org/10.1029/2000GL012745, 2001.

Environment and Climate Change Canada: The Canadian Earth System Model Large Ensembles, available at: https://open.canada.ca/data/en/dataset/aa7b6823-fd1e-49ff-a6fb-68076a4a477c (last access: 29 November 2021), 2020.

European Centre for Medium-Range Weather Forecasts: ERA Interim, Daily, available at: https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/ (last access: 29 November 2021), 2020.

Federal Ministry of Food and Agriculture: Trockenheit und Dürre 2018 – Überblick über Maßnahmen, available at: https://www.bmel.de/DE/Landwirtschaft/Nachhaltige-Landnutzung/Klimawandel/_Texte/Extremwetterlagen-Zustaendigkeiten.html (last access: 25 July 2019), 2018.

Folland, C. K., Knight, J., Linderholm, H. W., Fereday, D., Ineson, S., and Hurrell, J. W.: The Summer North Atlantic Oscillation: Past, Present, and Future, J. Climate, 22, 1082–1103, https://doi.org/10.1175/2008JCLI2459.1, 2009.

Goodfellow, I., Bengio, Y., and Courville, A.: Deep Learning, The MIT Press, Cambridge, MA, USA, 2016.

Guyon, I. and Elisseeff, A.: An Introduction to Variable and Feature Selection, J. Mach. Learn. Res., 3, 1157–1182, 2003.

Hao, K.: We analyzed 16,625 papers to figure out where AI is headed next, available at: https://www.technologyreview.com/2019/01/25/1436/we-analyzed-16625-papers-to-figure-out-where-ai-is-headed-next/ (last access: 21 March 2021), 2019.

Hao, Z., Singh, V. P., and Xia, Y.: Seasonal Drought Prediction: Advances, Challenges, and Future Prospects, Rev. Geophys., 56, 108–141, https://doi.org/10.1002/2016RG000549, 2018.

Hurrell, J. W., Kushnir, Y., Ottersen, G., and Visbeck, M. (Eds.): The North Atlantic Oscillation: Climatic Significance and Environmental Impact, American Geophysical Union, Washington, D.C., 2003.

IPCC: Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the

IPCC, Assessment report (Intergovernmental Panel on Climate Change), Working Group, Cambridge University Press, Cambridge, England, 2013.

Janocha, K. and Czarnecki W. M. Wojciech: On Loss Functions for Deep Neural Networks in Classification, Schedae Informaticae, 25, 49–59, https://doi.org/10.4467/20838476SI.16.004.6185, 2016.

Keyantash, J. and National Center for Atmospheric Research Staff: The Climate Data Guide: Standardized Precipitation Index (SPI), available at: https://climatedataguide.ucar.edu/climate-data/standardized-precipitation-index-spi (last access: 21 March 2021), 2018.

Kington, J. A.: Daily weather mapping from 1781, Climatic Change, 3, 7–36, https://doi.org/10.1007/bf02423166, 1980.

Kirchmeier-Young, M., Zwiers, F., and Gillett, N.: Attribution of Extreme Events in Arctic Sea Ice Extent, J. Climate, 30, 553–571, https://doi.org/10.1175/JCLI-D-16-0412.1, 2016.

Kumar, V. and Minz, S.: Feature selection: a literature review, SmartCR, 4, 211–229, 2014.

Kushner, P. J., Mudryk, L. R., Merryfield, W., Ambadan, J. T., Berg, A., Bichet, A., Brown, R., Derksen, C., Déry, S. J., Dirkson, A., Flato, G., Fletcher, C. G., Fyfe, J. C., Gillett, N., Haas, C., Howell, S., Laliberté, F., McCusker, K., Sigmond, M., Sospedra-Alfonso, R., Tandon, N. F., Thackeray, C., Tremblay, B., and Zwiers, F. W.: Canadian snow and sea ice: assessment of snow, sea ice, and related climate processes in Canada's Earth system model and climate-prediction system, The Cryosphere, 12, 1137–1156, https://doi.org/10.5194/tc-12-1137-2018, 2018.

Leduc, M., Mailhot, A., Frigon, A., Martel, J.-L., Ludwig, R., Brietzke, G. B., Giguère, M., Brissette, F., Turcotte, R., Braun, M., and Scinocca, J.: The ClimEx Project: A 50-Member Ensemble of Climate Change Projections at 12-km Resolution over Europe and Northeastern North America with the Canadian Regional Climate Model (CRCM5), J. Appl. Meteorol. Clim., 58, 663–693, https://doi.org/10.1175/JAMC-D-18-0021.1, 2019.

Lim, Y.-K.: The East Atlantic/West Russia (EA/WR) teleconnection in the North Atlantic: climate impact and relation to Rossby wave propagation, Clim. Dynam., 44, 3211–3222, https://doi.org/10.1007/s00382-014-2381-4, 2015.

Lundberg, S. and Lee, S.: A unified approach to interpreting model predictions, CoRR, arXiv [preprint], arXiv:1705.07874, 25 November 2017.

Maas, A. L.: Rectifier Nonlinearities Improve Neural Network Acoustic Models, in: Proceedings of the 30th International Conference on Machine Learning, 16–21 Juni 2013, Atlanta, Georgia, USA, 2013.

Martynov, A., Laprise, R., Sushama, L., Winger, K., Šeparović, L., and Dugas, B.: Reanalysis-driven climate simulation over CORDEX North America domain using the Canadian Regional Climate Model, version 5: model performance evaluation, Clim. Dynam., 41, 2973–3005, https://doi.org/10.1007/s00382-013-1778-9, 2013.

McGovern, A., Elmore, K. L., Gagne, D. J., Haupt, S. E., Karstens, C. D., Lagerquist, R., Smith, T., and Williams, J. K.: Using Artificial Intelligence to Improve Real-Time Decision-Making for High-Impact Weather, B. Am. Meteorol. Soc., 98, 2073–2090, https://doi.org/10.1175/BAMS-D-16-0123.1, 2017.

McKee, T., Doesken, N., and Kleist, J.: The relationship of drought frequency and duration to time scales, in: Proceedings of the 8th Conference on Applied Climatology, American Meteorological Society Boston, 17–22 January 1993, Anaheim, CA, USA, 1993.

Mikhailova, N. and Yurovsky, A.: The East Atlantic Oscillation: Mechanism and Impact on the European Climate in Winter, Phys. Oceanogr., 4, 27–36, https://doi.org/10.22449/1573-160X-2016-4-25-33, 2016.

Morid, S., Smakhtin, V., and Bagherzadeh, K.: Drought forecasting using artificial neural networks and time series of drought indices, Int. J. Climatol., 27, 2103–2111, https://doi.org/10.1002/joc.1498, 2007.

Russell, S. and Norvig, P.: Artificial Intelligence: A Modern Approach, 3rd edn., Prentice Hall Press, Upper Saddle River, NJ, USA, 2009.

Saito, T. and Rehmsmeier, M.: The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets, PLoS One, 10, 1–21, https://doi.org/10.1371/journal.pone.0118432, 2015.

Santos, J. F., Portela, M. M., and Pulido-Calvo, I.: Spring drought prediction based on winter NAO and global SST in Portugal, Hydrol. Process., 28, 1009–1024, https://doi.org/10.1002/hyp.9641, 2014.

Sasaki, Y.: The truth about of the F-measure, available at: https://www.cs.odu.edu/~mukka/cs795sum09dm/Lecturenotes/Day3/F-measure-YS-26Oct07.pdf (last access: 30 November 2021), 2007.

Sheffield, J. and Wood, E. F.: Drought: past problems and future scenarios, Earthscan, London, Washington, DC, 2011.

Sheffield, J., Andreadis, K. M., Wood, E. F., and Lettenmaier, D. P.: Global and Continental Drought in the Second Half of the Twentieth Century: Severity–Area–Duration Analysis and Temporal Variability of Large-Scale Events, J. Climate, 22, 1962–1981, https://doi.org/10.1175/2008JCLI2722.1, 2009.

Spinoni, J., Naumann, G., Vogt, J., and Barbosa, P.: Meteorological Drought in Europe: Events and Impacts: Past Trends and Future Projections, Publications Office of the European Union, Luxembourg, 2016.

Spinoni, J., Naumann, G., and Vogt, J. V.: Pan-European seasonal trends and recent changes of drought frequency and severity, Global Planet. Change, 148, 113–130, https://doi.org/10.1016/j.gloplacha.2016.11.013, 2017.

Trenberth, K.: Changes in Precipitation with Climate Change, Clim. Res., 47, 123–138, https://doi.org/10.3354/cr00953, 2011.

von Trentini, F., Aalbers, E. E., Fischer, E. M., and Ludwig, R.: Comparing interannual variability in three regional single-model initial-condition large ensembles (SMILEs) over Europe, Earth Syst. Dynam., 11, 1013–1031, https://doi.org/10.5194/esd-11-1013-2020, 2020.

World Meteorological Organization: Standardized Precipitation Index User Guide, WMO, Geneva, Switzerland, 2012.

Yoon, J.-H., Mo, K., and Wood, E. F.: Dynamic-model-based seasonal prediction of meteorological drought over the contiguous United States, J. Hydrometeorol., 13, 463–482, 2012.

Zargar, A., Sadiq, R., Naser, B., and Khan, F. I.: A review of drought indices, Environ. Rev., 19, 333–349, https://doi.org/10.1139/a11-013, 2011.

## 4.2 Paper II: Inter-seasonal connection of typical European heatwave patterns to soil moisture

**Reference:** Felsche, E., Böhnisch, A., Ludwig, R. (2023). Inter-seasonal connection of typical European heatwave patterns to soil moisture. npj Climate and Atmospheric Science, 6(1), 1.

**Plain language summary:** Prolonged heat periods have become a recurring feature of the European climate. Recent events like the 2003 heatwave in France, the 2010 Russian heatwave, and the 2019 European heatwave have caused considerable economic losses due to crop failure, imposed substantial stress on the health system, and caused thousands of heat-related deaths. Due to climate change, an increase in length and frequency of heatwaves has been observed since 1950 in most regions worldwide. However, until now, little knowledge is available on the generalised patterns of heatwaves since most studies focus on the analysis of single historical heatwave events. Increased knowledge will help to improve heatwave and drought prediction and mitigation. This study uses hierarchical agglomerative clustering to derive nine dominating spatial heatwave patterns from a 50-member regional climate model (Canadian Regional Climate Model version 5, CRCM5-LE). The heatwave patterns correspond well with clusters derived from an observational data set (E-OBS) and with extreme historical heatwave events. Moreover, we analyse the occurrence of heatwaves in the identified spatial patterns regarding a soil moisture deficit present before and after the event. We show that negative soil moisture anomalies in the preceding winter/spring (JFMA) can serve as a predictor for heatwaves in South Europe. For North Europe, we find a negative correlation between the number of heatwave days in summer and autumn (OND) soil moisture content.

**Author's contribution:** E.F. designed the research under the supervision of R.L. E.F . and A.B. performed the data analysis and led the writing of the paper. E.F. prepared all figures. All authors contributed to the interpretation of the findings and revision of the paper.

**Scope of the journal:** "npj Climate and Atmospheric Science is an open access journal covering the relevant physical, chemical and biological components of atmospheric and climate science. [...] Topics covered by the journal include climate dynamics, climate variability, weather and climate prediction, climate change, ocean dynamics, weather extremes, air pollution, atmospheric chemistry including aerosols, the hydrological cycle and atmosphere–ocean and atmosphere–land interactions. " (Nature, 2024)

**Status:** published

**Journal:** npj Climate and Atmospheric Science

**Impact Factor (2-Year):** 9.45

npj | climate and atmospheric science

www.nature.com/npjclimatsci

ARTICLE OPEN

Check for updates

# Inter-seasonal connection of typical European heatwave patterns to soil moisture

Elizaveta Felsche[1,2,3]✉, Andrea Böhnisch[2] and Ralf Ludwig[2]

Although prolonged heat periods have become a recurring feature of European climate, little knowledge is available on dominant spatial patterns of heatwaves and their influence on moisture-related processes. Increased knowledge will help to improve heatwave and drought prediction and mitigation. This study uses hierarchical agglomerative clustering to derive nine dominating spatial heatwave patterns from a 50-member regional climate model (Canadian Regional Climate Model version 5, CRCM5-LE). The heatwave patterns correspond well with clusters derived from an observational data set (E-OBS) and with extreme historical heatwave events. Moreover, we analyse the occurrence of heatwaves in the identified spatial patterns regarding a soil moisture deficit present before and after the event. We show that negative soil moisture anomalies in the preceding winter/spring (JFMA) can serve as a predictor for heatwaves in South Europe. For North Europe, we find a negative correlation between the number of heatwave days in summer and autumn (OND) soil moisture content.

*npj Climate and Atmospheric Science* (2023)6:1 ; https://doi.org/10.1038/s41612-023-00330-5

## INTRODUCTION

Heatwaves and droughts substantially impact human mortality, economic well-being, infrastructure, and natural ecosystems[1–3]. For example, the 2003 heatwave in Europe is estimated to have caused more than 70,000 deaths[4]. Globally, 2% of working hours are lost due to too-hot conditions[5]. Droughts that often accompany a heatwave have been estimated to cause losses of USD 621 Million on average per event between 1950 and 2014 in Europe[6]. The 2010 heatwave in Russia caused USD 15 Billion (1% gross domestic product) in total economic losses[7]. Since 1950, most regions worldwide have observed a significant increase in the number of heatwave days, maximum duration, and cumulative heat[8]. Climate model projections estimate that the described trends will continue throughout the 21st century[9,10]. There is a high need for operational seasonal forecasts of heatwaves and droughts to mitigate their impacts, e.g., to introduce measures for water saving or prepare navigation infrastructure for low flows[11,12]. Current forecasts offer limited predictive capabilities, underlining the importance of future studies to increase understanding of mechanisms causing the events[13–15]. Identifying heatwave patterns allows a meaningful way of dimensionality reduction, which is important for further research on driving physical mechanisms for heatwave occurrence.

Current research highlights that heatwaves and droughts are highly interrelated and caused by similar persistent large-scale atmospheric circulation patterns[16–18]. Moreover, the self-intensifying nature of extreme droughts and heatwaves has been suggested as central to their evolution[19–21]. There is a two-fold relationship. On the one hand, soil and vegetation dry with the occurrence of a heatwave, leading to reduced evaporation. Therefore, the likelihood of rainfall decreases, favouring the formation of drought[20,22]. On the other hand, evaporation decreases with the onset of drought. The reduced cloud cover leads to a larger fraction of solar radiation reaching the land surface, increasing the likelihood of heatwave formation[20,22]. Global-warming-induced changes in thermodynamic conditions account for 57.3% of Europe's increase in extreme heat occurrence[23].

The influence of precipitation and soil moisture anomalies on heatwave formation has been studied in different European regions. A rainfall deficit in the Mediterranean in spring is found to favour the formation of heatwaves in Northern Europe as the rainfall deficit propagates northward throughout the summer[16,24]. A recent study[17] confirmed that dry conditions in winter/spring seasons prevail prior to hot summers over Southern Europe. Other studies confirm that anomalously dry Western and Northern European summers significantly correlate with the occurrence of heatwaves in those regions[25]. Soil moisture and other precipitation-related indices correlate with the temperature extremes in South-Eastern Europe[26,27]. For the European heatwave of July 2019, land-atmosphere feedback and influences of northward propagation of dryness contributed to the exceptional intensity of the event[28].

Most research on heatwaves investigates historical events based on observational data[17,24,26,29], which rarely happen by definition. Moreover, only few studies have analysed generalised patterns of heatwaves to derive scientific findings applicable to coherent regions[24,29] instead of focusing on the causes of single events[30,31]. By 'coherent regions', we here and thereafter mean regions connected to similar atmospheric circulation patterns, such that the heatwaves occur simultaneously and over the same geographical region. Large climate model ensembles have proven their usefulness for the investigation of extreme events both in terms of extreme cold and wet, as well as in terms of hot and dry events[32–36]. They allow the assessment of the natural variability of extreme weather events and therefore facilitate to derive of statistically reliable findings. Moreover, regional climate models offer a finer spatial resolution, which allows for the resolution of finer spatiotemporal processes and therefore obtains spatial patterns on a regional and subregional level when compared to Global Climate Models[37,38].

[1]Technical University of Munich, Munich, Germany. [2]Department of Geography, Ludwig-Maximilans-Universität München, Munich, Germany. [3]Center for Digital Technology and Management, Munich, Germany. ✉email: felsche@cdtm.de

In this study, we want to take a regional approach to the investigation of heatwaves, as those usually cover a fraction of the continent. Therefore, we aim to find stable spatial patterns of heatwaves using the 50 members of a Single Model Initial-condition Large Ensemble (SMILE) CRCM5-LE over Europe. Canadian Earth System Model 2-Large Ensemble (CanESM2-LE) during the period 1950-2099 is used to derive the boundary conditions for the Canadian Regional Climate Model version 5-Large Ensemble (CRCM5-LE)[39]. CRCM5-LE obtains more realistic representations of climate over complex topologies, as in the southwest part of Scandinavia, the Iberian Peninsula, the Alps and the Pyrenees[39]. A study by Trentini et al.[40] confirms the applicability of the chosen model for heatwave research in the European domain. It investigates the interannual variability of three different large ensembles, with CRCM5-LE being one of them, and compares it to E-OBS data. The study shows that CRCM5-LE has a good representation of JJA temperature and the number of heatwave days per year. Another study[41] compares CRCM5-LE with the EURO-CORDEX ensemble and confirms the added value of the ensemble for the European domain. Moreover, CRCM5-LE was already used in a multitude of studies on European extreme events[32,42,43].

We use 1500 model years that correspond to the years 1981-2010 historical climate. A 'heatwave day' occurs when the local daily maximum temperature exceeds the 95th JJA percentile of the whole period. We use the three-day-running mean in order to obtain robust signals. In total, we obtain more than 50,000 heatwave days. For an exact definition of a heatwave day, see the section 'Methods'. Following previous studies on heatwave classification[24,29], in the first step, we apply hierarchical agglomerative clustering on the heatwave days in order to identify predominant heatwave patterns. The clustering algorithm starts by assigning each data point to its own cluster (agglomeration). Then it merges all the clusters using a defined similarity measure and builds a hierarchy between clusters, which is based on how similar they are to one another[44]. We use cosine similarity as our similarity measure. The optimal number of clusters is determined using the elbow method. It picks the number where the added information by creating one more cluster sharply drops[45]. This point is determined by calculating the knee of the curve. For a detailed description, see 'Methods'. Subsequently, the obtained spatial patterns of heatwaves are analysed in terms of the influence of soil moisture and precipitation conditions in spring and summer on heatwave formation and the influence of heatwave occurrences on dry conditions in the following fall/winter.

## RESULTS

### Typical European heatwave patterns

We focus our investigations on the European domain of the CRCM5-LE[39], as we are interested in regional heatwaves. We obtain a total of nine significant spatial patterns from CRCM5-LE for the years 1981–2010. Figure 1 shows the identified spatial patterns, which we order from West to East: Iberian Peninsula (IP), Western Europe 2 (WE2), Western Europe 1 (WE1), Britain and Ireland (BI), South-Eastern Europe (SEE), Greece and South Italy (GSI), Scandinavia (SCA), Central-Eastern Europe (CEE) and North-Eastern Europe (NEE).

The pattern significance is assessed via bootstrapping, which we apply according to the existing literature on heatwave clustering[24,29]. For bootstrapping, we divide the data set into a validation and training set 100 times so that one-hundredth of the data is assigned to the validation set and the rest to the training data set. We perform clustering using the training data set and then assign clustering classes to the validation data according to the nearest distance to data points within the training data set. The obtained labels are compared to the ones originating from clustering the whole data set. A stability score is calculated for each cluster. It corresponds to the number of correctly assigned events vs the total number of validation events per spatial pattern. The stability scores are compared to the ones from a Monte-Carlo pseudo-experiments, where we assign the validation data points 1000 times to one of the clusters in a random way. This allows us to estimate the probability density function of the null hypothesis that the clustering does not entail information. In Fig. 2, the mean stability scores per cluster derived from bootstrapping are compared with the ones from the Monte-Carlo pseudo experiments. The nine patterns are significant on the 99 % level according to a two-sided t-test; the least stable spatial patterns with a stability score below 0.9 are WE1, SEE and CEE.

The visual inspection of the spatial patterns confirms their meaningfulness since natural geographical boundaries like mountains serve as delimiting boundaries, as is the case for IP, WE2 and SEE. In order to characterise the heatwave patterns, we examine the mean maximum temperature and the mean calendar day of the first heatwave occurrence (see Fig. 1). We find three spatially related groups when looking at the mean first calendar day of the heatwave in a year. The earliest events happen in the BI pattern with the mean first calendar day of the event of 25th June, followed by Northern patterns of SCA, CEE, and NEE at the beginning of July. The mean first calendar day of heatwave is the latest in the Southern and Central European patterns of IP, WE2, WE1, SEE, and GSI, where the mean first calendar day of heatwave occurs in the second half of July. The mean maximum temperature is higher for the patterns with fewer events - e.g., WE2 and WE1. From that, we can derive that events that belong to those patterns have their hot spots over the same area, while, e.g., in the case of GSI, the maxima of the respective events match to a lesser percentage.

Next, we visually compare the patterns to observed historical heatwaves in Europe. We find that many patterns obtained from the analysis on CRCM5-LE reproduce historical events, even though those have not been part of the analysis. For example, the WE1 is similar to the French heatwave in the summer of 2003[1]. The record-breaking heatwave in the summer of 1976 in Britain can be matched with the BI pattern[1]. CEE pattern reproduces the heatwave of 1994 in Eastern Germany and Poland[1]. Finally, the events of 2007 in the Balkans and Greece and 2010 in Russia can be matched to SEE and NEE, respectively[1].

Additional validation is performed by comparing the spatial patterns from CRCM5-LE to the ones derived from the clustering of heatwaves derived from the observational data set E-OBS. The E-OBS' spatial patterns can be found in Fig. 3. To compare both clustering results, we calculate the cosine similarity between the spatial patterns obtained from CRCM5-LE and those from E-OBS and match them by the maximum value. The measure is chosen to stay consistent with the distance measure used for clustering. Cosine similarity corresponds to one when the input vectors are identical and to zero if they are orthogonal. The results are shown in Table 1. The patterns IP, BI, SCA and CEE, are in excellent correspondence, as can be seen visually and from the pattern cosine similarity. Furthermore, WE1 and WE2 combine to one common pattern in E-OBS - the WE, as indicated by the high similarity value. Therefore, we calculate the sum of the patterns by adding the values pixelwise. Similar behaviour can be seen in SEE and GSI, which divide into a Southern and a Northern part. In contrast, the patterns originating from E-OBS divide into West (Italy) and East (Balkans and Greece). Finally, the two North-Eastern patterns in E-OBS combine into the NEE pattern of CRCM5-LE. Supplementary Table 1 shows the pattern similarity values derived from the ERA-Interim-driven model run of CRCM5 (CRCM5/ERA) and CRCM5-LE. The patterns can be found in Supplementary Figure 1. The results are comparable to those for E-OBS. They confirm that the dominating spatial heatwave

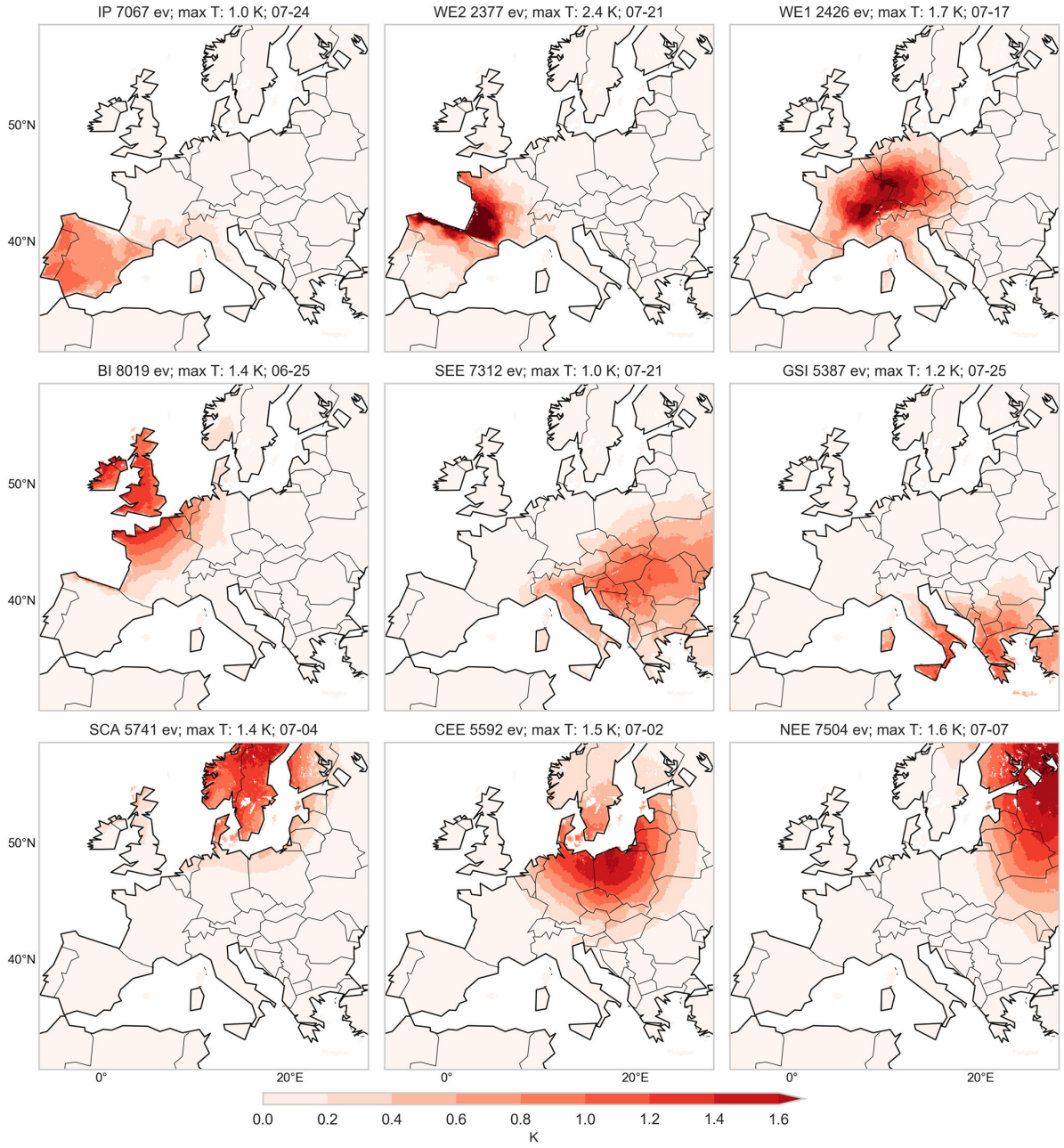


**Fig. 1 Nine typical heatwave patterns over Europe derived from CRCM5-LE.** Patterns obtained by hierarchical clustering of 1981-2010 Canadian Regional Climate Model 5 Large Ensemble (CRCM5-LE). In the title from left to right: pattern abbreviation, number of events belonging to the pattern (ev), mean maximum temperature in K, mean calendar day of the first heatwave occurrence. From left to right, from top to bottom: IP: Iberian Peninsula, WE2: Western Europe 2, WE1: Western Europe 1, BI: Britain and Ireland, SEE: South-Eastern Europe, GSI: Greece/Southern Italy, SCA: Scandinavia, CEE: Central/Eastern Europe, NEE: North-Eastern Europe.

patterns from CRCM5-LE are similar in the area they cover with those found when clustering observational data or the reanalysis-driven run of CRCM5. Given the difference in the number of events used as input for the analysis (1059 events from E-OBS vs 51.044 events from the 50 members of CRCM5-LE), we argue that the patterns originating from CRCM5-LE allow reliable statistical interpretability and robustness and are therefore used for further analysis.

Additionally, we test the robustness of our results in terms of domain choice. As we cannot pick a larger domain, we compare the resulting patterns for a smaller domain. We cut off ten boundary pixels on each side, thereby reducing the $280 \times 280$ grid to $260 \times 260$. The resulting heatwave patterns are similar in form and shape, as in Fig. 1, but without the grid cells at the domain's border. Therefore, we conclude that the resulting heatwave patterns do not depend on the domain choice. Moreover, we

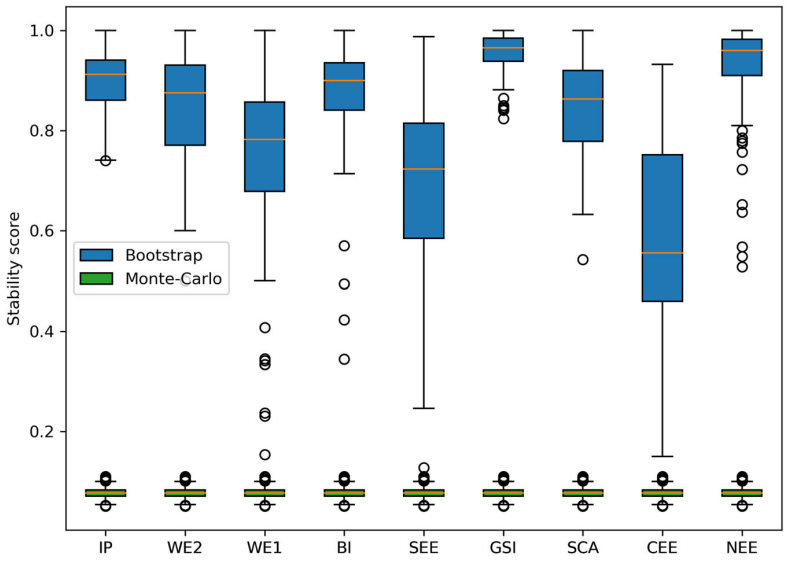


**Fig. 2 Mean stability score per heatwave pattern.** The stability score of bootstrap samples is compared to Monte-Carlo pseudo-experiments. The median in orange, end of the box, indicates the first and third quartiles. The boundaries of the whiskers indicate the 1.5 interquartile range.

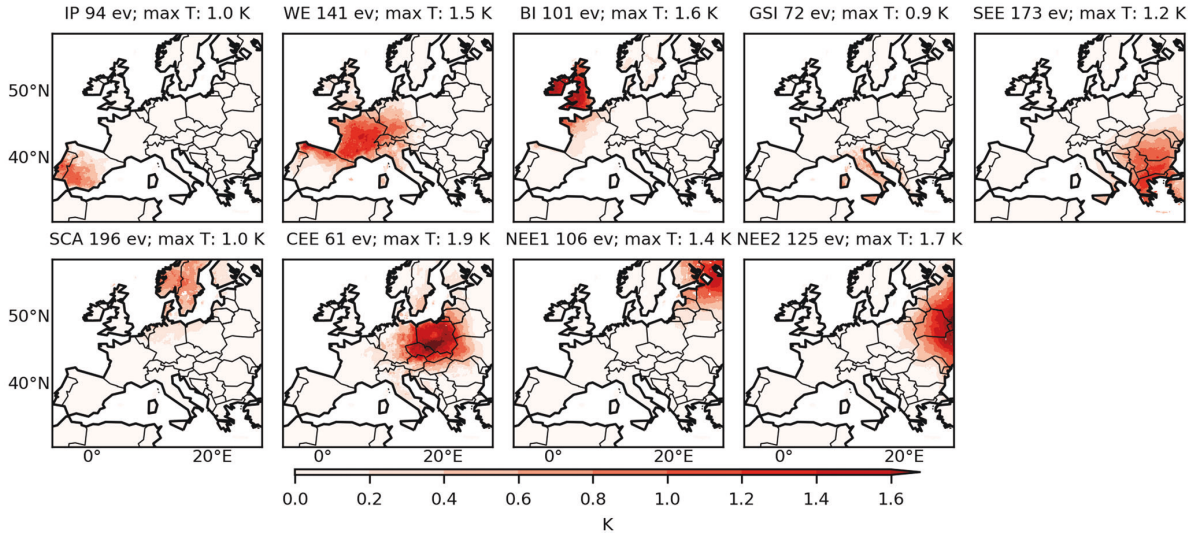


**Fig. 3 Nine typical heatwave patterns over Europe derived from E-OBS.** In the title from left to right: pattern abbreviation, number of events belonging to the pattern (ev), mean maximum temperature in K. Pattern names same are the same as in Fig. 1, except NEE1: North-Eastern Europe 1, NEE2: North-Eastern Europe 2.

compare the patterns when including sea grid cells. We calculate the spatial patterns with sea grid cells for the ERA-Interim driven run of CRCM5. The results are shown in Supplementary Fig. 3. We see that new patterns emerge over the sea areas that are not impacting land clusters. In 7 out of 9 cases, land patterns stay very similar—they cover the same area and, in some cases, add the sea areas along the coast (e.g., WE and BI). The CEE pattern is no longer present; however, it is constituted only out of 24 events without the sea grid cells and is, therefore, unstable. Moreover, IP splits up into IP1 and IP2. Therefore, we conclude that heatwave patterns over land are mainly unrelated to sea heatwaves, and we omit sea grid cells in further analysis.

## Seasonal connection to soil moisture and precipitation

Heatwaves and droughts are related phenomena that influence the formation of one another, as the hydrological cycle is inseparably connected to the heat-related processes in the atmosphere. Therefore, we inspect soil moisture anomalies and anomalies in seasonal precipitation before the heatwave occurrence (JFMA), after the heatwave (OND) and during the heatwave (MJJAS) in dependence on the number of heatwave days in every spatial pattern per summer.

The quantile regression method is applied to investigate the relationship between the number of heatwave days per summer season and the soil moisture or seasonal precipitation (for more

**Table 1.** E-OBS patterns assigned to CRCM5-LE patterns by the maximum value of cosine similarity.

| CRCM5-LE pattern | EOBS pattern | Similarity |
|---|---|---|
| IP | IP | 0.86 |
| BI | BI | 0.85 |
| SCA | SCA | 0.92 |
| CEE | CEE | 0.83 |
| SEE | SEE | 0.74 |
| GSI | SEE | 0.74 |
| NEE | NEE1 | 0.87 |
| WE1 | WE | 0.82 |
| WE2 | WE | 0.68 |
| WE1 + WE2 | WE | 0.93 |
| GSI + SEE | GSI + SEE | 0.92 |
| NEE | NEE1 + NEE2 | 0.95 |

information, see 'Methods' section). A scatter plot of soil moisture and precipitation versus the number of heatwave days for IP in JFMA is shown in Supplementary Fig. 2. We use the range of 10-90th percentile, which allows us to investigate if there is a link between the variables for the upper and hence more extreme quantiles. We expect that the relationship between soil moisture/precipitation and the number of heatwave days differs for upper quantiles. For each pattern, we plot the soil moisture anomalies for the 25 years (2% of most extreme events) with the highest number of heatwave days to obtain a visual validation for the correlations. The soil moisture anomalies in the upper portion of the soil column (0cm-10cm depth) are used instead of deeper soil moisture levels due to data availability. Additionally, we repeat the analysis using a model run where ERA-Interim is used as a boundary condition instead of CanESM2 to compare and validate the findings.

Figure 4a, b shows the quantile regression slopes of the number of heatwave days $N_{HW}$ in relation to soil moisture $mrsos_{JFMA}$ and precipitation anomaly $pr_{JFMA}$ in the preceding winter/spring season (JFMA). Statistically significant slopes with a 90% confidence level for a two-sided *t*-test are identified with black edge. We find gradually increasing negative slopes for an increasing number of heatwave days for North European patterns of BI and NEE and Southern-European patterns GSI, SEE, IP and WE2. In the case of the SEE pattern, there is a stronger influence of precipitation deficit on the number of heatwave days than when compared to other patterns. By contrast, we find no significant relationship between the soil moisture anomaly and precipitation deficit in winter/spring and the number of heatwave days in the Central European WE1 and CEE patterns, as well as in the Northern European pattern SCA. Our results suggest that there is a predictive power of soil moisture in the preceding winter/spring (JFMA) for heatwave occurrence in summer for South and North Europe. Moreover, our results suggest that there is predictive power of seasonal precipitation anomalies in winter/spring (JFMA) in SEE for heatwave occurrence in summer. In Supplementary Fig. 4, we show the results of the quantile regression analysis for the ERA-Interim-driven run of CRCM5. The results confirm the negative relationship between the number of heatwave days and mrsos anomaly in winter only for NEE, BI, SEE and IP, although none of the slopes is significant.

Figure 4c displays the spatial distribution of mrsos anomalies for the patterns IP, WE2, SEE, GSI, NEE and BI for the 25 years with the highest number of heatwave days. We choose the patterns that show a significant relationship in the quantile analysis. Significant anomalies are indicated with the black edge. We find that the extreme heatwave years in IP and SEE patterns are connected not only to local soil moisture and precipitation deficit in the pattern area but also in other parts of South Europe. We find that extreme heatwaves in the GSI pattern are connected to continental and Northern Europe soil moisture anomalies. Following Fig. 4a, the anomalies are insignificant for the Northern patterns (NEE, BI). The obtained results confirm findings concerning the positive influence of the dry winter season on hot summers[16,19,46]. We cannot find a significant dependency between soil moisture deficits in the South and heatwaves in the North of Europe in contrast to what is suggested by previous studies[16,24]. Results for the remaining patterns are displayed in Supplementary Fig. 5.

A deficit in soil moisture in the season during the heatwave (MJJAS) is present for all identified patterns in Fig. 5. We see significantly decreasing slopes in Fig. 5a of $N_{HW}$ versus $mrsos_{MJJAS}$ for all patterns. We find the same, although mostly non-significant, relationship when performing the analysis on the ERA-Interim dataset (see Supplementary Fig. 4b). These results are in accordance with a previous study that also found increasing negative slopes for the quantile analysis of soil moisture in relation to the percentage of heatwave days in Central and Eastern Europe[26]. We extend these results by finding this relationship also in Northern Europe. The results are similar in the case of seasonal precipitation as the dependent variable. We see that for the BI pattern, soil moisture has a bigger influence than for other patterns; for the IP pattern, precipitation anomalies are a more robust predictor compared to other patterns.

In Fig. 5c, the spatial patterns are displayed. A significant soil moisture deficit in SCA, CEE, and NEE for the 25 years with the highest number of heatwave days is also connected to a significant soil moisture increase in Western and Central Europe. A contrasting pattern is visible in the IP region: for the 25 years with the highest number of heatwave days, negative soil moisture anomalies are observed in South-Western Europe and positive in North-Eastern Europe. The dipolar structure is a well-known phenomenon: it has been shown in previous studies that positive phases of the North Atlantic Oscillation are connected to negative SPI averages in Southern Europe and positive averages in Northern Europe[47].

Extremely long periods of heatwaves pose substantial stress on the soil moisture visible in the following season (OND), as seen in Fig. 6. While for quantiles below 0.2, coefficients equal zero, the slopes turn negative for higher values for patterns SCA, NEE, SEE, CEE and IP. Out of those, significant soil moisture anomalies in the following season are present only for patterns SCA, NEE and SEE. This is confirmed by the analysis of the ERA-Interim-driven run (Supplementary Fig. 4c) apart from the highest quantile. These results serve as an indication of the memory of soil moisture in Northern Europe, as well as in the South-Eastern parts and suggest that there is a predictive power of the number of heatwave days per summer on dry anomalies in soil moisture in subsequent fall/winter (OND). The slopes of the quantile regression for precipitation anomalies are shown in Fig. 6b. None of the slope coefficients is negative; this suggests that hot summers do not lead to dry fall/winter in Europe. In contrast, we see a positive correlation between the upper quantiles of the number of hot days and seasonal precipitation in BI and NEE. The observed quantile regression coefficients are the lowest compared to the other seasons. For SCA, NEE and SEE, the negative soil moisture anomalies are also visible in the following winter season (see Fig. 6c). Results for the remaining patterns are displayed in Supplementary Fig. 5.

## DISCUSSION

Using cluster analysis, we identified (1) nine distinct patterns of European heatwaves, which we validated by comparing with E-OBS and applying bootstrapping. The spatial patterns offer not only the possibility to investigate regional heatwave characteristics, e.g., BI

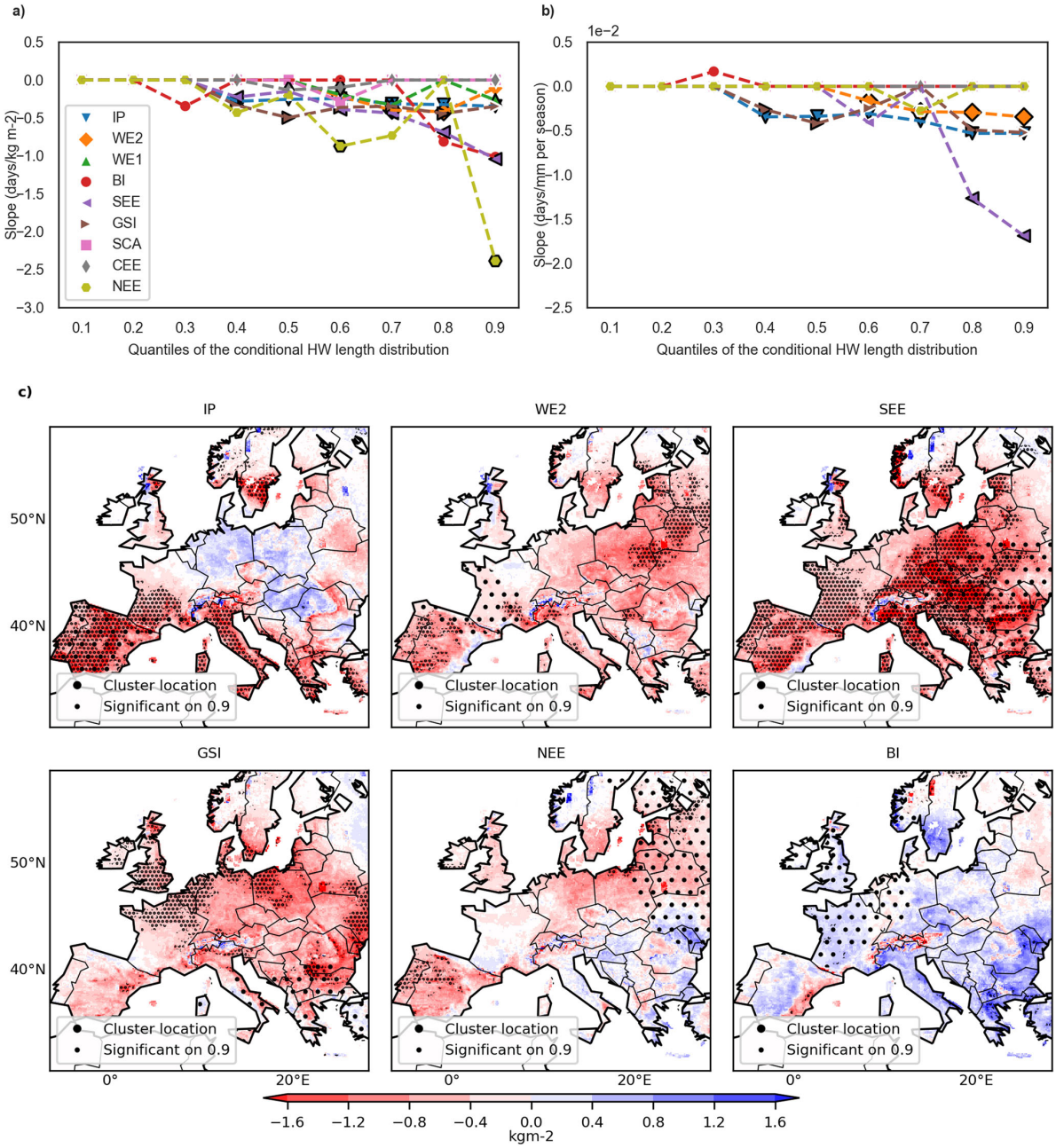


**Fig. 4 Quantile regression slopes for $N_{HW}$ in relation to moisture-related variables in winter before.** $N_{HW}$ versus $mrsos_{JFMA}$ (**a**) and $pr_{JFMA}$ (**b**). Statistically significant slopes with a 90% confidence level with a two-sided *t*-test are identified with black edge. $mrsos_{JFMA}$ for the 25 years with the highest number of heatwave days in chosen patterns with significant precipitation anomalies (**c**).

earliest heatwaves to latest in Southern parts of Europe, but also offer to understand further the seasonal influence of large-scale soil moisture anomalies and precipitation anomalies on the number of heatwave days in the chosen patterns and vice versa. We show that (2) soil moisture deficit in the preceding winter/spring (JFMA) can serve as a predictor for heatwaves in Southern (GSI, SEE, IP, WE2) and Northern (BI, NEE) Europe. Moreover, (3) all patterns show a significant negative relationship between soil moisture in the summer season (MJJAS) and the number of heatwave days. (4) The analysis of soil moisture anomalies in the following season (OND) shows a significant negative relationship for SCA, SEE and NEE. This shows that long heatwave events lead to a substantial soil moisture deficit preserved until the following season. For now, the obtained findings apply only to present-day climate; it has to be further investigated whether future climate change impacts these relationships.

In this study we perform a clustering analysis of heatwaves using a SMILE of a high-resolution RCM. Through the employment

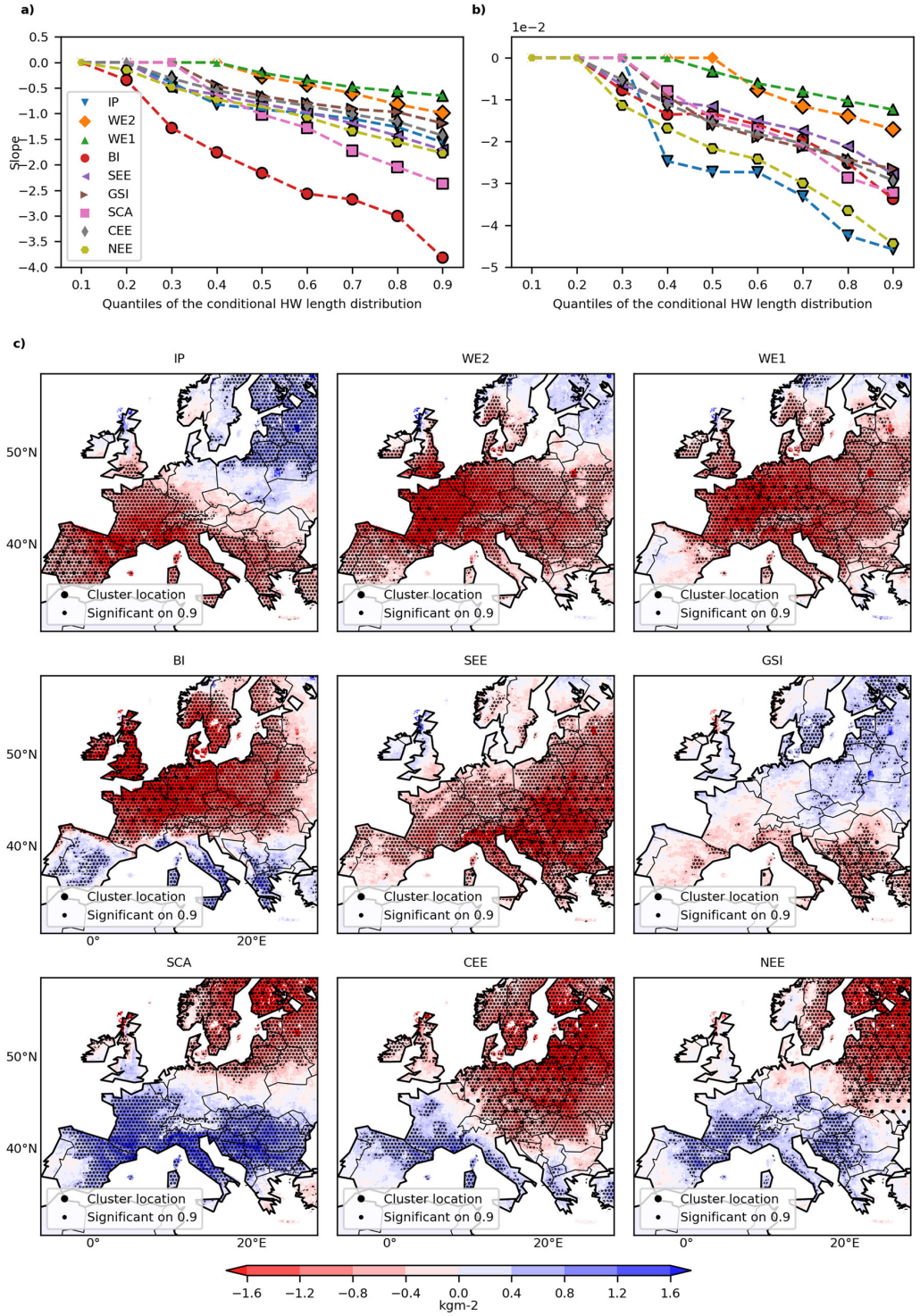


**Fig. 5 Quantile regression slopes for $N_{HW}$ in relation to moisture-related variables in summer.** $N_{HW}$ versus $mrsos_{MJAAS}$ (**a**) and $pr_{MJJAS}$ (**b**). Statistically significant slopes with a 90% confidence level with a two-sided *t*-test are identified with the black edge. $mrsos_{MJJAS}$ for 25 years with the highest number of heatwave days in (**c**).

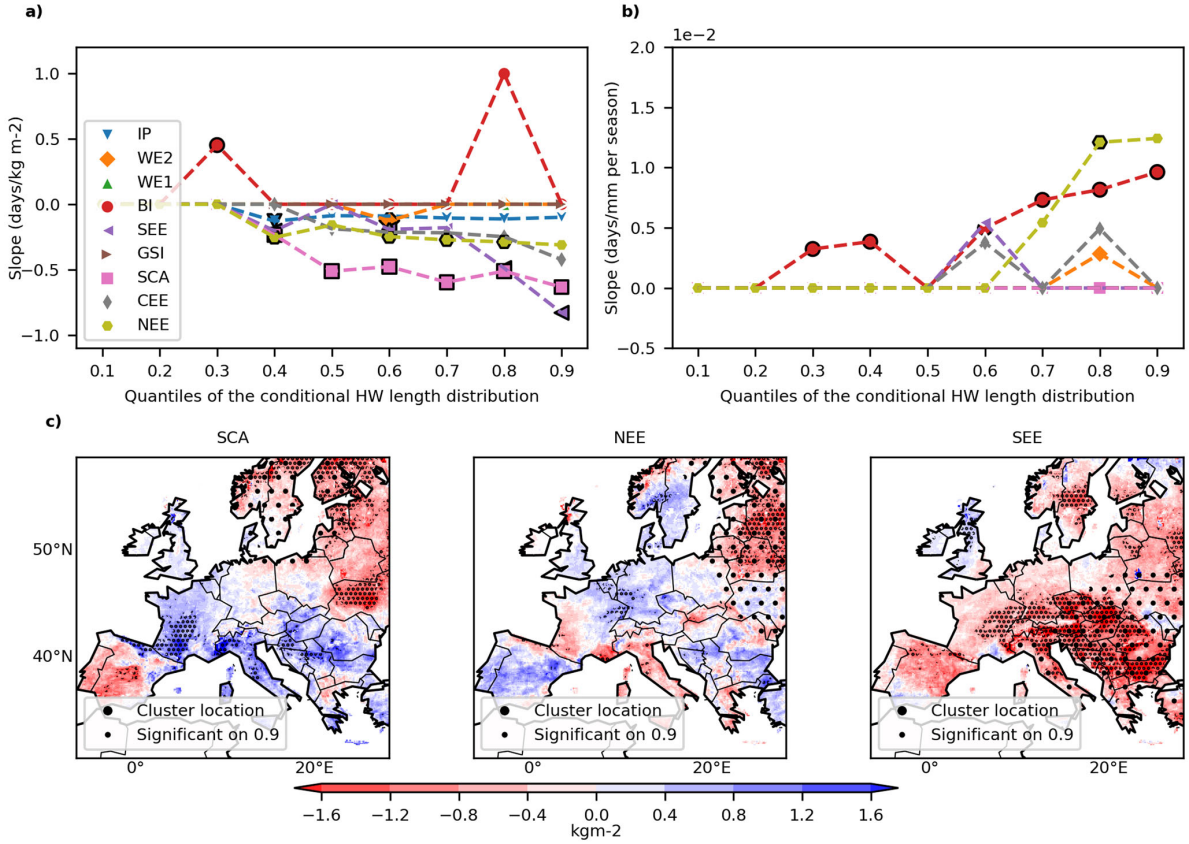


**Fig. 6 Quantile regression slopes for $N_{HW}$ in relation to moisture-related variables in fall after.** $N_{HW}$ versus $mrsos_{OND}$ (**a**) and $pr_{OND}$ (**b**). Statistically significant slopes with a 90% confidence level with a two-sided *t*-test are identified with black edging. $mrsos_{OND}$ for the 25 years with the highest number of heatwave days in chosen patterns with significant precipitation anomalies (**c**).

of the CRCM5-LE, we assess the natural variability of heatwaves and derive stable patterns of heatwaves. Regional Large Ensembles have proven useful in research on extreme events[15,32,33]. Nevertheless, it is known that the models are prone to biases regarding the modelling of land-atmosphere interactions that contribute to the formation of heat waves[48]. We find similar patterns when clustering using the E-OBS dataset or ERA-Interim driven run of CRCM5.

The classification into nine distinct heatwave patterns in Europe is a unique finding of this study. The study by Stefanon[24] finds six heatwave patterns, however, based on 78 heatwave events that consist of 643 heatwave days in contrast to more than 50,000 heatwave days used in our case. When we compare those patterns to the ones found by our analysis, we can assign them in the following way: 'Russian' pattern compares to NEE; 'Western Europe' pattern to WE1 and WE2; 'Eastern Europe' pattern to CEE, SEE, GSI; 'Iberian' pattern to IP; 'North Sea' pattern to BI and SCA; 'Scandinavian' pattern to none, however, its area is in huge parts outside of our domain. Therefore, we find similar heatwave patterns in both studies. Also, previous studies mostly used the percentage of rain days as a soil moisture proxy for the analysis instead of direct soil moisture, as in our case[16,17].

The coupling between spring soil moisture availability in Southern Europe (GSI, IP, WE2, SEE) and heatwave occurrence in summer has been described in previous studies[24,49]. The southern regions of Europe have a dry climate, where evaporation is soil moisture limited[50]. The described link between reduced soil moisture leading to fewer clouds and more solar radiation and, therefore, more heatwaves is valid for that region. It appears to be one of the main driving mechanisms for heatwave formation. The link for NEE is less significant, however present. It has to be investigated further how far other factors, such as snow cover, influence the link. The missing coupling between spring soil moisture and precipitation and the occurrence of a heatwave in SCA, WE1 and CEE can be explained by the fact that the vegetation system in those regions is rarely water-limited[51]. Therefore, even if there is a comparably dry spring, the soil still has enough moisture for evaporation and the formation of clouds.

We see a coupling between the heatwave occurrence in the Eastern and Northern parts of the domain (SCA, NEE, and SEE) and autumn drought occurrence. Those regions have a temperate climate and relatively high mean soil moisture values. This allows for a higher variability of soil moisture when compared to more Southern regions. Therefore, it takes the soil until the next season to recover after a prolonged heatwave. This relationship is missing for the Southern regions (IP, GSI, WE2). Those regions experience low mean precipitation in summer and, therefore, a lower expected and possible variability of soil moisture.

These results can, in most cases, be confirmed when performing the analysis on the ERA-Interim-driven run of CRCM5. The results mainly differ for upper quantiles (0.9). It can be explained by the fact that we analyse only 30 years; therefore, the upper quantile includes only three values, leading to high slope value uncertainty.

For future research, we suggest performing the analyses on deeper soil moisture levels, as those are known to show higher persistence[52]. Further, we suggest analysing for interdependencies

between heatwaves and precipitation/soil moisture deficits across different areas, as we see, e.g., in Fig. 6 for BI pattern, soil moisture over Central and Eastern Europe shows significant anomalies for the 25 years with the highest number of heatwave days. Also, a further heatwave-pattern-based investigation in terms of the effects of other variables, such as latent and sensible heat fluxes, would be of interest for future studies on the interrelation of heatwaves and droughts.

We suggest using the obtained patterns for heatwave analysis and predictability instead of pixelwise or even country-wise approaches in future studies. The obtained patterns allow a meaningful complexity reduction by finding spatially coherent regions instead of arbitrary grouping, e.g., by country. For agricultural research and the general public, the study's outcomes can enhance the predictability of heatwave events in Southern (GSI, SEE, IP, WE2) and Northern (BI, NEE) Europe on a seasonal scale.

Moreover, applying the described framework offers great potential for investigating other extreme events like droughts.

## METHODS

### Data sets

The central part of our analysis is based on the daily maximum temperature and monthly soil moisture and precipitation data from the single-model initial condition large ensemble (SMILE) consisting of 50 members, the Canadian Regional Climate Model 5 Large Ensemble (CRCM5-LE). The data was produced within the scope of the ClimEx Project (Ref. [39], www.climex-project.org). Dynamical downscaling via CRCM5-LE is applied to the data originating from the 50-member initial condition Canadian Earth System Model 2 (CanESM2)[53]. The data is provided at a resolution of 0.11° (12 km) and is produced for the years 1950–2099 for a European and an Eastern North America domain. Historical greenhouse gas concentrations are used for the years 1950–2005; starting from 2006, the RCP8.5[54] forcing scenario is used. We use the data from all 50 members for the years 1981–2010, translating to 1500 model years, which are analysed for heatwave events. A comparison of the CRCM5-LE to the E-OBS dataset has been performed in a previous study[39] and showed a temperature bias between −2 and +2 °C, while warm deviations mainly happen over highlands. For the validation of the obtained patterns, we use daily gridded observational data set E-OBS[55] for the years 1981-2010, as well as one model run of the CRCM5, which was driven by the global atmospheric reanalysis data set ERA-Interim via boundary conditions[39,56].

### Heatwave definition

Literature gives evidence for a wide range of similar heatwave definitions, which are adapted to the specific study goals[24,57–59]. In this study, we define heatwaves for land areas in continental Europe (EUR-11 domain) as prolonged periods of above-average temperatures in an extended area during the period 1981–2010. These heatwaves consist of at least three consecutive hot days, where hot days are characterised by a positive anomaly of daily maximum temperature (tasmax) to the local 95th JJA (1981–2010) percentile, allowing for comparability across the domain. In order to obtain robust signals, we use the 3-day-running mean to derive these anomalies. Negative anomalies are set to zero to focus on hot extremes[24]. Two heatwaves are separated by a minimum of three days below threshold[57]. We remove heatwave days consisting of patterns smaller than $9 \times 9$ grid cells. An additional filter is introduced to eliminate spatially small events covering an area of less than 1% of the land area (500 grid cells). Positive anomalies only occur during the months May-October in our data sets. The analysis is based on heatwave days fulfilling the above-mentioned criteria and amounts in the case of CRCM5-LE to around 50,000 heatwave days used as input for the clustering analysis.

### Clustering analysis

In literature, clustering has frequently been used to analyse and classify weather patterns in the mid-latitudes[60–62]. This study uses the obtained heatwave days as input to the agglomerative hierarchical clustering algorithm[24,29]. Distance between two vectors, $\boldsymbol{r}$ and $\boldsymbol{q}$, is defined here as follows:

$$d(\boldsymbol{r}, \boldsymbol{q}) = 1 - cs(\boldsymbol{r}, \boldsymbol{q}) \tag{1}$$

$$cs(\boldsymbol{r}, \boldsymbol{q}) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M} r_{i,j} q_{i,j}}{\left(\sum_{i=1}^{N}\sum_{j=1}^{M} r_{i,j}\right)^{1/2}\left(\sum_{i=1}^{N}\sum_{j=1}^{M} p_{i,j}\right)^{1/2}} \tag{2}$$

$cs(\boldsymbol{r}, \boldsymbol{q})$ refers to the cosine similarity measure between two vectors[63]. It is defined as 1 for parallel vectors and as 0 for orthogonal. For the clustering algorithm, the average linkage is used[63].

The optimal number of clusters is determined by applying the elbow method[64]: we compute the distortion score as the sum of squared distances to the assigned centre for every possible number of clusters and pick the number of clusters that corresponds to the knee of the curve[45,64].

Due to a large number of events, the obtained data set has, in absolute numbers, a higher number of a-typical events, which have a big distance to all other events of the data set. Filtering by preliminary clustering to 32 clusters is introduced to remove these events. Events belonging to so-called minority clusters with a small number of events (<0.1% of the data) are removed from the data set[65]. In total less than 1% of heatwave events are removed that way. Repeated clustering is performed on the resulting data set. We derive 12 clusters as the optimal number from the elbow method. The obtained clusters are cross-validated by 100 times dividing the data set into a verification period that amounts to 1/100 of the data set and the remaining 99/100 used as input to clustering. The nearest distance to the training clusters then determines the labels for the verification period. Those are then compared to the ones originating from the clustering on the whole data set. Finally, a stability score is computed per cluster that amounts to the number of correctly assigned validation events to the total number of events per cluster. The results are then compared to a Monte-Carlo pseudo-experiment, where the labels are assigned in a purely random way 1000 times. Three out of twelve clusters do not pass the described validation; nine are significant on the 99%-level according to a two-sided $t$-test.

### Quantile regression

In order to evaluate the impact of heatwave length on soil moisture and seasonal precipitation deficit, we use quantile regression as suggested by similar studies[17,26]. Quantile regression is a method that goes beyond standard linear regression, as it can be used when the linearity and independence of variables are not given. It estimates the conditional median of the target variable[66]. Here, we use a linear model for the conditional quantiles.

For the quantile regression, we use the following variables derived from the CRCM5-LE data set:

1. Number of heatwave days per pattern $N_{HW}$ per summer season of May, June, July, August, and September (MJJAS)
2. Mean soil moisture anomalies in the upper portion of the soil column (top 10 cm) averaged for the following three seasons in the pattern region: January, February, March, April (JFMA) $mrsos_{JFMA}$; May, June, July, August, September (MJJAS) $mrsos_{MJJAS}$ and October, November, December (OND) $mrsos_{OND}$
3. Summed precipitation anomalies for the same seasons as in (2): $pr_{JFMA}$, $pr_{MJJAS}$ and $pr_{OND}$ in the pattern region.

The pattern region for *mrsos* and *pr* is defined as the area of 100 land pixels around the maximum of the spatial pattern. Given the

spatial resolution of 12 km, which amounts to approximately 14400 $km^2$.

## DATA AVAILABILITY

The data is retrieved from the CRCM5-LE via https://www.climex-project.org/en/data-access/. The E-OBS data set from the EU-FP6 project UERRA, the Copernicus Climate Change Service, and the data providers in the ECA&D project can be accessed via https://www.ecad.eu. Derived data supporting the findings of this study are available from the corresponding author upon reasonable request.

## CODE AVAILABILITY

The source codes for the analysis of this study are available from the corresponding author upon reasonable request.

Received: 19 May 2022; Accepted: 10 January 2023;
Published online: 21 January 2023

## REFERENCES

1. Russo, S., Sillmann, J. & Fischer, E. M. Top ten European heatwaves since 1950 and their occurrence in the coming decades. *Environ. Res. Lett.* **10**, 124003 (2015).
2. Sedlmeier, K., Feldmann, H. & Schädler, G. Compound summer temperature and precipitation extremes over Central Europe. *Theor. Appl. Climatol.* **131**, 1493–1501 (2018).
3. Orth, R., O, S., Zscheischler, J., Mahecha, M. D. & Reichstein, M. Contrasting biophysical and societal impacts of hydro-meteorological extremes. *Environ. Res. Lett.* **17**, 014044 (2022).
4. Robine, J.-M. et al. Death toll exceeded 70,000 in Europe during the summer of 2003. *C. R. Biol.* **331**, 171–178 (2008).
5. Kjellström, T., Maître, N., Saget, C., Otto, M. & Karimova, T. *Working on a Warmer Planet: The Effect of Heat Stress on Productivity And Decent Work* (International Labour Organization, 2019).
6. Zink, M. et al. The German drought monitor. *Environ. Res. Lett.* **11**, 074002 (2016).
7. Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M. & García-Herrera, R. The hot summer of 2010: redrawing the temperature record map of Europe. *Science* **332**, 220–224 (2011).
8. Perkins-Kirkpatrick, S. & Lewis, S. Increasing trends in regional heatwaves. *Nat. Commun.* **11**, 1–8 (2020).
9. García-León, D. et al. Current and projected regional economic impacts of heatwaves in Europe. *Nat. Commun.* **12**, 1–10 (2021).
10. Ballester, J., Rodó, X. & Giorgi, F. Future changes in Central Europe heat waves expected to mostly follow summer mean warming. *Clim. Dyn.* **35**, 1191–1205 (2010).
11. Rossi, G., Cancelliere, A. & Giuliano, G. Case study: multicriteria assessment of drought mitigation measures. *J. Water Resour. Plan. Manag.* **131**, 449–457 (2005).
12. Lowe, D., Ebi, K. L. & Forsberg, B. Heatwave early warning systems and adaptation advice to reduce human health consequences of heatwaves. *Int. J. Environ. Res. Public Health* **8**, 4623–4648 (2011).
13. Hao, Z., Singh, V. P. & Xia, Y. Seasonal drought prediction: advances, challenges, and future prospects. *Rev. Geophys.* **56**, 108–141 (2018).
14. Wang, Y. et al. A random forest model to predict heatstroke occurrence for heatwave in China. *Sci. Total Environ.* **650**, 3048–3053 (2019).
15. Felsche, E. & Ludwig, R. Applying machine learning for drought prediction in a perfect model framework using data from a large ensemble of climate simulations. *Nat. Hazards Earth Syst. Sci.* **21**, 3679–3691 (2021).
16. Vautard, R. et al. Summertime European heat and drought waves induced by wintertime Mediterranean rainfall deficit. *Geophys. Res. Lett.* **34**, https://doi.org/10.1029/2006GL028001, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2006GL028001 (2007).
17. Quesada, B., Vautard, R., Yiou, P., Hirschi, M. & Seneviratne, S. I. Asymmetric European summer heat predictability from wet and dry southern winters and springs. *Nat. Clim. Change* **2**, 736–741 (2012).
18. Kornhuber, K., Petoukhov, V., Petri, S., Rahmstorf, S. & Coumou, D. Evidence for wave resonance as a key mechanism for generating high-amplitude quasi-stationary waves in boreal summer. *Clim. Dyn.* **49**, 1961–1979 (2017).
19. Seneviratne, S. I. et al. Investigating soil moisture–climate interactions in a changing climate: a review. *Earth Sci. Rev.* **99**, 125–161 (2010).
20. Miralles, D. G., Gentine, P., Seneviratne, S. I. & Teuling, A. J. Land–atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges. *Ann. N. Y. Acad. Sci.* **1436**, 19 (2019).
21. Schumacher, D. L. et al. Amplification of mega-heatwaves through heat torrents fuelled by upwind drought. *Nat. Geosci.* **12**, 712–717 (2019).
22. O, S. et al. The role of climate and vegetation in regulating drought-heat extremes. *J. Clim.* **1**, 1–21 (2022).
23. Horton, D. E. et al. Contribution of changes in atmospheric circulation patterns to extreme temperature trends. *Nature* **522**, 465–469 (2015).
24. Stefanon, M., D'Andrea, F. & Drobinski, P. Heatwave classification over Europe and the Mediterranean region. *Environ. Res. Lett.* **7**, 014023 (2012).
25. Della-Marta, P. M. et al. Summer heat waves over Western Europe 1880–2003, their relationship to large-scale forcings and predictability. *Clim. Dyn.* **29**, 251–275 (2007).
26. Hirschi, M. et al. Observational evidence for soil-moisture impact on hot extremes in Southeastern Europe. *Nat. Geosci.* **4**, 17–21 (2011).
27. Whan, K. et al. Impact of soil moisture on extreme maximum temperatures in Europe. *Weather Clim. Extremes* **9**, 57–67 (2015).
28. Sousa, P. M. et al. Distinct influences of large-scale circulation and regional feedbacks in two exceptional 2019 European heatwaves. *Commun. Earth Environ.* **1**, 1–13 (2020).
29. Wang, P., Tang, J., Shuyu, W., Xinning, D. & Juan, F. Regional heatwaves in China: a cluster analysis. *Clim. Dyn.* **50**, 1901–1917 (2018).
30. Katsafados, P., Papadopoulos, A., Varlas, G., Papadopoulou, E. & Mavromatidis, E. Seasonal predictability of the 2010 Russian heat wave. *Nat. Hazards Earth Syst. Sci.* **14**, 1531–1542 (2014).
31. Black, E. et al. Factors contributing to the summer 2003 European heatwave. *Weather* **59**, 217–223 (2004).
32. Böhnisch, A., Mittermeier, M., Leduc, M. & Ludwig, R. Hot spots and climate trends of meteorological droughts in Europe-Assessing the percent of normal index in a single-model initial-condition large ensemble. *Front. Water* **3** https://doi.org/10.3389/frwa.2021.716621 (2021).
33. Poschlod, B., Ludwig, R. & Sillmann, J. Ten-year return levels of sub-daily extreme precipitation over Europe. *Earth Syst. Sci. Data* **13**, 983–1003 (2021).
34. Perkins-Kirkpatrick, S. E., Fischer, E. M., Angélil, O. & Gibson, P. B. The influence of internal climate variability on heatwave frequency trends. *Environ. Res. Lett.* **12**, 044005 (2017).
35. Jeong, D. I., Yu, B. & Cannon, A. J. Links between atmospheric blocking and North American winter cold spells in two generations of Canadian earth system model large ensembles. *Clim. Dyn.* **57**, 2217–2231 (2021).
36. Deser, C. et al. Insights from earth system model initial-condition large ensembles and future prospects. *Nat. Clim. Change* **10**, 277–286 (2020).
37. Laprise, R. et al. Climate projections over CORDEX Africa domain using the fifth-generation Canadian regional climate model (CRCM5). *Clim. Dyn.* **41**, 3219–3246 (2013).
38. Lucas-Picher, P., Laprise, R. & Winger, K. Evidence of added value in north american regional climate model hindcast simulations using ever-increasing horizontal resolutions. *Clim. Dyn.* **48**, 2611–2633 (2017).
39. Leduc, M. et al. The ClimEx Project: a 50-member ensemble of climate change projections at 12-km resolution over Europe and Northeastern North America with the Canadian Regional Climate Model (CRCM5). *J. Appl. Meteorol. Clim.* **58**, 663–693 (2019).
40. von Trentini, F., Aalbers, E. E., Fischer, E. M. & Ludwig, R. Comparing interannual variability in three regional single-model initial-condition large ensembles (SMILEs) over Europe. *Earth Syst. Dyn.* **11**, 1013–1031 (2020).
41. von Trentini, F., Leduc, M. & Ludwig, R. Assessing natural variability in RCM signals: comparison of a multi model EURO-CORDEX ensemble with a 50-member single model large ensemble. *Clim. Dyn.* **53**, 1963–1979 (2019).
42. Poschlod, B., Zscheischler, J., Sillmann, J., Wood, R. R. & Ludwig, R. Climate change effects on hydrometeorological compound events over Southern Norway. *Weather Clim. Extremes* **28**, 100253 (2020).
43. Mittermeier, M., Braun, M., Hofstätter, M., Wang, Y. & Ludwig, R. Detecting climate change effects on Vb cyclones in a 50-member single-model ensemble using machine learning. *Geophys. Res. Lett.* **46**, 14653–14661 (2019).
44. Lukasová, A. Hierarchical agglomerative clustering procedure. *Pattern Recognit.* **11**, 365–381 (1979).
45. Kodinariya, T. M. & Makwana, P. R. Review on determining number of cluster in k-means clustering. *Int. J.* **1**, 90–95 (2013).
46. Wang, G., Dolman, A. & Alessandri, A. A summer climate regime over Europe modulated by the north Atlantic oscillation. *Hydrol. Earth Syst. Sci.* **15**, 57–64 (2011).
47. López-Moreno, J. I. & Vicente-Serrano, S. M. Positive and negative phases of the wintertime North Atlantic oscillation and drought occurrence over Europe: a multitemporal-scale approach. *J. Clim.* **21**, 1220–1243 (2008).
48. Sippel, S. et al. Refining multi-model projections of temperature extremes by evaluation against land–atmosphere coupling diagnostics. *Earth Syst. Dyn.* **8**, 387–403 (2017).

49. Zampieri, M. et al. Hot European summers and the role of soil moisture in the propagation of Mediterranean drought. *J. Clim.* **22**, 4747–4758 (2009).
50. Teuling, A. et al. A regional perspective on trends in continental evaporation. *Geophys. Res. Lett.* **36**, L02404, https://doi.org/10.1029/2008GL036584 (2009).
51. Koster, R. D. et al. Regions of strong coupling between soil moisture and precipitation. *Science* **305**, 1138–1140 (2004).
52. Xu, Z.-g, Wu, Z.-y, He, H., Guo, X. & Zhang, Y.-l Comparison of soil moisture at different depths for drought monitoring based on improved soil moisture anomaly percentage index. *Water Sci. Eng.* **14**, 171–183 (2021).
53. Martynov, A. et al. Reanalysis-driven climate simulation over CORDEX North America domain using the Canadian Regional Climate Model, version 5: model performance evaluation. *Clim. Dynam.* **41**, 2973–3005 (2013).
54. IPCC. *Climate Change 2013 - The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the IPCC*. Assessment report (Intergovernmental Panel on Climate Change): Working Group (Cambridge University Press, Cambridge, 2013).
55. Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M. & Jones, P. D. An ensemble version of the e-obs temperature and precipitation data sets. *J. Geophys. Res.* **123**, 9391–9409 (2018).
56. Dee, D. P. et al. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**, 553–597 (2011).
57. Keellings, D. & Moradkhani, H. Spatiotemporal evolution of heat wave severity and coverage across the United States. *Geophys. Res. Lett.* **47**, e2020GL087097 (2020).
58. Clemesha, R. E. S., Guirguis, K., Gershunov, A., Small, I. J. & Tardy, A. California heat waves: their spatial evolution, variation, and coastal modulation by low clouds. *Clim. Dyn.* **50**, 4285–4301 (2018).
59. Lhotka, O. & Kyselý, J. Spatial and temporal characteristics of heat waves over central Europe in an ensemble of regional climate model simulations. *Clim. Dyn.* **45**, 2351–2366 (2015).
60. Smyth, P., Ghil, M., Ide, K., Roden, J. & Fraser, A. Detecting atmospheric regimes using cross-validated clustering. In *KDD*, 61–66 (1997).
61. Machado, J. & Lopes, A. M. Rare and extreme events: the case of Covid-19 pandemic. *Nonlinear Dyn.* **100**, 2953–2972 (2020).
62. Hannachi, A., Straus, D. M., Franzke, C. L., Corti, S. & Woollings, T. Low-frequency nonlinearity and regime behavior in the northern hemisphere extratropical atmosphere. *Rev. Geophys.* **55**, 199–234 (2017).
63. Cheng, X. & Wallace, J. M. Cluster analysis of the northern hemisphere wintertime 500-hpa height field: Spatial patterns. *J. Atmos. Sci.* **50**, 2674–2696 (1993).
64. Jung, Y., Park, H., Du, D.-Z. & Drake, B. L. A decision criterion for the optimal number of clusters in hierarchical clustering. *J. Glob. Optim.* **25**, 91–111 (2003).
65. Li, Y. et al. Cluster-based data filtering for manufacturing big data systems. *J. Quality Technol.* 54:3, 290–302 (2021).
66. Koenker, R. & Hallock, K. F. Quantile regression. *J. Econ. Perspect.* **15**, 143–156 (2001).

## ACKNOWLEDGEMENTS

We acknowledge the E-OBS data set from the EU-FP6 project UERRA, the Copernicus Climate Change Service, and the data providers in the ECA&D project (https://www.ecad.eu). The CRCM5 LE was created within the ClimEx project, which was funded by the Bavarian State Ministry for the Environment and Consumer Protection. Computations of the CRCM5 LE were made on the SuperMUC supercomputer at Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities. We acknowledge Environment and Climate Change Canada for providing the CanESM2 LE driving data.

## AUTHOR CONTRIBUTIONS

E.F. designed the research under the supervision of R.L. E.F. and A.B. performed the data analysis and led the writing of the paper. E.F. prepared all figures. All authors contributed to the interpretation of the findings and revision of the paper.

## FUNDING

Open Access funding enabled and organized by Projekt DEAL.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41612-023-00330-5.

**Correspondence** and requests for materials should be addressed to Elizaveta Felsche.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## 4.3 Paper III: Will Present-Day Compound Hot and Dry European Summers Still Be Extreme in the Future?

**Reference:** Felsche, E., Böhnisch, A., Poschlod B., Ludwig, R. (2023). Will Past Compound Hot and Dry European Summers Still Be Extreme in the Future? Submitted to Communications Earth & Environment

**Plain language summary:** Heatwaves and dry spells are significant climate hazards with far-reaching implications for health, economy, agriculture, and ecosystems. The frequency of compound hot and dry summers in Europe has risen in recent years. In this study we first identify nine European regions where compound hot and dry events happen simultaneously. We than identify most extreme historical occurrences of compound hot and dry events for every identified region using historical reanalysis ERA5. Using this data we examine whether past compound events will remain rare climatological events under changing climate conditions. We compare it with model data on three Global Warming Levels (GWL): +1.2K, +2K and +3K for nine selected sub-regions. Key findings indicate a significant increase in the frequency of the most extreme past occurrences under GWL2 and GWL3. For specific events, the probability of occurrence rises by up to 5-6 times from GWL2 to GWL3. Moreover, our analysis unveils a significant northward shift in the climatology of hot and dry events under GWL3. The hot and dry climate observed in Eastern Europe under current conditions is anticipated to extend into substantial parts of the Baltic Sea Coast and Scandinavia.

**Author's contribution:** E.F. designed the research under the supervision of R.L.. E.F. , A.B. and B.P. regularly feedbacked the results and provided ideas for further development. E.F. performed the data analysis and led the writing of the paper. E.F. prepared all figures. All authors contributed to the interpretation of the findings and revision of the paper.

**Scope of the journal:** "Communications Earth & Environment is an open access journal from Nature Portfolio that publishes high-quality research, reviews and commentary in the Earth, environmental and planetary sciences. [...] Our scope covers all areas of the geosciences, climate and environmental sciences as well as planetary sciences, including those at the interface with ecology, sustainability and environmental social sciences." (Nature, 2024)

**Status:** submitted

**Journal:** Communications Earth & Environment

**Impact Factor (2-Year):** 7.9

# Will Present-Day Compound Hot and Dry European Summers Still Be Extreme in the Future?

**Elizaveta Felsche[1,2,3,*], Andrea Böhnisch[2], Benjamin Poschlod[4], and Ralf Ludwig[2]**

[1]Technical University of Munich, Munich, Germany
[2]Department of Geography, Ludwig-Maximilans-Universität München, Munich, Germany
[3]Center for Digital Technology and Management, Munich, Germany
[4]Research Unit Sustainability and Climate Risk, Center for Earth System Research and Sustainability (CEN), Universität Hamburg, Hamburg, Germany
[*]felsche@cdtm.de

## ABSTRACT

Heatwaves and dry spells are significant climate hazards with far-reaching implications for health, economy, agriculture, and ecosystems. The frequency of compound hot and dry summers in Europe has risen in recent years. This study examines whether past compound events will remain rare climatological events under changing climate conditions. We use reanalysis data (2001-2022) and compare it with model data on three Global Warming Levels (GWL): +1.2 K, +2 K and +3 K for nine selected sub-regions. Key findings indicate a significant increase in the frequency of the most extreme past occurrences under GWL2 and GWL3. For specific events, the probability of occurrence rises by up to 5-6 times from GWL2 to GWL3. Moreover, our analysis unveils a significant northward shift in the climatology of hot and dry events under GWL3. The hot and dry climate observed in Eastern Europe under current conditions is anticipated to extend into substantial parts of the Baltic states, Finnland and Scandinavia.

## 1 Introduction

Prolonged periods of heatwaves and droughts significantly affect human health, the economy, agriculture, and natural ecosystems. Compounding events - two or more separate co-occurring events - are the ones that are known to cause high impacts[1,2]. In recent years, Europe has experienced a series of compounding hot and dry events, e.g. 2003, 2015 and 2018[3,4], each of which has caused increased mortality[5,6], economic losses due to crop yield reduction, blockage of river transportation due to low flow, worker productivity decrease[7], and posed extreme stress on the natural ecosystems.

Quantifying the probability of historical compound hot and dry events is challenging due to the atmospheric and hydrological interrelation of temperature and precipitation[8]. Extreme heat and the absence of precipitation share the same atmospheric large-scale drivers as anticyclonic conditions, which lead to reduced cloudiness. Moreover, there is a self-intensifying feedback between hot and dry conditions[9,10]: with the onset of a heatwave, soil dries, and when falling below a certain threshold, a larger fraction of solar radiation is transformed into sensible heat. This leads to an increase in air temperature and evaporative demand, which in turn dries the soil even more and reduces precipitation due to lower evapotranspiration[11,12]. During hot and dry compound events, the Bowen Ratio of sensible heat to latent heat is enhanced compared to non-compound events in large parts of Europe, and lower soil moisture conditions are present[13]. Previous studies have shown that seasonal summer temperatures and precipitation in Europe are highly correlated[1,14], meaning that a higher number of compounding events happen than one would expect when looking at univariate probabilities and assuming independence. Moreover, looking to the future, rising summer average temperatures[15] and a drying trend[16] in the European summers might favour the intensification of the feedback mechanisms under changing climatic conditions[17,18].

Quantifying the changes in probability of historical compound hot and dry events for different global warming levels is essential for many reasons. Determining the probability of historical events under future conditions and comparing it to the present probability gives us insight into what we might expect to experience, contingent

upon global warming levels we might reach. This knowledge is invaluable for stakeholders like policy-makers, politicians, engineers, and farmers when managing water resources, adjusting agricultural practices or adapting to changing ecological minimum flow conditions[19]. A significant increase in probability strengthens the need for impact mitigation[20]. For instance, if compound hot and dry events should become a frequent feature of the European climate, there is a need to develop alternative solutions for cooling of thermal power plants, as they heavily depend on cooling water from natural resources[21], implementation of measures to conserve water[22,23], establishment of alternative transportation methods to shipping in summer and building green spaces like parks to reduce the urban heat island and urban sprawl effects to not additionally worsen the conditions in cities due to man-made issues[24,25].

Thus, the probability of occurrence of compound hot and dry events is essential to assess; however, those events are complex and multivariate in terms of drivers and feedbacks and happen rarely by definition. The probability quantification based on observational data is limited due to its temporal length. Applying a Single Model Initial Condition Large Ensemble (SMILE)[26] allows overcoming these limitations by providing a robust statistical estimation of extreme event occurrence. A SMILE consists of a multitude of simulations (ensemble members), each of which has the same forcing and the same physical model but differing initial conditions[27]. This allows to distinguish between the signal delivered by the internal chaotic nature of climate (natural climate variability) and a forced response due to effects like climate change. We argue for the application of SMILEs for extreme event probability quantification. Past studies have confirmed the usefulness of SMILEs for investigating compound extremes[4,13,27,28].

In this study we want to identify the most extreme compound hot and dry events on the European and regional scale of the past two decades, 2001-2022. For those events, we aim to quantify the probability of occurrence given the current climate and how this probability changes under projected global warming levels of +2 K or +3 K (GWL2 or GWL3, respectively). Seneviratne et al. and Gampe et al. used Global Warming Levels (GWL) to communicate climate change impacts at temperature targets relevant to policy and decision-making[29,30]. Moreover, they allow to compare results across forcing scenarios. In the study, we estimate the probability of the most extreme historical hot and dry events of the two past decades 2001-2022 in the European climate reanalysis dataset ERA5[31] by comparing it to a 50-member regional large ensemble, the Canadian Regional Climate Model, version 5, Large Ensemble (CRCM5-LE)[32] under RCP8.5. For the most extreme compound hot and dry events in the ERA5 dataset, we quantify the probability for those events to happen by using the following three periods out of CRCM5-LE:

- **PRES**: $+1.2K$ GWL model world representing the present conditions. It corresponds to the model years 2001-2020.
- **GWL2**: $+2K$ GWL model world representing the positive target set by the Paris Agreement; model years 2021-2040.
- **GWL3**: $+3K$ GWL model world, approximately representing the realistic perspective following current trend; model years 2042-2061.

We inspect temperature and precipitation and use seasonal summer averages (JJA) for the event definition. Due to the interrelation of those two variables, we cannot inspect those in univariate terms, but we have to model their interdependence using copulas[8]. In recent years, copulas have been widely used to study the interrelation between multiple variables[13,33,34].

Multivariate copulas are especially useful for assessing the occurrence probability and return periods of compound events[35–37]. We use the Survival Kendall probability $p_{SK}$, which estimates the probability of having an event at least as rare in probability as the one observed[38]. With the application on compound hot and dry events, we follow the study of Aghakouchak et al., which investigated the 2014 Californian heatwave and drought[35]. Survival Kendall probability can be understood as a critical layer $L$ separating the bivariate probability space into critical and non-critical regions. We use a Generalized Pareto Distribution (GPD) fit for values above the 95th percentile in temperature while using empirical distributions below. For precipitation, we use empirical distributions. The isolines dividing the probability space are shown for one arbitrary grid cell in Fig. 1. By spatially clustering the

resulting yearly $p_{SK}$ spatial maps of CRCM5-LE PRES, we identify nine European sub-regions most likely to experience a simultaneous event. We use the sub-regions to calculate local probabilities by averaging temperature and precipitation and then computing the $p_{SK}$ probability for the following regions: SWE: South-West Europe, CMD: Central Mediterranean, BP: Balkan Peninsula, AC: Atlantic Coast, CE: Central Europe, EE: Eastern Europe, NBS: North and Baltic Sea, NEE: North-East Europe and NSC: North Scandinavia. We obtained the regions by applying Hierarchical Agglomerative Clustering on all of the yearly events with $p_{SK} < 0.1$ that have a spatial extent of at least 500 grid cells ($\approx 1\%$ of the land area). More details can be found in Methods. Fig. 2 displays the entire European domain and sub-regions.

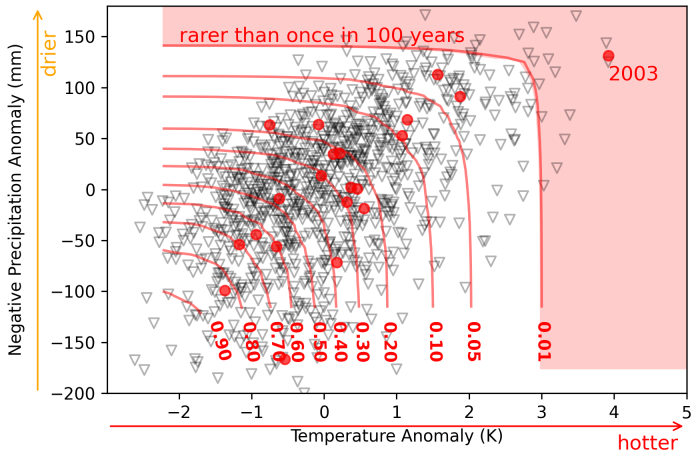


**Figure 1.** Temperature and negative precipitation scatter plot for one exemplary grid cell. Black triangles for CRCM5-LE data, red dots for ERA5 data. $p_{SK}$ isolines in the probability space are shown in red.

## 2 Results

### 2.1 Identifying most extreme historical compound events on the European scale

We first focus on identifying the most extreme event in the past two decades (2001-2022) on the European scale. Therefore, we detrend ERA5 for 1959-2022 and calculate its univariate quantiles for temperature and precipitation. We map the quantiles of ERA5 with those from CRCM5-LE PRES and adjust ERA5 according to the quantile. The authors argue in favour of this approach, in contrast to proceeding as usual and using the historical data as a reference, which would require three transformations on three periods of model data: GWL2, GWL3 and PRES. This would introduce a transformation error, as we have a limited historical data record of only 1959-2022. Moreover, for copula-based probability calculation, only ranks (hence quantiles) are used; therefore, there is no difference in the resulting probabilities in the analysis. We perform three statistical tests on temperature and precipitation between ERA5 and CRCM5-LE PRES on the sub-regional level. They confirm we can use the adjusted data, as it represents the univariate and bivariate structure well. We use 1,000 model years (50 members, 20 years each) for PRES, GWL2 and GWL3. This gives us a solid statistical estimation. Read more on the methodology in Methods.

We then calculate $p_{SK}$ of having an event of that magnitude. Therefore, we find the best fitting copula on the grid-cell-wise values and regionally averaged temperature and precipitation values. We plot the resulting $p_{SK}$ per year for the whole European domain in Fig. 3. The figure confirms the extreme rarity of 2003, as the probability is close to zero in the entire affected area. Moreover, to have an insight into which proportion of the domain is experiencing an extremely hot and dry event, we plot in red the fraction of land with a $p_{SK}$ value below the threshold of 10%, 5% and 1%. The $p_{SK}$ values for the latter case are calculated for every grid cell part of the domain. Also, 2003 is the most extreme event in this measure, with over 20%, 15% and 10% of the grid cells affected, respectively.

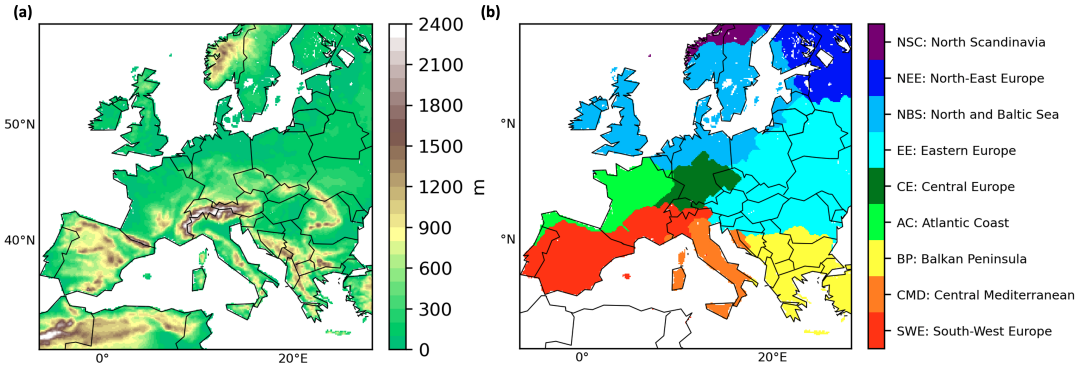


**Figure 2.** Study domain and sub-regions. The orography [m] over the European domain of the CRCM5-LE in 0.11 deg resolution in **(a)** and the sub-regions obtained by clustering events used in the analysis in **(b)**

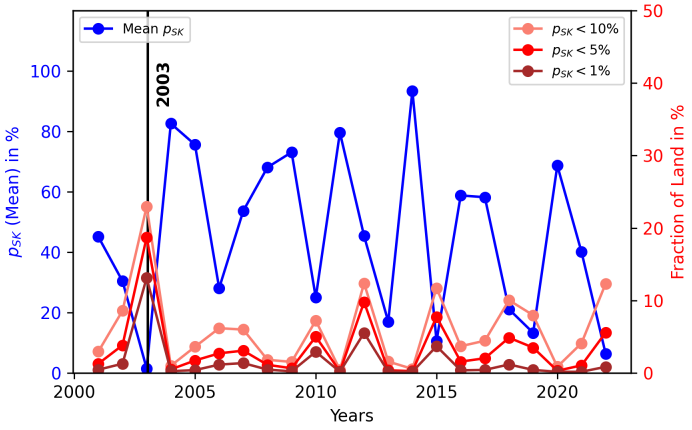


**Figure 3.** Mean $p_{SK}$ for the whole European domain for 2001-2022 in ERA5 and fraction of land affected by a compound hot and dry event per year. Mean $p_{SK}$ for averaged values of temperature and precipitation in blue. In red, the fraction of land is experiencing a $p_{SK}$ below a threshold noted in the legend. The vertical line highlights the event with the least $p_{SK}$ in black.

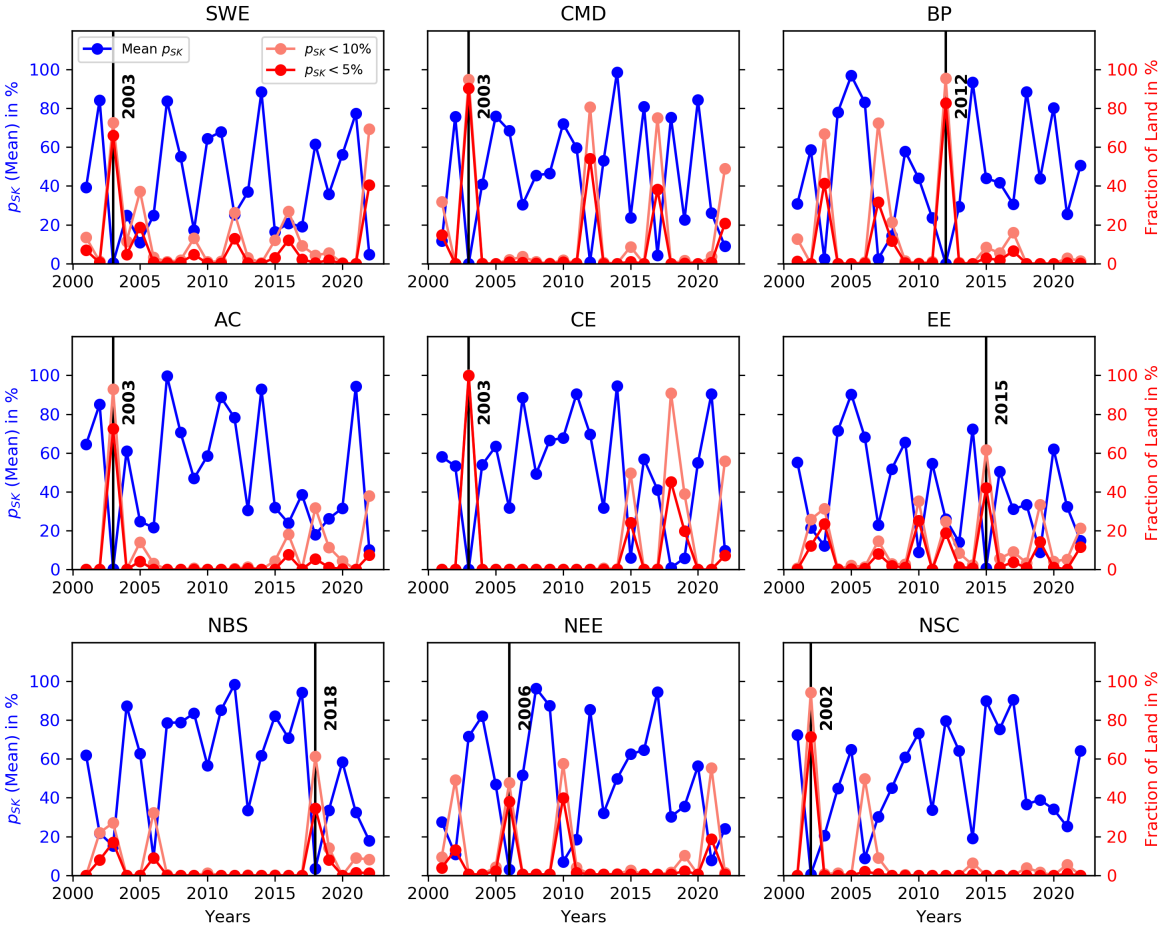


**Figure 4.** Mean $p_{SK}$ for the years 2001-2022 in ERA5 and fraction of land affected by a compound hot and dry event per year per region. Mean $p_{SK}$ averaged for every region per year in blue. Fraction of land experiencing an extremely rare hot and dry event per year in red. The vertical line highlights the event with the least $p_{SK}$ in black.

Additionally, we inspect the most extreme events on the regional scale. Therefore, we repeat the same analysis, averaging temperature and precipitation over the chosen sub-regions. The results can be seen in Fig. 4. We obtain that 2003 was also the most extreme in the southern parts of the domain, such as SWE, CMD, AC and CE. Moreover, we find 2012 for BP, 2015 in EE, 2018 in NBS, 2006 in NEE and 2002 in NSC as additional events to be inspected.

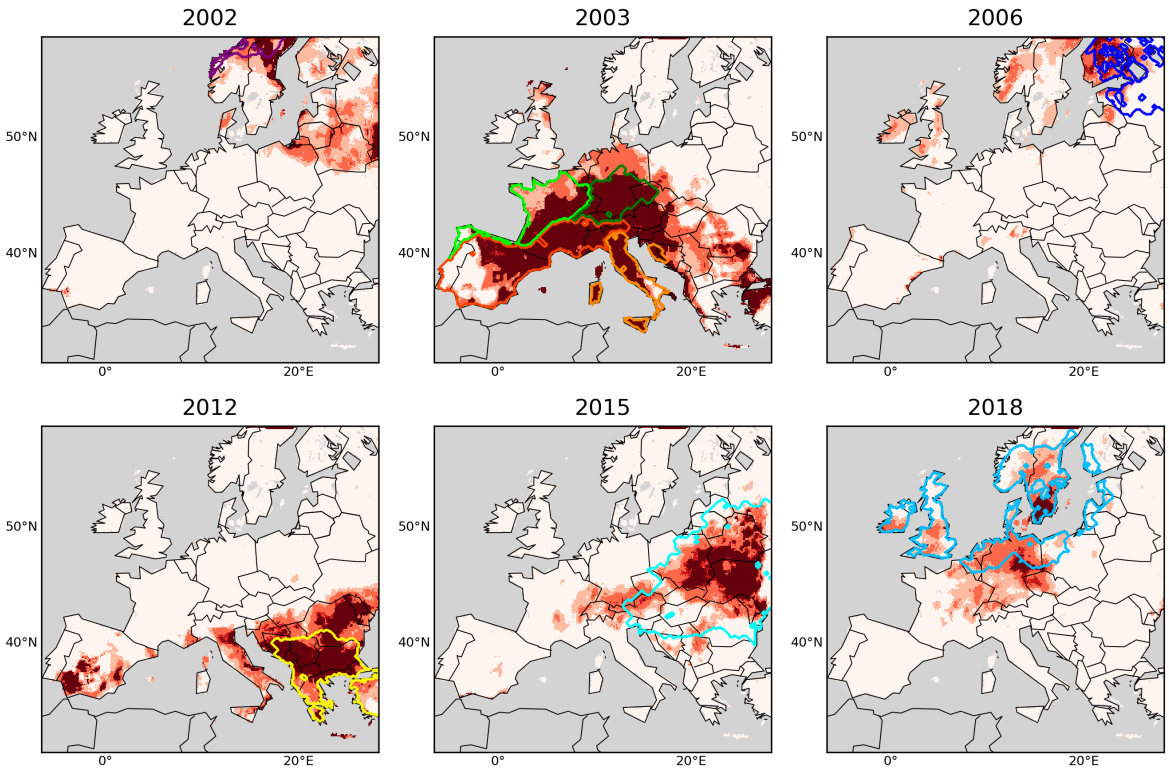


**Figure 5.** Spatial maps of $p_{SK}$ for the six most extreme compound hot and dry events on the European scale: 2002, 2003, 2006, 2012, 2015, 2018.

We plot the event maps for the six extraordinary hot and dry years on the European domain in Fig. 5. The displayed event maps confirm the extremity of the year 2003. No other event had a comparable spatial extent. We calculate the probabilities of occurrence in the three chosen periods for the areas where the events are most extreme. The probabilities in % for the affected regions are displayed in Table 1. The probabilities for all events are extremely low in the current climate, underlining the extraordinarity of the chosen events. In the present, the probability lies between 2.2% for NEE in 2006 and NBS in 2018 and 0.002 % for the 2012 event for the Balkan Peninsula. The estimated probabilities correspond to return periods between 45 and $>> 10,000$ years. Moreover, we inspect the change in probability to GWL2 and GWL3; we see a diverging pattern. Events of 2002, 2003 and 2018 seem to become a frequent feature of the future European climate as the event probabilities rise to 46% for GWL3, while other events like 2006, 2012 and 2015 experience only a slight rise in probability, remaining a rare feature of the climate in the future.

To better understand where the change in probability originates on the European scale, we look at the underlying distribution of temperature and precipitation in Fig. 6. In the following, we compare the events to future temperature and precipitation distributions under GWL2 and GWL3. We see that what used to be the most extreme events in the PRES period lie way closer to the centre of the distributions in the future periods of GWL2 and GWL3. The change is very pronounced, especially in temperature. The extreme temperatures during the 2003 event are below

| year | region | PRES | | GWL2 | | GWL3 | |
|---|---|---|---|---|---|---|---|
| | | $p_{SK}$ | $p_{tas}$/ $p_{pr}$ | $p_{SK}$ | $p_{tas}$/ $p_{pr}$ | $p_{SK}$ | $p_{tas}$/ $p_{pr}$ |
| **2002** | **NSC** | 0.1 | 0.1<br>27.5 | **5.5** | 4.6<br>25.4 | **17.1** | 21.8<br>21.3 |
| **2003** | **SWE** | 0.1 | 0.1<br>4.5 | 4.5 | 5.4<br>9.5 | **34.2** | 57.1<br>26 |
| **2003** | **CMD** | 0.4 | 0.1<br>6.9 | **12.6** | 3.9<br>11.7 | **46.0** | 47.3<br>22.7 |
| **2003** | **AC** | 0.1 | 0.1<br>33.1 | **6.4** | 4.7<br>44.5 | **35.9** | 34.6<br>66.2 |
| **2003** | **CE** | 0.1 | 0.1<br>0.5 | 3.4 | 1.0<br>2.1 | **11.9** | 13.3<br>4.9 |
| **2006** | **NEE** | 2.2 | 22.6<br>1.4 | 4.0 | 51.4<br>3.1 | **5.6** | 85.5<br>4.0 |
| **2012** | **BP** | 0.002 | 0.0<br>0.0 | 0.4 | 7.1<br>0.0 | 2.5 | 57.7<br>0.1 |
| **2015** | **EE** | 0.9 | 21.9<br>0.6 | 2.1 | 53.3<br>1.5 | 2.7 | 93.3<br>1.7 |
| **2018** | **NBS** | 2.2 | 1.4<br>3.9 | **9.0** | 21.5<br>5.1 | **18.8** | 66.7<br>6.9 |

**Table 1.** Event probability in % in regions of event occurrence for the three chosen periods. The probability corresponds to a $p_{SK}$ of a summer as hot and dry or hotter and drier than the historical event, Bold:$p_{SK} > 5\%$, return period $\leq 20$ years. $p_{tas}$ and $p_{pr}$ refer to univariate exceedance probabilities in temperature and precipitation.

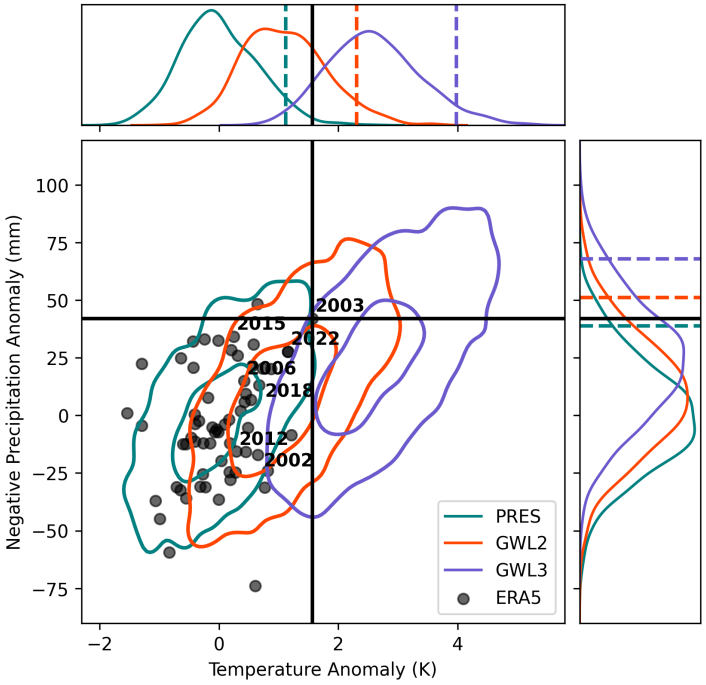


**Figure 6.** Bivariate distribution for average anomalies in temperature and negative precipitation for the three chosen periods for the whole European domain. CRCM5-LE PRES in dark green, GWL2 in orange and GWL3 in violet. The two lines encircle 95% and 50% of the data obtained by a kernel density estimate. ERA5 years 1959-2022 shown with black dots. Most extreme years from the previous analysis are notified with black text. The most extreme year of 2003 for the whole domain is highlighted with black horizontal and vertical lines. Marginal distributions of temperature anomaly for the three periods are displayed in the upper plot; negative precipitation anomaly in the right plot. Dashed lines in the marginal plots indicate the 95th percentile of the corresponding distribution.

the 95th percentile in GWL2 and would be considered a rather cold summer under GWL3. Although very much dominated by temperature, drying also contributes to the events' intensification. The dryness of 2003 will become less extreme and below the 95th percentile in negative precipitation (above the 5th percentile in precipitation) under GWL2 and GWL3. Hence, the event will become more common and less exceptional in the new climate, both in terms of temperature and precipitation.

## 2.2 Distributional shifts on the regional scale

Additionally, we investigate the changes in the distribution of temperature and precipitation anomaly on the regional scale. The results are displayed in Fig. 7. We also inspect the empirical univariate probabilities shown in Table 1 under $p_{tas}$ and $p_{pr}$.

In terms of general change in climatology, we observe a two-fold pattern. Firstly, in the northern regions, such as NBS, NSC and NEE, changes are primarily driven by shifts in temperature, while precipitation changes only a little. Meanwhile, in the South and Central European regions, including Southern Western Europe (SWE), Central Mediterranean (CMD), Balkan Peninsula (BP), Atlantic Coast (AC), Central Europe (CE), and Eastern Europe (EE), the climatic alterations are further compounded by a decrease in precipitation. This implies a more frequent occurrence of compound hot and dry extremes.

On the event level, we can now determine where the diverging intensification of event probabilities comes from. Some of the events like 2012 in BP and 2015 in EE were initially extreme in precipitation with the univariate probability $p_{pr}$ below 0.6% in PRES, as can be seen in the distribution plot in Fig. 7 and Table 1. Therefore, even under future climate conditions, those events are not to be expected more frequently than every 20 years or with a $p_{SK} <= 5\%$. Due to their initial intensity in temperature, all other events are expected to occur at least once or more often in the 20-year window under GWL3 climate conditions. Events like 2002, 2003 and 2018 are above the 5% threshold even under GWL2.

These findings shed light on the potential long-term changes in extreme hot and dry compound event probabilities, suggesting that certain events may define the "new normal" while others stay extreme in future expected climates.

## 2.3 Future compound hot and dry event climatology

We evaluated in previous sections that historical compound hot and dry events will become more frequent in the expected future climate. However, is it possible to find comparable compound hot and dry events to what is to be expected in future in the present in other regions? Therefore, we take the distribution of GWL3 compound hot and dry events and try to find the most similar region in the present climate regarding extreme temperature and precipitation distribution. By doing this, we can say that under GWL of $+3K$, compound hot and dry events, e.g., in South Germany, will be like they are nowadays in Eastern Europe in terms of values for temperature and precipitation. Therefore, we define the extreme event distribution by considering all events with $p_{SK} < 0.05$ and find the best match via the symmetrical Kullback-Leibler divergence[39]. We use the Kullback-Leibler divergence to measure the difference between two distributions. It is beneficial in our case, as it can be defined in bivariate terms—more on the definition in Methods.

The clustered regions include various climatic zones, e.g., the Alps and vast parts of Germany are part of the same cluster, as the clustering is optimised towards events happening simultaneously and not to regions with a similar hot and dry events distribution. Therefore, we chose a representative point/grid cell for the region in PRES time frame and use the event distribution at that point for further analysis. The representative grid cell is found via the smallest summed distance to all the other grid cells in the cluster. The distribution of the chosen representative points is shown in Fig. 8 **(a)**. NSC has the least dry and hot events compared to other regions' distribution. All the other regions are closer connected with the representative point in BP demonstrating the hottest and driest climate. The geographical location of the chosen representative points is displayed in Fig. 8 **(b)** in red.

Given the substantial climatic variations within geographical clusters, we choose to first calculate the matched climatologies of present-present (Figure 8 **(b)**) to then compare the results to the future (Figure 8 **(c)**). The comparison reveals a noteworthy northward shift of most hot and dry climatologies, as exemplified by BP, which is expected to cover most of Southern Europe and under GWL3 expands into more northern territories. Previously confined to

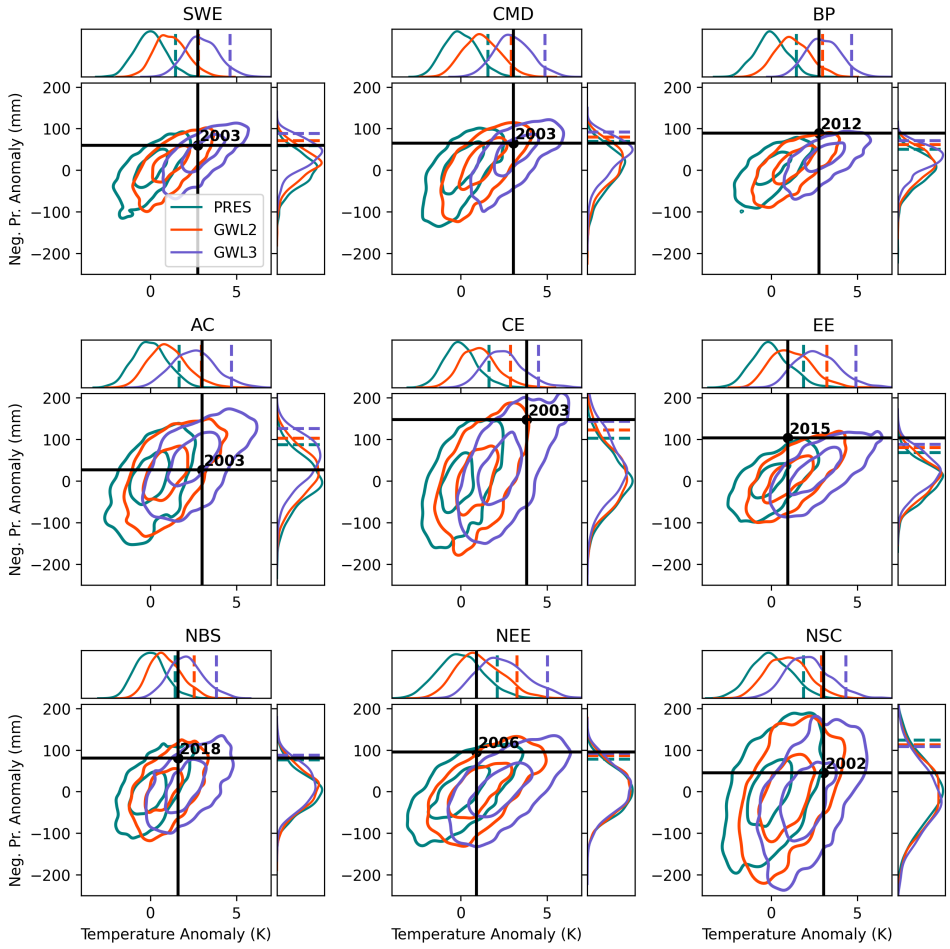


**Figure 7.** Bivariate distribution for average anomalies in temperature and negative precipitation for the three chosen periods for chosen sub-regions ordered from South to North from West to East. The most extreme ERA5-event based on $p_{SK}$ is highlighted with a black line.

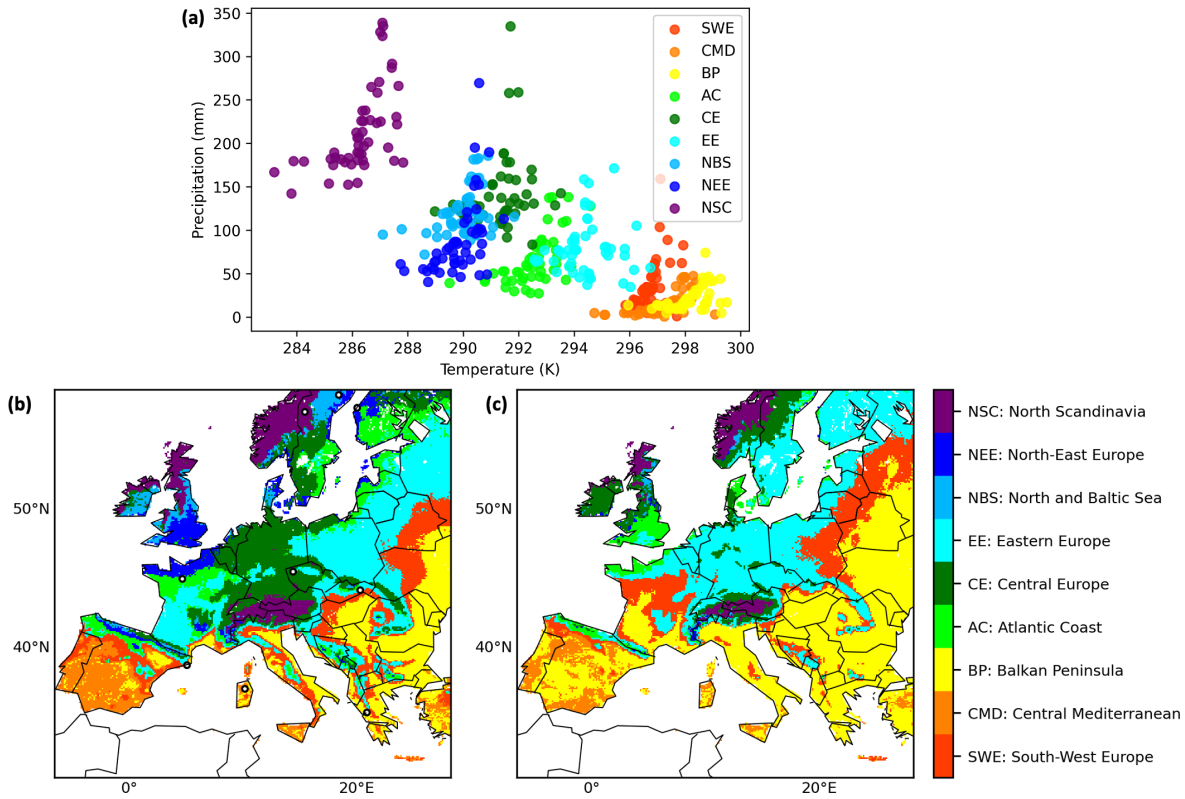


**Figure 8.** Changes in the climatology of compound hot and dry events from PRES to GWL3. **(a)** distribution of the chosen representative point for every geographical cluster in CRCM5-LE PRES data. **(b)** results of matching the representative point and grid cell values for CRCM5-LE PRES. Black-edged dots: location of the representative points. **(c)** results of grid-cell-wise matching between the event distribution in CRCM5-LE GWL3 and representative points in CRCM5-LE PRES.

the Balkan Peninsula, it extends into huge parts of the Iberian Peninsula, South France, Italy and Eastern Europe under GWL3. The area in the Alps, currently experiencing the relatively wet and cold NSC climate, shrinks, shifting in parts to a CE climate. Furthermore, the EE climate is gradually extending its presence into substantial parts of Central Europe and the Baltic Sea coast, including southern Sweden and Finland. These findings collectively indicate a significant northward shift in all hot and dry climate zones under GWL3. In Supplementary Fig. S1, we plot the most probable values for an event with $p_{SK} = 0.05$ for every region's chosen representative point. The plot confirms the shifts in climatic zones we see in Fig. 8, as, e.g., the GWL3 value of CE representative point is close to the PRES values of AC and EE. This is a change we can also see in Fig. 8 **(c)**.

## 3 Discussion

Our study highlights the significant implications of future climate conditions on compound hot and dry events in Europe. As a first step, we identify 9 European hot and dry events regions, where the events tend to happen simultaneously. For those regions, we identify the most extreme years out of the past 22 years of European climate: 2002 for NSC, 2003 for SWE, CMD, AC, CE; 2006 for NEE; 2012 for BP; 2015 for EE, and 2018 for CE. We confirm that the most extreme historical events will become more frequent features of the European climate under both scenarios - GWL2 and GWL3. For some events, the probability of occurrence rises 5-6 times from GWL2 to GWL3 and up to 46% under GWL3, translating into almost every second summer. This underlines the importance of sticking to the +2 degrees in mitigation policy. Moreover, we see a twofold pattern. Some past events are characterised by an extremity mainly in precipitation, as those in 2006, 2012, and 2015. Those events will experience a rise in frequency; however, they will remain extreme and rare even under future climate conditions for GWL2 and GWL3. Other past events such as 2002, 2003, and 2018 are projected to become immensely more frequent, with probabilities of up to 46%, as in the case of 2003 for CMD for GWL3. Additionally, our analysis reveals a significant northward shift in the climatology of hot and dry events under GWL3: the extremely hot and dry climate currently observed in BP is expected to extend into substantial parts of the spatial EE cluster, the relatively moist and cold NSC region in the Alps are contracting, while the EE compound event climate extends northward, reaching regions as far as the Baltic Sea and Scandinavia.

Our study contributes three key findings to the scientific field: (1) Our study shows the utility of a high-resolution large ensemble for compound event investigation, as it is better able to relate to past experienced events due to the amount of data and to resolve heterogeneities, e.g., for the investigation of the climatology shifts. We see the topography and the proximity to coastlines as governing effects on the hot and dry summer climates, which GCMs could not show due to the coarse resolution. (2) Our study presents a methodology to quantify probabilities of historical compound events with large ensembles, using an extreme event definition via Survival Kendall probability. (3) Last but not least, we see our contribution in presenting the shifts in the climatology of compound hot and dry events up to GWL3. The spatial analysis of hot and dry climate patterns and spatial shifts while evolving towards GWL3 are vital tools for illustrating the potential consequences of climate change. Only few studies have investigated changes in compound hot and dry events[13]. Investigating climatology changes of extreme events is best possible with the application of a SMILE, as no other datasets provide enough data on extremes. As such, this work represents a valuable contribution that can inform compound event research and enhance the communication efforts of stakeholders from science and policy.

A study by Ionita et al. confirms the chosen extreme compound hot and dry summers[33]: according to the study, 2003 ranks the hottest year over northern Spain, France, and south Germany; 2015 the driest over east Europe, 2012 the hottest in south-eastern Europe and 2006 and 2018 as one of the other extreme events in Europe. The correspondence is remarkable, as Ionita uses a different event definition, built out of criteria on the minimum length of the heatwave, temperature bigger than the 90th percentile, and three months' Standardized Precipitation Index smaller than minus one. The study confirms our event definition approach, as it uses a different dataset and a different definition of an event and arrives with a similar event set. The only exception is 2002, which is not included in the list of events. However, the NSC region is small; therefore, the results might vary on a differently cut domain.

To our knowledge, no studies exist that quantify the probability for all of the identified extreme hot and dry

compound events in bivariate terms. According to Charpentier et al.,[40] for 2003, the return period is estimated to be 115 to 37 years, corresponding to 0.8% to 2.7% in probability, depending on the model and heatwave definition. However, it must be noted that events were only defined in univariate terms in the study. Our results show a smaller probability, which is reasonable, as we include the extremity in precipitation. Rousi et al. find an increase in the probability of a summer like 2018 up to 96% in a $2^{\circ}C$ warmer world (corresponds in our definition to GWL2) by looking at the distribution of accumulated heat[3]. Our approach shows an eight-fold (2.2% to 18.8% ) increase in the probability of occurrence between PRES and GWL2 but reaches a return period of once in five years in contrast to almost every year (96% under GWL2). However, the bivariate event definition in our study goes beyond the univariate, purely temperature-based definition by Rousi et al. Given that the reported impacts of the 2018 event are linked to drought and heat[41,42], this underlines the importance of investigating the events in bivariate terms, as otherwise, there is a high chance of probability overestimation. A study by Boehnisch et al. confirms that sticking to GWL2 might reduce event occurrences multiple times[13]. The study also quantified compound hot and dry summer occurrences; however, using only CRCM5-LE[13]. Another study confirmed that changes in compound hot and dry event probability in Europe are driven by changes in temperature[43]. In terms of changes in climatic zones of hot and dry extremes, our study presents novel results. However, we can compare the spatial patterns and shifts to a study by Beck et al., which quantified the changes in climate by the end of the century using the Köppen-Geiger classification. This classification is based on monthly temperature and precipitation values and describes changes in the mean climate[44]. The general trend found by our study can be confirmed, disregarding the difference in the definitions we see the same effect for, e.g., the climate present in the Iberian Peninsula spreading into Central European regions.

The findings of this study should be viewed in the context of certain limitations. The reliance on a single large ensemble necessitates confidence in its representation of temperature and precipitation, as model and performed data adjustments biases might occur. Therefore, inspecting other high-resolution large ensembles in future research might be necessary. Focusing on two variables alone means that events may also have been extreme in other variables, e.g. wind, which could affect their probability. Moreover, the probabilities always rely on the event definition, and different event definitions will undoubtedly lead to other probabilities and return periods for event occurrence. A model-related limitation lies within the missing coupling of the dynamical downscaling between the RCM and the GCM. The high-resolution CRCM5-LE does not feed back into the driving CanESM2. Therefore, the higher-resolution land-atmosphere interaction cannot affect the boundary conditions of the driving climate model. Moreover, the GWL approach assumes independence from the chosen model scenario, which must be validated further.

Looking forward, there are several avenues for future research. In terms of compound hot and dry events, we have estimated their occurrence probabilities in future. However, there is still a need for research on future factors leading to extreme event formation and evolution. Our study on probability estimation can serve as a blueprint for analysing other compound extreme events. Furthermore, there is potential in the analysis of shifts in climate zones, which could be applied to different types of extreme events as well as to studies on the mean climate.

# 4 Methods

## 4.1 Bivariate probability assessment based on Survival Kendall probability and Copulas

In this study, we are investigating hot and dry extremes. We characterise the events by temperature and precipitation. Those variables are highly interrelated. Therefore, we cannot calculate the probability in univariate terms, as this would disregard their bivariate structure[34,35]. Consequently, we use Copulas to describe the interrelation of the variables. Moreover, for the event definition, we use the "Survival Kendall" hazard scenario[36], which has previously proven its usefulness for the application on extreme heat and dryness[35]. This approach divides the event probability space into "safe" and potentially dangerous events. All events on the separating line share the same probability. We calculate our Survival Kendall probability using the following relationship:

$$p_{SK} = Pr(\hat{C}(\bar{F}_X(x), \bar{F}_Y(y)) \geq t) \tag{1}$$

where $\bar{F}_X(x)$ and $\bar{F}_Y(y)$ are the marginal survival functions of the two variables, t the critical probability and $\hat{C}$ the survival copula. $\bar{F}_X(x) = 1 - F_X(x)$ and $\bar{F}_Y(x) = 1 - F_Y(y)$, where $F_X(x)$ and $F_Y(y)$ correspond to the marginal

cumulative distribution functions.

With the application on compound dry and hot extremes, we follow a previous study on the 2014 Californian drought[35].

We use the VineCopula package in R[45] for the calculation. Therefore, we first estimate the marginal probability functions and then transform those to uniform distributions. For precipitation, we use empirical probability functions. For temperatures above the 95-th percentile, we use a Generalized Pareto distribution fit to get a higher accuracy for rare events. Below the 95th percentile, we are using an empirical probability distribution. Different copula families exist that could be applied to bivariate structures. We fit 15 different copula families and use the Bayesian information criterion (BIC) to choose the best-fitting copula. All selected copulas pass the goodness-of-fit test based on Kendall's process[46,47] on 0.05 level.

### 4.2 Data sets

Assessing the occurrence probabilities of rare events requires an abundant database. To do this, we use seasonally averaged temperature and precipitation for June-July-August (JJA) originating from a single-model initial condition large ensemble (SMILE), the Canadian Regional Climate Model 5 Large Ensemble (CRCM5-LE). The CRCM5-LE has been produced within the scope of the ClimEx project[32] on a European and a North American domain at a spatial resolution of $0.11^\circ$ (12km) for the years 1950-2099. We use the European domain for our analysis, which is displayed in Fig. 2. In the data production, dynamical downscaling via CRCM5-LE is applied on the global atmospheric 50-member initial-condition model - Canadian Earth System Model 2 (CanESm2)[48]. After a few years, the 50 members are losing their dependency due to the chaotic nature of the weather while maintaining its greenhouse gas forcing. The model is driven by historical greenhouse gas emissions up to 2005. Starting from the year 2006, the RCP8.5 forcing scenario is used. For the analysis, we are using the data from the three time periods: 2001-2020 (PRES), 2021-2040 (GWL2) and 2042-2061 (GWL3). We use 20 years from the 50 members per period, which results in 1,000 years per period available for analysis.

To assess the observed hot and dry events, we use the fifth generation of the ECMWF atmospheric reanalysis ERA5[31]. We use linear interpolation on the European domain to make the data comparable with the grid from CRCM5-LE. For the fits, we are using the data for 1959-2022.

Temperatures in both datasets are linearly detrended for each period. As there is a model bias, we transfer ERA5 into the model world by transferring it according to its quantile and perform two tests to evaluate how well the representation is: we test via Kolmogorov-Smirnov, if both marginal distributions are well represented[49], and we perform a test of the bivariate structures via the TwoCop test in R. The latter test is developed by Rémillard and Scaillet,[50]. It is designed to assess whether the empirical copulas between two distributions could belong to the same underlying distribution. Therefore, we can use it to evaluate whether they significantly deviate from one another. All the statistical tests have been performed on the $\alpha = 0.05$ significance level. We follow the methodology of Zscheischler et al.[34].

### 4.3 Global Warming Levels

Global Warming Levels have proven useful in science communication for calculating climate change-induced impacts with relevant levels for policy and decision-makers in various fields. The approach is less dependent on the choice of model and scenario. GWLs are calculated as mean surface air temperature anomalies referencing the historical conditions 1850-1900[29]. They refer to the 20 years from the first year where the intended level was measured.

For our analysis, we follow the results of a previous study[13], that used the anomalies from the CanESM2 driving model from 1850-1900 and use the same three periods:

1. Present period (PRES) 2001-2020 with $GWL = +1.2^\circ C$
2. Future I period with $GWL = +2^\circ C$ (GWL2)
3. Future II period with $GWL = +3^\circ C$ (GWL3)

### 4.4 Finding regions where compound hot and dry events happen via Clustering

We apply the clustering technique to obtain regions to analyse hot and dry events. We aim to have regions experiencing events simultaneously, with a high spatial and temporal interrelation in event occurrence. This study uses the yearly and per grid cell calculated $p_{SK}$-probability of event occurrence as input to the agglomerative hierarchical clustering algorithm. We define the distance measure for the algorithm as

$$d(\mathbf{r},\mathbf{q}) = 1 - cs(\mathbf{r},\mathbf{q}) \tag{2}$$

where $\mathbf{r}$, $\mathbf{q}$ refer to two spatial events for which we are calculating the distance and cs to the cosine similarity measure[51,52]. We only include grid cells with $p_{SK} < 0.1$ for the clustering to cluster only on extreme occurrences. We estimate the optimal number of clusters by applying the elbow method[53]. We apply a majority cluster filtering method introduced in a previous study[52] and arrive at 9 clusters used as regions for the hot and dry compound event probability quantification. The resulting clusters are displayed in Fig. 2, right.

### 4.5 Estimating future compound hot and dry events climatology

To estimate the future compound hot and dry event climatology, we take all the events from the PRES and GWL3 period and filter $p_{SK} > 0.05$ to obtain the distribution of hot and dry compound events per grid cell.

For the construction of a distance measure, we use the Kullback-Leibler divergence[39]. The study by Perez et al. showed that it is possible to calculate the measure based on an empirical estimation from the samples using KD trees and nearest neighbours[39]. We use the Python implementation from Robert Kern[54]. As the Kullback-Leibler divergence is a non-symmetric measure, we symmetrise it by calculating the distance $D$ via: $D = 0.5KL(p,q) + 0.5KL(q,p)$, where $p$ and $q$ are bivariate distributions of two different grid cells.

To find the representative points for each cluster, we calculate the sum of the distance of every point to every other point in the spatial cluster. The point with the smallest summed distance is taken as a representative point for the spatial cluster. The distribution of the chosen points is shown in Fig. 8.

## DATA AVAILABILITY

The data is retrieved from the CRCM5-LE via https://www.climex-project.org/en/data-access/. The ERA5 data set was downloaded from https://cds.climate.copernicus.eu/. Derived data supporting the findings of this study are available from the corresponding author upon reasonable request.

## CODE AVAILABILITY

All codes used are direct implementations of standard methods, packages and techniques, described in detail in Methods. The code used to prepare the figures in this paper is available from the corresponding author upon reasonable request.


## ACKNOWLEDGMENTS

The CRCM5-LE was created within the ClimEx project, which was funded by the Bavarian State Ministry for the Environment and Consumer Protection. Computations of the CRCM5-LE were made on the SuperMUC supercomputer at Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities. We acknowledge Environment and Climate Change Canada for providing the CanESM2-LE driving data.


## COMPETING INTERESTS

The authors declare no competing financial or non-financial Interests.

## AUTHOR CONTRIBUTIONS

E.F. designed the research under the supervision of R.L.. E.F. , A.B. and B.P. regularly feedbacked the results and provided ideas for further development. E.F. performed the data analysis and led the writing of the paper. E.F. prepared all figures. All authors contributed to the interpretation of the findings and revision of the paper.

## References


1. Zscheischler, J. *et al.* Future climate risk from compound events. *Nat. Clim. Chang.* **8**, 469–477 (2018).
2. Masson-Delmotte, V. *et al.* Climate change 2021: the physical science basis. *Contribution working group I to sixth assessment report intergovernmental panel on climate change* **2** (2021).
3. Rousi, E. *et al.* 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, DOI: 10.5194/nhess-23-1699-2023 (2023).
4. Sedlmeier, K., Feldmann, H. & Schädler, G. Compound summer temperature and precipitation extremes over central europe. *Theor. applied climatology* **131**, 1493–1501 (2018).
5. Robine, J.-M., Cheung, S. L., Le Roy, S., Van Oyen, H. & Herrmann, F. R. Report on excess mortality in europe during summer 2003. *EU Community Action Programme for Public Heal. Grant Agreem.* **2005114**, 28 (2007).
6. Robine, J.-M. *et al.* Death toll exceeded 70,000 in europe during the summer of 2003. *Comptes rendus biologies* **331**, 171–178 (2008).
7. Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M. & García-Herrera, R. The hot summer of 2010: redrawing the temperature record map of europe. *Science* **332**, 220–224 (2011).
8. Zscheischler, J. & Seneviratne, S. I. Dependence of drivers affects risks associated with compound events. *Sci. Adv.* **3**, e1700263, DOI: 10.1126/sciadv.1700263 (2017). https://www.science.org/doi/pdf/10.1126/sciadv.1700263.
9. Miralles, D. G., Gentine, P., Seneviratne, S. I. & Teuling, A. J. Land–atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges. *Annals New York Acad. Sci.* **1436**, 19 (2019).
10. O, S. *et al.* The role of climate and vegetation in regulating drought-heat extremes. *J. Clim.* **1**, 1–21 (2022).
11. Vogel, M. M., Zscheischler, J. & Seneviratne, S. I. Varying soil moisture–atmosphere feedbacks explain divergent temperature extremes and precipitation projections in central europe. *Earth Syst. Dyn.* **9**, 1107–1125, DOI: 10.5194/esd-9-1107-2018 (2018).
12. Seneviratne, S. I. *et al.* Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Rev.* **99**, 125–161 (2010).
13. Boehnisch, A., Felsche, E., Mittermeier, M., Poschlod, B. & Ludwig, R. Future hotspots of compound dry and hot summers emerge in european agricultural areas. *Authorea Prepr.* (2023).
14. Trenberth, K. E. & Shea, D. J. Relationships between precipitation and surface temperature. *Geophys. Res. Lett.* **32**, DOI: https://doi.org/10.1029/2005GL022760 (2005).
15. Thompson, V. *et al.* The most at-risk regions in the world for high-impact heatwaves. *Nat. Commun.* **14**, 2152 (2023).
16. Böhnisch, A., Mittermeier, M., Leduc, M. & Ludwig, R. Hot spots and climate trends of meteorological droughts in europe–assessing the percent of normal index in a single-model initial-condition large ensemble. *Front. Water* **3**, DOI: 10.3389/frwa.2021.716621 (2021).
17. Almendra-Martn, L. *et al.* Analysis of soil moisture trends in europe using rank-based and empirical decomposition approaches. *Glob. Planet. Chang.* **215**, 103868, DOI: https://doi.org/10.1016/j.gloplacha.2022.103868 (2022).
18. Perkins-Kirkpatrick, S. E., Fischer, E. M., Angélil, O. & Gibson, P. B. The influence of internal climate variability on heatwave frequency trends. *Environ. Res. Lett.* **12**, 044005 (2017).
19. Owen, A. L., Conover, E., Videras, J. & Wu, S. Heat waves, droughts, and preferences for environmental policy. *J. Policy Analysis Manag.* **31**, 556–577 (2012).

**20.** Bittner, M.-I., Matthies, E. F., Dalbokova, D. & Menne, B. Are european countries prepared for the next big heat-wave? *The Eur. J. Public Heal.* **24**, 615–619 (2014).

**21.** Kromp-Kolb, H. *et al.* Climate change: Assessment of the vulnerability of nuclear power plants and approaches for their adaptation. Tech. Rep., Organisation for Economic Co-Operation and Development (2021).

**22.** Rossi, G., Cancelliere, A. & Giuliano, G. Case study: multicriteria assessment of drought mitigation measures. *J. Water Resour. Plan. Manag.* **131**, 449–457 (2005).

**23.** Lowe, D., Ebi, K. L. & Forsberg, B. Heatwave early warning systems and adaptation advice to reduce human health consequences of heatwaves. *Int. journal environmental research public health* **8**, 4623–4648 (2011).

**24.** Ward, K., Lauf, S., Kleinschmit, B. & Endlicher, W. Heat waves and urban heat islands in europe: A review of relevant drivers. *Sci. Total. Environ.* **569**, 527–539 (2016).

**25.** Habibi, S. & Asadi, N. Causes, results and methods of controlling urban sprawl. *Procedia Eng.* **21**, 133–141 (2011).

**26.** Deser, C. *et al.* Insights from earth system model initial-condition large ensembles and future prospects. *Nat. Clim. Chang.* **10**, 277–286 (2020).

**27.** Bevacqua, E. *et al.* Advancing research on compound weather and climate events via large ensemble model simulations. *Nat. Commun.* **14**, 2145, DOI: 10.1038/s41467-023-37847-5 (2023).

**28.** Poschlod, B., Zscheischler, J., Sillmann, J., Wood, R. R. & Ludwig, R. Climate change effects on hydrometeorological compound events over southern norway. *Weather. Clim. Extrem.* **28**, 100253 (2020).

**29.** Seneviratne, S., Zhang, X., Adnan, M. *et al.* Chapter 11: weather and climate extreme events in a changing climate. climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change, vol (2021).

**30.** Gampe, D. *et al.* Applying global warming levels of emergence to highlight the increasing population exposure to temperature and precipitation extremes. *EGUsphere* **2023**, 1 (2023).

**31.** Hersbach, H. e. a. The era5 global reanalysis. *Q. J. Royal Meteorol. Soc.* DOI: 10.1002/qj.3803 (2020).

**32.** Leduc, M. *et al.* The ClimEx Project: A 50-Member Ensemble of Climate Change Projections at 12-km Resolution over Europe and Northeastern North America with the Canadian Regional Climate Model (CRCM5). *J APPL METEOROL CLIM* **58**, 663–693, DOI: 10.1175/JAMC-D-18-0021.1 (2019).

**33.** Ionita, M., Caldarescu, D. E. & Nagavciuc, V. Compound hot and dry events in europe: variability and large-scale drivers. *Front. Clim.* **3**, 688991 (2021).

**34.** Zscheischler, J. & Fischer, E. M. The record-breaking compound hot and dry 2018 growing season in germany. *Weather. Clim. Extrem.* **29**, 100270 (2020).

**35.** AghaKouchak, A., Cheng, L., Mazdiyasni, O. & Farahmand, A. Global warming and changes in risk of concurrent climate extremes: Insights from the 2014 California drought: Global Warming and Concurrent Extremes. *Geophys. Res. Lett.* **41**, 8847–8852, DOI: 10.1002/2014GL062308 (2014).

**36.** Salvadori, G., Durante, F., De Michele, C., Bernardi, M. & Petrella, L. A multivariate copula-based framework for dealing with hazard scenarios and failure probabilities. *Water Resour. Res.* **52**, 3701–3721 (2016).

**37.** Zscheischler, J. & Lehner, F. Attributing compound events to anthropogenic climate change. *Bull. Am. Meteorol. Soc.* **103**, E936–E953 (2022).

**38.** Salvadori, G., Durante, F. & De Michele, C. Multivariate return period calculation via survival functions. *Water Resour. Res.* **49**, 2308–2311, DOI: 10.1002/wrcr.20204 (2013).

**39.** Pérez-Cruz, F. Kullback-leibler divergence estimation of continuous distributions. In *2008 IEEE international symposium on information theory*, 1666–1670 (IEEE, 2008).

**40.** Charpentier, A. On the return period of the 2003 heat wave. *HAL, Work. Pap.* **109**, DOI: 10.1007/s10584-010-9944-0 (2010).

**41.** Bastos, A. *et al.* Direct and seasonal legacy effects of the 2018 heat wave and drought on european ecosystem productivity. *Sci. Adv.* **6**, eaba2724, DOI: 10.1126/sciadv.aba2724 (2020). https://www.science.org/doi/pdf/10.1126/sciadv.aba2724.

**42.** Dirmeyer, P. A., Balsamo, G., Blyth, E. M., Morrison, R. & Cooper, H. M. Land-atmosphere interactions exacerbated the drought and heatwave over northern europe during summer 2018. *AGU Adv.* **2**, e2020AV000283 (2021).

**43.** Manning, C. *et al.* Increased probability of compound long-duration dry and hot events in europe during summer (1950–2013). *Environ. Res. Lett.* **14**, 094006 (2019).

**44.** Beck, H. E. *et al.* Present and future köppen-geiger climate classification maps at 1-km resolution. *Sci. data* **5**, 1–12 (2018).

**45.** Nagler, T. *et al.* Vinecopula: statistical inference of vine copulas. *R package version* **2** (2019).

**46.** Wang, W. & Wells, M. T. Model selection and semiparametric inference for bivariate failure-time data. *J. Am. Stat. Assoc.* **95**, 62–72 (2000).

**47.** Genest, C., Quessy, J.-F. & Rémillard, B. Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scand. J. Stat.* **33**, 337–366 (2006).

**48.** Martynov, A. *et al.* Reanalysis-driven climate simulation over CORDEX North America domain using the Canadian Regional Climate Model, version 5: model performance evaluation. *CLIM DYNAM* **41**, 2973–3005, DOI: 10.1007/s00382-013-1778-9 (2013).

**49.** Berger, V. W. & Zhou, Y. Kolmogorov–smirnov test: Overview. *Wiley statsref: Stat. reference online* (2014).

**50.** Rémillard, B. & Scaillet, O. Testing for equality between two copulas. *J. Multivar. Analysis* **100**, 377–386 (2009).

**51.** Cheng, X. & Wallace, J. M. Cluster analysis of the northern hemisphere wintertime 500-hpa height field: Spatial patterns. *J. atmospheric sciences* **50**, 2674–2696 (1993).

**52.** Felsche, E., Böhnisch, A. & Ludwig, R. Inter-seasonal connection of typical european heatwave patterns to soil moisture. *npj Clim. Atmospheric Sci.* **6**, 1 (2023).

**53.** Jung, Y., Park, H., Du, D.-Z. & Drake, B. L. A decision criterion for the optimal number of clusters in hierarchical clustering. *J. Glob. Optim.* **25**, 91–111 (2003).

**54.** Kern, R. Kullback-leibler divergence. https://gist.github.com/atabakd/ed0f7581f8510c8587bc2f41a094b518 (2011).

**Will Present-Day Compound Hot and Dry European Summers Still Be Extreme in the Future?, E. Felsche et al.**

***Supplementary material***

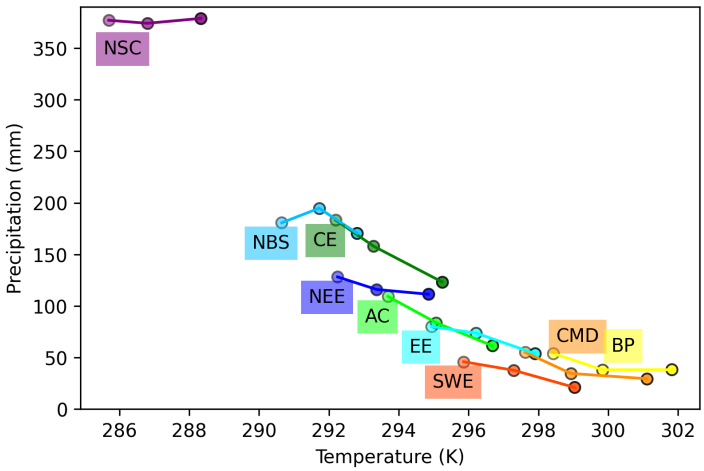


**Figure S1.** Most probable values for $p_{SK} = 0.05$ in temperature and precipitation for the chosen representative points in the nine sub-regions for the three periods: PRES (left, most transparent), GWL2 (middle, medium transparent), GWL3 (right, non-transparent)

# 5. Synthesis

In the following the key findings of the three studies are formulated as answers to the research questions introduced in section 3. The overarching research question providing the basis for the entire thesis is answered below.

## Paper I: Drought Prediction

*Q1.1: What is the potential of machine learning in drought prediction? What opportunities for research does it offer?*

The methodology applied in Paper I showed that ANNs can be a useful tool for investigating the predictability of extreme events like droughts. The trained models achieve an accuracy of 55% and 57% for the Munich and Lisbon domains, respectively. The study found the best-performing architecture and parameters for each of the domains. The accuracy shows that there is relevant information that allows for correct prediction in more than half of the cases. Potentially better results can be achieved when applying larger and more sophisticated models. The methodology allows the inclusion of a big set of variables for prediction and the exploration which of those offer the most relevant information via explainable AI methods. While physics-based evaluation of the variables potentially important for the prediction has to be done before the training, only by analysing the contribution of the variables will we know which information is used by the trained model for the correct prediction.

*Q1.2: What are the relevant variables for drought prediction in Lisbon? What are the relevant variables for drought prediction in Munich?*

Paper I provides an analysis of the variables most important for correct prediction. For the Lisbon domain, we find only variables one month before the event in the top five predictors. Sea level pressure (*psl*), surface air pressure (*ps*), and NAO are the top three relevant variables with the highest contribution, therefore showing the strong influence of the atmospheric pressure system

on drought formation in Lisbon. The first two relate to local conditions, while NAO refers to the general state of the atmosphere over the Atlantic Ocean and Europe. The next two variables for the Lisbon domain with the strongest contribution to the prediction are northward near-surface wind (*vas*) and evaporation (*evspsbl*). In the city, northern wind prevails; therefore, the variable *vas* is closely connected to the presence of wind in general (Alcoforado et al., 2006). As the whole Iberian Peninsula is a moisture-limited system, the presence of the wind is connected to the presence of convective clouds that are built in other areas and can discharge over Lisbon. *evspsbl* has a very direct influence on the formation of drought given that if evaporation is getting lower, the probability of formation of rain clouds also decreases (Sheffield and Wood, 2011).

For the Munich domain, the highest contribution is found for the following variables: NAO1, *psl*1, EAWR5, *ps*1 and SCA5. While NAO, *psl* and *ps* refer to the atmospheric pressure, similar to Lisbon, it is particularly interesting that there is an influence on drought formation five months before the event from teleconnections indices EAWR and SCA. The study finds that EAWR and SCA five months before the event have the most influence on the prediction of droughts happening in spring and summer, therefore referring to atmospheric conditions in winter. Both patterns indicate anticyclonic conditions over Central Europe. This can be confirmed by Träger-Chatterjee et al., who found that low precipitation amounts and high solar radiation in previous late winter and spring are connected to dry and hot summers in Central Europe.

*Q1.3: What role does seasonality play in the prediction? Are there particular seasons where droughts are better predictable?*

For the Munich domain, better drought predictability is given for spring, fall and summer months, while the classifier shows lower Precision and Recall in winter. In contrast, for the Lisbon domain, the lowest predictability is found in summer. For Lisbon, higher overall accuracy, precision and recall are given compared to Munich.

## Paper II: Heatwave patterns and soil moisture influence

*Q2.1: How can we identify typical patterns of heatwaves in Europe with the help of machine-learning-powered clustering?*

In Paper II, nine dominant patterns of heatwaves in Europe are found: Iberian Peninsula (IP), West-

ern Europe 1 and 2 (WE1 and WE2), Britain and Ireland(BI), South-Eastern Europe (SEE), Greece and South Italy (GSI), Scandinavia (SCA), Central-Eastern Europe (CEE) and North-Eastern Europe (NEE). Those patterns correspond well with those gained from observational data and when compared to previous studies and historical extreme heatwave events. Hierarchical agglomerative clustering in combination with cosine similarity has proven to be a useful technique for clustering. As some of the events in the dataset have rare spatial features, a majority cluster filtering was introduced that excluded events dissimilar to other data. The excluded events constitute $< 0.1\%$ of the data.

*Q2.2: How does a soil moisture deficit in spring influence the occurrence of a heatwave in summer in the identified regions?*

The study identifies a significant relationship between a soil moisture deficit and precipitation deficit in the preceding winter/spring (JFMA) and the number of heatwave days in summer in Southern Europe (GSI, SEE, IP, WE2). Southern Europe is a moisture-limited region. Multitude of positive feedbacks like reduced formation of convective clouds and increased solar radiation contribute to maintaining the drought in those regions and amplifying the coupling between drought and heat, which is transferred to the following season (Quesada et al., 2012). This finding is confirmed by previous studies that outlined the governing role of spring soil moisture over the Mediterranean on the heatwave formation in summer (Zampieri et al., 2009).

*Q2.3: What is the influence of a heatwave in summer on the soil moisture deficit in the following fall for each of the identified regions?*

The analysis of soil moisture anomalies in the following season (OND) shows a significant negative relationship for SCA, SEE and NEE. This indicates that long heatwaves in summer lead to a substantial soil moisture deficit preserved until the following season in those regions. SCA, NEE and SEE in part are energy-limited evaporation regimes showing a higher overall variability of soil moisture. Therefore, the study shows that a soil moisture deficit that builds up in summer takes longer to return to normal levels in the following spring. This finding is new to our knowledge.

# Paper III: Compound hot and dry events

*Q3.1: How can we identify typical patterns of compound hot and dry events in Europe with the help of machine-learning-powered clustering? What are the most extreme historical compound hot and dry events in those regions?*

Publication III highlights the significant implications of future climate conditions on compound hot and dry events in Europe. In Paper III, we show that the clustering methodology from Paper II is applicable to compound hot and dry events. We find nine typical patterns of compound hot and dry events in Europe: South-West Europe (SWE), Central Mediterranean (CMD), Balkan Peninsula (BP), Atlantic Coast (AC), Central Europe (CE), Eastern Europe (EE), North and Baltic Sea (NBS), North-East Europe (NEE) and North Scandinavia (NSC). The probability is defined in bivariate terms via copulas and the Survival Kendall probability definition. We identify 2003 as the most extreme compound hot and dry summer on the European scale and in the following regions: SWE, CMD, AC and CE. 2012 is the hottest summer for BP, 2015 for EE, 2018 for NBS, 2006 for NEE and 2002 for NSC.

*Q3.2: How will the probability of historical compound hot and dry events change under the influence of climate change under two possible scenarios of GWL2 and GWL3?*

The study results show that the most extreme historical events will become more frequent features of the European climate under both scenarios - GWL2 and GWL3. For some events, the probability of occurrence rises 5-6 times from GWL2 to GWL3 and up to 46% under GWL3, translating into almost every second summer. This underlines the importance of sticking to the +2 degrees policy in regulation. Especially past events of 2002, 2003, and 2018 with an initial extremity in temperature are projected to become immensely more frequent, with probabilities of up to 46%, as in the case of 2003 for Central Mediterranean for GWL3.

*Q3.3: How will the climatology of compound hot and dry events change under future climate?*

Publication III performs an analysis of the changes in the climatology of compound hot and dry events by comparing the current climate to the future climate of hot and dry events under GWL3. The study shows a northward shift in the climatology of hot and dry events under GWL3: the hot and dry climate currently observed in the Balkan Peninsula is expected to extend into substantial parts of Eastern Europe. The study also shows that the moist and cold regions in the Alps are

shrinking under future climate and that the Baltic Sea and Scandinavia will experience the hot and dry events climate now present in Eastern Europe.

## Overarching research question

*Can machine learning approaches facilitate the research on prediction and interrelation of heatwaves and droughts in Europe?*
*If yes, what can we learn from it about the physical nature of the events and the expected changes in the future?*

This thesis illustrates throughout the three publications that machine learning is a helpful tool in the investigation of heatwaves and droughts in Europe. Publication I demonstrates a framework for machine-learning-powered drought prediction for two locations: Munich and Lisbon. Better prediction accuracy is achieved for the Lisbon domain, which can be an indication of better overall predictability in Lisbon. From the trained models, we then learn that global (NAO) and local air pressure variables (*ps, psl*) one month before the event provide skill to the model in both domains. Additionally, in Lisbon northward near-surface wind (*vas*) one month before the event, in Munich SCA and EAWR five months before the event, are variables important for the prediction.

Papers II and III show that clustering, an unsupervised machine learning approach, can be a helpful tool for a meaningful dimensionality reduction of spatial events. The research identifies nine typical patterns of heatwaves in Europe and nine typical patterns of compound heatwaves and drought in Europe. Paper II uses those patterns to investigate the interrelation of heatwaves with a soil moisture deficit in those regions in the previous spring and the following fall. Paper III uses the identified patterns to study the probability of occurrence of most extreme past compound hot and dry events. The research shows an immense increase in the probability of future compound hot and dry events. In many cases, there is a doubling in event probabilities derived for GWL2 to GWL3.

Machine Learning approaches are especially useful when, as in our case, working with a large dataset of events provided by a SMILE.

This thesis has shown that machine learning and artificial intelligence approaches have huge potential in investigating drivers of extremes with the application on heatwaves and droughts solely, as well as on compounding events. Other studies confirmed the potential of machine learning for

enhanced understanding of drivers of other hydro-meteorological extreme events like freezing rain (Mittermeier et al., 2022), floods (Mosavi et al., 2018), as well as demonstrated the usability of machine learning for applications on prediction of impacts like crop yield (Schmidt and Felsche, 2024).

Much of the future potential in climate research lies in artificial intelligence approaches that can enhance the climate models and offer valuable analysis tools for observational and model data investigation. The key to a successful application lies in integrating domain knowledge and meteorological knowledge into the machine learning model. As machine learning is a black-box method, the key challenge lies in the interpretation of the results. Explainable and interpretable AI approaches applied in this thesis allow for understanding the results and learning about the investigated events.

# References


Ahmed, K., Sachindra, D., Shahid, S., Iqbal, Z., Nawaz, N., and Khan, N. (2020). Multi-model ensemble predictions of precipitation and temperature using machine learning algorithms. *Atmospheric Research*, 236:104806.

Alcoforado, M.-J., H, A., Lopes, A., Vasconcelos, J., and Vieira, R. (2006). Observational studies on summer winds in Lisbon (Portugal) and their influence on daytime regional and urban thermal patterns. *Merhavim*, 6.

Asthana, T., Krim, H., Sun, X., Roheda, S., and Xie, L. (2021). Atlantic hurricane activity prediction: A machine learning approach. *Atmosphere*, 12(4).

Baatz, R., Hendricks Franssen, H., Euskirchen, E., Sihi, D., Dietze, M., Ciavatta, S., Fennel, K., Beck, H., De Lannoy, G., Pauwels, V., et al. (2021). Reanalysis in earth system science: Toward terrestrial ecosystem reanalysis. *Reviews of Geophysics*, 59(3):e2020RG000715.

Ballester, J., Rodó, X., and Giorgi, F. (2010). Future changes in central Europe heat waves expected to mostly follow summer mean warming. *Climate dynamics*, 35(7-8):1191–1205.

Barnston, A. G. and Livezey, R. E. (1987). Classification, Seasonality and Persistence of Low-Frequency Atmospheric Circulation Patterns. *Monthly Weather Review*, 115(6):1083–1126.

Bednar-Friedl, B., Biesbroek, R., Schmidt, D., Alexander, P., Borsheim, K., Carnicer, J., Georgopoulou, E., Haasnoot, M., Cozannet, G. L., Lionello, P., Lipka, O., Möllmann, C., Muccione, V., Mustonen, T., Piepenburg, D., and Whitmarsh, L. (2022). *Europe*, pages 1817–1927. Cambridge University Press, Cambridge, UK and New York, USA.

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., and Tian, Q. (2023). Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619(7970):533–538.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg.

Bochenek, B. and Ustrnul, Z. (2022). Machine learning in weather prediction and climate analyses;applications and perspectives. *Atmosphere*, 13(2).

Boehnisch, A., Felsche, E., Mittermeier, M., Poschlod, B., and Ludwig, R. (2023). Future hotspots of compound dry and hot summers emerge in European agricultural areas. *Authorea Preprints*.

Böhnisch, A., Mittermeier, M., Leduc, M., and Ludwig, R. (2021). Hot spots and climate trends of meteorological droughts in Europe–assessing the percent of normal index in a single-model initial-condition large ensemble. *Frontiers in Water*, 3:716621.

Bravar, L. and Kavvas, M. (1991). On the physics of droughts. analysis and simulation of the interaction of atmospheric and hydrologic processes during droughts. *Journal of Hydrology*, 129(1):299–330.

Bueh, C. and Nakamura, H. (2007). Scandinavian pattern and its climatic impact. *Quarterly Journal of the Royal Meteorological Society*, 133(629):2117–2131.

Cheng, X. and Wallace, J. M. (1993). Cluster analysis of the Northern Hemisphere wintertime 500-hPa height field: Spatial patterns. *Journal of Atmospheric Sciences*, 50(16):2674–2696.

Della-Marta, P. M., Luterbacher, J., von Weissenfluh, H., Xoplaki, E., Brunet, M., and Wanner, H. (2007). Summer heat waves over western Europe 1880–2003, their relationship to large-scale forcings and predictability. *Climate Dynamics*, 29(2):251–275.

Domeisen, D. I., Eltahir, E. A., Fischer, E. M., Knutti, R., Perkins-Kirkpatrick, S. E., Schär, C., Seneviratne, S. I., Weisheimer, A., and Wernli, H. (2023). Prediction and projection of heatwaves. *Nature Reviews Earth & Environment*, 4(1):36–50.

Farinosi, F., Dosio, A., Calliari, E., Seliger, R., Alfieri, L., and Naumann, G. (2020). Will the Paris Agreement protect us from hydro-meteorological extremes? *Environmental Research Letters*, 15(10):104037.

Felsche, E., Böhnisch, A., and Ludwig, R. (2023). Inter-seasonal connection of typical European heatwave patterns to soil moisture. *npj Climate and Atmospheric Science*, 6(1):1.

Felsche, E., Böhnisch, A., Poschlod, B., and Ludwig, R. (2024). Will present-day compound hot and dry European summers still be extreme in the future? Submitted to *Communications Earth & Environment*.

Felsche, E. and Ludwig, R. (2021). Applying machine learning for drought prediction in a perfect model framework using data from a large ensemble of climate simulations. *Natural Hazards and Earth System Sciences*, 21(12):3679–3691.

Folland, C. K., Knight, J., Linderholm, H. W., Fereday, D., Ineson, S., and Hurrell, J. W. (2009). The Summer North Atlantic Oscillation: Past, Present, and Future. *Journal of Climate*, 22(5):1082–1103.

Fraley, C. and Raftery, A. E. (1998). How many clusters? which clustering method? answers via model-based cluster analysis. *The computer journal*, 41(8):578–588.

Garcia-Leon, D., Casanueva, A., Standardi, G., Burgstall, A., Flouris, A. D., and Nybo, L. (2021). Current and projected regional economic impacts of heatwaves in Europe. *Nature communications*, 12(1):1–10.

Grillakis, M. G. (2019). Increase in severe and extreme soil moisture droughts for Europe under climate change. *Science of The Total Environment*, 660:1245–1255.

Hanel, M., Rakovec, O., Markonis, Y., Máca, P., Samaniego, L., Kyselỳ, J., and Kumar, R. (2018). Revisiting the recent European droughts from a long-term perspective. *Scientific reports*, 8(1):9499.

Heinermann, J. and Kramer, O. (2016). Machine learning ensembles for wind power prediction. *Renewable Energy*, 89:671–679.

Hirschi, M., Seneviratne, S. I., Alexandrov, V., Boberg, F., Boroneant, C., Christensen, O. B., Formayer, H., Orlowsky, B., and Stepanek, P. (2011). Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. *Nature Geoscience*, 4(1):17–21.

Holzinger, A., Saranti, A., Molnar, C., Biecek, P., and Samek, W. (2022). Explainable AI methods-a brief overview. In *International workshop on extending explainable AI beyond deep models and classifiers*, pages 13–38. Springer.

Hsu, H. and Dirmey-er, P. (2023). Soil moisture-evaporation coupling shifts into new gears under increasing CO2. *Nature Communications*, 14.

Huntingford, C., Jeffers, E. S., Bonsall, M. B., Christensen, H. M., Lees, T., and Yang, H. (2019). Machine learning and artificial intelligence to aid climate change research and preparedness. *Environmental Research Letters*, 14(12):124007.

Jacques-Dumas, V., Ragone, F., Borgnat, P., Abry, P., and Bouchet, F. (2022). Deep learning-based extreme heatwave forecast. *Frontiers in Climate*, 4.

Janocha, K. and Czarnecki, W. M. (2017). On loss functions for deep neural networks in classification. *arXiv preprint arXiv:1702.05659*.

Jiang, W. and Luo, J. (2022). An evaluation of machine learning and deep learning models for drought prediction using weather data. *Journal of Intelligent & Fuzzy Systems*, 43(3):3611–3626.

Jung, Y., Park, H., Du, D.-Z., and Drake, B. L. (2003). A decision criterion for the optimal number of clusters in hierarchical clustering. *Journal of Global Optimization*, 25(1):91–111.

Kaspar, F., Mächel, H., Jacob, D., and Kottmeier, C. (2017). Beobachtung von Klima und Klimawandel in Mitteleuropa und Deutschland. *Klimawandel in Deutschland: Entwicklung, Folgen, Risiken und Perspektiven*, pages 17–26.

Kautz, L.-A., Martius, O., Pfahl, S., Pinto, J. G., Ramos, A. M., Sousa, P. M., and Woollings, T. (2022). Atmospheric blocking and weather extremes over the Euro-Atlantic sector–a review. *Weather and climate dynamics*, 3(1):305–336.

Kendon, E. J., Jones, R. G., Kjellström, E., and Murphy, J. M. (2010). Using and designing gcm–rcm ensemble regional climate projections. *Journal of Climate*, 23(24):6485 – 6503.

Kodinariya, T. M. and Makwana, P. R. (2013). Review on determining number of cluster in k-means clustering. *International Journal*, 1(6):90–95.

Kueh, M.-T. and Lin, C.-Y. (2020). The 2018 summer heatwaves over northwestern Europe and its extended-range prediction. *Scientific reports*, 10(1):19283.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., and Battaglia, P. (2023). Learning skillful medium-range global weather forecasting. *Science*, 0(0):eadi2336.

Leduc, M., Mailhot, A., Frigon, A., Martel, J.-L., Ludwig, R., Brietzke, G. B., Giguère, M., Brissette, F., Turcotte, R., Braun, M., and Scinocca, J. (2019). The ClimEx Project: A 50-Member Ensemble of Climate Change Projections at 12-km Resolution over Europe and Northeastern North America with the Canadian Regional Climate Model (CRCM5). *Journal of Applied Meteorology and Climatology*, 58(4):663–693.

Lee, C. C. and Sheridan, S. C. (2012). A six-step approach to developing future synoptic classifications based on GCM output. *International Journal of Climatology*, 32(12):1792–1802.

Lhotka, O., Kyselỳ, J., and Farda, A. (2018). Climate change scenarios of heat waves in Central Europe and their uncertainties. *Theoretical and applied climatology*, 131(3-4):1043–1054.

Li, M., Yao, Y., Simmonds, I., Luo, D., Zhong, L., and Chen, X. (2020). Collaborative impact of the NAO and atmospheric blocking on European heatwaves, with a focus on the hot summer of 2018. *Environmental Research Letters*, 15(11):114003.

Liang-Liang, L., Jian, L., and Ru-Cong, Y. (2022). Evaluation of CMIP6 HighResMIP models in simulating precipitation over Central Asia. *Advances in Climate Change Research*, 13(1):1–13.

Lilensten, J., Wit, T. D. D., and Matthes, K. (2016). *Earth's climate response to a changing sun*. EDP Sciences.

Lim, Y.-K. (2015). The East Atlantic/West Russia (EA/WR) teleconnection in the North Atlantic: climate impact and relation to Rossby wave propagation. *Climate Dynamics*, Online first.

Linardatos, P., Papastefanopoulos, V., and Kotsiantis, S. (2021). Explainable AI: A review of machine learning interpretability methods. *Entropy*, 23(1).

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., et al. (2023). Summary of chatgpt-related research and perspective towards the future of large language models. *Meta-Radiology*, page 100017.

Livingston, I. (2023). Wintry blast in Europe has brought record snow and bitter cold. *The Washington Post*.

Lundberg, S. M. and Lee, S. (2017). A unified approach to interpreting model predictions. *CoRR*, abs/1705.07874.

Lupo, A., Kininmonth, W., Armstrong, J., and Green, K. (2013). Global climate models and their limitations. *Climate change reconsidered II: Physical science*, 9:148.

Maher, N., Milinski, S., and Ludwig, R. (2021). Large ensemble climate model simulations: introduction, overview, and future prospects for utilising multiple types of large ensemble. *Earth System Dynamics*, 12(2):401–418.

Mikhailova, N. and Yurovsky, A. (2016). The East Atlantic Oscillation: Mechanism and Impact on the European Climate in Winter. *Physical Oceanography*, 4.

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.

Mittermeier, M., Weigert, M., Rügamer, D., Küchenhoff, H., and Ludwig, R. (2022). A deep learning based classification of atmospheric circulation types over Europe: projection of future changes in a CMIP6 large ensemble. *Environmental Research Letters*, 17(8):084021.

Mosavi, A., Ozturk, P., and Chau, K.-w. (2018). Flood prediction using machine learning models: Literature review. *Water*, 10(11):1536.

Mukherjee, S. and Mishra, A. K. (2021). Increase in compound drought and heatwaves in a warming world. *Geophysical Research Letters*, 48(1):e2020GL090617. e2020GL090617 2020GL090617.

Murtagh, F. and Contreras, P. (2017). Algorithms for hierarchical clustering: an overview, ii. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(6):e1219.

Notz, D. (2015). How well must climate models agree with observations? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373(2052):20140164.

Osman, M., Zaitchik, B., and Winstead, N. (2022). Cascading drought-heat dynamics during the 2021 Southwest United States heatwave. *Geophysical Research Letters*, 49(12):e2022GL099265.

Perkins-Kirkpatrick, S. and Lewis, S. (2020). Increasing trends in regional heatwaves. *Nature communications*, 11(1):3357.

Quesada, B., Vautard, R., Yiou, P., Hirschi, M., and Seneviratne, S. I. (2012). Asymmetric European summer heat predictability from wet and dry southern winters and springs. *Nature Climate Change*, 2(10):736–741.

Roedel, W. (2013). *Physik unserer Umwelt: Die Atmosphäre*. Springer-Verlag.

Rokach, L. and Maimon, O. (2005). Clustering methods. *Data mining and knowledge discovery handbook*, pages 321–352.

Rumelhart, D. E., Durbin, R., Golden, R., and Chauvin, Y. (2013). Backpropagation: The basic theory. In *Backpropagation*, pages 1–34. Psychology Press.

Russell, S. and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*. Prentice Hall Press, Upper Saddle River, NJ, USA, 3rd edition.

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):124003.

Saltikoff, E., Friedrich, K., Soderholm, J., Lengfeld, K., Nelson, B., Becker, A., Hollmann, R., Urban, B., Heistermann, M., and Tassone, C. (2019). An overview of using weather radar for climatological studies: successes, challenges, and potential. *Bulletin of the American Meteorological Society*, 100(9):1739–1752.

Schmidt, M. and Felsche, E. (2024). The effect of climate change on crop yield anomaly in Europe. *Climate Resilience and Sustainability*, 3(1):e261.

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A. J. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, 99(3):125–161.

Sharifani, K. and Amini, M. (2023). Machine learning and deep learning: A review of methods and applications. *World Information Technology and Engineering Journal*, 10(07):3897–3904.

Sheffield, J. and Wood, E. (2011). Drought: Past problems and future scenarios. *Drought: Past Problems and Future Scenarios*, pages 1–234.

Shinde, P. P. and Shah, S. (2018). A review of machine learning and deep learning applications. In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, pages 1–6.

Spinoni, J., Naumann, G., and Vogt, J. V. (2017). Pan-European seasonal trends and recent changes of drought frequency and severity. *Global and Planetary Change*, 148:113–130.

Stefanon, M., D'Andrea, F., and Drobinski, P. (2012). Heatwave classification over Europe and the Mediterranean region. *Environmental Research Letters*, 7(1):014023.

Stryhal, J. and Plavcová, E. (2023). On using self-organizing maps and discretized Sammon maps to study links between atmospheric circulation and weather extremes. *International Journal of Climatology*.

Suarez-Gutierrez, L., Müller, W. A., Li, C., and Marotzke, J. (2020). Dynamical and thermodynamical drivers of variability in European summer heat extremes. *Climate Dynamics*, 54:4351–4366.

Teuling, A., Hirschi, M., Ohmura, A., Wild, M., Reichstein, M., Ciais, P., Buchmann, N., Ammann, C., Montagnani, L., Richardson, A., et al. (2009). A regional perspective on trends in continental evaporation. *Geophysical Research Letters*, 36(2).

Tippett, M. K., Allen, J. T., Gensini, V. A., and Brooks, H. E. (2015). Climate and hazardous convective weather. *Current Climate Change Reports*, 1:60–73.

Truedinger, A. J., Jamshed, A., Sauter, H., and Birkmann, J. (2023). Adaptation after extreme flooding events: Moving or staying? the case of the Ahr Valley in Germany. *Sustainability*, 15(2):1407.

Tuel, A. and Eltahir, E. A. (2021). Mechanisms of European summer drying under climate change. *Journal of Climate*, 34(22):8913–8931.

Vautard, R., Yiou, P., D'Andrea, F., de Noblet, N., Viovy, N., Cassou, C., Polcher, J., Ciais, P., Kageyama, M., and Fan, Y. (2007). Summertime European heat and drought waves induced by wintertime mediterranean rainfall deficit. *Geophysical Research Letters*, 34(7).

von Trentini, F., Leduc, M., and Ludwig, R. (2019). Assessing natural variability in RCM signals: comparison of a multi model EURO-CORDEX ensemble with a 50-member single model large ensemble. *Climate Dynamics*, 53:1963–1979.

Wehrli, K., Guillod, B. P., Hauser, M., Leclair, M., and Seneviratne, S. I. (2019). Identifying key driving processes of major recent heat waves. *Journal of Geophysical Research: Atmospheres*, 124(22):11746–11765.

Whan, K., Zscheischler, J., Orth, R., Shongwe, M., Rahimi, M., Asare, E. O., and Seneviratne, S. I. (2015). Impact of soil moisture on extreme maximum temperatures in. *Weather and Climate Extremes*, 9:57–67.

Willingham, E. (2023). AIs victories in go inspire better human game playing. *Scientific American*.

Zampieri, M., D'andrea, F., Vautard, R., Ciais, P., de Noblet-Ducoudré, N., and Yiou, P. (2009). Hot European summers and the role of soil moisture in the propagation of Mediterranean drought. *Journal of Climate*, 22(18):4747–4758.

Zhang, Y. (2010). *New advances in machine learning*. BoD–Books on Demand.