SPATIO-TEMPORAL STRUCTURES IN HEATWAVE AND DROUGHT ASSESSMENT ACROSS EUROPE:

EXPLAINING INTERNAL VARIABILITY AND MULTIVARIATE RELATIONS FROM A REGIONAL CLIMATE MODEL LARGE ENSEMBLE

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

vorgelegt von

ANDREA BÖHNISCH

eingereicht am 15.01.2024, München

Supervisor: Prof. Dr. Ralf Ludwig,
Department of Geography,
Ludwig-Maximilians-Universität München,
Munich, Germany

Second Supervisor: Dr. Martin Leduc,
(1) Ouranos,
Montréal, Canada
(2) 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: 02.05.2024

"In nature nothing exists alone."
- Rachel Carson, Silent Spring

Acknowledgements

These are the results of a four-year voyage to seek out new statistical methods and connections, to boldly go where no geographer has gone before (or at least, only few). When I embarked on this enterprise, I did not know what lay ahead. Speaking from a SMILE point of view, there were uncountable members to unfold (imagine parallel time lines, seeded in winter 2019). I feel I followed a good one, with the right people granting support, providing a pleasant working atmosphere, and pushing me sometimes beyond the frontiers of my comfort zone.

Among these people and first of all, I would like to thank my supervisor Prof. Ralf Ludwig for guiding me on the first steps of growing into a scientist. I am extremely grateful for the opportunity to freely develop a topic that perfectly suited my interests. Thank you for encouraging me to pursue these ideas although I was committed to other project work as well, for excellent scientific supervision and kind motivation, and for opening doors to scientific exchange at various conference and workshop visits that made my curiosity thrive.

Many thanks go also to my second supervisor Martin Leduc for his always helpful and fast responses to any kind of questions during the last years and his willingness to revise this PhD thesis. This thesis being tightly related to the Bavarian-Quebec project ClimEx, I am also profoundly indebted to the entire ClimEx team for the opportunity to work on their data.

During a large part of my PhD, the world had shrunk due to the pandemic. So I started looking for topics and collaborations more closely. And what expertise and inspirations did I find within the working group! In particular, I owe many thanks to my colleagues: Magdalena, Alex, Ben, Elizaveta, and Raul, and the entire basement working group. Thank you for introducing me to scientific and technical topics, for brightening lunch time discussions, and pleasant conference trips. You really made me look forward to returning to the office after the pandemic.

Special thanks go to Vera, definitely the secret core of the group, for having an open ear to all kinds of administrative questions, providing emergency fuel in form of chocolates, and keeping the gears running in the group.

II

Last, but not least, I wish to thank my family and friends for continuous support during my PhD, including absolutely necessary times off, and uncountable cups (or rather pots) of tea in precisely the right moments.

Earl Grey, of course.

Summary

Within the complex climate system, no variable exists in isolation. Manifold kinds of relationships connect them. These relationships may link different variables (e.g., teleconnection indices and responses) or the same variable at different locations (e.g., air pressure over the Azores and Iceland). In this PhD thesis, these relationships are understood as structures among variables that are empirically quantified from climate data.

This thesis focuses on structures of atmospheric hazards in Europe, i.e., of phenomena with the potential for adverse effects on natural and human systems. The hazards of interest are heatwaves and droughts: Both are of major importance to Europe because they cause considerable loss and damages (e.g., in 2018, 2022). Delving deeper into facets that may foster understanding or prediction of these hazards is thus crucial for adaptation planning. Here, structures come into play. Heatwaves and (meteorological) droughts are subject to robust climate variability and strong forced trends due to anthropogenic global warming. Structures related to them can thus be investigated under obviously changing background conditions. Additionally, both hazards are strongly linked mechanistically, thus bearing the potential for finding structures among them as well. The structures under consideration in this thesis include a teleconnection and its structure to responses, heatwave tracks, facets of droughts, and dry and hot compound summers. In parts, they describe causal relations, i.e., one component causing the other(s).

This work attempts to provide empirical answers for the question of *how data-based structures may contribute meaningfully and in an actionable way to regional climate change assessment of heatwaves and droughts.* Given the origin of structures in temporally varying and changing variables, this includes an assessment of internal variability and trends of structures. So far, spatio-temporal structures were hardly investigated regarding their naturally occurring variability, but rather assumed to be stable, albeit with a certain level of (sampling) uncertainty. In some cases, e.g., related to teleconnections, structural changes due to global warming were acknowledged. Given that climate change conditions affect the structure components, assessment of potential structural changes is also required. A robust assessment of potentially transiently changing struc-

tures though demands appropriate estimates of their inherent uncertainty or variability. For empirically derived structures, the lack of sufficient data samples may undermine this requirement (e.g., short or non-stationary time series).

Within the research area of climate model ensembles, though, a large treasure of earth- and climate-related data has emerged recently and awaits being retrieved. In particular, so-called Single Model Initial-Condition Large Ensembles (SMILEs) provide abundant data to robustly investigate extreme and rare climate events, internal climate variability, forced trends – and relationships among variables as will be shown here. SMILEs are collections of climate simulations, driven by a single climate model with slightly differing atmospheric states in the beginning, but under the same forcing scenario and model parameterizations. The employed SMILE increases the number of simulated years by factor 50 compared to observations or single simulations. It is based on a regional climate model (Canadian Regional Climate Model, version 5) and was obtained in the ClimEx (Climate Change and Hydrological Extremes) project by dynamically downscaling the SMILE of a global climate model until 2100 (Canadian Earth System Model, version 2, under the emissions concentration scenario RCP8.5). Hazards like heatwaves and droughts were thus supposed to be resolved in high geographical detail at impact-relevant spatial scales.

Four papers – three peer-reviewed publications and one submitted manuscript, sidelined by several co-authored publications and manuscripts – provide the scientific basis for concluding that heatwave and drought structures – just like their contributing variables – are subject to internal variability and change transiently. These structures are shown to be useful for further assessment of regional climate change. The first paper addresses internal variability of teleconnection–response structures (European temperature and precipitation). Also, large-scale atmospheric patterns in the regional SMILE are compared against the global SMILE in this setting: For changes of recurrent heatwave and drought features to be investigated on a regional scale, climate drivers and their internal variability should be represented reasonably in the data used for investigation. The second paper gives evidence for robust meteorological drought trends and variability across Europe using an easy-to-communicate index. Recurrent heatwave tracks, as so-far unknown structures, are distilled, described, and associated to driving atmospheric conditions in the third paper. In the fourth paper, dry and hot compound extreme

events are related to soil moisture impacts. Their structural changes are considered in addition to temperature and precipitation trends.

Papers I–IV deliver five major findings to introduce the structures for further regional climate assessment:

1. Large-scale atmospheric patterns and internal variability of teleconnection–response structures are consistent between the driving Canadian Earth System Model version 2 Large Ensemble (CanESM2-LE) and the nested Canadian Regional Climate Model version 5 Large Ensemble (CRCM5-LE) (paper I).

2. Against this background, structures of heatwaves and droughts can be robustly identified (e.g., causal discovery) and described (e.g., by means of correlations, copula) in the regional SMILE with data-driven methods (papers I–IV).

3. The structures as well as spatial patterns of (potentially) driving variables are subject to internal climate variability and may change over time. It is thus useful and necessary to employ large samples, e.g., SMILEs, for their derivation and change assessment (papers I–IV).

4. To support these data structures mechanistically, investigating atmospheric drivers and structure components is useful, but causal interpretations require caution (papers III, IV).

5. As these relationships are abstract, they require good strategies for (intra-)science communication (paper II, III).

Moreover, the results illustrate that those structures may be used meaningfully as quantities (cf. trends, averages) of hazards. For example, heatwave tracks or responses to teleconnections are shown to be subject to internal variability. Using different periods for their analyses may thus result in inconsistent results. As suggested in literature, structures based on, e.g., causal discovery, may also improve further model evaluation or evaluate model improvement. In this thesis, the teleconnection–response structure is, e.g., used for evaluating the Global Climate Model (GCM)–Regional Climate Model (RCM) propagation of variability. For comparison of structures in observations and simulations though, considering the range of their internal variability is crucial, as is also shown

in this thesis. As a last example, the multivariate structure of compound dry and hot events allowed estimating their joint extremeness and then relating them to impact-relevant soil moisture contents. Altogether, hazards like heatwaves and droughts may thus be contextualized and interpreted in novel contexts when using structures (e.g., intra-seasonally, spatially).

Since these results are primarily based on data-driven findings, several limitations have to be considered, such as the importance of recurring to correct and complete data. To minimize the risk of using spurious data-based relationships, a framework for considering causality is proposed: It merges both data-driven derivation of structures and aligning potentially causal drivers (also to connect the structure components) or impacts. The findings of this thesis thence also contribute to the assessment of chances and limitations of purely data-driven inferences. Furthermore, this thesis stresses that internal variability and transient trends affect relationships among variables, i.e., that they are not stationary. In parts, these variations of structures occurred in addition to the trends and variability of the components. These findings point towards the importance of reasonably sampling and discussing of, e.g., periods under consideration: Any quantification of relationships is potentially subject to internal variability and hence a certain degree of uncertainty.

Especially physical geography with its efforts to connect various compartments of the joint nature-human Earth system could benefit from considering relationships among variables from "climate big data" and from investigating their impacts and drivers.

Zusammenfassung

Im Erdsystem stehen klimatische Variablen auf vielfältige Weise miteinander in Beziehung. Diese Beziehungen verbinden unterschiedliche Variablen (z. B. atmosphärische Indizes und deren klimatische Effekte) oder dieselbe Variable an unterschiedlichen Orten (z. B. Luftdruck im Bereich der Azoren und Island). Im Rahmen dieser Dissertation werden solche Beziehungen als persistente Strukturen aufgefasst. Sie können empirisch durch Wetter- und Klimadaten quantifiziert werden.

Im Zentrum dieser Arbeit stehen Strukturen von Hitzewellen und Dürren und damit Phänomenen, die das Potenzial für negative Auswirkungen auf natürliche und anthropogene Systeme bergen (*hazards*): Sowohl Hitzewellen als auch Dürren führten in den vergangenen Jahren wiederholt zu erheblichen Schäden und Verlusten in Europa (z. B. 2018, 2022). Eine tiefergehende Betrachtung von verschiedenen Facetten dieser Naturgefahren kann daher die Planung zur Anpassung an sie unterstützen. Die erwähnten Strukturen repräsentieren hier Beispiele für derartige Facetten. Hitzewellen und Dürren basieren auf Variablen mit deutlicher Klimavariabilität und robusten Klimawandeltrends, sodass sich ihre Strukturen unter veränderlichen Hintergrundbedingungen untersuchen lassen. Da zudem beide Ereignistypen durch ähnliche Prozesse angetrieben werden, kann die Untersuchung ihrer Beziehungen untereinander als Testfall für die Erfassung und Analyse von empirischen Strukturen dienen. Die hier betrachteten Beispiele für Strukturen umfassen den Zusammenhang zwischen einer atmosphärischen Fernwirkung (Telekonnektion) und ihren klimatischen Auswirkungen, typische Bewegungsrichtungen von Hitzewellen sowie Aspekte von Dürren oder multivariaten „heißen Dürren" (*dry and hot compound events*).

Insbesondere soll die Frage beantwortet werden, *wie datenbasiert ermittelte Strukturen auf relevante und verwertbare Weise zu regionalen Untersuchungen von Hitzewellen und Dürren beitragen können*. Da Strukturen auf zeitlich veränderlichen Variablen beruhen, beinhaltet dies auch eine Betrachtung von interner Variabilität und Klimawandeltrends der Strukturen. Im Allgemeinen wurden Strukturen bislang nicht selbst hinsichtlich ihrer internen Variabilität untersucht, sondern – mit gewissen Unsicherheiten behaftet – als stabil betrachtet. Strukturelle Änderungen von Telekonnektionen im Rahmen des Klimawandels etwa erfahren jedoch Beachtung in der wissenschaftlichen Literatur. Dies ist

relevant, da auch die Variablen, aus denen sich diese Strukturen zusammensetzen, Klimawandeltrends unterliegen. Eine robuste Abschätzung solcher transienten Zustandsänderungen von Strukturen erfordert wiederum eine angemessene Berücksichtigung ihrer natürlicherweise auftretenden Variabilität. Diese jedoch fehlt oft. Teilweise lässt sich dies auf mangelnde Datenverfügbarkeit (z. B. einzelne kurze oder von Schwankungen überlagerte Zeitreihen) zurückführen.

Im Forschungsbereich der Klimamodell-Ensembles ist über die vergangenen Jahre ein (Klima-)Datenschatz entstanden, der diese Einschränkung obsolet macht. Insbesondere sogenannte *single model initial-condition large ensembles* (SMILEs) bieten umfangreiche Datengrundlagen für die Betrachtung von seltenen oder extremen Ereignissen, interne Klimavariabilität und Klimawandeltrends – und Beziehungen zwischen Variablen, wie hier gezeigt werden soll. SMILEs stellen Sammlungen mehrerer Klimasimulationen desselben Modells und Szenarios dar, die sich lediglich durch die atmosphärischen Initialbedingungen unterscheiden. Das hier verwendete SMILE verfünfzigfacht dadurch die Datenbasis im Vergleich zu Beobachtungen oder einfachen Simulationen. Die Besonderheit dieses SMILEs besteht darin, dass es auf einem regionalen Klimamodell beruht (*Canadian Regional Climate Model, version 5*), welches damit relevante Skalen für Klimawandelauswirkungen besser auflöst als das globale, antreibende Klimamodell (*Canadian Earth System Model, version 2*). Insbesondere können räumliche Muster von Hitzewellen und Dürren in höherer geographischer Genauigkeit abgebildet werden. Das regionale SMILE wurde im Rahmen des bayerisch-quebecer Projekts ClimEx (Klimawandel und Extremereignisse) erzeugt und liefert für Europa unter dem Emissionskonzentrationsszenario RCP8.5 Klimaprojektionen bis 2100.

Vier wissenschaftliche Fachartikel – drei in Fachzeitschriften bereits veröffentlichte Publikationen und ein eingereichtes Manuskript, unterstützt durch weitere mitverfasste Fachartikel – bilden die Basis für die Schlussfolgerung, dass Strukturen von Hitzewellen und Dürren sich für weiterführende Klimawandelanalysen eignen. Darüber hinaus zeigen sie, dass Strukturen genau wie ihre zugrundeliegenden Variablen interner Klimavariabilität und transienten Änderungen unterworfen sind. Der erste Fachartikel befasste sich mit der internen Klimavariabilität von Strukturen eines Telekonnektionsindex und seiner klimatischen Effekte in Europa (Temperaturen und Niederschlag). Zudem wurden dort großskalige atmosphärische Muster im regionalen und globalen SMILE abgeglichen.

Ihre korrekte Abbildung im verwendeten Datensatz ist Grundlage für eine erfolgreiche Untersuchung von wiederkehrenden Hitze- und Dürremustern auf regionaler Ebene. Der zweite Fachartikel untersuchte Trends und Variabilität von meteorologischen Dürren in Europa auf eine Weise, welche eine Dissemination der Ergebnisse in die breite Öffentlichkeit unterstützen soll. Wiederkehrende Zugbahnen von Hitzewellen stellten den Fokus des dritten Fachartikels dar: Methoden zur Herleitung, Beschreibung und Anknüpfung an zugrundeliegende physikalische Prozesse wurden hier für diese bislang unbekannten Strukturen erprobt. Der vierte Fachartikel betrachtete multivariate heiße Dürren (*compound dry and hot events*) und deren Bezug zu Bodenfeuchtedürren. Insbesondere die treibenden Faktoren für Zunahmen der Ereignisse im Zuge des Klimawandels – Trends in den Komponenten und strukturelle Änderungen – wurden dabei beleuchtet.

Aus den vier Fachartikeln lassen sich fünf grundlegende Punkte ableiten, um Strukturen für weitergehende regionale Klimawandelanalysen einzuführen:

1. Groß-skalige atmosphärische Muster und die interne Variabilität von klimatischen Auswirkungen einer Telekonnektion sind konsistent zwischen dem globalen und regionalen SMILE abgebildet (Fachartikel I).

2. Vor diesem Hintergrund lassen sich im CRCM5-LE Strukturen von Hitzewellen und Dürren datengetrieben robust identifizieren (z. B. mit *causal discovery*-Methoden) und beschreiben (z. B. mit Copulas) (Fachartikel I–IV).

3. Die Strukturen unterliegen interner Variabilität und Klimatrends. Die umfangreiche Datenstichprobe des SMILE stellt daher eine nützliche und erforderliche Grundlage dar, um Strukturen von Hitzewellen und Dürren zu untersuchen (Fachartikel I–IV).

4. Zur physikalischen Begründung dieser Strukturen bieten atmosphärische Treiber und die Strukturkomponenten gute Anhaltspunkte; sie erfordern jedoch Vorsicht bei einer kausalen Interpretation (Fachartikel III, IV).

5. Zuletzt bedürfen die teils abstrakten Strukturen angemessener Strategien für ihre Darstellung im Bereich der Kommunikation (Fachartikel II, III).

Jenseits dieser fünf grundlegenden Ergebnisse illustrieren alle vier Fachartikel, dass die Strukturen wie andere Aspekte von Naturgefahren auch (z. B. Trends, Mittelwerte)

nützliche Größen darstellen. Beispielsweise lässt sich zeigen, dass Telekonnektionseffekte nicht stationär auftreten. Ähnliches gilt für die im zweiten Fachartikel hergeleiteten Hitzewellenzugbahnen: Die Betrachtung unterschiedlicher Zeiträume für die Definition von Strukturen kann daher in inkonsistenten Resultaten münden. Für Klimawandeltrends ist dieser Effekt bereits gut belegt. Auch für den Vergleich von Strukturen aus Beobachtungen mit jenen aus Simulationen muss der Schwankungsbereich der internen Variabilität Berücksichtigung finden. Darüber hinaus erlauben Strukturen eine verbesserte Modellevaluation im Hinblick auf die Abbildung physikalischer Zusammenhänge. So unterstützte hier die Struktur aus Telekonnektionsindex und Klimawirkung in Mitteleuropa die Evaluierung der Kette aus globalem und regionalem SMILE. Weiterhin wurde mit Hilfe einer Struktur, welche die Seltenheit multivariater heißer Dürren jenseits empirischer Grenzen darzustellen erlaubt, ein Zusammenhang zwischen diesen Ereignissen und ihren Auswirkungen auf die Bodenfeuchte untersucht. Durch die Betrachtung von Strukturen lassen sich Ereignisse wie Hitzewellen und Dürren folglich auf neue Weisen kontextualisieren und interpretieren (z. B. jahreszeitenübergreifend, räumlich).

Da die hier gezeigten Ergebnisse hauptsächlich durch die Auswertung existierender Datensätze (insbesondere Modelldaten) erzielt wurden, ergeben sich spezifische Einschränkungen: So steht und fällt die Aussagekraft der Auswertungen mit der Qualität und Vollständigkeit der genutzten Daten. Daher schließt die Dissertation mit einem Vorschlag zu einem strukturierten Vorgehen, um die datengetriebene Ableitung von Strukturen mit (mutmaßlich) kausal verknüpften Treibern und Effekten zu verbinden. Dies ist erforderlich, auch um Scheinzusammenhänge innerhalb der Strukturen zu vermeiden. Diese Arbeit trägt damit dazu bei, Chancen und Grenzen von rein datenbasiert gewonnenen Informationen abzuwägen. Weiterhin weist sie auf die Bedeutung sinnvoll ausgewählter Untersuchungszeiträume hin, da auch Strukturen selbst natürlicher Variabilität und Trends unterliegen – teilweise zusätzlich zu den Trends und Variationen ihrer Komponenten. Damit ist jede datenbasierte Quantifizierung von Zusammenhängen zwischen Klimavariablen unter dem Vorbehalt von interner Klimavariabilität und damit einhergehender Unsicherheit zu betrachten.

Untersuchungen von datenbasierten Variablenzusammenhängen sowie ihre Verknüpfung mit Treibervariablen können schließlich insbesondere für die Physische Geographie von Bedeutung sein, wird dort doch häufig das Zusammenspiel verschiedener Komponenten im Erdsystem betrachtet.

Contents

List of Figures

List of Acronyms

AR	Assessment Report
CanESM2	Canadian Earth System Model version 2
CanESM2-LE	Canadian Earth System Model version 2 Large Ensemble
CDHE	Compound Dry and Hot Event
ClimEx	Climate Change and Hydrological Extremes
CMIP	Coupled Model Intercomparison Project
CORDEX	Coordinated Downscaling Experiment
CRCM5	Canadian Regional Climate Model version 5
CRCM5-LE	Canadian Regional Climate Model version 5 Large Ensemble
DAG	Directed Acyclic Graph
EGU	European Geosciences Union
ENSO	El Niño/Southern Oscillation
GCM	Global Climate Model
GHG	Greenhouse Gas
IPCC	Intergovernmental Panel on Climate Change
MMLEA	Multi Model Large Ensemble Archive
NAO	North Atlantic Oscillation
PC	Peter and Clark (developer surnames)
PNI	Percent of Normal Index
RCM	Regional Climate Model
RCP	Representative Concentration Pathway
SLP	sea level pressure
SMI	Soil Moisture Index
SMILE	Single Model Initial-Condition Large Ensemble
SPEI	Standardized Precipitation Evapotranspiration Index
SPI	Standardized Precipitation Index
SRI	Standardized Runoff Index
SSP	Shared Socioeconomic Pathway

1   Climate (Variability) in Space, Time, and Data

Welcome to the near future!

At the time of writing this thesis, we have reached the 2020s, a period still considered as a near-term horizon in the most recent, sixth Assessment Report (AR) of the Intergovernmental Panel on Climate Change (IPCC) (2021–2040 in Lee et al. 2021). Research shows that the climate system is transitioning to a new state due to considerable human influence: Given a global warming of roughly 1 °C in 2001–2020 relative to the pre-industrial period, more and more climate indicator signals robustly emerge from the noise of naturally occurring climate variability (e.g., Ossó et al. 2022; Masson-Delmotte et al. 2021) that hitherto served as an excuse to ignore or deny climate trends. The conclusion is clear: We are now living amidst anthropogenic climate change and experiencing climate conditions that resemble projections from earlier decades (Rahmstorf et al. 2012; Stouffer & Manabe 2017; Carvalho et al. 2022).

Undoubtedly, the climate system, in which this change is occurring, is highly complex: Atmospheric processes interact with land and ocean – including life (biosphere) and ice (cryosphere) – with basically all compartments experiencing anthropogenic disturbance (Masson-Delmotte et al. 2021).

Zooming into the atmospheric compartment of the climate system, a distinction is taken between immediate atmospheric effects and aggregated statistical conditions. The first one refers to the instantaneous state and evolution of meteorological variables like air pressure, temperature, and precipitation (or sunshine). This is weather. Climate, on the contrary, encompasses temporally (and often spatially) aggregated information on the state of meteorological variables, including their averages, variability, and trends. Weather can thus be directly experienced by looking into the sky, while climate features more abstractly in statistics.

Both weather and climate interact on numerous scales: From a temporal perspective, chaotic properties of the atmosphere introduce the above mentioned natural variability which also features in long-term climate time series (e.g., fluctuations in global mean temperatures). External forcing like naturally ejected volcanic aerosol or anthropogenic Greenhouse Gas (GHG) emissions from fossil fuel burning or land use/land cover changes superimpose trends. The anthropogenic contribution in particular pushes the climate

system to new states (i.e., climate change, Masson-Delmotte et al. 2021). Therefore, "all weather events are affected by climate change because the environment in which they occur is warmer and moister than it used to be" (Trenberth 2012, p. 283).

This temporal variability opens a spread between extreme climate characteristics: For instance, ambient temperature reaches extremely high or low values at times, but generally oscillates around its (seasonal) mean value. Rare extreme events may have detrimental impacts on human and natural systems as the prevailing heat conditions and associated extreme dryness in Europe during the recent years 2018, 2019, and 2022 showed. Adaptation to these events requires changes among others in the health system, agricultural practices, energy systems, or transportation (see Section 2). Those specific atmospheric phenomena with the potential for adverse effects on natural and human systems are called hazards (e.g., Zscheischler et al. 2018). Climate change is projected (and already shown) to create conditions under which previously extreme events, in particular heat-waves and droughts, become more frequent, more prevailing, and more intense, hence a "new normal state" (Masson-Delmotte et al. 2021).

In such a world, it becomes mandatory to assess climate change impacts for adaptation strategies to sustain human well-being – not least because climate change and its impacts vary strongly in space as well. This spatial variability can be studied in great detail in the heterogeneous landscapes of Europe with their mixture of flatland to mountainous areas, maritime to continental, and subpolar to subtropical climate conditions. Hence, for impact-relevant conclusions to be drawn, regional analyses of extreme events and their variability are necessary.

What aspects of the climate system, though, are subject to climate variability? Is variability restricted to temporal or (static) spatial characteristics of given hazards? Geographical research questions in particular may also address remote effects and relationships among drivers and effects in the climate system. In this thesis, these (empirical) relationships among multiple variables are termed structures (e.g., an index quantifying atmospheric flow onto Europe *related to* resulting European winter temperatures). These structures represent the form of relationship that is enclosed in the hyphen in, e.g., temperature–precipitation relationship (Figure 1).

If already known, spatio-temporal structures may be quantified by correlation or regression coefficients (Figure 1 (a)), fitted joint functions (Figure 1 (b)), or similar.

(a)

(b)
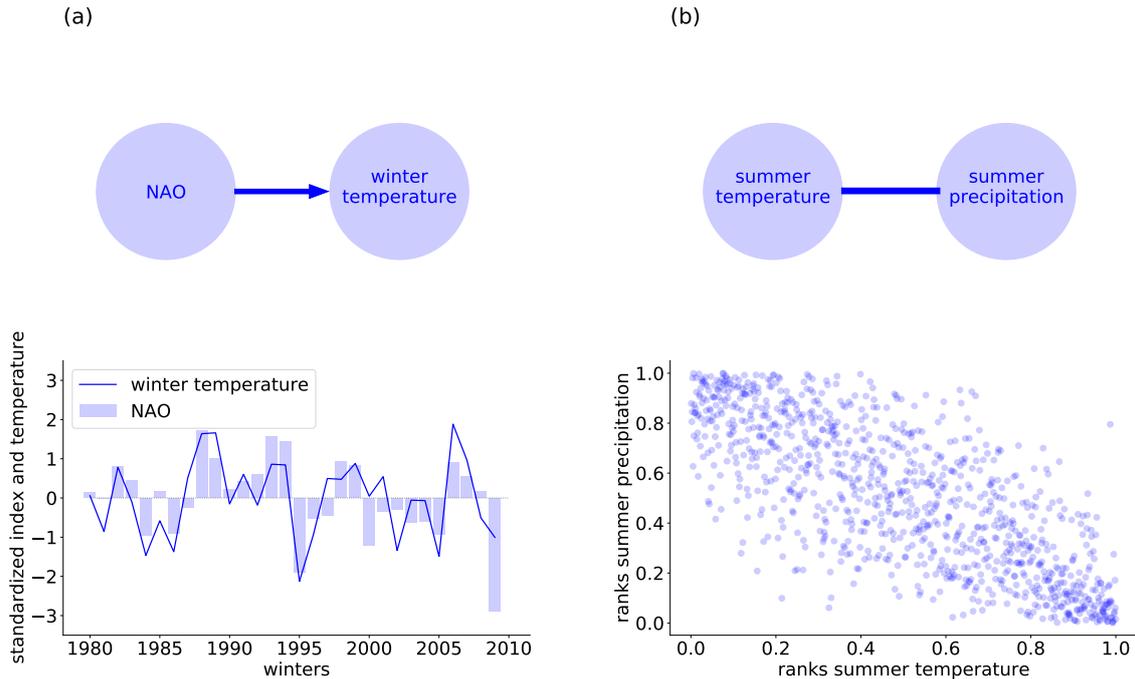
Figure 1: The "structure" concept of this thesis. (a) North Atlantic Oscillation (NAO)–winter temperature structure: time series of the NAO index and standardized winter temperature at an arbitrary location in central Europe (data: Canadian Regional Climate Model version 5 (CRCM5) driven by reanalysis). (b) Summer temperature–precipitation structure: temperature and precipitation ranks of 1000 summers in an arbitrary region south of the Alps (data: CRCM5-LE). The empirical copula of the bivariate temperature–precipitation distribution represents this structure.

Otherwise, they may be extracted from available climate data using, e.g., data-driven methods, but based on theory-guided hypotheses. Given their origin in variables subject to natural climate variability and climate change, this thesis suggests that structures be affected by them as well. Their evaluation promises deeper understanding on (recurrent) climate hazard properties and connections both in time and space. For instance, how do these structures change under global warming conditions? As alluded to above, robust signal detection requires emergence from climate noise, i.e., climate variability. Therefore, climate variability of structures merits attention as well. To answer these questions, though, a sufficient representation of the climate system in terms of data is required.

Luckily, observations and simulations provide a plethora of available information on the state and changes of the Earth system both in time and space. Global and regional climate models allow to understand complex relations in, and test assumptions on the

climate system. Large single-model ensembles, i.e., collections of simulations by the same model and external forcing, but with minuscule differences in the starting conditions, are exploited for robust analyses of rare (extreme) events and estimates of climate variability. Here, they will also serve to obtain spatio-temporal structures belonging to (land) heatwaves, droughts, and their combined occurrences.

This cumulative PhD thesis examines data-based spatio-temporal structures pertinent to geographical questions. This allows to highlight new facets of heatwaves and droughts in the context of climate change assessment and provide novel types of contextualization or interpretation. Additionally, it thereby enriches the collection of metrics of which climate variability is investigated. The particularity of the analyses conducted here is the investigation of trends and variability of heatwave and drought structures in a yet unique regional Single Model Initial-Condition Large Ensemble (SMILE) (Leduc et al. 2019; Maher et al. 2021).

The following section will introduce some fundamental definitions and background knowledge on heatwaves and droughts in Europe (Section 2). Next, the state of the art of the employed methods points to emerging knowledge gaps (Section 3) which lead to the main and satellite research questions of this thesis (Section 4). Three first-authored peer-reviewed publications and one submitted manuscript then provide the scientific base for finding answers (Section 5). These answers (Section 6) guide towards a synthesis (Section 7) and result in the proposition of a framework on including causal considerations in data-driven analyses of spatio-temporal structures (Section 8).

2 Hot and Dry Extremes over Europe

The first part of this section introduces the major hazards featured in the scientific publications. Although at first glance more invisibly striking than extreme precipitation or floods, heatwaves, droughts, and their combined occurrence (i.e., Compound Dry and Hot Events (CDHEs)) are among the deadliest and most expensive hazards in Europe. Drivers of their variability at different scales are presented in the second part of this section.

2.1 On Heat and Droughts

Extreme heat and drought repeatedly resulted in high fatalities, crop failures, or wildfires in Europe during the last years (Section 2.1.1 and 2.1.2). Global mitigation efforts notwithstanding, investigation and prediction of, as well as adaptation to these events is thus of paramount importance in order to sustain livable conditions in Europe for approximately 740 million inhabitants (UN 2022).

2.1.1 Heatwaves: Summer Nightmares

Heatwaves are periods of unusually hot weather (WMO 2023a). Typically, they last for several days to weeks. In agreement with common perception, heatwaves are defined relative to the warmest temperatures during a year. Following WMO (2023a), this separates them from warm spells (too warm with respect to the time of year). In general, thermodynamic (i.e., driven by radiation and surface fluxes) and dynamic processes (i.e., related to atmospheric motions like large-scale circulation) contribute to heat accumulation, the relative importance of which varies regionally (Röthlisberger & Papritz 2023; Miralles et al. 2014; Suarez-Gutierrez et al. 2020a; Kraus 2004).

Heatwaves may be quantified by the duration of days exceeding a certain local temperature threshold, their cumulative heat (magnitude), or spatial extent (WMO 2023a). An extensive list of metrics is given in Barriopedro et al. (2023). Any threshold-based definition is relative to a certain reference period. This allows to investigate, e.g., frequency, intensity, or duration changes over time with respect to climate conditions to which infrastructure, administration, and society are adapted (e.g., infrastructure constructed

after World War II in large parts of Europe, and thus under outdated climate conditions, Ossó et al. 2022). Progressing global warming between reference and projected future epochs, though, may shift and likely change the shape of the temperature distribution, thus resulting in more (and hotter) heatwaves (Suarez-Gutierrez et al. 2020a).

Potential impacts of (land) heatwaves comprise among others adverse effects on the health sector, agriculture, and ecology with direct ties to economy (e.g., Vogel et al. 2019; Barriopedro et al. 2011). While for human health high absolute temperatures are most detrimental (e.g., prolonged exposure to temperatures above the body core temperature of 37 °C or to high nighttime temperatures that inhibit recovery from daytime heat stress, Sherwood & Huber 2010; Asseng et al. 2021), other impacts relate to the extremeness based on the local temperature distribution (e.g., wherever a system is adopted to the typical regional temperature regime, such as vegetation). In both cases, the cooling of human bodies or plants by transpiration requires sufficient water supply. Some countries implemented or currently implement national (e.g., the Netherlands, RIVM n.d.) or communal (e.g., Germany, StMGP n.d.) heatwave emergency plans to warn and relieve affected population.

During the last years, Europe was increasingly hit by several heatwaves, a selection of which follows (see Russo et al. 2015, for more examples): In 2022, more than 16 300 extreme weather-related fatalities were attributed to heatwaves (WMO 2023b). For the first time on record, the UK experienced temperatures above 40 °C and temperature records were shattered in various locations (WMO 2023b). Only a few years before, 80 heat-related deaths per million inhabitants were recorded in Europe in 2018 (van Daalen et al. 2022). Across the entire Northern Hemisphere, heat-related wildfires, crop losses, infrastructural damages (e.g., roads melting), and anomalous animal mortality were reported (Vogel et al. 2019). Northeastern Europe and Russia were struck by a particularly enduring event in 2010 with more than 55 000 excess deaths in Russia. More than 2 million km² were covered by unprecedented temperatures. Russian annual crop failure amounted to 25 % and the total economic loss equaled roughly 1 % of the gross domestic product (Barriopedro et al. 2011). Although already 20 years ago, the 2003 heatwave and its accompanying dry conditions are still considered as exceptional, not at last because of a record of 108 heat-related deaths per million inhabitants (van Daalen et al. 2022). Gross and net primary (plant) productivity were severely reduced,

with some forest sites even turning into CO_2 sources (Ciais et al. 2005). Lastly, leaping several decades, 1976 saw an intense heatwave in Great Britain which was considered a benchmark up to very recently (Baker et al. 2021): Climate change since the 1970s has increased the probability of a 1976-type heat by factor 10 in 2019.

While some of these events were confined to smaller regions (e.g., 1976), others spread across large parts of Europe during the course of several weeks (e.g., 2003, 2022). Knowing about preferential propagation directions in advance may thus aid early warning and, e.g., emergency plan activation against the imminent impacts of heatwaves.

2.1.2 Droughts: Creeping Disasters

Like heatwaves, droughts occur globally. Opposed to arid conditions that climatologically characterize regions in which precipitation cannot maintain vegetation (WMO 2023a), droughts describe multifaceted non-permanent water deficit events. According to the system affected, history provides numerous examples for (i) meteorological droughts (i.e., precipitation deficit), (ii) soil-moisture or agricultural droughts (i.e., deficits in soil water or plant-available water), (iii) hydrological droughts (i.e., streamflow deficits), and (iv) socio-economic droughts (i.e., human access to water is anthropogenically precluded). These categories were first suggested by Wilhite & Glantz (1985). Depending on the purpose of studies, several sub-types are used, e.g., snowmelt deficit droughts (Brunner et al. 2022). A fifth category, ecological droughts representing a lack of ecologically available water that pushes ecosystems beyond the limits of their adaptive capacity, was suggested by Crausbay et al. (2017).

Common drought features include a creeping onset (except for so-called flash droughts which recently caught attention, Otkin et al. 2018), long duration, and complex land–atmosphere feedbacks. Often, it is difficult to determine their ending. One reason for this is their tendency for propagating from, e.g., meteorological droughts to agricultural, hydrological, or socio-economic droughts (WMO 2023a; van Loon 2015).

Most literature on droughts define this hazard by distribution-based indices like the Standardized Precipitation Index (SPI) (McKee et al. 1993) and Standardized Precipitation Evapotranspiration Index (SPEI) (Vicente-Serrano et al. 2010) for meteorological drought, the Soil Moisture Index (SMI) for agricultural droughts (Zink et al. 2016), or

the Standardized Runoff Index (SRI) for hydrological droughts (Shukla & Wood 2008). Besides these definitions, the Percent of Normal Index (PNI) relates precipitation during a specific period to the long-term mean and is considered as easy to communicate (Willeke & Hosking 1994; Nikbakht et al. 2013; Falzoi et al. 2019).

Typically, droughts impact multiple sectors at different timescales: While precipitation may have returned to normal after a period of deficit, it may have been insufficient to refill soil water storage or raise groundwater levels. Thus, while the meteorological drought is past, soil water and groundwater drought continue and may prevail until the next meteorological drought or even pre-condition subsequent events (e.g., van Loon 2015).

In many cases, listing major drought years in Europe means revisiting some of the previously mentioned heatwave years: In 1976, water supplies in the UK were already depleted following an exceptionally dry winter, resulting in considerable water shortages in various sectors and subsequently water use restrictions (Baker et al. 2021; Rodda & Marsh 2011). Next, the hot and dry summer of 2003 saw financial impacts of roughly 11 billion Euro in central-western European agriculture and forestry (De Bono et al. 2004). 25 000 wildfires destroying several hundred thousands of hectare land and shutdown of nuclear reactors in France due to cooling water shortage were among further consequences of this year (De Bono et al. 2004). In 2010, a major drought event accompanied the heatwave in western Russia. Fatally, its timing and location coincided with winter and spring crop flowering, hence strongly reducing wheat production and export, and thus raising wheat prices in importing countries such as Egypt, Tunisia, or Turkey (Hunt et al. 2021). 2018 was characterized by particularly long drought conditions in precipitation, soil moisture, and river flow, with groundwater deficits prevailing beyond 2018 in southwestern Germany (Tijdeman et al. 2022). Germany further experienced particularly low soil moisture conditions, forest and crop damages, wildlife population collapsing, forest fires, and cooling water shortages (Conradt et al. 2023; Vogel et al. 2019). Lastly, in 2022 nearly all of Europe was subject to major drought conditions. Again, water shortages resulted in water-use restrictions (e.g., for irrigation), reduced power production, low reservoir or river levels, and considerable crop yield reductions (WMO 2023b).

Due to the harmful and often inter-seasonal impacts of droughts, these hazards are monitored meticulously in Germany (*Dürremonitor Deutschland,* UFZ n.d.) and Europe (European Drought Observatory, EMS 2023). These platforms also allow to track the increasing drought frequency and intensity during the last years.

The preceding paragraphs introduced heatwaves and droughts separately and illustrated the toll they have on human and natural systems. Along the way, though, the tight connection of both hazards became apparent. This co-occurrence bears the risk of compounding impacts of both extreme event types. Therefore, events or seasons with both extremely dry and hot conditions can be considered as compound events of climate hazards (e.g., CDHEs, Zscheischler et al. 2020). In terms of preparation and adaptation, a temporal co-occurrence imposes considerable challenges: During heatwaves, the demand for water rises (e.g., cooling, transpiration), yet its availability drops during droughts. Thus given their apparent co-occurrences, how are heatwaves and droughts interrelated in Europe?

2.2   Drivers of Atmospheric Variability at Different Scales

Climate extremes arise from natural climate variability within the chaotic climate system. Atmospheric conditions fluctuate on various spatial and temporal scales – from daily local land–atmosphere processes (Section 2.2.1) to persistent weather patterns (Section 2.2.2) and seasonal hemispheric variability mode regimes (Section 2.2.3). These fluctuations are mostly manifestations of physical processes evening out energy imbalances within the Earth's climate system (Lehner & Deser 2023). Their interplay can inhibit or facilitate the occurrence of climate hazards like heatwaves or (meteorological) droughts in Europe as shown in the following survey on current research regarding heatwave and drought drivers.

2.2.1   Land–Atmosphere Feedbacks

At the interface of land surface and atmosphere, surface properties play an important role in governing heat and moisture transports (Figure 2, lower right). Land forcing, governed by different land cover or management such as irrigation, and aerosol effects further contribute to heating processes on the local scale (Barriopedro et al. 2023, also
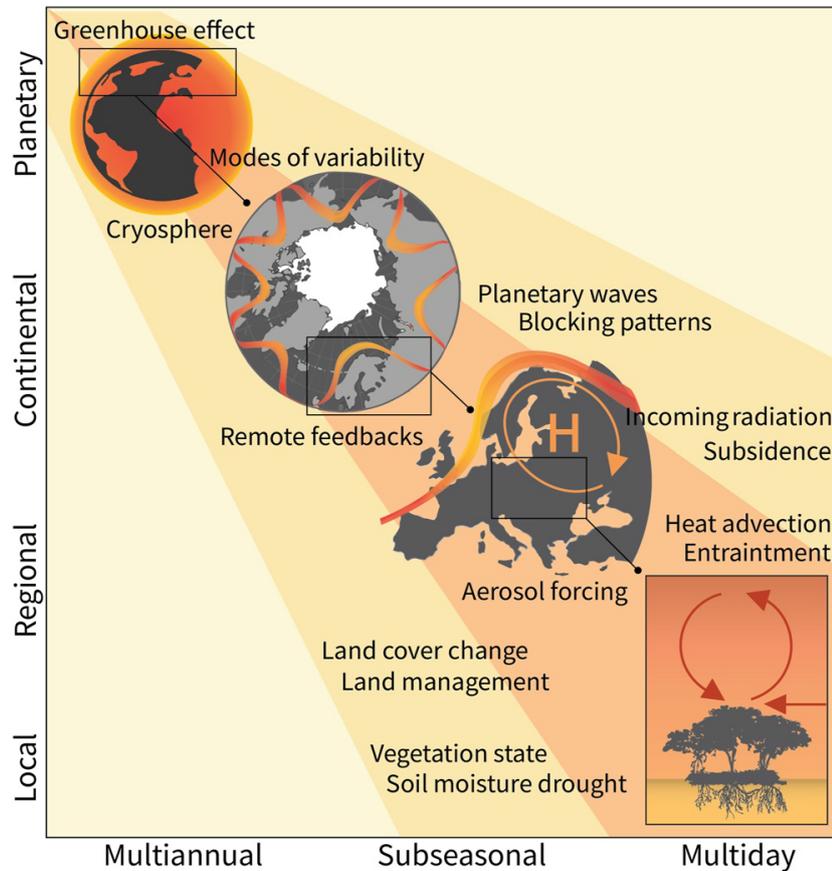
Figure 2: Drivers of heatwaves at various spatio-temporal scales (adapted from Barriopedro et al. 2023). (Meteorological) Drought formation is related to similar drivers.

for further details on heatwave drivers). This results in high spatial heterogeneity of temperature and humidity (Schönwiese 2020). Most important to energy and moisture transport, the turbulent heat flux, a component of the Earth's energy budget, is separated into sensible and latent heat fluxes. Being contingent on the specific heat capacity of a surface or body, the sensible heat flux describes heat that is provided to the environment by raising its ambient temperature. The latent heat flux describes warmth "hidden" within the process of evapotranspiration and "(re-)appearing" in condensation (Schönwiese 2020).

Evapotranspiration, i.e., evaporation from soil or open water and vegetation transpiration, plays a crucial role with regards to heatwave and drought development. Connecting the energy and water cycle, this process is governed among others by the interplay of radiation trends, precipitation, and available soil moisture (Teuling et al. 2009; Schönwiese 2020; Seneviratne et al. 2010). Evapotranspiration results in ambient cooling due

to heat extraction in the process. Its tight relation to incoming shortwave radiation leads to a pronounced diurnal cycle, hence adding to variability also on short temporal scales (e.g., Schönwiese 2020). In Europe, evapotranspiration is typically limited by insufficient energy availability, whereas some dry regions can be characterized as moisture-limited (Schwingshackl et al. 2017; Teuling et al. 2009; Zscheischler & Seneviratne 2017). These regimes occasionally change due to meteorological conditions (Schwingshackl et al. 2017; Zscheischler et al. 2015).

Soil moisture governs the partitioning into sensible and latent heat fluxes (Zscheischler et al. 2015). In particular, the fraction of sensible heat increases over dry soils (with less evapotranspiration) compared to wet soils with increasing incoming radiation, wherefore temperatures rise (Denissen et al. 2021; Seneviratne et al. 2010). However, heating over dry soils may also be attenuated by transpiration if vegetation access deeper (and occasionally moister) soil layers (Williams & Torn 2015). Often, low soil moisture conditions, especially in deeper horizons, prevail for a long period: Legacy or memory effects, e.g., dryness in the previous season, considerably contribute to heatwave intensity by reducing the potential for latent cooling (Fischer et al. 2007; Seneviratne et al. 2010). Higher temperatures, however, correspond to higher evaporative demand in the atmosphere due to enhanced water holding capacity (Seneviratne et al. 2010). Since evapotranspiration is fueled by soil moisture – also via vegetation – it may ultimately deplete soils, given sufficient energy availability (e.g., Seneviratne et al. 2010).

Local and remote effects of land–atmosphere coupling may further amplify heatwaves, by, e.g., elongating them (Fischer et al. 2007). For instance, the expansion of heatwaves and droughts is linked to upwind droughts via land–atmosphere feedback (Schumacher et al. 2019, 2022). Elongation and intensification may also result from heat accumulation over several days that is facilitated by night-time heat storage in higher atmospheric layers (Miralles et al. 2014). Regarding current extreme temperature variability, the role of local thermodynamic drivers is minor compared to dynamic conditions when considering both driving types together (Suarez-Gutierrez et al. 2020a).

With global warming, net radiation availability is projected to rise in Europe (Denissen et al. 2022). Its effects on soil moisture availability are difficult to determine: Presumably, ecosystems increasingly encounter less energy-limited conditions and more (soil) moisture-limited conditions, partly owing to intensification of current conditions

or regime shifts (Denissen et al. 2022). Thermodynamic effects are projected to regionally outpace the impacts of dynamic conditions on variability changes of monthly temperature extremes (Suarez-Gutierrez et al. 2020a).

2.2.2 Atmospheric Circulation Patterns

Leaving the land surface and climbing several geopotential height levels, one may define the typical meso-scale atmospheric circulation patterns governing every-day European weather (Figure 2, center right). Atmospheric circulation plays a major role in regional climate (e.g., Shepherd 2014). Corresponding patterns describe recurrent positions of extra-tropical cyclones (low pressure systems) and anticyclones (high pressure systems), troughs and ridges to each other. The most common patterns of relevance to European weather were cataloged by Hess & Brezowsky (1969) and recently re-categorized using pattern correlation techniques (James 2006) and a deep learning algorithm (Mittermeier et al. 2022). The interplay and sequence of cyclones and anticyclones governs the transport of moisture and heat onto Europe and shapes the local weather.

For heat, drought, and compounds thereof, middle to upper troposphere anticyclonic conditions are of relevance (Röthlisberger & Martius 2019; Pfahl & Wernli 2012; Black et al. 2004; Stefanon et al. 2012; Kornhuber et al. 2019). Anticyclonic conditions often prevail for several weeks, with pattern persistence fostering the occurrence and intensity of heatwaves and droughts (e.g., Hoffmann et al. 2021; Kyselý 2008; Röthlisberger & Martius 2019).

Often, quasi-stationary anticyclones block the zonal atmospheric flow onto Europe wherefore these conditions are termed atmospheric blocking (Rex 1950; Barriopedro et al. 2006; Kautz et al. 2022). Hence, precipitation-bearing frontal systems are deviated and rainfall deficits ensue below the blocking system whereas neighboring regions experience above-average precipitation (Sousa et al. 2017; Kautz et al. 2022). As opposed to low pressure conditions, tropospheric anticyclonic and blocking conditions are associated with subsidence processes (Kraus 2004; Kautz et al. 2022). Subsidence, i.e., descending air masses, results in *in situ* diabatic or adiabatic heating (Sousa et al. 2018; Bieli et al. 2015; Pfahl & Wernli 2012). Under these conditions, clouds dissolve and convection is inhibited. On one side, these conditions impede precipitation. On another,

cloud free conditions enhance incoming shortwave radiation (Bieli et al. 2015). Radiation changes in turn may affect soil moisture depletion and enhance heating (Section 2.2.1).

Consequently, heat extremes and droughts tend to coincide spatially with anticyclonic conditions, as, e.g., heatwaves in 1976, 2003, 2010, and 2018 or the 2003 and 2010 droughts (Kornhuber et al. 2019; Suarez-Gutierrez et al. 2020a; Kautz et al. 2022). In some cases, though, heatwaves and droughts are also associated to advection of hot and dry air masses, e.g., commonly on the Iberian Peninsula (Stefanon et al. 2012) or in central Europe in 2018 (Hoy et al. 2020).

Climate change is projected to alter the atmospheric circulation, pattern sequences, or their dynamical properties (e.g., persistence, Kautz et al. 2022; Mittermeier et al. 2022; Hoffmann & Spekat 2021). For instance, Coumou et al. (2018) argue that Arctic amplification, i.e., the faster warming of high latitudes under climate change, supports the development of hot and dry conditions: Due to the decreasing equator-to-pole temperature gradient, summer storm tracks weaken which is advantageous for developing hot and dry conditions in Europe. However, while thermodynamic indicators show robust trends associated with high confidence, dynamic aspects are subject to high uncertainty with low confidence in their trends and hardly distinguishable from climate variability (Shepherd 2014). Hence, projections of hazards under dynamic controls, e.g., regional precipitation, also suffer from these uncertainties (Shepherd 2014). As already mentioned, Suarez-Gutierrez et al. (2020a) project a decreasing dominance of high pressure conditions in driving extreme temperature variability with progressing climate change.

2.2.3 Atmospheric Modes and Teleconnections

Lastly, on hemispheric to global spatial scales and seasonal to decadal temporal scales, intrinsic atmospheric modes govern natural climate variability (e.g., Hurrell & Deser 2009; Delworth & Zeng 2016; Deser et al. 2017). They introduce low-frequency fluctuations that change the atmospheric background state against which short-term weather occurs (Figure 2, center left). These background states, regimes, may modify the magnitude or frequency of extreme events (Hurrell 1995; Scaife et al. 2008; Trenberth 2012).

A highly volatile mode of atmospheric variability with direct impacts on European climate is the North Atlantic Oscillation (NAO). First described by Walker & Bliss (1932)

and reproduced in various studies since, this teleconnection pattern represents a merid-ional sea level pressure (SLP) gradient over the North Atlantic between the Azores High and Icelandic Low that can be related mechanistically to atmospheric wave breaking (Wallace & Gutzler 1981; Barnston & Livezey 1987; Benedict et al. 2004; Roedel & Wagner 2011). Teleconnections generally refer to "contemporaneous correlations be-tween geopotential heights on a given pressure surface at widely separated points on earth" (Wallace & Gutzler 1981, p. 784). In case of the NAO, a strong gradient (NAO positive) strengthens the zonal atmospheric flow onto Europe including a northeast shift of Atlantic storm-tracks (Trigo et al. 2002; Hurrell & Deser 2009). On the con-trary, a weak or even inverted gradient (NAO negative) leads to dominant influence from Eurasian air masses in Europe (Hurrell 1995; Hurrell & Deser 2009; Iles & Hegerl 2017). Therefore, NAO positive winters tend to be mild and stormy in northern Eu-rope due to warm moist air mass advection, and cooler and drier in the south (inverted during NAO negative winters, Hurrell 1995; Trigo et al. 2002; Roedel & Wagner 2011). The NAO occurs year-round, but correlates most strongly with weather and climate in winter (Hurrell & Deser 2009). Negative phases dominate in summers, positive phases in winters (Hoffmann & Spekat 2021).

On the European scale, several other SLP seesaws operate, typically affecting weather by alternately advecting air masses of different temperature and moisture properties. Examples are the Scandinavian and the East Atlantic pattern, both of which also inter-play with the NAO (e.g., Comas-Bru & McDermott 2014). The Atlantic Multidecadal Oscillation induces (multi-)decadally lasting warmer or cooler phases in close connection with the ocean and correlates strongly with summer temperatures in Europe (Hoy et al. 2020). The most famous ocean–atmosphere teleconnective pattern, El Niño/Southern Oscillation (ENSO), causes global temperatures to rise to extreme heights due to heat releases in the Pacific ocean (e.g., 1997/98, 2015, or 2023, ECMWF 2023; Latif & Keenly-side 2009).

Depending on the background atmospheric mode regime, certain circulation patterns are favored: For instance, a moderate, positive relationship between positive summer NAO and anticyclonic conditions in northwest Europe is suggested by Folland et al. (2009). As mentioned previously, this corresponds to cloud-free, dry and warm conditions which can be enhanced by land–atmosphere feedback (Section 2.2.1 and 2.2.2).

The impact of climate change on atmospheric modes is difficult to discern (Figure 2, upper left). In case of the NAO, shrinking Arctic sea ice extents were associated to more frequent negative phases (Warner 2018). The sea ice extent in turn was related to ocean heat transport governed by NAO variations (Delworth et al. 2016). Moreover, atmospheric modes may reinforce or attenuate climate change effects. For instance, ENSO and NAO are supposed to have enhanced global warming during several decades (e.g., Hurrell & Deser 2009). Opposed to that, the NAO was shown to have masked regional warming trends in northern Europe between 1988–2012 (more so in winter, less in summer, Iles & Hegerl 2017).

Projections of future regime prevalence are subject to high uncertainty, mostly due to high variability of the modes (Deser et al. 2017; Maher et al. 2018). For instance, while Cattiaux et al. (2013) show an increase of NAO negative phases under global warming in single simulations of Coupled Model Intercomparison Project (CMIP)5 models, the results of McKenna & Maycock (2021) suggest a prevalence of positive NAO phases in a suite of SMILEs (see Section 3.3). For summers, mostly circulation patterns corresponding to positive NAO phases are projected (Folland et al. 2009).

To summarize Section 2, there is evidently no single cause for typically high-impact heatwaves or droughts to occur, but an interplay of various favorable drivers and conditions at different spatial and temporal scales. This interplay in turn is subject to considerable knowledge gaps: Statistical relations being established, the mechanisms linking (dynamic) drivers and, e.g., heatwave occurrence still require further investigation (Barriopedro et al. 2023), just as land–atmosphere feedbacks do (Seneviratne et al. 2010).

3   Distilling Spatio-Temporal Structures of Climate Hazards from Big Data

Section 3 provides an overview on the state of the art of tools and concepts to explain structures from multidimensional data (i.e., space, time, and model members). As mentioned in Section 1, structures refer to empirical, often directed, sometimes causal relationships among variables which presumably relate to distinct physical processes. They answer the question: How are specific components connected? This thesis investigates structures among multiple variables (e.g., teleconnection index and temperature) or the same variable at different locations (e.g., time series of temperature above a heatwave threshold).

A survey on already used dependence structures among variables opens this section (Section 3.1). The next part is dedicated to global and regional climate modeling with a focus on the considered scales (Section 3.2), leading to an introduction of large climate model ensembles (Section 3.3). The last part introduces to causal concepts in data sciences (Section 3.4). Each section closes with related research gaps.

3.1   Structures Among Multiple Variables

As alluded to in Section 2, summer land temperature and precipitation are known to be anticorrelated, such that warm summers often coincide with dryness, but rarely with humid conditions (Trenberth & Shea 2005). This structure can be related to physical processes. Ignoring the co-variability of, e.g., temperature and precipitation may confound the interpretation of their univariate time series (Trenberth & Shea 2005).

The (mutual) dependence of variables contributes to the complexity of analyzing the climate system. Often, it is a function of the univariate distributions: For instance, summer temperatures and precipitation regionally show a stronger dependence in their hot and dry distributional tail (i.e., their extremes) than in the main part of the distribution (Figure 1 (b), Wazneh et al. 2020; Zscheischler et al. 2020).

To address these characteristics, compound event analysis brought forth a suite of tools such as fitting multivariate distributions (copulas) to the dependent variables. Compound events generally describe a co-occurrence of several (often extreme) events

(Zscheischler et al. 2020): These can be (i) multivariate, e.g., hot and dry summers, (ii) pre-conditioned, e.g., rain-on-snow floods, (iii) temporally compounding, e.g., temporal clustering of storms, or (iv) spatially concurring, e.g., floods in several parts of Europe. Often, the joint impacts of their single event components are enhanced under these compound conditions (e.g., Leonard et al. 2014, Section 2.1.2). This underlines the high relevance of compound events for regional adaptation planning because accounting for dependence among hazards provides a better estimate of their occurrence probability and may render related extreme impacts more foreseeable (Zscheischler et al. 2018; Zscheischler & Seneviratne 2017).

Adaptation planning for instance strongly relies on the concept of return levels. Return levels describe the event magnitude that occurs with a fixed probability per year. A univariate example is the volume of water during a flood that occurs once every hundred years on average (100 years being its return period then). This value serves as benchmark for flood protection infrastructure in Bavaria (with return levels changing under non-stationary climate conditions, StMUV n.d.; LfU n.d.). The multivariate case, pertinent to compound events, is more complex. For one, multivariate return periods are not unique because several combinations of univariate values can result in the same return period (e.g., Yue & Rasmussen 2002; Brunner et al. 2016). For another, defining an event in a multivariate manner typically results in a more specific description since multiple conditions are to be met (e.g., both temperature and precipitation exceeding a certain threshold, AghaKouchak et al. 2014; Zscheischler & Fischer 2020). Therefore, precisely defining multivariate structures is crucial for inferring actionable metrics to base adaptation on, especially since multivariate events tend to be more rare than univariate events and hence gaining generalized information on them is subject to considerable uncertainty (AghaKouchak et al. 2014; Bevacqua et al. 2017; Zscheischler & Fischer 2020).

By virtue of their high impacts and relevance to adaptation planning, compound events thus provide a good occasion to apply the structure concept. Aside from compound events, structures also exist in the concept of teleconnections when atmospheric regimes condensed to index time series are related to climate impacts (e.g., via correlation or regression Hurrell & Deser 2009, Section 2.2.3). These are often used to explain climate variability in, e.g., Europe (Hurrell & Deser 2009; Comas-Bru & McDermott 2014).

Furthermore, remote effects, as, e.g., in dry conditions intensifying distant heatwaves (Section 2.2.1, Schumacher et al. 2019), may be considered as spatial structures and then evaluated themselves (e.g., are they recurrent? varying? transferable?). Knowledge on multivariate structures is further relevant when addressing bias correction for, e.g., impact modeling: To retain coherent relations among variables, their interrelations require consideration (Cannon 2018).

Aside from teleconnection studies, established structures so far were only rarely examined themselves (exceptions for compound events include, e.g., Wazneh et al. 2020, but on a very small area). How do they vary in space on regional scales? Mostly aggregated or coarsely resolved regions or single grid cells were investigated so far. How robust are they against sampling uncertainties and internal climate variability? Copulas were often fitted to limited observational records in order to interpolate and extrapolate compound events of varying degree of extremeness. Yet, the dependence structure itself may also be contingent on the available data sample or epoch under consideration as was sometimes acknowledged by bootstrapping, though from the same sample (e.g., Poschlod et al. 2020). Lastly, how do these dependence structures change under climate change conditions as suggested by Zscheischler et al. (2020)? While, e.g., tail dependence changes were investigated in some cases on the global scale (Zscheischler & Seneviratne 2017), spatially distributed information on a regional scale are still missing.

3.2   On the Brink of (Spatial) Resolutions

Why do scales matter? Weather phenomena and their impacts occur on a wide range of characteristic temporal and spatial scales as shown in Section 2.2 (e.g., Clark 1985; Schönwiese 2020): For example, while a tornado in the wake of thunderstorm cells swirls through very narrow corridors (scale of several meters) during the course of minutes, a hurricane originates from meso-scale synoptic cyclones and brings havoc to kilometer-wide areas for several days. When synthetically reproducing these hazards, i.e., modeling them, for understanding and projection purposes, their scale becomes important. Global Climate Models (GCMs) typically provide output at the scale of several dozens to hundreds of kilometers with time steps of multiple hours. Clearly, the hurricane can be resolved in space and time in a GCM, while the tornado requires higher spatial and temporal resolution.

Higher spatial and temporal resolutions thus offer the potential, but also the need to consider processes on a more detailed basis in climate modeling (e.g., Dudhia 2014): To simulate the above-mentioned tornado, a model in question is required to resolve convective processes governing the thunderstorm. For the hurricane (sub-)tropical circulation processes are sufficient, though not necessarily for all related impacts. As a second example, the large-scale dynamic drivers of heatwaves and droughts are spatially rather homogeneous on small scales, whereas land–atmosphere coupling depends on the heterogeneous land surface properties. These properties comprise, e.g., inland water bodies, urban areas, or orography leaving a "signature" in the climate (Rummukainen 2016). Balaji et al. (2022), though, argue that there is no "target resolution" to resolve all relevant processes in the climate system.

3.2.1   From Global Climate Modeling ...

Anthropogenic climate change is effective globally and at characteristic time scales of several centuries (Schönwiese 2020; Clark 1985). Next to observational records, the most important tools to investigate climate change at global and decadal scales, both in the past or in the future, under plausible or counterfactual scenarios, are climate models. Basically, they translate the fundamental knowledge on climate into code (Balaji et al. 2022). Climate models, or more specifically, general circulation models and Earth system models, are comprised of interconnected collections of physical (and chemical) equations (Roedel & Wagner 2011). Based on fundamental laws of physics, they mirror in varying degrees of coupling and detail atmosphere, ocean, land, and ice and feedbacks thereof (e.g., Shepherd 2014). Climate models are designed to conserve mass and energy in process representations of coupled model components (Balaji et al. 2022; Roedel & Wagner 2011). The global land surface, the ocean, and the atmosphere are discretized into homogeneous volume increments for which the model equations are solved numerically (Roedel & Wagner 2011). Some components, e.g., cloud processes, are parameterized due to insufficient spatial (horizontal, vertical), temporal, and process-related resolution, giving rise to model structural uncertainty in simulations (Balaji et al. 2022; Roedel & Wagner 2011). The difference between general circulation and Earth system models being irrelevant for this thesis, both will be referred to as Global Climate Models (GCMs) as opposed to Regional Climate Models (RCMs) hereafter.

High complexity and degree of detail notwithstanding, GCMs simplify real world relationships. This is less a shortcoming, but rather a strength of these tools as it allows to test hypotheses on process representations without the need to consider every detail (e.g., Balaji et al. 2022). A good climate model is able to reproduce atmospheric variables as empirically observed. The better it reproduces current conditions, the more trustful it is considered (Roedel & Wagner 2011). Balaji et al. (2022) provide a discussion of shortcomings and criticisms on climate models, including problems related to unresolved terms (i.e., parameterization), model calibration, and the emerging structural uncertainty among multiple models. However, the authors conclude by considering climate models as "indispensable" (p. 9).

For statements on the future of the global climate, scientists recur to various scenarios of societal development. These narratives describe hypothetical pathways (e.g., the Shared Socioeconomic Pathways (SSPs)) or GHG emissions resulting in varying radiative forcing (e.g., the "older" Representative Concentration Pathways (RCPs) that are now combined with SSPs) and hence offer a wide range of potential futures (O'Neill et al. 2017, 2020; Riahi et al. 2017; van Vuuren et al. 2011). Radiative forcing describes the influence of natural or anthropogenic drivers on the top-of-atmosphere energy balance (Forster et al. 2021). The climate system responds to changes in radiative forcing by adjusting temperature with humidity following, whereas dynamic responses occur indirectly via eddy feedback (Shepherd 2014). Based on the scenarios, any climate trajectory unfolding in climate simulations responds to the question "what could happen if", not "what will happen". Nevertheless, it is possible to compare observations against scenario projections. For the past years that were covered by projections from earlier decades, CMIP3–6 simulations tended to underestimate global warming (independent of scenarios, Carvalho et al. 2022). While Carvalho et al. (2022) thus suggest that, so far, observations follow at best the worst-case scenarios, it is possible that the observed trajectory will diverge from them in the future. Their plausibility and especially their widespread use are contested (Hausfather & Peters 2020b; Schwalm et al. 2020) since it is argued that yet recoverable coal resources would not suffice to raise global emissions to the levels of, e.g., RCP8.5 (Ritchie & Dowlatabadi 2017; Hausfather & Peters 2020a). RCP8.5, i.e., a non-mitigation scenario, prescribes GHG emissions resulting in 8.5 W/m² additional radiative forcing in 2100, with global temperature increases of 3 °C to 5 °C (van Vuuren et al. 2011; IPCC 2014).
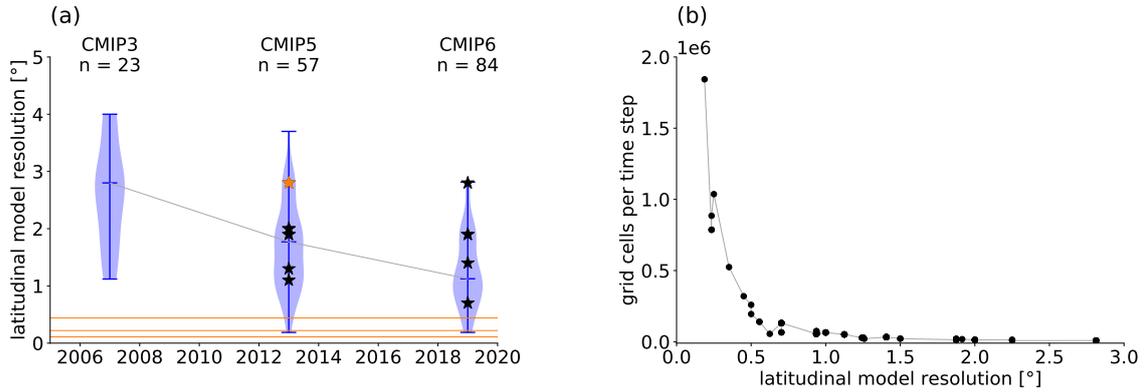
Figure 3: Horizontal resolutions of the atmospheric component in GCMs featuring in CMIP3, CMIP5, and CMIP6 (data sources: Appendix). (a) Violin plots of the latitudinal spatial resolution at publication dates of IPCC AR4 (2007), AR5 (2013/14), and AR6 (2021/2022). Stars mark the spatial resolutions of global SMILEs (orange star: CanESM2-LE). Orange lines at 0.44°, 0.22°, and 0.11° indicate typical spatial resolutions of RCMs featuring in CORDEX (see Section 3.3). Some CMIP6 models are part of the High Resolution Model Intercomparison Project (Haarsma et al. 2016). (b) Global grid cells per time step corresponding to the indicated latitudinal model resolution of CMIP6 models.

Yet, for impact-relevant studies the typical resolutions of GCMs are too coarse. Impacts are to be considered on a regional scale to account for implications from heterogeneous background fields and hence to provide meaningful information for process-understanding and adaptation planning (e.g., Rummukainen 2016). This requires sufficient representation of topography and land–sea contrasts. Therefore, it may be useful to pull out a magnifying glass to discern finer spatial patterns.

3.2.2   ... to Regional Climate Modeling

During the last years, spatial (also temporal, not shown) resolutions of GCMs continuously increased as illustrated in Figure 3 (a) for the models of CMIP phases 3–6. 50 km is a typical spatial resolution of recent GCM (Balaji et al. 2022). High spatial resolutions come at the cost of data volume inflation per time step (Figure 3 (b)). Thus, confining high-resolution investigations to distinct areas of interest is reasonable.

For regional understanding of the climate system various approaches exist. Here, the focus is set on dynamical downscaling performed by RCMs, i.e., limited area models, that are nested in GCM simulations (Roedel & Wagner 2011). As opposed to, e.g., statistical downscaling, dynamical downscaling offers physically consistent climate real-

izations (Rummukainen 2016). Nesting describes the coupling between the GCM and RCM, with one-way nesting referring to the GCM influencing the RCM only and two-way nesting referring to mutual information exchange (e.g., Rummukainen 2010; Lorenz & Jacob 2005). In general, the driving GCM provides lateral boundary conditions of which the RCM develops its own climate at higher spatial and temporal resolutions over a specific domain (for spatial RCM resolutions see orange lines in Figure 3 (a), Leps et al. 2019; Becker et al. 2015). Lateral boundary conditions provide inflow from the driving data to the RCM and allow for outflow of small-scale features from the domain (Leps et al. 2019). Often, a transition zone on the domain boundaries is excluded from further analyses owing to potential errors originating from spatial resolution jumps (e.g., Matte et al. 2017; Leduc et al. 2019; Leps et al. 2019).

Any deficiency in the driving GCM data propagates to the RCM ("garbage in, garbage out", e.g., Diaconescu et al. 2007; Giorgi & Mearns 1999). The control of lateral boundary conditions on the RCM climate, though, diminishes inside the domain with increasing distance to the inflow side and elapsing time since inflow (Lucas-Picher et al. 2008a,b): The longer air parcels stay within the domain, the stronger is the effect exerted by the RCM on them, thus enlarging biases towards the driving data. In the European domain, this corresponds to strongest control in the west due to the prevailing westerly wind belt. However, since large-scale atmospheric features, such as blocking, are considered reliable proxies for extreme events (Kay et al. 2015), their correct representation is crucial. Therefore, some models enable nudging the RCM climatology towards the GCM climatology within the domain for bias reduction (Leduc et al. 2019; Omrani et al. 2013).

The added value of RCM compared to GCM simulations is based on the assumption that large-scale features of the lateral boundary conditions are preserved, while small- and short-scale features and higher amplitudes of (extreme) events are developed in the RCM climate (e.g., Rummukainen 2016). Further, (i) surface forcing is supposed to be more realistically included by virtue of higher spatial resolution, (ii) the simulation of atmospheric circulation is improved due to better equation discretization, and (iii) more fine-scale processes are resolved explicitly (Lucas-Picher et al. 2017).

For instance, topographic features like the Alps may shape the delineation of areas impacted by hazards (e.g., Torma et al. 2015): While in RCM resolutions Alpine valleys
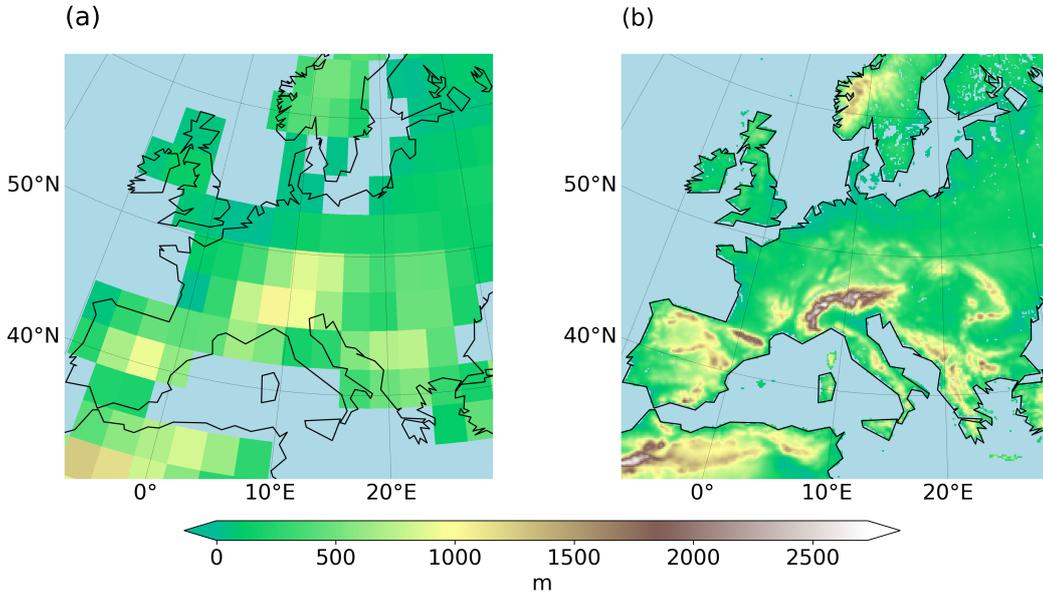
Figure 4: European topography in a coarse GCM (2.8°) and an RCM (0.11°).

exist within high mountains, the Alps are a four-grid cell lower mountain range in a coarsely resolved GCM model world (including, e.g., lower temperatures at the top of the mountains in the RCM; see Figure 4 for (a) the Canadian Earth System Model version 2 (CanESM2) and (b) the CRCM5). Similarly, the European coastline is well represented within the RCM, while the coarse GCM hardly knows of Italy, Greece, or the Adriatic Sea. The latter of course hampers small-scale effects due to the land–sea contrast (e.g., heat advection from oceans, Röthlisberger & Papritz 2023) and excludes regions of high concern to drought and heatwave examination.

In addition, regional heating or drying processes may result in more realistic event magnitudes (e.g., amplification) owing to more detailed spatial representation of atmospheric processes (e.g., Teichmann et al. 2013; Torma et al. 2015). Added value (mostly) increases with finer spatial resolutions which suggests that physically consistent processes drive the added value (Torma et al. 2015; Lucas-Picher et al. 2017; Rummukainen 2016). Added value is considered to vary in space (Barriopedro et al. 2023). It also emerges when aggregating RCM simulations to the coarser GCM resolution, but relies on the availability of high-quality observational data against which RCM simulations can be compared (Torma et al. 2015). However, e.g., heatwave representation does not generally improve when comparing coarsely (e.g., 0.44°) to highly (0.11°) resolved RCM

simulations, possibly because even 12 km is too coarse for some processes (Figure 3 (a), orange lines, Vautard et al. 2013; Sangelantoni et al. 2023).

Compared to GCMs, RCMs systematically reduce biases towards observations, presumably due to their enhanced spatial resolution, their parameterization, and tuning towards observations – though not necessarily when applying bias correction to GCMs or RCMs (Sørland et al. 2018; Paeth & Mannig 2013; Eden et al. 2014). When comparing GCMs and RCMs at similar (original) spatial resolutions (which is possible since recently, see Figure 3 (a)), both provide comparably adequate information (Demory et al. 2020). RCMs are capable of distilling climate change signals in high geographical detail, instead of drowning it in noise that is potentially induced by the downscaling procedure (Paeth & Mannig 2013). With very high resolutions and further explicit process representation instead of parameterization (of, e.g., convection in so-called convection permitting models), though, biases towards observations may further decrease (Hundhausen et al. 2023; Sangelantoni et al. 2023).

Yet, with all their opportunities and deficiencies in regional climate representation, are RCMs capable of representing hazards and structures thereof reasonably – both with respect to their driving data and observations? Since RCMs are considered to improve hazard simulations and, furthermore, structures constitute important properties of hazards, challenging the ability of RCMs to represent structures in high geographical detail is crucial to employing RCMs as tools for hazard projections and thus adaptation planning.

3.3   Explaining Internal Variability from Climate Model Ensembles

Projections of climate change as provided by climate models are subject to considerable uncertainty. This uncertainty is typically partitioned in three categories: (i) model or climate response uncertainty, (ii) radiative forcing or scenario uncertainty, and (iii) internal variability (Hawkins & Sutton 2009; Lehner et al. 2020). Relative contributions of these uncertainty sources vary from near-term to long-term projections (Hawkins & Sutton 2009, 2011; Lehner et al. 2020).

While the first two can be addressed by compilations of simulations (so-called ensembles) from several models and forcing scenarios, the third one is more challenging. Strictly speaking, it even adulterates results on (i) and (ii) if only one realization per model and scenario is employed since it cannot be ruled out that any effect (e.g., model A being "hotter" than model B, or scenario 1 leading to more heatwaves than scenario 2) occurs simply by chance during the period under consideration due to climate variability (e.g., Maher et al. 2019; Kay et al. 2015; Deser et al. 2020; Taylor et al. 2012b). With the rise of computational power and storage resources, however, climate models can be run several times with the same parameterizations and external forcing, but varied initial conditions. These so-called SMILEs already proved to be highly useful tools in investigating forced climate change and internal climate variability.

In the context of SMILEs, internal variability refers to unforced climate variability around a certain mean state arising from the (internal) chaotic nature of the climate system (e.g., von Trentini et al. 2020; Lehner & Deser 2023). It may be triggered by thermodynamic and dynamic processes (Section 2.2) and, as claimed by Lehner & Deser (2023), "affects virtually every aspect of the climate system" (p. 1). Internal variability was shown to vary across models (also between GCM and RCM, Suarez-Gutierrez et al. 2021; Maher et al. 2021), over time (Wood et al. 2021), and it is suspected to depend on the external forcing (Deser et al. 2020; Lehner et al. 2020).

The following sections introduce the reader to the creation of SMILEs (Section 3.3.1) and their current use in climate change research (Section 3.3.2). A dedicated focus will be set on the generation and use of the CRCM5-LE, a regional SMILE produced in the Climate Change and Hydrological Extremes (ClimEx) project (Leduc et al. 2019).

3.3.1 How to Set Up an Ensemble (Fast)

The ensemble approach in climate science was first dominated by multi-model ensembles of opportunity: For instance, the CMIP initiatives aimed at collocating comparable model outputs by prescribing guidelines for modeling groups. These served as a data base for IPCC reports AR4 (CMIP3, Solomon et al. 2007), AR5 (CMIP5, Taylor et al. 2012b; Stocker et al. 2013), and AR6 (CMIP6, Eyring et al. 2016; Masson-Delmotte et al. 2021). Coordinated simulation protocols with respect to domain choice, spatial resolution, or model outputs were also developed for regional simulations, namely in the

Coordinated Downscaling Experiment (CORDEX) project (Giorgi et al. 2009). With multi-model ensembles of opportunity, the dependence of results on a single model and its model structure can be circumvented, provided that all models run the same forcing scenario. However, many models share modules or parameterizations. Their realizations may be not independent and hence aggregating across ensembles of opportunity without considering these dependencies poses the risk of biased inferences (Abramowitz et al. 2019).

As opposed to multi-model ensembles, the members of a SMILE differ only by their initial conditions, everything else held constant (e.g., Deser et al. 2020; Kay et al. 2015). Initial conditions may describe strongly diverging ocean states (macro initialization) or small perturbations of the atmosphere (micro initialization) (Hawkins et al. 2016; Deser et al. 2020). Owing to the chaotic properties of the climate system, tiny perturbations result in multiple randomly phased storylines around a mean state that become independent after a few years when running the same model several times (except maybe for deep ocean states, Deser et al. 2020; Leduc et al. 2019; Fyfe et al. 2017; Kay et al. 2015; Lehner & Deser 2023). One storyline may show a slight drying trend over central Europe, while another projects wetting in all but northern Europe, in parts also opposing the anthropogenically or naturally forced trends. Nevertheless, both are equally plausible under the same external forcing (Leduc et al. 2019; Hawkins & Sutton 2009; von Trentini et al. 2020).

Several global SMILEs based on different models and external forcing scenarios are available (Deser et al. 2020). They allow to disentangle all types of uncertainty mentioned by Hawkins & Sutton (2009). Among them, the CanESM2-LE was one of the first to be run 50 times (Kirchmeier-Young et al. 2017; Fyfe et al. 2017). It uses a combination of macro and micro initialization that has been suggested to be of particular scientific value in estimating internal variability (Deser et al. 2020).

Its then outstanding ensemble size predestined the RCP8.5-run CanESM2-LE to be selected for a unique experiment: dynamically downscaling a global SMILE to two regional domains (Europe, northeastern North America, Leduc et al. 2019). In the Bavarian-Québec project ClimEx, hydrometeorological extreme events and their impacts are investigated in high spatial and temporal detail, which require an extensive data base (Leduc et al. 2019). Therefore, each CanESM2 member provided the lateral boundary

conditions to a regional climate simulation with the CRCM5. Only few other (smaller) regional SMILEs exist so far (e.g., Aalbers et al. 2018). Comparing the CRCM5-LE to the multi-model ensemble CORDEX, von Trentini et al. (2019) found that a large portion of the multi-model ensemble spread is actually due to internal variability, a finding that corroborates results on global models by Deser et al. (2012). The authors suggest that each model contribution to the multi-model ensemble should be thought of as being embedded in a "cloud" representing internal variability (von Trentini et al. 2019) – which, on a global scale, can be achieved by means of the Multi Model Large Ensemble Archive (MMLEA) (Deser et al. 2020).

Global SMILEs are extensively validated (e.g., Suarez-Gutierrez et al. 2021). But while the existing regional SMILEs are compared against observations or other ensembles (e.g., von Trentini et al. 2019), estimates of climate variability in the GCM–RCM pipeline are missing. Since internal climate variability, though, is shown to be (global) model dependent (Suarez-Gutierrez et al. 2021; Deser et al. 2020; Maher et al. 2021), it may be suspected that a regional model also affects variability – and subsequently anything that may be found within its ensemble spread.

3.3.2   SMILE-ing Climate Changes: On Trends, Internal Variability, and Method Test Beds

The emergence of SMILEs gave rise to a wide range of applications in climate sciences. First and foremost, the numerous simulations of virtual weather sequences that share comparable climate statistics allow to capture internal climate variability more comprehensively than single model realizations (Figure 5, Kay et al. 2015; Deser et al. 2012; Leduc et al. 2019). As shown by the diverging dotted lines, internal variability may change transiently. When evaluating SMILEs, internal variability turns out to be a challenge: In reality, there is only one realization of climate. Given natural climate variability, how can it be related to several members of an ensemble? No single ensemble realization can be expected to meet the characteristics of the observation due to different phasing of ensemble internal variability, let alone the cross-ensemble mean or median that average out short-term variability (e.g., Kay et al. 2015; Maher et al. 2019). Therefore, Suarez-Gutierrez et al. (2021) and Maher et al. (2019) proposed a framework that compares the distributions of observations and SMILEs, including as-
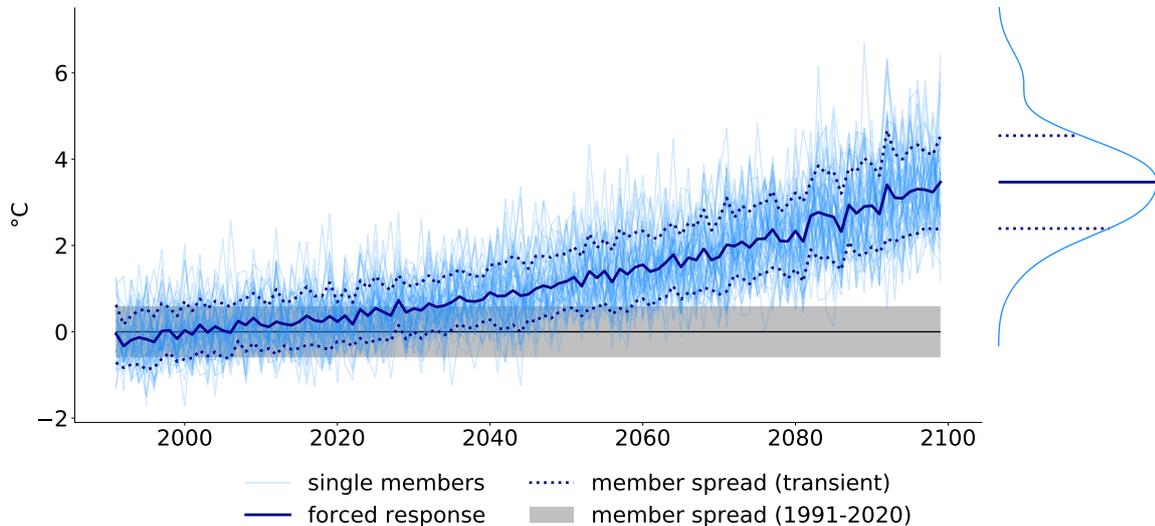
Figure 5: Internal variability confounding the forced response in the large ensemble data of the CRCM5-LE. Member spread refers to the (often used) cross-member standard deviations and represents transient internal variability. Blue curve on the right: (vertical) cross-member distribution in 2100. The variable used for illustration is mean heatwave temperature changes under RCP8.5 at a grid cell representing Munich/Germany.

sessments of variability (i.e., ensemble spread) over- and underestimation. However, the ensemble spread should not be seen as an envelope with sharp boundaries outside of which the occurrence probability of events drops to zero. Rather, probabilities approach zero asymptotically (Figure 5, vertical curve on the right side), thus mirroring that the envelope is only an estimate based on a limited number of members.

Secondly, a robust estimate of internal variability allows to separate climate change signals from noise (Figure 5). Without the need for fitting trend functions, averaging across the members or pooling all members into one long time series typically results in robust trends (depending on the variable under consideration) that contrast against internal variability represented by the member spread (cf. the time of emergence concept) (e.g., Deser et al. 2020, 2012; Hawkins & Sutton 2009; Maher et al. 2019; Lehner et al. 2017). The cross-member distribution of values at a certain point in time (Figure 5, right), e.g., of heatwave temperatures in a given year from all members, also provides a means for estimating uncertainty associated with single values (e.g., 10 % of all members show values above 20 °C) or of the ensemble mean estimate (e.g., 15 °C $\pm$ 1 °C). Alike analyses in multi-model ensembles are impeded due to their incorporation of physically differing and

sometimes even unrealistic models (Meehl et al. 2020). Trends of thermodynamic variables like temperature are strong enough to emerge rapidly in small SMILEs, whereas precipitation, dynamically driven variables, or complex index values require larger ensemble sizes (Maher et al. 2018; Wood et al. 2021). Especially for temperature and (heavy) precipitation extremes, internal variability constitutes a considerable fraction of total uncertainty until the mid to end of the 21st century (Blanusa et al. 2023).

Thirdly, internal variability of SMILEs encompasses extreme events (peaks of single members in Figure 5), some of which may exceed observed maxima (Fischer et al. 2021). Abundant data samples provide the means to investigate low-probability events and their frequency changes due to climate change. They allow empirically assigning an extremeness estimate like 1-in-100 year frequency instead of recurring to extreme value statistics to parameterize the often insufficiently sampled distributional tails (Willkofer et al. 2023; van der Wiel et al. 2019, for floods). Further examples of low-probability, but high-impact event investigation in SMILEs include extreme heat (e.g., Suarez-Gutierrez et al. 2020b), rare compound events (Poschlod et al. 2020), seasonal extreme precipitation (e.g., Wood & Ludwig 2020), or storm intensity in tropical cyclones (Bruyère et al. 2022).

The ability of SMILEs to compare phenomena across members further allows for the testing and employment of novel, "data-greedy" methods to search complex spatio-temporal patterns. For example, Vb cyclones, a specific weather pattern that is related to extreme precipitation in southern Germany, were tracked in the ensemble using machine learning techniques (Mittermeier et al. 2019). Additionally, the combination of several types of SMILEs allows for robust attribution estimates of, e.g., changes due to anthropogenic or natural forcing or spatial resolution changes (Maher et al. 2021).

To summarize, there is a lot to be found within the margins of the ensemble range. Among others, skillfully defining internal climate variability is crucial to deciding whether a new climate state of a given variable is reached. Most recent investigations focused on certain parts of value distributions (e.g., central values, extremes, and variability) and attempted to gain significant (change) signals, sometimes taking ensemble spreads as uncertainty range due to internal variability into account. Relations between variables, however, were mostly examined within the pooling approach (e.g., compound events, teleconnections). While this allowed for statistically robust estimation of relationships,

internal variability of these structures remained often neglected. This challenges, e.g., the evaluation of observed relationships and modeled ones as it remains unclear whether potential discrepancies arise due to model deficiency or internal variability. As stated often in literature, the one "climate realization" that is observed in the real world cannot be expected to follow the ensemble mean. However, it should be encompassed in the ensemble spread (e.g., Suarez-Gutierrez et al. 2021). This refers to time series, mean values, or extremes, but should also be investigated for spatio-temporal structures. Moreover, the rise of purely data-driven methods calls for assessing the physical coherence of their results.

3.4   Causality and Causal Discovery

This section addresses a specific kind of dependence among variables, namely, causal relationships. Causality can be approached from various perspectives: Pearl & Mackenzie (2018)'s "Ladder of Causation" includes three steps, i.e., (i) association of variables, (ii) targeted interventions to evaluate their outcomes, and (iii) counterfactual examination of problems. In general, causality is related to understanding *what* factors contribute to an effect and *how* they contribute (Campaner 2011). Section 3.4.1 provides an overview on causality perspectives that are of relevance for this purpose. Subsequently, Section 3.4.2 introduces statistical methods for inferring causal relationships from data.

3.4.1   Two (and a Half) Flavors of Causality

The first concept of causality under consideration takes on a mechanistic perspective on cause and effect. This means that "all steps underlying an observed correlation are made explicit" (Campaner 2011, p. 1071). Mechanisms in this definition refer to single steps of a process and help in closing an explanatory gap between (putative) cause and effect. For instance, an important mechanism in relating anthropogenic GHG emissions to global warming is the absorption and reemission of radiation by CO_2 (e.g., Kiehl & Trenberth 1997; Schönwiese 2020). Mechanistic causality is often tied to theoretical knowledge on the relationship between cause and effect (Russo & Williamson 2007; Williamson 2019). This helps to explain and generalize their dependence (Russo & Williamson 2007). Researching causality of phenomena adds to process understanding, forecast, and model optimization (Camps-Valls et al. 2023). Yet often, identifying mechanisms from

observations and relating them in a multi-faceted mosaic of interactions and feedbacks remains challenging (e.g., Campaner 2011). Employing (climate) models may hence contribute in estimating whether sufficient mechanistic knowledge on processes exists (see Williamson (2019) for various kinds of mechanism establishing, though in medicine).

Besides the mechanistic definition, the statistical or difference-making perspective is widely used. It accounts for the question "Does the (putative) cause make a difference to the effect?", and allows for predicting (Russo & Williamson 2007). This perspective will be split in two flavors here, the first of which is a probabilistic perspective. Assuming X and Y to be two events, it can be defined that (i) Y can only occur if X occurred as well (necessary causality, counterfactual), and (ii) X always leads to Y (sufficient causality). In case (i), other factors than X may also contribute to Y, while (ii) cannot rule out that Y occurs without X preceding. Both (i) and (ii) can be assigned with a probability, which is 1 under perfectly deterministic conditions (Hannart et al. 2016).

This approach is well-established in climate science, especially in the field of attribution research. Here, the basic assumption is that event characteristics differ in worlds with or without a certain "manipulation" (i.e., an intervention in an experiment). In the most prominent case, this manipulation is running simulations with and without anthropogenic forcing (e.g., Hannart et al. 2016). External forcing being the only difference between the experiments, it follows that any change in the GHG including simulation is caused by the external forcing, assuming no differences due to internal variability (see van Oldenborgh et al. 2021, for a review on attribution research). This type of attribution research aims at pinning climate impacts to anthropogenic climate change to support climate litigation (e.g., Hannart et al. 2016) or at contextualizing extraordinary weather events (e.g., recent European heatwaves and droughts, Stott et al. 2004; Otto et al. 2012; Vogel et al. 2019; Schumacher et al. 2023). Ultimately, these studies aim to infer "whether anthropogenic forcing was necessary for the event to occur and whether it is sufficient for such events to continue to occur repeatedly in the future" (Kirchmeier-Young et al. 2017, p. 554). Other examples include modeling studies in which model realizations with a certain particularity are compared to realizations without it (e.g., Fischer et al. 2007, for the role of land–atmosphere coupling in heatwave amplification).

A related flavor describes a concept of predictive causality: "We say that Y_t is causing X_t, if we are better able to predict X_t using all available information than if the information

apart from Y_t had been used" (Granger 1969, p. 428). This "Granger causality" is conditioned on a specified set of variables within a linear model, but only bivariately and for stationary time series (Verdes 2005; Mosedale et al. 2006; Granger 1969). Still, there may be no physical causal mechanism linking X_t and Y_t (Kaufmann & Stern 1997). As an example from climate sciences, Kodra et al. (2011) showed that radiative forcing by GHGs "Granger causes" global warming, possibly supported by natural processes.

Typically, causality cannot be understood in a purely deterministic sense: Multiple causes usually contribute to effects (e.g., interactions of local and regional climate effects, or atmospheric interactions in addition to external forcing, Fischer & Knutti 2015; Campaner 2011). Hence, attribution studies of rare events often succeed in carving out necessary conditions (e.g., "this would [likely] not have happened without climate change"), while struggling in finding sufficient conditions (e.g., "climate change always causes this") (cf. Hannart et al. 2016; Kirchmeier-Young et al. 2017).

3.4.2   Discovering Causality

When assessing drivers for events, scientists often recur to correlating other variables to a given phenomenon, e.g., composites of atmospheric conditions or circulation indices (e.g., Hurrell & Deser 2009). However, correlation is not causation, it only provides a measure of symmetric similarity among variables (Deng & Ebert-Uphoff 2014; Verdes 2005). Williamson (2019) offers a list for various kinds of relationships that may underlie correlation among variables, but also warns against assuming correlation to be generally following from mechanistic causality.

The flavors of causality given in Section 3.4.1 imply a direction between cause and effect that cannot be flipped without inverting the causal chain. For example, while we claim that initially rising (anthropogenic) GHG emissions *cause* global warming, the opposite, i.e., initial global warming *causing* rising (anthropogenic) GHG emissions, is unequivocally considered implausible among climate scientists (e.g., Masson-Delmotte et al. 2021). Correlation, however, does not allow to decide on either direction or to unravel hidden common drivers (Kretschmer et al. 2016). Yet, dependent variables are generally assumed to be related causally either directly or via hidden common drivers (*Reichenbach's common cause principle*, e.g., Camps-Valls et al. 2023).

When quantifying causality based on data, physical processes (as, e.g., given in Section 2.2) are implicitly assumed to generate fingerprints in the data that can be discovered (Nowack et al. 2020). Causal discovery then is the attempt to "learn causal relationships from observational data" (Deng & Ebert-Uphoff 2014, p. 193) and "estimate the relative importance of various causal factors" (Spirtes & Glymour 1991, p. 62). Causal inference is often used synonymously to describe the identification of driving and driven systems (e.g., Palus & Vejmelka 2007). It addresses the quantification of systematic structures in the data (Runge et al. 2019a). An extensive review on causal discovery methods is provided by Camps-Valls et al. (2023).

Causal discovery as used in this thesis is strongly based on predictive causality as, e.g., in Granger causality. A big drawback of this foremost bivariate method, though, is that it cannot account for hidden common drivers (Runge et al. 2019b). However, multivariate dependence inference is covered by the Peter and Clark (developer surnames) (PC) algorithm. The PC algorithm is based on a Directed Acyclic Graph (DAG) (Spirtes & Glymour 1991). A DAG is a graphical model, comprising a set of vertices (nodes), connected by edges with a unique direction each that do not allow for cyclic paths along the edges (Ebert-Uphoff & Deng 2012b; Kalisch et al. 2012; Spirtes & Glymour 1991). The PC algorithm produces a sparse graph linking the variables under consideration (Spirtes & Glymour 1991; Deng & Ebert-Uphoff 2014; Kretschmer et al. 2017). The resulting structure indicates information flows, e.g., an edge like $X \longrightarrow Y$ implies an information flow from X (e.g., NAO index and its variations) to Y (e.g., winter temperature and its variations), or, put differently, "X is likely to have a causal effect on variable Y" (Deng & Ebert-Uphoff 2014, p. 194). If two nodes or variables are not linked directly by edges, it is assumed that no direct causal association exists (Runge et al. 2019a).

Algorithms applied to the climate system typically encounter high-dimensional data sets and abundant interdependencies among different variables or the same variable at different locations (Runge et al. 2019b). While the PC algorithm allows for linear, lagged relationships only (Kretschmer et al. 2017), extensions to this algorithm also include contemporaneous links among variables, as well as undirected ones (e.g., Runge 2018). These extensions are considered particularly useful for autocorrelated time series (Kretschmer et al. 2016; Runge et al. 2019a; Runge 2020).

Causal discovery is mostly discussed for observations (Runge et al. 2019a; Runge 2018). Applying causal discovery methods to complex spatio-temporal data typically includes (i) hypothesizing on potentially interacting variables, (ii) defining nodes within spatio-temporal data (dimension reduction) to circumvent, e.g., spatial auto-correlation, (iii) obtaining a network among nodes without spurious links or indirect paths using, e.g., conditional independence tests, (iv) quantification of causal links by, e.g., path analytic methods, and (v) theory-guided hypothesis testing of causal mechanisms underlying the links (Runge et al. 2015; Ebert-Uphoff & Deng 2012a,b; Kretschmer et al. 2016; Bahrenberg et al. 2008; Di Capua et al. 2020). As opposed to correlation networks, causal discovery methods retain significant, direct links only (Runge et al. 2019a).

Recent applications in climate sciences include causal hypothesis testing among multiple variables (e.g., Kretschmer et al. 2016), causal network analyses (Ebert-Uphoff & Deng 2012a), process-oriented causal model evaluation (Nowack et al. 2020; Karmouche et al. 2023b), analyses on information flow and hubs (e.g., the duration and distance to which the influence from a given location extends, Deng & Ebert-Uphoff 2014; Ebert-Uphoff & Deng 2012a; Runge et al. 2015), teleconnection analyses (Ebert-Uphoff & Deng 2012b; Karmouche et al. 2023b; Di Capua et al. 2020), land–atmosphere coupling (Almendra-Martín et al. 2022), or prediction based on causal precursors (Kretschmer et al. 2017).

Among the major challenges in causal analyses, small sample sizes rank high (e.g., Camps-Valls et al. 2023). Further, the above mentioned studies mostly ignored the potentially confounding influence of internal variability on causal structures or potential changes related to global warming. However, they acknowledged the strong dependence on the data set used for causal discovery. As stated by Lehner & Deser (2023), though, internal variability affects nearly all quantities in the climate system. Assuming that relationships among these variables vary as well, e.g., depending on the period under consideration or in different realizations of climate trajectories (e.g., the "blinking edges" among nodes referred to by Ebert-Uphoff & Deng (2010)) is thus reasonable. Yet, investigations of causal discovery structures within large ensembles considering internal variability are clearly missing. The latter, however, is required if potential changes of those structures due to global warming are addressed.

4   Research Questions

Physical geography is particularly interested in interconnected human and natural systems in space and time with a focus on the physical compartments. Climate change is one of the major challenges spreading into these systems. Where climatology and physical geography meet, though, knowledge on large-scale processes or phenomena cannot be obtained easily from real-world experiments. In consequence, researchers recur to observations or simulation experiments in a virtual world. As more profuse data from earth observation and modeling become available, joining forces among environmental and data science opens new pathways to exploit this "climate big data".

Most research on heatwaves and droughts focuses on mean states or variability and their change in space and time. Sometimes more advanced strategies are pursued as shown in the previous sections, such as the investigation of multivariate hazards. There, dependencies are commonly used to describe distinct events. However, they also carry information in themselves. This is rarely examined, mostly because it requires large data sets in order to obtain persistent structures and robust (change) information.

As outlined in the previous sections, current research shows that

- recurrent features of heatwaves and droughts (e.g., location, tracks, multivariate relationships) require investigation of their changes (Section 2 and 3.1),

- for meaningful regional analyses, relevant climate features should be represented reasonably in climate (model) data (Section 3.2 and 3.3),

- properties of variables, including structures, can be assumed to be masked or confounded by internal variability (Section 3.1 and 3.3),

- data-driven approaches provide promising means to unearth recurrent and significant properties from spatio-temporal data that can be indicative of relevant physical relationships (Section 3.4).

Altogether, while profound analyses on mean and extreme values of variables are performed, relationships among variables, i.e., structures, are less investigated. Potential trends in structures due to global warming (and hence new states to adapt to) are difficult to discern without knowledge on their variability. Here, SMILEs can contribute.

In this context, this thesis poses the main research question of *how data-based structures may contribute meaningfully and in an actionable way to regional climate change assessment of heatwaves and droughts.*

A suite of supportive questions addresses at first the description and explanation of spatio-temporal structures, including their internal variability and potential trends. Implications for assessing regional climate change are enclosed in the answers. Therefore, the advantages of the CRCM5-LE, a regional SMILE produced in the ClimEx project (Section 3.3.1), are exploited.

Q1 | Does an RCM incorporate large-scale atmospheric regimes as present in the driving data and reproduce realistic (remote) responses?

Without this information, the quality of spatially distributed climate indicators in the RCM remains questionable. Q1 addresses the suitability of the employed data.

- *Q1.a: General performance – Can the driving CanESM2-LE correctly reproduce the NAO and its climatic implications for central Europe?*

- *Q1.b: Nesting approach – To what extent do large-scale SLP patterns in the RCM diverge from the driving GCM data?*

Q2 | Can data-driven approaches robustly describe spatio-temporal structures of heatwaves and droughts hidden in a SMILE?

Given affirmation, data-driven approaches may be considered helpful in capturing and analyzing complex hazard properties.

- *Q2.a: New structure discovery – Can a causal discovery algorithm adequately capture spatio-temporal hazard properties, in particular heatwave tracks?*

- *Q2.b: Known structure description – How can already known structures be described robustly for further assessment?*

Q3 | Do internal climate variability and forced trends affect structures among variables?

In view of the dominant role of internal variability in any aspect of climate hazards and its origin in processes governing, e.g., heatwaves and droughts as well, the derived structures are also supposed to be subject to spatio-temporal variations.

- *Q3.a: GCM–RCM variability – How does the range of driver–response structures (e.g., NAO and climate responses) diverge among a global and regional SMILE?*

- *Q3.b: Spatio-temporal variability – Are spatio-temporal structures of heatwave and drought characteristics subject to internal variability?*

- *Q3.c: Structures and climate change – Can trends of spatio-temporal structures of, e.g., heat and drought characteristics be obtained robustly?*

Q4 | How can the obtained structures be verified physically?
Data-driven approaches cannot derive structures which exist in the real world, but not in the data. Yet, they could produce spurious relationships. Hence, obtained statistical associations require thorough evaluation and, if possible, association to corresponding mechanisms.

- *Q4.a: Linking to physical processes – Can independent drivers account for the obtained spatio-temporal structures of, e.g., heatwave tracks?*

- *Q4.b: Structure components – What do we learn about structures, e.g., bivariate CDHE, from their single components?*

Q5 | How can complex results of SMILE analyses regarding internal variability and extreme events be conveyed effectively?
Communicating complex results to target audiences requires clear definition and illustration of the key aspects. Especially seemingly abstract heatwave or CDHE structures benefit from suitable presentation. To tailor results for outreach, a dimension reduction (e.g., across space, time, or members) is advantageous, yet drops relevant information.

- *Q5.a: Regional hot spots – How do specific conclusions on high-dimensional data profit from assigning regional "hot spots"?*

- *Q5.b: Result robustness – Can robust structures or spatial patterns of heat and drought be presented while conveying relevant information on internal variability?*

- *Q5.c: Result Presentation – How can we illustrate and communicate complex multidimensional results from a SMILE?*

5   Scientific Publications

This cumulative PhD thesis comprises three first-authored peer-reviewed publications and one submitted first-authored manuscript at the intersection of physical geography, climatology, and data science (Figure 6 (a)).
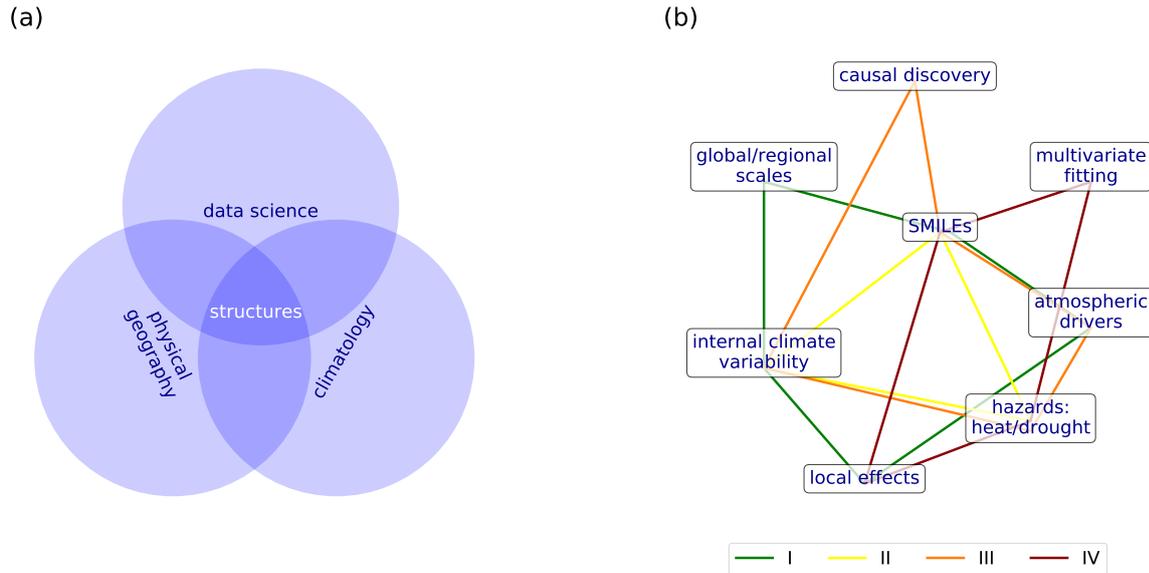

Figure 6: (a) Positioning of structure derivation and examination in the intersection of physical geography, climatology, and data science. (b) Keyword-based positioning of papers I–IV in the intersection region of the three spheres shown in (a). Keywords belonging to the same paper are connected by colored lines.

Paper I verifies the suitability of the regional SMILE regarding the representation of large-scale atmospheric patterns in the driving GCM and driven RCM. Papers I–IV employ data-driven approaches on a regional SMILE in order to extract and describe the effect size of recurrent spatio-temporal structures based on assumed relationships, seek to find ways of connecting said structures with drivers and effects, and discuss their advantages and limitations (Figure 6 (b)). Aside from paper I, the focus is set on a climatological perspective of heat and drought. Paper II describes future changes of European droughts against the backdrop of internal climate variability, paper III distills and examines recurrent heatwave tracks, and paper IV merges heat and dryness by investigating bivariate CDHE under climate change conditions. Paper II additionally explores ways to effectively communicate multi-dimensional results.

5.1 Paper I: Using a Nested Single-Model Large Ensemble to Assess the Internal Variability of the North Atlantic Oscillation and its Climatic Implications for Central Europe

Reference: Böhnisch, A., Ludwig, R., and Leduc, M. (2020): Using a Nested Single-Model Large Ensemble to Assess the Internal Variability of the North Atlantic Oscillation and its Climatic Implications for Central Europe, *Earth Syst. Dynam.*, 11, 617–640, DOI: https://doi.org/10.5194/esd-11-617-2020.

Transition to paper I: In literature, simulations from both the global CanESM2-LE and regional CRCM5-LE were so far evaluated regarding their respective variable means and variability against observations/reanalysis. Yet, to evaluate the reliability of the employed regional SMILE, assessing variability propagation from the driving GCM into the driven RCM is mandatory as well; in particular, when considering drivers from outside the regional domain. In this study, the propagation is investigated using an atmospheric mode of variability with high relevance to European climate, namely the NAO, by employing the advantages of SMILEs. Since the NAO governs the zonal flow onto Europe, it is particularly appropriate for comparing large-scale atmospheric conditions among GCM and RCM. As the cause–effect direction is comprehensively documented in literature (e.g., Benedict et al. 2004), the study focuses on the effect size: Pearson correlation and least-squares regression are used to quantify NAO responses of winter temperature and precipitation. During winter, the NAO impact reaches its maximum (e.g., Hurrell & Deser 2009). The NAO–impact structure in the CRCM5-LE is evaluated against reanalysis and the driving CanESM2-LE. Furthermore, the large-scale representation of winter SLP across Europe in the CRCM5-LE is investigated. Showing that the RCM performs well in representing both the large-scale SLP and the climatic conditions during variable atmospheric regimes would raise trust in the simulation of driving conditions and associated hazards like heatwaves and droughts. The purpose of this study is thus an evaluation of the regional SMILE's suitability across Europe.

Author's contribution: This study was conceptualized by AB under supervision of RL. Formal analysis, visualization of results, and writing of the original draft was performed by AB. All authors: interpretation of the findings and paper revision.

Status: published in Earth System Dynamics (Journal Impact Factor: 7.3, 12/2023)

Earth System
Dynamics

Using a nested single-model large ensemble to assess the internal variability of the North Atlantic Oscillation and its climatic implications for central Europe

Andrea Böhnisch[1]**, Ralf Ludwig**[1]**, and Martin Leduc**[2,3]

[1]Department of Geography, Ludwig-Maximilians-Universität München, Munich, Germany
[2]Ouranos, Montréal, Québec, Canada
[3]Centre ESCER, Université du Québec à Montréal, Montréal, Québec, Canada

Correspondence: Andrea Böhnisch (a.boehnisch@lmu.de)

Abstract. Central European weather and climate are closely related to atmospheric mass advection triggered by the North Atlantic Oscillation (NAO), which is a relevant index for quantifying internal climate variability on multi-annual timescales. It remains unclear, however, how large-scale circulation variability affects local climate characteristics when downscaled using a regional climate model. In this study, 50 members of a single-model initial-condition large ensemble (LE) of a nested regional climate model are analyzed for a NAO–climate relationship. The overall goal of the study is to assess whether the range of NAO internal variability is represented consistently between the driving global climate model (GCM; the Canadian Earth System Model version 2 – CanESM2) and the nested regional climate model (RCM; the Canadian Regional Climate Model version 5 – CRCM5). Responses of mean surface air temperature and total precipitation to changes in the NAO index value are examined in a central European domain in both CanESM2-LE and CRCM5-LE via Pearson correlation coefficients and the change per unit index change for historical (1981–2010) and future (2070–2099) winters. Results show that statistically robust NAO patterns are found in the CanESM2-LE under current forcing conditions. NAO flow pattern reproductions in the CanESM2-LE trigger responses in the high-resolution CRCM5-LE that are comparable to reanalysis data. NAO–response relationships weaken in the future period, but their inter-member spread shows no significant change. The results stress the value of single-model ensembles for the evaluation of internal variability by pointing out the large differences of NAO–response relationships among individual members. They also strengthen the validity of the nested ensemble for further impact modeling using RCM data only, since important large-scale teleconnections present in the driving data propagate properly to the fine-scale dynamics in the RCM.

1   Introduction

One of the major sources of uncertainty regarding short-term future climate projections is internal climate variability, while model climate response and greenhouse gas concentration scenarios become more important sources of uncertainty on a longer-term time horizon (Hawkins and Sutton, 2009, 2011). The term "internal variability" denotes climate variability which is not forced by external processes (either anthropogenic or natural) but arises from the chaotic properties of the climate system itself (Leduc et al., 2019; Deser et al., 2012), i.e., from varying sequences of weather events under identical external forcings. These sequences of weather events may be altered by global atmospheric modes of variability through the linking between large-scale circulation and local weather characteristics (like surface air temperature and precipitation). Such large-scale atmospheric modes can thereby establish periods of discernible states on multi-annual timescales.

Among these modes, the North Atlantic Oscillation (NAO) is particularly important for Northern Hemisphere climate. Its two states, positive and negative, are evoked by planetary wave breaking in the polar front, leading to antagonistic pressure behavior of two centers over the North Atlantic: one located within the subtropical high-pressure belt ("Azores High", AH) and the second in subpolar regions ("Icelandic Low", IL) (Benedict et al., 2004). The resulting pressure gradient, which is stronger during positive and weaker during negative phases, affects large-scale extratropical circulation, especially the strength and position of midlatitude westerly winds connected to the jet stream, and air mass advection during boreal winter (Deser et al., 2017; Hurrell and Deser, 2010). Compared with neutral conditions, the positive NAO state leads to warmer and moister winters in northern Europe but cooler and drier conditions in the south, and vice versa in the negative state (e.g., Hurrell and Deser, 2010; Pokorná and Huth, 2015; Woollings et al., 2015).

The NAO is commonly quantified with an index that makes use of the air pressure or geopotential height gradient between AH and IL. The index may be calculated as a normalized difference of station measurements, spatially averaged values of preset regions, or the region of highest variance is obtained by principal component analysis (PCA) (Pokorná and Huth, 2015; Hurrell and Deser, 2010; Stephenson et al., 2006; Hurrell, 1995; Rogers, 1984). Each method has its advantages and limitations. For example, station-based or fixed-in-space indices do not reproduce shifting NAO patterns and may be affected by microclimatic noise and other teleconnection patterns (Hurrell and Deser, 2010; Osborn, 2004). Indices based on PCA on the other hand are dependent on the chosen data domain for calculation and on the data set itself (Osborn, 2004). The different approaches, however, lead to highly similar index time series (see, e.g., Pokorná and Huth, 2015, for a detailed survey of various approaches).

While the typical NAO pattern and its impacts are usually correctly reproduced in global climate models (GCMs) (Stephenson et al., 2006; Ulbrich and Christoph, 1999; Reintges et al., 2017), its fidelity in a future climate remains uncertain: the NAO is found as intensifying but also counteracting global warming in the Northern Hemisphere ("global warming hiatus"; Iles and Hegerl, 2017; Deser et al., 2017; Delworth et al., 2016). Similarly, the findings regarding the prevalence of future positive or negative states lack unity. Some analyses of CMIP5 models, for example, suggest more positive phases under rising greenhouse gas concentrations until 2100 (e.g., Kirtman et al., 2013; Christensen et al., 2013), others favor an increase of negative phases (Cattiaux et al., 2013).

In most of these studies, it was common to rely on one simulation per model and estimate the model's performance regarding the NAO by this single run. This approach allows for comparing different models (and observations). However, it is not possible to robustly evaluate the range of NAO index values and evolution in a projected future climate, or whether the chosen simulation is a good representation of how this model simulates the phenomenon in question (Leduc et al., 2019). Relying on single realizations possibly deteriorates the assessment of a given model, as single realizations may vary considerably among themselves due to internal variability (and also deviate from the climate evolution observed in reality). One way to sample realizations is to perturb the initial conditions of the model, leading to multiple simulations with identical external forcing which only differ due to internal variability. Examples for recent GCM initial-condition large ensembles of transient simulations are the 100-member Max Planck Institute Grand Ensemble (MPI-GE; Maher et al., 2019), the 50-member Canadian Earth System Model Large Ensemble (CanESM2-LE, e.g., Kirchmeier-Young et al., 2017; Fyfe et al., 2017) or the 40-member Community Earth System Modelling Large Ensemble (CESM-LE Kay et al., 2015) which were, among others, used for various analyses of internal variability or extreme events. Such initial-condition ensembles also allow a more robust distribution of atmospheric modes to be sampled, as was done, e.g., for El Niño–Southern Oscillation (ENSO) in Maher et al. (2018). That is why the present study is investigating the NAO pattern in a single-model large ensemble of a GCM.

However, when interested in NAO impacts on a regional scale, like central Europe, the GCM is not sufficient for fine-scale responses. Due to their coarse spatial resolution, GCMs poorly resolve land–water contrasts and topographic characteristics which may be highly relevant in climate impact studies over heterogeneous landscapes (Leduc et al., 2019). Thus, dynamical downscaling of the GCM members using a regional climate model (RCM) is advised (Leduc et al., 2019). The downscaling of a GCM single-model large ensemble, CanESM2-LE, was performed within the Climate Change and Hydrological Extremes project (ClimEx; http://www.climex-project.org/, last access: 3 July 2020; Leduc et al., 2019).

Examples of analyses on the separation of the forced signal from internal variability within a 16-member single-model initial-condition GCM–RCM ensemble of EC-Earth and RACMO2 were performed by Aalbers et al. (2018) for various extreme precipitation indices.

Combining the driving GCM and nested RCM (i.e., driven by lateral boundary conditions of the GCM) large ensembles (LEs) allows for analyzing the spread of NAO states and responses within one model chain. In doing so, it is possible to establish the range of internal variability of the NAO and find robust NAO and response patterns by significantly reducing uncertainty associated with internal variability in the ensemble.

The present study targets the research question of how global circulation variability, in this case the NAO teleconnection, affects local climate characteristics when downscaled using an RCM. It specifically aims at evaluating

whether the range of internal variability is represented consistently between the driving GCM-LE and the driven high-resolution RCM-LE. The latter may be important for impact modelers who work with RCM data on internal variability without taking the driving GCM into account.

To answer these research questions, this study focuses on four topics and related key questions:

a. Regarding the general performance of the model chain, can the driving GCM resolve the NAO correctly and are climatic implications for central Europe reproduced?

b. In terms of the nesting approach, does the RCM correctly incorporate the NAO pattern present in the driving data and produce realistic response patterns?

c. For internal variability, what is the range of possible NAO patterns and responses, expressed by the inter-member spread among the 50 members of the GCM-LE and the RCM-LE?

d. With climate change, how do topics (a)–(c) change in transient climate simulations that extend until 2099 using an RCP8.5 emissions scenario?

2   Data and methodology

2.1   Data

Data from three different sources are employed in this study (Table 1). The major source is the RCM-LE data set of the ClimEx project which is described in detail in Leduc et al. (2019). The ClimEx project is conducted in a Québec–Bavarian cooperation and targets issues of hydrological extreme events in the time horizon of 1950–2099, using a nested high-resolution 50-member single-model initial-condition large ensemble with an RCP8.5 emissions scenario from 2006 onwards (Leduc et al., 2019). Five members of CanESM2-LE (2.8° spatial resolution; Fyfe et al., 2017) with different ocean initial conditions were slightly perturbed in 1950, leading to 10 members per ocean family. The members are assumed to become independent about 5 years after their initialization in 1950 (spin-up period) (Leduc et al., 2019).

Regarding the atmospheric circulation, Fig. 1 shows that, due to the chaotic nature of the atmospheric system, the daily NAO index seems to lose dependence on the initial conditions within the course of 1 month after initialization (see Leduc et al., 2019, for a similar presentation of member independence).

As described in Leduc et al. (2019), the 50 CanESM2 members were dynamically downscaled using the Canadian Regional Climate Model version 5 (CRCM5 Large Ensemble, 0.11° spatial resolution) over two domains covering Europe and northeastern North America, each sized 280 × 280 grid cells on a rotated grid. Large-scale spectral nudging
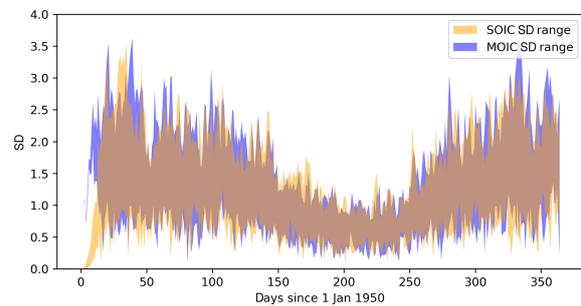
Figure 1. Inter-member standard deviation of a daily NAO index in CanESM2-LE starting on 1 January 1950 as a function of time. The inter-member standard deviation (SD) is derived from 10 groups of five members with the same ocean initial conditions (SOICs) and 10 groups of five members with mixed ocean initial conditions (MOICs, following an approach in Leduc et al., 2019).

of the horizontal wind field was applied during the nesting process (Leduc et al., 2019). This single-RCM 50-member ensemble allows for internal variability and extreme events to be detected in high spatial and temporal resolution within a total of 7500 modeled years (Leduc et al., 2019).

Comparing the internal variability of the CRCM5 members with the inter-member spread of a subset of the multi-model EURO-CORDEX (Coordinated Regional Climate Downscaling Experiment) ensemble regarding regionally integrated European winter temperature and precipitation, von Trentini et al. (2019) showed that both ensemble spreads are of comparable magnitude. The CORDEX ensemble consists of several GCM–RCM combinations set up in a coordinated modeling framework and aims at evaluating uncertainty due to model configuration (Giorgi et al., 2009). The comparison of the single-model and multi-model spreads suggests that a large fraction of the CORDEX ensemble spread regarding temperature and precipitation can be explained by internal variability, despite the fact that it was not explicitly sampled within the CORDEX framework (where most models provided a single simulation; von Trentini et al., 2019). At smaller regional scales, however, single-model and multi-model spreads may show considerable and in parts temporally changing differences which may partly be induced by model response uncertainties (von Trentini et al., 2019).

In the present study, model data are compared with the ERA-Interim (ERA-I) reanalysis data set of the European Centre for Medium-Range Weather Forecasts (Dee et al., 2011, ECMWF). Additionally, a CRCM5 run driven by ERA-I is used to evaluate CRCM5 under "perfect" (as far as ERA-I can be assumed to represent reality) lateral boundary conditions, i.e., without the potential CanESM2 data input error.

The relevant variables for this study are

Table 1. Overview of used data sets, their spatial resolution, the number of members and the employed variables.

| Data name	Model type	Spatial resolution	Members	Model output variable names	Institution
ERA-I	Reanalysis	$0.75° \times 0.75°$	1	msl (Pa), t2m (K), tp (m)	ECMWF
CRCM5/ERA-I	RCM	$0.11° \times 0.11°$	1	psl (Pa), tas (K), pr (kg m^{-2} s^{-1})	Ouranos
CanESM2-LE	GCM	$2.8° \times 2.8°$	50	psl (Pa), tas (K), pr (kg m^{-2} s^{-1})	CCCma*
CRCM5-LE	RCM	$0.11° \times 0.11°$	50	psl (Pa), tas (°C), pr (mm)	Ouranos

* CCCma – Canadian Centre for Climate Modelling and Analysis.

– (mean) sea level air pressure (referred to as "SLP", converted to hPa);

– near-surface air temperature (referred to as "nSAT", converted to K); and

– total precipitation including liquid and solid precipitation from all types of clouds (referred to as "PR", converted to mm).

ERA-I variables t2m, tp and msl (Table 1) are chosen as they are assumed to most accurately represent the GCM and RCM variables. As the variables derived from the three data sources are available at different temporal resolutions (3-hourly for tas and psl in RCM, hourly for pr in RCM, daily for psl, pr and tas in GCM, 6-hourly for ERA-I t2m and msl analysis, and 12-hourly for ERA-I tp forecast data), they are all aggregated to daily values first.

In the Appendix, Fig. A1 shows that CRCM5 tends to underestimate (overestimate) mean winter nSAT mean in the northern (southern) part of the domain, regardless of the driving data (Fig. A1a for ERA-I and Fig. A1c for CanESM2), whereas winter PR sums are overestimated in nearly the entire domain, with strongest values in the south-eastern part (Fig. A1d and f). Displaying the opposite bias of CRCM5, CanESM2 overestimates (underestimates) mean winter nSAT in the northern (southern) part of the domain (Fig. A1b), whereas winter PR sum is underestimated in the eastern half of the domain and overestimated on the western side of the Alps (Fig. A1e). As this study will focus on responses of nSAT and PR induced by the NAO (see Sect. 2.2.4), aside from regions with particularly high PR sum values, it is found that such NAO responses are generally insensitive to these biases.

Commonly, NAO impact studies focus on seasonally aggregated values of the analyzed variables or extreme events (e.g., Stephenson et al., 2006). Yet the NAO, which accounts for variations in the mean zonal atmospheric flow towards Europe, can be assumed to influence not only winter mean values but also their interannual variability. So, in addition to analyses of winter mean temperature (nSAT mean) and precipitation sums (PR sum), selected analyses are also performed on winter mean monthly standard deviations of daily mean temperature (nSAT SD) as a measure of temperature variability.
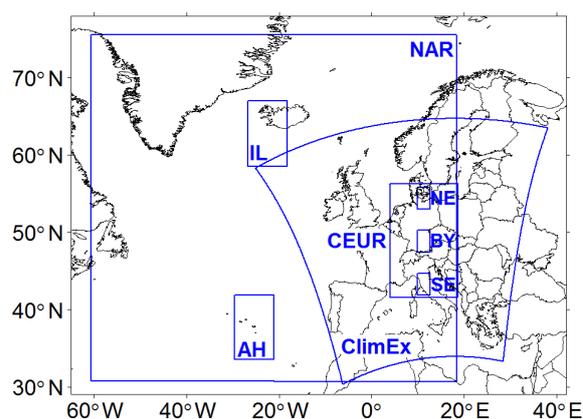
Figure 2. Regions of interest. Abbreviations and domain sizes in terms of GCM grid cells (2.8°) are as follows: AH – Azores High (3×3); IL – Icelandic Low (3×3); NAR – large-scale North Atlantic region (28 × 16); CEUR – central Europe (5 × 5); NE – northern Europe (1); BY – Bavaria (1); SE – southern Europe (1); ClimEx – domain used in ClimEx project (extent approximately 22 × 12 after resampling to GCM grid).

2.2   Methodology

2.2.1   Regions of interest and time horizon selection

Analyses are performed on time series of spatially averaged information (nSAT mean, PR sum for response variables and SLP for index calculation) as well as on spatially explicit data (nSAT mean, nSAT SD, PR sum). All data are provided as NetCDF files, and most pre-processing is performed using the Climate Data Operators (CDOs) of the Max Planck Institute for Meteorology (Schulzweida, 2017).

The regions of interest and their names used in this study are displayed in Fig. 2. The formation of the NAO over the North Atlantic (NAR, AH, IL regions) is analyzed in the ERA-I and CanESM2-LE data, while responses over central Europe (CEUR, NE, BY, SE regions) are evaluated in ERA-I, CRCM5/ERA-I, CRCM5-LE and CanESM2-LE data.

AH and IL regions are centered over Ponta Delgada (Azores) and Reykjavik (Iceland), two commonly used stations for NAO index calculations. To avoid microclimatic impacts and sampling uncertainties of a single grid cell and

to account for moving SLP centers (see, e.g., Moore et al., 2013), both NAO core regions are extended to 3×3 GCM grid cell matrices. In preliminary analyses conducted for the present study, the NAO index has proven to be very robust towards the exact shape of the core regions.

The central European domain (CEUR) is defined in CanESM2-LE by selecting a 5×5 GCM grid cell matrix centered over Munich (Germany). This CEUR domain extends from Denmark in the north to central Italy in the south and from Poland to France in the east–west direction. The corresponding CEUR region within the ClimEx European domain is used to quantify the impacts of the NAO in the CRCM5-LE data. It lies downstream of the westerly flows initiated by the NAO, so the following analyses set a special focus on the incorporation of large-scale inflow from the western side into the nested RCM.

As the responses to the NAO are expected to vary over the CEUR domain, it seems favorable to analyze spatial structures explicitly in addition to analyses of time series over several subset regions. These subset regions (see, e.g., Déqué et al., 2007) denote small-scale sample areas inside the CEUR domain, sized one GCM grid cell each, with expected typical "northern European" (NE) and "southern European" (SE) NAO responses for a more detailed statistical analysis. A third GCM grid cell is chosen to represent the transition zone between NE and SE. Coincidentally, it closely represents the region of Bavaria, which is why the name "BY" is assigned to it. ERA-I and RCM data (3×4 and 26×26 grid cells, respectively) are spatially aggregated to GCM resolution for this part of the analysis.

This study focuses on interannual analyses which are conducted for two time horizons covering 30 years each. The historical (hist; 1981–2010) period is used to establish reference statistics in the ERA-I data and the ERA-I-driven CRCM5 run which are then evaluated in the GCM-LE and the RCM-LE. Links and relationships established for the historical period are also investigated in a far future horizon (fut; 2070–2099).

The chosen period length is assumed to include major fluctuations, like internal climate variations or several solar cycles, which might affect NAO phases (Andrews et al., 2015). Thus, their influence can be assumed to be represented by the sampled NAO time series. Relationships between the NAO and response variables most probably vary on different timescales (Hurrell and Deser, 2010; Woollings et al., 2015; Xu et al., 2016; Hurrell and Van Loon, 1997). However, as 30-year periods are not long enough for analyses of multidecadal (> 30-year) NAO–response variability (Woollings et al., 2015), stationarity in NAO–impact relationships is assumed for simplicity reasons.

Since the NAO is known to be strongest in winter (Hurrell and Deser, 2010) and the connection between station-based indices and NAO responses tends to be best in winter (see Pokorná and Huth, 2015, for DJF months), analyses are performed for this season only. Preliminary tests within

this study have shown that correlations and links between the NAO index and the climate variables are more distinct from noise, if March is included as well. That is why an extended winter season is used here (DJFM; see also Iles and Hegerl, 2017; Hurrell, 1995; Osborn, 2004).

All data (spatially explicit and subset time series) are aggregated to the seasonal timescale for further use (winter means for nSAT and winter sums for PR).

2.2.2   Deriving a NAO index

The NAO index is derived from ERA-I and CanESM2-LE data, resulting in 1 ERA-I and 50 GCM realizations. As the CRCM5 ClimEx domain does not cover the AH and IL regions (see Fig. 2), the index is not derived from this data source. The NAO is quantified in this study with an index which is closest to a station based or zonally averaged index. It therefore directly represents the winter SLP gradient over the North Atlantic.

The time series of AH and IL originate from the temporally shortened and spatially averaged SLP time series of both grid cell matrices. Daily SLP values are averaged to monthly means (Cropper et al., 2015) and scaled to obtain mean $\mu = 0$ and standard deviation of 1σ, as outlined in Osborn (2004) and Hurrell and Van Loon (1997), by subtracting the 1981–2010 seasonal mean (overbar) and dividing by the 1981–2010 seasonal standard deviation (s_{IL}, s_{AH}):

$$\text{NAO index} = \frac{\text{AH} - \overline{\text{AH}}}{s_{\text{AH}}} - \frac{\text{IL} - \overline{\text{IL}}}{s_{\text{IL}}}. \tag{1}$$

Monthly indices are next averaged to DJFM means. This approach is similar to Woollings et al. (2015) and Jones et al. (2013).

The ERA-I NAO index calculated this way shows high agreement with often-cited NAO indices like the time series of Hurrell (Pearson correlation of $r = 0.95$ with ERA-I NAO index; index available at https://climatedataguide.ucar.edu/climate-data/hurrell-north-atlantic-oscillation-nao-index-station-based, last access: 3 July 2020). For further analyses, it will therefore serve as a reference.

To compare future with historical index values, the future time series of AH and IL are standardized with the historical SLP standard deviations (see also Ulbrich and Christoph, 1999; Hansen et al., 2017) and mean values. The standardization of each GCM member is carried out individually.

2.2.3   Evaluation of the large-scale SLP pattern in RCM data

To estimate whether the NAO may be seen as being correctly represented in the nested RCM data, the reproduction of interannual SLP pattern variations in the CRCM5 data is verified. Therefore, monthly mean SLP data of CRCM5 (both driving data sets) and ERA-I are linearly interpolated

to GCM resolution over the ClimEx domain. During interpolation, small scales are automatically filtered such that the remaining large scales of driving data and RCM data may be compared. As a next step, a root mean square difference (RMSD) of the difference time series between monthly mean driving and RCM data over the hist and fut time periods is obtained across all members and winter months:

$$\text{RMSD}(i, j) = \langle\langle\frac{\sqrt{\langle D_\text{m}(i, j, t, n)^2\rangle_t}}{\sqrt{\text{VarDrive}_\text{m}(i, j, n)}}\rangle_n\rangle_m \tag{2}$$

$$\text{VarDrive}_\text{m}(i, j, n) = \langle(\text{Drive}_\text{m}(i, j, t, n)$$
$$- \langle\text{Drive}_\text{m}(i, j, t, n)\rangle_t)^2\rangle_t, \tag{3}$$

where $\langle\cdot\rangle$ is the averaging operator over a given index, D_m is the difference between monthly mean driving data and RCM data; Drive_m is driving SLP data; VarDrive_m is the variance of SLP driving data over the 30-year periods; i, j are spatial grid coordinates, m indicates months 12, 1–3, n indicates ensemble members 1–50 for CanESM2 and member 1 for ERA-I, and t indicates years in 1981–2010 and 2070–2099. The normalization by the square root of the temporal variance of the driving data provides a measure relative to the interannual variability of the SLP pattern in a given location.

2.2.4   Climatic changes associated with NAO

All data sources (Table 1) are used to obtain response patterns of the variables nSAT and PR. Climatic changes associated with the NAO are evaluated using Pearson correlation coefficients and a slope parameter obtained by linear regression.

ERA-I and CRCM5/ERA-I nSAT and PR data are correlated with the ERA-I index, CanESM2 and CRCM5 members are correlated with the CanESM2 index calculated for the corresponding member.

The correlation analysis assumes (symmetric) linear relationships between the NAO index and nSAT or PR. The associated response of the variables to NAO changes may then be quantified by a linear equation (Iles and Hegerl, 2017; Stephenson et al., 2006; Hurrell, 1995):

$$Y = \alpha_1 X + \alpha_0 + \varepsilon_Y, \tag{4}$$

with Y being the (response) variable at a given grid cell that is partly explained by the NAO (X, the predictor) and by any other influences (ε_Y; Stephenson et al., 2006; von Storch and Zwiers, 2003). The coefficient α_1 is estimated on each grid cell using ordinary least squares regression with the R function lm (https://www.rdocumentation.org/, last access: 3 July 2020). It represents mean change in nSAT or PR that accompanies unit index change during the time period under consideration (Iles and Hegerl, 2017). The line offset α_0 in Eq. (4) is equal to the long-term mean. The α_1 coefficients may be computed with respect to normalized index series (von Storch and Zwiers, 2003), but in this study the non-normalized index time series is preferred in order to take into account the member-specific index units. The NAO–response relationship is analyzed individually for each GCM and RCM member (as is done, e.g., in Woollings et al., 2015).

2.2.5   Addressing internal variability

In this study, the GCM–RCM combination allows to set a focus on the internal variability of an RCM ensemble and the driving GCM ensemble. Climate modes tend to show high internal variability (see, e.g., Maher et al., 2018, for an analysis of ENSO internal variability in CMIP5 models and two single-model large ensembles). The present study targets the NAO-related internal variability within a single GCM–RCM combination.

In general, natural internal variability may be understood from different angles. When looking into single realizations of time series of a given variable, internal variability may be seen as represented by the oscillation around the long-term mean evolution, i.e., the residuals (Frankcombe et al., 2015; Hawkins and Sutton, 2009, 2011). In this case, the amplitude of internal variability is usually calculated as a time-invariant quantity for the period under consideration (Hawkins and Sutton, 2009, 2011).

Another way is investigating transient internal variability in initial-condition ensembles, e.g., in Maher et al. (2019). In this case, the ensemble establishes ranges of possible weather event sequences by superposing single realizations which are equally likely by construction of the ensemble.

In the present study, the latter approach is used within the 50-member CanESM2-LE and CRCM5-LE. This allows to sample internal variability at single points in time as the range of the members' values, i.e., across members (e.g., Maher et al., 2018). While internal variability is assumed to be stationary within both 30-year periods for this study, the use of a LE allows to detect potential changes in internal variability between both analysis periods.

Internal variability is expressed as the across-member standard deviation, i.e., the inter-member spread of CanESM2-LE and CRCM5-LE (see also Leduc et al., 2019; Déqué et al., 2007; Aalbers et al., 2018) among the 30-year means, rather than computing a transient internal variability at each time step as was done, e.g., in Maher et al. (2019). Aggregations to ensemble means (like in Deser et al., 2017; Aalbers et al., 2018) of NAO responses are only performed for illustrating purposes in order to avoid masking model internal variability (Zwiers and von Storch, 2004).

3   Results

The result section is structured in two large parts: Sect. 3.1 deals with the representation of the NAO and climatic responses in the GCM and RCM, and Sect. 3.2 targets internal variability in the GCM and RCM.

3.1   NAO within the ClimEx data set

Naturally, the first step when evaluating the NAO in a model ensemble is to analyze its representation and index distribution in the model data of interest.

3.1.1   NAO index and SLP conditions

CanESM2-LE produces NAO index values which follow a distribution comparable to the ERA-I data (similar to a normal distribution with $\mu = 0$, $\sigma = 1$; Fig. 3a), but the CanESM2-LE distribution appears smoother due to a larger sample size ($n = 1500$ for CanESM2-LE and $n = 30$ for ERA-I). Maximum and minimum index values (x axis in Fig. 3a) of some of the 50 members exceed those of the ERA-I realization; thus, ERA-I which serves as a reference realization lies well within the ensemble inter-member spread. The future NAO index shows a similar distribution of values but with slightly less positive and more negative values (red curve in Fig. 3a).

For the following analyses, independence of the 50 members is critical to interpreting the inter-member spread as a proxy for internal variability. In evaluating this, it is important to recall that the 50-member CanESM2-LE was constructed in two steps (Fyfe et al., 2017; Leduc et al., 2019). First, independent atmosphere–ocean states in 1850 were used to launch five historical simulations integrated forward until 1950. Second, in 1950, each of these five ensemble members was used to launch 10 individual simulations by applying a small perturbation to the atmosphere and integrated forward until 2099, thereby producing the 50-member large ensemble.

As a consequence, for this study, members between each of the five groups of 10 are expected to be independent. However, members within each group of 10 are highly correlated in 1950 and progressively increase their independence beyond their 1950 starting point. To evaluate whether the 10 members within each of the five groups have become sufficiently independent by the two 30-year periods of interest (1981–2010 and 2070–2099), correlations among member time series are applied to two groups following Leduc et al. (2019): (i) correlations among the 10 members from the same group (same ocean initial conditions, SOICs; $n = 225$ cases, dotted lines in Fig. 3b) and (ii) correlations between each member and the 40 members from the four other groups (MOICs, $n = 1000$ cases, solid lines in Fig. 3b).

These correlations approximately follow a normal distribution with $\mu = 0$ and $\sigma = 0.2$. There is a slight surmount of low positive correlations in the SOIC group compared with the MOIC group which is (not significantly) stronger in the fut time horizon (see red and black dotted lines in Fig. 3b). Although zero correlations do not necessarily imply independence, clear correlations among members would contradict the assumption of independence. In general, the members are thus not seen as being dependent.

As will be discussed below, the SLP pattern over the North Atlantic changes slightly in the future period. So the direct comparison between historical and future SOIC and MOIC correlations remains difficult. The members also show no systematic correlation with the ERA-I NAO index despite similar statistics (see also Fig. 9). Thus, the ERA-I and GCM indices can be seen as not dependent realizations drawn from the same distribution.

In order to further evaluate the NAO representation in CanESM2-LE, Fig. 4 presents the large-scale SLP patterns in the NAR region during neutral, positive and negative NAO conditions. Positive (negative) index years are chosen, if the respective index value exceeds 1 (-1) as in Rogers (1984). The neutral conditions refer to the 30-year SLP average. Regions with strong sampling uncertainties, i.e., where the standard error is larger than the anomaly, are indicated with stippling in Fig. 4b, c, e and f.

Under neutral NAO conditions, the North Atlantic region is characterized by a pressure dipole. This structure is intensified and tilted clockwise in the CanESM2-LE ensemble mean (Fig. 4d) compared with ERA-I (Fig. 4a). The mean SLP difference between the CanESM2-LE mean and ERA-I reaches up to 10 hPa in both directions. SLP values are higher over Greenland and lower over the North Sea in CanESM2-LE compared with ERA-I (Fig. 4a and d). Long-term neutral states of both driving data sources show robust signals in the entire NAR region (i.e., no stippling). This suggests that the different patterns in GCM and reference data are not singularly artifacts arising from different sample sizes but rather robust features.

The GCM multi-member composites of positive and negative phases show less pronounced SLP anomalies than the reference data (Fig. 4b, c, e and f). Transition regions between the AH and IL nodes are marked by high uncertainty in ERA-I, whereas the SLP anomalies at the NAO centers of action show less uncertainty. The GCM patterns are more robustly assessed (i.e., less prone to sampling uncertainty) as can be seen by the very small area with stippling in which the sign of the anomaly may not be assessed robustly in Fig. 4e and f. So the difference between CanESM2-LE and ERA-I NAO anomalies may be due to the fact that ERA-I composites are derived from 3 negative and 4 positive years, whereas the GCM data provide 264 negative and 263 positive years during 1981–2010.

The difference between SLP anomalies in positive and negative years representing the pressure variability is indicated by white lines. These NAO centers of action reach GCM (ERA-I) SLP differences between positive and negative conditions of about 12.5 (17.5) hPa in the IL region and 7.5 (10.0) hPa in the AH region. They do not coincide with the highest and lowest SLP values in the neutral state but are situated near the 3×3 GCM grid cell matrices used for index calculation (see Fig. 2). This supports the choice of these SLP centers for index calculation.
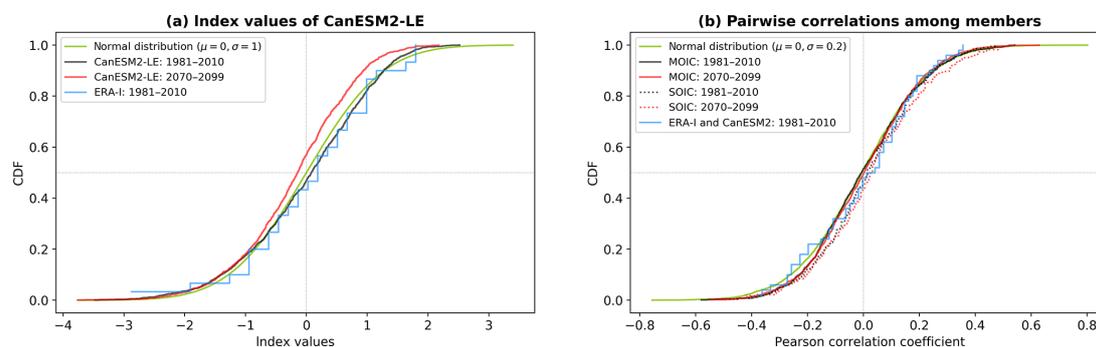
Figure 3. Cumulative density functions (CDFs) of NAO index values. **(a)** Distribution of all CanESM2-LE ($n = 50 \times 30$ per period) and ERA-I ($n = 30$) NAO index values. **(b)** Pairwise correlations among member NAO index time series from the same ocean families (SOICs, dotted lines, $n = 225$), from different ocean families (MOICs, solid lines, $n = 1000$) and between ERA-I and all CanESM2 members ($n = 50$). Black: 1981–2010 CanESM2-LE, red: 2070–2099 CanESM2-LE, blue: 1981–2010 ERA-I, green: normal distribution with $\mu = 0$ and $\sigma = 1$ in panel **(a)** and $\sigma = 0.2$ in panel **(b)**.

Under projected future climate conditions, SLP rises over large parts of the North Atlantic and shows less variability (see Fig. 4g–i). Future positive phases tend to be weaker as SLP shows a marked increase in the northern NAO node region. Negative phases exhibit SLP decreases in both node regions, although with larger changes near IL, resulting in negative phases to become slightly weaker as well.

Having established a reasonably plausible representation of the NAO in the driving data, the next step is to evaluate the large-scale NAO pattern in the RCM data. This is achieved by analyzing the deviations of RCM and driving data SLP variability. Figure 5 maps the RMSD between driving data and RCM SLP during 1981–2010 for driving data ERA-I (Fig. 5a) and CanESM2-LE (Fig. 5b), and CanESM2-LE in 2070–2099 (Fig. 5c). An $O(1)$ value of RMSD would indicate a poor reproduction of the SLP signal in the RCM because the RMSD between the RCM and driving data SLP is of the same order as the variability of the SLP in the driving data. Values of RMSD $\ll 1$, on the other hand, would indicate a good reproduction of the SLP signal in the RCM because it suggests that the RCM is tracking the variability in the driving data. With this understanding, it can be seen that the large-scale SLP pattern is reasonably well represented in most parts of the entire ClimEx domain for both driving data sets and both periods (significant at $p \leq 0.05$ using a t test with a false detection rate < 0.1 to account for multiple hypothesis testing, see Wilks, 2016). All subpanels in Fig. 5 show an RMSD increase towards the south, indicating that in these regions the control exerted by the lateral boundary conditions on the CRCM5 internal solution appears to be weaker. The RMSD is larger in the CanESM2/CRCM5 combination than in the ERA-I/CRCM5 combination and slightly increases in the future period in the southern parts (Fig. 5c). The differences of the spatial patterns are most likely due to different large-scale SLP patterns in both driving data sets which are in parts visible in Fig. 4a and d. In the CEUR domain (red box in Fig. 5), however, errors are low in general, and therefore the NAO pattern of the driving data may be assumed to be correctly incorporated there. It is thus reasonable to continue with the evaluation of nSAT and PR responses in the CEUR domain.

3.1.2   Local climate response to the NAO

nSAT and PR spatial responses as revealed in the ERA-I data are generally reproduced under current climate conditions in CanESM2-LE and CRCM5-LE (see Figs. 6–8). The highest magnitudes of the NAO responses (i.e., the slope of the regression line, α_1, introduced in Eq. 4) occur in the CRCM5/ERA-I run for all variables. In general, CRCM5 produces stronger α_1 response values at the local scale than the driving data. Regarding the absolute α_1 values, the CRCM5-LE mean meets the ERA-I better than the CRCM5/ERA-I run.

Positive NAO conditions are accompanied by winters with warmer temperatures (up to $+2$ K per unit index change; Fig. 6) and less day-to-day nSAT variability compared to neutral conditions (Fig. 7). The mainly positive relationship between nSAT mean and the NAO (Fig. 6) is strongest in the northeastern parts of the domain. Regionally, the NAO explains up to 40 %–60 % of nSAT mean variability (see also Fig. A2, where the nSAT mean α_1 share of the entire winter standard deviation of daily temperature values is shown). Explained variance is highest in the CRCM5/ERA-I run and lowest in CanESM2-LE.

The reduction of nSAT variability reaches up to 0.4–0.6 K in the northeastern continental section, while it is near zero in the southern part of the domain (Fig. 7).

In comparison to the neutral state, positive phases are also accompanied by more humid conditions in the north and drier conditions in the south of the CEUR domain (Fig. 8). The strength of the NAO–PR relationship, expressed by a
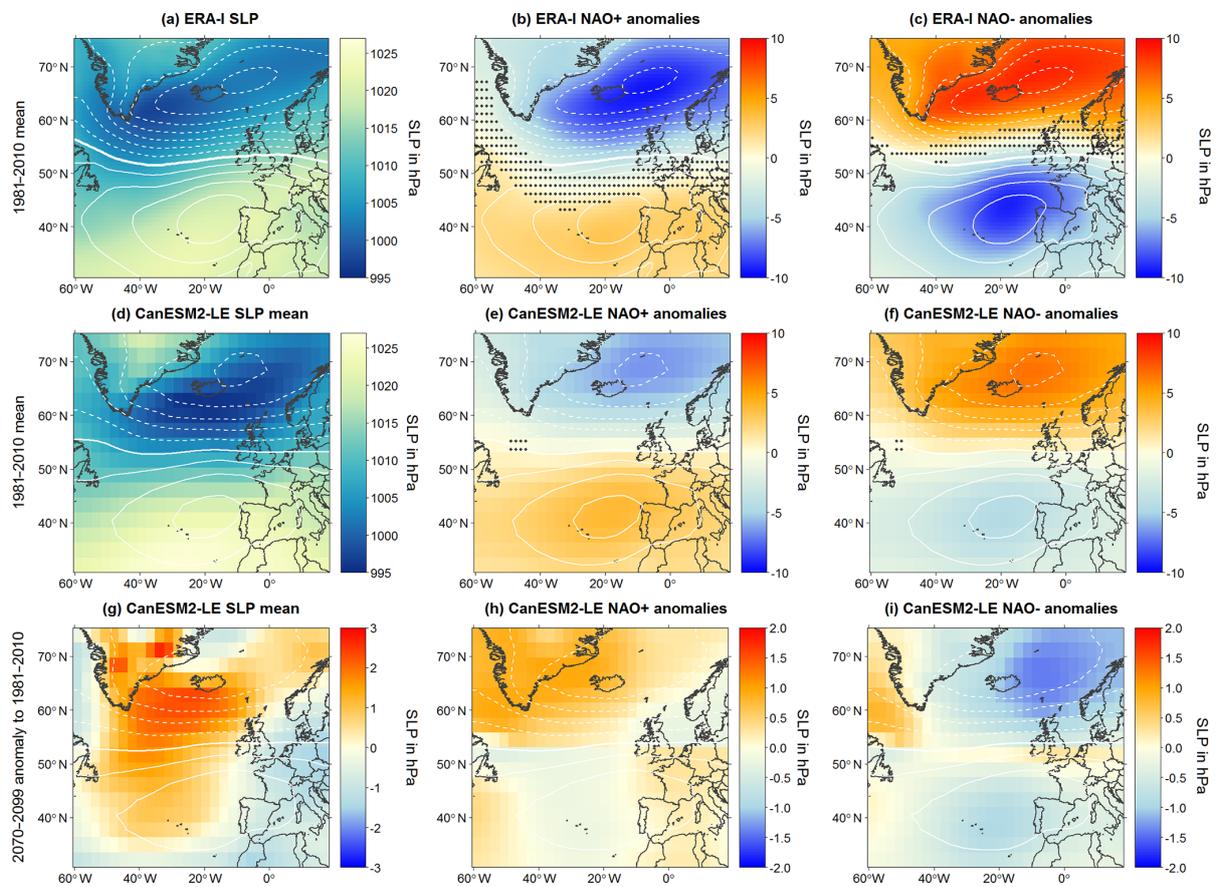
Figure 4. NAR winter mean SLP (hPa) composites in ERA-I **(a–c)** and CanESM2-LE **(d–i)** data showing long-term neutral conditions **(a, d, g)**, NAO positive **(b, e, h)** and negative anomalies **(c, f, i)**, **(a–f)** for 1981–2010 and **(g–i)** 2070–2099 changes with respect to 1981–2010 in GCM data. White isolines indicate the difference between positive and negative anomalies by a step of 2.50 hPa, as, e.g., in Hurrell (1995); solid: positive, dashed: negative, bold line: zero. Grey stippling in subpanels **(b–c)** and **(e–f)** indicates regions where the anomaly is smaller than the standard error of the composite samples.
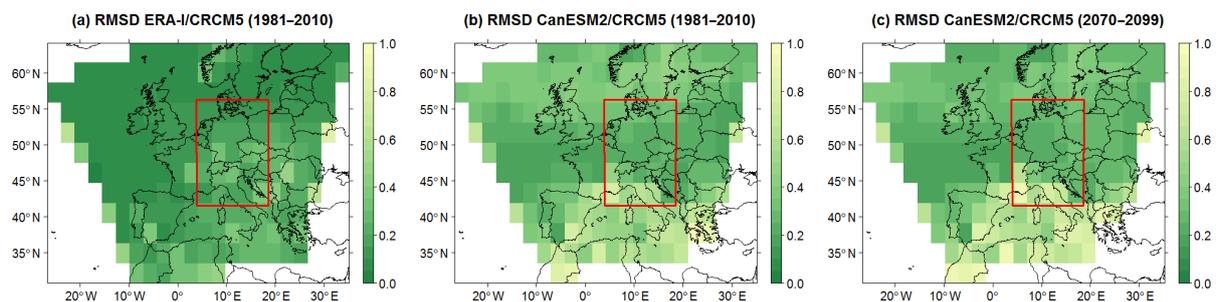
Figure 5. RMSD of monthly SLP differences between driving data and CRCM5 members, calculated following Eq. (2). Coloring: RMSD ≤ 1 significant at $p \leq 0.05$ with a false detection rate smaller than 0.1 (see Wilks, 2016): **(a)** for driving data ERA-I (1981–2010, one realization), **(b)** for driving data CanESM2-LE (1981–2010, 50 members) and **(c)** for driving data CanESM2-LE (2070–2099, 50 members). Red box: position of CEUR domain.
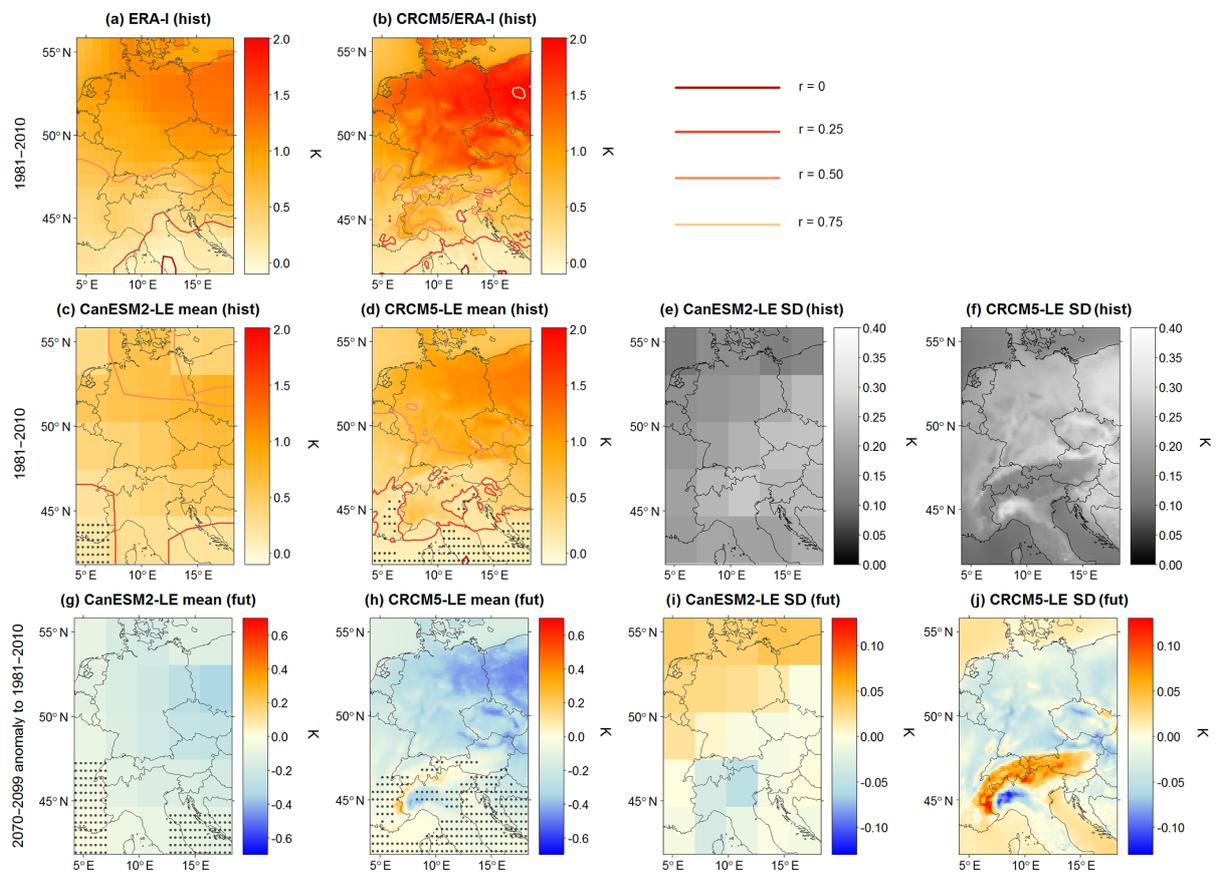
Figure 6. Spatial patterns of change in nSAT mean (α_1 in K) for a unit change in the NAO index for ERA-I, CRCM5/ERA-I, CanESM2-LE and CRCM5-LE in 1981–2010 **(a–f)** and the change in 2070–2099 with respect to 1981–2010 **(g–j)**. Both 50-member ensembles are represented with ensemble mean **(c–d, g–h)** and standard deviation (SD, **e–f, i–j**) representing the inter-member spread. Reddish lines in the ensemble mean maps represent the Pearson correlation between nSAT mean and the NAO index at an increment of 0.25; red shadings – see legend in upper right panel. Grey stippling in the ensemble mean maps indicates regions where SNR < 1, with SNR being the signal-to-noise ratio between the 30-year ensemble mean and SD of GCM- and RCM-LEs in both time periods. Stippling will be explained in more detail in Sect. 3.2.2.

Pearson correlation coefficient (r), is not affected by topography in any of the models within the domain; only the pivotal line crossing Europe is following the Alpine ridges (see solid dark line in Fig. 8a–d). The change between positive and negative r and α_1 occurs within a very narrow region. Within CanESM2-LE, this zero-line is shifted northwards compared with ERA-I, CRCM5/ERA-I and CRCM5-LE. As is visible in Fig. 8, higher α_1 values in mountainous regions indicate strong NAO responses related to orography. Regionally, the NAO accounts for 40 %–50 % of total PR sum variance, in both positively and negatively correlated regions. In CRCM5-LE, single spots in mountainous regions (e.g., in the Dinaric Alps) show extremely high PR sum α_1 values (up to ±220 mm per unit index change). In these parts the long-term mean PR sums are also very high. This stresses the more detailed production of geographical features but also the ten-

dency to evolve local extreme values in the high-resolution RCM (see similar results for local daily extreme precipitation in Leduc et al., 2019), which may even be noted in the (spatially aggregated) bias towards the GCM (see Fig. A1f). PR sum shows only weak correlations in the central region of the CEUR domain.

The mean state of nSAT and PR changes in the transient climate simulation towards warmer and moister conditions with less intra-seasonal variability of nSAT. For a detailed description of the future climate evolution (though for 2080–2099) in Europe within CRCM5-LE, see Leduc et al. (2019). Future NAO–climate relationships weaken in general compared with the historical ones for all variables as can be inferred from the ensemble mean changes in panels of Figs. 6g–h, 7g–h and 8g–h. The spatial patterns of NAO-induced change do not change considerably between both
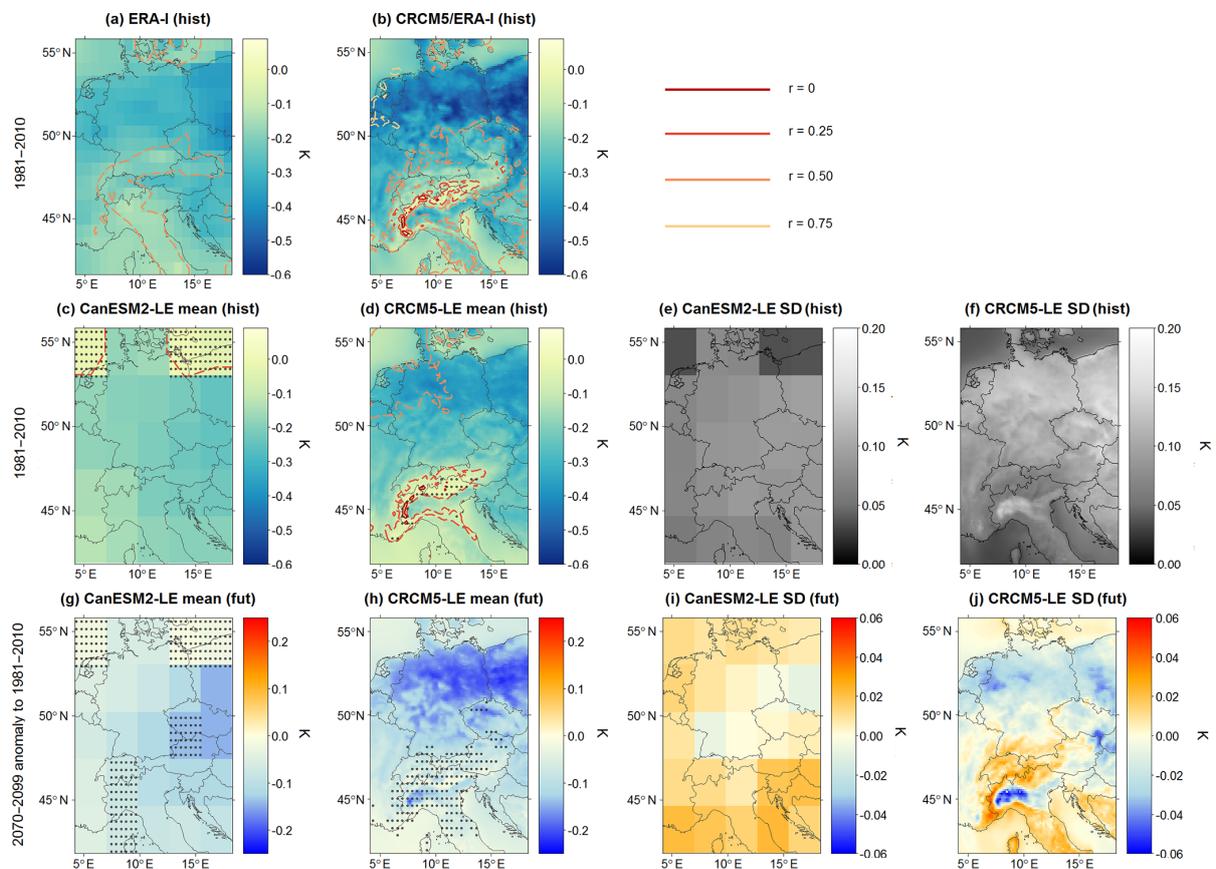
Figure 7. Like Fig. 6 but for nSAT SD (α_1 in K). Dashed lines of correlation coefficients indicate negative values. Note that the difference maps for CanESM2-LE and CRCM5-LE mean are calculated using absolute values.

periods. The response to the NAO, α_1, is clearly reduced in nSAT mean as well as nSAT SD, and there is also a reduction in PR sum change (Figs. 6g–h, 7g–h and 8g–h.).

### 3.2 Internal variability at the GCM and RCM scale

The next section focuses less on the ensemble mean changes but rather on their internal variability. The representation of internal variability in the GCM and RCM regarding the responses to the NAO in CEUR and subset regions NE, BY, SE is assessed via the inter-member spreads of CRCM5-LE and CanESM2-LE, and their differences.

3.2.1 Multi-member ensemble

CanESM2-LE reproduces typical NAO index characteristics. Figure 9 summarizes several statistics for all 50 GCM members as multiples of the reference, i.e., ERA-I, value. Generally, the ensemble meets the ERA-I value in all aspects of the NAO index. However, some GCM members only reach

half of the ERA-I teleconnectivity values (minimum correlation between AH/IL time series: $r = -0.281$, not significantly different from zero at $p \leq 0.05$ using a t test; ERA-I $r = -0.699$). This finding is especially interesting as this metric quantifies the strength of the NAO within the individual members. The inter-member spread of the teleconnection strength does not change significantly over time, despite the SLP changes over the North Atlantic. The 2070–2099 NAO index exhibits less interannual variability, less positive phases, more neutral phases and a relative increase of negative phases but with reduced mean values (see also Fig. 3a).

The spatial NAO responses also show a considerable degree of internal variability. Its spatial distribution expressed by diverging ensemble members can be derived from Figs. 6e–f, 7e–f and 8e–f, presenting spatially distributed ensemble SD as a measure of inter-member spread. Locally, the RCM shows considerably higher spreads than the GCM. Largest deviations for nSAT mean are found in continental regions of CEUR, but they do not simply correspond to high
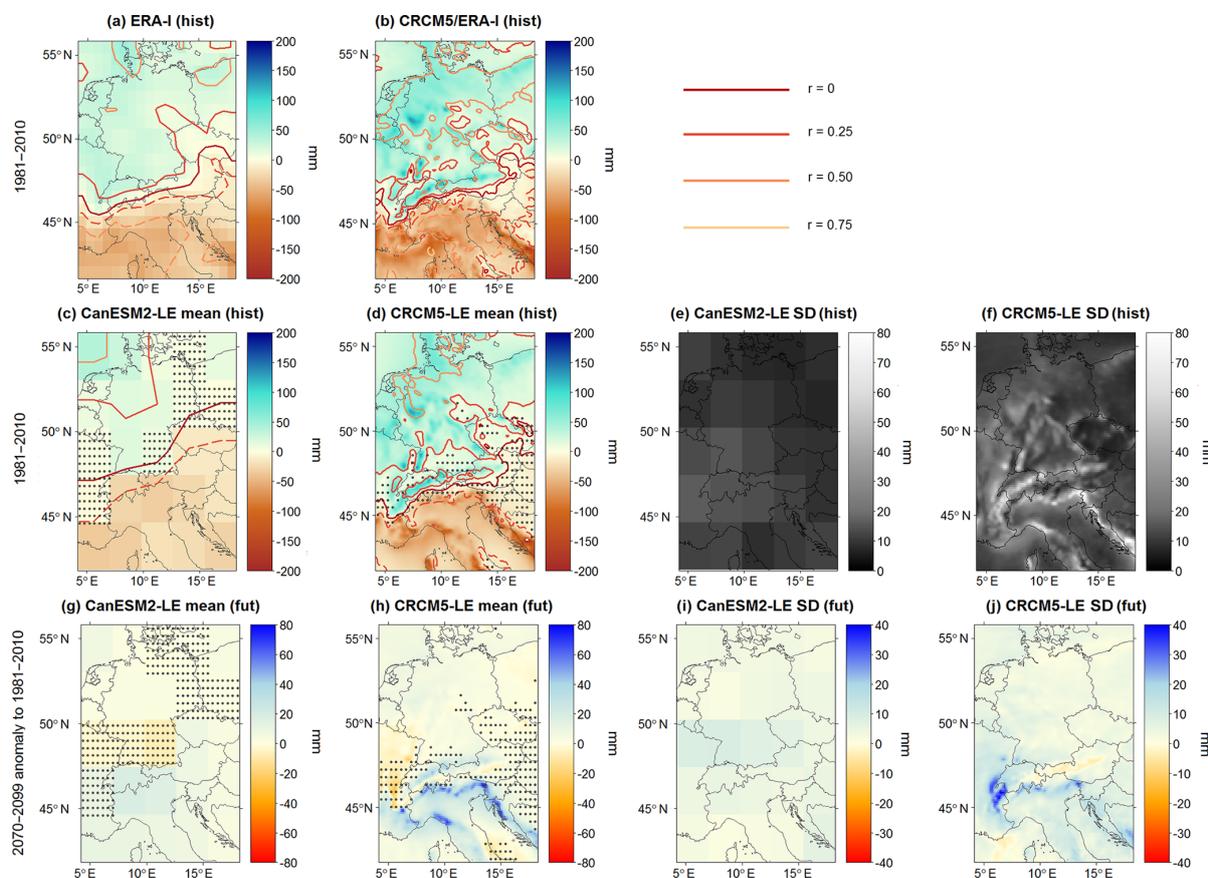
Figure 8. Like Figs. 6 and 7 but for PR sum (α_1 in mm). Dashed lines of correlation coefficients indicate negative values. Note that the difference maps for CanESM2-LE and CRCM5-LE mean are calculated using absolute α_1 values and that the color bar in the bottom row is flipped compared with Figs. 6 and 7.

or low α_1 (see also Fig. A3a–d). Low inter-member spread corresponds mostly to Alpine and sea regions.

The stippling in Figs. 6–8c–d and g–h indicates regions where the variability among the members is larger than the ensemble mean response, i.e., where the signal-to-noise ratio (SNR) between ensemble mean and SD lies below 1. For nSAT mean, the SNR exceeds 1 in most regions north of the Alps (Fig. 6c, d, g and h). nSAT SD shows SNR < 1 in the northern parts of the CanESM2-LE data (Fig. 7c, g) and in the Alpine region of the CRCM5-LE data (Fig. 7d, h). This variable shows a strong linear relationship between LE mean and SD (Fig. A3e–h). Regarding PR sum, RCM members vary most in regions with highest absolute α_1 values and altitudes, but there is no clear dependence in the GCMs (Fig. A3i–l). An east–west corridor of SNR values below 1 accompanies rather low α_1 values of PR sum (Fig. 8c, d, g, and h).

In addition to future changes in the NAO responses ensemble means, there is also a change in the spatial distribu-

tion of the inter-member spread expressed as ensemble SD (Figs. 6i–j, 7i–j and 8i–j).

To further investigate the inter-member spread, Fig. 10 illustrates the Pearson correlation coefficients r between the NAO index and subset regions for nSAT mean or PR sum in GCM- and RCM-LEs separately. In these boxplots, the variability among the members is illustrated by the box size, i.e., the interquartile distance. Both ensemble inter-member spreads generally envelope the ERA-I value (dashed line) of the given region, apart from GCM hist in Fig. 10b. This general finding does not change in the projected future climate: most boxes and whiskers keep their size, only GCM nSAT in the NE region is characterized by a larger range in the future (significant at $p \leq 0.05$, using an F test for comparison of variances). Some of the ensemble mean values exhibit a significant shift towards lower r values in the future for both models for nSAT mean and PR sum (see text insertions: CanESM2(hist, fut) and CRCM5(hist, fut)). An unpaired Mann–Whitney/U test is applied here as the samples
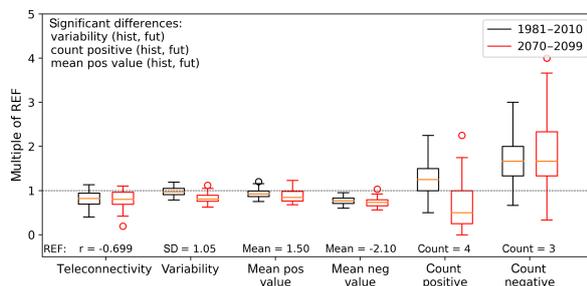
Figure 9. Several index statistics of all 50 CanESM2-LE members expressed as multiples of the respective ERA-I value (ERA-I value set to 1.0): teleconnectivity (Pearson correlation between AH and IL time series), index variability (expressed as temporal standard deviation of index time series), mean value of all positive (negative) phases and count of all positive (negative) phases per realization. Positive (negative) years are defined by an absolute index value exceeding 1. Text denotes combinations of which the differences are significant at $p \leq 0.05$ using an unpaired Mann–Whitney/U test. Orange line in boxplots is the median.

from hist and fut are seen as being drawn from different climates (since the null hypothesis of independence between hist and fut periods could not be rejected at $p \leq 0.05$ using a χ^2 test).

3.2.2   Change of scales

Having analyzed GCM and RCM separately so far, the next step is to compare both ensembles. A χ^2 test reveals that GCM and RCM samples of r can be seen as significantly dependent in both time frames. The amount of variance explained by the NAO is generally higher in the ERA-I reference than in the RCM ensemble mean. CRCM5-LE enhances the relationship showing higher r and α_1 values than CanESM2-LE (see Fig. 10 for r and Figs. 6–8 for α_1). This enhancement by CRCM5 is notably independent of the driving data: for both variables, the CRCM5/ERA-I r value (dotted lines in Fig. 10) is also found to be higher than the ERA-I value in most regions (dashed lines in Fig. 10). In all subset regions, the CRCM5/ERA-I r value lies in the upper part (stronger correlations) of the CRCM5-LE ensemble values.

Figure 10 shows that mean r values of RCM (grey filling) and GCM (white) members are significantly different in all subset regions for nSAT mean in both time horizons but only in the NE and BY regions for PR sum; in the SE region, only weak differences between GCM and RCM PR sum r distributions are visible. In NE and BY regions this difference is expressed by higher r values in RCM data, whereas in the SE region lower r values are found in the RCM data (only for nSAT mean). Apart from PR sum in the NE region (both time horizons), no significant difference between the spread amplitudes of GCM and RCM is visible ($p \leq 0.05$, F test). The inter-member spread of the correlation between NAO and re-

sponse variables is not generally altered during the nesting process.

To evaluate the covariability of CanESM2 and CRCM5 data in the subset regions, time series of the response variables originating from both data sources are correlated member-wise (see Fyfe et al., 2017, for a similar approach). As can be seen in Fig. 11, highest accordance on average is reached for nSAT mean in both periods, indicating that CanESM2-LE and CRCM5-LE show very similar temporal variability for this variable. The covariability of GCM and RCM time series is weaker for PR sum (Fig. 11b) and nSAT SD (Fig. 11c) than for nSAT mean (Fig. 11a) in both periods. Also, the inter-member spread is larger for PR sum and nSAT SD than for nSAT mean. This finding suggests that there is a larger discrepancy in portraying PR sum and nSAT SD in the RCM with respect to the GCM compared with nSAT mean; i.e., the RCM does not generally track the variability induced by the GCM for these variables. The correlations of nSAT mean and PR sum between CanESM2 and CRCM5 subset regions are in general significantly lower under future climate conditions compared with the historical ones, apart from nSAT mean in the BY region and PR sum in the SE region (see text in Fig. 11). For nSAT SD a significant shift of the distribution of r towards higher values is visible, apart from nSAT SD in the BY region. All variables exhibit a future inter-member spread increase, but not all subset regions are affected (e.g., nSAT mean in BY or nSAT SD in SE, Fig. 11). This suggests that under future climate conditions a potential reduction of GCM–RCM covariability needs to be considered, at least for PR sum and (weaker) for nSAT mean.

4   Discussion

4.1   General performance of the model chain

The ClimEx climate data ensemble is able to reproduce a NAO-like pattern with realistic temporal and spatial characteristics over the North Atlantic and corresponding response patterns in central Europe. Ensemble mean information aggregates several realizations, and so differences towards the single ERA-I realization are to be expected. However, results show that the ERA-I pattern may in general be seen as being "embedded" in the RCM or GCM inter-member spread, implying that the GCM, RCM and reference data share comparable climate statistics.

Regarding temperature, Europe is commonly seen as divided into a region with positive NAO–response correlations in the north and negative correlations in the south (see, e.g., Woollings et al., 2015). The first is found in the here-presented results; the latter is not clearly visible in the chosen domain. nSAT SD is correlated negatively with the NAO, pointing towards less temperature variability in winters with positive NAO phases and a higher variability during negative phases. Correlations of PR sums and NAO are in accordance with the prevalence of large-scale (frontal) precipitation in
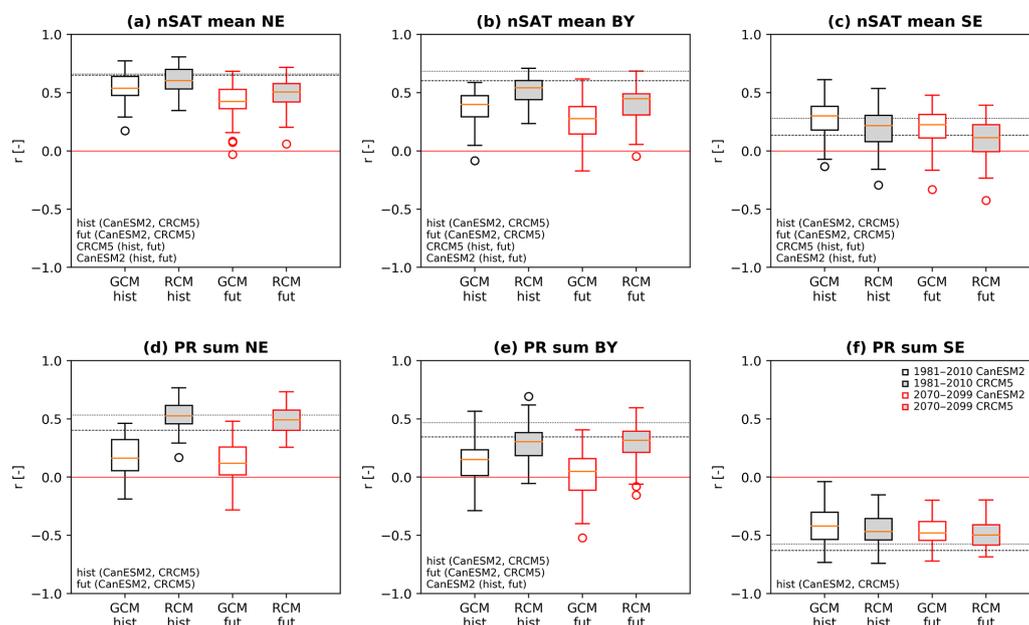
Figure 10. Boxplots of nSAT mean **(a–c)** and PR sum **(d–f)** showing Pearson correlation (r) with the NAO index of 50 CanESM2-LE (white filling) and CRCM5-LE (grey filling) realizations for three regions (NE, BY, SE) in historical (black outlines) and future (red outlines) time horizons. Dashed (dotted) horizontal lines indicate the ERA-I (CRCM5/ERA-I) value; text denotes combinations of which the differences are significant at $p \leq 0.05$ using an unpaired Mann–Whitney/U test for the comparison between hist and fut periods and a paired Wilcoxon test for the comparison between CanESM2-LE and CRCM5-LE. Orange line in boxplots is the median. For regions NE, BY and SE, see Fig. 2.
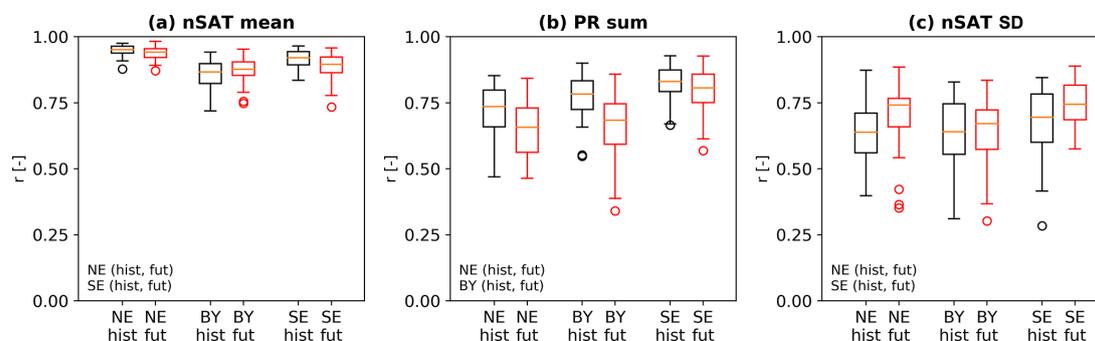
Figure 11. Temporal covariability of CanESM2-LE and CRCM5-LE subset regions in all 50 members. Each boxplot represents 50 Pearson correlation coefficients of the time series of variables nSAT mean **(a)**, PR sum **(b)** and nSAT SD **(c)** in the subset regions between CanESM2-LE members and the corresponding CRCM5-LE members. Time periods used for correlations are 1981–2010 (hist, black) and 2070–2099 (fut, red). Text denotes combinations of which the differences are significant at $p \leq 0.05$ using an unpaired Mann–Whitney/U test. Orange line in boxplots is the median. For regions NE, BY and SE, see Fig. 2.

winter which might be affected if the large-scale circulation is altered due to the NAO.

The strong SLP gradient under neutral NAO conditions over the North Atlantic noted in CanESM2-LE suggests an overestimation of the local atmospheric circulation with too strong westerlies. Similar model biases are widely reported (see, e.g., Ruprich-Robert and Cassou, 2015; Stephenson

et al., 2006; Reintges et al., 2017; Ulbrich et al., 2008). Since the NAO index was obtained from raw SLP data, it contains the contribution of the NAO but possibly also of microclimatic noise or other teleconnection patterns like the East Atlantic (EA) and the Scandinavian pattern (SCA) which interact with the NAO and exert a notable control on the North Atlantic SLP gradient according to Moore et al. (2013). These

authors investigated the contributions of the North Atlantic teleconnections NAO, EA and SCA in reanalysis data by separating them with empirical orthogonal functions. The authors found that the "pure" NAO accounts for about one-third of winter SLP variability, and the second and third leading modes for roughly 20 % and 15 %, respectively (see also Comas-Bru and McDermott, 2014). Thus, the results shown here may be seen as representing the superposition of these atmospheric modes.

The fidelity of NAO responses further depends on two aspects: (i) the goodness of representation of the large-scale NAO-related SLP pattern in CEUR and (ii) the strength of the linear relationship between the NAO and the response variables. The first point is addressed by a good representation of the SLP pattern in RCM data (see Fig. 5). The second point may be targeted by a combination of the strength of the responses (correlations r) and the response values themselves (α_1): NAO responses in the CEUR domain of all data sets are most reliable in regions where a strong linear relationship between the NAO and the response variable may be assumed. This may be the case if the correlation coefficient between the NAO index and the variable time series on the given grid cells is significantly different from zero. However, linearity does not apply under all conditions. For example, particularly strong negative NAO phases with low-ice conditions in the Arctic coincide with cooling in Europe that is weaker than expected from a linear relationship due to an accompanying warming over Siberia (Screen, 2017). Low correlation values may also suggest that climate variability in these regions is only to a small fraction influenced by the NAO in this data set and period under consideration. In these cases, the NAO as expressed by the North Atlantic SLP gradient in this study is not the most important contributor and the noise, ε_Y in Eq. (4), is dominant.

Historical changes induced by the NAO (α_1, all data sources) are generally in accordance with observed composite anomalies (see also Fig. A4) but most so in regions with significant correlations. Thus, the future change of nSAT and PR per unit index change is most valid where correlations are high and where the NAO related responses emerge from internal variability (i.e., SNR > 1). Of course, α_1 and composite maps are not identical, as on the one hand the average index value that accompanies nSAT and PR anomalies is not the same (±1 for α_1 but $+1.498$ and -2.103 for ERA-I composites; see Fig. 9). On the other hand, α_1 estimates a change which is singularly generated by the NAO index in a linear relationship, while composite maps originate from raw data which might include further influences.

4.2   Nesting approach

NAO response patterns are similar within CanESM2-LE and CRCM5-LE, but some deviations remain due to differences in model parameterization and spatial resolution. Another possible explanation could be that the control exerted by

CanESM2 through the CRCM5 lateral boundary conditions is insufficient, but this is unlikely given the relatively small CRCM5 domain implying stronger lateral boundary conditions control (Leduc and Laprise, 2009), in addition to the strong spectral nudging of large scales that was applied in the production of CRCM5-LE (Leduc et al., 2019). Also, the large-scale SLP pattern over CEUR shows no large errors in CRCM5-LE with respect to its driving data sources (see Fig. 5) and temporal correlation of GCM and RCM time series are generally high. Nevertheless, the influence of the lateral boundary conditions regarding SLP appears to vary over the CRCM5 domain, being a bit weaker in the southern part. It is worth noting that this feature is less pronounced when CRCM5 is driven by ERA-I as compared with CanESM2, highlighting the importance to investigate further the interactions between global atmospheric circulation, surface forcings (e.g., topography and land–sea contrasts) and local feedbacks.

CRCM5 reproduces the response structures much finer than CanESM2 and adds some robust high-resolution geographical features which are clearly visible within the ensemble mean. Apart from the coarser pattern resolution, there is also a shift in the spatial climate patterns in CanESM2-LE within the CEUR domain with respect to ERA-I data which are not found in CRCM5-LE: for example, typical continental climate features, such as high nSAT variability (as indicated in Fig. 6), are shifted southwards in CanESM2-LE with respect to CRCM5-LE data (or ERA-I). This shift may be explained by the fact that due to coarser spatial resolution the GCM topography shows land grid cells where the Mediterranean or the Baltic Sea extend in ERA-I and CRCM5; thus, in the GCM, the continent Europe also occupies a region which is sea in ERA-I. Assuming that the land–sea distribution affects the climate evolution, the GCM also experiences a geographical shift of climatic characteristics (such as continental properties) compared with the ERA-I and RCM data within the study domain. Another example is the dividing line for NAO–PR sum relations (see Fig. 8), which shows a displacement in the GCM compared with the RCM. This displacement is related to the GCM orography which deviates due to the coarser spatial resolution in shape, position and height from the RCM orography. These findings suggest that similar responses of GCM and RCM to the NAO may not be visible at the same geographical location (i.e., coordinates) but under similar geographical conditions (exposition, altitude, distance to sea). Continuing this thought, the RCM reproducing the spatial climatic patterns in the "correct" location is another expression of the RCM added value for regional- or local-scale analyses. However, for general statements on this issue, analyses on a larger domain would be necessary.

On the regional scale, the correlations in CRCM5 are significantly stronger in several regions than in CanESM2 (Figs. 6–8). These differences are not evened out by spatial aggregation. Thus, in CRCM5-LE, more variance is ex-

plained by the NAO (i.e., by large-scale circulation) than in CanESM2-LE. Explained variance is also higher in the single realizations of ERA-I and CRCM5/ERA-I than in the ensemble mean of GCM and RCM.

4.3   Internal variability

In general, the 50 NAO signals from the atmospheric "inflow" as given by the GCM boundary conditions are correctly translated into 50 regional responses of the RCM regarding the range of internal variability.

The large ensemble internal variability favors a smoothing of structures in the ensemble mean. Nevertheless, as the ensemble mean (GCM and RCM) reproduces patterns very similar to the observed ones, the atmospheric dynamics behind can be regarded as correctly reproduced in all members.

When looking at spatially explicit ensemble SD, the RCM-LE exhibits higher ensemble SD values than the GCM. This is in accordance with Giorgi et al. (2009) who stated that internal variability at finer scales tends to be larger compared with coarser scales. However, the amplitude of the inter-member spread of NAO–response correlations in the aggregated RCM and GCM subset regions is similar. Thus, the range of internal variability regarding the strength of the NAO–response relationship is transferred during nesting and the added CRCM5 internal variability (Leduc et al., 2019) does not significantly alter it. However, the ensemble values are shifted towards significantly higher r values in the RCM compared with the GCM in both time frames but not in the SE region.

4.4   Climate change

The results show that historical and projected future climate statistics deviate such that the comparison of relationships in both periods remains difficult: the NAO pattern changes, NAO index variability and nSAT and PR responses are reduced in the future climate simulation. Also the uncertainty range of the signals does not change significantly in the future horizon.

With the here-presented results, it can be argued that the internal variability of more complex parameters (such as the NAO–response relationship quantified via Pearson correlation) shows no significant changes between historical and future periods. When looking at the spatial distribution of α_1 ensemble SD, however, several regions show slight future increases or decreases which are not necessarily consistent between GCM and RCM.

It has to be added that this study evaluated two 30-year blocks rather than continuous time series, treating the NAO–response relationship and the inter-member spread as stationary during these blocks such that the inter-member spread of both periods represents generalized conditions for 1981–2010 and 2070–2099. According to Comas-Bru and McDermott (2014), potential non-stationarity in NAO–response re-

lationships can at least partly be attributed to influences of the EA/SCA patterns on the NAO and especially the geographical position of the North Atlantic SLP gradient.

The relative prevalence of negative index phases in the future period occurs in correspondence to a generally strengthened high-pressure ridge over the North Atlantic and especially Greenland (see Fig. 4g). The latter feature is supposed to be related to the emergence of negative index phases (Hanna et al., 2015; Woollings et al., 2010; Gillett and Fyfe, 2013; Cattiaux et al., 2013; Screen, 2017). Another relationship ties the emergence of negative NAO index phases to reduced sea ice extents: Warner (2018) found that particularly October sea ice extent over the Barents/Kara Sea is positively correlated with the NAO in that it leads to strengthened IL and AH. Consequently, a reduced sea ice extent is associated with negative NAO phases, but this relationship is not simply linear (Warner, 2018). For example, Screen (2017) notes that negative NAO events tend to be stronger during winters with low sea ice extents. The NAO–sea ice relationship may follow from sea ice effects on the stratospheric polar vortex or from tropospheric Arctic amplification which reduces the meridional temperature gradient leading to a weakened, more wavy jet stream in the midlatitudes (Warner, 2018). CanESM2-LE is known to show a low bias regarding Arctic sea ice in all seasons compared with observations (Kushner et al., 2018), but it follows quite correctly the observed downward trend (Kirchmeier-Young et al., 2017) and leads to a clear reduction of sea ice in the 2070–2099 horizon compared with 1981–2010 in the entire Arctic and also the Barents/Kara Sea, as is verified with the CanESM2 variable "sea ice concentration" for this study (not shown).

An increasing frequency (relative to positive phases) of negative NAO events as noted in the presented results favors more cold and harsh winters in theory due to the advection of continental Eurasian air masses (Screen, 2017) which is in great contradiction to projected future background conditions (warmer, moister, see Leduc et al., 2019) that would rather, likewise following from theory, accompany positive phases. On the other hand, the response to NAO impulses is clearly reduced for nSAT mean, PR sum and nSAT SD. A coherent explanation for this discrepancy might be that as correlations weaken, the Eurasian influence (advection of cold, dry air masses) during negative phases may be repressed or weaker in its occurrence than now or, as indicated by Screen (2017), is actually increasing warmer air mass advection. As less nSAT and PR variance is explained by the NAO in the future climate projections than in the historical period, the influence of this climate mode on CEUR climate may be seen as potentially reduced.

5   Conclusions

In this study, an RCM single-model initial-condition large ensemble is analyzed with a special focus on the downscaled

responses to a teleconnection, the NAO, that is present in the driving data. For proper assessment, the driving GCM ensemble is also included in the study. With regard to the key questions raised in the introduction, the following can be stated:

a. The ClimEx RCM-LE and its driving GCM-LE are able to depict a robust NAO pattern under current forcing conditions. Each member represents a distinct climate evolution while sharing comparable statistics with all other 49 realizations and producing NAO and response patterns that are more robust than patterns of individual realizations. The ensemble also shows climate statistics that are comparable with the reference time series and patterns. The clearly visible connection of the NAO with nSAT mean and PR sum follows well-known patterns and allows to derive robust information on the influence of the NAO on nSAT variability (nSAT SD).

b. The RCM is able to reproduce the large-scale SLP pattern and realistic response patterns in the analyzed domain. Clearly more topographic features are visible in CRCM5-LE than in CanESM2-LE, which suggests added value by the RCM regarding the evaluation of small-scale NAO impacts. Deviations of nSAT and PR responses between members vary spatially within the domain and are found mostly in regions with strongest NAO responses.

c. Internal variability of the NAO pattern is expressed very well within the 50-member single-model ensembles and easily spans the observations regarding various indicators. The range of NAO responses is represented consistently between the driving GCM and the nested RCM. The spread is shifted towards stronger NAO–nSAT/PR relations in the RCM compared with the GCM in both time horizons.

d. Concerning climate change, several changes go hand in hand: the winter index variability is reduced, the overall winter variability of nSAT and PR and also the fraction of NAO-explained nSAT is reduced, the relationship between NAO and response variables is weakened, and the covariability of CanESM2 and CRCM5 subset regions for all variables is reduced.

While these results are especially valid for the analyzed GCM–RCM combination, they allow drawing some general conclusions. The results strengthen the validity of this GCM–RCM combination for further applications, as important large-scale teleconnections only present in the GCM propagate properly to the fine-scale dynamics in the RCM. The RCM does not alter the spread of driving GCM data which is valuable information for impact modeling with a focus on internal variability. The results also stress the importance of single-model ensembles for evaluating and estimating internal variability since single realizations show considerable variations among themselves and also deviations from the ensemble mean. So the ensemble mean and the ensemble spread together are needed for robust assessment of climate modes and whether a given model is able to reproduce the phenomenon of interest.

Appendix A
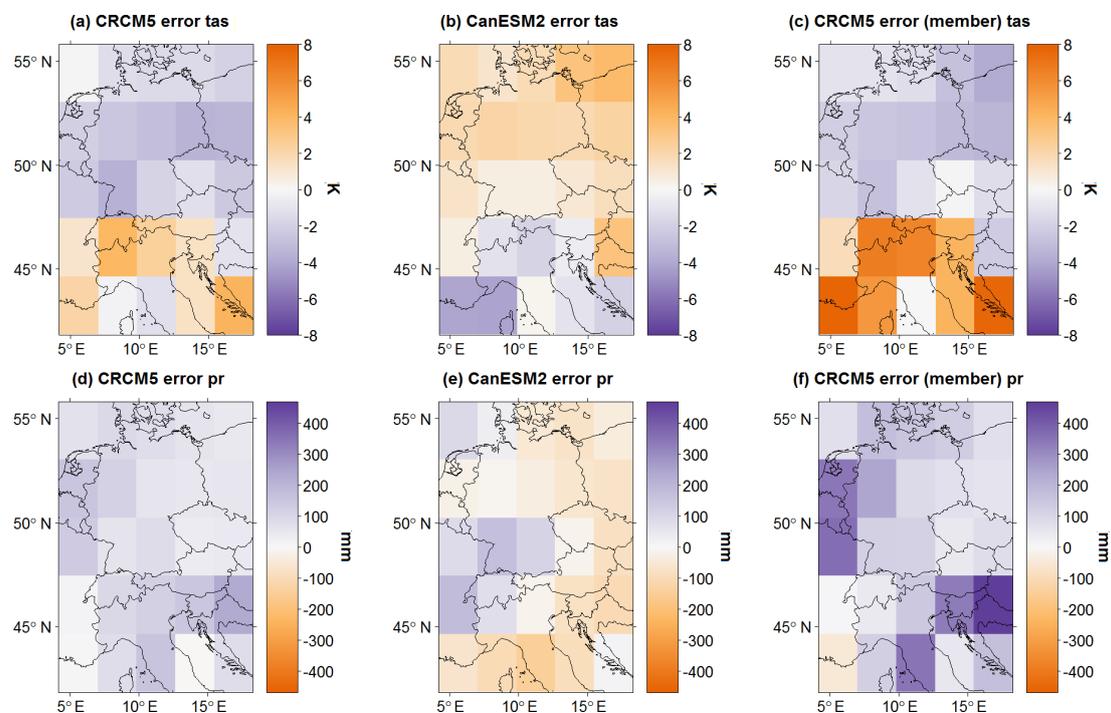
Figure A1. Model deviation for the 1981–2010 winter mean nSAT mean **(a–c)** and winter mean PR sum **(d–f)** in GCM resolution (2.8°). **(a, d)** Error of CRCM5 under ERA-I boundary conditions (difference between CRCM5/ERA-I and ERA-I). **(b, e)** Error of CanESM2-LE towards ERA-I data (ensemble mean of differences between CanESM2-LE members and ERA-I). **(c, f)** CRCM5 error under CanESM2-LE boundary conditions (ensemble mean of differences between CRCM5 members and corresponding CanESM2 members).

A. Böhnisch et al.: NAO response variability in an RCM large ensemble   635
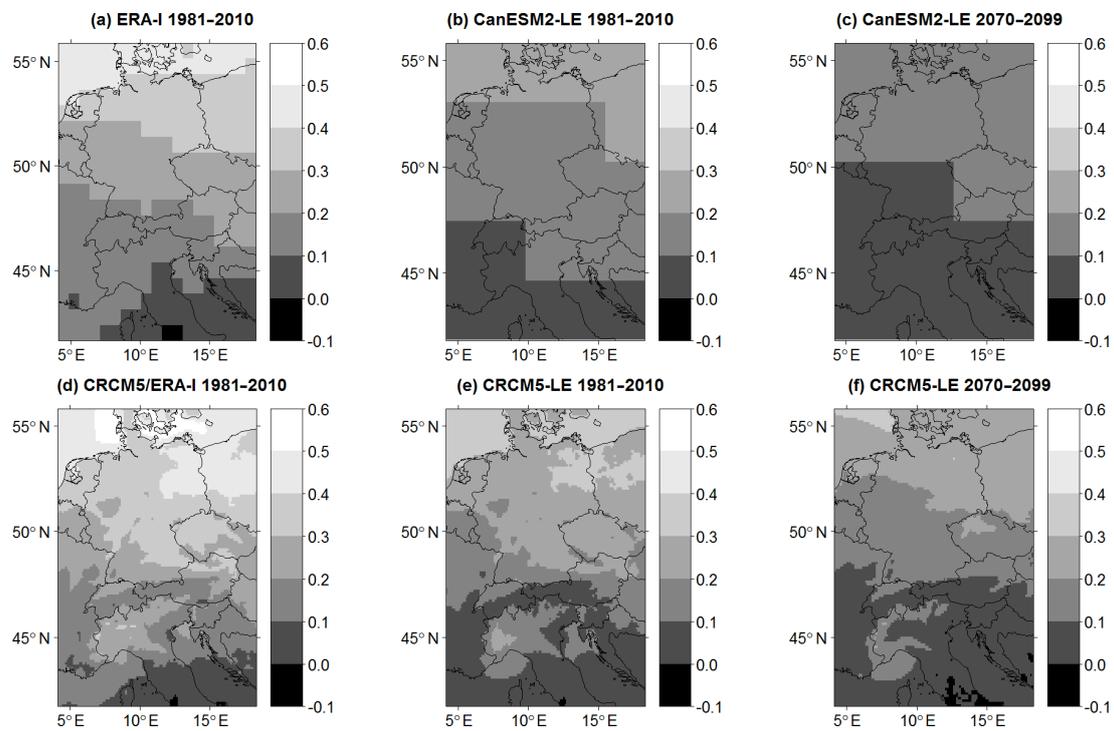
Figure A2. Ratio of nSAT α_1 and winter mean daily standard deviation of nSAT for driving data **(a–c)** and RCM data **(d–f)** during historical **(a, b, d, e)** and future **(c, f)** conditions. The panels show the proportion of nSAT α_1 in winter mean daily standard deviation of nSAT.
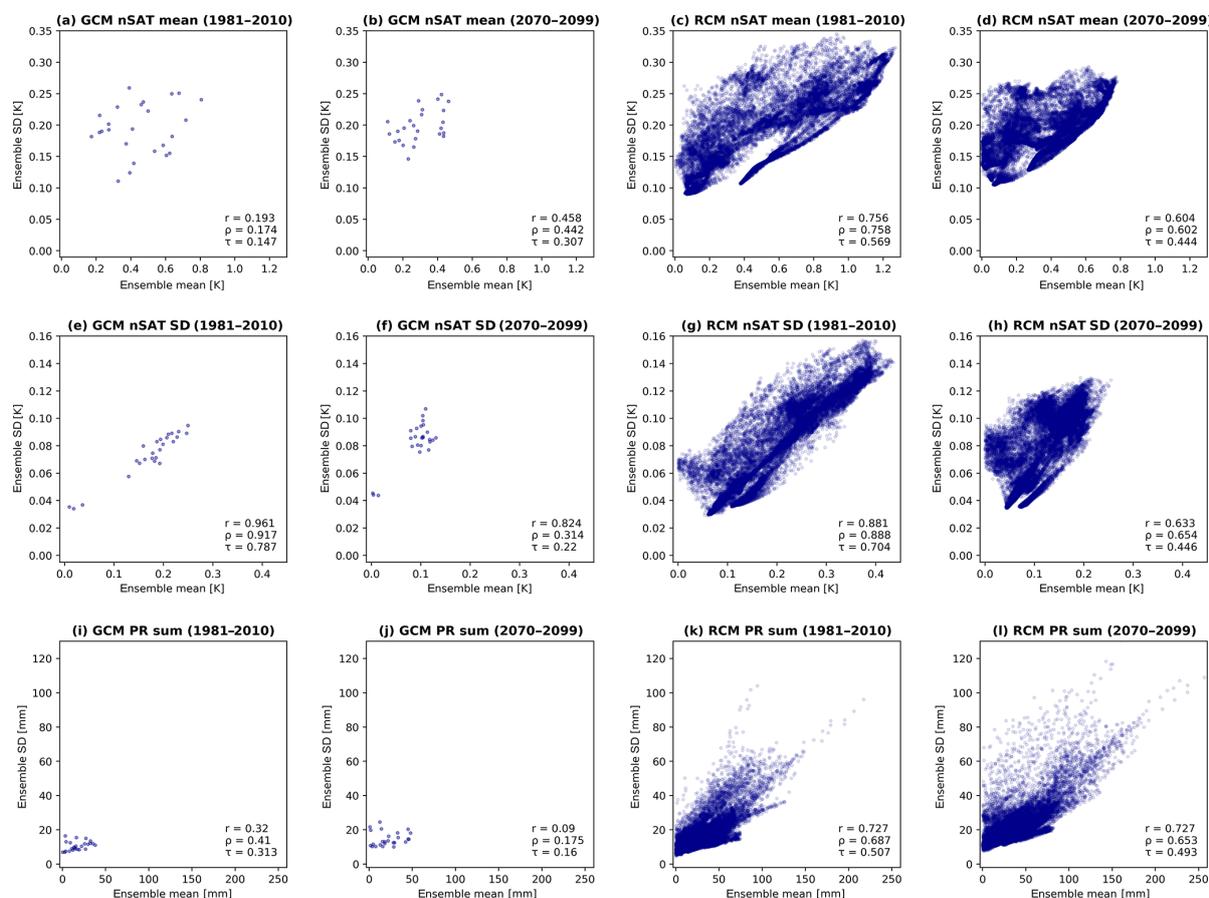
Figure A3. Relationship between LE mean and SD values of CanESM2-LE (GCM) and CRCM5-LE (RCM) for variables nSAT mean **(a–d)**, nSAT SD **(e–h)**, PR sum **(i–l)** for hist and fut periods. Lower right corner: r – Pearson correlation coefficient, ρ – Spearman rank correlation coefficient, τ – Kendall's tau.

(a) nSAT mean NAO+   **(b) nSAT mean NAO-**   **(c) PR sum NAO+**   **(d) PR sum NAO-**

(e) nSAT mean NAO+ (SD)   **(f) nSAT mean NAO- (SD)**   **(g) PR sum NAO+ (SD)**   **(h) PR sum NAO- (SD)**
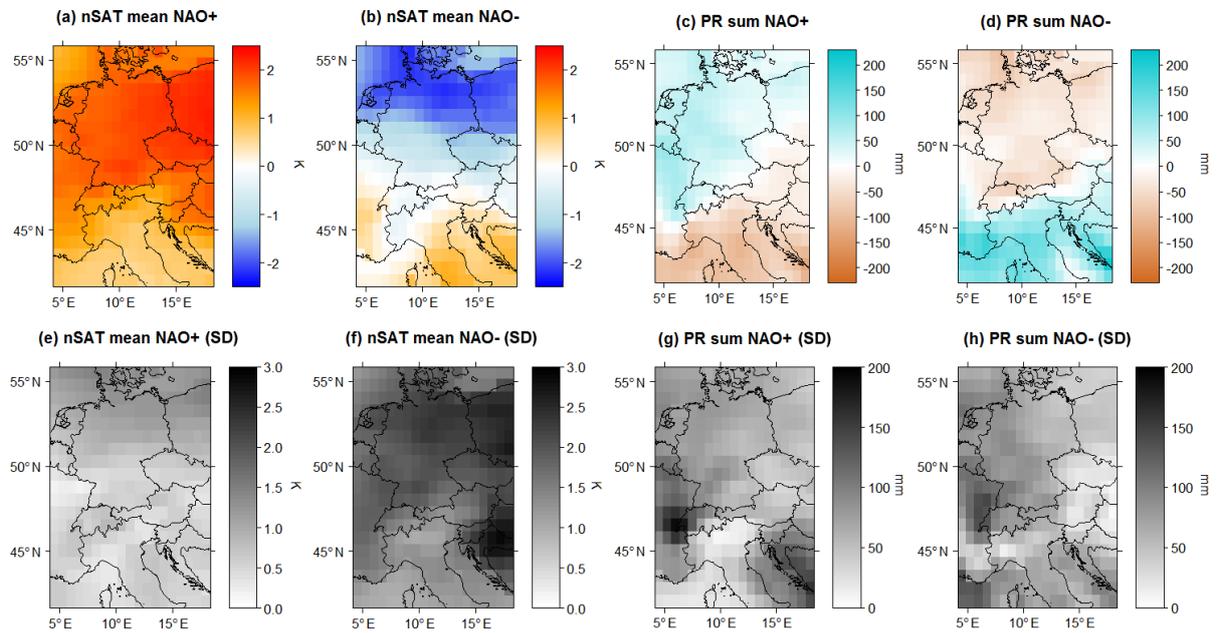
Figure A4. ERA-I anomalies from the long-term mean of nSAT mean in K and PR sum in mm in NAO positive (1989, 1990, 1994, 1995) and negative (1996, 2001, 2010) winters. Mean index value for positive (negative) NAO phases is $+1.498$ (-2.103).

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

Competing interests. The authors declare that they have no conflict of interest.

Special issue statement. This article is part of the special issue "Large Ensemble Climate Model Simulations: Exploring Natural Variability, Change Signals and Impacts". It is not associated with a conference.

References

Aalbers, E. E., Lenderink, G., van Meijgaard, E., and van den Hurk, B. J. J. M.: Local-scale changes in mean and heavy precipitation in Western Europe, climate change or internal variability?, Clim. Dynam., 50, 4745–4766, https://doi.org/10.1007/s00382-017-3901-9, 2018.

Andrews, M. B., Knight, J. R., and Gray, L. J.: A simulated lagged response of the North Atlantic Oscillation to the solar cycle over the period 1960–2009, Environ. Res. Lett., 10, 054022, https://doi.org/10.1088/1748-9326/10/5/054022, 2015.

Benedict, J. J., Lee, S., and Feldstein, S. B.: Synoptic View of the North Atlantic Oscillation, J. Atmos. Sci., 61, 121–144, https://doi.org/10.1175/1520-0469(2004)061<0121:SVOTNA>2.0.CO;2, 2004.

Cattiaux, J., Douville, H., and Peings, Y.: European temperatures in CMIP5: origins of present-day biases and future uncertainties, Clim. Dynam., 41, 2889–2907, https://doi.org/10.1007/s00382-013-1731-y, 2013.

Christensen, J. H., Krishna Kumar, K., Aldrian, E., An, S.-I., Cavalcanti, I. F. A., de Castro, M., Dong, W., Goswami, P., Hall, A., Kanyanga, J. K., Kitoh, A., Kossin, J., Lau, N.-C. and Renwick, J. and Stephenson, D. B. and Xie, S.-P., and Zhou, T: Climate Phenomena and their Relevance for Future Regional Climate Change, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, UK and New York, USA, 1271–1308, 2013.

Comas-Bru, L. and McDermott, F.: Impacts of the EA and SCA patterns on the European twentieth century NAO–winter climate relationship, Q. J. Roy. Meteorol. Soc., 140, 354–363, https://doi.org/10.1002/qj.2158, 2014.

Cropper, T., Hanna, E., Valente, M. A., and Jónsson, T.: A daily Azores-Iceland North Atlantic Oscillation index back to 1850, Geosci. Data J., 2, 12–24, https://doi.org/10.1002/gdj3.23, 2015.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteorol. Soc., 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Delworth, T. L., Zeng, F., Vecchi, G. A., Yang, X., Zhang, L., and Zhang, R.: The North Atlantic Oscillation as a driver of rapid climate change in the Northern Hemisphere, Nat. Geosci., 9, 509–512, https://doi.org/10.1038/ngeo2738, 2016.

Déqué, M., Rowell, D. P., Lüthi, D., Giorgi, F., Christensen, J. H., Rockel, B., Jacob, D., Kjellström, E., de Castro, M., and van den Hurk, B.: An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections, Climatic Change, 81, 53–70, https://doi.org/10.1007/s10584-006-9228-x, 2007.

Deser, C., Phillips, A., Bourdette, V., and Teng, H.: Uncertainty in climate change projections: the role of internal variability, Clim. Dynam., 38, 527–546, https://doi.org/10.1007/s00382-010-0977-x, 2012.

Deser, C., Hurrell, J. W., and Phillips, A. S.: The role of the North Atlantic Oscillation in European climate projections, Clim. Dynam., 49, 3141–3157, https://doi.org/10.1007/s00382-016-3502-z, 2017.

Environment and Climate Change Canada: CanESM2 Large Ensembles Output, available at: https://open.canada.ca/data/en/dataset/aa7b6823-fd1e-49ff-a6fb-68076a4a477c, last access: 3 July 2020.

European Centre for Medium-Range Weather Forecasts: ERA-Interim, available at: https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/, last access: 3 July 2020.

Frankcombe, L. M., England, M. H., Mann, M. E., and Steinman, B. A.: Separating Internal Variability from the Externally Forced Climate Response, J. Climate, 28, 8184–8202, https://doi.org/10.1175/JCLI-D-15-0069.1, 2015.

Fyfe, J. C., Derksen, C., Mudryk, L., Flato, G. M., Santer, B. D., Swart, N. C., Molotch, N. P., Zhang, X., Wan, H., Arora, V. K., Scinocca, J., and Jiao, Y.: Large near-term projected snowpack loss over the western United States, Nat. Commun., 8, 14996, https://doi.org/10.1038/ncomms14996, 2017.

Gillett, N. P. and Fyfe, J. C.: Annular mode changes in the CMIP5 simulations, Geophys. Res. Lett., 40, 1189–1193, https://doi.org/10.1002/grl.50249, 2013.

Giorgi, F., Jones, C., and Asrar, G. R.: Addressing climate information needs at the regional level: the CORDEX framework, WMO Bull., 58, 175–183, 2009.

Hanna, E., Cropper, T. E., Jones, P. D., Scaife, A. A., and Allan, R: Recent seasonal asymmetric changes in the NAO (a marked summer decline and increased winter variability) and associated changes in the AO and Greenland Blocking Index, Int. J. Climatol., 35, 2540–2554, https://doi.org/10.1002/joc.4157, 2015.

Hansen, F., Greatbatch, R. J., Gollan, G., Jung, T., and Weisheimer, A.: Remote control of North Atlantic Oscillation predictability via the stratosphere, Q. J. Roy. Meteorol. Soc., 143, 706–719, https://doi.org/10.1002/qj.2958, 2017.

Hawkins, E. and Sutton, R.: The Potential to Narrow Uncertainty in Regional Climate Predictions, B. Am. Meteorol. Soc., 90, 1095–1108, https://doi.org/10.1175/2009BAMS2607.1, 2009.

Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in projections of regional precipitation change, Clim. Dynam., 37, 407–418, https://doi.org/10.1007/s00382-010-0810-6, 2011.

Hurrell, J. W.: Decadal Trends in the North Atlantic Oscillation: Regional Temperatures and Precipitation, Science, 269, 676–679, https://doi.org/10.1126/science.269.5224.676, 1995.

Hurrell, J. W. and Deser, C.: North Atlantic climate variability: The role of the North Atlantic Oscillation, J. Mar. Syst., 79, 231-244 https://doi.org/10.1016/j.jmarsys.2009.11.002, 2010.

Hurrell, J. W. and Van Loon, H.: Decadal variations in climate associated with the North Atlantic Oscillation, Climatic Change, 36, 301–326, https://doi.org/10.1023/A:1005314315270, 1997.

Iles, C. and Hegerl, G.: Role of the North Atlantic Oscillation in decadal temperature trends, Environ. Res. Lett., 12, 114010, https://doi.org/10.1088/1748-9326/aa9152, 2017.

Jones, P. D., Osborn, T. J., and Briffa, K. R.: Pressure-Based Measures of the North Atlantic Oscillation (NAO): A Comparison and an Assessment of Changes in the Strength of the NAO and its Influence on Surface Climate Parameters, in: The North Atlantic Oscillation: Climatic Significance and Environmental Impact, edited by: Hurrell, J. W., Kushnir, Y., Ottersen, G., and Visbeck, M., AGU – American Geophysical Union, Washington, D.C., 51–62, 2013.

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S. C., Danabasoglu, G., Edwards, J., Holland, M., Kushner, P., Lamarque, J.-F., Lawrence, D., Lindsay, K., Middleton, A., Munoz, E., Neale, R., Oleson, K., Polvani, L., and Vertenstein, M.: The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability, B. Am. Meteorol. Soc., 96, 1333–1349, https://doi.org/10.1175/BAMS-D-13-00255.1, 2015.

Kirchmeier-Young, M. C., Zwiers, F. W., and Gillett, N. P.: Attribution of Extreme Events in Arctic Sea Ice Extent, J. Climate, 30, 553–571, https://doi.org/10.1175/JCLI-D-16-0412.1, 2017.

Kirtman, B., Power, S. B., Adedoyin, J. A., Boer, G. J., Bojariu, R., Camilloni, I., Doblas-Reyes, F. J., Fiore, A. M., Kimoto, M., Meehl, G. A., Prather, M., Sarr, A., Schär, C., Sutton, R., van Oldenborgh, G. J., Vecchi, G., and Wang, H. J.: Near-term Climate Change: Projections and Predictability, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, UK and New York, USA, 953–1028, 2013.

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. and Laprise, R.: Regional climate model sensitivity to domain size, Clim. Dynam., 32, 833–854, https://doi.org/10.1007/s00382-008-0400-z, 2009.

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.

Maher, N., Matei, D., Milinski, S., and Marotzke, J.: ENSO Change in Climate Projections: Forced Response or Internal Variability?, Geophys. Res. Lett., 45, 11390–11398, https://doi.org/10.1029/2018GL079764, 2018.

Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., Kröger, J., Takano, Y., Ghosh, R., Hedemann, C., Li, C., Li, H., Manzini, E., Notz, D., Putrasahan, D., Boysen, L., Claussen, M., Ilyina, T., Olonscheck, D., Raddatz, T., Stevens, B., and Marotzke, J.: The Max Planck Institute Grand Ensemble: Enabling the Exploration of Climate

System Variability, J. Adv. Model. Earth Syst., 11, 2050–2069, https://doi.org/10.1029/2019MS001639, 2019.

Moore, G. W. K., Renfrew, I. A., and Pickart, R. S.: Multidecadal Mobility of the North Atlantic Oscillation, J. Climate, 26, 2453–2466, https://doi.org/10.1175/JCLI-D-12-00023.1, 2013.

Osborn, T. J.: Simulating the winter North Atlantic Oscillation: the roles of internal variability and greenhouse gas forcing, Clim. Dynam., 22, 605–623, https://doi.org/10.1007/s00382-004-0405-1, 2004.

Ouranos: CRCM5-LE ClimEx, available at: https://climex-data.srv.lrz.de/Public/, last access: 3 July 2020.

Pokorná, L. and Huth, R.: Climate impacts of the NAO are sensitive to how the NAO is defined, Theor. Appl. Climatol., 119, 639–652, https://doi.org/10.1007/s00704-014-1116-0, 2015.

Reintges, A., Latif, M., and Park, W.: Sub-decadal North Atlantic Oscillation variability in observations and the Kiel Climate Model, Clim. Dynam., 48, 3475–3487, https://doi.org/10.1007/s00382-016-3279-0, 2017.

Rogers, J. C.: The Association between the North Atlantic Oscillation and the Southern Oscillation in the Northern Hemisphere, Mon. Weather Rev., 112, 1999–2015, https://doi.org/10.1175/1520-0493(1984)112<1999:TABTNA>2.0.CO;2, 1984.

Ruprich-Robert, Y. and Cassou, C.: Combined influences of seasonal East Atlantic Pattern and North Atlantic Oscillation to excite Atlantic multidecadal variability in a climate model, Clim. Dynam., 44, 229–253, https://doi.org/10.1007/s00382-014-2176-7, 2015.

Schulzweida, U.: CDO User Guide, Climate Data Operators Version 1.9.1, October 2017, MPI for Meteorology, available at: https://code.mpimet.mpg.de/projects/cdo/embedded/cdo.pdf (last access: 1 April 2018), 2017.

Screen, J. A.: The missing Northern European winter cooling response to Arctic sea ice loss, Nat. Commun., 8, 14603, https://doi.org/10.1038/ncomms14603, 2017.

Stephenson, D. B., Pavan, V., Collins, M., Junge, M. M., and Quadrelli, R., and Participating CMIP2 Modelling Groups: North Atlantic Oscillation response to transient greenhouse gas forcing and the impact on European winter climate: a CMIP2 multi-model assessment, Clim. Dynam., 27, 401–420, https://doi.org/10.1007/s00382-006-0140-x, 2006.

Ulbrich, U. and Christoph, M.: A shift of the NAO and increasing storm track activity over Europe due to anthropogenic greenhouse gas forcing, Clim. Dynam., 15, 551–559, https://doi.org/10.1007/s003820050299, 1999.

Ulbrich, U., Pinto, J. G., Kupfer, H., Leckebusch, G. C., Spangehl, T., and Reyers, M.: Changing Northern Hemisphere Storm Tracks in an Ensemble of IPCC Climate Change Simulations, J. Climate, 21, 1669–1679, https://doi.org/10.1175/2007JCLI1992.1, 2008.

von Storch, H. and Zwiers, F. W.: Statistical Analysis in Climate Research, Cambridge University Press, Cambridge, 2003.

von Trentini, F., Leduc, M., and 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. Dynam., 53, 1963–1979, https://doi.org/10.1007/s00382-019-04755-8, 2019.

Warner, J. L.: Arctic sea ice – a driver of the winter NAO?, Weather, 73, 307–310, https://doi.org/10.1002/wea.3399, 2018.

Wilks, D. S.: "The Stippling Shows Statistically Significant Grid Points": How Research Results are Routinely Overstated and Overinterpreted, and What to Do about It, B. Am. Meteorol. Soc., 97, 2263–2273, https://doi.org/10.1175/BAMS-D-15-00267.1, 2016.

Woollings, T., Franzke, C., Hodson, D. L. R., Dong, B., Barnes, E. A., Raible, C. C., and Pinto, J. G.: A Regime View of the North Atlantic Oscillation and Its Response to Anthropogenic Forcing, J. Climate, 23, 1291–1307, https://doi.org/10.1175/2009JCLI3087.1, 2010.

Woollings, T., Franzke, C., Hodson, D., Dong, B., Barnes, E., Raible, C., and Pinto, J.: Contrasting interannual and multidecadal NAO variability, Clim. Dynam., 45, 539–556, https://doi.org/10.1007/s00382-014-2237-y, 2015.

Xu, T., Shi, Z., Wang, H., and An, Z.: Nonstationary impact of the winter North Atlantic Oscillation and the response of mid-latitude Eurasian climate, Theor. Appl. Climatol., 124, 1–14, https://doi.org/10.1007/s00704-015-1396-z, 2016.

Zwiers, F. W. and von Storch, H.: On the role of statistics in climate research, Int. J. Climatol., 24, 665–680, https://doi.org/10.1002/joc.1027, 2004.

5.2   Paper II: Hot Spots and Climate Trends of Meteorological Droughts in Europe – Assessing the Percent of Normal Index in a Single-Model Initial-Condition Large Ensemble

Reference: 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. *Front. Water* 3:716621. DOI: 10.3389/frwa.2021.716621

Transition to paper II: With the CRCM5-LE reliably reproducing large-scale SLP variability, paper II systematically examines meteorological droughts in Europe for the first time in a regional SMILE. Confirming coarse-resolution multi-model trends (Spinoni et al. 2020), the study provides robust indication that climate change increases summer drought frequency, length, and intensity by comparing a future non-mitigation scenario (RCP8.5) against CRCM5-LE simulations without climate change (i.e., pre-industrial CO_2 concentrations). Internal and intra-annual climate variability are considered by estimating signal uncertainty and seasonally diverging drought trends, respectively. The study carves out European regions particularly subject to droughts (i.e., hot spots). In many of them, summers constitute the major precipitation season while also showing strongest drying trends. A temporally directed structure, the sequences of (non-)drought winters and summers, illustrates changing inter-seasonal drought characteristics. A dedicated goal of this paper is also the experimental use of scientific sound, yet accessible ways to convey internal variability and spatial patterns: Since results from SMILE analyses are often complex, communication requires dimension reduction while pertaining correctness. Using an easy-to-interpret index, the PNI, is considered to be central to this effort as well.

Author contributions: AB & MM: conception of the study, statistical analyses, and manuscript preparation in equal parts. Focus above: AB's analyses. RL: founder & head of the ClimEx project. ML: simulation of CRCM5-LE and pi-control runs. RL & ML monitoring and support of research process, revision of the manuscript. All authors contributed to the article and approved the submitted version.

Status: published in Frontiers in Water (Impact Factor: 2.9, 12/2023)

frontiers in Water

Check for updates

Hot Spots and Climate Trends of Meteorological Droughts in Europe–Assessing the Percent of Normal Index in a Single-Model Initial-Condition Large Ensemble

Andrea Böhnisch[1†], Magdalena Mittermeier[1†], Martin Leduc[2] and Ralf Ludwig[1]*

[1] *Department of Geography, Ludwig-Maximilians-Universität München (LMU), Munich, Germany, [2] Ouranos Consortium, Montreal, QC, Canada*

Drought, caused by a prolonged deficit of precipitation, bears the risk of severe economic and ecological consequences for affected societies. The occurrence of this significant hydro-meteorological hazard is expected to strongly increase in many regions due to climate change, however, it is also subject to high internal climate variability. This calls for an assessment of climate trends and hot spots that considers the variations due to internal variability. In this study, the percent of normal index (PNI), an index that describes meteorological droughts by the deviation of a long-term reference mean, is analyzed in a single-model initial-condition large ensemble (SMILE) of the Canadian regional climate model version 5 (CRCM5) over Europe. A far future horizon under the Representative Concentration Pathway 8.5 is compared to the present-day climate and a pre-industrial reference, which is derived from pi-control runs of the CRCM5 representing a counterfactual world without anthropogenic climate change. Our analysis of the SMILE reveals a high internal variability of drought occurrence over Europe. Considering the high internal variability, our results show a clear overall increase in the duration, number and intensity of droughts toward the far future horizon. We furthermore find a strong seasonal divergence with a distinct increase in summer droughts and a decrease in winter droughts in most regions. Additionally, the percentage of summer droughts followed by wet winters is increasing in all regions except for the Iberian Peninsula. Because of particularly severe drying trends, the Alps, the Mediterranean, France and the Iberian Peninsula are suggested to be considered as drought hot spots. Due to the simplicity and intuitivity of the PNI, our results derived from this index are particularly appropriate for region-specific communication purposes and outreach.

Keywords: drought, SMILE, pre-industrial, percent of normal, internal variability, large ensemble, Europe

1. INTRODUCTION

With progressing climate change droughts have become a critical high-impact hydro-meteorological hazard globally, and particularly in Europe, in recent years (Spinoni et al., 2018). In combination with high temperature anomalies, the deficit of precipitation has caused large economic, social and environmental costs in the years 2003, 2010 and 2018 (Bastos et al., 2020), e.g., through crop losses (Spinoni et al., 2018) or a drop in (renewable) power generation production (Naumann et al., 2021). Climate projections show that more damages are expected for the end of the twenty-first century (Spinoni et al., 2018). However, regional differences between single drought events are high (Beillouin et al., 2020) and thus call for the identification of geographical hot-spots of droughts under climate change.

Due to the complexity of drought impacts on diverse different sectors like water supply, agriculture, and ecosystems (Spinoni et al., 2018; Liu et al., 2019), a universal definition of droughts that satisfies all users is impractical (Lloyd-Hughes, 2014). Instead, droughts are classified as meteorological, hydrological, agricultural and socio-economic droughts (Mishra and Singh, 2010; Liu et al., 2019). Meteorological droughts focus on the deficit of precipitation over a region and period of time compared to normal conditions (Mishra and Singh, 2010; Sheffield and Wood, 2012) and do not consider factors like streamflow, soil moisture or water demand (Mishra and Singh, 2010; Lloyd-Hughes, 2014). Because meteorological droughts are a potential predecessor of other drought types, their investigation is key for diverse sectors of implications nonetheless. Within each category a variety of drought indices are in use, which enable the assessment of drought frequency, duration, severity and spatial extent (Mishra and Singh, 2010). A common index for meteorological droughts is the Palmer Drought Severity Index (PDSI), which includes the effects of temperature and evapotranspiration (Palmer, 1965). A popular drought index solely based on precipitation is the Standardized Precipitation Index (SPI; McKee et al., 1993), which compares the precipitation sum of a given period with a long-term reference by fitting it to a probability distribution and then transforming it to a normal distribution. The SPI expresses the deviation of a given period from the reference in measures of the standard deviation (Mishra and Singh, 2010). A less complex index for meteorological droughts is the Percent of Normal Index (PNI; Werick et al., 1994; Willeke and Hosking, 1994). It directly represents the percentage of precipitation of a specific period compared to the long-term mean. Due to its simple calculation and intuitive meaning it can serve for communication and outreach purposes (Smakhtin and Hughes, 2004; Nikbakht et al., 2013).

In the study of Spinoni et al. (2018), climate projections using simulations from the Coordinated Regional Downscaling Experiment (CORDEX) show an increase in the frequency and severity of droughts over the entire European continent under the Representative Concentration Pathway 8.5 (RCP 8.5): This increase occurs especially in spring and summer while during winters in northern Europe the occurrence of droughts is projected to decrease. The regions with the highest increase in drought frequency in the mentioned study are the Iberian Peninsula, southern Europe, France, the British Isles and north-eastern Scandinavia. Using another type of scenarios, Lehner et al. (2017) stress the potential to reduce mean dryness conditions and drought lengths in Europe in a scenario with 1.5°C increase with respect to pre-industrial conditions as opposed to a 2°C increase scenario by employing the PDSI in a 10-member ensemble. While there is a high agreement among different climate projections on a general drying trend over Europe in the course of the twenty-first century (Stagge et al., 2015; Dai and Zhao, 2017; Cook et al., 2020), results differ significantly between studies in regards to a potential drying trend in historical data (Blenkinsop and Fowler, 2007; Dai and Zhao, 2017; Stagge et al., 2017; Hänsel et al., 2019; Vicente-Serrano et al., 2021). The study by Stagge et al. (2017) reveals diverging drying trends over Europe from 1970 to 2014 depending on the choice of the drought index. Hänsel et al. (2019) state that the observed trend largely depends on the chosen study period due to a large temporal variability of drought conditions.

Several studies show that the occurrence of droughts is subject to a high natural variability (Bonaccorso et al., 2003; Santos et al., 2010; Hawkins and Sutton, 2011; Cook et al., 2016; Zhao and Dai, 2017; Mikšovský et al., 2019; Vicente-Serrano et al., 2021). Vicente-Serrano et al. (2021) state that current long-term drought trends in Western Europe are dominated by internal variability. Zhao and Dai (2017) show the same effect for several regions worldwide, whereas they find a consistency of observed drying trends and simulated forced signal over Southern Europe. These studies underline that natural variability is an important source of uncertainty and the understanding of the temporal and spatial variability of drought occurrence is a crucial knowledge that may help to enhance the management practices of these complex extreme events e.g., with respect to the management of water resources (Santos et al., 2010). A recent study of Spinoni et al. (2020) used a very large multi-model ensemble of regional climate models (RCMs) on a grid of 0.44° spatial resolution. While such a model-setup allows robust climate projections, it cannot distinguish between model uncertainty and internal variability. A single-model initial-condition large-ensemble (SMILE), which consists of several model runs of the same model driven by the same boundary conditions but slightly differed initial conditions, offers the possibility of studying the internal variability of extreme events under climate change (Kay et al., 2015).

In this study, we present the analysis of meteorological droughts in a SMILE that consists of 50 simulations of a RCM over Europe. We take advantage of the SMILE dataset in order to assess climate change signals of drought occurrence over Europe under the consideration of their high internal variability. We analyze regional hot-spots and the frequency, severity and duration of droughts using the PNI. Climate trends for the far future under the RCP8.5 scenario are compared to present-day and pre-industrial climate. The comparison to a pre-industrial climate state is possible through the employment of novel pi-control simulations, which accompany the SMILE. These allow to relate present-day drought conditions and future trends to a

counterfactual world without anthropogenic climate change. The choice of a simple drought index and the comparison with a pre-industrial reference directly aligns this study for communication and outreach purposes. To our knowledge this study is the first region-specific analysis of droughts over Europe that employs a SMILE of a high-resolution RCM in order to assess the uncertainty of internal climate variability.

2. DATA AND METHODS

2.1. CRCM5-LE

Precipitation based indices are known to show strong internal variability (e.g., Vicente-Serrano et al., 2021). They are also highly sensitive to orographic processes (Basist et al., 1994) and so a sufficient spatial resolution of the data used is crucial for meaningful analyses inside a topographically heterogeneous domain like Europe. We therefore use data from a SMILE of a RCM in order to assess the variability of meteorological droughts over Europe at high spatial resolution. The Canadian Regional Climate Model version 5 (CRCM5; Martynov et al., 2013; Šeparovic et al., 2013) Large Ensemble (CRCM5-LE) provides 50 members at 0.11° spatial resolution, which were obtained by dynamically downscaling the Canadian Earth System Model Large Ensemble (CanESM2-LE; Fyfe et al., 2017) during the period of 1950–2099 (Leduc et al., 2019). The CRCM5-LE and the CanESM2-LE were produced as described in Fyfe et al. (2017) and Leduc et al. (2019): The CanESM2-LE was produced by applying small random atmospheric perturbations to a 1,000-year equilibrium run (CMIP5 pi-control; Arora et al., 2011) in order to obtain five historical simulations starting in 1850. In 1950 new atmospheric perturbations were applied to each of these five families resulting in 50 CanESM2-LE members, which are (after a 5-year spin-off phase) independent realizations of the modeled climate. From 2006 onward, the RCP8.5 scenario serves as external forcing. These 50 CanESM2 members were used as boundary conditions for the CRCM5-LE during the dynamical downscaling process. By construction, all members share the same climatology, but span a range which allows to estimate internal climate variability. von Trentini et al. (2019) showed that the CRCM5-LE variability of precipitation change between 2070–2099 and 1980–2009 reaches the spread of a multi-model ensemble of 22 different RCM setups.

Furthermore, a set of dynamically downscaled pi-control runs exists, which consists of 35 simulations of the CRCM5 driven by the CanESM2-pi-control run in pre-industrial mode (Arora et al., 2011). The atmospheric CO_2 level of those pi-control runs corresponds to 285 ppm in the year 1850. The aerosol concentration of the pi-control runs also represents the pre-industrial mode (IPCC, 2013). The CanESM2-pi-control run covers 1,000 virtual years. During the downscaling with the CRCM5, 35 members with 22 years each are generated, whereby the first 2-years are considered as spin-up phase. The control runs thus provide a pre-industrial ensemble with a total of 700 years, which represents a climate without anthropogenic global warming.

TABLE 1 | Drought categories for certain PNI value ranges.

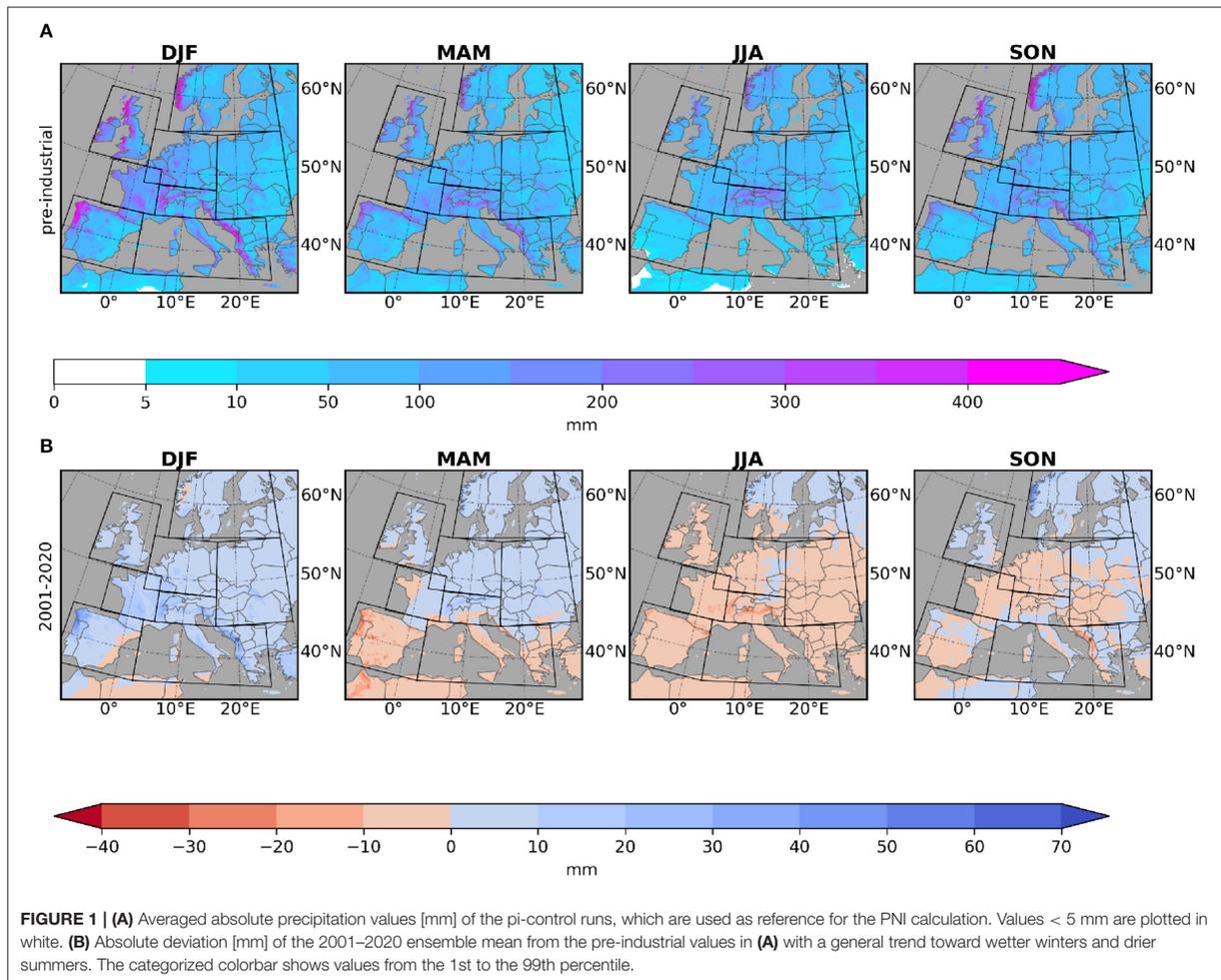| Drought category	PNI threshold [%]	Class width [%]
Slight drought	< 80	10
Moderate drought	< 70	15
Severe drought	< 55	15
Extreme drought	< 40	40

2.2. PNI

The percent of normal index (PNI) is a precipitation-based index for the evaluation of meteorological droughts (Werick et al., 1994; Willeke and Hosking, 1994). The PNI reveals the percentage of precipitation in a given period compared to the normal precipitation in the reference period (Falzoi et al., 2019). Due to its simplicity it is particularly appropriate for region-specific communication purposes and outreach. The PNI is computed for monthly or seasonal precipitation sums and is calculated by dividing the actual precipitation sum of a given period (P) by the climatological mean of that period in the reference period (P_{ref}) and multiplying it with 100 (Falzoi et al., 2019):

$$PNI = \frac{P}{P_{ref}} * 100 \tag{1}$$

PNI values below 80% are considered as drought conditions and are further categorized into four classes of increasing severity based on Falzoi et al. (2019) (see **Table 1**). Note that the classes cover uneven percentages of PNI values. The PNI is calculated individually for each pixel in the European domain of the CRCM5. A disadvantage of the PNI is that it implies by construction that the precipitation sums would follow a normal distribution, for which mean and median are the same. As this is not the case, misleading cases can occur when the median of a certain period and region is $\geq$ 20% lower than the mean. In this case, the median has a PNI of $\leq$ 80% and thus 50% of the distribution would appear as dry compared to the mean (Hayes, 2002; Yihdego et al., 2019). The monthly and seasonal precipitation sums are derived from hourly data. The drizzle is removed with a threshold of 1 mm per day (Kjellström et al., 2010).

2.3. Pre-industrial Reference

We use the pre-industrial climate state derived from pi-control runs of the CRCM5 as reference for the PNI calculation. This allows the comparison of future scenarios and the present climate to a reference climate state without any anthropogenic global warming. The relevance of such a pre-industrial reference increases with progressing climate change. Since 2017 global warming has reached approximately 1.0°C above pre-industrial levels (IPCC, 2018). Using a pre-industrial reference period ensures that this substantial amount of warming is considered. To further allow a comparison of future scenarios with present-day climate we analyze two 20-year periods: 2001–2020 as present-day and 2080–2099 in the far future. This provides

FIGURE 1 | (A) Averaged absolute precipitation values [mm] of the pi-control runs, which are used as reference for the PNI calculation. Values < 5 mm are plotted in white. **(B)** Absolute deviation [mm] of the 2001–2020 ensemble mean from the pre-industrial values in **(A)** with a general trend toward wetter winters and drier summers. The categorized colorbar shows values from the 1st to the 99th percentile.

50 members × 20 years per analysis period and therefore a large database to investigate extreme events. The difference between the climatological mean of the ensemble in a present-day climate vs. a pre-industrial climate state is illustrated in **Figure 1**. **Figure 1A** shows the averaged precipitation sums per season of 20-years of the 35 CRCM5 pi-control runs. This represents a pre-industrial climate state. High precipitation values are visible in mountain regions and on the west coasts of land masses (e.g., Norway, Great Britain, Galicia, the Balkans), especially in winter (DJF). The averaged precipitation sums of the present-day period 2001–2020 in the 50 members of the CRCM5-ensemble differ from that in a range of −45–74 mm (see **Figure 1B**). This difference reveals the climate change signal between present-day and pre-industrial and mainly shows a drying trend over large parts of Europe in summer and a trend toward wetter conditions in winter. The spatial patterns of this change go in line with near-future climate projections e.g., in the study of Kjellström et al. (2018), who compare a future 1.5°C warm world with the reference 1971–2000. **Supplementary Figures 1**, **2** show timeseries of the PNI values of 20-year slices in the ensemble in

comparison to the pre-industrial PNI values. For summer this reveals in most regions a trend toward higher PNI values in the early period of 1955–1974 compared to pre-industrial, before a continuous decreasing trend applies until the end of the century. This intermediate increase in summer PNI values might be due to a lower radiative forcing because of higher aerosol concentrations in 1955–1974 as shown in IPCC (2013).

2.4. Region-Specific Analysis

For region-specific analysis the European domain is divided into eight regions with different climate characteristics: the British Isles, Scandinavia, Mid-Europe, the Alps, Eastern Europe, France, the Mediterranean and the Iberian Peninsula (see **Figure 2**). The regions are based on the defined regions of the PRUDENCE project (Christensen and Christensen, 2007), but clipped to the smaller ClimEx-domain as previously done in von Trentini et al. (2019). If not stated otherwise, only land surface pixels within these regions are taken into account in order to avoid that drought conditions over landmasses are weighted out by precipitation over sea pixels for maritime regions. In order to
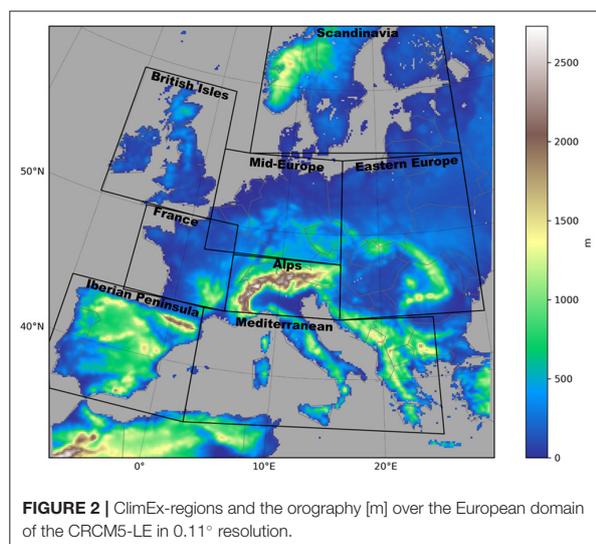
FIGURE 2 | ClimEx-regions and the orography [m] over the European domain of the CRCM5-LE in 0.11° resolution.

summarize the state of drought over the regions, the pixel-based PNI values are summarized into the category, whose PNI threshold is fulfilled by at least 60% of the pixel. For example, if 60% of the pixel in the Iberian Peninsula have PNI values ≤55%, then the region is classified to the severe drought category. The threshold of 60% is derived from Stefanon et al. (2012).

2.5. Signal Maps

Areas of particularly robust and strong changes in a quantity may be efficiently visualized using the signal maps approach by Pfeifer et al. (2015): these maps compile the information of ensembles by using the local ensemble mean change signal with respect to a reference period under the condition that a pre-defined percentage of ensemble members agrees on change sign and shows a robust signal, which is obtained by statistically testing the significance of local changes. Since the PNI by construction is an index of change, we adjust this method in the following way: rather than analyzing the difference of our quantity between a future and the reference period, we use the average drought categories of the PNI in the future period, if more than 66% of members agree on the sign, i.e., PNI < 100 or PNI > 100, and show a category equal to or stronger than the indicated one. Furthermore, we assume a particularly robust signal if more than 90% of members agree on a given category instead of using a statistical test. This map type allows to identify distinct regions of strong changes at a glance. The PNI used for this type of analysis is calculated seasonally (DJF, MAM, JJA, SON), i.e., seasonal precipitation sums with respect to seasonal climatological means, and for each pixel individually.
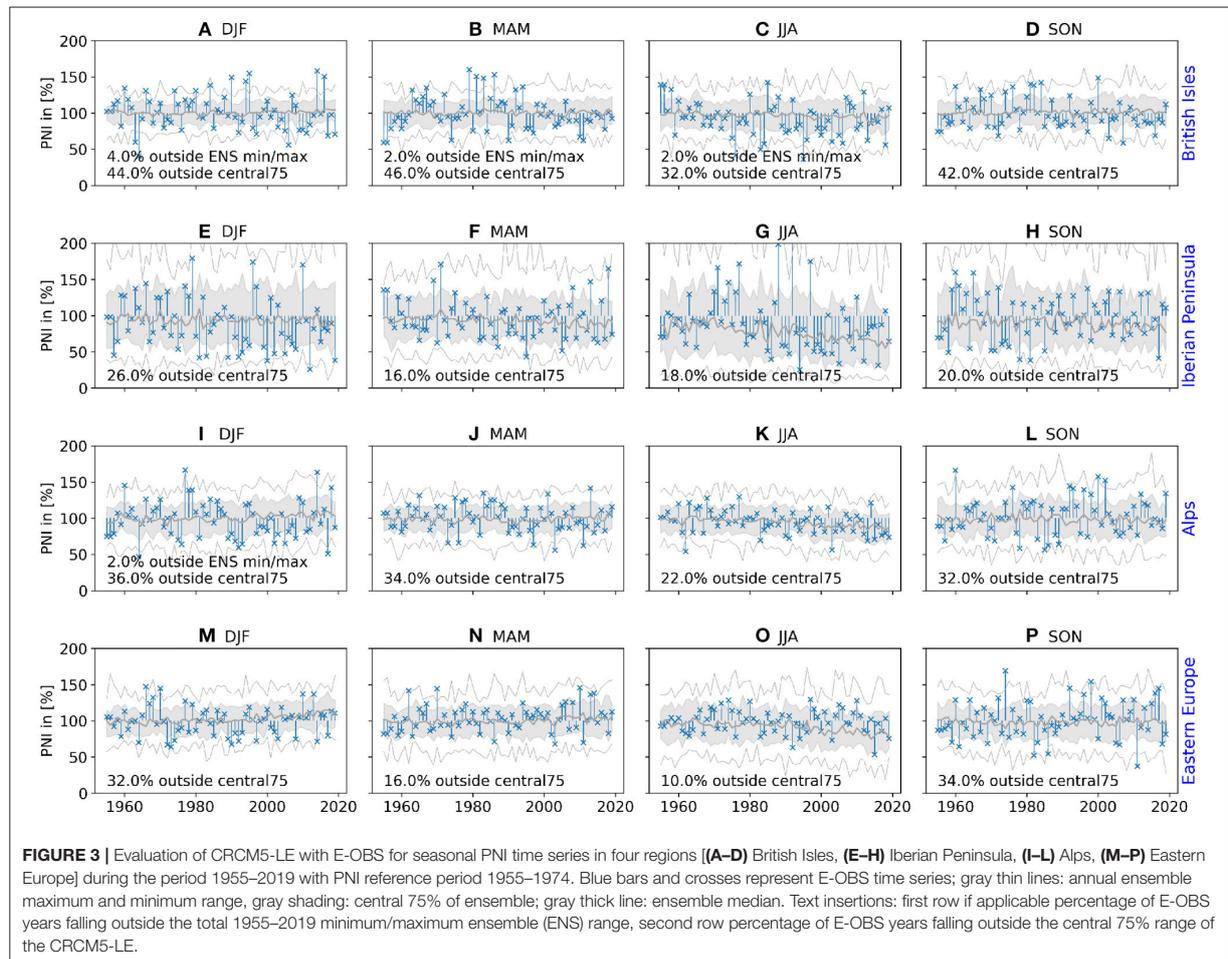
3. RESULTS

3.1. PNI Evaluation

We first assess PNI internal variability in the CRCM5-LE using an approach that is described in Suarez-Gutierrez et al. (2021) and

Wood et al. (2021): underestimation of ensemble PNI variability is identified if observed PNI values exceed the ensemble 12.5–87.5th percentile range (i.e., the central 75% of the distribution) in more than 80% of the analyzed years. Additionally, the amount of observed PNI values falling outside the ensemble range, thus being more extreme than covered by the ensemble, is calculated. To this end, seasonal PNI values are calculated for 1955–2019 with the E-OBS gridded dataset (daily precipitation sums of E-OBS variable RR, version 22.0e at 0.1° spatial resolution, regridding to CRCM5 grid followed by drizzle removal; Haylock et al., 2008; Cornes et al., 2018) and the CRCM5-LE with respect to 1955–1974. The comparison of CRCM5-LE and E-OBS climatological means reveals a wet bias in all seasons (strongest in winter) and regions (especially in mountainous regions, where winters are also subject to a warm bias) which is documented in **Supplementary Figure 3** (for a single member and another time period see Leduc et al., 2019). In most of the regions and seasons, however, E-OBS PNI variability is fully covered by the CRCM5-LE as can be seen in **Figure 3**. The central 75% of the ensemble encompass between 60% (e.g., British Isles MAM and SON) and 90% (e.g., Eastern Europe JJA) of all E-OBS years. For this figure, the seasonal PNI value was aggregated over all land pixels of the regions to derive time series showing trends and variability of the PNI. Single E-OBS years outside the ensemble range in a given year are to be expected, so the percentage numbers **Figures 3A-C,I** refer to years exceeding the total minimum/maximum range of the ensemble during 1955–2019. Only the British Isles show a slight underestimation of the variability according to the total ensemble range of CRCM5-LE. These time series also give a good indication of the amount of inter-annual variability that is to be expected within the depicted regions: whereas for the Alps and British Isles PNI variability in all seasons is low (PNI values of 50–150% during most seasons), the Iberian Peninsula is affected by large PNI variability, ranging from almost 0 to >200%. This is mostly tied to the low absolute precipitation values in the latter case. The four regions in **Figure 3** all show very weak trends (e.g., Iberian Peninsula and Alps in JJA), if any, which are masked by the large variability, during the period under consideration. Whilst PNI values of **Figure 3** are not directly comparable to subsequent **Figure 4** due to different reference periods and seasonal aggregation, a division into a period of little trends (before the 2001–2020 period) and a period with strong trends (after the 2001–2020 period) becomes apparent. A corresponding figure with the four missing regions can be found in the **Supplementary Figure 4**.

3.2. Seasonality: Seasonal Hot Spots and Drought Categories

Inspired by the warming stripes by Hawkins (2018), **Figure 4** provides an overview of regionally averaged ensemble median monthly PNI values. Each ring of these *drying stripes* represents a single year, starting at the center of the circles in the year 2000. Green colors correspond to a PNI>100%, i.e., wetter conditions than normal, whereas brown colors symbolize a drying trend (PNI < 100%). Using the ensemble median allows to picture the long-term trend; internal variability, e.g., wetter

FIGURE 3 | Evaluation of CRCM5-LE with E-OBS for seasonal PNI time series in four regions [**(A–D)** British Isles, **(E–H)** Iberian Peninsula, **(I–L)** Alps, **(M–P)** Eastern Europe] during the period 1955–2019 with PNI reference period 1955–1974. Blue bars and crosses represent E-OBS time series; gray thin lines: annual ensemble maximum and minimum range, gray shading: central 75% of ensemble; gray thick line: ensemble median. Text insertions: first row if applicable percentage of E-OBS years falling outside the total 1955–2019 minimum/maximum ensemble (ENS) range, second row percentage of E-OBS years falling outside the central 75% range of the CRCM5-LE.

years during drier conditions, is nevertheless visible during the shown period of 2000–2099. This figure also highlights months with particularly strong PNI reductions (e.g., summer months in **Figures 4B,C,F–G**) and PNI increases (e.g., winter months in **Figures 4F,H**). Note that the PNI trends in the later years tend to be stronger than in the earlier years. This is especially visible in **Figures 4B,C,F–G** for summer months and **Figure 4H** for winter months. Most regions experience a precipitation decrease during summer months and an increase during the winter months of differing intensity (weak in the British Isles region, strong in Eastern Europe, with a general tendency for a stronger divergence between summer and winter toward the eastern regions, i.e., with increasing continentality). The largest exceptions can be found in the Iberian Peninsula (general decrease) and in Scandinavia (general increase).

In the following, analyses focus on distinct drought categories (see **Table 1**) to better assess drought severity. The **Figures 5, 6** show spiderplots for four regions each, which illustrate the seasonality of each drought category (slight, moderate, severe,

and extreme). Two spidernets are plotted for the two periods: present-day from 2001 to 2020 (blue) and the far future 2080–2099 (orange). The line of the spidernet illustrates the percentage of years in the ensemble and period that belongs to this drought category in the particular month. One hundred percent represent 1,000 years (50 members × 20 years). A blue value of 10% in August in the upper left spiderplot for example means that under the present-day climate in the British Isles 10% of all months August in the ensemble belong to the category of a slight drought. Keep in mind, that the drought categories are defined based on pre-industrial precipitation sums. For the British Isles and Scandinavia, the percentages of droughts are low (<20%) even in summer in the far future (see **Figure 5**). In the British Isles moderate and severe droughts are becoming more frequent by up to 10% between far future and present-day. In Mid-Europe the occurrence probability of an extreme drought strongly increases in the far future amounting to 25% in August. In the Alps (see **Figure 5**) and in Eastern Europe (see **Figure 6**) severe as well as extreme droughts have higher probabilities in the far future

FIGURE 4 | Ensemble medians of monthly PNI values from 2000 to 2099 averaged over the ClimEx regions [(**A–H**) *drying stripes*]. The blue axis corresponds to a time axis, starting at 2000 in the center and extending to 2099 at the outer limit of the discs. Monthly labels are indicated clockwise.
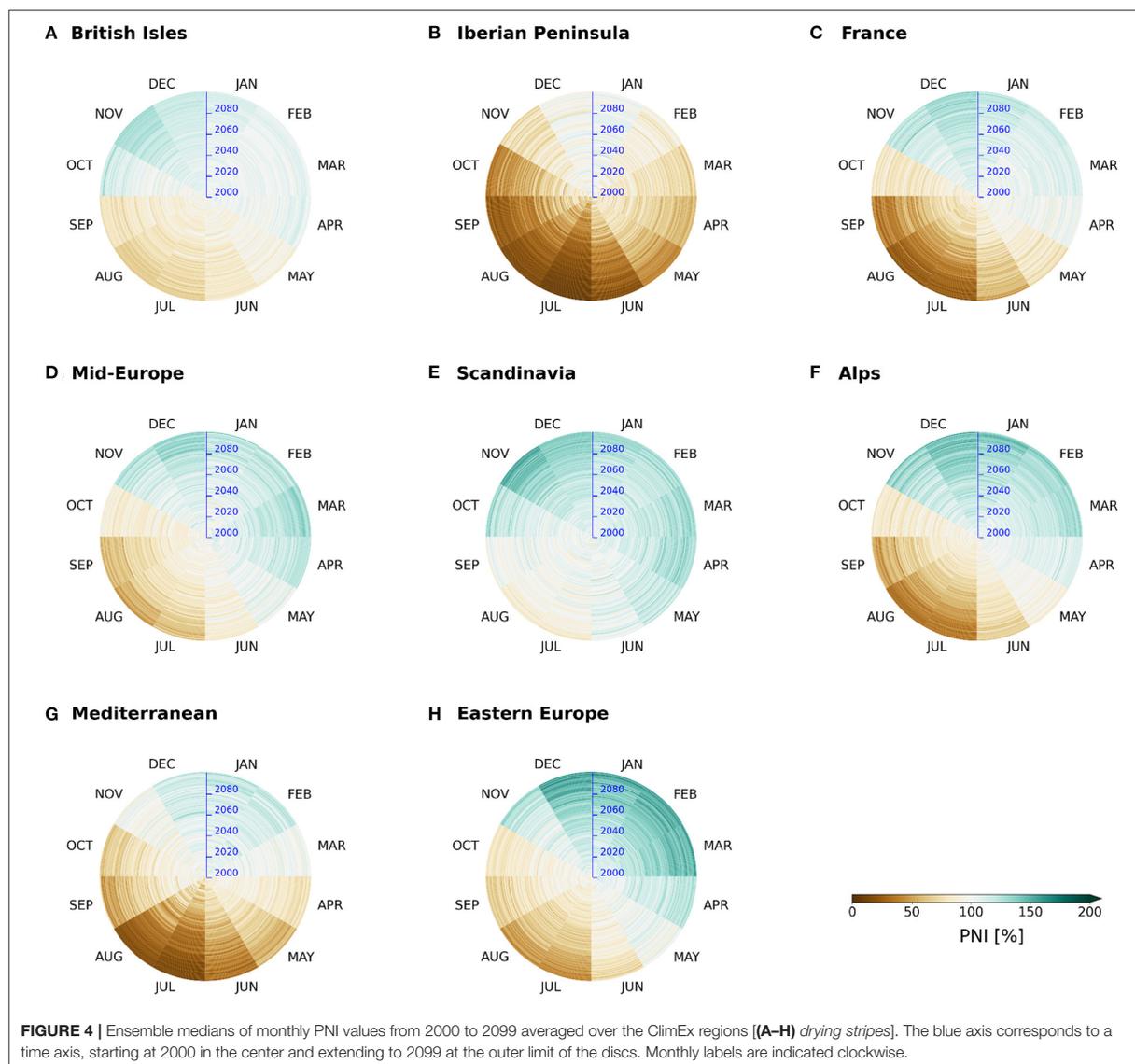
with values around 20% (severe), respectively 40% (extreme) in July and August. In France (see **Figure 6**), on the contrary, there is hardly a growth in the number of severe droughts in summer and even a decrease for the moderate drought category. This is due to the strong increase of extreme droughts, because summer droughts in the far future seem to fall in the category of highest severity most of the time. The percentages range up to >60% in July and August. In spring and fall, however, moderate and severe droughts are increasing, although only slightly. This effect is even stronger for the Mediterranean region and the Iberian Peninsula. In the Mediterranean, the percentages of extreme droughts in the far future reach around 80% in July and August. The percentage of severe droughts is around 15% in May, June, September and October. In the Iberian Peninsula

the percentage of extreme droughts in the far future is the highest of all regions, reaching 96% in July. In August the value is 88 and 76% in June and September. However, it is important to keep in mind that the absolute precipitation in the Iberian Peninsula in summer is very low and contributes only to 2% of the yearly precipitation in July and August (see **Figure 8B**). Considering this, the high value in June is even more remarkable as it contributes about 6% to the annual precipitation in present-days. Overall, the results regarding the PNI values during drought events reveal an increase in the intensity of summer droughts in the far future in all regions. For the winter months, a considerable decrease in slight and moderate droughts is shown for Scandinavia and Mid-Europe and less intense for several further regions e.g., the Alps, Eastern
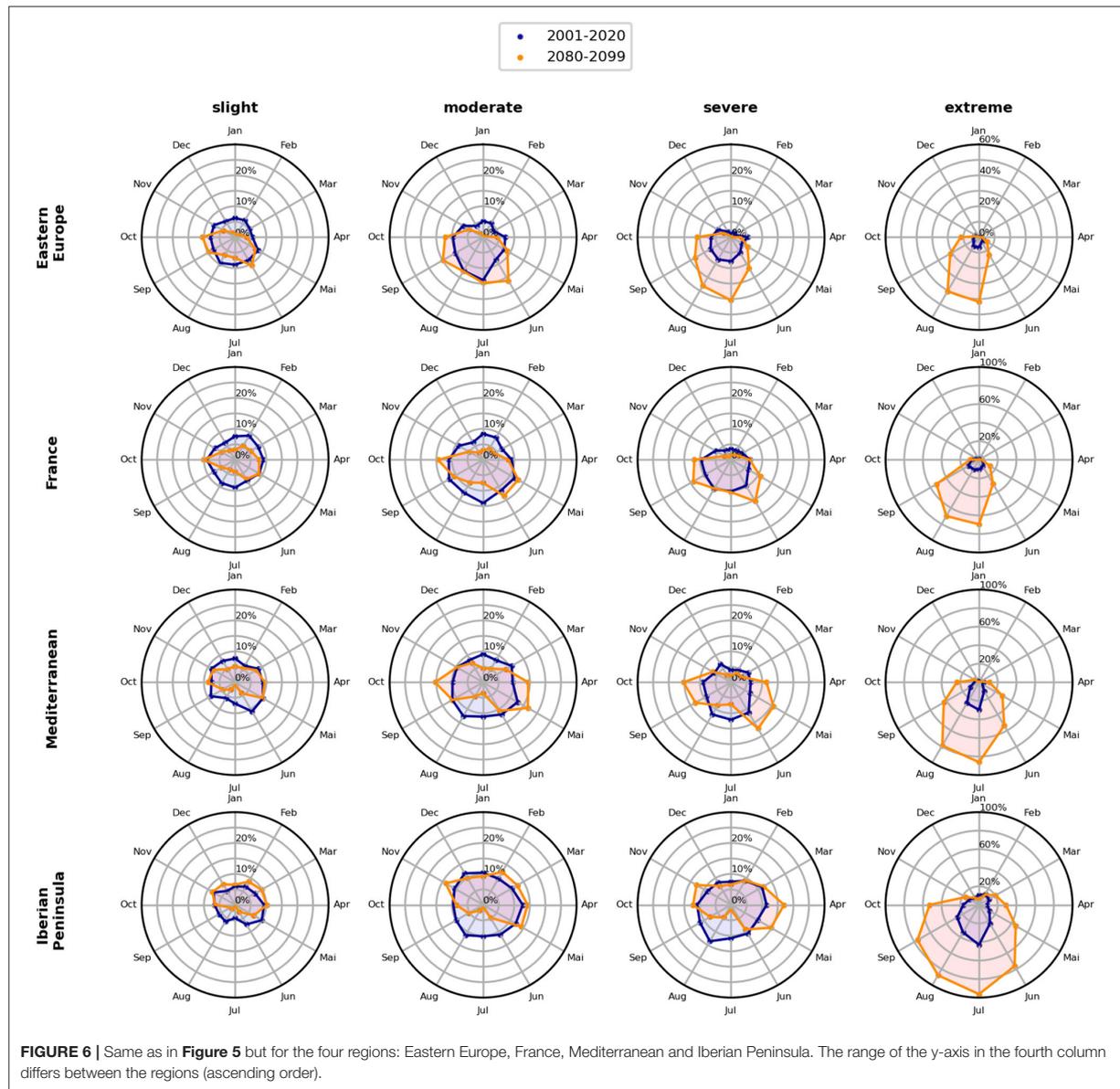
FIGURE 5 | Spiderplots showing the percentage of months in the CRCM5-LE that belong to the specific drought category (columns: slight, moderate, severe, extreme) for the two periods: present-day (2001–2020; blue) and far future (2080–2099; orange) for each of the four regions: British Isles, Scandinavia, Mid-Europe and Alps. One hundred percent stands for 1,000 values (20 years × 50 members). The range of the y-axis in the fourth column differs between the regions (ascending order).

Europe, France and the Mediterranean between the far future and present-day.

3.3. Seasonality of the Spatial Distribution

The seasonally varying spatial distribution of drought categories is visualized in **Figure 7**. These maps are adjusted versions of signal maps described by Pfeifer et al. (2015). Compared to the pre-industrial reference, most parts of Europe show a reduced PNI, with the largest reduction occurring in JJA in

Southern Europe. For example, the mean precipitation state during 2080–2099 over the Iberian Peninsula is robustly classified as an extreme drought compared to the pre-industrial reference. Central Europe also shows slight to moderate drought conditions with high agreement among all ensemble members, equaling 55–80% of the reference summer precipitation (see hatching in **Figure 7**). During the winter months, the largest part of Europe is classified as non drought, apart from single regions in Northern Africa, parts of the Mediterranean coast and a small region in

FIGURE 6 | Same as in **Figure 5** but for the four regions: Eastern Europe, France, Mediterranean and Iberian Peninsula. The range of the y-axis in the fourth column differs between the regions (ascending order).

Norway with a weak PNI reduction. Spring sees a wide extension of slight to severe droughts in the southwestern regions of the domain, whereas additionally in SON mountainous regions are affected by a weak PNI reduction. Since colored and hatched regions show generally a strong correspondence (apart from JJA eastern regions and SON), this seasonal derivation of local hot spots may be seen as highly robust. A similar map showing the very weak changes between the pre-industrial period and 2001–2020 is provided in the **Supplementary Figure 5**. We address the influence of low absolute precipitation sums, especially in the southern regions, in the subsequent sections.

3.4. Drought Intensities

Figure 8 shows four selected regions: the Alps with its high altitude climate, the Iberian Peninsula with the most pronounced trend toward extreme droughts, Eastern Europe with a continental climate and the British Isles with strong maritime influences. The remaining regions are shown in the **Supplementary Figure 6**. The analysis in **Figure 8** looks in more detail at the drought months (at least 60% of the regions' pixel have PNI values <80%). For all drought months, the mean PNI value of the drought pixels (PNI value <80%) is calculated. This gives insights about the severity of the drought

FIGURE 7 | Seasonal signal maps for the period 2080–2099 with respect to the pre-industrial period. Colors show where at least 66% of all members agree on the indicated PNI category (or higher severity), hatching indicates an agreement of more than 90% of members.
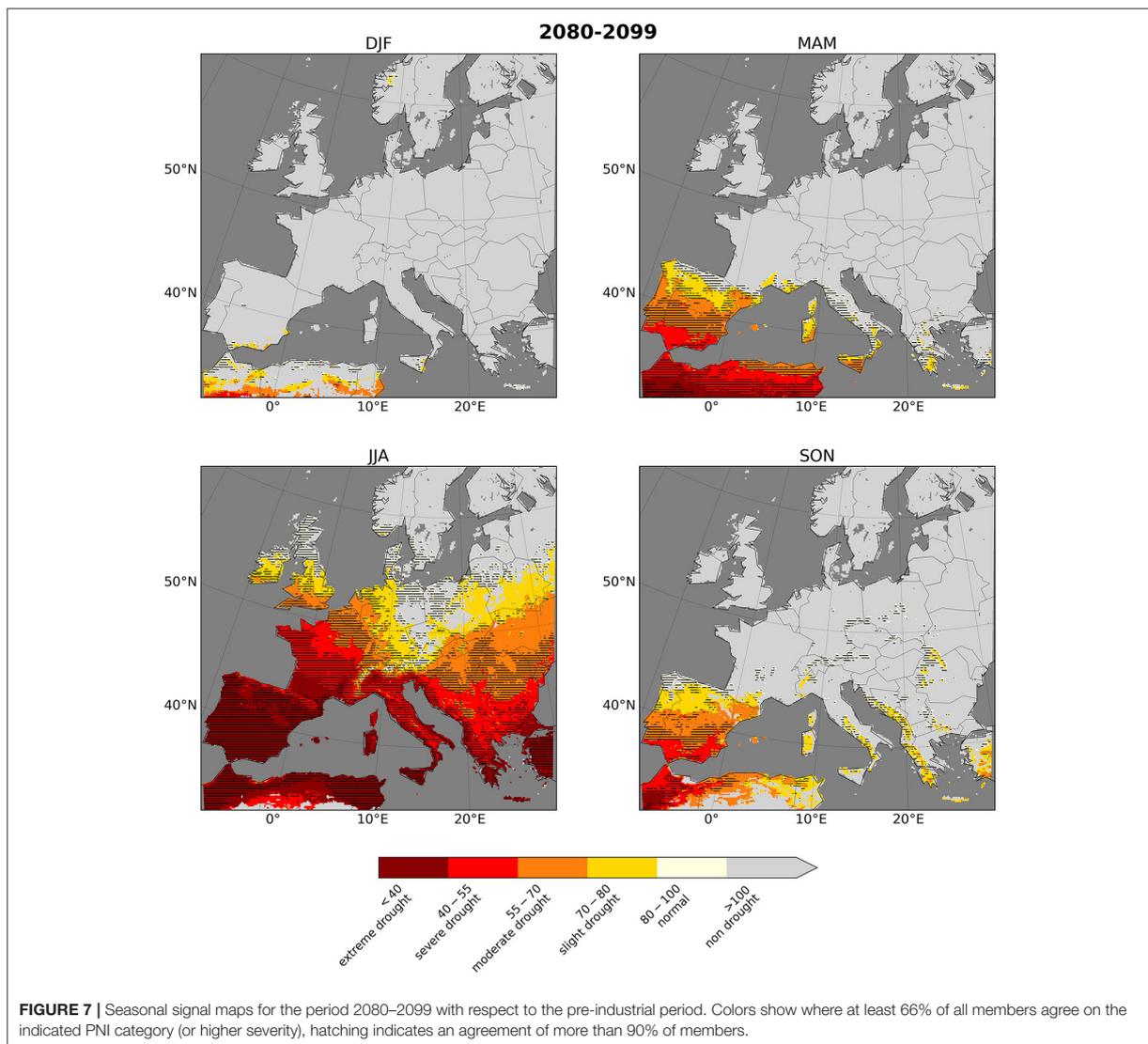
event, directly expressed as percentage of precipitation relative to the pre-industrial normal condition. The boxplots illustrate the distribution of the mean PNI values of all drought months in the specific period in the CRCM5-ensemble. The highest theoretically possible sample size for a boxplot is 1,000 (20 years × 50 members). The barplots below the boxplots delineate the percentage that each month contributes to the yearly precipitation sum in 2001–2020. This information is crucial for months that have low monthly precipitation sums and a small share of the yearly precipitation. A low PNI value, which represents a major relative change, is in such cases triggered by only a minor reduction in the monthly precipitation sum. The actual sample size closest to the theoretical maximum is 940 and occurs in the far future over the Iberian Peninsula in July (see

Figure 8B). This means that in the far future over the Iberian Peninsula only 60 out of 1,000 years in the ensemble have a July that does not show a drought condition. In other words, 94% of all future months July have a total precipitation <80% of what used to be normal in pre-industrial times. The median PNI value of those drought events amounts to 7.6%, which means that 50% of the distribution have even lower PNI values. However, the total precipitation sums of the summer months in the Iberian Peninsula are low and contribute in the case of July only 2% to the yearly precipitation sum (see **Figure 8B**). All four regions show the tendency of lower PNI values during droughts in summer than in winter. This is the case in the present-day period and the far future and particularly strong in the Iberian Peninsula. In the far future this effect amplifies in all four regions. The internal
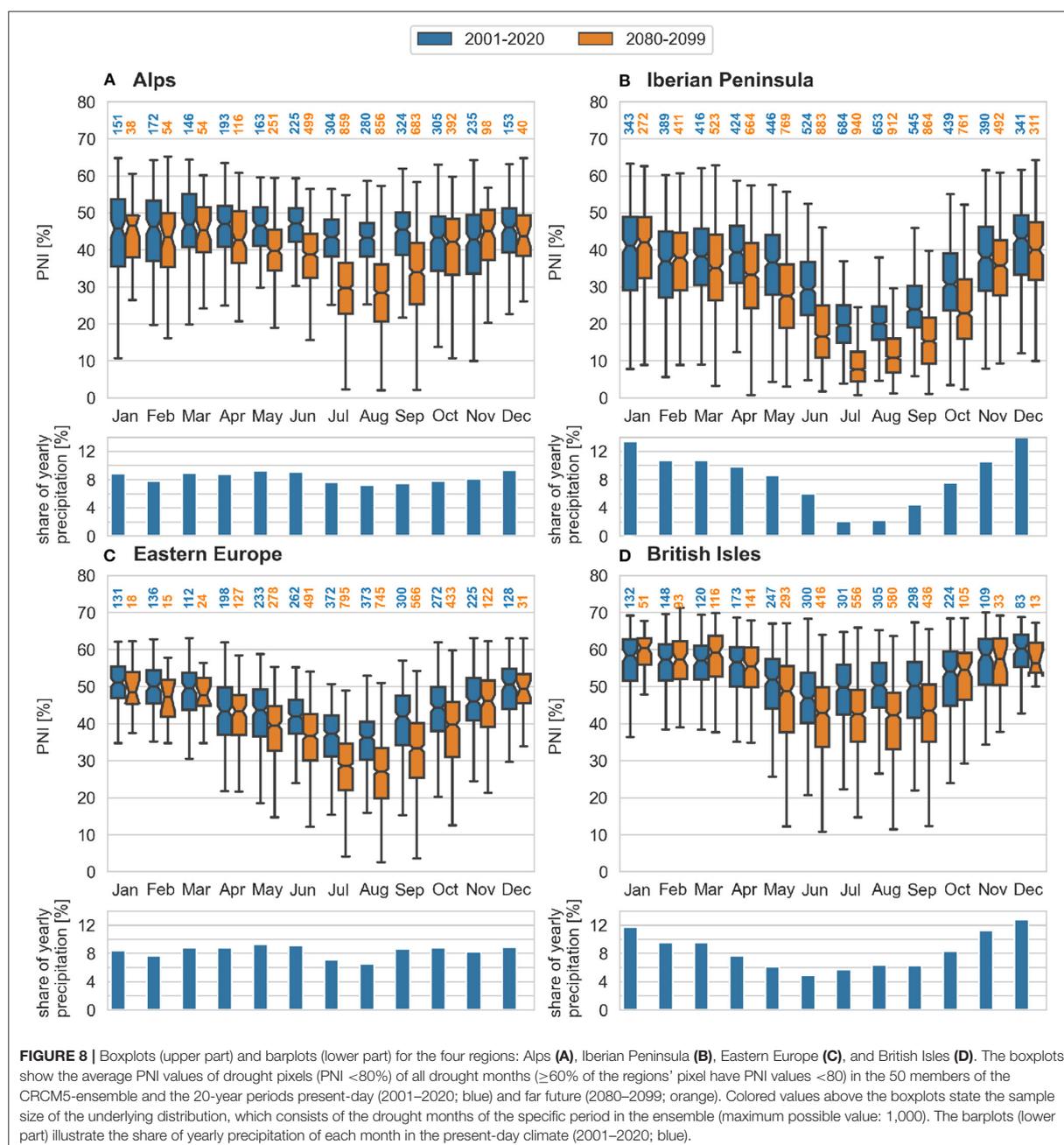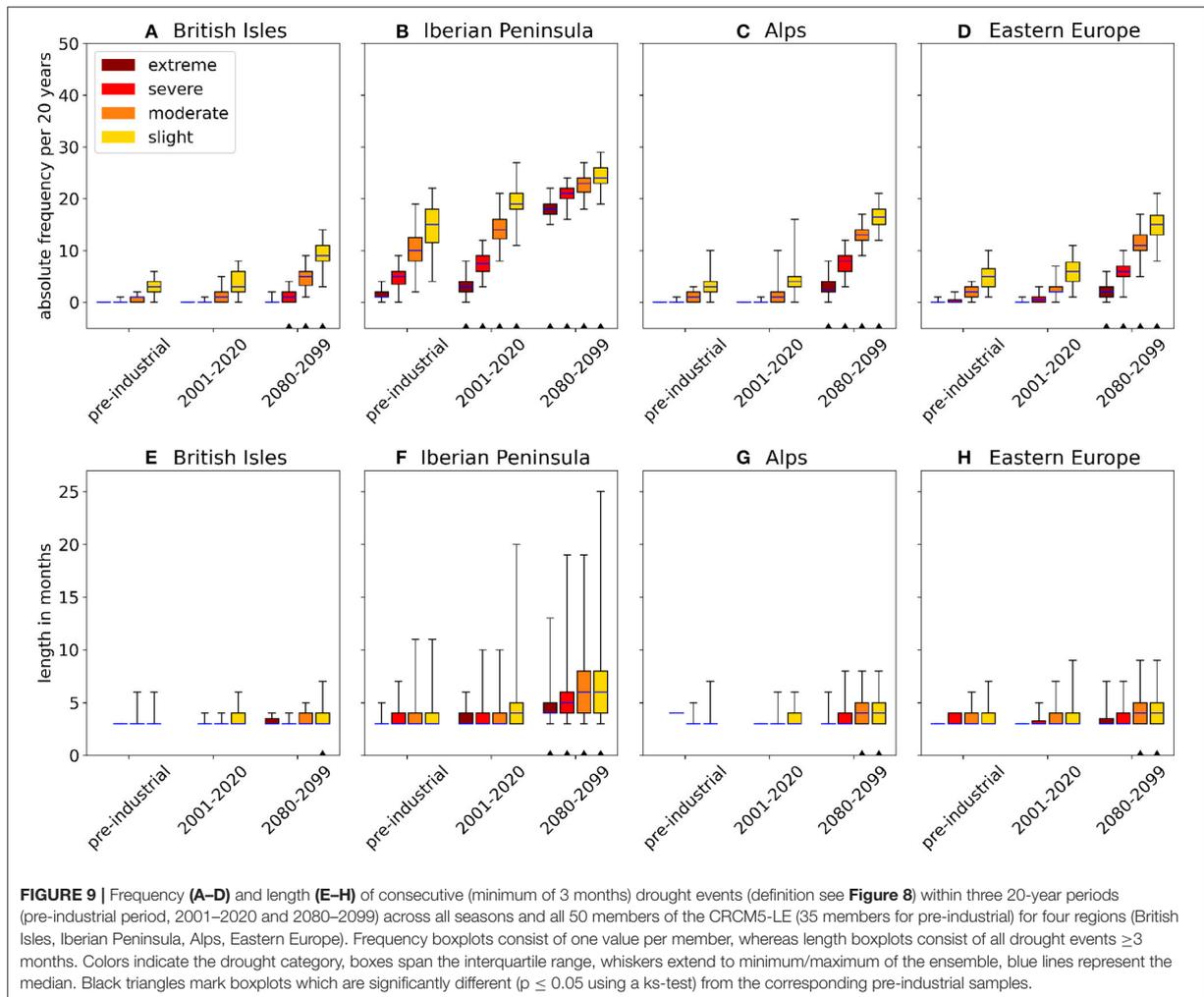
FIGURE 8 | Boxplots (upper part) and barplots (lower part) for the four regions: Alps **(A)**, Iberian Peninsula **(B)**, Eastern Europe **(C)**, and British Isles **(D)**. The boxplots show the average PNI values of drought pixels (PNI <80%) of all drought months (≥60% of the regions' pixel have PNI values <80) in the 50 members of the CRCM5-ensemble and the 20-year periods present-day (2001–2020; blue) and far future (2080–2099; orange). Colored values above the boxplots state the sample size of the underlying distribution, which consists of the drought months of the specific period in the ensemble (maximum possible value: 1,000). The barplots (lower part) illustrate the share of yearly precipitation of each month in the present-day climate (2001–2020; blue).

variability of PNI values of droughts is large, which is shown in a large boxplot spread. For example in the far future in the British Isles (see **Figure 8D**) the mean PNI value of drought pixels of a May under drought condition ranges from almost 70% to about 10 %. In the Alps the decrease of the median of the far-future PNI values compared to present-day is almost 15% in July and August and thus particularly strong. The spread is large and covers PNI

values from 59 to 2%. Such an extremely low PNI value denotes a tremendous precipitation deficit as July and August hold about 7.6 and 7.3% of the yearly share.

3.5. Drought Length and Frequency

Since the harmful potential of droughts is connected to their duration, we next investigate the length of events due

FIGURE 9 | Frequency **(A–D)** and length **(E–H)** of consecutive (minimum of 3 months) drought events (definition see **Figure 8**) within three 20-year periods (pre-industrial period, 2001–2020 and 2080–2099) across all seasons and all 50 members of the CRCM5-LE (35 members for pre-industrial) for four regions (British Isles, Iberian Peninsula, Alps, Eastern Europe). Frequency boxplots consist of one value per member, whereas length boxplots consist of all drought events ≥3 months. Colors indicate the drought category, boxes span the interquartile range, whiskers extend to minimum/maximum of the ensemble, blue lines represent the median. Black triangles mark boxplots which are significantly different (p ≤ 0.05 using a ks-test) from the corresponding pre-industrial samples.

to climate change. Therefore, only drought events with a minimum length of three consecutive months are considered. We choose this minimum length since it equals the length of the previously analyzed seasons. **Figure 9** presents the frequency of occurrences (**Figures 9A–D**) and the length (**Figures 9E–H**) of these continuous drought events for four regions. Insights on the remaining four regions are provided in the **Supplementary Figure 7**. The categories refer to the least severe category within a continuous drought event, e.g., a continuous severe drought event may contain drought months of the category extreme and severe, but not moderate. This is why frequency and length are generally higher for the low-severity categories. In general, all regions experience an increase in frequency and length of these continuous drought events across all categories. This increase is considerably stronger between 2080–2099 and 2001–2020 than between 2001 and 2020 and the pre-industrial period. This finding is supported

by the fact that most frequency boxplots in the far future are significantly different from the pre-industrial boxplots, whereas the present-day boxplots are not (black triangles in **Figure 9**). The simultaneous increase of event frequency and length suggests, e.g., over the Iberian Peninsula (**Figures 9B,F**), that shorter events do not simply merge to longer events. The median shows 2–5 times more frequent slight/moderate drought events in the far future horizon, which results on average in 1 event in 2 years over the British Isles, yearly events in the Alps and Eastern Europe and more than 1 event per year over the Iberian Peninsula. Due to internal variability, however, the frequency may even be higher (or lower). The Iberian Peninsula also experiences yearly events of extreme droughts longer than 3 months which most likely occur during the summer months (see **Figure 7**). In Eastern Europe and the Alps, extreme droughts are projected to occur in 2–4 years out of 20 (**Figures 9C,D**), whereas during the pre-industrial

TABLE 2 | Consecutive seasons of dry summers (winters) and wet winters (summers) in eight regions for pre-industrial (PI), present day (PRES, 2001–2020) and far future (FF, 2080–2099) horizons.

Region	JJA (PI)	JJA (PRES)	JJA (FF)	DJF (PI)	DJF (PRES)	DJF (FF)
	P(wet DJF | drought JJA) in [%]			P(wet JJA | drought DJF) in [%]		
British isles	0.30	12.65	27.60	25.93	9.10	0.0
Iberian peninsula	27.31	33.20	27.01	27.36	15.69	0.47
France	15.15	19.35	52.88	26.97	14.13	0.0
Mid-Europe	18.57	25.16	59.13	16.04	14.39	0.0
Scandinavia	12.5	30.85	67.63	26.23	12.96	0.0
Alps	25.49	28.77	76.99	15.79	8.51	0.0
Mediterranean	20.31	29.74	48.48	24.32	18.26	0.0
Eastern Europe	17.43	26.11	91.29	12.77	9.30	0.0
	No. of droughts			No. of droughts		
British isles	67	166	529	27	33	7
Iberian peninsula	238	509	970	201	255	213
France	132	310	904	89	92	10
Mid-Europe	70	155	597	106	132	15
Scandinavia	56	94	173	61	54	3
Alps	51	146	865	76	47	3
Mediterranean	192	390	951	111	115	33
Eastern Europe	109	203	689	47	43	1

The upper half of the table provides conditional probabilities of having a wet DJF (JJA) given a drought in JJA (DJF). These values are in [%]. JJA header in this table refer to summer-winter sequences, DJF to winter-summer. The lower half of the table provides the absolute number of droughts in JJA and DJF across all members and years per period (maximum value possible is 1,000). Drought JJA and DJF refer to seasons where at least 60% of the land pixels within a region show PNI < 80%. Subsequent seasons' PNI values are obtained by averaging the same pixels and classified as drought or wet, if this average is PNI< 80% and PNI> 120%, respectively.

and present period they only occurred occasionally. Drought length reaches more than 6 months in all regions except the British Isles for all drought categories with an absolute maximum over the Iberian Peninsula with up to 1 year of continuous extreme drought and around 2 years of continuous drought with at least slight severity. In this region, boxplots become very similar during the far future period, indicating that especially extreme droughts increase in frequency and length. Ensemble ranges show that on the one hand the projected frequencies are quite robust (short boxes and whiskers in the 2080–2099 horizon for most cases in **Figures 9A–D**). On the other hand, the probability of longer drought event duration increases with wider distributions (larger boxes during 2080–2099 compared to pre-industrial and 2001–2020 in **Figures 9E–H**). Changes in these properties from pre-industrial to 2080–2099 are reflected by the black triangles, indicating the rejection of a null-hypothesis on same distributions ($p \leq 0.05$, using a Kolmogorov-Smirnov test).

3.6. Compensation of Drought Seasons

As **Figure 9** indicates, e.g., for the Iberian Peninsula region, some drought events may extend over 1–2 years, thereby affecting usually drier and wetter seasons. It is highly relevant to ask whether drought events of one season may be—in principle—compensated by subsequent wetter seasons. This is not meant to be a quantitative water budget analysis, but rather an estimate of the general tendency since e.g., evapotranspiration increases due to rising temperature may by larger than potential

projected precipitation gains (see e.g., Spinoni et al., 2020). Specifically, what is the probability of having a wet winter (summer), given a precedent drought summer (winter)? Winters and summers were chosen since these seasons tend to show the highest share of annual precipitation and the strongest PNI reductions, respectively (**Figure 8**). **Table 2** provides the corresponding values for the pre-industrial, present day and far future periods. A table containing a similar analysis with respect to fall and spring is provided in the **Supplementary Table 1**. A season is defined as drought if at least 60% of the land pixels within a region show PNI < 80%. The PNI for the next but one season is spatially averaged over the same pixels. This subsequent season is considered as drought as well if PNI < 80%. Since dry conditions of the PNI are defined as more than 20% less precipitation than normal conditions (**Table 1**), we also decide to use a threshold of more than 20% higher than normal conditions for wet seasons, i.e., PNI > 120%. We first consider the probability of having a drought in a given summer (or winter): What is the probability that the following winter (summer) will be wet (upper half of **Table 2**)? The first notion is a general increase in the case of wet winters following dry summers with a maximum value of 91% during 2080–2099 in Eastern Europe and a minimum of 27% over the Iberian Peninsula and the British Isles. The Iberian Peninsula is the only region, in which the percentage of summer droughts followed by wet winters is not increasing in the far future, but is slightly lower than in the present-day climate. In all other

regions the trend points clearly toward an increasing percentage of summer droughts that are compensated by higher than normal precipitation in winter. This mirrors the general tendency for wetter winters and drier summers with a larger seasonal gradient in the more southeastern regions as mentioned previously (e.g., **Figure 4**). Wet summers following a winter drought, occurring at a similar probability to the inverse combination during pre-industrial times for most regions, disappear for all but one region in the far future horizon. So if a given winter is classified as a drought, the next summer will not show wet conditions. The lower half of **Table 2** provides the absolute numbers of drought events within the ensemble per period to put the conditional probabilities in perspective. JJA droughts become more frequent in all regions whereas the projected increase between the 80 years between far future and present day is remarkably stronger than between present day and pre-industrial times. At the same time, DJF droughts show a general decrease between pre-industrial times and far future with a peak during the present time in some regions. With more summer droughts and less winter droughts, the decrease in wet summers following a winter drought seems reasonable. The increase in wet winters following a summer drought points toward an increasing number of wet winters at a comparable or even stronger magnitude as dry summers. So on the one hand, although in far future, a winter drought will most likely not be compensated by subsequent wet summers, the case of winter droughts becomes less frequent itself. On the other hand, inter-seasonal variability regarding precipitation tends to increase.

4. DISCUSSION

4.1. Seasonal Trends

The most pronounced seasonal trends identified in this study are an increase in summer drought frequency and intensity and a decreasing number of winter droughts in several regions of various climates. This goes in line with a general tendency toward larger differences between winter and summer precipitation, with precipitation increases during winter and precipitation decreases during summer in the far future horizon. The identified decrease of winter droughts in several regions are in accordance with Spinoni et al. (2018) apart from the regions France and the Mediterranean, where these authors do not find a drought decrease in winter.

A deficit in precipitation is rated most critical for months that have a high or above-average share of the regions' annual precipitation. This is for example the case for the summer months in the Alps or spring and winter in the Iberian Peninsula. In this regard, the question if a drought season is compensated by a wetter than normal subsequent season is of high relevance, especially if the season of the drought shows a high share of annual precipitation under normal circumstances. For example in a region like the Iberian Peninsula, high absolute winter precipitation sums bear the potential to compensate precipitation deficits of dry summers by a subsequent wet winter. Wet summers, however, correspond to absolute precipitation amounts lower than the winters, such that they likely lack the potential to compensate winter droughts. In a region like the

Alps, where precipitation is less concentrated in a single season, a compensation depends more strongly on the magnitude of precipitation deficit and surplus.

Wet seasons experiencing precipitation increases and dry seasons experiencing precipitation losses result in a higher inter-seasonal variability of precipitation in the far future. There is one caveat to this particular conclusion because of the wet bias in the CRCM5-LE, which is most pronounced during winter and in the mountainous regions whereas it is small during summer. However, as our evaluation shows, the ensemble reflects well observed PNI variability and minimum/maximum values such that the projected changes are considered robust and reliable.

4.2. Regional Trends

Most regions share the seasonally divergent PNI trends, with differences regarding the annual distribution. Exceptions are the Iberian Peninsula with a drying tendency throughout the year and Scandinavia with precipitation increase in all seasons but July and August at the very end of the twenty-first century. This is generally in line with the findings in Spinoni et al. (2018). Additionally, we find longer drought periods in most regions by the end of the twenty-first century. While these tend to show around 6 months in some regions, in line with the aforementioned seasonal trends of more wet winters following drought summers, hyperannual droughts are also projected to occur. This is partly tied to the overall increase of drought occurrences with respect to our reference. Long lasting meteorological droughts, however, hold the potential to turn into e.g., hydrological droughts, which have strong impacts on water supply (Van Lanen et al., 2016). Lorenzo-Lacruz et al. (2010) for example state an increasing correlation between drought indices and hydrological variables in central Spain for longer drought periods. Nevertheless, for impact assessment other indices, accounting e.g., for the effects of temperature or evapotranspiration, may be better suited, e.g., in conjunction with the use of impact models like hydrological models (Dobson et al., 2020). Lorenzo-Lacruz et al. (2010) find that the SPEI shows more severe droughts within their Spanish study area than the precipitation-only based SPI, suggesting that without temperature the (hydrological) impact of droughts may be underestimated. In line with this, Vicente-Serrano et al. (2014) show for the entire Iberian Peninsula that an increase of temperatures and therefore of evaporative demand in the atmosphere already resulted in higher (SPEI) drought intensities since 1961.

In addition to drought length, the spatial extent of droughts is of importance. Very localized droughts of high intensity may have weaker impacts than moderate droughts of larger spatial extent (Dobson et al., 2020). We address this point by using the spatial constraints (of having droughts in at least 60% of a given region) to ensure that we investigate events of large extent.

4.3. Hot Spots

The four regions selected for a detailed analysis in this study show diverging future drought trends across various climates in Europe: whereas the Alps and Eastern Europe, corresponding to high mountain and continental climate conditions, respectively,

are marked by increasing differences between winter and summer precipitation, the Iberian Peninsula, located in a Mediterranean climate, sees an overall drying in all seasons and the British Isles, subject to maritime climate conditions, exhibit only slight trends. Taking all regions considered in this study into account, particularly strong changes and drought increases are found in the Alps, the Iberian Peninsula, but also France and the Mediterranean. These regions can therefore be considered as drought hot spots. All four regions show a strong decrease of summer PNI values (see **Figure 4**) and a severe increase in the number of extreme droughts in summer (see **Figures 5, 6**). The Iberian Peninsula, the Mediterranean and France are furthermore highlighted as hot spots in **Figure 7** with a high agreement among the members of the CRCM5-LE. A characteristic of the Alps is the high contribution of the summer months to the yearly precipitation sum (see **Figure 8**), which makes summer precipitation deficits even more critical. The frequency and lengths of consecutive drought events (of at least 3 months) is increasing in all four regions (see **Figure 9** and **Supplementary Figure 7**). The Iberian Peninsula, in contrast to the other regions, is additionally affected by a decreasing percentage of summer droughts compensated by wet winters (see **Table 2**) and can, inter alia, therefore be particularly postulated as a hot spot. In general, the drying trend we identified in the Iberian Peninsula goes in line with Spinoni et al. (2018), who find similarly frequent droughts using a different drought metric in the EURO-CORDEX multi-model ensemble.

4.4. Index Choice

The advantage of the PNI lies in its intuitive meaning as it directly represents the percentage of precipitation in a certain month or season compared to the long-term mean. This simplicity in turn also involves the main disadvantage of the PNI. Because no normalization is undertaken in the calculation of this index, the PNI implies that the precipitation sums follow a normal distribution (Hayes, 2002; Yihdego et al., 2019). Still, our findings on climate signals on droughts are in line with the study by Spinoni et al. (2018), who employ a normalized drought index.

In analyzing meteorological droughts based on precipitation deficit only, we investigate the framing conditions for further impact assessment, which may employ drought definitions that are specifically tailored toward a certain impact. Using a relative measure for estimating meteorological droughts reduces the impact of model bias because we also refer to model climatological references. Drought impact assessment on the other hand requires well-chosen model bias adjustment (Ruffault et al., 2014).

4.5. Perspectives Due to Pre-industrial Reference

The relative measure also allows to relate future changes with familiar conditions to which most human systems and behaviors are adapted. This we achieved by comparing a present day period with a far future period. However, we showed that for most drought characteristics this present day period is already affected by climate change. As previously discussed, this is a good reason for using a pre-industrial reference period. By doing this, we have two perspectives from the present day period: we show recent,

comparably small climate changes since the pre-industrial period that are already inherent nowadays, and we can relate them with what is projected for a far future horizon. These changes regarding drought frequency and length are non-significant for the present day period, in line with findings as in Vicente-Serrano et al. (2021), where no clear trend during the historical period is found. Internal variability of drought characteristics during this period is thus too large to identify clear trends, even within a SMILE.

4.6. Internal Variability of Drought Trends

The ensemble provides two key messages: first, we see robust increases in overall drought numbers and severity over Europe using an extreme RCP8.5 emissions scenario. Second, we note a higher inter-seasonal variability of wet seasons becoming wetter, dry seasons becoming drier on the one hand, and more long lasting drought events on the other hand. The latter tend to be rare events, e.g., ≥ 12 months in Iberian Peninsula results in 1, 2, 7, and 32 events per thousand years for categories extreme, severe, moderate and slight, respectively. We even find one, albeit slight, event in thousand years of more than 2 years in length (**Figure 9**). The SMILE provides the means to detect these as well. The spread of drought intensities throughout the CRCM5-LE is also high and stresses the high internal variability of droughts. We also note that results concerning the most extreme drought classes tend to be more robust than less severe categories. This refers to the projected numbers of droughts as well as to the spatial location of most extreme PNI changes.

5. CONCLUSION

To summarize, we find an overall increase in drought numbers, and a high internal variability of drought intensities both in the present day period and the projected far future across various European climate regions. Drought lengths are projected to show higher values and variability in the future. The changes in length and frequency are significant in most regions only in the far future horizon. Additionally, summer droughts are projected to become more intense with more droughts reaching the category "extreme," while winter droughts become less frequent in several regions. The percentage of summer droughts followed by wet winters increases in all regions, except for the Iberian Peninsula. Our results suggest the consideration of four pronounced hot spots due to especially strong drought trends: the Alps, France, the Mediterranean and the Iberian Peninsula, whereof the Iberian Peninsula is particularly affected by the drying trend toward the far future.

This study is to our knowledge the first region-specific analysis of droughts over Europe using the SMILE of a high-resolution RCM. It assesses hot spots and region-specific climate signals of droughts over Europe against the backdrop of a high internal variability in drought occurrence. Through the employment of the 50-member SMILE we assess the change in drought occurrence under the consideration of this major source of uncertainty. The choice of the simple PNI index furthermore aligns this study directly for communication and outreach purposes, especially with respect to the plot type of *drying stripes*.

DATA AVAILABILITY STATEMENT

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frwa.2021.716621/full#supplementary-material

REFERENCES

Arora, V. K., Scinocca, J., Boer, G. J., Christian, J. R., Denman, K. L., Flato, G. M., et al. (2011). Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophys. Res. Lett.* 38:L05805. doi: 10.1029/2010GL046270

Basist, A., Bell, G. D., and Meentemeyer, V. (1994). Statistical relationships between topography and precipitation pattern. *J. Clim.* 7, 1305–1315. doi: 10.1175/1520-0442(1994)007andlt;1305:SRBTAPandgt;2.0.CO;2

Bastos, A., Fu, Z., Ciais, P., Friedlingstein, P., Sitch, S., Pongratz, J., et al. (2020). Impacts of extreme summers on European ecosystems: a comparative analysis of 2003, 2010 and 2018. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 375:20190507. doi: 10.1098/rstb.2019.0507

Beillouin, D., Schauberger, B., Bastos, A., Ciais, P., and Makowski, D. (2020). Impact of extreme weather conditions on European crop production in 2018. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 375:20190510. doi: 10.1098/rstb.2019.0510

Blenkinsop, S., and Fowler, H. (2007). Changes in European drought characteristics projected by the PRUDENCE regional climate models. *Int. J. Climatol.* 27, 1595–1610. doi: 10.1002/joc.1538

Bonaccorso, B., Bordi, I., Cancelliere, A., Rossi, G., and Sutera, A. (2003). Spatial variability of drought: an analysis of the SPI in Sicily. *Water Resour. Manag.* 17, 273–296. doi: 10.1023/A:1024716530289

Christensen, J., and Christensen, O. (2007). A summary of the PRUDENCE model projections of changes in European climate by the end of this century. *Clim. Change* 81, 7–30. doi: 10.1007/s10584-006-9210-7

Cook, B., Mankin, J., Marvel, K., Williams, A., Smerdon, J., and Anchukaitis, K. (2020). Twenty-first century drought projections in the CMIP6 forcing scenarios. *Earths Future* 8:e2019EF001461. doi: 10.1029/2019EF001461

Cook, B. I., Anchukaitis, K. J., Touchan, R., Meko, D. M., and Cook, E. R. (2016). Spatiotemporal drought variability in the Mediterranean over the last 900 years. *J. Geophys. Res. Atmospheres* 121, 2060–2074. doi: 10.1002/2015JD023929

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., and Jones, P. D. (2018). An ensemble version of the E-OBS temperature and precipitation data sets. *J. Geophys. Res. Atmospheres* 123, 9391–9409. doi: 10.1029/2017JD028200

Dai, A., and Zhao, T. (2017). Uncertainties in historical changes and future projections of drought. Part I: Estimates of historical drought changes. *Clim. Change* 144, 519–533. doi: 10.1007/s10584-016-1705-2

Dobson, B., Coxon, G., Freer, J., Gavin, H., Mortazavi-Naeini, M., and Hall, J. W. (2020). The spatial dynamics of droughts and water scarcity in england and wales. *Water Resour. Res.* 56:e2020WR027187. doi: 10.1029/2020WR027187

Falzoi, S., Gleeson, E., Lambkin, K., Zimmermann, J., Marwaha, R., O'Hara, R., et al. (2019). Analysis of the severe drought in Ireland in 2018. *Weather* 74, 368–373. doi: 10.1002/wea.3587

Fyfe, J. C., Derksen, C., Mudryk, L., Flato, G. M., Santer, B. D., Swart, N. C., et al. (2017). Large near-term projected snowpack loss over the western United States. *Nat Commun.* 8:14996. doi: 10.1038/ncomms14996

Hänsel, S., Ustrnul, Z., Łupikasza, E., and Skalak, P. (2019). Assessing seasonal drought variations and trends over Central Europe. *Adv. Water Resour.* 127, 53–75. doi: 10.1016/j.advwatres.2019.03.005

Hawkins, E. (2018). *Warming Stripes*. Available online at: http://www.climate-lab-book.ac.uk/2018/warming-stripes/ (accessed May 8, 2021).

Hawkins, E., and Sutton, R. (2011). The potential to narrow uncertainty in projections of regional precipitation change. *Clim. Dyn.* 37, 407–418. doi: 10.1007/s00382-010-0810-6

Hayes, M. J. (2002). *Drought Indices*. National Drought Mitigation Center, University of Nebraska.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M. (2008). A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *J. Geophys. Res. Atmospheres* 113:D20. doi: 10.1029/2008JD010201

IPCC (2013). "Anthropogenic and natural radiative forcing," in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, eds T. Stocker, D. Qin, G. -K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. V. B. Xia, and P. Midgley (Cambridge: Cambridge University Press), 659–740.

IPCC (2018). "Summary for policymakers," in *Global Warming of 1.5°C. An IPCC Special Report on the Impacts of Global Warming of 1.5°C Above Pre-Industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change, Sustainable Development, and Efforts to Eradicate Poverty*, eds V. Masson-Delmotte, P. Zhai, H. -O. Pörtner, D. Roberts, J. Skea, P. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. Matthews, Y. Chen, X. Zhou, M. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield Geneva: (World Meteorological Organization), 32.

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. (2015). The community earth system model (CESM) large ensemble project: a community resource for studying climate change in the presence of internal climate variability. *Bull. Am. Meteorol. Soc.* 96, 1333–1349. doi: 10.1175/BAMS-D-13-00255.1

Kjellström, E., Boberg, F., Castro, M., Christensen, J. H., Nikulin, G., and Sánchez, E. (2010). Daily and monthly temperature and precipitation statistics as performance indicators for regional climate models. *Clim. Res.* 44, 135–150. doi: 10.3354/cr00932

Kjellström, E., Nikulin, G., Strandberg, G., Christensen, O. B., Jacob, D., Keuler, K., et al. (2018). European climate change at global mean temperature increases of 1.5 and 2 C above pre-industrial conditions as simulated by the EURO-CORDEX regional climate models. *Earth Syst. Dyn.* 9, 459–478. doi: 10.5194/esd-9-459-2018

Leduc, M., Mailhot, A., Frigon, A., Martel, J.-L., Ludwig, R., Brietzke, G. B., et al. (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). *J. Appl. Meteorol. Climatol.* 58, 663–693. doi: 10.1175/JAMC-D-18-0021.1

Lehner, F., Coats, S., Stocker, T. F., Pendergrass, A. G., Sanderson, B. M., Raible, C. C., et al. (2017). Projected drought risk in 1.5 c and 2 c warmer climates. *Geophys. Res. Lett.* 44, 7419–7428. doi: 10.1002/2017GL074117

Liu, Y., Zhu, Y., Ren, L., Singh, V. P., Yong, B., Jiang, S., et al. (2019). Understanding the spatiotemporal links between meteorological and hydrological droughts from a three-dimensional perspective. *J. Geophys. Res. Atmospheres* 124, 3090–3109. doi: 10.1029/2018JD028947

Lloyd-Hughes, B. (2014). The impracticality of a universal drought definition. *Theor. Appl. Climatol.* 117:3–4. doi: 10.1007/s00704-013-1025-7

Lorenzo-Lacruz, J., Vicente-Serrano, S., López-Moreno, J., Beguería, S., García-Ruiz, J., and Cuadrat, J. (2010). The impact of droughts and water management on various hydrological systems in the headwaters of the Tagus River (central Spain). *J. Hydrol.* 386, 13–26. doi: 10.1016/j.jhydrol.2010.01.001

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

McKee, T. B., Doesken, N. J., and Kleist, J. (1993). "The relationship of drought frequency and duration to time scales," in *Proceedings of the 8th Conference on Applied Climatology, Vol. 17* (Boston), 179–183.

Mikšovskỳ, J., Brázdil, R., Trnka, M., and Pišoft, P. (2019). Long-term variability of drought indices in the Czech Lands and effects of external forcings and large-scale climate variability modes. *Clim. Past* 15, 827–847. doi: 10.5194/cp-15-827-2019

Mishra, A. K., and Singh, V. P. (2010). A review of drought concepts. *J. Hydrol.* 391, 202–216. doi: 10.1016/j.jhydrol.2010.07.012

Naumann, G., Cammalleri, C., Mentaschi, L., and Feyen, L. (2021). Increased economic drought impacts in Europe with anthropogenic warming. *Nat. Clim. Chang* 11, 485–491. doi: 10.1038/s41558-021-01044-3

Nikbakht, J., Tabari, H., and Talaee, P. H. (2013). Streamflow drought severity analysis by percent of normal index (PNI) in northwest Iran. *Theor. Appl. Climatol.* 112, 565–573. doi: 10.1007/s00704-012-0750-7

Palmer, W. C. (1965). *Meteorological Drought*. U.S. Weather Bureau Research Paper. No. 45. 58pp.

Pfeifer, S., Bülow, K., Gobiet, A., Hänsler, A., Mudelsee, M., Otto, J., et al. (2015). Robustness of ensemble climate projections analyzed with climate signal maps: seasonal and extreme precipitation for Germany. *Atmosphere* 6, 677–698. doi: 10.3390/atmos6050677

Ruffault, J., Martin-StPaul, N. K., Duffet, C., Goge, F., and Mouillot, F. (2014). Projecting future drought in Mediterranean forests: bias correction of climate models matters! *Theor. Appl. Climatol.* 117, 113–122. doi: 10.1007/s00704-013-0992-z

Santos, J. F., Pulido-Calvo, I., and Portela, M. M. (2010). Spatial and temporal variability of droughts in Portugal. *Water Resour. Res.* 46:W03503. doi: 10.1029/2009WR008071

Šeparovi,c, L., Alexandru, A., Laprise, R., Martynov, A., Sushama, L., Winger, K., et al. (2013). Present climate and climate change over North America as simulated by the fifth-generation Canadian regional climate model. *Clim. Dynamics* 41, 3167–3201. doi: 10.1007/s00382-013-1737-5

Sheffield, J., and Wood, E. F. (2012). *Drought: Past Problems and Future Scenarios*. Routledge.

Smakhtin, V. U., and Hughes, D. (2004). *Review, Automated Estimation and Analyses of Drought Indices in South Asia*. Working Paper 83. Colombo: International Water Management Institute.

Spinoni, J., Barbosa, P., Bucchignani, E., Cassano, J., Cavazos, T., Christensen, J. H., et al. (2020). Future global meteorological drought hot spots: a study based on CORDEX Data. *J. Clim.* 33, 3635–3661. doi: 10.1175/JCLI-D-19-0084.1

Spinoni, J., Vogt, J. V., Naumann, G., Barbosa, P., and Dosio, A. (2018). Will drought events become more frequent and severe in Europe? *Int. J. Climatol.* 38, 1718–1736. doi: 10.1002/joc.5291

Stagge, J., Tallaksen, L., and Rizzi, J. (2015). "Future meteorological drought: projections of regional climate models for Europe," in *EGU General Assembly Conference Abstracts* (Vienna), 7749.

Stagge, J. H., Kingston, D. G., Tallaksen, L. M., and Hannah, D. M. (2017). Observed drought indices show increasing divergence across Europe. *Sci. Rep.* 7, 1–10. doi: 10.1038/s41598-017-14283-2

Stefanon, M., D'Andrea, F., and Drobinski, P. (2012). Heatwave classification over Europe and the Mediterranean region. *Environ. Res. Lett.* 7, 9. doi: 10.1088/1748-9326/7/1/014023

Suarez-Gutierrez, L., Milinski, S., and Maher, N. (2021). Exploiting large ensembles for a better yet simpler climate model evaluation. *Clim. Dyn.* doi: 10.1007/s00382-021-05821-w

Van Lanen, H. A., Laaha, G., Kingston, D. G., Gauster, T., Ionita, M., Vidal, J.-P., et al. (2016). Hydrology needed to manage droughts: the 2015 European case. *Hydrol. Process* 30, 3097–3104. doi: 10.1002/hyp.10838

Vicente-Serrano, S. M., Domínguez-Castro, F., Murphy, C., Hannaford, J., Reig, F., Peña-Angulo, D., et al. (2021). Long-term variability and trends in meteorological droughts in Western Europe (1851-2018). *Int. J. Climatol.* 41, E690–E717. doi: 10.1002/joc.6719

Vicente-Serrano, S. M., Lopez-Moreno, J.-I., Beguería, S., Lorenzo-Lacruz, J., Sanchez-Lorenzo, A., García-Ruiz, J. M., et al. (2014). Evidence of increasing drought severity caused by temperature rise in southern europe. *Environ. Res. Lett.* 9:044001. doi: 10.1088/1748-9326/9/4/044001

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. *Clim. Dyn.* 53, 1963–1979. doi: 10.1007/s00382-019-04755-8

Werick, W., Willeke, G., Guttman, N., Hosking, J., and Wallis, J. (1994). National drought atlas developed. *Eos* 75, 89–90. doi: 10.1029/94EO00706

Willeke, M., and Hosking, J. (1994). *The National Drought Atlas Institute for Water Resources Rep*. Fort Belvoir, VA: Army Corps of Engineers, 587.

Wood, R. R., Lehner, F., Pendergrass, A. G., and Schlunegger, S. (2021). Changes in precipitation across time scales in multiple global climate model large ensembles. *Environ. Res. Lett.* 16, 084022. doi: 10.1088/1748-9326/ac10dd

Yihdego, Y., Vaheddoost, B., and Al-Weshah, R. A. (2019). Drought indices and indicators revisited. *Arabian J. Geosci.* 12, 69. doi: 10.1007/s12517-019-4237-z

Zhao, T., and Dai, A. (2017). Uncertainties in historical changes and future projections of drought. Part II: model-simulated historical and future drought changes. *Clim. Change* 144, 535–548. doi: 10.1007/s10584-016-1742-x

5.3   Paper III: European Heatwave Tracks: Using Causal Discovery to Detect Recurring Pathways in a Single-Regional Climate Model Large Ensemble

Reference: 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. *Env. Res. Lett.*, 18(1), 014038. DOI:10.1088/1748-9326/aca9e3

Transition to paper III: In some European regions, droughts not only span several seasons, but recurrently precede or follow summer heatwaves on an inter-seasonal time scale. This was shown in a co-authored companion paper that clustered European heatwave core regions in the CRCM5-LE (Felsche et al. 2023). Paper III then investigates how heatwaves propagate among these core regions during summers by finding a structure to describe their recurrent tracks: A causal discovery algorithm is for the first time applied to derive the dominant spatio-temporal structure of regional heatwave onsets in the SMILE. As an advantage, the SMILE provides numerous events for this exploratory work. The strength of the structure is quantified by conditional probabilities (e.g., onset in A followed by onset in B and onset in B preceded by onset in A). Strongest tracks occur in western and southeastern Europe. Being among the first studies to employ a causal discovery algorithm in a SMILE, this work also assesses the internal variability of spatio-temporal heatwave track structures and compares the observational structure to the model. Robust estimates of heatwave tracks are further sidelined by composites of the accompanying (though not solely causing) large-scale geopotential height patterns at 500 hPa. In doing so, paper III goes beyond evaluations of trends and internal variability of known hazard structures or static patterns (papers I and II) by distilling and evaluating a yet unknown heatwave structure in a data-driven way from the CRCM5-LE.

Author's contribution: AB conceptualized this study and performed the calculations and analyses with support by EF. RL monitored and supervised the research process. All authors: revision of the manuscript and approval of the submitted version.

Status: published in Environmental Research Letters (Journal Impact Factor: 6.7, 12/2023)

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

CrossMark

European heatwave tracks: using causal discovery to detect recurring pathways in a single-regional climate model large ensemble

A Böhnisch[1,*] , E Felsche[1,2] and R Ludwig[1]

[1] Department of Geography, Ludwig-Maximilians University Munich, Luisenstrasse 37, Munich, Germany
[2] Center for Digital Technology and Management, Munich, Germany
[*] Author to whom any correspondence should be addressed.

E-mail: a.boehnisch@lmu.de

Abstract

Summer heatwaves repeatedly affect extended regions in Europe, resulting in adverse economic, social, and ecological impacts. Recent events, e.g. the 2022 heatwave, also attract interest regarding the spatial shifts of their impact centers. Evaluations so far either investigated heatwave passages at pre-defined locations or employed algorithms to spatio-temporally track their core regions. Usually, the latter focus on single events, and thus often fail to generalize spatial heatwave tracks or ignore track characteristics. Here, we use a data-driven approach employing causal discovery to robustly characterize European heatwave tracks in single-model initial condition large ensemble (SMILE) climate simulations to overcome sampling uncertainties of observational records. This enables us to identify specific recurrent heatwave tracks, evaluate their preferential seasonal occurrence, and associate them with moving high pressure centers. Additionally, the evaluation of heatwave track representation in the SMILE extends standard model evaluation, which is mostly based on static statistics. We provide the first comprehensive analysis on heatwave tracks considering internal climate variability conducted within a SMILE, promoting the latter as a methodological testbed in climate extremes research.

1. Introduction

Due to their frequent and often persisting occurrences in extended European regions, heatwaves accounted for substantial economic, social, health, and ecologic impacts and loss during the past years.

Local heatwave occurrence is commonly associated with various dynamic or thermodynamic drivers. Dynamic drivers include anticyclonic, sometimes quasi-stationary ('blocking') conditions fostering local heating during subsidence processes or heat advection (Lhotka *et al* 2018, Simmonds 2018, Shafiei Shiva *et al* 2019, Li *et al* 2020, Suarez-Gutierrez *et al* 2020). Blocking is known to be connected to Northern Hemisphere temperature extremes (see also Zhuo *et al* 2022). Thermodynamic drivers refer to soil moisture conditions affecting evaporative cooling (Hirschi *et al* 2011, Jaeger and Seneviratne 2011) or local

and upwind land-atmosphere coupling (Fischer *et al* 2007, Lhotka and Kyselý 2015b, Schumacher *et al* 2019) which may both be influenced by preceding seasonal precipitation deficits (Della-Marta *et al* 2007, Vautard *et al* 2007, Bastos *et al* 2020, Felsche *et al* 2022). Associated to the latter, anthropogenic land use changes like albedo effects, humidity and radiation budget changes due to imperviousness alteration, i.e. urbanization, regionally explain considerable portions of heatwave variability (Li *et al* 2021, Wu *et al* 2021). Studies show that favorable projected changes in these drivers or their combinations imply increases in heatwave number, duration, and intensity under changing climate conditions which stresses the importance of quantifying heatwave characteristics (Meehl and Tebaldi 2004, Shafiei Shiva *et al* 2019, Molina *et al* 2020, Suarez-Gutierrez *et al* 2020) for adaptation.

Additionally, scientific knowledge on geographical heatwave characteristics is key to adaptation, be it the derivation of distinct geographical heatwave regions in Europe by means of clustering (Stefanon *et al* 2012), the integration of spatio-temporal characteristics in a heat severity index (Keellings and Moradkhani 2020), inter-model differences of typical spatial patterns (Gibson *et al* 2017, Molina *et al* 2020), or ranking of historical events regarding their extent and intensity (Russo *et al* 2015, Lhotka and Kyselý 2015a, 2015b). Depending on the size and shape of regions affected by heatwaves, people exposure and energy demand for cooling may increase, especially in regions not acclimatized to high temperatures (Smith *et al* 2013, Lyon *et al* 2019). While the spatial extent of heatwaves is captured by temporally aggregated measures like the *heatwave magnitude index daily* (HWMId, Russo *et al* 2015), they miss the temporal evolution of heatwaves.

Increased knowledge on the spatio-temporal evolution of heatwaves offers the opportunity for warning and preparation along their propagation pathways (Clemesha *et al* 2018, Lo *et al* 2021). Thus, not only defining static heatwave regions, but also considering their time-dependent positions is crucial in heatwave prediction (Clemesha *et al* 2018), analysis (Sánchez-Benítez *et al* 2020, Lo *et al* 2021), or impact assessment. For example, Stefanon *et al* (2012) present the temporal sequence of spatial European heatwave clusters, but without further investigating potentially preferential sequences. Clemesha *et al* (2018) investigate the propagation and characteristics of observed California heatwaves when passing pairs of distinct geographical regions. Others follow a Lagrangian perspective by determining major heating processes within European heatwave trajectories (Bieli *et al* 2015, Zschenderlein *et al* 2019), or by tracking global drought displacement (Herrera-Estrada *et al* 2017, Herrera-Estrada and Diffenbaugh 2020). Spensberger *et al* (2020) find that one observed transition among heatwaves in distant regions, Central European–Scandinavian successions in 2003 and 2018, occurred coincidentally based on dynamics investigations.

These analyses do not generalize heatwave track characteristics and fail to explain the reasons of specific tracks. Thus an approach to (a) derive typical heatwave tracks, and (b) summarize track specifics is needed, in order to (c) associate tracks with explanatory variables. We attempt to extend and generalize the idea of establishing dynamical links among heatwaves occurring in close temporal and spatial vicinity.

An elegant solution to derive directed links among spatio-temporal phenomena, including the information on link significance, effect size, and temporal lags of signal propagation, is provided by causal discovery algorithms. Causal discovery or causal inference allows harvesting the potential of climate big data to gain knowledge on processes by deriving

hypotheses on causal interdependence hidden in (observational) data (Ebert-Uphoff and Deng 2012a, Deng and Ebert-Uphoff 2014, Runge *et al* 2019a, 2019b). More specifically, Runge *et al* (2019b) mention causal hypothesis testing, climate network analysis, driver identification with respect to extreme impacts or model evaluation as potential applications in climate research. So far, causal discovery was used to obtain networks of information flow within the atmosphere, e.g. portraying westerly atmospheric flows (Ebert-Uphoff and Deng 2012b), or the identification of major atmospheric perturbation 'gateways' related to teleconnections (Runge *et al* 2015). Almendra-Martín *et al* (2022) use a causal discovery algorithm to identify the influence of climate modes on soil moisture variability. Very recently, first climate model ensemble based investigations of teleconnections were performed (Galytska *et al* 2022, Karmouche *et al* 2022).

Transferring the concept of information flow (Runge *et al* 2015) to the heatwave propagation question, we suggest interpreting heatwave movements as temporally lagged heatwave onset in distinct regions. We propose a generalized approach to explain heatwave propagation based on causal discovery and examine potential explanations for derived pathways. This study builds on results from a previous study (Felsche *et al* 2022), where recurrent heatwave core regions were derived by means of clustering. We hypothesize that distinct recurring propagation tracks can be identified beyond single events in Europe. To generalize from observed event sequences we employ a single-model initial condition large ensemble (SMILE) of a regional climate model. This increases the sample of potential heatwave tracks and allows to evaluate observed tracks with respect to naturally occurring climate variability in high spatio-temporal resolution.

2. Methodology

Our approach to track heatwaves consists of four steps: (a) define each day within the analysis period meeting preset criteria as a heatwave (hot day), (b) cluster all hot days with respect to their geographical extent into core regions, (c) aggregate these core regions into time series and (d) apply a causal discovery algorithm to derive directed links among the core region time series.

2.1. Data and validation

Observational records provide a limited amount of heatwave events in Europe. Thus, employing a SMILE allows to increase the sample size: By using the Canadian Regional Climate Model, version 5, Large Ensemble (CRCM5-LE, Leduc *et al* 2019) we obtain 50 members of comparable climate statistics to derive robust results at high spatial resolution ($0.11°$, $12.5\,\text{km}$) while considering internal climate

variability. While the CRCM5-LE was already extensively used for analyses on rare (hydrometeorological) extremes (Champagne *et al* 2020, Wood and Ludwig 2020, Böhnisch *et al* 2021, Brunner *et al* 2021, Poschlod and Ludwig 2021), the present study is the first to apply causal inference where large samples are also of importance (Spirtes and Glymour 1991, Runge *et al* 2019b).

We pool the period 1981–2010 from 50 members into one 1500 year time series. Since all members are designed to share comparable climate statistics, typical patterns and subsequently pathways of phenomena are assumed to be comparable across all members (and to reality), only varying due to internal climate variability (Leduc *et al* 2019).

For heatwave definition, we use daily maximum temperature (tasmax). Daily 500 hPa geopotential height (Z500) is employed for large-scale atmospheric pattern definition. We first perform heatwave clustering and track derivation on a ERA-Interim driven (Dee *et al* 2011) run of CRCM5 (CRCM5/ERA, Leduc *et al* 2019) in order to evaluate them against historical events while using the same regional climate model. Additionally, an external evaluation on the E-OBS gridded dataset (version 22.0e at 0.1° spatial resolution; Haylock *et al* 2008, Cornes *et al* 2018) is performed (figure S1). All analyses are confined to land areas within a European domain (Leduc *et al* 2019).

2.2. Heatwave definition

Heatwave definitions in literature vary according to study perspectives (Perkins and Alexander 2013, Smith *et al* 2013, Shafiei Shiva *et al* 2019). To properly assess heatwave tracks, our heatwave definition requires information for each time step (day) under consideration. We define heatwaves relative to local climatology (Della-Marta *et al* 2007) as a minimum of three consecutive hot days, separated by at least three non-hot days (Keellings and Moradkhani 2020, Spensberger *et al* 2020) from the preceding/following hot day: a hot day occurs if the centered three-day-running mean of tasmax exceeds the local 95th JJA (1981–2010) percentile (similar to Lhotka and Kyselý 2015a). Negative anomalies are omitted. Whereas the percentile in a single run or observations/reanalysis may be poorly estimated due to the limited sample size (30 years), uncertainty due to internal variability is reduced in the SMILE by deriving the percentile from a 1500 year sample. Additionally, percentile based anomalies reduce potential inconsistencies among the data sources induced by climate model bias.

Spurious artifacts ($\leqslant 9 \times 9$ pixels) and isolated hot day occurrences are removed. A time step is considered as belonging to an extended heatwave event, if hot days occur in at least 1% of the land area in the domain (i.e. 500 grid cells). All matching time steps are next used to derive core regions by clustering (Felsche *et al* 2022). Preliminary analyses revealed hot

days in months May–October, such that we splice this period over all years and members (Ebert-Uphoff and Deng 2012a).
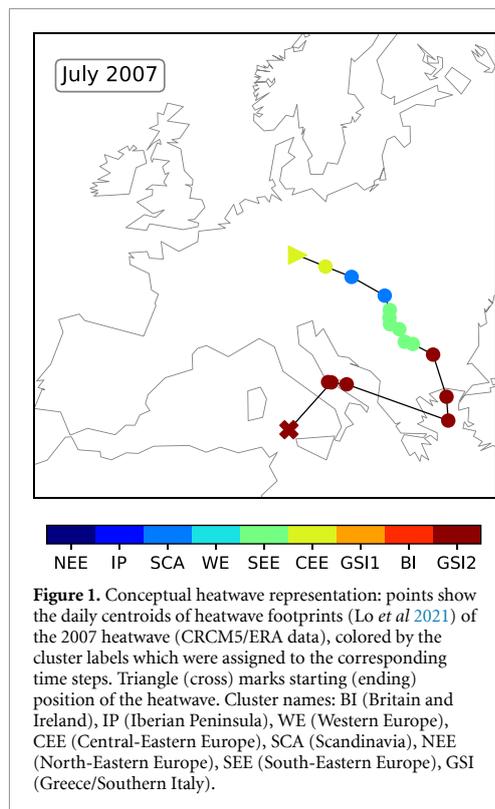
2.3. Location of core regions

For core region definition, we refer to spatially coherent recurring patterns of European heatwaves by means of clustering (Felsche *et al* 2022).

Clustering is frequently employed for midlatitude weather pattern analysis and classification (Smyth *et al* 1997, Stefanon *et al* 2012, Lhotka and Kyselý 2015a, Hannachi *et al* 2017, Wang *et al* 2018, Keellings and Moradkhani 2020, Machado and Lopes 2020). We use daily maps of hot day occurrences as input to a two-step agglomerative hierarchical clustering (Felsche *et al* 2022). The optimal number of clusters is determined by computing the distortion score for every possible number of clusters and piking the knee of the curve (Jung *et al* 2003). When comparing cross-validation clustering results against cluster results from a Monte–Carlo pseudo-experiment, nine spatial heatwave patterns are obtained which are significant on the 99% level based on a two-sided t-test (Felsche *et al* 2022). These cluster footprints correspond well with historical events captured by other heatwave indexes (e.g. Russo *et al* 2015). Felsche *et al* (2022) shows high consistency among clusters in CRCM5/ERA and CRCM5-LE. We follow their naming conventions for CRCM5-LE: BI (Britain and Ireland), IP (Iberian Peninsula), WE (Western Europe), CEE (Central-Eastern Europe), SCA (Scandinavia), NEE (North-Eastern Europe), SEE (South-Eastern Europe), GSI (Greece/Southern Italy). Sequences of time steps labeled according to their respective clusters then represent single heatwaves (figure 1).

Next, clustered time steps are spatially averaged, using the normalized intensity of temporally averaged cluster footprints as weights for their centroid coordinates (Herrera-Estrada *et al* 2017, Herrera-Estrada and Diffenbaugh 2020, Sánchez-Benítez *et al* 2020). This provides nine time series of positive temperature anomalies, located at the cluster footprint centers (i.e. their centroids) and representing the components (Runge *et al* 2015) among which spatio-temporal links are to be established. A sensitivity experiment using different weights for aggregation and therefore time series positioning revealed no major changes in link directions. Since CRCM5/ERA and CRCM5-LE provide slightly different cluster footprints, the positions of these time series diverge as well (see figure 2(a) and figure S1 for E-OBS).

2.4. Transitions among core regions by means of causal discovery

In order to analyze heatwave tracks between heatwave onset and offset, we split them into transitions among core regions. This captures direct movements between various regions. Location changes within

Figure 1. Conceptual heatwave representation: points show the daily centroids of heatwave footprints (Lo *et al* 2021) of the 2007 heatwave (CRCM5/ERA data), colored by the cluster labels which were assigned to the corresponding time steps. Triangle (cross) marks starting (ending) position of the heatwave. Cluster names: BI (Britain and Ireland), IP (Iberian Peninsula), WE (Western Europe), CEE (Central-Eastern Europe), SCA (Scandinavia), NEE (North-Eastern Europe), SEE (South-Eastern Europe), GSI (Greece/Southern Italy).

cluster regions or growth and shrinkage of the spatial extent without heatwave core shifts are considered as stationary.

A heatwave moves from region A to region B, if heatwave onset in region A is followed by heatwave onset in region B, and if heatwave onset in region B was preceded by heatwave onset in region A: The heatwave stays in region A until day t and moves to region B at day t+1. We thus aim to explain a later position of a heatwave by its previous position (i.e. effect preceded by its cause).

Generally in causal discovery, distinct nodes representing single processes (variables or locations of the same variable) are derived prior to link establishing, by e.g. sampling of equally spaced grid points or rotated principal component analysis (Ebert-Uphoff and Deng 2012b, Deng and Ebert-Uphoff 2014, Runge *et al* 2015, 2019a, Nowack *et al* 2020). Here, we employ time series assigned to nine spatially distributed clusters of heatwave core regions.

To define linear spatio-temporal links among heatwave core regions we employ a causal discovery algorithm provided with the *Tigramite 5.1* Python package, namely the PCMCI algorithm (Runge *et al* 2019a, Nowack *et al* 2020). It is based on the PC algorithm introduced by Spirtes and Glymour (1991): First, a fully connected graph of all variables (nodes) under consideration, e.g. time series at different locations, is created. Then, conditionally independent

links (edges) are deleted, while for the remaining links directions are identified (e.g. by considering a temporal ordering of cause and effect) (Ebert-Uphoff and Deng 2012a, Deng and Ebert-Uphoff 2014, Runge *et al* 2019b). An amelioration with respect to time series is provided by the two-step PCMCI algorithm, adding a momentary conditional independence test to the PC algorithm (Runge *et al* 2019a, Nowack *et al* 2020): it rejects spurious or indirect links by testing for independence conditional on the common past of network nodes. Thus, only direct links among network nodes are kept which goes beyond pure pairwise correlation graphs (Runge *et al* 2019b). The PCMCI algorithm can be more powerful than correlation analysis in determining relationships between variables, even in finding links which may not be obvious in classic correlation analysis (Almendra-Martín *et al* 2022). Additionally, it bears the potential to find indirect links and consider common drivers as opposed to other forms of causal discovery like Granger causality (Granger 1969, Runge *et al* 2019b, Galytska *et al* 2022).

In this present study, conditional independence tests are conducted based on partial linear correlations (ParCorr) with a significance threshold of $\alpha = 10^{-3}$. We use a minimum temporal lag of one day, reflecting transitions, and a maximum of three days, reflecting the three days separating two heatwaves. Additionally, we focus on positively correlated links because negative correlations would correspond to the non-occurrence of heatwaves in region B after occurrence in region A.

2.5. Explaining heatwave tracks

The importance of heatwave drivers varies regionally (Zschenderlein *et al* 2019), with dynamical mechanisms generally dominating over local thermodynamical ones (Suarez-Gutierrez *et al* 2020). As mentioned previously, high pressure, i.e. positive Z500 height anomalies, may lead to subsidence, cloud dispersal and subsequently enhanced heating at the surface. Additionally, prevailing dry conditions, e.g. low soil moisture contents, amplify the heating process by reducing the latent heat flux. Both drivers may interact and intensify, e.g. in feedback loops (Fischer *et al* 2007), since prevailing heat in turn may increase evaporation, thereby leading to further soil drying (Schumacher *et al* 2019).

We focus on associated large-scale Z500 conditions to investigate why heatwaves follow a given track. Patterns similar to the Z500 anomaly composites alongside selected heatwave transitions are searched during months May–October. Thereupon we calculate the probabilities of (a) having this pattern during an eight-day period centered at the indicated transition and (b) of having this transition during the indicated pattern. Pattern similarity is assessed by means of masking: if at least 66% of grid cells at an arbitrary time step show positive anomalies inside the

composite positive anomaly region, and if at the same time an analogous agreement is achieved within the composite negative anomaly region, this time step is considered as being similar to the composite pattern. Consecutive days of similar patterns are summarized as one occurrence, which is related to atmospheric circulation persistence (see Jézéquel *et al* 2018, who relate atmospheric patterns to hot days by means of Euclidean distance calculation).

3. Results

3.1. Spatial propagation of heatwaves

During 1981–2010, 115 European heatwaves with at least three days are identified in CRCM5/ERA compared to 5425 in the 50-member CRCM5-LE. Most heatwaves occur during summer months JJA, starting earlier in the West than in the East (Felsche *et al* 2022). The relative frequency of heatwaves ending in the Eastern parts of Europe is increased compared to the frequency of heatwaves starting there (figure 2(a)). Accordingly, heatwaves start more frequently in the Western parts of Europe. Taken together, these first findings indicate a general west-to-east movement of heatwaves in both CRCM5/ERA and CRCM5-LE, which mirrors the dominating westerly flow onto Europe (e.g. Zschenderlein *et al* 2019). Heatwaves starting outside the defined core regions, e.g. outside the domain, over Northern Africa or the ocean, are not captured.

Most heatwaves consist of several transitions (figures 2(b) and (c)). While in general the number of transitions increases with heatwave length, very stationary heatwaves extend up to 20 days within one cluster and highly mobile heatwaves encounter 6 transitions in ten days. CRCM5-LE shows a stronger variability among transition–length combinations than CRCM5/ERA, representing natural variability of heatwave characteristics. The three historical examples 2003, 2006, and 2010 exhibit extreme lengths with respect to the majority of events (e.g. three transitions in 44 days in 2010).

A more detailed network of statistical robust sequences among neighboring core regions is provided in figures 3(a) and (b). It represents a mixture of latitudinal and longitudinal transition directions like a spine of the continental landmasses: While in the West, a South-Northward direction prevails, the East experiences the opposite direction. Within the scope of our study, we focus on links with lag = 1 among two core regions at a time. In order to obtain robust results, links were first derived on the entire 1500-summer period. To include internal variability of these, they were also calculated for each member separately, allowing to count their frequency. In figure 3, only links occurring in at least 15 members (i.e. 30% of the ensemble) are presented with colors indicating their frequency (compare Galytska *et al* 2022, Karmouche *et al* 2022, for alternative
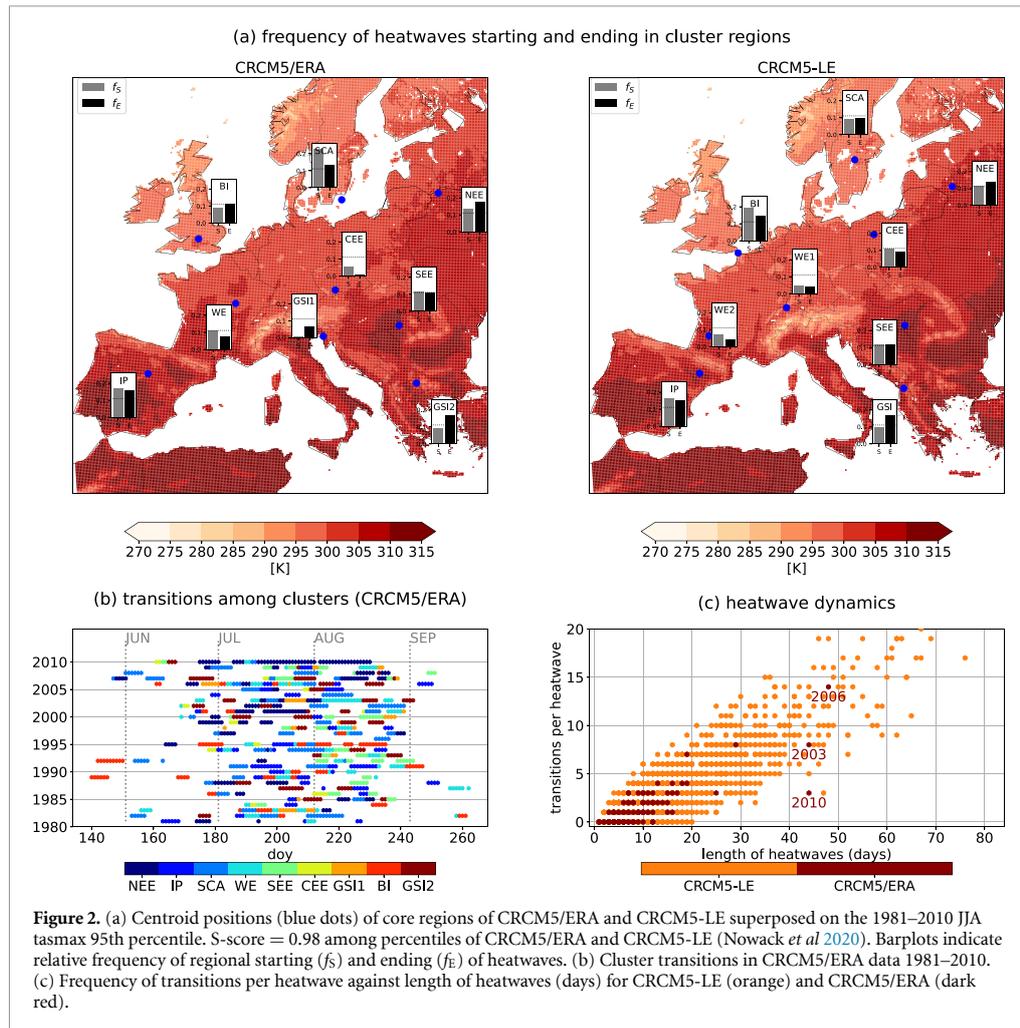
presentation methods in ensembles). As a real-world example, the July 2007 heatwave (figure 1) is represented within the SEE–GSI link.

In order to further evaluate the obtained links, we next compare the graph based on model data with reanalysis and observation based links (compare Karmouche *et al* 2022). While most of these links figure in the CRCM5/ERA and E-OBS network as well (figures S1(a) and (c)), several links from historical data are missing within the robust CRCM5-LE links, being thus (most likely) rare outliers. Others are shifted compared to CRMC5-LE since the position of core regions diverges. The CRCM5-LE links show higher correspondence to the representation in CRCM5/ERA (F1-scores among members and reference above 0.5, figure S2) than to the one in E-OBS (F1-scores close to 0.2; most likely due to clearly shifted or split core regions like, e.g. GSI). Additionally, no link connecting CEE and SCA can be established in CRCM5/ERA as was also found in Spensberger *et al* 2020. Since CRCM5-LE shows these links in 16 and 30 members, respectively, it is likely not observed in CRCM5/ERA due to potential errors in reference causal graph reconstruction (Galytska *et al* 2022) or internal variability. Exact agreement between reference data and model network cannot be expected due to internal climate variability. However, larger ensemble sizes increase the likelihood of capturing observed relationships within the ensemble range (compare Karmouche *et al* 2022).

Due to algorithm construction, the absence of links is a more robust result than their existence: if no statistical link can be established among two nodes, it is unlikely that a physical process links both under the assumption of data faithfully representing the underlying physical processes (Runge *et al* 2019a). All links are highly invariant towards leaving out arbitrary nodes in calculation (except for nodes contained in a given link, figure 3(b)).

We next analyze pairs of core regions before and after transitions, i.e. the links derived in figure 3. All selected transitions occur most often during July and August (figure 4(a)); however, transitions in Western and North-Western regions tend to occur earlier (beginning of July) than South-Eastern regions (beginning of August) with the exception of transition IP–WE2. This is connected to the seasonal preference of heatwave clusters themselves (Felsche *et al* 2022).

The links found in figure 3 are next evaluated against all possible links between the first and second cluster in a transition (k_1 and k_2). All of them rank among the top three most frequent transitions from their respective starting cluster (figures 4(b) and (c)). Transition occurrences of IP–WE2, WE2–BI and SEE–GSI, i.e. comparably short distances, exceed the expected value of randomly occurring transitions (i.e. k_1 and k_2 being independent) by factor 3 or 4. Among these, SEE–GSI mirrors source region–target region

Figure 2. (a) Centroid positions (blue dots) of core regions of CRCM5/ERA and CRCM5-LE superposed on the 1981–2010 JJA tasmax 95th percentile. S-score = 0.98 among percentiles of CRCM5/ERA and CRCM5-LE (Nowack *et al* 2020). Barplots indicate relative frequency of regional starting (f_S) and ending (f_E) of heatwaves. (b) Cluster transitions in CRCM5/ERA data 1981–2010. (c) Frequency of transitions per heatwave against length of heatwaves (days) for CRCM5-LE (orange) and CRCM5/ERA (dark red).
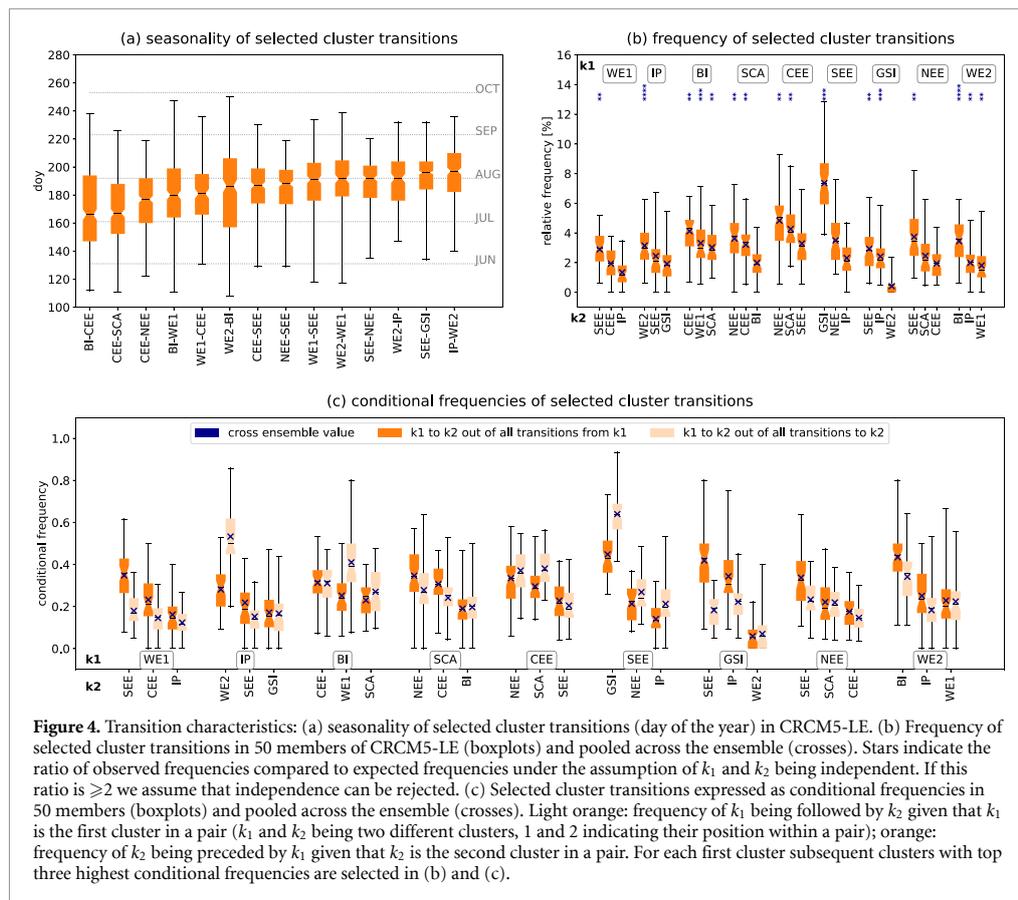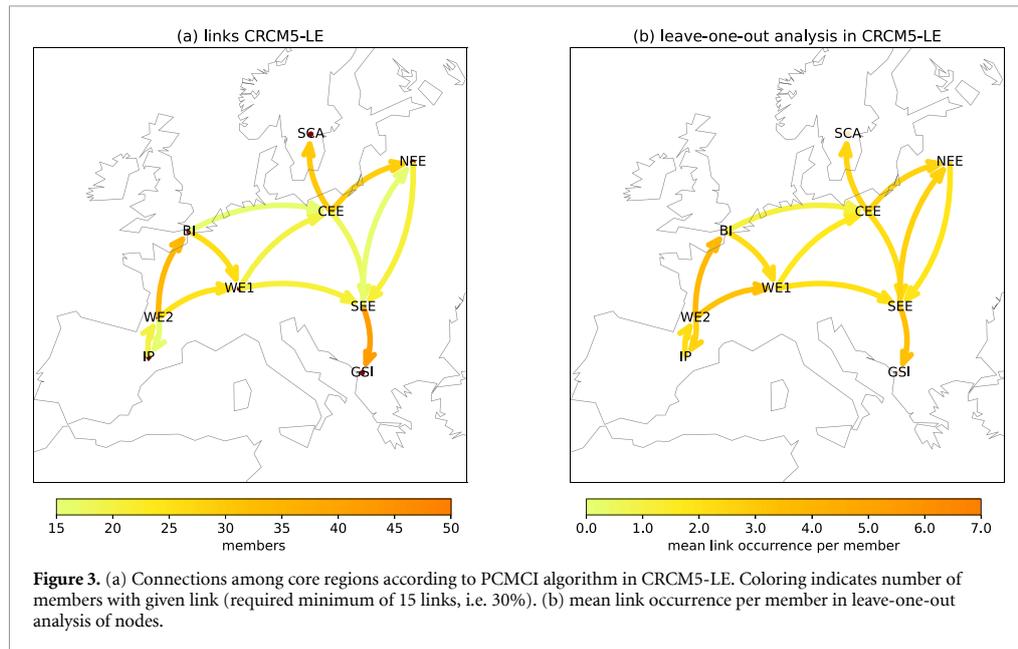
directions of trajectories found in Bieli *et al* 2015. The authors also identify trajectory source regions for heatwaves in the BI region west to Great Britain, which is not represented in figure 3. Generally, source regions (i.e. start regions in our analysis) of air masses are located close to heatwave occurrence regions (Bieli *et al* 2015).

Expressing transition strength in terms of conditional frequencies, figure 4(c) supports the results from the causal discovery network: Values of 1 would indicate a strict connection between both clusters, i.e. k_1 is always followed by k_2 and k_2 being always preceded by k_1. In other words, heatwave occurrence in k_1 is likely a sufficient and necessary condition for heatwave occurrence in k_2. Figure 4(c) shows high cross-ensemble values (and thus suggests strong connections) for, e.g. transitions SEE–GSI or IP–WE2, but also large variability among members: In 45% of all SEE occurrences, the heatwave propagates to GSI. Equally close connections characterize transitions WE2–BI (43%) or GSI–SEE

(42%). This indicates a preference of heatwaves moving from the given first core region to the corresponding second core region within the analysis period.
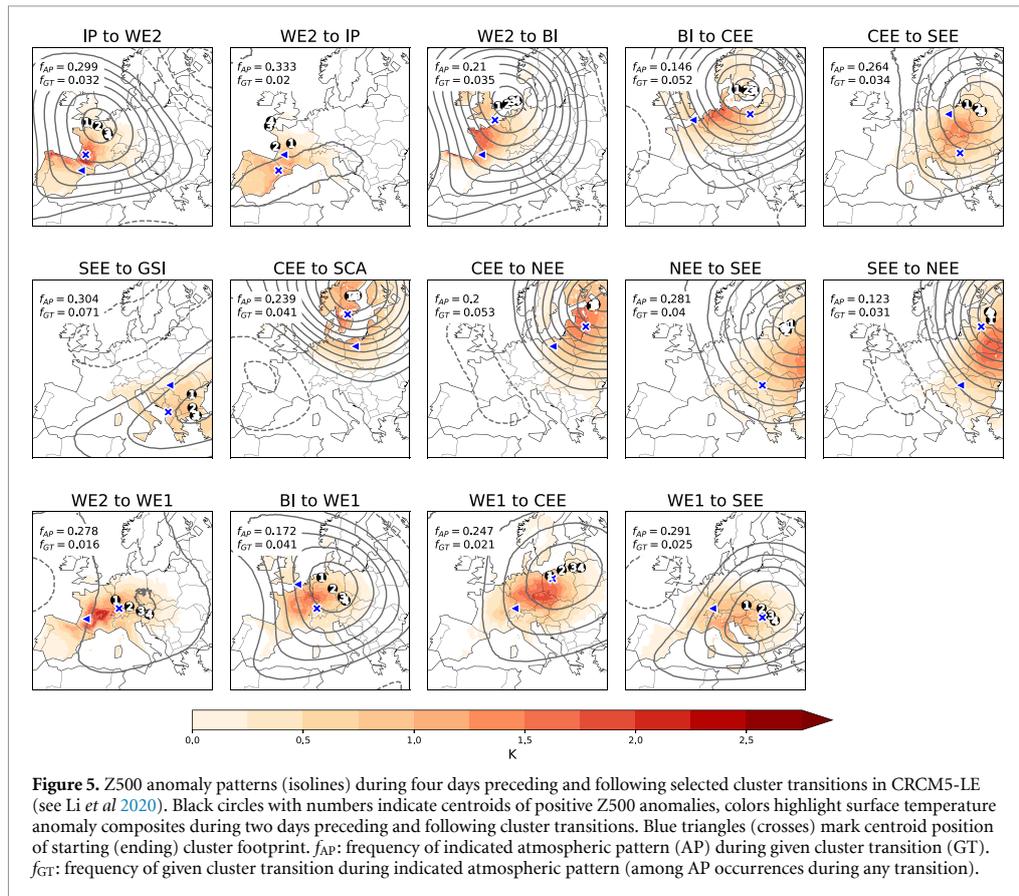
3.2. Associated dynamic conditions

We next seek to connect these previously discussed transitions with potential drivers explaining their pathways (figure 5). Since negative soil moisture anomalies prevail before and after the transitions in the core regions (figure S3), we focus on centroid positions of positive Z500 anomalies during the two days preceding and two days following an indicated transition (figure 5). Our hypothesis states that a moving center of high pressure anomalies, as strong predictors of heat (Suarez-Gutierrez *et al* 2020), draws heatwaves with it in a similar direction. Acknowledging the fact that single events may deviate from these composites and the link between composites and heatwave occurrence may be not specific (Boschat *et al* 2016,

Figure 3. (a) Connections among core regions according to PCMCI algorithm in CRCM5-LE. Coloring indicates number of members with given link (required minimum of 15 links, i.e. 30%). (b) mean link occurrence per member in leave-one-out analysis of nodes.

Figure 4. Transition characteristics: (a) seasonality of selected cluster transitions (day of the year) in CRCM5-LE. (b) Frequency of selected cluster transitions in 50 members of CRCM5-LE (boxplots) and pooled across the ensemble (crosses). Stars indicate the ratio of observed frequencies compared to expected frequencies under the assumption of k_1 and k_2 being independent. If this ratio is $\geqslant 2$ we assume that independence can be rejected. (c) Selected cluster transitions expressed as conditional frequencies in 50 members (boxplots) and pooled across the ensemble (crosses). Light orange: frequency of k_1 being followed by k_2 given that k_1 is the first cluster in a pair (k_1 and k_2 being two different clusters, 1 and 2 indicating their position within a pair); orange: frequency of k_2 being preceded by k_1 given that k_2 is the second cluster in a pair. For each first cluster subsequent clusters with top three highest conditional frequencies are selected in (b) and (c).

Clemesha *et al* 2018, Zschenderlein *et al* 2019), we only examine regions where the signal exceeds the standard error among composite components. As was found for stationary heatwave patterns (Stefanon *et al* 2012), Z500 anomaly composites and temperature patterns correspond well in their spatial extent during

Figure 5. Z500 anomaly patterns (isolines) during four days preceding and following selected cluster transitions in CRCM5-LE (see Li *et al* 2020). Black circles with numbers indicate centroids of positive Z500 anomalies, colors highlight surface temperature anomaly composites during two days preceding and following cluster transitions. Blue triangles (crosses) mark centroid position of starting (ending) cluster footprint. f_{AP}: frequency of indicated atmospheric pattern (AP) during given cluster transition (GT). f_{GT}: frequency of given cluster transition during indicated atmospheric pattern (among AP occurrences during any transition).

transitions. Some heatwave footprints, however, are observed to be larger than their associated synoptic patterns (Spensberger *et al* 2020). The IP–WE2 Z500 and heat anomaly patterns as well as the propagation tracks are similar to the 'European Cluster' in Sánchez-Benítez *et al* 2020.

In general, Z500 anomalies are located to the North-East of the heatwave core regions (compare crosses and triangles in figure 5). Blocking is known to intensify hot extremes southwest of its occurrence by various heating processes and advecting warm continental air masses (Pfahl 2014). For BI–CEE, CEE–SEE, BI–WE1, WE1–SEE, and SEE–GSI, centroid positions of Z500 anomalies lie between start and end positions of the transitions, in parts clearly mirroring the movement direction (see positions of 1–4 relative to each other, figure 5, and figure S3 for a comparison of selected links with CRCM5/ERA).

These associated patterns tend to appear during 12%–33% of the respective transitions in CRCM5-LE. Unsurprisingly, relative occurrences of the indicated transitions during all pattern occurrences remain low (0.1%–0.8%), given a high frequency of these patterns during May–October. Thus, the specific Z500 pattern occurrence is very likely not a sufficient cause for transition occurrence, but may be treated

as a necessary condition (see Boschat *et al* 2016, Hannart *et al* 2016). Among pattern occurrence during any heatwave transition, however, the indicated transitions occur up to ten times more frequently, ranging from 2% to 7%. To summarize, associated pattern occurrence alone cannot be interpreted as a clear sufficient or necessary condition to causally explain heatwave transitions, although they tend to propagate in the same direction as contemporaneous heatwaves.

4. Discussion

By employing a framework involving causal discovery, we find indications of distinct heatwave tracks in Europe during 1981–2010. A very recent example of their meaningfulness is the July 2022 heatwave in Europe which started on the Iberian Peninsula, propagated towards Western Europe, Great Britain, Central-Eastern Europe and last to South-Eastern Europe (ECMWF 2022).

In parts, heatwave tracks are explained by large-scale circulation patterns (land-atmosphere or sea surface temperature related drivers were outside the study scope). Our approach allows to extract information on heatwave direction, dislocation velocity and

seasonal occurrence along these tracks. These analyses are intended to understand a facet of spatially distributed extreme events, especially heatwaves, but not specifically to add to local prediction. Defining core regions based on naturally occurring geographical patterns of heatwaves rather than analyzing at arbitrarily chosen coordinates allows to perform a meaningful complexity reduction by grouping coherent spatial patterns to one core region. Causal discovery facilitates the comprehensive connection of heatwave regions across space and time and thus merges two dimensions which are commonly investigated separately. Moreover, it isolates dominant heatwave directional movements from a large dataset and identifies target regions of heatwave tracks.

Further investigation of recurrent heatwave tracks regarding the spatially varying interplay (Suarez-Gutierrez *et al* 2020) and importance of their respective drivers (as e.g. in Wu *et al* 2021) bears the potential to improve heatwave prediction or attenuation of regional driving factors by e.g. suitable land cover changes (Miralles *et al* 2019).

There are some limitations to our approach. First, employing a causal discovery algorithm requires certain conditions to be met (Spirtes and Glymour 1991, Ebert-Uphoff and Deng 2012a, Runge *et al* 2015): In order to correctly derive a network of directed links, all variables of relevance need to be included in the analysis (causal sufficiency). A spatially constrained domain may exclude regions of concern to the phenomenon. We attempt to evade this problem by constraining our analysis to land masses of Europe, i.e. heatwaves starting and ending in Europe. Moreover, large scale air mass flow—as shown to be related to heatwave tracks—dominantly follows a west-east path in Europe, thus leaving our domain towards the cut-off side. In order to clarify the influence of potentially missing nodes or variables outside the domain, we performed a leave-one-out analysis mimicking different 'domain sizes'. The analysis showed no considerable changes to the network (figure 3(b)). Thus, even if pathways entering/exiting the domain were missing, links within the domain would remain widely undisturbed. Consequently, the obtained links are valid with respect to the included nodes only as is also mentioned in Karmouche *et al* 2022.

This leads to the question why using a regional domain and not a global one, which may have avoided this problem. However, the selection of a small spatial domain reduced the risk of having several heatwaves with no common origin at the same time. Moreover, using a spatial resolution of 0.11° as opposed to, e.g. 1° or 2.8°, leads to higher precision of spatial pattern delineation in heterogeneous landscapes (Molina *et al* 2020) and thus improves core region location.

Using a SMILE allows to derive a large variety of events and, consequently, of transitions among regions. This offers the opportunity to investigate

likelihoods of links: For example, WE2–BI, SEE–GSI and CEE–NEE are more robust against the backdrop of internal variability than BI–CEE, CEE–SEE or NEE–SEE. Furthermore, in single realizations of reanalysis/observations long-lasting heatwaves may bias the clustering towards these events. For example, during the persistent 2010 heatwave in Russia, two clusters dominate (figure 2(b)), whereas clustering among 1500 rather than 30 years increases the likelihood of having various different spatial heatwave patterns as a baseline. Additionally, since heatwave tracks are extremely dependent on the exact weather patterns, small samples may bias expectations for larger samples. This was also found when investigating causal discovery links of single SMILE members (e.g. regarding CEE–SCA). Possibly large multi-model uncertainties (Nowack *et al* 2020) are beyond the scope of this study, but should be further investigated, e.g. in SMILEs of different models.

5. Summary and conclusions

In this study, we not only use causal discovery to derive observed spatio-temporal propagation of a phenomenon, but also to investigate spatio-temporal extremes on a more general ground for the first time in a SMILE. Thus the SMILE serves as a methodological testbed to investigate the power of causal discovery in abstracting spatio-temporal links from single event analyses. Additionally, we are able to infer that spatio-temporal links are represented plausibly in our SMILE which adds to common model evaluation based on static statistical metrics.

Causality in this study is understood as explaining the occurrence of heat anomalies at a given location by occurrences at more distant locations. Similar to established methods of heatwave tracking, our method is not confined to this type of spatial extremes and may be extended to investigate dominating landfall heatwave/drought paths.

We suggest that as high pressure anomalies propagate across Europe, associated high temperatures move along. Surface-atmospheric processes, e.g. moisture availability, evaporation, or heating of dry landscapes, may enhance or reduce temperatures, leading to a decoupling of Z500 and temperature anomalies after some time. A more detailed analysis of heat origins or temperature evolution during heatwaves (e.g. associated to Z500 timeseries) may extend the understanding of heatwave propagation. Further studies could also include investigations as to how heatwaves tracks and their connection to large-scale atmospheric patterns evolve under changing climate conditions or in different seasons. In order to verify model representation of heatwave tracks, we propose to repeat similar analyses in further (high-resolution) SMILEs. This could also help to better estimate the influence of domain choice.

If communicated broadly, recurring connections among core regions may foster awareness among people living in a downstream region whenever an upstream region is already affected by a heatwave.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

ORCID iD

A Böhnisch ⬤ https://orcid.org/0000-0001-7496-1478

References

Almendra-Martín L, Martínez-Fernández J, Piles M, González-Zamora A, Benito-Verdugo P and Gaona J 2022 Influence of atmospheric patterns on soil moisture dynamics in Europe *Sci. Total Environ.* **846** 157537

Bastos A *et al* 2020 Direct and seasonal legacy effects of the 2018 heat wave and drought on European ecosystem productivity *Sci. Adv.* **6** eaba2724

Bieli M, Pfahl S and Wernli H 2015 A lagrangian investigation of hot and cold temperature extremes in Europe *Q. J. R. Meteorol. Soc.* **141** 98–108

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 *Front. Water* **3** 716621

Boschat G, Simmonds I, Purich A, Cowan T and Pezza A B 2016 On the use of composite analyses to form physical hypotheses: An example from heat wave—SST associations *Sci. Rep.* **6** 29599

Brunner M I, Swain D L, Wood R R, Willkofer F, Done J M, Gilleland E and Ludwig R 2021 An extremeness threshold determines the regional response of floods to changes in rainfall extremes *Commun. Earth Environ.* **2** 173

Champagne O, Leduc M, Coulibaly P and Arain M A 2020 Winter hydrometeorological extreme events modulated by large-scale atmospheric circulation in southern ontario *Earth Syst. Dyn.* **11** 301–18

Clemesha R E S, Guirguis K, Gershunov A, Small I J and Tardy A 2018 California heat waves: their spatial evolution, variation and coastal modulation by low clouds *Clim. Dyn.* **50** 4285–301

Cornes R C, van der Schrier G, van den Besselaar E J M and Jones P D 2018 An ensemble version of the e-obs temperature and precipitation data sets *J. Geophys. Res.: Atmos.* **123** 9391–409

Dee D P *et al* 2011 The era-interim reanalysis: configuration and performance of the data assimilation system *Q. J. R. Meteorol. Soc.* **137** 553–97

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 *Clim. Dyn.* **29** 251–75

Deng Y and Ebert-Uphoff I 2014 Weakening of atmospheric information flow in a warming climate in the community climate system model *Geophys. Res. Lett.* **41** 193–200

Ebert-Uphoff I and Deng Y 2012b A new type of climate network based on probabilistic graphical models: Results of boreal winter versus summer *Geophys. Res. Lett.* **39** L19701

Ebert-Uphoff I and Deng Y 2012 Causal discovery for climate research using graphical models *J. Clim.* **25** 5648–65

ECMWF 2022 Update on European heatwave of July 2022 (available at: www.ecmwf.int/en/about/media-centre/focus/2022/update-european-heatwave-july-2022) (Accessed 22 July 2022)

Felsche E, Böhnisch A and Ludwig R 2022 Inter-seasonal connection of typical European heatwave patterns to soil moisture Research Square pp 1–25 (https://doi.org/10.21203/rs.3.rs-1673999/v1)

Fischer E M, Seneviratne S I, Lüthi D and Schär C 2007 Contribution of land-atmosphere coupling to recent European summer heat waves *Geophys. Res. Lett.* **34** L06707

Galytska E, Weigel K, Handorf D, Jaiser R, Köhler R H, Runge J and Eyring V 2022 Causal model evaluation of arctic-midlatitude teleconnections in cmip6 *Earth Space Sci. Arch.* 1–33

Gibson P B, Perkins-Kirkpatrick S E, Alexander L V and Fischer E M 2017 Comparing australian heat waves in the cmip5 models through cluster analysis *J. Geophys. Res.: Atmos.* **122** 3266–81

Granger C W J 1969 Investigating causal relations by econometric models and cross-spectral methods *Econometrica* **37** 424–38

Hannachi A, Straus D M, Franzke C L E, Corti S and Woollings T 2017 Low-frequency nonlinearity and regime behavior in the northern hemisphere extratropical atmosphere *Rev. Geophys.* **55** 199–234

Hannart A, Pearl J, Otto F E L, Naveau P and Ghil M 2016 Causal counterfactual theory for the attribution of weather and climate-related events *Bull. Am. Meteorol. Soc.* **97** 99–110

Haylock M R, Hofstra N, Klein Tank A M G, Klok E J, Jones P D and New M 2008 A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006 *J. Geophys. Res.: Atmos.* **113** D20119

Herrera-Estrada J E and Diffenbaugh N S 2020 Landfalling droughts: Global tracking of moisture deficits from the oceans onto land *Water Resour. Res.* **56** e2019WR026877

Herrera-Estrada J E, Satoh Y and Sheffield J 2017 Spatiotemporal dynamics of global drought *Geophys. Res. Lett.* **44** 2254–63

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 *Nat. Geosci.* **4** 17–21

Jaeger E B and Seneviratne S I 2011 Impact of soil moisture–atmosphere coupling on European climate extremes and trends in a regional climate model *Clim. Dyn.* **36** 1919–39

Jézéquel A, Cattiaux J, Naveau P, Radanovics S, Ribes A, Vautard R, Vrac M and Yiou P 2018 Trends of atmospheric circulation during singular hot days in Europe *Environ. Res. Lett.* **13** 054007

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

Karmouche S, Galytska E, Runge J, Meehl G A, Phillips A S, Weigel K and Eyring V 2022 Regime-oriented causal model evaluation of atlantic-pacific teleconnections in cmip6 *EGUsphere* **2022** 1–42

Keellings D and Moradkhani H 2020 Spatiotemporal evolution of heat wave severity and coverage across the united states *Geophys. Res. Lett.* **47** e2020GL087097

Leduc M *et al* 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) *J. Appl. Meteorol. Climatol.* **58** 663–93

Lhotka O and Kyselý J 2015 Characterizing joint effects of spatial extent, temperature magnitude and duration of heat waves and cold spells over central Europe *Int. J. Climatol.* **35** 1232–44

Lhotka O and Kyselý J 2015 Spatial and temporal characteristics of heat waves over central Europe in an ensemble of regional climate model simulations *Clim. Dyn.* **45** 2351–66

Lhotka O, Kyselý J and Plavcová E 2018 Evaluation of major heat waves' mechanisms in EURO-CORDEX RCMs over Central Europe *Clim. Dyn.* **50** 4249–62

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 *Environ. Res. Lett.* **15** 114003

Li X, Fan W, Wang L, Luo M, Yao R, Wang S and Wang L 2021 Effect of urban expansion on atmospheric humidity in Beijing-Tianjin-Hebei urban agglomeration *Sci. Total Environ.* **759** 144305

Lo S-H, Chen C-T, Russo S, Huang W-R and Shih M-F 2021 Tracking heatwave extremes from an event perspective *Weather Clim. Extremes* **34** 100371

Lyon B, Barnston A G, Coffel E and Horton R M 2019 Projected increase in the spatial extent of contiguous US summer heat waves and associated attributes *Environ. Res. Lett.* **14** 114029

Machado J and Lopes A M 2020 Rare and extreme events: the case of covid-19 pandemic *Nonlinear Dyn.* **100** 2953–72

Meehl G A and Tebaldi C 2004 More intense, more frequent and longer lasting heat waves in the 21st century *Science* **305** 994–7

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 *Ann. New York Acad. Sci.* **1436** 19–35

Molina M O, Sánchez E and Gutiérrez C 2020 Future heat waves over the mediterranean from an euro-cordex regional climate model ensemble *Sci. Rep.* **10** 8801

Nowack P, Runge J, Eyring V and Haigh J D 2020 Causal networks for climate model evaluation and constrained projections *Nat. Commun.* **11** 1415

Perkins S E and Alexander L V 2013 On the measurement of heat waves *J. Clim.* **26** 4500–17

Pfahl S 2014 Characterising the relationship between weather extremes in Europe and synoptic circulation features *Nat. Hazards Earth Syst. Sci.* **14** 1461–75

Poschlod B and Ludwig R 2021 Internal variability and temperature scaling of future sub-daily rainfall return levels over Europe *Environ. Res. Lett.* **16** 064097

Runge J *et al* 2019 Inferring causation from time series in earth system sciences *Nat. Commun.* **10** 2553

Runge J, Nowack P, Kretschmer M, Flaxman S and Sejdinovic D 2019 Detecting and quantifying causal associations in large nonlinear time series dataset *Sci. Adv.* **5** eaau4996

Runge J, Petoukhov V, Donges J F, Hlinka J, Jajcay N, Vejmelka M, Hartman D, Marwan N, Paluš M and Kurths J 2015 Identifying causal gateways and mediators in complex spatio-temporal systems *Nat. Commun.* **6** 8502

Russo S, Sillmann J and Fischer E M 2015 Top ten European heatwaves since 1950 and their occurrence in the coming decades *Environ. Res. Lett.* **10** 124003

Sánchez-Benítez A, Barriopedro D and García-Herrera R 2020 Tracking Iberian heatwaves from a new perspective *Weather Clim. Extremes* **28** 100238

Schumacher D L, Keune J, van Heerwaarden C C, Vilà-Guerau de Arellano J, Teuling A J and Miralles D G 2019 Amplification of mega-heatwaves through heat torrents fuelled by upwind drought *Nat. Geosci.* **12** 712–7

Shafiei Shiva J, Chandler D G and Kunkel K E 2019 Localized changes in heat wave properties across the United States *Earth's Future* **7** 300–19

Simmons I 2018 What causes extreme hot days in Europe? *Environ. Res. Lett.* **13** 071001

Smith T T, Zaitchik B F and Gohlke J M 2013 Heat waves in the United States: definitions, patterns and trends *Clim. Change* **118** 811–25

Smyth P, Ghil M, Ide K, Roden J and Fraser A 1997 Detecting atmospheric regimes using cross-validated clustering *KDD'97: Proc. 3rd Int. Conf. on Knowledge Discovery and Data Mining* pp 61–66

Spensberger C, Madonna E, Boettcher M, Grams C M, Papritz L, Quinting J F, Röthlisberger M, Sprenger M and Zschenderlein P 2020 Dynamics of concurrent and sequential central European and Scandinavian heatwaves *Q. J. R. Meteorol. Soc.* **146** 2998–3013

Spirtes P and Glymour C 1991 An algorithm for fast recovery of sparse causal graphs *Soc. Sci. Comput. Rev.* **9** 62–72

Stefanon M, D'Andrea F and Drobinski P 2012 Heatwave classification over Europe and the mediterranean region *Environ. Res. Lett.* **7** 014023

Suarez-Gutierrez L, Müller W A, Li C and Marotzke J 2020 Dynamical and thermodynamical drivers of variability in European summer heat extremes *Clim. Dyn.* **54** 4351–66

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 *Geophys. Res. Lett.* **34** L07711

Wang P, Tang J, Wang S, Dong X and Fang J 2018 Regional heatwaves in china: a cluster analysis *Clim. Dyn.* **50** 1901–17

Wood R R and Ludwig R 2020 Analyzing internal variability and forced response of subdaily and daily extreme precipitation over Europe *Geophys. Res. Lett.* **47** e2020GL089300

Wu X, Wang L, Yao R, Luo M and Li X 2021 Identifying the dominant driving factors of heat waves in the north china plain *Atmos. Res.* **252** 105458

Zhuo W, Yao Y, Luo D, Simmonds I and Huang F 2022 Combined impact of the cold vortex and atmospheric blocking on cold outbreaks over east asia and the potential for short-range prediction of such occurrences *Environ. Res. Lett.* **17** 084037

Zschenderlein P, Fink A H, Pfahl S and Wernli H 2019 Processes determining heat waves across different European climates *Q. J. R. Meteorol. Soc.* **145** 2973–89

5.4   Paper IV: Future Hotspots of Compound Dry and Hot Summers Emerge in European Agricultural Areas

Reference: Böhnisch, A., Felsche, E., Mittermeier, M., Poschlod, B., Ludwig, R. Future Hotspots of Compound Dry and Hot Summers Emerge in European Agricultural Areas. Submitted to Earth's Future. doi: 10.22541/essoar.169447466.61538021/v1

Transition to paper IV: After paper II investigating droughts and paper III focusing on heatwaves, paper IV merges both hazards in addressing dry and hot summers. In particular, the relationship among CDHE summers and soil moisture droughts under three global warming levels is investigated for the first time in high geographical detail in a regional SMILE. Soil moisture droughts are understood as facilitators among meteorological hazards and impacts on, e.g., crops. Here, a two-step structure extraction process is employed: First, the (known) dependence of temperature and precipitation is modeled using bivariate copulas. Next, this bivariate structure is translated into a univariate extremeness measure for each model summer and related to soil moisture droughts for physical quantification. The direction of relationships is not defined since the meteorological hazard and soil moisture contents mutually influence each other. Structural changes, e.g., of the tail dependence, and trends of the structure components are considered to assess CDHE frequency increases. In order to enhance understanding of processes governing heating, drying, and soil depletion, energy flux composites of CDHE summers are compared to non-CDHE summers. Like papers I–III, paper IV builds on the (compared to observations) enhanced sample size of rare events and profits from the high spatial resolution of the CRCM5-LE. After paper I addressing large-scale atmospheric conditions and papers II and III focusing on regional meteorological hazards and upper level drivers, paper IV thereby completes the assessment of spatio-temporal structures in heat and droughts with analyses down on the land surface.

Author's contribution: AB: Conceptualization, formal analysis, investigation, methodology, project administration, visualization, writing original draft, EF and MM: formal analysis, investigation, BP: methodology, software, validation, RL: Supervision, funding acquisition; all authors: review & editing of the original draft.

Status: submitted to Earth's Future (Impact Factor: 8.2, 12/2023)

Future hotspots of compound dry and hot summers emerge in European agricultural areas

Andrea Böhnisch[1], Elizaveta Felsche[1,2], Magdalena Mittermeier[1], Benjamin Poschlod[3], Ralf Ludwig[1]

[1]Department of Geography, Ludwig-Maximilians-Universität München, Munich, Germany
[2]Center for Digital Technology and Management, Munich, Germany
[3]Research Unit Sustainability and Climate Risk, Center for Earth System Research and Sustainability,
Universität Hamburg, Hamburg, Germany

Key Points:

- During compound dry and hot extreme (CDHE) summers, latent heat flux is markedly reduced in widespread areas of the European continent.
- The frequency increase of CDHE events, associated with extremely low soil moisture, doubles under GWL3 compared to GWL2.
- CDHE frequency increases are predominantly driven by rising temperature, with regional contributions of bivariate tail dependence increases.

Corresponding author: Andrea Böhnisch, `a.boehnisch@lmu.de`

Abstract

Compound dry and hot extremes (CDHE, such as recent summers 2015, 2018 and 2022 in Europe) have wide ranging impacts: Heat exacerbates moisture shortages during dry periods whereas water demand rises. Climate change will likely increase the intensity, frequency, and duration of CDHE events in Europe. However, current studies focus on drivers and impacts in coarse-resolution global climate models and likely miss spatial details of CDHE characteristics. To overcome this issue, we exploit a regional 50-member single-model initial condition large ensemble (SMILE) at 12 km spatial resolution. Hence 1000 model years per 20 year-periods provide an extensive database of CDHE and robust estimations of their occurrence changes across Europe in high geographical detail. CDHE occurrences are investigated in a current climate and at two global warming levels (+2 ˚C, +3 ˚C). We identify Northern France, Southern Germany, Switzerland, Southern Ireland, and the western coasts of the Black Sea with currently low CDHE frequencies as emerging hotspots. These regions experience a tenfold occurrence increase under global warming conditions. Apart from Western Europe, temperature is the dominant contributor to frequency increases. Furthermore, tail dependencies strengthen in regions with high CDHE frequency increases. In European agricultural areas, soil moisture shows very strong negative correlations with CDHE extremeness. Last, our results suggest a halving of CDHE in a +2 ˚C world compared to a +3 ˚C world, highlighting the necessity of climate mitigation with respect to this hazard type.

Plain Language Summary

During the last years, summers tended to be exceptionally dry and hot at the same time. Dry and hot conditions affect various economic and ecologic sectors, for example agriculture by soil moisture reduction. Assessing their frequency and intensity under climate change conditions is hence pivotal to develop effective adaptation strategies. The particularity of this study is a so-called regional climate model large ensemble: Its 50 simulations from the same model are equally probable realizations of climate trajectories. We thus investigate 1000 model years for a current climate, a +2˚C and +3˚C warmer world at high geographical detail. This allows for robust analysis as numerous events occur per period. We show that hot and dry summers become more frequent, mostly because of warming with some regions affected by both warming and drying. Furthermore, we find a strengthening link between high temperature and low precipitation, which is often not considered in studies. Additionally, lower soil moisture conditions in agricultural areas coincide with more extreme dry and hot summers. In a +3˚C world, these events are projected to occur at least twice as frequent as in a +2˚C world. This stresses the relevance of climate change mitigation efforts.

1   Introduction

Triggered by an accumulation of recent events, the temporal co-occurrence of extremely dry and hot conditions has sparked a large literature body. Globally, but especially in Europe, simultaneous droughts and heatwaves rank first among multivariate hazard investigations (Afroz et al., 2023). Up to 20 % of heatwaves coincided with droughts since the 1980s (rising trend; Mukherjee & Mishra, 2021). In Europe, droughts during the warm season – often accompanied by heatwaves – increasingly emerge as the dominant drought type (Markonis et al., 2021). For instance, the year 2018 exhibited unprecedented dry and hot conditions during spring to summer in the northern hemisphere (Buras et al., 2020). Vegetation, thriving from suitable growing conditions in spring, aggravated soil depletion by summer due to enhanced transpiration (Bastos et al., 2020).

Heatwaves and droughts share common drivers, albeit on different effective time scales (Miralles et al., 2019). This is reflected in the general negative correlation of temperature and precipitation (Zscheischler & Fischer, 2020; Trenberth & Shea, 2005). For

example, in 2018 anticyclonic blocking through April–October over central Europe, in particular a stationary pattern that was recurrently associated with heat anomalies over Europe and North America, favored persistent dry and hot conditions (Buras et al., 2020; Toreti et al., 2019; Rousi et al., 2023; Kornhuber et al., 2019). Buras et al. (2020) also show the close spatial correspondence of high pressure, hot extremes (which typically occur below anticyclonic conditions, Kornhuber et al., 2019), and water budget deficits. This context can be explained by drying and warming in descending air masses, which exacerbate atmospheric evaporative demand such that subsequently increased evapotranspiration may reduce soil moisture (e.g., Zscheischler et al., 2020). Dry soils in turn heat up more quickly and thus support the sensible heat flux (e.g., Schwingshackl et al., 2017). The warming effect in humid areas during hot and dry conditions due to enhanced net radiation is dampened by evaporative cooling, which is induced by vegetation transpiration and soil evaporation (O et al., 2022). In arid areas, generally low soil water contents and dry vegetation constrain latent heat and amplify temperature increases via enhanced sensible heat fluxes (O et al., 2022). Locally, drought conditions precede extreme heat in summers (Felsche et al., 2023), while simultaneous drought conditions may prolong heatwaves via land-atmospheric coupling (Fischer et al., 2007).

This relationship is mutual: Manning et al. (2019) suggest that enduring and intense hot and dry conditions also trigger soil moisture droughts, and Mukherjee et al. (2023) find amplifying soil effects in both drought–heat and heat–drought cascades. In Germany, soil moisture depletion and precipitation deficits during summer 2018 resulted in a shift from commonly energy-limited to moisture-limited evaporative regimes (Rousi et al., 2023). Soil moisture deficits, however, considerably hamper vegetation productivity (Bastos et al., 2020). In summer 2018, the general water budget was more strongly affected in European agricultural and pasture regions than in forests, but vegetation degraded in both arable and forest regions (Buras et al., 2020). Crop yields of major plants in Northern and central Europe were halved compared to the preceding 5 years (Toreti et al., 2019). In the similarly hot and dry summer of 2003, European gross and net primary production decreased by up to 30 % and 20 %, respectively (Ciais et al., 2005). While heat was shown to mostly affect crop yields, droughts additionally kill the plants (Lesk et al., 2016). Thus a co-occurrence of both extremes also bears the potential to merge impacts, especially by affecting soil moisture as a pre-condition for crop development.

The impacts of compounding extremes are hence amplified compared to its single components. This holds also true for compound dry and hot extreme (CDHE) events, as mentioned previously. Literature describes various kinds of compound events, e.g., pre-conditioned, temporally or spatially compounding, and multivariate types (e.g., Zscheischler et al., 2020). CDHE can be considered as multivariate, in that two hazards co-occur simultaneously in time and space due to their common drivers, or as pre-conditioned if, e.g., soil moisture conditions of previous seasons were taken into account (Zscheischler et al., 2020). Identifying compound events with joint distributions, in this case of temperature and precipitation, allows their investigation via multivariate probability distribution functions, i.e., copulas (Bevacqua et al., 2017; Zscheischler et al., 2020). These represent dependencies among the variables and can be used to derive multivariate extreme value probabilities (Zscheischler et al., 2020). Event occurrence probabilities in turn can be expressed as return periods. For instance, return periods for the CDHE growing season 2018 exceed several thousand years for certain event definitions (Zscheischler & Fischer, 2020). Especially in situations where adaptation and decision making rely on return periods, such as water resources management, bivariate analyses are essential. Without considering the bivariate dependence structure, there is a risk of both overestimating or underestimating the occurrence of events (Bevacqua et al., 2017): For instance, bivariate return periods of the 2014 California winter drought, one of the first CDHE to be investigated bivariately, were shown to be higher than univariate precipitation deficit return periods owing to extremely high winter temperatures (AghaKouchak et al., 2014).

119   Most studies on bivariate events focus on prominent cases without gaining gener-
120   alized knowledge on the event–impact relationships by, e.g., aligning event extremeness
121   with impact extremeness. Examples for this approach include the calculation of (stan-
122   dardized) temperature and precipitation ratios or products (Hao et al., 2018; Mukher-
123   jee & Mishra, 2021), but without considering the variable dependencies. Others employ
124   water budget deficits as CDHE intensity surrogate (Buras et al., 2020). In this study,
125   we consider bivariate return periods as an intensity surrogate. Since they indicate the
126   joint extremeness of the considered variables, higher return periods also correspond to
127   higher temperatures and lower precipitation in the CDHE case. To illustrate the inten-
128   sity of the bivariate return periods, we align soil moisture to the CDHE.

129   In order to evaluate low-frequency compound events and derive meaningful knowl-
130   edge on their effects on soil moisture, observational records provide too few events. Hence,
131   ensembles of climate model simulations are beneficial to enlarge the event sample. How-
132   ever, for the investigation of compound events, it is advisable to be sure about compa-
133   rable process representation in all used simulations (e.g., regarding the joint temperature–
134   precipitation distribution). Both issues can be addressed by accessing single-model ini-
135   tial condition large ensembles (SMILEs) (e.g., Maher et al., 2021). SMILEs consist of
136   several simulations of the same model under the same external forcing conditions (i.e.,
137   scenario), differing only due to their initial conditions. Global SMILEs proved to be a
138   skillful tool for the reduction of uncertainty due to internal variability in multivariate
139   event attribution (Bevacqua et al., 2023). However, it is a known issue that compound
140   events require finer spatial resolution if realistic information for adaptation planning on
141   a regional scale is sought (François & Vrac, 2023).

142   The goal of this study is thus to (a) obtain and explain spatially explicit frequency
143   changes in European CDHE summers (June–August, JJA) under three global warming
144   levels and (b) relate the ranked events with soil moisture as a relevant condition for im-
145   pacts on agriculture. In order to reduce sampling uncertainties from a statistical per-
146   spective and address internal climate variability, we employ a regional high resolution
147   SMILE.

2   Materials and Methods

2.1   Regional Large Ensemble Data for robust sampling

150   Investigating low-probability compound events of extremes requires an abundant
151   data base. We therefore employ the regional SMILE of the Canadian Regional Climate
152   Model, version 5 (CRCM5-LE; Leduc et al., 2019). The CRCM5-LE was developed within
153   the ClimEx project: 50 members of the Canadian Earth System Model, version 2, Large
154   Ensemble (CanESM2-LE; Fyfe et al., 2017; Kirchmeier-Young et al., 2017) were dynam-
155   ically downscaled with the CRCM5 to obtain 50 high-resolution (0.11 °, corresponding
156   to 12.5 km) time series of 1950–2099 over two domains, Europe and Northeastern North
157   America (Leduc et al., 2019). The original members of the CanESM2-LE were constructed
158   by applying small random perturbations to the long-term control run in 1850 and sub-
159   sequently in 1950. After a few years, the 50 members are considered to be independent
160   due to the chaotic nature of weather sequences, while still following the same forcing con-
161   ditions (RCP8.5 from 2006 onward) and thus pertaining comparable climate statistics
162   (Leduc et al., 2019).

163   The CRCM5-LE already proved its value for compound analyses of hydro-meteorological
164   extremes, namely rain on saturated soil and rain-on-snow events (Poschlod et al., 2020).
165   Further, this regional SMILE was used for investigation of heatwaves (Böhnisch et al.,
166   2023), droughts (Böhnisch et al., 2021), and heat and drought linkage at an inter-seasonal
167   scale (Felsche et al., 2023).

2.2 Global Warming Levels in a regional climate model

168

169   We employed global warming levels (GWL) for our analysis of future climate pro-
170   jections. This approach has been widely applied because it has the advantage of being
171   less sensitive to the selected model and scenario. Furthermore, it allows to directly com-
172   pare the warming rate to the goal of the Paris Agreement of limiting global warming to
173   "(...) well below 2 °C above pre-industrial levels and to pursue efforts to limit the tem-
174   perature increase to 1.5 °C (...)" (UNFCCC, 2015). The GWLs were calculated as anoma-
175   lies in the yearly global mean surface air temperature (*tas*) to the pre-industrial refer-
176   ence period 1850–1900 (Hauser et al., 2022; Seneviratne et al., 2021). GWLs refer to a
177   20-year period centered around the first year, in which the warming level is exceeded (*tas*
178   > GWL). The methodology is based on Hauser et al. (2022), which was used for the Sixth
179   Assessment Report of the Intergovernmental Panel on Climate Change (IPCC). We adopted
180   the code for the use in the CanESM2-LE. To this end, we pooled all 50 members before
181   calculating the anomalies to 1850–1900.

182   Our reference period 2001–2020 translates to GWL= +1.2 °C (GWL1.2) in CanESM2-
183   LE (observed approximately 1 °C; Gulev et al., 2021). This is less an effect of the forc-
184   ing scenario for RCP8.5 was shown to be in high agreement with observed emissions (Schwalm
185   et al., 2020). Instead, it mirrors the model's rather high equilibrium climate sensitivity
186   (3.7 K; Swart et al., 2019). Comparing modeled global *tas* with observational global mean
187   temperature though may result in an overestimation partly due to insufficient observa-
188   tional data coverage and blending air temperature over land with sea surface temper-
189   atures over ocean areas in observations (Richardson et al., 2016; Vogel et al., 2019).

190   Future periods in our study are represented by 20-year slices centered at GWL=
191   +2 °C (GWL2, Paris Agreement; UNFCCC, 2015) and GWL= +3 °C (GWL3, close to
192   the most realistic end-of-century temperature of 2.8 °C under current trends in climate
193   policy; Liu & Raftery, 2021).

194   Time periods corresponding to a given GWL were calculated within the global SMILE,
195   and adopted for use in the regional SMILE.

2.3 Definition and Bivariate Evaluation of Compound Events

196

2.3.1 Event Definition

197

198   This study takes a multivariate perspective on dry and hot extremes, since we are
199   particularly interested in the combined occurrences of these hazards. We employed thus
200   the "AND" hazard scenario to connect both univariate extremes (Zscheischler & Fischer,
201   2020): the temporal co-occurrence of linearly detrended summer mean temperatures and
202   (negative) precipitation sums exceeding the respective 95th percentile of 2001–2020 (with
203   the 95th percentile of negative precipitation equaling the 5th percentile; see Supplemen-
204   tary figure S1). By definition, these events are expected to be very rare because both
205   variables have to exceed a high threshold. However, since JJA temperature and nega-
206   tive precipitation show strong correlations in most parts of Europe, which intensified dur-
207   ing the 21st century, CDHE occur more often than would be implied by independence
208   (Zscheischler & Seneviratne, 2017). This implies that warm summers are commonly dry
209   and wet summers are cool (see also Trenberth & Shea, 2005; Wang et al., 2021). Due
210   to the extensive large ensemble database, 1000 years instead of 20 years (see fig. 1 (a))
211   are available per analysis period and allow for robust baseline definition (i.e., percentile
212   estimates across all 50 ensemble members) and event characteristic estimation (e.g., fre-
213   quency changes, associated behavior).

214   In order to characterize CDHE summer energy partitioning compared to non-CDHE
215   summers, we employed the Bowen Ratio (BR, Bowen, 1926). The BR describes the ra-
216   tio of sensible heat flux and latent heat flux, which are negatively coupled (e.g., Schwing-

217   shackl et al., 2017). For this analysis, we used the model variables surface upward la-
218   tent heat flux and surface upward sensible heat flux.

2.3.2   Estimation of Bivariate Return Periods

220   In order to estimate the joint extremeness of CDHEs, we calculated bivariate re-
221   turn periods. Generally, return periods are the inverse of the (annual) exceedance prob-
222   ability p of a given event intensity, the return level z_p. Hence, the return level z_p is ex-
223   pected to be exceeded every $1/p$ years, defining thus the return period $T = 1/p$ (Coles,
224   2001). Bivariate return periods however remain ambiguous and become larger than their
225   univariate component return periods due to the second variable that is required to meet
226   the extremes condition as well (AghaKouchak et al., 2014; Zscheischler & Fischer, 2020).
227   In large samples like the CRCM5-LE, (annual) event occurrences per time period can
228   be counted and inverted to obtain the return period (Zscheischler & Fischer, 2020). This
229   empirical approach is generally limited by the time series length. With 1000 years avail-
230   able, 10 events with $T = 100$ are to be expected statistically, while the most extreme
231   case would be $T = 1000$. Any inference on this level would be highly uncertain since
232   it is based on a single event (e.g., Zscheischler & Fischer, 2020). For shorter time series,
233   the maximum empirical T also decreases such that extreme event estimation suffers from
234   high uncertainties (Bevacqua et al., 2017). Instead of event counting, we here fitted cop-
235   ulas, i.e., multivariate probability distributions, to the bivariate distributions (Zscheischler
236   & Fischer, 2020). The large advantage of distribution fitting is the option for pushing
237   the rareness boundaries of the empirical approach.

238   For the procedure in this study we used the R package *VineCopula* (Nagler et al.,
239   2023). First, we transformed the empirical marginals of summer temperature and pre-
240   cipitation (multiplied with -1 for calculation purposes) to uniform distributions on [0,1].
241   Next, the most suitable copula family was estimated using the Bayesian Information Cri-
242   terion (BIC) and fitted to the data. For this study, we chose the locally best fitting cop-
243   ula family from eight single-parametric copula families (fig. S3).

244   Following the relation in Brunner et al. (2016), the return period T was obtained
245   by:

$$T(u,v) = \frac{\mu}{1 - u - v + C(u,v)} \tag{1}$$

246   giving the probability for jointly exceeding the event defining thresholds in the de-
247   nominator, with u, v corresponding to univariate probabilities of exceeding the respec-
248   tive threshold, $C(u,v)$ being the copula at (u,v), and the mean interarrival time $\mu =$
249   1 in our case since we investigated annual events (Zscheischler & Fischer, 2020; Zscheis-
250   chler & Seneviratne, 2017; Brunner et al., 2016).

2.3.3   Distributional Change Assessments

252   Both changes in temperature and precipitation may lead to frequency changes by
253   shifting the bivariate distribution compared to the reference period. Additionally, the
254   bivariate (tail) dependence structure may change over time.

255   In order to address the first point, we here propose a method to disentangle the
256   dominating drivers of frequency changes. Horizontal shifts of the distribution (along the
257   orange line in fig. 1 (b)) indicate temperature changes as sole drivers whereas vertical
258   shifts (along the blue line in fig. 1 (b)) point to precipitation changes. Any change with
259   both a horizontal and vertical component thus is due to a combination of temperature
260   and precipitation changes. For the definition of the dominating driver, we used the av-
261   erage JJA drying per degree warming (fig. 1 (b)): In Europe, the slope of this relation-
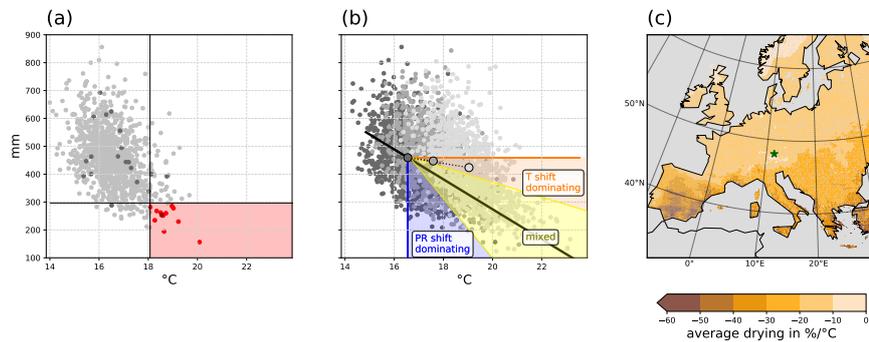
Figure 1. (a) Precipitation and temperature of 1000 summers (50 members for 2001–2020) over a grid cell representing Munich/Germany (star in (c)). Dark grey and dark red dots show the limited sample of one arbitrary member. Black lines indicate the 95th percentile of temperature (vertical) and 5th percentile of precipitation (i.e., the 95th percentile of negative precipitation; horizontal) with the red area highlighting all summers meeting the definition criterion for a CDHE. (b) Definition of temperature (orange) and precipitation (blue) dominance in distributional shifts under climate change conditions. Yellow indicates mixed contributions of temperature and precipitation (see text). Grey shaded point clouds correspond to current, GWL2, and GWL3 climates for the same pixel as in (a). The black line represents the local average summer drying scaled with warming. (c) Average summer drying scaled with warming expressed as slopes of a linear line fitted to the local bivariate distribution.

ship follows a North–South gradient with highest values in the Mediterranean area and especially over the Iberian Peninsula where summer precipitation is very low (fig. 1 (c)). Distributional shifts along this slope represent the occurrence of more extreme events by heating and drying following the current relationship. If the center of the distribution is shifted within the orange sector of fig. 1 (b), temperature is identified as dominating driver, while it is precipitation for shifts into the blue sector. Since we are also interested in simultaneous changes of temperature and precipitation, we introduced a buffer zone between a line with half the local slope and a line with twice the local slope to account for uncertainties in slope estimation (yellow sector). This combination is further referred to as mixed drivers. This approach is based on correlation of the full distributions, which, as Zscheischler and Seneviratne (2017) argue, can serve as an indicator for the likelihood of CDHE if the percentile threshold for event definition is not too high.

To account for dependencies in the distribution extremes, tail (= extremal) dependence above the 95th univariate percentiles (chi(0.95); Coles et al., 1999) were calculated for each period separately using the R package *extRemes* (Gilleland, 2022). Confidence intervals at the 0.05 level were obtained by bootstrapping 1000 times.

2.4 Assessment of CDHE Impacts on Soil Moisture

In one of the first compound event definitions by Leonard et al. (2014), compound events are defined by the extremeness of impacts originating from multiple contributing hazards. While our CDHE definition rather follows a hazard-based perspective, we nevertheless aim to assess CDHE effects in this study. Our (univariate) target variable is soil moisture, classified as the soil moisture index (SMI) of Zink et al. (2016), which also forms the basis of the German Drought Monitor. The SMI is based on soil mois-

286 ture percentiles of a reference period (2001–2020 in our case). We used JJA soil mois-
287 ture in the upper portion of the soil column to assess agricultural droughts during cur-
288 rent climate, GWL2, and GWL3. Soil moisture is especially useful when assessing event
289 impacts, for soil moisture droughts have large agricultural and ecosystem-specific impacts.
290 Assessing soil moisture conditions is hence most relevant in areas where they potentially
291 have an impact. Therefore, we confined our analyses of CDHE–soil moisture relation-
292 ships on European agricultural areas. These comprise Corine Land Cover (CLC2018 ver-
293 sion 2020_20u1, linearly regridded to CRCM5-LE spatial resolution; EEA, 2020) level-
294 2 classes *arable land*, *permanent crops*, and *heterogeneous agricultural areas*.

3   Results

3.1   Bowen Ratio Increases During CDHE

297 CDHE and non-CDHE summers differ with respect to the energy-partitioning of
298 sensible and latent heat flux. In order to illustrate these differences in a spatially explicit
299 way, we first look at the Bowen Ratio during summer under current climate conditions.
300 During non-CDHE summers, the latent heat flux, i.e., evaporative cooling (O et al., 2022),
301 is dominating over the sensible heat flux in large areas of Europe (fig. 2 (a)–(b)). These
302 coincide with the wet evapotranspiration regions (energy-limited) of Schwingshackl et
303 al. (2017). The dominating low BR conditions favor widespread cloud formation and sum-
304 mer precipitation. In CDHE summers (fig. 2 (b)), however, BR increases in large areas.
305 High BR occurs in their wet/transition regions (moisture-limited). Zscheischler et al. (2015)
306 state that under dry conditions, evapotranspiration and temperature are strongly dom-
307 inated by soil moisture. Especially the Mediterranean regions, the lower course of the
308 Danube and coastal regions of the Black Sea experience BR > 10. Under these condi-
309 tions, a reduced latent heat flux (and hence evaporation) suggests low soil moisture avail-
310 ability, while temperatures rise (Mukherjee et al., 2023). Consequently, cloud convec-
311 tion and precipitation are inhibited.

312 We find no BR inversions or only small increases during CDHE in Northern and
313 central Europe as well as in mountainous regions (fig. 2 (a)–(b)). However, these regions
314 are characterized by evaporation increases (and hence soil drying) during CDHE sum-
315 mers (fig. 2 (c)). This suggests an increase in latent heat flux and, potentially, a reduced
316 temperature increase due to evaporative cooling (O et al., 2022). These regions are char-
317 acterized by an energy-limited evapotranspiration regime (Teuling et al., 2009), where
318 higher temperatures in CDHE summers compared to non-CDHE summers favor evap-
319 oration. The remainder of the domain, largely defined by soil-moisture limited evapo-
320 transpiration regimes (Teuling et al., 2009), experiences major evaporation reductions
321 (fig. 2 (c)), presumably due to moisture limitations in comparison to non-CDHE sum-
322 mers. High BR values, i.e., low latent heat flux compared to sensible heat flux, may re-
323 sult from low soil moisture conditions (Trenberth & Shea, 2005). Since soil moisture and
324 evaporation mutually influence each other and CDHE affect evaporation (Miralles et al.,
325 2019), we conclude here that soil moisture is affected by CDHE occurrences as well.

326 The described relationships for CDHE and non-CDHE hold true for GWL2 and
327 GWL3 (see supplementary fig. S2 for BR evolution under GWL2 and GWL3).

3.2   CDHE Frequency Increases

329 CDHE occur rarely under current climate conditions (fig. 3 (a)). Assuming no de-
330 pendence between temperature and precipitation, the occurrence probability of a CDHE
331 would amount to $0.05 \times 0.05 = 0.0025 = 0.25$ events per 100 years. This corresponds
332 to a 1-in-400 year event. This very rare frequency is however exceeded over most of Eu-
333 rope. Assuming total dependence, the frequency has an upper limit at 5 events per 100
334 years by definition of the CDHE events, equaling a 1-in-20 year event. In the CRCM5-
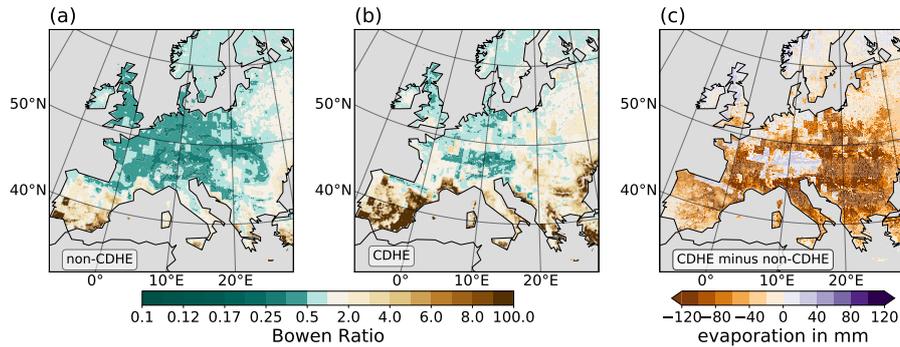
Figure 2.   Bowen Ratio for non-CDHE summers (a) and CDHE summers (b) under current climate conditions. The median across all ensemble members is shown per category. Brownish colors indicate regions with sensible heat > latent heat, greenish colors indicate regions with sensible heat < latent heat. (c) evaporation increases (purple) and decreases (orange) in CDHE summers compared to non-CDHE summers under current conditions.

LE, highest event frequencies reach 3.5 events per 100 years in central eastern Europe (roughly 1-in-28 year event). On the contrary, parts of the Mediterranean, Aegean and Black Sea coastal regions as well as Southern Ireland, Northern France, and mountainous regions in central and Northern Europe encounter < 0.5 events per 100 years which corresponds to a 1-in-200 year event.

For GWL2, event frequencies regionally double to triple, with strongest increases in Southern Europe and weakest changes in Northern and central eastern Europe as well as the Western Iberian Peninsula (fig. 3 (b)). No decreases are detected. Interestingly, while some regions with highest event frequencies under current conditions, e.g., central eastern Europe, encounter only increases by < 3 events per 100 years, Southeastern France both shows high frequencies under current conditions and strong increases under GWL2. Contrasting to that, the coastal areas of the Mediterranean, Aegean and Black Sea with low event occurrences under current conditions experience an even higher increase by 6–9 events per 100 years.

With further ascending GWL, event frequencies surge (fig. 3 (c)): Especially in mountainous forelands of Northern/Northeastern Spain and central/Southwestern France more than 1 out of 4 years under GWL3 qualify as a CDHE with respect to current percentile definitions (adding frequencies in fig. 3 (a) and (c)). The same holds true for the Po Valley in Northern Italy. Regions north of the Alps, in Northern France, Southern Ireland or the Western Iberian Peninsula with currently very few events (< 0.5 per 100 years) experience up to > 15 events per 100 years in addition to current frequencies. Eastern Europe and the Balkans are characterized by a North–South gradient of increases. Lowest gains are found in Scandinavia, Northeastern Europe, the highest Alpine ridges, and Southern Spain. To put these numbers into perspective, Toreti et al. (2019) show that 2018-like droughts mirror typical summer conditions by the 2040s, using a multi-model ensemble under RCP8.5.

3.3 Drivers of CDHE Frequency Increases

What is driving these frequency increases? In fig. 4, we investigate changes in the bivariate distribution of temperature and precipitation. First, fig. 4 (a)–(b) demonstrate the prevalent dominance of temperature increases in shifting the distribution into the
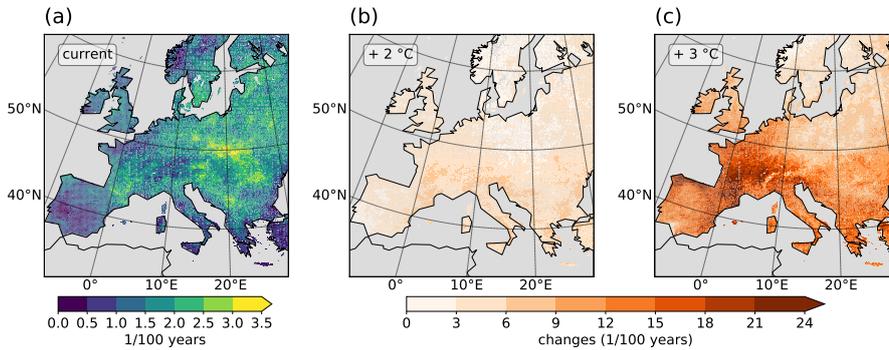
Figure 3. CDHE frequency for three global warming levels (absolute values for present climate (a) and changes under GWL2 (b) and GWL3 (c)). Events are defined as local exceedance of the current (2001–2020) 95th percentile of temperature and (negative) precipitation.

defined CDHE diagram space (see also fig. 1 (b)) under both GWL2 and GWL3. Precipitation dominates in mountainous Norway and Northern Spain. In the Atlantic regions of Western Europe, temperature and precipitation changes jointly foster frequency increases. Under GWL3 conditions, these areas with mixed drivers expand towards the East. In addition, precipitation dominance emerges from previously mixed driver regions. This finding mirrors earlier emergence of (mean summer) temperature trends compared to higher uncertainty and variability in precipitation trends (e.g., von Trentini et al., 2019; Seneviratne et al., 2021). For large parts of Europe, precipitation variability defines hence whether a CDHE occurs, if (nearly) every year exceeds the present temperature threshold of event definition (consistent with e.g., Zscheischler & Fischer, 2020).

Secondly, we consider the dependence structure of the distributions (fig. 4 (c)–(e)). As stated above, a tail dependence of 1 implies that each temperature extreme (as defined here) is associated with a precipitation extreme and vice versa. The joint occurrence probability of CDHE is thus 0.05 (i.e., 5 events per 100 years) and hence the same as for univariate extremes in our definition. On the contrary, a tail dependence of 0 implies independent behavior of temperature and precipitation extremes and thus a probability of $0.05 \times 0.05 = 0.0025$ (i.e., 0.25 events per 100 years in our case). It follows that the spatial distribution in fig. 4 (c) mirrors the spatially distributed CDHE frequencies (fig. 3 (a)) with highest tail dependence corresponding to highest event frequencies in central eastern Europe and bivariate tail independence in mountainous Norway, Northern France, Southern Ireland, inner Alpine regions, and Mediterranean coastal regions with very rare CDHE occurrence. Under GWL2, the tail dependence exceeds the current 95 % confidence interval especially in regions with currently low tail dependence values (e.g., Northeastern France and Northern Italy, the Danube delta or mountainous Norway, fig. 4 (d)). In these regions, the tail dependence increase may add to event frequency. Tail dependence reductions are found on the western Iberian Peninsula with already low values and, notably, in central eastern Europe with currently highest values. More spatially distinct clusters emerge under GWL3 (fig. 4 (e)), where robust tail dependence increases occur in Northern France, Southern UK and Ireland, the Alpine (foreland) and Cantabrian Mountain regions, and Scandinavia. Tail dependence decreases, e.g., in Southern Sweden, parts of the Iberian Peninsula, and central eastern Europe. In South-western Spain, this decrease may contribute to the rather low CDHE occurrence increase under GWL3 conditions (see fig. 3 (c)). Tail dependence changes are reflected by changes in the underlying copula family (supplementary fig. S3 (a)–(c)): For example, tail depen-
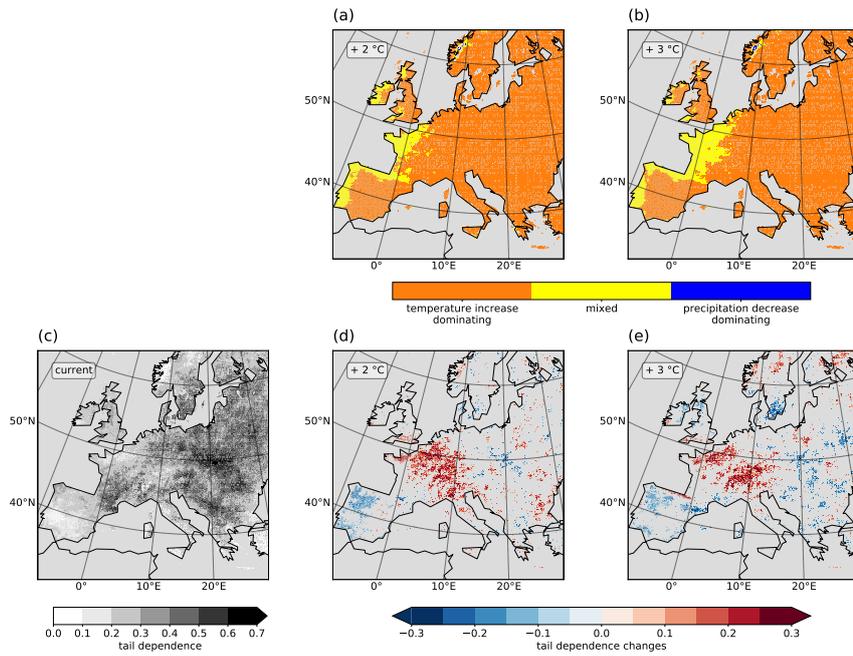
Figure 4.   Changes in combined temperature and precipitation distributions. (a)–(b) distributional shifts due to temperature increases (orange), precipitation decreases (blue) or both (yellow) following the approach from fig. 1 (b)). Only land areas with significant correlations of JJA temperature and precipitation are colored. (c)–(e) tail dependence of temperature and (negative) precipitation: (c) current absolute values, changes for GWL2 (d) and GWL3 (e). For GWL2 and GWL3 only regions with changes exceeding the present 95 % confidence interval are shown. Note: The tail dependence refers to the tails above the respective 95th temperature and (negative) precipitation percentile of each period.

³⁹⁹ dence increases mostly correspond to switches from symmetric copula families (mostly
⁴⁰⁰ Gaussian or Frank) to asymmetric families (e.g., Gumbel which only occur in regions with
⁴⁰¹ BR < 1 under current conditions). Decreases are associated with the inverted switch.
⁴⁰² Symmetric families represent regions with amplified tail dependence in the hot-dry and
⁴⁰³ cold-wet tail, whereas asymmetric families include only one tail with enhanced depen-
⁴⁰⁴ dence. Note that the bivariate structure is generally weak to moderate in most regions
⁴⁰⁵ (theoretical Kendall's τ with $0.2 < \tau < 0.5$, fig. S2 (d)–(f)), pointing towards rather
⁴⁰⁶ similar bivariate distributions. With increasing GWL, τ increases in Western Europe,
⁴⁰⁷ hence strengthening the differences between the joint summer temperature–precipitation
⁴⁰⁸ distributions.

⁴⁰⁹   The tail dependence also allows for a quick change of perspective: Since it is cal-
⁴¹⁰ culated with respect to each period (current, GWL2, GWL3), we are also able to infer
⁴¹¹ that CDHEs defined relative to the percentiles of each period occur more (less) frequently
⁴¹² where tail dependence increases (decreases).

3.4 Soil Moisture Scaling with CDHE Extremeness

⁴¹⁴   To account for the risk that agricultural droughts, i.e., soil moisture deficits, pose
⁴¹⁵ on crops, we focus our further assessment on European agricultural regions.

⁴¹⁶   We start our assessment with return periods T of CDHE in current, GWL2, and
⁴¹⁷ GWL3 conditions (fig. 5 (a)–(c)). Therefore, we ask the question: How extreme would
⁴¹⁸ a future CDHE be in relation to the current temperature and precipitation distribution?
⁴¹⁹ Since higher return periods correspond to hotter and drier summers with respect to cur-
⁴²⁰ rent CDHE, they are interpreted as surrogates for joint event intensity. T is obtained
⁴²¹ for the 95th percentile of temperature and (negative) precipitation of the respective pe-
⁴²² riods from the copula fitted to the present bivariate distribution. Hence under current
⁴²³ conditions (fig. 5 (a)), the distribution again mirrors the current tail dependence (fig. 4 (c))
⁴²⁴ and event frequency distribution (fig. 3 (a)). The theoretical minimum return period of
⁴²⁵ the current period is $T = 20$ (perfect tail dependence), the maximum $T = 400$ (inde-
⁴²⁶ pendence). Consistent with that, we find among the CDHE just passing both thresh-
⁴²⁷ olds return periods of $T = 30$ to $T = 300$ in the current period. Under GWL2 condi-
⁴²⁸ tions (fig. 5 (b)), return periods increase to several hundreds to thousands of years with
⁴²⁹ respect to the current distribution. In single grid cells (dark red), the extremeness of these
⁴³⁰ CDHE is unprecedented ($T = $ inf.). In these cases, (mostly) future temperature or pre-
⁴³¹ cipitation lie outside the margins of the current distribution. Hence CDHE of this ex-
⁴³² tremeness did not occur at all in the current period of the CRCM5-LE. Under GWL3
⁴³³ (fig. 5 (c)), these CDHE are dominating across Europe: We find $T = 1000$ to $T = 3000$
⁴³⁴ years in eastern Germany, Poland, and the Baltics, whereas the remainder of Europe is
⁴³⁵ subject to CDHE with a current occurrence probability $p = 0$. To generalize, the con-
⁴³⁶ ditions of CDHE definition correspond to highly unlikely current conditions when con-
⁴³⁷ sidering GWL2, and unprecedented conditions in GWL3.

⁴³⁸   During all summers exceeding the respective CDHE definition in current, GWL2,
⁴³⁹ and GWL3 climates (fig. 5 (d)–(f)), extreme (below 5th percentile) or exceptional droughts
⁴⁴⁰ (below 2nd percentile) prevail in European agricultural regions. Exceptions are very southerly
⁴⁴¹ parts (Southern Spain, Turkey) where the soil moisture content corresponds to moder-
⁴⁴² ate (below 20th percentile) or severe (below 10th percentile) droughts. However, since
⁴⁴³ SMI classes are calculated with respect to the local distribution and the local distribu-
⁴⁴⁴ tions do not always range from total depletion to total saturation, the 'less severe' cat-
⁴⁴⁵ egories may represent low absolute soil moisture conditions as well, while more severe
⁴⁴⁶ drought conditions in humid regions may represent higher absolute soil moisture con-
⁴⁴⁷ ditions. With rising GWL, virtually all European agricultural areas experience excep-
⁴⁴⁸ tional drought conditions during future CDHE.
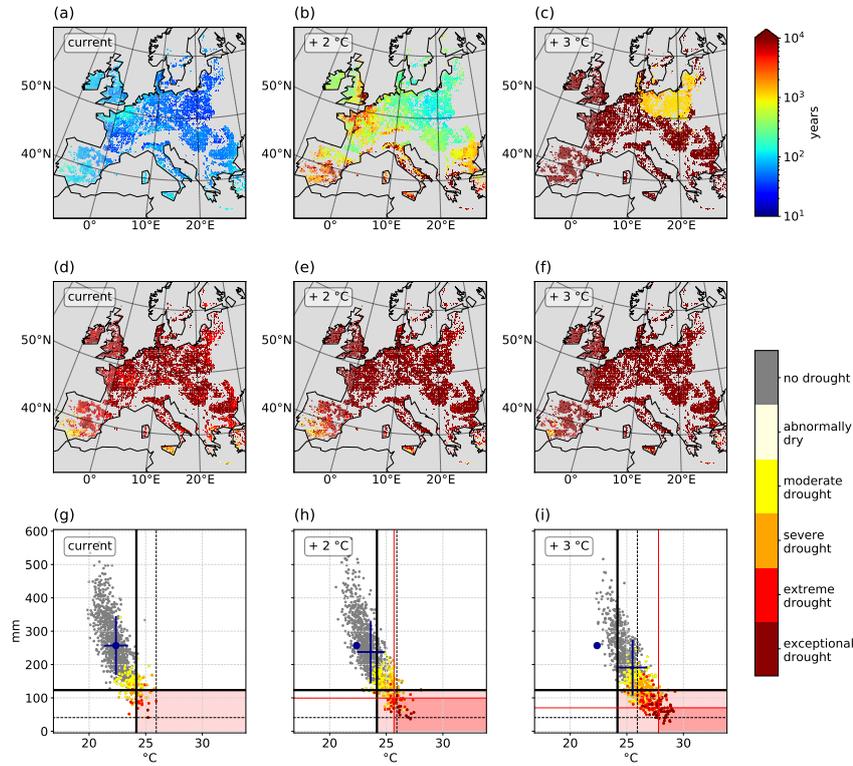
Figure 5. CDHE intensity for current, GWL2 and GWL3 conditions in European agricultural regions. (a)–(c) return period of summers with temperatures and (negative) precipitation at the GWL-specific 95th percentile (crosses of thick black lines in (g) and red lines in (h) and (i). (d)–(f) average SMI categories during all summers exceeding the GWL-specific 95th percentiles of temperature and (negative) precipitation. (g)–(i) scatter plots of summer precipitation against summer temperature for an example region (Po Valley, Northern Italy). Thick (thin) black lines show the present 5th and 95th percentiles (minimum and maximum) for precipitation and temperature, respectively. Red lines mark the 5th and 95th percentiles for GWL2 and GWL3. Light red background highlights current CDHE summers; strong red background CDHE summers for GWL2 and GWL3 percentiles. Blue dots show the current mean, crosses span one standard deviation of the respective periods for temperature and precipitation. Colors in (d)–(i) indicate soil moisture drought categories (percentiles) with respect to the current period following Zink et al. (2016).

449   Figures 5 (g)–(i) further show the relationship among soil moisture droughts and
450 compound events in an example region (Po Valley, south of the Alps) to illustrate the
451 relationship between temperature, precipitation and SMI in all summers: Summers within
452 the shaded diagram space (i.e., CDHE) are affected by more extreme SMI categories in
453 all periods; under GWL3 the majority of CDHE summers corresponds to 'exceptional
454 drought' (fig. 5 (i)). Soil moisture drought extremeness follows the distributional axis,
455 (i.e., not dominantly along the temperature or precipitation axis). With progressing global
456 climate change, distribution shifts towards warmer and drier conditions (see crosses rel-
457 ative to blue dots in (h) and (i)) increase the frequency of summers within the light red
458 shaded diagram space and also more extreme SMI. The majority of CDHE summers in
459 GWL2 and GWL3 is characterized by unprecedented temperatures (dotted black ver-
460 tical line) and numerous future events undercut the driest current summer as well (dot-
461 ted black horizontal line). This fact illustrates why this region is colored in dark red in
462 fig. 5(c). CDHE frequencies even increase with respect to the future percentiles (dark
463 red shaded diagram space) which aligns with risen tail dependence in this region (fig. 4 (h)–
464 (i)). Overall, figs. 5 (g)–(i) suggest a stable relationship of high (low) absolute temper-
465 ature (precipitation) values and soil moisture drought categories.

466   Last, how is bivariate extremeness of summers related to SMI? Figures 6 (a)–(c)
467 provide Spearman rank correlations well below -0.8 in most of European agricultural ar-
468 eas. This strong relationship implies that more extreme CDHE translate to lower mois-
469 ture conditions. Note that the correlation does not allow to conclude whether CDHE are
470 triggered or enhanced by low SMI values or vice versa, e.g., via land-atmosphere feed-
471 backs. As discussed in Manning et al. (2019) and Mukherjee et al. (2023), both is plau-
472 sible and most likely interconnected. In addition, soil moisture effects from previous sea-
473 sons or years (Felsche et al., 2023; Bastos et al., 2020) may confound the effect of CDHE
474 on soil moisture conditions of the same summer. The correlation is highly linear in all
475 GWLs (fig. 6 (d)–(f)), with a shift from low event extremeness and high soil moisture
476 in the example region during current conditions to high event extremeness and low soil
477 moisture conditions under GWL3. Again, this mirrors large projected CDHE frequency
478 increases both in absolute terms and relative to all summers of a given GWL epoch. These
479 summers hence pose a triple hazard to ecosystems and agriculture in the affected regions,
480 arising from low soil moisture, high temperature and thus high water demand for tran-
481 spiration, and low precipitation.

4 Discussion

483   In this study, we assessed frequency increases of European CDHE within a regional
484 SMILE, drivers of these increases, and the association of CDHE with soil moisture droughts.
485 The study does not provide insights in the causal directions of the SMI–CDHE relation-
486 ship, i.e., answer the question whether low soil moisture results in or from CDHE occur-
487 rence.

488   Defining CDHE based on summer precipitation percentiles comes at a cost as we
489 found in our results: In very dry regions, precipitation fluctuates on a low level. Hence,
490 due to the local JJA precipitation distribution, absolute differences between years be-
491 low or above the percentile threshold are rather small. Here, temperature variability de-
492 fines whether a CDHE occurs during a given period. Note that this is a different effect
493 than precipitation variability driving CDHE occurrence in areas where regional warm-
494 ing induces yearly exceedance of the temperature threshold. Compared to the remain-
495 der of the domain, lag effects may play a more important role in soil moisture contents
496 in areas with very low JJA precipitation sums. In general, CDHE may be more precisely
497 defined with a Survival Kendall hazard definition instead of the AND definition (see, e.g.,
498 in fig. 5 (g)–(i), Salvadori et al., 2016). However, the correlation of SMI and CDHE ex-
499 tremeness is highly linear even in our simplified event definition.
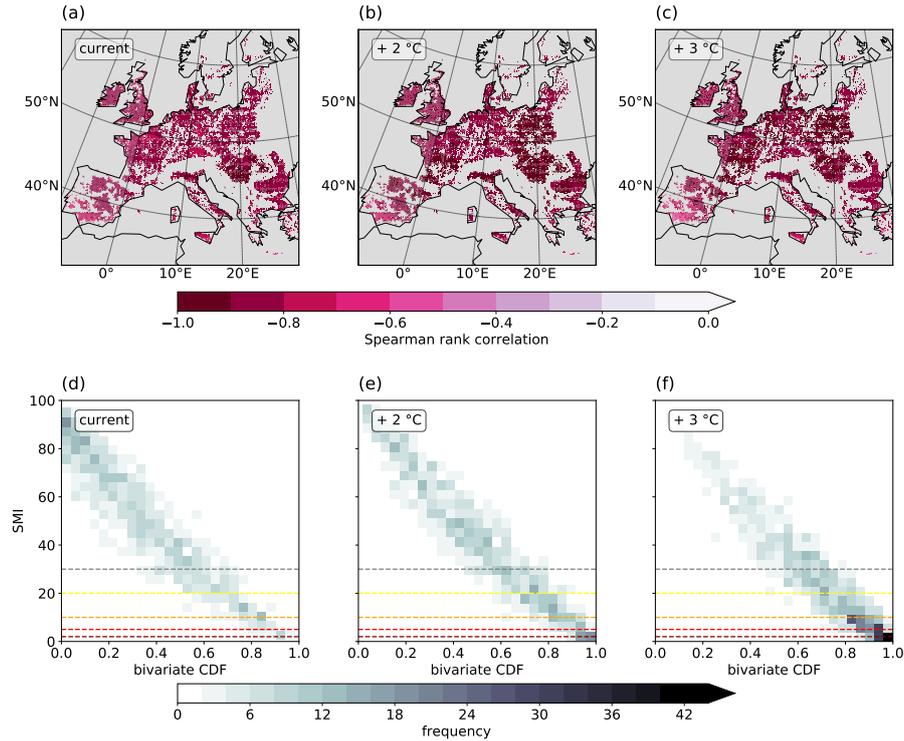
Figure 6. Relationship between CDHE extremeness (relative to conditions of the current period) and SMI values. (a)–(c) Spatially distributed Spearman rank correlation of CDHE extremeness and SMI values. (d)–(f) bivariate histograms of spatially aggregated CDHE extremeness and SMI in an example region (Po Valley, Northern Italy). Colors indicate the amount of summers in a given square. Dashed lines correspond to abnormally dry (grey), moderate drought (yellow), severe drought (orange), extreme drought (red), and exceptional drought (dark red) SMI conditions expressed as percentiles following Zink et al. (2016).

500   For explaining CDHE frequency increases, we focused on temperature and precip-
501   itation mean shifts, i.e., no variability or higher-order distributional changes which are
502   represented, e.g., in the marginal changes in François and Vrac (2023). Inspections of
503   local distributions showed that for summer CDHE variability changes only marginally
504   under GWL2 and GWL3 (e.g., fig. 5 (d)–(f)). Shifts of the joint distributions alone were
505   shown to considerably increase CDHE frequencies – not only in arid regions as done by
506   Hao et al. (2018) and Mukherjee and Mishra (2021), but also in transitional/humid re-
507   gions. Our approach is limited by the margins of the current temperature and precip-
508   itation distributions since we relate future events to the current distribution. Neverthe-
509   less, we showed that the joint increase of hot and dry extremeness can be used as a qual-
510   itative intensity measure. Beyond that, Wang et al. (2021) pointed to regionally inten-
511   sifying negative correlations between temperature and precipitation over the last decades
512   which led to an increase of CDHE, especially in the form of more heat events during droughts.
513   However, we show that not only correlation of the full distribution is projected to change
514   with rising GWL, but also the distributional tails and the entire dependence structure.
515   Bivariate dependence structures in models though require cautious consideration. Zscheischler
516   and Fischer (2020) point towards an underestimation of temperature and precipitation
517   tail dependence in CMIP5 models. This would imply a potential underestimation of CDHE.
518   A more detailed investigation into bivariate distributional characteristics in model and
519   observational data is hence advisable for locally specific assessments.

520   By reaching GWL3 in the middle of the 21st century (2042–2061) under RCP8.5,
521   the CanESM2 driving the CRCM5-LE proves to be a rather hot global climate model.
522   We therefore used a relative model- and scenario-independent measure of time, i.e., the
523   GWL, to overcome the effect of an intrinsically 'hot' global climate model with a high-
524   emission scenario. Assessing uncertainties related to this approach requires comparative
525   studies in other model SMILEs and with other scenarios. Yet, so far, there is only a very
526   limited number of regional SMILEs (typically with only few members) available (e.g.,
527   Aalbers et al., 2018).

528   As argued in Jha et al. (2023), the selection of warming levels and models explains
529   most of the uncertainty in CDHE changes over Europe. The choice of copula families
530   contributes the least in their assessment, while Zscheischler and Fischer (2020) argue that
531   event definition and copula fitting affect the final probability and therefore extremeness
532   of events. In our study, we attempted to reduce this kind of uncertainty by not focus-
533   ing on single events. Instead, the SMILE served as a basis for investigating general char-
534   acteristics of a large number of events, thus reducing the influence of outliers. Testing
535   several copula families helped to find the locally best fitting bivariate distribution. Fur-
536   ther, while in principle the SMILE provides the required size to sample low-probability
537   events ($T = 1000$), we found that future events tend to be clearly more rare than cur-
538   rent 1-in-1000 year events. Hence, even the large ensemble is insufficient for empirical
539   estimations and distributional sampling is necessary.

540   Using the SMILE though allows for a robust sampling of internal variability which
541   potentially masks dependence changes in setups with few members (Bevacqua et al., 2023).
542   In addition, differing states of large-scale atmospheric modes prevalent in single mem-
543   bers during the selected period of investigation may trigger differences in compound event
544   frequencies (Bevacqua et al., 2023). This shows the high importance of internal variabil-
545   ity in the evaluation of low-probability events and justifies the use of a SMILE.

546   While the CRCM5-LE provides high geographical detail in the spatial distribution
547   of frequency (changes), results are affected by coarse resolution geophysical inputs as is
548   visible in fig. 2: The tiling pattern resolution (1°) is coarser than the CRCM5 resolution,
549   but finer than the spatial resolution of the driving general circulation model CanESM2.
550   In central Europe, high bedrock depths (i.e., large soil column) coincide particularly well
551   with low BR in fig. 2 (b) and high evaporation in fig. 2 (c). Presumably, a large soil col-
552   umn contributes more strongly to evaporation than neighboring areas with thin soil columns.

However, this assumption requires further investigation, as well as implications on the reliability of other variables. For instance, this effect is also visible in the upper distributional tail of temperature at high temporal resolution (see also Miller et al., 2023). In spite of this, the regional SMILE allowed to highlight hotspots of event frequency (changes) and regionally varying driver dominance in high geographical detail. This is a large advantage of our study over similar analyses with coarse-resolution global SMILEs: For example, a distinction of coastal or mountainous regions would not be possible on a coarse grid since the small-scale features cannot be resolved. Hence, the derivation of relevant drivers or dependence changes would have been impeded.

Given considerable frequency increases of CDHE and their association to low soil moisture contents, we argue that the relationship between both deserves further investigation. Denissen et al. (2022) show that soil moisture limited conditions represent the new normal under a high-emission global warming scenario in that they intensify and expand in length. Since it has been shown that heatwaves, droughts or compound CDHE can be triggered by depleted soils (Fischer et al., 2007), investigating CDHE effects on soil moisture is also crucial in bringing forth the research on potential legacy effects on subsequent seasons or years (e.g., CDHE triggering subsequent CDHE mediated by prevailing soil depletion). CDHE may exert influence not only on temporally, but also spatially distant events: Li et al. (2023) show that dry soils in upwind regions may lead to propagation of events and, adding onto local land-atmosphere coupling, affect crop yields downwind of events. For example, these authors found that maize failure in Southeastern Europe and wheat failure in Italy tend to be associated with dry and hot conditions.

5   Conclusions

We find that European compound hot and dry summers are characterized by an increase of evaporative demand in the atmosphere, but with reduced evaporation in most regions, presumably due to soil moisture deficits. Mountainous regions experience increased evaporation, most likely due to higher temperatures and still dominant energy limitation of their evaporation regime. The frequency of CDHE summers increases considerably in Europe under climate change conditions. Owing to the high spatial resolution of our SMILE, we robustly identify regions in Southern France and Northern Spain as hotspots due to highest absolute increases, whereas, e.g., Southern Germany, Northern France, Southern Ireland, or the southwestern Black Sea coast can be identified as currently low-frequency areas with highest multiplication of events under climate change. Apart from Western European regions, Northern Spain and mountainous Norway, frequency increases can be mostly attributed to rising temperatures. Yet, climate change also affects the bivariate dependence structure of temperature and precipitation, fostering tail dependencies and hence the co-occurrence of dry and hot conditions. Further, events intensify with respect to the current conditions of precipitation and temperature. Soil moisture during CDHE is projected to remain extremely low under GWL2 and GWL3 in agricultural regions and shows particularly strong negative correlations with bivariate summer intensity.

This study finds newly emerging CDHE hotspots in European areas with yet unseen combinations of extremely hot and dry conditions. Regardless of the causal directions in the SMI–CDHE relationship, the tight relationship of low soil moisture and CDHE therefore poses an increasing risk to agriculture that requires consideration in adaptation planning.

This study also shows an ordering of temperature and precipitation changes in driving the frequency increases: For GWL2, temperature increase is the major driver of CDHE frequency increases. For GWL3, precipitation decrease additionally emerge as important driver (in the form of mixed contributions). Here, it would be interesting to further investigate the processes and mechanisms driving local dependence increases or decreases.

⁶⁰⁴   The regional SMILE is particularly apt for analyzing compound events in the ex-
⁶⁰⁵ treme tails of the bivariate distribution. Climate change is shown to produce events that
⁶⁰⁶ are much rarer than any observed summer, while currently extremely rare events become
⁶⁰⁷ the new normal. Fitting distributions instead of counting the summers that meet the
⁶⁰⁸ event definition criteria hence allows to avoid a saturation effect related to the maximum
⁶⁰⁹ empirical event rareness under current conditions (i.e., $T = 1000$ years). Using SMILEs,
⁶¹⁰ further research can elucidate potential benefits of increasing sample sizes in reducing
⁶¹¹ the uncertainty ranges of distribution fitting for extremely rare events.

⁶¹²   Last, we conclude that limiting global warming to $+2$ °C considerably reduces CDHE
⁶¹³ hazards in Europe, which regionally then results in half the amount of summers with ex-
⁶¹⁴ tremely low soil moisture availability. Since the risk of impacts on human systems de-
⁶¹⁵ pends on resilience structures in the affected regions (e.g., Lesk et al., 2016), hazard re-
⁶¹⁶ duction should be accompanied by fostering resilience towards CDHE effects as well.

Open Research Section

⁶¹⁸   The CRCM5-LE data used for all performed analyses is described in Leduc et al.
⁶¹⁹ (2019) and available at https://www.climex-project.org/en/data-access/
⁶²⁰   Corine land cover data is provided by the European Union, Copernicus Land Mon-
⁶²¹ itoring Service 2018, European Environment Agency (EEA) at https://land.copernicus.eu/pan-
⁶²² european/corine-land-cover
⁶²³   Codes to perform the presented analyses and obtain the figures will be shared through
⁶²⁴ a public repository upon publication of the manuscript.

Acknowledgments

References

⁶³⁸ Aalbers, E. E., Lenderink, G., van Meijgaard, E., & van den Hurk, B. J. J. M.
⁶³⁹   (2018).   Local-scale changes in mean and heavy precipitation in western eu-
⁶⁴⁰   rope, climate change or internal variability?   *Climate Dynamics*, *50*, 4745-
⁶⁴¹   4766.   Retrieved from `https://doi.org/10.1007/s00382-017-3901-9`   doi:
⁶⁴²   10.1007/s00382-017-3901-9
⁶⁴³ Afroz, M., Chen, G., & Anandhi, A. (2023). Drought- and heatwave-associated com-
⁶⁴⁴   pound extremes: A review of hotspots, variables, parameters, drivers, impacts,
⁶⁴⁵   and analysis frameworks.   *Frontiers in Earth Science*, *10*.   Retrieved from
⁶⁴⁶   `https://www.frontiersin.org/articles/10.3389/feart.2022.914437`
⁶⁴⁷   doi: 10.3389/feart.2022.914437
⁶⁴⁸ AghaKouchak, A., Cheng, L., Mazdiyasni, O., & Farahmand, A.   (2014).   Global
⁶⁴⁹   warming and changes in risk of concurrent climate extremes: Insights from
⁶⁵⁰   the 2014 california drought.   *Geophysical Research Letters*, *41*(24), 8847-8852.
⁶⁵¹   Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/`

652   `10.1002/2014GL062308` doi: https://doi.org/10.1002/2014GL062308

653   Bastos, A., Ciais, P., Friedlingstein, P., Sitch, S., Pongratz, J., Fan, L., . . . Zaehle, S.
654   (2020).   Direct and seasonal legacy effects of the 2018 heat wave and drought
655   on european ecosystem productivity. *Science Advances*, *6*(24), eaba2724. Re-
656   trieved from `https://www.science.org/doi/abs/10.1126/sciadv.aba2724`
657   doi: 10.1126/sciadv.aba2724

658   Bevacqua, E., Maraun, D., Hobæk Haff, I., Widmann, M., & Vrac, M. (2017). Multi-
659   variate statistical modelling of compound events via pair-copula constructions:
660   analysis of floods in ravenna (italy).   *Hydrology and Earth System Sciences*,
661   *21*(6), 2701–2723. Retrieved from `https://hess.copernicus.org/articles/`
662   `21/2701/2017/` doi: 10.5194/hess-21-2701-2017

663   Bevacqua, E., Suarez-Gutierrez, L., Jézéquel, A., Lehner, F., Vrac, M., Yiou, P., &
664   Zscheischler, J. (2023). Advancing research on compound weather and climate
665   events via large ensemble model simulations.   *Nature Communications*, *14*,
666   2145. Retrieved from `https://doi.org/10.1038/s41467-023-37847-5` doi:
667   10.1038/s41467-023-37847-5

668   Bowen, I. S. (1926). The ratio of heat losses by conduction and by evaporation from
669   any water surface. *Physical Review*, *27*, 779-787.

670   Brunner, M. I., Seibert, J., & Favre, A.-C. (2016). Bivariate return periods and their
671   importance for flood peak and volume estimation.   *WIREs Water*, *3*(6), 819-
672   833. Retrieved from `https://wires.onlinelibrary.wiley.com/doi/abs/10`
673   `.1002/wat2.1173` doi: https://doi.org/10.1002/wat2.1173

674   Buras, A., Rammig, A., & Zang, C. S.   (2020).   Quantifying impacts of the 2018
675   drought on european ecosystems in comparison to 2003. *Biogeosciences*, *17*(6),
676   1655–1672. Retrieved from `https://bg.copernicus.org/articles/17/1655/`
677   `2020/` doi: 10.5194/bg-17-1655-2020

678   Böhnisch, A., Felsche, E., & Ludwig, R.   (2023).   European heatwave tracks: using
679   causal discovery to detect recurring pathways in a single-regional climate
680   model large ensemble.   *Environmental Research Letters*, *18*(1), 014038.
681   Retrieved from `https://dx.doi.org/10.1088/1748-9326/aca9e3`   doi:
682   10.1088/1748-9326/aca9e3

683   Böhnisch, A., Mittermeier, M., Leduc, M., & Ludwig, R.   (2021).   Hot spots and
684   climate trends of meteorological droughts in europe–assessing the percent of
685   normal index in a single-model initial-condition large ensemble.   *Frontiers*
686   *in Water*, *3*.   Retrieved from `https://www.frontiersin.org/articles/`
687   `10.3389/frwa.2021.716621` doi: 10.3389/frwa.2021.716621

688   Ciais, P., Reichstein, N., M.and Viovy, Granier, A., Ogée, J., Allard, V., Aubinet,
689   M., . . . Valentini, R.   (2005).   Europe-wide reduction in primary productivity
690   caused by the heat and drought in 2003. *Nature*, *437*, 529–533. Retrieved from
691   `https://doi.org/10.1038/nature03972` doi: 10.1038/nature03972

692   Coles, S. (2001). *An introduction to statistical modeling of extreme values.* Springer
693   London.

694   Coles, S., Heffernan, J., & Tawn, J. (1999). Dependence measures for extreme value
695   analyses. *Extremes*, *2*, 339–365. Retrieved from `https://doi.org/10.1023/`
696   `A:1009963131610` doi: 10.1023/A:1009963131610

697   Denissen, J. M. C., Teuling, A. J., Pitman, A. J., Koirala, S., Migliavacca, M.,
698   Li, W., . . . Orth, R.   (2022).   Widespread shift from ecosystem energy to
699   water limitation with climate change.   *Nature Climate Change*, *12*, 677–
700   684. Retrieved from `https://doi.org/10.1038/s41558-022-01403-8` doi:
701   10.1038/s41558-022-01403-8

702   EEA, E. E. A.   (2020).   *Corine land cover (clc) 2018, version*
703   *2020_20u1.*   (https://land.copernicus.eu/pan-european/corine-land-
704   cover/clc2018?tab=metadata. Last accessed: 9 June 2023)

705   Felsche, E., Böhnisch, A., & Ludwig, R.   (2023).   Inter-seasonal connection of typi-
706   cal european heatwave patterns to soil moisture. *npj Climate and Atmospheric*

707   *Science*, *6*(1), 1. doi: https://doi.org/10.1038/s41612-023-00330-5

708   Fischer, E. M., Seneviratne, S. I., Lüthi, D., & Schär, C.   (2007).   Contribution of
709   land-atmosphere coupling to recent european summer heat waves.   *Geophysi-*
710   *cal Research Letters*, *34*(6). Retrieved from `https://agupubs.onlinelibrary`
711   `.wiley.com/doi/abs/10.1029/2006GL029068`   doi: https://doi.org/10.1029/
712   2006GL029068

713   François, B., & Vrac, M.   (2023).   Time of emergence of compound events: contribu-
714   tion of univariate and dependence properties.   *Natural Hazards and Earth Sys-*
715   *tem Sciences*, *23*(1), 21–44. Retrieved from `https://nhess.copernicus.org/`
716   `articles/23/21/2023/`   doi: 10.5194/nhess-23-21-2023

717   Fyfe, J. C., Derksen, C., Mudryk, L., Flato, G. M., Santer, B. D., Swart, N. C.,
718   . . . Jiao, Y.   (2017).   Large near-term projected snowpack loss over the
719   western United States.   *Nature Communications*, *8*.   Retrieved from
720   `https://doi.org/10.1038/ncomms14996`   doi: 10.1038/ncomms14996

721   Gilleland, E. (2022). *extremes: Extreme value analysis. r package version 2.1-3.* Re-
722   trieved from `https://staff.ral.ucar.edu/ericg/extRemes/`

723   Gulev, S., Thorne, P., Ahn, J., Dentener, F., Domingues, C., Gerland, S., . . . Vose,
724   R.   (2021).   Changing state of the climate system.   In V. Masson-Delmotte et
725   al. (Eds.), *Climate change 2021: The physical science basis. contribution of*
726   *working group i to the sixth assessment report of the intergovernmental panel*
727   *on climate change.*   Cambridge University Press, Cambridge, United Kingdom
728   and New York, NY, USA. doi: 10.1017/9781009157896.004

729   Hao, Z., Hao, F., Singh, V. P., & Zhang, X.   (2018).   Changes in the severity of
730   compound drought and hot extremes over global land areas.   *Environmen-*
731   *tal Research Letters*, *13*(12), 124022.   Retrieved from `https://dx.doi.org/`
732   `10.1088/1748-9326/aaee96`   doi: 10.1088/1748-9326/aaee96

733   Hauser, M., Engelbrecht, F., & Fischer, E. M.   (2022).   *Transient global warming lev-*
734   *els for CMIP5 and CMIP6.*   Zenodo.   Retrieved from `https://doi.org/10`
735   `.5281/zenodo.7390473`   doi: 10.5281/zenodo.7390473

736   Jha, S., Gudmundsson, L., & Seneviratne, S. I.   (2023).   Partitioning the uncertain-
737   ties in compound hot and dry precipitation, soil moisture, and runoff extremes
738   projections in cmip6.   *Earth's Future*, *11*(3), e2022EF003315.   Retrieved
739   from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/`
740   `2022EF003315`   (e2022EF003315 2022EF003315) doi: https://doi.org/10.1029/
741   2022EF003315

742   Kirchmeier-Young, M. C., Zwiers, F. W., & Gillett, N. P.   (2017).   Attribution of
743   extreme events in arctic sea ice extent.   *Journal of Climate*, *30*(2), 553 - 571.
744   Retrieved from `https://journals.ametsoc.org/view/journals/clim/30/2/`
745   `jcli-d-16-0412.1.xml`   doi: https://doi.org/10.1175/JCLI-D-16-0412.1

746   Kornhuber, K., Osprey, S., Coumou, D., Petri, S., Petoukhov, V., Rahmstorf, S., &
747   Gray, L. (2019). Extreme weather events in early summer 2018 connected by a
748   recurrent hemispheric wave-7 pattern. *Environmental Research Letters*, *14*(5),
749   054002. Retrieved from `https://doi.org/10.1088/1748-9326/ab13bf`   doi:
750   10.1088/1748-9326/ab13bf

751   Leduc, M., Mailhot, A., Frigon, A., Martel, J.-L., Ludwig, R., Brietzke, G. B.,
752   . . . Scinocca, J.   (2019).   The climex project: A 50-member ensemble
753   of climate change projections at 12-km resolution over europe and north-
754   eastern north america with the canadian regional climate model (crcm5).
755   *Journal of Applied Meteorology and Climatology*, *58*(4), 663–693.   doi:
756   10.1175/JAMC-D-18-0021.1

757   Leonard, M., Westra, S., Phatak, A., Lambert, M., van den Hurk, B., McInnes, K.,
758   . . . Stafford-Smith, M. (2014). A compound event framework for understand-
759   ing extreme impacts. *WIREs Climate Change*, *5*(1), 113-128. Retrieved from
760   `https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcc.252`   doi:
761   https://doi.org/10.1002/wcc.252

Lesk, C., Rowhani, P., & Ramankutty, N.   (2016).   Influence of extreme weather disasters on global crop production.   *Nature*, *529*, 84–87.   Retrieved from https://doi.org/10.1038/nature16467   doi: 10.1038/nature16467

Li, H., Keune, J., Smessaert, F., Nieto, R., Gimeno, L., & Miralles, D. G.   (2023). Land–atmosphere feedbacks contribute to crop failure in global rainfed breadbaskets.   *npj Climate and Atmospheric Science*, *6*, 51.   Retrieved from https://doi.org/10.1038/s41612-023-00375-6   doi: 10.1038/ s41612-023-00375-6

Liu, P. R., & Raftery, A. E.   (2021).   Country-based rate of emissions reductions should increase by 80% beyond nationally determined contributions to meet the 2°c target.   *Communications Earth & Environment*, *2*, 29. Retrieved from https://doi.org/10.1038/s43247-021-00097-8   doi: 10.1038/s43247-021-00097-8

Maher, N., Milinski, S., & 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.   Retrieved from https://esd.copernicus.org/articles/12/401/2021/   doi: 10.5194/esd-12-401-2021

Manning, C., Widmann, M., Bevacqua, E., Loon, A. F. V., Maraun, D., & Vrac, M. (2019).   Increased probability of compound long-duration dry and hot events in europe during summer (1950–2013).   *Environmental Research Letters*, *14*(9), 094006.   Retrieved from https://dx.doi.org/10.1088/1748-9326/ab23bf doi: 10.1088/1748-9326/ab23bf

Markonis, Y., Kumar, R., Hanel, M., Rakovec, O., Máca, P., & AghaKouchak, A. (2021).   The rise of compound warm-season droughts in europe.   *Science Advances*, *7*(6), eabb9668.   Retrieved from https://www.science.org/doi/abs/ 10.1126/sciadv.abb9668   doi: 10.1126/sciadv.abb9668

Miller, J., Böhnisch, A., Ludwig, R., & Brunner, M. I.   (2023).   Climate change impacts on regional fire weather in heterogeneous landscapes of central europe. *Natural Hazards and Earth System Sciences Discussions*, *2023*, 1–25.   Retrieved from https://nhess.copernicus.org/preprints/nhess-2023-51/ doi: 10.5194/nhess-2023-51

Miralles, D. G., Gentine, P., Seneviratne, S. I., & 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.   Retrieved from https://nyaspubs.onlinelibrary .wiley.com/doi/abs/10.1111/nyas.13912   doi: https://doi.org/10.1111/ nyas.13912

Mukherjee, S., & Mishra, A. K.   (2021).   Increase in compound drought and heatwaves in a warming world.   *Geophysical Research Letters*, *48*(1), e2020GL090617.   Retrieved from https://agupubs.onlinelibrary.wiley .com/doi/abs/10.1029/2020GL090617   (e2020GL090617 2020GL090617) doi: https://doi.org/10.1029/2020GL090617

Mukherjee, S., Mishra, A. K., Zscheischler, J., & Entekhabi, D.   (2023).   Interaction between dry and hot extremes at a global scale using a cascade modeling framework.   *Nature Communications*, *14*, 277.   Retrieved from https:// doi.org/10.1038/s41467-022-35748-7   doi: 10.1038/s41467-022-35748-7

Nagler, T., Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Erhardt, T., ... Vatter, T.   (2023).   *Vinecopula: Statistical inference of vine copulas. r package version 2.4.5.*   Retrieved from https://github.com/tnagler/ VineCopula

O, S., Bastos, A., Reichstein, M., Li, W., Denissen, J., Graefen, H., & Orth, R. (2022).   The role of climate and vegetation in regulating drought–heat extremes.   *Journal of Climate*, *35*(17), 5677 - 5685.   Retrieved from https:// journals.ametsoc.org/view/journals/clim/35/17/JCLI-D-21-0675.1.xml

817   doi: https://doi.org/10.1175/JCLI-D-21-0675.1

818   Poschlod, B., Zscheischler, J., Sillmann, J., Wood, R. R., & Ludwig, R. (2020). Cli-
819   mate change effects on hydrometeorological compound events over southern
820   norway. *Weather and Climate Extremes*, *28*, 100253. Retrieved from `https://`
821   `www.sciencedirect.com/science/article/pii/S2212094719301574`   doi:
822   https://doi.org/10.1016/j.wace.2020.100253

823   Richardson, M. I., Cowtan, K. D., Hawkins, E., & Stolpe, M. B.   (2016).   Rec-
824   onciled climate response estimates from climate models and the energy
825   budget of earth.   *Nature Climate Change*, *6*, 931-935.   Retrieved from
826   `https://doi.org/10.1038/nclimate3066` doi: 10.1038/nclimate30660

827   Rousi, E., Fink, A. H., Andersen, L. S., Becker, F. N., Beobide-Arsuaga, G., Breil,
828   M., . . . Xoplaki, E. (2023). The extremely hot and dry 2018 summer in central
829   and northern europe from a multi-faceted weather and climate perspective.
830   *Natural Hazards and Earth System Sciences*, *23*(5), 1699–1718.   Retrieved
831   from `https://nhess.copernicus.org/articles/23/1699/2023/`   doi:
832   10.5194/nhess-23-1699-2023

833   Salvadori, G., Durante, F., De Michele, C., Bernardi, M., & Petrella, L.   (2016).   A
834   multivariate copula-based framework for dealing with hazard scenarios and
835   failure probabilities.   *Water Resources Research*, *52*(5), 3701-3721.   Retrieved
836   from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/`
837   `2015WR017225` doi: https://doi.org/10.1002/2015WR017225

838   Schwalm, C. R., Glendon, S., & Duffy, P. B.   (2020).   Rcp8.5 tracks cumulative
839   co$_2$ emissions.   *Proceedings of the National Academy of Sciences*,
840   *117*(33), 19656-19657.   Retrieved from `https://www.pnas.org/doi/abs/`
841   `10.1073/pnas.2007117117` doi: 10.1073/pnas.2007117117

842   Schwingshackl, C., Hirschi, M., & Seneviratne, S. I.   (2017).   Quantifying spa-
843   tiotemporal variations of soil moisture control on surface energy balance and
844   near-surface air temperature.   *Journal of Climate*, *30*(18), 7105 - 7124.   Re-
845   trieved from `https://journals.ametsoc.org/view/journals/clim/30/18/`
846   `jcli-d-16-0727.1.xml` doi: https://doi.org/10.1175/JCLI-D-16-0727.1

847   Seneviratne, S., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., . . .
848   Zhou, B.   (2021).   Weather and climate extreme events in a changing climate.
849   In V. Masson-Delmotte et al. (Eds.), *Climate change 2021: The physical sci-*
850   *ence basis. contribution of working group i to the sixth assessment report of the*
851   *intergovernmental panel on climate change.* doi: 10.1017/9781009157896.013

852   Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett,
853   N. P., . . . Winter, B.   (2019).   The canadian earth system model version 5
854   (canesm5.0.3).   *Geoscientific Model Development*, *12*(11), 4823–4873.   Re-
855   trieved from `https://gmd.copernicus.org/articles/12/4823/2019/`   doi:
856   10.5194/gmd-12-4823-2019

857   Teuling, A. J., Hirschi, M., Ohmura, A., Wild, M., Reichstein, M., Ciais, P., . . .
858   Seneviratne, S. I.   (2009).   A regional perspective on trends in continental
859   evaporation.   *Geophysical Research Letters*, *36*(2).   Retrieved from `https://`
860   `agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008GL036584`   doi:
861   https://doi.org/10.1029/2008GL036584

862   Toreti, A., Belward, A., Perez-Dominguez, I., Naumann, G., Luterbacher, J., Cronie,
863   O., . . . Zampieri, M.   (2019).   The exceptional 2018 european water seesaw
864   calls for action on adaptation.   *Earth's Future*, *7*(6), 652-663.   Retrieved
865   from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/`
866   `2019EF001170` doi: https://doi.org/10.1029/2019EF001170

867   Trenberth, K. E., & Shea, D. J.   (2005).   Relationships between precipitation and
868   surface temperature.   *Geophysical Research Letters*, *32*(14).   Retrieved
869   from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/`
870   `2005GL022760` doi: https://doi.org/10.1029/2005GL022760

871   UNFCCC.   (2015).   *Adoption of the paris agreement. report no.*

fccc/cp/2015/l.9/rev.1.   Retrieved from `http://unfccc.int/resource/` `docs/2015/cop21/eng/l09r01.pdf`

Vogel, M. M., Zscheischler, J., Wartenburger, R., Dee, D., & Seneviratne, S. I. (2019).   Concurrent 2018 hot extremes across northern hemisphere due to human-induced climate change.   *Earth's Future*, *7*(7), 692-703.   Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/` `2019EF001189` doi: https://doi.org/10.1029/2019EF001189

von Trentini, F., Leduc, M., & 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.   Retrieved from `https://doi.org/10.1007/s00382-019-04755-8`   doi: 10.1007/s00382-019-04755-8

Wang, R., Lü, G., Ning, L., Yuan, L., & Li, L.   (2021).   Likelihood of compound dry and hot extremes increased with stronger dependence during warm seasons.   *Atmospheric Research*, *260*, 105692.   Retrieved from `https://` `www.sciencedirect.com/science/article/pii/S0169809521002489`   doi: https://doi.org/10.1016/j.atmosres.2021.105692

Zink, M., Samaniego, L., Kumar, R., Thober, S., Mai, J., Schäfer, D., & Marx, A. (2016).   The german drought monitor. *Environmental Research Letters*, *11*(7), 074002.   Retrieved from `https://dx.doi.org/10.1088/1748-9326/11/7/` `074002` doi: 10.1088/1748-9326/11/7/074002

Zscheischler, J., & Fischer, E. M.   (2020).   The record-breaking compound hot and dry 2018 growing season in germany.   *Weather and Climate Extremes*, *29*, 100270. Retrieved from `https://www.sciencedirect.com/science/article/` `pii/S2212094719302002` doi: https://doi.org/10.1016/j.wace.2020.100270

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., ... Vignotto, E.   (2020).   A typology of compound weather and climate events.   *Nature Reviews Earth & Environment*, *1*, 333–347.   Retrieved from `https://doi.org/10.1038/s43017-020-0060-z`   doi: 10.1038/s43017-020-0060-z

Zscheischler, J., Orth, R., & Seneviratne, S. I.   (2015).   A submonthly database for detecting changes in vegetation-atmosphere coupling.   *Geophysical Research Letters*, *42*(22), 9816-9824.   Retrieved from `https://agupubs` `.onlinelibrary.wiley.com/doi/abs/10.1002/2015GL066563`   doi: https://doi.org/10.1002/2015GL066563

Zscheischler, J., & Seneviratne, S. I.   (2017).   Dependence of drivers affects risks associated with compound events. *Science Advances*, *3*(6), e1700263. Retrieved from `https://www.science.org/doi/abs/10.1126/sciadv.1700263` doi: 10.1126/sciadv.1700263

6   Responses to the Research Questions

The following paragraphs summarize the main answers to the research questions outlined in Section 4. Each paragraph closes by highlighting the respective significance for this thesis.

Heatwaves and droughts turned out to be particularly suitable examples for structure analyses: Aside from both hazards being of major importance to Europe (Section 2), they can be related to strong forced trends due to anthropogenic global warming (well established for heat, shown for droughts in paper II). Therefore, structures related to these hazards can be investigated under clearly changing background conditions. Since both hazards are strongly linked mechanistically, investigating their mutual relationship may further serve as a blueprint for distilling and analyzing empirical structures (i.e., proof of concept). Lastly, from a methodological point of view, both hazards can easily be defined (based on literature), thus allowing for traceability.

Q1 | Does an RCM incorporate large-scale atmospheric regimes as present in the driving data and reproduce realistic (remote) responses?

Q1.a – General performance: Compared to reanalysis, the chain of CanESM2-LE and CRCM5-LE reproduces winter NAO patterns plausibly in paper I, with varying NAO strengths. NAO–response structures and their inter-member variability are properly represented in the global and regional SMILE. However, in amplifying signals from the driving GCM, both in high-resolution spatial patterns and aggregated time series, the regional SMILE produces NAO–response structures closer to reanalysis.

Q1.b – Nesting approach: As shown in paper I, discrepancies in SLP between the driving and driven SMILE that are larger than half the inter-annual SLP variability of the GCM arise in the southern part of the domain, with negligible errors in a west-east corridor during winter. Spatially varying discrepancies in the CRCM5-LE towards driving GCM data are non-stationary under external forcing and larger than towards driving reanalysis data.

Significance for the thesis: Q1 can be affirmed regarding the NAO in paper I. Divergences in the large-scale atmospheric field among driving data and RCM climatology
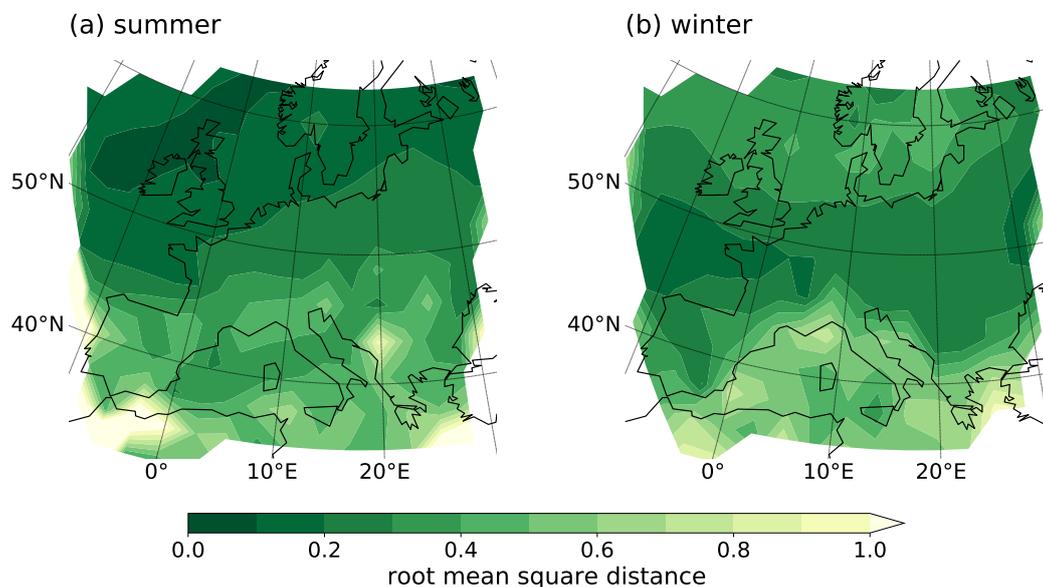
Figure 7: Root mean square distance between CRCM5-LE and CanESM2-LE SLP patterns for summer (a) and winter (b). Calculation following Böhnisch et al. (2020).

were investigated for the first time in two related SMILEs. Mostly, the discrepancies of large-scale SLP (stand-in for atmospheric conditions) among RCM and driving data are significantly smaller than inter-annual variability of the driving data. Thus, fine-scale features in the CRCM5-LE are developed against the background of consistently (i.e., widely unbiased) transferred large-scale features within the CRCM5-LE (possibly also facilitated by spectral nudging, Leduc et al. 2019). Thereby, they illustrate the RCM added value.

While paper I focused on winters, conclusions on the large-scale pattern divergences within the RCM–GCM chain can be transferred to summers using the same error metric (Figure 7): During June–August, errors increase towards the southern and continental parts of the domain, presumably due to decreasing influence of fast moving zonal inflow and hence a dominance of fine-scale local effects on seasonal climate (e.g., convective activity).

Given the strong connection of atmospheric patterns to hazards like heatwaves and droughts, the findings of Q1 increase trust in the regional analyses of papers II–IV. These conclusions are limited to the described model chain (including scenario choice). Yet, among the new generation of CMIP6 SMILEs, none were dynamically downscaled

to a regional domain so far which still renders the CMIP5-based 50-member CRCM5-LE a unique data treasure for this type of analyses.

Q2 | Can data-driven approaches robustly describe spatio-temporal structures of heatwaves and droughts hidden in a SMILE?

Q2.a – New structure discovery: In paper III, a causal discovery algorithm was successfully employed to obtain recurrent heatwave tracks in Europe. No *a priori* knowledge on track directions was passed to the algorithm. It was solely building on the temporally lagged onset of heatwaves in separated node regions, thus linking time series of the same variable at different locations. These tracks figure in both observational data and the CRCM5-LE and allow to group heatwaves, e.g., by their onset time from west to east.

Q2.b – Known structure description: Aside from the discovered structure in paper III, known multivariate relationships were also described: e.g., the directed NAO–response relationship in paper I and the mutual (tail) dependence among heat and drought during summers in paper IV. In the first case, regression coefficients were used to quantify climate variability due to NAO fluctuations. The second case allowed to rank the bivariate events and relate them to an impact variable. In papers II and III, conditional probabilities described intra-seasonal drought–non-drought sequences and heatwave strengths, respectively. These descriptions provided metrics representing the structures that can be mapped in space and compared across epochs.

Significance for the thesis: Q2 can be affirmed for the NAO–response structure (paper I), heatwaves (paper III), and CDHE (paper IV). Paper III employed a structure discovery tool in a novel way for the investigation of recurrent spatio-temporal heatwave tracks. Here, a previously unknown structure among various locations of the same variable was unraveled from a regional SMILE and later quantified by means of conditional probabilities. As suggested in Camps-Valls et al. (2023), the causal discovery algorithm proved to usefully automate the process of learning from available data. Meaningful interpretation of the causal discovery-derived structure assumes (close to) time-invariant time lags between the putative cause and its effect. Depending on the relationship under consideration, however, this is not always the case in the non-linear climate system. To consider time-dependent lags or feedbacks, methods like convergent cross-mapping may

be more appropriate (van Nes et al. 2015). Papers I and IV addressed known structures among two variables at the same location in terms of the structure concept: While both types of structure description are often used in an auxiliary way, the relationships themselves were objects of further analyses here.

Furthermore, paper III touched on a general challenge in evaluating results from data-driven derivation: Contrary to synthetic data with purposely constructed structures (e.g., Runge et al. 2019b), *a priori* knowledge on the underlying structures is scarce in real data. Evaluation may thus be conducted via plausibility checks. For instance, in paper III, general knowledge on the westward drift in central Europe allowed to argue that the prevailing west–east tracks obtained by the algorithm seem plausible. Since model and observational data agreed on this finding, the newly discovered structure was considered trustworthy. The answers to Q2 thus show the established and innovative use of data-based methods to derive and describe structures as a novel facet of hazards for subsequent investigation.

Q3 | Do internal climate variability and forced trends affect structures among variables?

Q3.a – GCM–RCM variability: Paper I gives evidence for considerable variability of the NAO–response structure in both CanESM2-LE and CRCM5-LE, with the NAO centers of action located in the GCM only (i.e., outside the RCM domain). Given small differences among GCM and RCM, seasonal winter climate variability due to NAO phases is also considered to propagate correctly from the GCM to the RCM.

Q3.b – Spatio-temporal variability: Q3.b can be readily answered with "yes". All papers showed considerable variability or uncertainty among members or samples, e.g., in the NAO–response relationship or recurrent heatwave tracks. Some heatwave tracks in paper III appeared in all members, others more rarely. This suggested a prevalence of certain tracks and allowed to assess the robustness of the tracks obtained from the (single) observed realization of reality. In paper II, internal variability was not addressed for structures, but seasonal spatial patterns, while in paper IV structure changes (see Q3.c) were so far related to the present confidence interval instead of member-derived internal variability.

Q3.c – Structures and climate change: The NAO structure strength shows no robust change, but the NAO–temperature structure decreases significantly under climate change conditions (paper I). Paper II showed clear frequency changes of inter-seasonal drought–non drought structures, but focused on spatially explicit drought frequencies. In paper III, structural trends were not evaluated. Lastly, in paper IV, structural changes over time regionally exceed the present 95 % confidence levels and are thus considered to be robust. The CDHE extremeness–SMI structure intensified slightly over time.

Significance for the thesis: It was shown that internal climate variability affects heatwave and drought structures and spatial patterns just like any other statistical measure describing climate variables, thus affirming Q3. New insights on variability and trends were gained in high geographical detail. Three different types of structures were primarily analyzed in papers I, III, and IV: the directed NAO–response structure in paper I, the likewise directed heatwave tracks in paper III, and the mutual temperature–precipitation structure (as well as the relationship between CDHE and SMI) in paper IV. While the changing inter-seasonal drought–non-drought sequences in paper II may also be considered as structures (cf. Section 7.1), the focus in this study was on spatial patterns.

Thanks to the 50 members of the high-resolution CRCM5-LE, structure variability can be assessed more robustly than by using observations or single simulations. Paper III presented for the first time evidence on internal variability of the spatio-temporal heat-wave track structure itself. This form of internal climate variability is typically masked when employing aggregate measures for heatwaves or single (e.g., observed) time series. It is thus recommended to also investigate structures of other hazards with regards to their (potential) internal variability, e.g., explicitly for temperature–precipitation relationships of CDHE.

Furthermore, depending on the structure under consideration, trends due to climate change (RCP8.5) emerge from variability (papers I and IV). The robustness of these changes could, e.g., be further quantified using signal to noise ratios or related to time or global warming via the time of emergence concept (see Section 3.3.2). Changes in the structure may also mirror altered physical relationships of the contributing variables (e.g., temperature–precipitation, NAO–response).

Using a single SMILE, though, allowed no conclusions on scenario or model structural influence on the results (e.g., land surface scheme for the temperature–precipitation structure in paper IV). However, so far there is only one regional SMILE of this size based on a single scenario (RCP8.5). To facilitate comparability towards other scenarios and models, global warming levels may be used instead of fixed epochs (as in Böhnisch et al. 2023b; Gampe et al. 2023).

All in all, the answers to Q3 illustrate the all-embracing statement by Lehner & Deser (2023), with respect to a previously underrepresented facet of heat and drought assessment.

Q4 | How can the obtained structures be verified physically?

Q4.a – Linking to physical processes: To answer this question, assumed drivers were linked to structures based on theoretical considerations: Paper III, e.g., investigated whether the composite pattern of 500 hPa geopotential height occurred specifically during the indicated heatwave tracks. By means of conditional probabilities, it was shown that composites of the assumed driver possibly constitute necessary conditions for the hazard structure under consideration, but fail to represent a sufficient condition. Associating the tracks to soil moisture did not provide a significant relation. In paper IV, processes governing the CDHE were also briefly considered (energy flux partitioning).

Q4.b – Structure components: In multivariate hazard structures, the univariate components and their dependence alter event occurrences, as shown in high geographical detail for the first time in paper IV: Temperature increases, but locally also precipitation decreases shift the joint distribution towards more frequent CDHEs summers under climate change conditions. Changes of the tail dependence regionally amplify this effect. In paper I, the weakening of NAO–response structures under global warming is associated to more NAO negative phases in the CanESM2-LE.

Significance for the thesis: While evaluation of model-based structures against observed structures was already covered by Q2, this research question aims at backing data-based structures with mechanistic processes. In papers I, III and IV, the attempts to physically verify the structures constituted of (i) investigating simultaneously occurring composite patterns of variables which are known to be related to the structures or parts of them and (ii) recurring to the variables that are linked in the structures.

When relating a structure that is based solely on data to a driver, thorough and theory-guided reasoning is pivotal. In case of paper IV, trends of the compound event components were investigated, though not explicitly regarding their fractional contributions. In addition to the components, the structure was shown to change (see also in paper I, the NAO–response structure changes). In other cases, analyses of composites may be useful in explaining a phenomenon or structure (e.g., the NAO SLP gradient resulting in flow changes onto Europe and resulting temperature/precipitation changes; heatwaves following the tracks of high pressure systems due to enhanced heating below them; energy fluxes during CDHEs). However, they do not necessarily allow to predict the phenomenon: While for the NAO, composites were based on the driver (i.e., positive/negative NAO phase) and resulting temperature/precipitation responses are investigated, the driving 500 hPa geopotential height patterns of heatwave tracks were selected conditioned on the heatwave occurrence. This illustrates some caveats on composite use raised by Boschat et al. (2016): First, while a precisely defined associated pattern may occur simultaneously to the hazard, it is unclear whether it also occurs without the hazard or with a different hazard. Second, the association is often not mutual, since associated patterns are conditioned on the occurrence of the phenomenon in question. Lastly, the user by design makes a choice of which driver they investigate.

Structures are also highly dependent on the data source. In case of paper I, the particular CanESM2-LE evolution of the NAO under climate change (negative NAO phase prevalence, most likely linked to early sea ice degradation) affects the regression coefficients which quantify the NAO–response structure. Repeating this study in one of the SMILEs of McKenna & Maycock (2021) that project mostly positive NAO phases (with some uncertainty) may result in a different NAO–response structure evolution.

While the answers to Q4 provide some suggestions for explaining structure-related findings, they are not exhaustive: Among others, they address associations, but not full mechanistic explanations of single process steps.

Q5 | How can complex results of SMILE analyses regarding internal variability and extreme events be conveyed effectively?

Q5.a – Regional hot spots: Hot spots mark regions with above-average changes or relevance to a question. While a (physical) geographer's first instinct is to map hazards

in space, spatial aggregation showed that hot spots allow purposeful reduction of specific SMILE data dimensions in papers I–IV: (i) comparing the NAO–response relationship over members and time among GCM and RCM in paper I, (ii) PNI evolution across categories, members, time, and space in paper II, (iii) defining coherent heatwave core regions in Felsche et al. (2023) to connect in paper III, and (iv) exemplarily conceptualizing the multivariate relationship of temperature, precipitation, and soil moisture droughts in paper IV.

Q5.b – Result robustness: In paper I, maps of ensemble averages and standard deviations were used to retain the climate signal and variability. Paper II merged both in signal maps (Pfeifer et al. 2015) of a regional SMILE to highlight areas of particularly robust drought category increases. Heatwave tracks in paper III were based on a robust structure derived from all members, but weighted by their representation within the members. In papers III and IV, the event rareness required all members to be pooled for robust results, such that here the "member axis" was oppressed.

Q5.c – Result Presentation: In addition to basing study figures on established types for spatially explicit mapping of change signals considering member divergences (signal maps, Pfeifer et al. 2015) or iconic charts for spatially and cross-member averaged information in European regions (*drying stripes* based on the *warming stripes* by Hawkins 2018), paper II aimed at providing results on droughts in an easy-to-communicate way and mentioned this goal explicitly. For outreach purposes, a press release accompanied the study. In papers I, III and IV, dimensions were investigated pairwise (member, spatial, temporal) to enhance readability.

Significance for the thesis: Structures are abstract concepts, but represent relevant hazard properties (e.g., spatio-temporal tracks, driver–response relationships, multivariate events). Therefore, a clear representation of structures, their variability and change signals is of utmost importance to avoid misinterpretation. Hot spots may be derived from pre-selected regions or from patterns of the spatially mapped variable. In a SMILE context, they summarize multiple dimensions and highlight regions that merit dedicated investigation at higher detail. Therefore, they are recommended for conveying SMILE results.

Papers I–IV included attempts to represent spatially distributed structures, but paper II specifically took the perspective of challenging common chart types to overcome "averag-

ing out" of hazard properties along the member axis. In doing so, paper II reached both the broad public (interview requests to both first authors) and scientists from different disciplines interested in general drought trends (cf. citations of paper II).

While creative chart design may raise attention, convey information in a surprising way, and adjust conventional illustration to new data types, structures, and research questions, its foremost goal remains to inform in a scientific sound and clear way about study results.

Research Question: Transient Changes of Actionable Data-Based Structures

Large-scale atmospheric pattern variability propagates from the driving CanESM2-LE to the regional CRCM5-LE, though with regionally varying errors (Q1). Against this background, relationships among variables are extracted from the regional SMILE, both known and unknown structures of droughts and heatwaves. For the derivation of the latter, a data-driven causal discovery algorithm has proven useful (Q2). The structures, as well as spatial patterns of (potentially) driving variables, are subject to internal climate variability and may change over time. It is thus useful and necessary to employ large samples, e.g., SMILEs, for their (change) assessment (Q3). To verify these data structures, investigating atmospheric drivers and structure components is useful, but causal interpretations require caution (Q4). Lastly, as these relationships are abstract, they call for good strategies for (intra-)science communication (Q5). Across all questions, papers I–IV showed meaningful applications for structures. Specific implications from these structures are discussed in more detail in Section 7.1. Overall, the implicit research hypothesis on structures being actionable tools for regional climate change assessment thus proved its value for the considered region and hazards.

Papers I–IV illustrate that under certain conditions data-driven approaches are suitable to infer knowledge on recurrent structures among heatwaves, droughts, or CDHE. These conditions include (i) a reasonable conception of the potential structure, e.g., inferred from observations or *a priori* knowledge, (ii) a precise definition of the structure under consideration and their components, (iii) a large sample of cases pertaining to the structure for robustly sampling its features and allowing for its variability to be captured, (iv) thorough plausibility checks in observation(-based) data and evaluation of model data,

and (v) a profound hypothesis and reasoning for linking the structure components and linking the structure to potential drivers or impacts. The last point was addressed in all papers, sometimes by referring to relevant literature. When associating composites, the relationship between driver and hazard should be evaluated by, e.g., using conditional probabilities as in Böhnisch et al. (2023a). It is generally recommended to first derive a structure in observational data before investigating models to avoid simply reproducing model structural (conceptual) properties.

Depending on the means to extract a structure, conclusions can be drawn on associations, mutual or directed statistical relationships, and their significance. Yet, they are conditional on user choices of at least the variables of interest (sometimes also the direction of relationships or time lags) and metric to quantify the structure (e.g., full distribution versus tail dependence for CDHE). Various assumptions about the data are required (see Runge 2018; Runge et al. 2019a, for a detailed discussion), among which causal sufficiency is of particular importance: the selection of all variables of relevance to a certain research question. For instance, too many variables may decrease explaining effect sizes (Runge et al. 2019b; Bahrenberg et al. 2008), whereas ignoring variables may bias the interpretation (Granger 1969; Runge et al. 2015). In addition, when using, e.g., DAG-based methods, the derivation of a given causal structure may be ambiguous because several models may fit the data (e.g., Bahrenberg et al. 2008), stressing the need for theoretical knowledge on the relationship.

Further, all studies highly rely on the assumption of data faithfully reproducing the (real physical) structure of interest. This refers to observational (e.g., correct temporal or spatial sampling to capture the underlying processes; measurement errors) and model data alike (e.g., correct process implementation or input data). Different structures in simulations and observations may be attributed to diverging process representation or internal variability. Moreover, when evaluating structures against observations, the question of whether the observed record provides a best-guess representation of a structure (e.g., of how temperature responds to NAO behavior) arises. The period under consideration could represent a biased estimate of the structure (e.g., too strong compared to other plausible historical "storylines" that could have evolved within the spread of internal variability, Lehner & Deser 2023). Potential reasons include the superimposed influence of particular atmospheric patterns (e.g., Comas-Bru & McDermott 2014). For evalu-

ation in the contexts of this thesis, observations (or reanalysis data) were considered a benchmark for practicality reasons.

The specific results on heatwaves, droughts, and CDHE of papers I–IV depend on the data source and are thus valid for the investigated model chain, including its parameterization, land use scheme, and the chosen scenario. In particular, the CanESM2 may be considered a "hot model" owing to its high equilibrium climate sensitivity of 3.7 °C, thus resulting in particularly strong warming by the end of the 21st century (Swart et al. 2019; Suarez-Gutierrez et al. 2021). Its temperature and precipitation variability though are considered well-represented (Suarez-Gutierrez et al. 2021; Wood et al. 2021). Furthermore, papers I–IV focused on relative measures, e.g., percentiles and anomalies instead of absolute values, thence avoiding bias correction. By focusing on reference periods several decades after SMILE initialization, all members can be assumed to be independent from their initial conditions (Leduc et al. 2019; Böhnisch et al. 2020).

The regional SMILE proved particularly useful for its large sample size and high spatial resolution of heat and drought extremes. Following the results in Böhnisch et al. (2020), large-scale atmospheric patterns from the CanESM2-LE are best reproduced in the CRCM5-LE at mid-latitudes in Europe during winter (high latitudes during summers, Figure 7). Outside of these areas, the RCM influence on the representation of large-scale features increases. In general, the model data allowed to investigate physically consistent events and drivers originating from plausible climate trajectories. Two perspectives on exploiting them in a SMILE were taken in this thesis: (i) pooling all members of a period under consideration into a long time series of constant climate, (ii) treating each member as a (short) single realization of reality. While (i) is particularly suitable for finding extreme events or robust relations between multiple variables, (ii) allows to estimate internal variability of various quantities by simply examining the inter-member spread. All papers took advantage of both perspectives.

In general, papers I–IV gave insights on how to consider data-based structures, with a particular focus on heat and drought assessment over Europe within a regional SMILE. As a large advantage, those data-driven approaches to work on structures among variables are transparent (each step is traceable and reproducible), they can be tailored to individual research questions, and they are transferable to other locations, epochs, hazards, or data sets.

7 Opening a Black Box? – Chances and Challenges in Hazard Structure Analyses

Decades ago, Davidson (1976) warned to "[n]ever take an empirically-derived relationship at its face value" (p. 35). Thus, wherever data-driven approaches are applied, some restrictions remain (Section 6). Structures can be understood as a kind of black box at first: They explain how variables are connected (e.g., positively correlated), but not how the connection is established mechanistically. In that light, Section 7 discusses their implications (Section 7.1), positions this thesis' findings in geographical research (Section 7.2), and points to potential applications (Sections 7.1–7.3).

7.1 Implications from Structure Analyses

Having shown that some structures describing hazards are subject to internal variability and change transiently, what can we learn from the structures of heatwaves and droughts? To provide a range of answers, papers I–IV are re-evaluated with a focus on further implications from their structure use.

In **paper I**, the considered structure is the NAO–response relationship that was shown to vary across all CanESM2-LE and CRCM5-LE members. This structure was used to evaluate internal variability propagation among a global and a regional SMILE. Additionally, the NAO itself can be seen as a structure, described by the negative correlation of the Azores High and Icelandic Low pressure systems. This structure also differs among ensemble members (Böhnisch et al. 2020). Since the same process parameterizations are inherent to all members of a SMILE, this difference describes variations of the NAO itself due to internal variability. Hence, they raise the question whether in principle the one observed realization of NAO teleconnection strength may vary in time (i.e., if identifying internal variability with inter-annual variability) or whether it is stationary, but in principle other strengths may be possible as well[1]. If the strength of the structure itself varies strongly due to internal variability, i.e., is subject to considerable uncertainty, relating mechanisms becomes more complicated. Assuming for instance a

---

[1]Note that in this study, the NAO index was defined based on the same coordinate boxes in all members. In theory, single members could produce NAO centers of actions (i.e., the two nodes with highest anti-correlation) outside of these coordinate boxes. This could affect the index calculation and subsequently the NAO–response relationship.

weak (e.g., $r = -0.28$, lowest ensemble value in Böhnisch et al. 2020) anti-correlation of Azores High and Icelandic Low in one SMILE member and strong anti-correlation in another (e.g., $r = -0.70$, as in reanalysis data, Böhnisch et al. 2020), the first case could in the extreme rule out the existence of the NAO. Thus finding mechanisms of this NAO driving, e.g., European winter temperature, stands on rather shaky grounds in the first case compared to the second. Generally, changes in teleconnections, i.e., the structures among centers of action, may mirror large-scale regime changes (e.g., Karmouche et al. 2023a). The NAO being also a commonly used driver in heatwave or drought prediction (e.g., Felsche & Ludwig 2021), internal variability of this atmospheric mode structure should be considered by, e.g., constraining simulated NAO strengths with observation-based values, if it is used in an explanatory context.

Apart from winters, seasons with drought conditions were shown to increase considerably in Europe under a high-emissions scenario without any mitigation in **paper II** (Böhnisch et al. 2021). While this study focused on robust spatial drought characteristic changes and less on structures, the analyzed cross-seasonal relationship of subsequent dry and wet seasons may be considered as such. This relationship was described by means of conditional probabilities of having a wet winter season following a summer drought (or vice versa) across all pooled ensemble members for eight European regions. This results in robust estimates by increasing the sample, but disregards the inter-member spreads as uncertainty tools. It was shown that in the considered future period (2080–2099) more drought summers are followed by wet winters, but with large spatial differences. In principle, water availability shortages could thus be compensated by winter precipitation (considering a simplified and qualitative perspective, excluding, e.g., water storage in snow or evapotranspiration). This is most likely in regions where most of the annual precipitation falls during winter, but less so in regions with the largest share of annual precipitation during summer. If winters encounter droughts, the subsequent summers were shown to be less likely wet under future conditions. In these cases, it is likely that drought conditions aggravate during the subsequent summer season due to, e.g., seasonally enhanced evapotranspiration, potentially resulting in drought propagation into soil moisture droughts (see Section 2.1.2). Here, considering a structure of an inter-seasonal relationship between summer and winter PNI values helped to contextualize droughts and assess their implications.

For heatwaves, the structure allowed to group single events and investigate their common characteristics in **paper III** (Böhnisch et al. 2023a). Since heatwaves, just like droughts, may occur everywhere, high variability of their locations is logical. Finding typical paths thus requires consideration of deviating cases and to define when a certain event does not follow the typical path. In this study, the object of investigation were single heatwave tracks, i.e., structures rather than alternative heatwave characteristics length, frequency, or intensity. The tracks themselves were quantified by means of mutual conditional probabilities, e.g., the probability of heatwaves in central Europe being preceded by heatwaves in France and heatwaves in France being followed by heatwaves in central Europe. These values are typically moderate to high for the tracks that feature in most ensemble members. A similar approach was applied to the connection of geopotential height patterns and specific heatwave tracks. Connecting the tracks to accompanying geopotential height patterns at 500 hPa allowed to suggest potential drivers for the found tracks which are plausible based on established processes (Section 2.2). Yet, while for most tracks the indicated patterns could be considered as necessary conditions (tracks mostly occurring under these patterns), they fail to be sufficient (patterns leading always to these tracks). Alike findings are prevalent in climate science (e.g., in attribution research, Section 3.4.1) and point to "webs of causes" (e.g., Campaner 2011) instead of mono-causal relationships. Connecting hazard and driver in this way, however, fails to reveal whether heatwave propagation occurs due to local heating within the area affected per heatwave day, due to hot air advection, a combination of both, or other mechanisms.

In the case of CDHE, the structure among temperature and precipitation allowed to obtain joint properties in **paper IV** (Böhnisch et al. 2023b). Transiently changing multivariate structures are of importance when, e.g., considering multivariate bias correction for application in modeling future impacts. While the structure was, strictly speaking, not necessary to derive joint return periods as a measure of extremeness in principle, it was required to push the boundaries of empirical return periods (i.e., with 1000 summers under consideration, the most extreme event cannot exceed a return period of 1000 years). In particular, future temperatures were projected to rise extremely, regionally to the point that the "coldest" future summers resemble the hottest present (2001–2020) summers. Combined with projected summer drying and structural changes, this results in largely unprecedented CDHE, corresponding to much higher return periods than 1000

years under present conditions. In fact, regionally more than half of the future summers were shown to correspond to a present return period of more than 1000 years. Here, the empirical counting approach reached saturation because of the SMILE size. Hence, for relating future events to present conditions and differentiating beyond the empirical boundary, considering the full compound event structure (i.e., the copula) was indeed necessary. The structure thus is an important hazard facet to be considered for adaptation planning (e.g., how extreme could a CDHE get?). For communication purposes in this particular case though, it may be suitable to exclude the most extreme values. For instance, the uncertainty of return periods of more than 100 000 years, corresponding to an annual occurrence probability of less than 10^{-5}, increases while the information content for application purposes decreases.

Lastly, based on the primary assumption that physical mechanisms leave fingerprints in the data, evaluating these in simulations provides the means for advanced model evaluation. Model skill under present conditions is assumed to be indicative of future model skill. This, however, holds not true in all cases because "models can be right for the wrong reasons" (Nowack et al. 2020, p. 7). Commonly, model evaluation consists of comparing trends, mean states, or variability and spatial patterns of a given variable among model simulations and observed values. Causal discovery ambitiously aims at representing dynamical coupling mechanisms that result from physical processes. Since these are assumed to be valid also under projected future conditions, they provide more robust and precise quantities for model evaluation (Nowack et al. 2020; Runge et al. 2019a). For example, Nowack et al. (2020) found that models with better causal structure representation produce, e.g., precipitation patterns more skillfully. Adding structure evaluation to common model evaluation may attenuate the possible problem of models producing correct climatologies, but owing to insufficient or wrong model formulations (Camps-Valls et al. 2023). Additionally, Barriopedro et al. (2023) argue for using modern data-driven methods to distill relevant driver variables from simulations. SMILEs qualify as promising tools in this regard. Their evaluation could be improved by testing for their ability to produce hazard structures that are found in observations and thoroughly related to physical processes. Put more generally, this type of model evaluation tests the physical knowledge on climate that is currently encoded in climate models (Section 3.2.1, cf. Camps-Valls et al. 2023).

Discovering, describing, and evaluating structures as well as connecting them to potential causes hence contributes, as asked in the major research question, to meaningful and actionable regional climate change assessment, not least, because it provides hints where to optimize climate (but also impact) models. Models in turn may support conclusions on mechanistic relationships by simulating the impact of manipulating a supposed causal precursor variable – thus, figuratively speaking, unlatching the lid of the structure black box. So, is there a future for structure analyses in geography, or maybe already a present?

7.2   A Geographical Method

Data-driven structure analyses provide the potential for further use in geographical research. Firstly, the structures allow to connect various components (i.e., different variables), time, and space (i.e., variables at different locations) within the joint human–natural earth system. Sometimes, these links may be unobvious from observations alone while in other cases assumed relationships can be verified in the data. Secondly, transient structure changes reveal currently underrepresented facets of hazard changes. For instance, assumptions on constant temperature–precipitation relationships would confound conclusions on future extreme event frequencies and hence underestimate exposure of people or infrastructure (cf. Böhnisch et al. 2023b). Robust change assessment in turn requires considering internal variability. Lastly, said structures allow more advanced model evaluation than mere temporal or spatial averages of variables (Section 7.1). This is especially important in research areas that rely strongly on model simulations and projections.

If the structure of interest is unknown, causal discovery may be used to build the skeleton upon which a traditional quantitative geographical method as path analysis can be applied: While path analysis requires hypotheses on the direction of cause–effect relationships (e.g., information on temporal order of effects, Bahrenberg et al. 2008), causal discovery may provide such a skeleton with directed links or reveal that no directional relationship exists. In the latter case (e.g., Runge et al. 2019a), hypotheses on links, data, algorithm (application), and potentially internal variability require thorough revision. Contingent on causal sufficiency, i.e., all relevant variables being known, causal
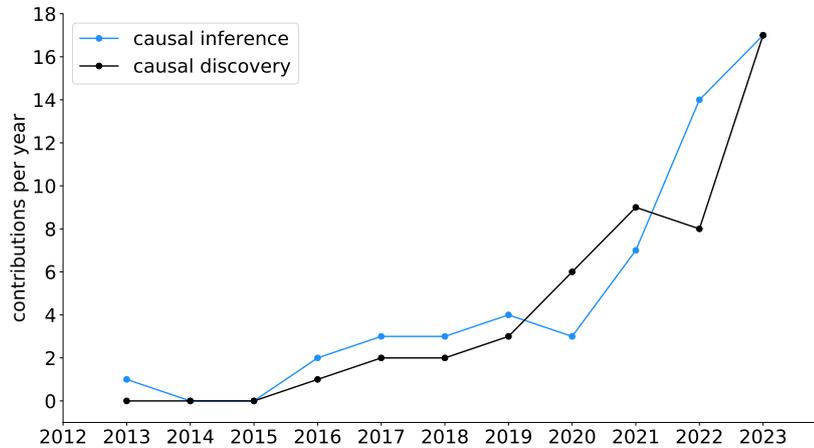
Figure 8: Geoscientific contributions employing causal inference and causal discovery at EGU General Assemblies 2013–2023 (data: search for "causal inference" and "causal discovery" in EGU General Assembly online programs, accessed: 30 May 2023).

pathway analysis may even provide an extensive picture of mechanisms mediating the causal effects of interest (Camps-Valls et al. 2023).

Causal discovery increasingly takes hold in mostly climate-related earth sciences. Figure 8 provides a snapshot of its dissemination in geosciences: The (absolute) number of abstracts, sessions, or short courses including corresponding key words at the European Geosciences Union (EGU) General Assemblies increased during 2013–2023. The majority of contributions cover climatological or meteorological topics, like teleconnections (Galytska et al. 2023; Herman & Runge 2023) or atmospheric driver identification (Miersch et al. 2023). Being representatives of a (potentially?) emerging branch in atmospheric research, these contributions often evaluate known structures or hypotheses. Largely, they are also driven by exploratory intentions (inherent to the term "causal *discovery*").

First steps into merging these approaches with established tools are currently undertaken: Applying causal discovery in SMILEs (e.g., Böhnisch et al. 2023a; Karmouche et al. 2023b) is particularly promising since the approach benefits from large sample sizes without trends in the period under consideration. Further application to questions aiming at disentangling driver–effect relationships (e.g., Almendra-Martín et al. 2022, with respect to land–atmosphere coupling), or deriving spatio-temporal (hazard) pat-

terns (as in Böhnisch et al. 2023a; Ebert-Uphoff & Deng 2012a) is thus suggested. Other propositions by Runge et al. (2015) include the detection of relations where the exact equations describing a phenomenon or process are unknown and hence no modeling is possible yet.

7.3   A Call for Investigating "Spatial Pattern Variability"

This thesis is centered around the apparent variability of multivariate structures in heatwave and drought assessment. Similarly, spatial NAO responses to the same phase were found to differ strongly between model members (Böhnisch et al. 2020). The spatial extent of heatwaves in the same region is also highly event-dependent as was found when preparing Felsche et al. (2023) and Böhnisch et al. (2023a). The purpose of this section is thus to point out briefly some applications of addressing the naturally occurring fluctuations of spatial patterns. To start, let us refrain from investigating internal variability per grid cell and address the variable extent of the coherent spatial hazard pattern. For what kind of analyses could this perspective be useful?

A quick first answer points to research questions on coupling related to local properties (e.g., heatwaves intensifying over dry soils, or afternoon rain preferentially falling over dry soils due to locally fostered convection, Taylor et al. 2012a; Miralles et al. 2019). Considering varying event extents also allows for investigations of remote spatial effects such that the hazard affects distant regions in different ways (i.e., strongly, weakly, or not) depending on its exact extent. For instance, as suggested in various studies (e.g., Miralles et al. 2019; Schumacher et al. 2019, 2022), droughts and heatwaves self-propagate by drying and heating downwind regions. Thus depending on the exact extent of a supposed heatwave, e.g., land–atmosphere coupling favors different expansion or propagation directions. In addition to dryness, orography, or land–sea conditions may as well come into play, depending on the exact heatwave location and extent. When working on climate model data, this point requires proper reproduction of the (supposedly) driving processes, including a sufficient spatial resolution to resolve the spatial features of concern.

Secondly, given that atmospheric blocking acts as precursor for co-located heatwaves and droughts (Kautz et al. 2022), its exact extent or position may trigger or intensify

local extreme events. Further, depending on the underlying topography, also events flanking the blocking pattern may evolve differently (e.g., orographic precipitation).

Thirdly, composite analysis could benefit from considering potentially diverging patterns. While composites are often used for driver association (Section 6), care must be taken when inferring physical relationships from mere associations of typical patterns and hazards as outlined by Boschat et al. (2016): Besides challenges mentioned in the answer to Q4, composites mask deviating behavior of single events which may question the robustness of the association on the one hand. Yet, on the other hand, given that instances of a phenomenon assembled in a composite vary and that this variance is used to obtain a robust pattern – what is the "typical" spatial pattern of a hazard without considering its internal spatial pattern variability?

Lastly, differing spatial extents of heatwaves or droughts due to internal variability may alter the percentage of affected population or agricultural areas and hence conclusions in impact studies. When considering climate change, hazards may extend to previously unaffected regions (see Lenton et al. 2023; Gampe et al. 2023, for changing population and area exposure to extreme heat) – the estimation of robust spatial pattern changes in turn can be assumed to require knowledge on variability during the reference period. Investigating the spatial extent adds onto local assessment in that it provides information on simultaneous occurrence of a hazard.

Examining the ability of models to reproduce the typical spatial patterns of observed hazards is thus highly advisable. Therefore, an assessment of the internal pattern variability is required. Knowing where hazard frequency or intensity increase is especially relevant for impact studies. Hence, it is necessary to rely on the ability of the model to reproduce patterns of hazard properties – including potential multivariate structures – plausibly.

8 A Framework Proposition for Going Beyond the Frontiers of Spatio-Temporal Averages
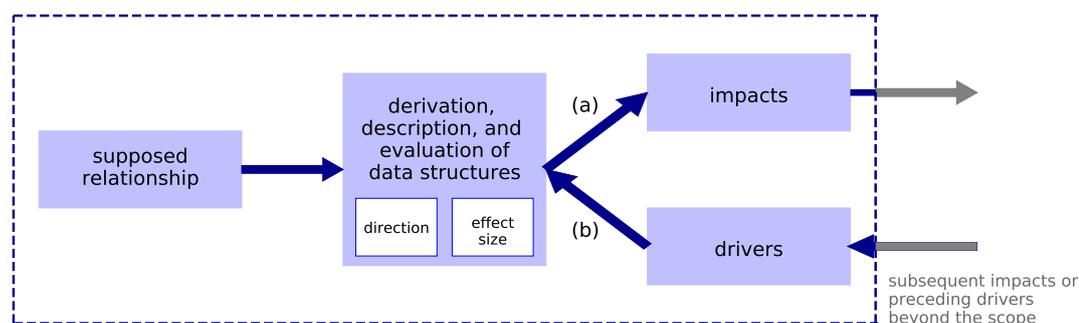

Figure 9: Conceptual framework to relate data-driven spatio-temporal structure derivation to (physical) processes. Arrows indicate the direction of the (at best mechanistic) relationship of the identified structures and their impacts (a) or drivers (b).

SMILEs clearly pave the way to investigating spatio-temporal variability for a more comprehensive picture of climate hazards like heatwaves and droughts. In general, findings from structures are not salient in common spatial or temporal averaging when describing heatwaves or droughts. Three papers of this thesis roughly followed a systematics from supposed structures among climate variables (either of different nature or spatial location) via data-driven derivation and description of said structures towards their connection with physical processes (e.g., atmospheric dynamics) or related impacts. This systematics on using structures is summarized in the proposed framework given by Figure 9. The derivation of data structures requires first an objective definition of cause–effect directions which is often assumed based on anecdotes or theory, but can be inferred by, e.g., causal discovery algorithms. In a highly complex feedback system, this may be a single partial relationship (see Seneviratne et al. 2010, for suggestions on soil moisture–precipitation interactions). Secondly, a description of effect sizes ensues (e.g., regression coefficients). If derived from observational data, the structures require at least scrutinizing investigation of whether they represent a real interrelation. In case of model data-derived structures, comparison to observations is required in addition. The empirical structures are next related to impacts or drivers, depending on the particular

research scope. Arrows (a) and (b) representing driver–hazard (including structures) and hazard–impact chains in Figure 9 require a thorough process-guided reasoning to break the glass ceiling towards mechanistic causality.

Borrowing an idea from epidemiological research, this point refers to merging two different concepts of causality: "The idea is that probabilistic evidence needs to be accounted for by an underlying mechanism before the causal claim can be established" (Russo & Williamson 2007, p. 159). Now, can relationships established by, e.g., causal discovery serve as stand-ins for probabilistic evidence? Following Williamson (2019), the answer is "yes" since the author calls for the putative cause and effect to be "probabilistically dependent conditional on potential confounders" (p. 36) which mirrors the concept of, e.g., the algorithm described in Runge et al. (2019b). For a relationship to be established reliably, it is important that the relation is found in the full population, which indicates that it is not necessarily found in a given sample even if a relationship can be assumed mechanistically (Williamson 2019). Internal climate variability may thus be a candidate for masking probabilistic evidence of causality.

For this thesis, this means: First, structures among variables may be derived by, e.g., causal discovery methods. These methods are based on the predictive causality perspective. Secondly, aligning the process steps governing these structures takes on a mechanistic causality perspective. A comprehensive chain of mechanisms or theory is sought to explain how it comes that the structure exists, as well as relating it to further impacts. But how to find mechanisms relating cause and effect? To better explain the heatwave propagation in Böhnisch et al. (2023a), e.g., analyses of air parcel trajectories can be useful: They allow to infer whether the heat within the spatial extent at a given time step originates from the heatwave extent of the previous time step or from a different location, or results from local heating (cf. Röthlisberger & Papritz 2023). As a second example, CDHE frequency increases in Böhnisch et al. (2023b) were related, among others, to structure changes. While it was possible to show that warmer summers dominate these increases, detailed knowledge on the land–atmosphere feedback driving the temperature–precipitation relationship is lacking. Here, the application of manipulative modeling (i.e., keeping all parameters but one constant) could provide the means to dive deeper into the driver analysis. Therefore, of course, models are required to adequately represent the structures of interest in the first place.

Finding the point at which (mechanistic or probabilistic) causality is considered as being established is non-trivial and depends on the concrete question (e.g., Russo & Williamson 2007). Being capable of establishing a dose-response relationship as, e.g., in health sciences (Waud 1975; EPA 2005) could be a good starting point. A basic example from climate science may be: Low "doses" of GHG emissions result in low responses of global warming. Also, at some point digging deeper for a causes becomes problematic: Either (i) we continue with infinitely chaining causes as each cause needs to be explained itself (which is called infinite regress), (ii) causes become circular, or (iii) end with dogmatic statements (*Münchhausen Trilemma*, Schurz 2021). Figure 9 acknowledges this point by the grey arrows outside the box, which represents the scope of a given research question. A pragmatic solution is thus precisely defining the problem at hand and communicating up to which point causal precursors are sought.

Analyzing data-based structures should by no means be considered as a substitute for model simulations, field observations, or laboratory experiments that help in establishing both dependence and mechanisms (cf. Kalisch et al. 2012; Runge et al. 2019a; Russo & Williamson 2007). Instead, structures are recommended as tools to complement these knowledge creating approaches by accessing and assessing relationships hidden in their data. The structures may serve as a node to tie into a web of causes (cf. Campaner 2011), i.e., to relate them mechanistically to their drivers or impacts. Structures also represent an additional facet of hazards beyond (univariate) average or variation measures. Moreover, they provide hints which variables are likely to be related in a mechanistic way themselves. Finding mechanisms then is important to rule out that these empirical relationships are spurious (Russo & Williamson 2007).

These structures often remain missed when employing temporal or spatial averages as is commonly done to overcome sampling uncertainty or internal climate variability. However, as shown here, they describe relevant and actionable features of heatwaves and droughts (as well as teleconnections), are subject to climate variability, and change over time. If SMILEs can be shown to faithfully represent physical processes and naturally occurring climate variability, they help in finding these structures among variables within their member spread.

In that sense, the framework proposed here, which attempts to narrow the gaps between data-driven relations and mechanistic explanation, may support opening the lid of the black box of purely data-driven knowledge generation in (physical) geography.

Appendix

Data sources for spatial resolution of GCMs as given in Figure 3 (accessed: 05 June 2023). For translation of model spectral resolutions to lat×lon see NCAR (2017).

SMILEs:

- Deser et al. (2020)

CMIP6:

- https://wcrp-cmip.github.io/CMIP6_CVs/docs/CMIP6_source_id.html

CMIP5:

- model overview: https://pcmdi.llnl.gov/mips/cmip5/availability.html

- Flato et al. (2013), Vicente-Serrano et al. (2022), Wu et al. (2019), von Salzen et al. (2013), Loganathan & Mahindrakar (2020), Chen et al. (2019), Si et al. (2021), Wu et al. (2020)

- https://www.metoffice.gov.uk/research/approach/modelling-systems/unified -model/climate-models/hadcm3

- https://www.nccs.nasa.gov/services/data-collections/coupled-products/ geos5-decadal

- https://www.gfdl.noaa.gov/atmospheric-model/

CMIP3:

- https://pcmdi.llnl.gov/ipcc/model_documentation/ipcc_model_documentation. php

- Gordon et al. (2002)

References

AALBERS, E. E., LENDERINK, G., VAN MEIJGAARD, E. & VAN DEN HURK, B. J. J. M. (2018): Local-scale changes in mean and heavy precipitation in Western Europe, climate change or internal variability? *Climate Dynamics*, 50, 4745–4766.

ABRAMOWITZ, G., HERGER, N., GUTMANN, E., HAMMERLING, D., KNUTTI, R., LEDUC, M., LORENZ, R., PINCUS, R. & SCHMIDT, G. A. (2019): ESD Reviews: Model dependence in multi-model climate ensembles: weighting, sub-selection and out-of-sample testing. *Earth System Dynamics*, 10(1), 91–105.

AGHAKOUCHAK, A., CHENG, L., MAZDIYASNI, O. & FARAHMAND, A. (2014): Global warming and changes in risk of concurrent climate extremes: Insights from the 2014 California drought. *Geophysical Research Letters*, 41(24), 8847–8852.

ALMENDRA-MARTÍN, L., MARTÍNEZ-FERNÁNDEZ, J., PILES, M., ÁNGEL GONZÁLEZ-ZAMORA, BENITO-VERDUGO, P. & GAONA, J. (2022): Influence of atmospheric patterns on soil moisture dynamics in Europe. *Science of The Total Environment*, 846, 157537.

ASSENG, S., SPÄNKUCH, D., HERNANDEZ-OCHOA, I. M. & LAPORTA, J. (2021): The upper temperature thresholds of life. *The Lancet Planetary Health*, 5, e378–e385.

BAHRENBERG, G., GIESE, E., NIPPER, J. & MEVENKAMP, N. (2008): Statistische Methoden in der Geographie. Band 2: Multivariate Statistik. Gebrüder Borntraeger Verlagsbuchhandlung, Berlin, Stuttgart.

BAKER, L., SHAFFREY, L. & HAWKINS, E. (2021): Has the risk of a 1976 north-west European summer drought and heatwave event increased since the 1970s because of climate change? *Quarterly Journal of the Royal Meteorological Society*, 147(741), 4143–4162.

BALAJI, V., COUVREUX, F., DESHAYES, J., GAUTRAIS, J., HOURDIN, F. & RIO, C. (2022): Are general circulation models obsolete? *Proceedings of the National Academy of Sciences*, 119(47), e2202075119.

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

BARRIOPEDRO, D., FISCHER, E. M., LUTERBACHER, J., TRIGO, R. M. & GARCÍA-HERRERA, R. (2011): The Hot Summer of 2010: Redrawing the Temperature Record Map of Europe. *Science*, 332(6026), 220–224.

BARRIOPEDRO, D., GARCÍA-HERRERA, R., LUPO, A. R. & HERNÁNDEZ, E. (2006): A Climatology of Northern Hemisphere Blocking. *Journal of Climate*, 19(6), 1042–1063.

BARRIOPEDRO, D., GARCÍA-HERRERA, R., ORDÓÑEZ, C., MIRALLES, D. G. & SALCEDO-SANZ, S. (2023): Heat Waves: Physical Understanding and Scientific Challenges. *Reviews of Geophysics*, 61(2), e2022RG000780.

BECKER, N., ULBRICH, U. & KLEIN, R. (2015): Systematic large-scale secondary circulations in a regional climate model. *Geophysical Research Letters*, 42(10), 4142–4149.

BENEDICT, J., LEE, S. & FELDSTEIN, S. (2004): Synoptic View of the North Atlantic Oscillation. *Journal of the Atmospheric Sciences*, 61(2), 121–144.

BEVACQUA, E., MARAUN, D., HOBÆK HAFF, I., WIDMANN, M. & VRAC, M. (2017): Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in Ravenna (Italy). *Hydrology and Earth System Sciences*, 21(6), 2701–2723.

BIELI, M., PFAHL, S. & WERNLI, H. (2015): A Lagrangian investigation of hot and cold temperature extremes in Europe. *Quarterly Journal of the Royal Meteorological Society*, 141(686), 98–108.

BLACK, E., BLACKBURN, M., HARRISON, G., HOSKINS, B. & METHVEN, J. (2004): Factors contributing to the summer 2003 European heatwave. *Weather*, 59(8), 217–223.

BLANUSA, M. L., LÓPEZ-ZURITA, C. J. & RASP, S. (2023): Internal variability plays a dominant role in global climate projections of temperature and precipitation extremes. *Climate Dynamics*, 61, 1931–1945.

BOSCHAT, G., SIMMONDS, I., PURICH, A., COWAN, T. & PEZZA, A. B. (2016): On the use of composite analyses to form physical hypotheses: An example from heat wave – SST associations. *Scientific Reports*, 6, 29599.

BRUNNER, M. I., SEIBERT, J. & FAVRE, A.-C. (2016): Bivariate return periods and their importance for flood peak and volume estimation. *WIREs Water*, 3(6), 819–833.

BRUNNER, M. I., VAN LOON, A. F. & STAHL, K. (2022): Moderate and Severe Hydrological Droughts in Europe Differ in Their Hydrometeorological Drivers. *Water Resources Research*, 58(10), e2022WR032871.

BRUYÈRE, C. L., BUCKLEY, B., JAYE, A. B., DONE, J. M., LEPLASTRIER, M., ALDRIDGE, J., CHAN, P., TOWLER, E. & GE, M. (2022): Using large climate model ensembles to assess historical and future tropical cyclone activity along the Australian east coast. *Weather and Climate Extremes*, 38, 100507.

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

BÖHNISCH, A., FELSCHE, E., MITTERMEIER, M., POSCHLOD, B. & LUDWIG, R. (2023b): Future hotspots of compound dry and hot summers emerge in European agricultural areas. ESS Open Archive [preprint], submitted to Earth's Future.

BÖHNISCH, A., LUDWIG, R. & LEDUC, M. (2020): Using a nested single-model large ensemble to assess the internal variability of the North Atlantic Oscillation and its climatic implications for central Europe. *Earth System Dynamics*, 11(3), 617–640.

BÖHNISCH, A., MITTERMEIER, M., LEDUC, M. & 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.

CAMPANER, R. (2011): Mechanistic causality and counterfactual-manipulative causality: recent insights from philosophy of science. *Journal of epidemiology and community health*, 65, 1070–1074.

CAMPS-VALLS, G., GERHARDUS, A., NINAD, U., VARANDO, G., MARTIUS, G., BALAGUER-BALLESTER, E., VINUESA, R., DIAZ, E., ZANNA, L. & RUNGE, J. (2023): Discovering causal relations and equations from data. *Physics Reports*, 1044, 1–68.

CANNON, A. J. (2018): Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables. *Climate Dynamics*, 50, 31–49.

CARVALHO, D., RAFAEL, S., MONTEIRO, A., RODRIGUES, V., LOPES, M. & ROCHA, A. (2022): How well have CMIP3, CMIP5 and CMIP6 future climate projections portrayed the recently observed warming. *Scientific Reports*, 12, 11983.

CATTIAUX, J., DOUVILLE, H. & PEINGS, Y. (2013): European temperatures in CMIP5: origins of present-day biases and future uncertainties. *Climate Dynamics*, 41, 2889–2907.

CHEN, S., WU, R., CHEN, W. & SONG, L. (2019): Performance of the CMIP5 models in simulating the Arctic Oscillation during boreal spring. *Climate Dynamics*, 53, 2083–2101.

CIAIS, P., REICHSTEIN, N., M.AND VIOVY, GRANIER, A., OGÉE, J., ALLARD, V., AUBINET, M., BUCHMANN, N., BERNHOFER, C., CARRARA, A., CHEVALLIER, F., DE NOBLET, N., FRIEND, A. D., FRIEDLINGSTEIN, P., GRÜNWALD, T., HEINESCH, B., KERONEN, P., KNOHL, A., KRINNER, G., LOUSTAU, D., MANCA, G., MATTEUCCI, G., MIGLIETTA, F., OURCIVAL, J. M., PAPALE, D., PILEGAARD, K., RAMBAL, S., SEUFERT, G., SOUSSANA, J. F., SANZ, M. J., SCHULZE, E. D., VESALA, T. & VALENTINI, R. (2005): Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature*, 437, 529–533.

CLARK, W. C. (1985): Scales of climate impacts. *Climatic Change*, 7, 5–27.

COMAS-BRU, L. & MCDERMOTT, F. (2014): Impacts of the EA and SCA patterns on the European twentieth century NAO–winter climate relationship. *Quarterly Journal of the Royal Meteorological Society*, 140(679), 354–363.

CONRADT, T., ENGELHARDT, H., MENZ, C., VICENTE-SERRANO, S. M., FARIZO, B. A., PEÑA-ANGULO, D., DOMÍNGUEZ-CASTRO, F., EKLUNDH, L., JIN, H., BOINCEAN, B., MURPHY, C. & LÓPEZ-MORENO, J. I. (2023): Cross-sectoral impacts of the 2018–2019 Central European drought and climate resilience in the German part of the Elbe River basin. *Regional Environmental Change*, 23, 32.

COUMOU, D., DI CAPUA, G., VAVRUS, S., WANG, L. & WANG, S. (2018): The influence of Arctic amplification on mid-latitude summer circulation. *Nature Communications*, 9, 2959.

CRAUSBAY, S. D., RAMIREZ, A. R., CARTER, S. L., CROSS, M. S., HALL, K. R., BATHKE, D. J., BETANCOURT, J. L., COLT, S., CRAVENS, A. E., DALTON, M. S., DUNHAM, J. B., HAY, L. E., HAYES, M. J., MCEVOY, J., MCNUTT, C. A., MORITZ, M. A., NISLOW, K. H., RAHEEM, N. & SANFORD, T. (2017): Defining Ecological Drought for the Twenty-First Century. *Bulletin of the American Meteorological Society*, 98(12), 2543 – 2550.

VAN DAALEN, K. R., ROMANELLO, M., ROCKLÖV, J., SEMENZA, J. C., TONNE, C., MARKANDYA, A., DASANDI, N., JANKIN, S., ACHEBAK, H., BALLESTER, J., BECHARA, H., CALLAGHAN, M. W., CHAMBERS, J., DASGUPTA, S., DRUMMOND, P., FAROOQ, Z., GASPARYAN, O., GONZALEZ-REVIRIEGO, N., HAMILTON, I., HÄNNINEN, R., KAZMIERCZAK, A., KENDROVSKI, V., KENNARD, H., KIESEWETTER, G., LLOYD, S. J., LOTTO BATISTA, M., MARTINEZ-URTAZA, J., MILÀ, C., MINX, J. C., NIEUWENHUIJSEN, M., PALAMARCHUK, J., QUIJAL-ZAMORANO, M., ROBINSON, E. J. Z., SCAMMAN, D., SCHMOLL, O., SEWE, M. O., SJÖDIN, H., SOFIEV, M., SOLARAJU-MURALI, B., SPRINGMANN, M., TRIÑANES, J., ANTO, J. M., NILSSON, M. & LOWE, R. (2022): The 2022 Europe report of the Lancet Countdown on health and climate change: towards a climate resilient future. *The Lancet Public Health*, 7, e942–e965.

DAVIDSON, N. (1976): Causal inferences from dichotomous variables. In: CATMOG (Concepts and Techniques in Modern Geography), volume 9, Boston, 37 pp.

DE BONO, A., PEDUZZI, P., KLUSER, S. & GIULIANI, G. (2004): Impacts of Summer 2003 Heat Wave in Europe. Environment Alert Bulletin 2.

DELWORTH, T., ZENG, F., VECCHI, G., YANG, X., ZHANG, L. & ZHANG, R. (2016): The North Atlantic Oscillation as a driver of rapid climate change in the Northern Hemisphere. *Nature Geoscience*, 9(7), 509–512.

DELWORTH, T. L. & ZENG, F. (2016): The Impact of the North Atlantic Oscillation on Climate through Its Influence on the Atlantic Meridional Overturning Circulation. *Journal of Climate*, 29(3), 941 – 962.

DEMORY, M.-E., BERTHOU, S., FERNÁNDEZ, J., SØRLAND, S. L., BROGLI, R., ROBERTS, M. J., BEYERLE, U., SEDDON, J., HAARSMA, R., SCHÄR, C., BUONOMO, E., CHRISTENSEN, O. B., CIARLO, J. M., FEALY, R., NIKULIN, G., PEANO, D., PUTRASAHAN, D., ROBERTS, C. D., SENAN, R., STEGER, C., TEICHMANN, C. & VAUTARD, R. (2020): European daily precipitation according to EURO-CORDEX regional climate models (RCMs) and high-resolution global climate models (GCMs) from the High-Resolution Model Intercomparison Project (HighResMIP). *Geoscientific Model Development*, 13(11), 5485–5506.

DENG, Y. & EBERT-UPHOFF, I. (2014): Weakening of atmospheric information flow in a warming climate in the Community Climate System Model. *Geophysical Research Letters*, 41(1), 193–200.

DENISSEN, J. M. C., ORTH, R., WOUTERS, H., MIRALLES, D. G., VAN HEERWAARDEN, C. C., DE ARELLANO, J. V.-G. & TEULING, A. J. (2021): Soil moisture signature in global weather balloon soundings. *npj Climate and Atmospheric Science*, 4, 13.

DENISSEN, J. M. C., TEULING, A. J., PITMAN, A. J., KOIRALA, S., MIGLIAVACCA, M., LI, W., REICHSTEIN, M., WINKLER, A. J., ZHAN, C. & ORTH, R. (2022): Widespread shift from ecosystem energy to water limitation with climate change. *Nature Climate Change*, 12, 677–684.

DESER, C., HURRELL, J. & PHILLIPS, A. (2017): The role of the North Atlantic Oscillation in European climate projections. *Climate Dynamics*, 49, 3141–3157.

DESER, C., LEHNER, F., RODGERS, K. B., AULT, T., DELWORTH, T. L., DINEZIO, P. N., FIORE, A., FRANKIGNOUL, C., FYFE, J. C., HORTON, D. E., KAY, J. E., KNUTTI, R., LOVENDUSKI, N. S., MAROTZKE, J., MCKINNON, K. A., MINOBE, S., RANDERSON, J., SCREEN, J. A., SIMPSON, I. R. & TING, M. (2020): Insights from Earth system model initial-condition large ensembles and future prospects. *Nature Climate Change*, 10, 277–286.

DESER, C., PHILLIPS, A., BOURDETTE, V. & TENG, H. (2012): Uncertainty in climate change projections: the role of internal variability. *Climate Dynamics*, 38, 527–546.

DI CAPUA, G., KRETSCHMER, M., DONNER, R. V., VAN DEN HURK, B., VELLORE, R., KRISHNAN, R. & COUMOU, D. (2020): Tropical and mid-latitude teleconnections interacting with the Indian summer monsoon rainfall: a theory-guided causal effect network approach. *Earth System Dynamics*, 11(1), 17–34.

DIACONESCU, E. P., LAPRISE, R. & SUSHAMA, L. (2007): The impact of lateral boundary data errors on the simulated climate of a nested regional climate model. *Climate Dynamics*, 28, 333–350.

DUDHIA, J. (2014): A history of mesoscale model development. *Asia-Pacific Journal of Atmospheric Sciences*, 50, 121–131.

EBERT-UPHOFF, I. & DENG, Y. (2010): Causal Discovery Methods for Climate Networks. Technical report, Georgia Institute of Technology, School of Mechanical Engineering, Atlanta. Res. Rep. GT-ME-2010-001.

EBERT-UPHOFF, I. & DENG, Y. (2012a): A new type of climate network based on probabilistic graphical models: Results of boreal winter versus summer. *Geophysical Research Letters*, 39(19).

EBERT-UPHOFF, I. & DENG, Y. (2012b): Causal Discovery for Climate Research Using Graphical Models. *Journal of Climate*, 25(17), 5648 – 5665.

ECMWF (2023): Surface air temperature for October 2023. Climate bulletin. European Center for Medium-Range Weather Forecasts (ECMWF), Copernicus, European Commussion. accessed: 10 November 2023.

EDEN, J. M., WIDMANN, M., MARAUN, D. & VRAC, M. (2014): Comparison of GCM- and RCM-simulated precipitation following stochastic postprocessing. *Journal of Geophysical Research: Atmospheres*, 119(19), 11,040–11,053.

EMS (2023): European Drought Observatory. Copernicus Emergency Management Service (EMS), accessed: 02 July 2023.

EPA (2005): Guidelines for Carcinogen Risk Assessment. United States Environmental Protection Agency (EPA). Technical report, Risk Assessment Forum U.S. Environmental Protection Agency, Washington, DC. EPA/630/P-03/001F.

EYRING, V., BONY, S., MEEHL, G. A., SENIOR, C. A., STEVENS, B., STOUFFER, R. J. & TAYLOR, K. E. (2016): Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5), 1937–1958.

FALZOI, S., GLEESON, E., LAMBKIN, K., ZIMMERMANN, J., MARWAHA, R., O'HARA, R., GREEN, S. & FRATIANNI, S. (2019): Analysis of the severe drought in Ireland in 2018. *weather*, 74(11), 368–373.

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.

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.

FISCHER, E. & KNUTTI, R. (2015): Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. *Nature Climate Change*, 5, 560–564.

FISCHER, E. M., SENEVIRATNE, S. I., LÜTHI, D. & SCHÄR, C. (2007): Contribution of land-atmosphere coupling to recent European summer heat waves. *Geophysical Research Letters*, 34(6).

FISCHER, E. M., SIPPEL, S. & KNUTTI, R. (2021): Increasing probability of record-shattering climate extremes. *Nature Climate Change*, 11, 689–695.

FLATO, G., MAROTZKE, J., ABIODUN, B., BRACONNOT, P., CHOU, S., COLLINS, W., COX, P., DRIOUECH, F., EMORI, S., EYRING, V., FOREST, C., GLECKLER, P., GUILYARDI, E., JAKOB, C., KATTSOV, V., C., R. & M., R. (2013): Evaluation of Climate Models. In: STOCKER, T., QIN, D., PLATTNER, G.-K., TIGNOR, M., ALLEN, S., BOSCHUNG, J., NAUELS, A., XIA, Y., V., B. & MIDGLEY, P. (eds.), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 741–866.

FOLLAND, C. K., KNIGHT, J., LINDERHOLM, H. W., FEREDAY, D., INESON, S. & HURRELL, J. W. (2009): The Summer North Atlantic Oscillation: Past, Present, and Future. *Journal of Climate*, 22(5), 1082 – 1103.

FORSTER, P., STORELVMO, T., ARMOUR, K., COLLINS, W., DUFRESNE, J.-L., FRAME, D., LUNT, D., MAURITSEN, T., PALMER, M., WATANABE, M., WILD, M. & ZHANG, H. (2021): The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity. In: MASSON-DELMOTTE, V., ZHAI, P., PIRANI, A., CONNORS, S., PÉAN, C., BERGER, S., CAUD, N., CHEN, Y., GOLDFARB, L., GOMIS, M., HUANG, M., LEITZELL, K., LONNOY, E., MATTHEWS, J., MAYCOCK, T., WATERFIELD, T., YELEKÇI, O., YU, R. & ZHOU, B. (eds.), Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 923–1054.

FYFE, J. C., DERKSEN, C., MUDRYK, L., FLATO, G. M., SANTER, B. D., SWART, N. C., MOLOTCH, N. P., ZHANG, X., WAN, H., ARORA, V. K., SCINOCCA, J. & JIAO, Y. (2017): Large near-term projected snowpack loss over the western United States. *Nature Communications*, 8, 14996.

GALYTSKA, E., WEIGEL, K., HANDORF, D., JAISER, R., KÖHLER, R., RUNGE, J. & EYRING, V. (2023): Causal model evaluation of Arctic-midlatitude process during the boreal cold season in CMIP6. EGU General Assembly 2023, Vienna, Austria, 24-28 Apr 2023, EGU23-11159.

GAMPE, D., SCHWINGSHACKL, C., BÖHNISCH, A., MITTERMEIER, M., SANDSTAD, M. & WOOD, R. (2023): Applying global warming levels of emergence to highlight the increasing population exposure to temperature and precipitation extremes. ESS Open Archive [preprint], submitted to Earth System Dynamics.

GIORGI, F., JONES, C. & ASRAR, G. R. (2009): Addressing climate information needs at the regional level: the CORDEX framework. *WMO Bulletin*, 58(3), 175–183.

GIORGI, F. & MEARNS, L. O. (1999): Introduction to special section: Regional Climate Modeling Revisited. *Journal of Geophysical Research: Atmospheres*, 104(D6), 6335–6352.

GORDON, H., ROTSTAYN, L., MCGREGOR, J., DIX, M., KOWALCZYK, E., O'FARRELL, S., WATERMAN, L., HIRST, A., WILSON, S., COLLIER, M., WATTERSON, I. & ELLIOTT, T. (2002): The CSIRO Mk3 climate system model. *Commonwealth Scientific and Industrial Research Organisation Atmospheric Research*, 60.

GRANGER, C. W. J. (1969): Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3), 424–438.

HAARSMA, R. J., ROBERTS, M. J., VIDALE, P. L., SENIOR, C. A., BELLUCCI, A., BAO, Q., CHANG, P., CORTI, S., FUČKAR, N. S., GUEMAS, V., VON HARDENBERG, J., HAZELEGER, W., KODAMA, C., KOENIGK, T., LEUNG, L. R., LU, J., LUO, J.-J., MAO, J., MIZIELINSKI, M. S., MIZUTA, R., NOBRE, P., SATOH, M., SCOCCIMARRO, E., SEMMLER, T., SMALL, J. & VON STORCH, J.-S. (2016): High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. *Geoscientific Model Development*, 9(11), 4185–4208.

HANNART, A., PEARL, J., OTTO, F. E. L., NAVEAU, P. & GHIL, M. (2016): Causal Counterfactual Theory for the Attribution of Weather and Climate-Related Events. *Bulletin of the American Meteorological Society*, 97(1), 99 – 110.

HAUSFATHER, Z. & PETERS, G. (2020a): Emissions – the 'business as usual' story is misleading. *Nature*, 577, 618–620.

HAUSFATHER, Z. & PETERS, G. P. (2020b): RCP8.5 is a problematic scenario for near-term emissions. *Proceedings of the National Academy of Sciences*, 117(45), 27791–27792.

HAWKINS, E. (2018): Warming stripes. Https://www.climate-lab-book.ac.uk/2018/warming-stripes/, accessed: 10 June 2023.

HAWKINS, E., SMITH, R. S., GREGORY, J. M. & STAINFORTH, D. A. (2016): Irreducible uncertainty in near-term climate projections. *Climate Dynamics*, 46, 3807–3819.

HAWKINS, E. & SUTTON, R. (2009): The Potential to Narrow Uncertainty in Regional Climate Predictions. *Bulletin of the American Meteorological Society*, 90(8), 1095 – 1108.

HAWKINS, E. & SUTTON, R. (2011): The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dynamics*, 37, 407–418.

HERMAN, R. & RUNGE, J. (2023): Using Causal Discovery to Clarify Observed and Simulated Relationships Between ENSO and Other Ocean Basins, EGU General Assembly 2023. EGU General Assembly 2023, Vienna, Austria, 24-28 Apr 2023, EGU23-13335.

HESS, P. & BREZOWSKY, H. (1969): Katalog der Großwetterlagen Europas. 2. neu bearbeitete und ergänzte Auflage. *Ber. Dt. Wetterd.*, 5.

HOFFMANN, P., LEHMANN, J., FALLAH, B. & HATTERMANN, F. F. (2021): Atmosphere similarity patterns in boreal summer show an increase of persistent weather conditions connected to hydro-climatic risks. *Scientific Reports*, 11, 22893.

HOFFMANN, P. & SPEKAT, A. (2021): Identification of possible dynamical drivers for long-term changes in temperature and rainfall patterns over Europe. *Theoretical and Applied Climatology*, 143, 177–191.

HOY, A., HÄNSEL, S. & MAUGERI, M. (2020): An endless summer: 2018 heat episodes in Europe in the context of secular temperature variability and change. *International Journal of Climatology*, 40(15), 6315–6336.

HUNDHAUSEN, M., FELDMANN, H., LAUBE, N. & PINTO, J. G. (2023): Future heat extremes and impacts in a convection-permitting climate ensemble over Germany. *Natural Hazards and Earth System Sciences*, 23(8), 2873–2893.

HUNT, E., FEMIA, F., WERRELL, C., CHRISTIAN, J. I., OTKIN, J. A., BASARA, J., ANDERSON, M., WHITE, T., HAIN, C., RANDALL, R. & McGAUGHEY, K. (2021): Agricultural and food security impacts from the 2010 Russia flash drought. *Weather and Climate Extremes*, 34, 100383.

HURRELL, J. W. (1995): Decadal Trends in the North Atlantic Oscillation: Regional Temperatures and Precipitation. *Science*, 269(5224), 676–679.

HURRELL, J. W. & DESER, C. (2009): North Atlantic climate variability: The role of the North Atlantic Oscillation. *Journal of Marine Systems*, 78(1), 28–41.

ILES, C. & HEGERL, G. (2017): Role of the North Atlantic Oscillation in decadal temperature trends. *Environmental Research Letters*, 12(11), 114010.

IPCC (2014): Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. Technical report, Geneva, Switzerland.

JAMES, P. M. (2006): An assessment of European synoptic variability in Hadley Centre Global Environmental models based on an objective classification of weather regimes. *Climate Dynamics*, 27, 215–231.

KALISCH, M., MÄCHLER, M., COLOMBO, D., MAATHUIS, M. H. & BÜHLMANN, P. (2012): Causal Inference Using Graphical Models with the R Package pcalg. *Journal of Statistical Software*, 47(11), 1–26.

KARMOUCHE, S., GALYTSKA, E., MEEHL, G. A., RUNGE, J., WEIGEL, K. & EYRING, V. (2023a): Changing effects of external forcing on Atlantic-Pacific interactions. EGUsphere [preprint].

KARMOUCHE, S., GALYTSKA, E., RUNGE, J., MEEHL, G. A., PHILLIPS, A. S., WEIGEL, K. & EYRING, V. (2023b): Regime-oriented causal model evaluation of Atlantic–Pacific teleconnections in CMIP6. *Earth System Dynamics*, 14(2), 309–344.

KAUFMANN, R. K. & STERN, D. I. (1997): Evidence for human influence on climate from hemispheric temperature relations. *Nature*, 388, 39–44.

KAUTZ, L.-A., MARTIUS, O., PFAHL, S., PINTO, J. G., RAMOS, A. M., SOUSA, P. M. & WOOLLINGS, T. (2022): Atmospheric blocking and weather extremes over the Euro-Atlantic sector – a review. *Weather and Climate Dynamics*, 3(1), 305–336.

KAY, J. E., DESER, C., PHILLIPS, A., MAI, A., HANNAY, C., STRAND, G., ARBLASTER, J. M., BATES, S. C., DANABASOGLU, G., EDWARDS, J., HOLLAND, M., KUSHNER, P., LAMARQUE, J.-F., LAWRENCE, D., LINDSAY, K., MIDDLETON, A., MUNOZ, E., NEALE, R., OLESON, K., POLVANI, L. & VERTENSTEIN, M. (2015): The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability. *Bulletin of the American Meteorological Society*, 96(8), 1333 – 1349.

KIEHL, J. T. & TRENBERTH, K. E. (1997): Earth's Annual Global Mean Energy Budget. *Bulletin of the American Meteorological Society*, 78(2), 197 – 208.

KIRCHMEIER-YOUNG, M. C., ZWIERS, F. W. & GILLETT, N. P. (2017): Attribution of Extreme Events in Arctic Sea Ice Extent. *Journal of Climate*, 30(2), 553 – 571.

KODRA, E., CHATTERJEE, S. & GANGULY, A. R. (2011): Exploring Granger causality between global average observed time series of carbon dioxide and temperature. *Theoretical and Applied Climatology*, 104.

KORNHUBER, K., OSPREY, S., COUMOU, D., PETRI, S., PETOUKHOV, V., RAHMSTORF, S. & GRAY, L. (2019): Extreme weather events in early summer 2018 connected by a recurrent hemispheric wave-7 pattern. *Environmental Research Letters*, 14(5), 054002.

KRAUS, H. (2004): Die Atmosphäre der Erde. Eine Einführung in die Meteorologie. Springer-Verlag, Berlin Heidelberg New York.

KRETSCHMER, M., COUMOU, D., DONGES, J. F. & RUNGE, J. (2016): Using Causal Effect Networks to Analyze Different Arctic Drivers of Midlatitude Winter Circulation. *Journal of Climate*, 29(11), 4069 – 4081.

KRETSCHMER, M., RUNGE, J. & COUMOU, D. (2017): Early prediction of extreme stratospheric polar vortex states based on causal precursors. *Geophysical Research Letters*, 44(16), 8592–8600.

KYSELÝ, J. (2008): Influence of the persistence of circulation patterns on warm and cold temperature anomalies in Europe: Analysis over the 20th century. *Global and Planetary Change*, 62(1), 147–163.

LATIF, M. & KEENLYSIDE, N. S. (2009): El Niño/Southern Oscillation response to global warming. *Proceedings of the National Academy of Sciences*, 106(49), 20578–20583.

LEDUC, M., MAILHOT, A., FRIGON, A., MARTEL, J.-L., LUDWIG, R., BRIETZKE, G. B., GIGUÈRE, M., BRISSETTE, F., TURCOTTE, R., BRAUN, M. & 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, J.-Y., MAROTZKE, J., BALA, G., CAO, L., CORTI, J., S. DUNNE, ENGEL-BRECHT, F., FISCHER, E., FYFE, J., JONES, C., MAYCOCK, A., MUTEMI, J., NDIAYE, O., PANICKAL, S. & ZHOU, T. (2021): Future Global Climate: Scenario-Based Projections and Near-Term Information. In: MASSON-DELMOTTE, V., ZHAI, P., PIRANI, A., CONNORS, S., PÉAN, C., BERGER, S., CAUD, N., CHEN, Y., GOLDFARB, L., GOMIS, M., HUANG, M., LEITZELL, K., LONNOY, E., MATTHEWS, J., MAYCOCK, T., WATERFIELD, T., YELEKÇI, O., YU, R. & ZHOU, B. (eds.), Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 553–672.

LEHNER, F. & DESER, C. (2023): Origin, importance, and predictive limits of internal climate variability. *Environmental Research: Climate*, 2(2), 023001.

LEHNER, F., DESER, C., MAHER, N., MAROTZKE, J., FISCHER, E. M., BRUNNER, L., KNUTTI, R. & HAWKINS, E. (2020): Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth System Dynamics*, 11(2), 491–508.

LEHNER, F., DESER, C. & TERRAY, L. (2017): Toward a New Estimate of "Time of Emergence" of Anthropogenic Warming: Insights from Dynamical Adjustment and a Large Initial-Condition Model Ensemble. *Journal of Climate*, 30(19), 7739 – 7756.

LENTON, T. M., XU, C., ABRAMS, J. F., GHADIALI, A., LORIANI, S., SAKSCHEWSKI, B., ZIMM, C., EBI, K. L., DUNN, R. R., SVENNING, J.-C. & SCHEFFER, M. (2023): Quantifying the human cost of global warming. *Nature Sustainability*, 6, 1237–1237.

LEONARD, M., WESTRA, S., PHATAK, A., LAMBERT, M., VAN DEN HURK, B., MCINNES, K., RISBEY, J., SCHUSTER, S., JAKOB, D. & STAFFORD-SMITH, M. (2014): A compound event framework for understanding extreme impacts. *WIREs Climate Change*, 5(1), 113–128.

LEPS, N., BRAUCH, J. & AHRENS, B. (2019): Sensitivity of Limited Area Atmospheric Simulations to Lateral Boundary Conditions in Idealized Experiments. *Journal of Advances in Modeling Earth Systems*, 11(8), 2694–2707.

LFU (n.d.): Amtliche Festsetzung von Überschwemmungsgebieten. Bayerisches Landesamt für Umwelt (LfU), accessed: 24 October 2023.

LOGANATHAN, P. & MAHINDRAKAR, A. B. (2020): Assessment and ranking of CMIP5 GCMs performance based on observed statistics over Cauvery river basin – Peninsular India. *Arabian Journal of Geosciences*, 13, 1200.

VAN LOON, A. F. (2015): Hydrological drought explained. *WIREs Water*, 2(4), 359–392.

LORENZ, P. & JACOB, D. (2005): Influence of regional scale information on the global circulation: A two-way nesting climate simulation. *Geophysical Research Letters*, 32(18).

LUCAS-PICHER, P., CAYA, D., BINER, S. & LAPRISE, R. (2008a): Quantification of the Lateral Boundary Forcing of a Regional Climate Model Using an Aging Tracer. *Monthly Weather Review*, 136(12), 4980 – 4996.

LUCAS-PICHER, P., CAYA, D., DE ELÍA, R. & LAPRISE, R. (2008b): Investigation of regional climate models' internal variability with a ten-member ensemble of 10-year simulations over a large domain. *Climate Dynamics*, 31, 927–940.

LUCAS-PICHER, P., LAPRISE, R. & WINGER, K. (2017): Evidence of added value in North American regional climate model hindcast simulations using ever-increasing horizontal resolutions. *Climate Dynamics*, 48, 2611–2633.

MAHER, N., MATEI, D., MILINSKI, S. & MAROTZKE, J. (2018): ENSO Change in Climate Projections: Forced Response or Internal Variability? *Geophysical Research Letters*, 45(20), 11390–11398.

MAHER, N., MILINSKI, S. & 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.

MAHER, N., MILINSKI, S., SUAREZ-GUTIERREZ, L., BOTZET, M., DOBRYNIN, M., KORNBLUEH, L., KRÖGER, J., TAKANO, Y., GHOSH, R., HEDEMANN, C., LI, C., LI, H., MANZINI, E., NOTZ, D., PUTRASAHAN, D., BOYSEN, L., CLAUSSEN, M., ILYINA, T., OLONSCHECK, D., RADDATZ, T., STEVENS, B. & MAROTZKE, J. (2019): The Max Planck Institute Grand Ensemble: Enabling the Exploration of Climate System Variability. *Journal of Advances in Modeling Earth Systems*, 11(7), 2050–2069.

MASSON-DELMOTTE, V., ZHAI, P., PIRANI, A., CONNORS, S., PÉAN, C., BERGER, S., CAUD, N., CHEN, Y., GOLDFARB, L., GOMIS, M., HUANG, M., LEITZELL, K., LONNOY, E., MATTHEWS, J., MAYCOCK, T., WATERFIELD, T., YELEKÇI, O., YU, R. & ZHOU, B. (2021): Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

MATTE, D., LAPRISE, R., THÉRIAULT, J. M. & LUCAS-PICHER, P. (2017): Spatial spin-up of fine scales in a regional climate model simulation driven by low-resolution boundary conditions. *Climate Dynamics*, 49, 563–574.

MCKEE, T. B., DOESKEN, N. J. & KLEIST, J. (1993): The relationship of drought frequency and duration to time scales. In: Proceedings of the 8th Conference on Applied Climatology, Boston, 179–183.

MCKENNA, C. M. & MAYCOCK, A. C. (2021): Sources of Uncertainty in Multimodel Large Ensemble Projections of the Winter North Atlantic Oscillation. *Geophysical Research Letters*, 48(14), e2021GL093258.

MEEHL, G. A., SENIOR, C. A., EYRING, V., FLATO, G., LAMARQUE, J.-F., STOUFFER, R. J., TAYLOR, K. E. & SCHLUND, M. (2020): Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models. *Science Advances*, 6(26), eaba1981.

MIERSCH, P., JIANG, S., RAKOVEC, O. & ZSCHEISCHLER, J. (2023): Identifying drivers of river floods using causal inference. EGU General Assembly 2023, Vienna, Austria, 24-28 Apr 2023, EGU23-12948.

MIRALLES, D. G., GENTINE, P., SENEVIRATNE, S. I. & 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.

MIRALLES, D. G., TEULING, A. J., VAN HEERWAARDEN, C. C. & VILÀ-GUERAU DE ARELLANO, J. (2014): Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. *Nature Geoscience*, 7, 345–349.

MITTERMEIER, M., BRAUN, M., HOFSTÄTTER, M., WANG, Y. & LUDWIG, R. (2019): Detecting Climate Change Effects on Vb Cyclones in a 50-Member Single-Model Ensemble Using Machine Learning. *Geophysical Research Letters*, 46(24), 14653–14661.

MITTERMEIER, M., WEIGERT, M., RÜGAMER, D., KÜCHENHOFF, H. & 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.

MOSEDALE, T. J., STEPHENSON, D. B., COLLINS, M. & MILLS, T. C. (2006): Granger Causality of Coupled Climate Processes: Ocean Feedback on the North Atlantic Oscillation. *Journal of Climate*, 19(7), 1182 – 1194.

NCAR (2017): The Climate Data Guide: Common Spectral Model Grid Resolutions. National Center for Atmospheric Research Staff (Eds). accessed: 06 June 2023.

van Nes, E. H., Scheffer, M., Brovkin, V., Lenton, T. M., Ye, H., Deyle, E. & Sugihara, G. (2015): Causal feedbacks in climate change. *Nature Climate Change*, 5, 445–448.

Nikbakht, J., Tabari, H. & Talaee, P. H. (2013): Streamflow drought severity analysis by percent of normal index (PNI) in northwest Iran. *Theoretical and applied climatology*, 112(3), 565–573.

Nowack, P., Runge, J., Eyring, V. & Haigh, J. D. (2020): Causal networks for climate model evaluation and constrained projections. *Nature Communications*, 11, 1415.

van Oldenborgh, G. J., van der Wiel, K., Kew, S., Philip, S., Otto, F., Vautard, R., King, A., Lott, F., Arrighi, J., Singh, R. & van Aalst, M. (2021): Pathways and pitfalls in extreme event attribution. *Climatic Change*, 166, 13.

Omrani, H., Drobinski, P. & Dubos, T. (2013): Optimal nudging strategies in regional climate modelling: investigation in a Big-Brother experiment over the European and Mediterranean regions. *Climate Dynamics*, 41, 2451–2470.

Ossó, A., Allan, R. P., Hawkins, E., Shaffrey, L. & Maraun, D. (2022): Emerging new climate extremes over Europe. *Climate Dynamics*, 58, 487–501.

Otkin, J. A., Svoboda, M., Hunt, E. D., Ford, T. W., Anderson, M. C., Hain, C. & Basara, J. B. (2018): Flash Droughts: A Review and Assessment of the Challenges Imposed by Rapid-Onset Droughts in the United States. *Bulletin of the American Meteorological Society*, 99(5), 911 – 919.

Otto, F. E. L., Massey, N., van Oldenborgh, G. J., Jones, R. G. & Allen, M. R. (2012): Reconciling two approaches to attribution of the 2010 Russian heat wave. *Geophysical Research Letters*, 39(4).

O'Neill, B. C., Carter, T. R., Ebi, K., Harrison, P. A., Kemp-Benedict, E., Kok, K., Kriegler, E., Preston, B. L., Riahi, K., Sillmann, J., van Ruijven, B. J., van Vuuren, D., Carlisle, D., Conde, C., Fuglestvedt, J.,

GREEN, C., HASEGAWA, T., LEININGER, J., MONTEITH, S. & PICHS-MADRUGA, R. (2020): Achievements and needs for the climate change scenario framework. *Nature Climate Change*, 10, 1074–1084.

O'NEILL, B. C., KRIEGLER, E., EBI, K. L., KEMP-BENEDICT, E., RIAHI, K., ROTHMAN, D. S., VAN RUIJVEN, B. J., VAN VUUREN, D. P., BIRKMANN, J., KOK, K., LEVY, M. & SOLECKI, W. (2017): The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. *Global Environmental Change*, 42, 169–180.

PAETH, H. & MANNIG, B. (2013): On the added value of regional climate modeling in climate change assessment. *Climate Dynamics*, 41, 1057–1066.

PALUS, M. & VEJMELKA, M. (2007): Directionality of coupling from bivariate time series: How to avoid false causalities and missed connections. *Phys. Rev. E*, 75, 056211.

PEARL, J. & MACKENZIE, D. (2018): The Book of Why. Basic Books, New York.

PFAHL, S. & WERNLI, H. (2012): Quantifying the relevance of atmospheric blocking for co-located temperature extremes in the Northern Hemisphere on (sub-)daily time scales. *Geophysical Research Letters*, 39(12).

PFEIFER, S., BÜLOW, K., GOBIET, A., HÄNSLER, A., MUDELSEE, M., OTTO, J., RECHID, D., TEICHMANN, C. & JACOB, D. (2015): Robustness of Ensemble Climate Projections Analyzed with Climate Signal Maps: Seasonal and Extreme Precipitation for Germany. *atmosphere*, 6(5), 677–698.

POSCHLOD, B., ZSCHEISCHLER, J., SILLMANN, J., WOOD, R. R. & LUDWIG, R. (2020): Climate change effects on hydrometeorological compound events over southern Norway. *Weather and Climate Extremes*, 28, 100253.

RAHMSTORF, S., FOSTER, G. & CAZENAVE, A. (2012): Comparing climate projections to observations up to 2011. *Environmental Research Letters*, 7(4), 044035.

REX, D. F. (1950): Blocking Action in the Middle Troposphere and its Effect upon Regional Climate. *Tellus*, 2(4), 275–301.

RIAHI, K., VAN VUUREN, D. P., KRIEGLER, E., EDMONDS, J., O'NEILL, B. C., FUJIMORI, S., BAUER, N., CALVIN, K., DELLINK, R., FRICKO, O., LUTZ, W., POPP, A., CUARESMA, J. C., KC, S., LEIMBACH, M., JIANG, L., KRAM, T., RAO, S., EMMERLING, J., EBI, K., HASEGAWA, T., HAVLIK, P., HUMPENÖDER, F., DA SILVA, L. A., SMITH, S., STEHFEST, E., BOSETTI, V., EOM, J., GERNAAT, D., MASUI, T., ROGELJ, J., STREFLER, J., DROUET, L., KREY, V., LUDERER, G., HARMSEN, M., TAKAHASHI, K., BAUMSTARK, L., DOELMAN, J. C., KAINUMA, M., KLIMONT, Z., MARANGONI, G., LOTZE-CAMPEN, H., OBERSTEINER, M., TABEAU, A. & TAVONI, M. (2017): The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, 42, 153–168.

RITCHIE, J. & DOWLATABADI, H. (2017): Why do climate change scenarios return to coal? *Energy*, 140, 1276–1291.

RIVM (n.d.): How does the National Heatwave Plan work? Rijksinstituut voor Volksgezondheid en Milieu Ministerie van Volksgezondheid, Welzijn en Sport (RIVM), accessed: 12 October 2023.

RODDA, J. C. & MARSH, T. J. (2011): The 1975-75 Drought - a contemporary and retrospective review. Technical report, Centre for Ecology and Hydrology, Wallingford, UK. 58 pp.

ROEDEL, W. & WAGNER, T. (2011): Physik unserer Umwelt: Die Atmosphäre. Springer-Verlag, Berlin Heidelberg.

RUMMUKAINEN, M. (2010): State-of-the-art with regional climate models. *WIREs Climate Change*, 1(1), 82–96.

RUMMUKAINEN, M. (2016): Added value in regional climate modeling. *WIREs Climate Change*, 7(1), 145–159.

RUNGE, J. (2018): Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(7), 075310.

RUNGE, J. (2020): Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In: PETERS, J. & SONTAG, D. (eds.), Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI), volume 124 of *Proceedings of Machine Learning Research*, PMLR, 1388–1397.

RUNGE, J., BATHIANY, S., BOLLT, E., CAMPS-VALLS, G., COUMOU, D., DEYLE, E., GLYMOUR, C., KRETSCHMER, M., MAHECHA, M. D., MUÑOZ-MARÍ, J., VAN NES, E. H., PETERS, J., QUAX, R., REICHSTEIN, M., SCHEFFER, M., SCHÖLKOPF, B., SPIRTES, P., SUGIHARA, G., SUN, J., ZHANG, K. & ZSCHEISCHLER, J. (2019a): Inferring causation from time series in Earth system sciences. *Nature Communications*, 10, 2553.

RUNGE, J., NOWACK, P., KRETSCHMER, M., FLAXMAN, S. & SEJDINOVIC, D. (2019b): Detecting and quantifying causal associations in large nonlinear time series dataset. *Science advances*, 5, eaau4996.

RUNGE, J., PETOUKHOV, V., DONGES, J. F., HLINKA, J., JAJCAY, N., VEJMELKA, M., HARTMAN, D., MARWAN, N., PALUŠ, M. & KURTHS, J. (2015): Identifying causal gateways and mediators in complex spatio-temporal systems. *Nature Communications*, 6, 8502.

RUSSO, F. & WILLIAMSON, J. (2007): Interpreting Causality in the Health Sciences. *International Studies in the Philosophy of Science*, 21(2), 157–170.

RUSSO, S., SILLMANN, J. & FISCHER, E. M. (2015): Top ten European heatwaves since 1950 and their occurrence in the coming decades. *Environmental Research Letters*, 10(12), 124003.

RÖTHLISBERGER, M. & MARTIUS, O. (2019): Quantifying the Local Effect of Northern Hemisphere Atmospheric Blocks on the Persistence of Summer Hot and Dry Spells. *Geophysical Research Letters*, 46(16), 10101–10111.

RÖTHLISBERGER, M. & PAPRITZ, L. (2023): Quantifying the physical processes leading to atmospheric hot extremes at a global scale. *Nature Geoscience*, 16, 210–216.

VON SALZEN, K., SCINOCCA, J. F., MCFARLANE, N. A., LI, J., COLE, J. N. S., PLUMMER, D., VERSEGHY, D., READER, M. C., MA, X., LAZARE, M. & SOLHEIM, L. (2013): The Canadian Fourth Generation Atmospheric Global Climate

Model (CanAM4). Part I: Representation of Physical Processes. *Atmosphere-Ocean*, 51(1), 104–125.

Sangelantoni, L., Sobolowski, S., Lorenz, T., Hodnebrog, O., Cardoso, R. M., Soares, P. M. M., Ferretti, R., Lavín-Gullón, A., Fernandez, J., Goergen, K., Milovac, J., Katragkou, E., Kartsios, S., Coppola, E., Pichelli, E., Adinolfi, M., Mercogliano, P., Berthou, S., de Vries, H., Dobler, A., Belusic, D., Feldmann, H., Tölle, M. H. & Bastin, S. (2023): Investigating the representation of heatwaves from an ensemble of km-scale regional climate simulations within CORDEX-FPS convection. *Climate Dynamics*.

Scaife, A., Folland, C., Alexander, L., Moberg, A. & Knight, J. (2008): European Climate Extremes and the North Atlantic Oscillation. *Journal of Climate*, 21(1), 72–83.

Schumacher, D. L., Keune, J., Dirmeyer, P. & Miralles, D. G. (2022): Drought self-propagation in drylands due to land–atmosphere feedbacks. *Nature Geoscience*, 15, 262–268.

Schumacher, D. L., Keune, J., van Heerwaarden, C. C., Vilà-Guerau de Arellano, J., Teuling, A. J. & Miralles, D. G. (2019): Amplification of mega-heatwaves through heat torrents fuelled by upwind drought. *Nature Geoscience*, 12, 712–717.

Schumacher, D. L., Zachariah, M., Otto, F., Barnes, C., Philip, S., Kew, S., Vahlberg, M., Singh, R., Heinrich, D., Arrighi, J., van Aalst, M., Hauser, M., Hirschi, M., Bessenbacher, V., Gudmundsson, L., Beaudoing, H. K., Rodell, M., Li, S., Yang, W., Vecchi, G. A., Harrington, L. J., Lehner, F., Balsamo, G. & Seneviratne, S. I. (2023): Detecting the human fingerprint in the summer 2022 West-Central European soil drought. *EGUsphere [preprint]*, 2023, 1–41.

Schurz, G. (2021): Erkenntnistheorie: Eine Einführung, chapter Der Begriff des Wissens. J.B. Metzler, Stuttgart, 15–31.

SCHWALM, C. R., GLENDON, S. & DUFFY, P. B. (2020): Reply to Hausfather and Peters: RCP8.5 is neither problematic nor misleading. *Proceedings of the National Academy of Sciences*, 117(45), 27793–27794.

SCHWINGSHACKL, C., HIRSCHI, M. & SENEVIRATNE, S. I. (2017): Quantifying Spatiotemporal Variations of Soil Moisture Control on Surface Energy Balance and Near-Surface Air Temperature. *Journal of Climate*, 30(18), 7105 – 7124.

SCHÖNWIESE, C.-D. (2020): Klimatologie. Ulmer, Stuttgart, Deutschland.

SENEVIRATNE, S. I., CORTI, T., DAVIN, E. L., HIRSCHI, M., JAEGER, E. B., LEHNER, I., ORLOWSKY, B. & TEULING, A. J. (2010): Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, 99(3), 125–161.

SHEPHERD, T. G. (2014): Atmospheric circulation as a source of uncertainty in climate change projections. *Nature Geoscience*, 7, 703–708.

SHERWOOD, S. C. & HUBER, M. (2010): An adaptability limit to climate change due to heat stress. *Proceedings of the National Academy of Sciences*, 107(21), 9552–9555.

SHUKLA, S. & WOOD, A. W. (2008): Use of a standardized runoff index for characterizing hydrologic drought. *Geophysical Research Letters*, 35(2).

SI, W., LIU, H., ZHANG, X. & ZHANG, M. (2021): Double Intertropical Convergence Zones in Coupled Ocean-Atmosphere Models: Progress in CMIP6. *Geophysical Research Letters*, 48(23), e2021GL094779.

SOLOMON, S., QIN, D., MANNING, M., CHEN, Z., MARQUIS, M., AVERYT, K., TIGNOR, M. & MILLER, H. (2007): Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

SOUSA, P. M., TRIGO, R. M., BARRIOPEDRO, D., SOARES, P. M. M., RAMOS, A. M. & LIBERATO, M. L. R. (2017): Responses of European precipitation distributions and regimes to different blocking locations. *Climate Dynamics*, 48, 1141–1160.

Sousa, P. M., Trigo, R. M., Barriopedro, D., Soares, P. M. M. & Santos, J. A. (2018): European temperature responses to blocking and ridge regional patterns. *Climate Dynamics*, 50, 457–477.

Spinoni, J., Barbosa, P., Bucchignani, E., Cassano, J., Cavazos, T., Christensen, J. H., Christensen, O. B., Coppola, E., Evans, J., Geyer, B., Giorgi, F., Hadjinicolaou, P., Jacob, D., Katzfey, J., Koenigk, T., Laprise, R., Lennard, C. J., Kurnaz, M. L., Li, D., Llopart, M., McCormick, N., Naumann, G., Nikulin, G., Ozturk, T., Panitz, H.-J., da Rocha, R. P., Rockel, B., Solman, S. A., Syktus, J., Tangang, F., Teichmann, C., Vautard, R., Vogt, J. V., Winger, K., Zittis, G. & Dosio, A. (2020): Future Global Meteorological Drought Hot Spots: A Study Based on CORDEX Data. *Journal of Climate*, 33(9), 3635–3661.

Spirtes, P. & Glymour, C. (1991): An Algorithm for Fast Recovery of Sparse Causal Graphs. *Social Science Computer Review*, 9(1), 62–72.

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

StMGP (n.d.): Hitze – Gesundheitsschutz bei hohen Temperaturen. Bayerisches Staatsministerium für Gesundheit und Pflege, accessed: 12 October 2023.

StMUV (n.d.): Hochwasserrisiko. Bayerisches Staatsministerium für Umwelt und Verbraucherschutz (StMUV), accessed: 24 October 2023.

Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V. & Midgley, P. (2013): Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Stott, P. A., Stone, D. A. & Allen, M. R. (2004): Human contribution to the European heatwave of 2003. *Nature*, 432, 610–614.

Stouffer, R. J. & Manabe, S. (2017): Assessing temperature pattern projections made in 1989. *Nature Climate Change*, 7, 163–165.

Suarez-Gutierrez, L., Milinski, S. & Maher, N. (2021): Exploiting large ensembles for a better yet simpler climate model evaluation. *Climate Dynamics*, 57, 2557–2580.

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

Suarez-Gutierrez, L., Müller, W. A., Li, C. & Marotzke, J. (2020b): Hotspots of extreme heat under global warming. *Climate Dynamics*, 55, 429–447.

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Hanna, S., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Sigmond, M., Solheim, L., von Salzen, K., Yang, D. & Winter, B. (2019): The Canadian Earth System Model version 5 (CanESM5.0.3). *Geoscientific Model Development*, 12(11), 4823–4873.

Sørland, S. L., Schär, C., Lüthi, D. & Kjellström, E. (2018): Bias patterns and climate change signals in GCM-RCM model chains. *Environmental Research Letters*, 13(7), 074017.

Taylor, C. M., de Jeu, R. A. M., Guichard, F., Harris, P. P. & Dorigo, W. A. (2012a): Afternoon rain more likely over drier soils. *Nature*, 489, 423–426.

Taylor, K. E., Stouffer, R. J. & Meehl, G. A. (2012b): An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, 93(4), 485 – 498.

Teichmann, C., Eggert, B., Elizalde, A., Haensler, A., Jacob, D., Kumar, P., Moseley, C., Pfeifer, S., Rechid, D., Remedio, A. R., Ries, H., Petersen, J., Preuschmann, S., Raub, T., Saeed, F., Sieck, K. & Weber, T. (2013): How Does a Regional Climate Model Modify the Projected Climate Change Signal of the Driving GCM: A Study over Different CORDEX Regions Using REMO. *Atmosphere*, 4(2), 214–236.

TEULING, A. J., HIRSCHI, M., OHMURA, A., WILD, M., REICHSTEIN, M., CIAIS, P., BUCHMANN, N., AMMANN, C., MONTAGNANI, L., RICHARDSON, A. D., WOHLFAHRT, G. & SENEVIRATNE, S. I. (2009): A regional perspective on trends in continental evaporation. *Geophysical Research Letters*, 36(2).

TIJDEMAN, E., BLAUHUT, V., STOELZLE, M., MENZEL, L. & STAHL, K. (2022): Different drought types and the spatial variability in their hazard, impact, and propagation characteristics. *Natural Hazards and Earth System Sciences*, 22(6), 2099–2116.

TORMA, C., GIORGI, F. & COPPOLA, E. (2015): Added value of regional climate modeling over areas characterized by complex terrain—Precipitation over the Alps. *Journal of Geophysical Research: Atmospheres*, 120(9), 3957–3972.

TRENBERTH, K. E. (2012): Framing the way to relate climate extremes to climate change. *Climatic Change*, 115, 283–290.

TRENBERTH, K. E. & SHEA, D. J. (2005): Relationships between precipitation and surface temperature. *Geophysical Research Letters*, 32(14).

VON TRENTINI, F., AALBERS, E. E., FISCHER, E. M. & LUDWIG, R. (2020): Comparing interannual variability in three regional single-model initial-condition large ensembles (SMILEs) over Europe. *Earth System Dynamics*, 11(4), 1013–1031.

VON TRENTINI, F., LEDUC, M. & 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.

TRIGO, R., OSBORN, T. & CORTE-REAL, J. (2002): The North Atlantic Oscillation influence on Europe: climate impacts and associated physical mechanisms. *Climate Research*, 20(1), 9–17.

UFZ (n.d.): Dürremonitor Deutschland. Helmholtz Zentrum für Umweltforschung (UFZ), accessed: 02 July 2023.

UN (2022): Total population by sex. United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects: The 2022 Revision. accessed: 11 December 2023. Https://population.un.org/wpp/.

VAUTARD, R., GOBIET, A., JACOB, D., BELDA, M., COLETTE, A., DÉQUÉ, M., FERNÁNDEZ, J., GARCÍA-DÍEZ, M., GOERGEN, K., GÜTTLER, I., HALENKA, T., KARACOSTAS, T., KATRAGKOU, E., KEULER, K., KOTLARSKI, S., MAYER, S., VAN MEIJGAARD, E., NIKULIN, G., PATARČIĆ, M., SCINOCCA, J., SOBOLOWSKI, S., SUKLITSCH, M., TEICHMANN, C., WARRACH-SAGI, K., WULFMEYER, V. & YIOU, P. (2013): The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project. *Climate Dynamics*, 41, 2555–2575.

VERDES, P. F. (2005): Assessing causality from multivariate time series. *Phys. Rev. E*, 72, 026222.

VICENTE-SERRANO, S. M., BEGUERÍA, S. & LÓPEZ-MORENO, J. I. (2010): A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *Journal of Climate*, 23(7), 1696 – 1718.

VICENTE-SERRANO, S. M., GARCÍA-HERRERA, R., PEÑA-ANGULO, D., TOMAS-BURGUERA, M., DOMÍNGUEZ-CASTRO, F., NOGUERA, I., CALVO, N., MURPHY, C., NIETO, R., GIMENO, L., GUTIERREZ, J. M., AZORIN-MOLINA, C. & EL KENAWY, A. (2022): Do CMIP models capture long-term observed annual precipitation trends? *Climate Dynamics*, 58, 2825–2842.

VOGEL, M. M., ZSCHEISCHLER, J., WARTENBURGER, R., DEE, D. & SENEVIRATNE, S. I. (2019): Concurrent 2018 Hot Extremes Across Northern Hemisphere Due to Human-Induced Climate Change. *Earth's Future*, 7(7), 692–703.

VAN VUUREN, D. P., EDMONDS, J., KAINUMA, M., RIAHI, K., THOMSON, A., HIBBARD, K., HURTT, G. C., KRAM, T., KREY, V., LAMARQUE, J.-F., MASUI, T., MEINSHAUSEN, M., NAKICENOVIC, N., SMITH, S. J. & ROSE, S. K. (2011): The representative concentration pathways: an overview. *Climatic Change*, 109, 5.

WALKER, G. & BLISS, E. (1932): World Weather V. *Memoirs of the Royal Meteorological Society*, 4(36), 53–84.

WALLACE, J. M. & GUTZLER, D. S. (1981): Teleconnections in the Geopotential Height Field during the Northern Hemisphere Winter. *Monthly Weather Review*, 109(4), 784 – 812.

WARNER, J. L. (2018): Arctic sea ice – a driver of the winter NAO? *Weather*, 73(10), 307–310.

WAUD, D. R. (1975): Analysis of Dose—Response Curves, chapter Analysis of Dose–Response Curves. Springer US, Boston, MA, 471–506.

WAZNEH, H., ARAIN, M. A., COULIBALY, P. & GACHON, P. (2020): Evaluating the Dependence between Temperature and Precipitation to Better Estimate the Risks of Concurrent Extreme Weather Events. *Advances in Meteorology*, 2020, 8763631.

VAN DER WIEL, K., WANDERS, N., SELTEN, F. M. & BIERKENS, M. F. P. (2019): Added Value of Large Ensemble Simulations for Assessing Extreme River Discharge in a 2 °C Warmer World. *Geophysical Research Letters*, 46(4), 2093–2102.

WILHITE, D. A. & GLANTZ, M. H. (1985): Understanding the Drought Phenomenon: The Role of Definitions. *Water International*, 10(3), 111–120.

WILLEKE, M. & HOSKING, J. (1994): The national drought Atlas Institute for water resources rep. *Army Corps of Engineers, Fort Belvoir*, 587.

WILLIAMS, I. N. & TORN, M. S. (2015): Vegetation controls on surface heat flux partitioning, and land-atmosphere coupling. *Geophysical Research Letters*, 42(21), 9416–9424.

WILLIAMSON, J. (2019): Establishing Causal Claims in Medicine. *International Studies in the Philosophy of Science*, 32(1), 33–61.

WILLKOFER, F., WOOD, R. R. & LUDWIG, R. (2023): Assessing the impact of climate change on high return levels of peak flows in Bavaria applying the CRCM5 Large Ensemble. *EGUsphere [preprint]*, 2023, 1–31.

WMO (2023a): Guidelines on the Definition and Characterization of Extreme Weather and Climate Events. Technical report, World Meteorological Organization (WMO), Switzerland. WMO-No. 1310.

WMO (2023b): State of the Climate in Europe 2022. Technical report, World Meteorological Organization (WMO), Switzerland. WMO-No. 1320.

Wood, R. R., Lehner, F., Pendergrass, A. G. & Schlunegger, S. (2021): Changes in precipitation variability across time scales in multiple global climate model large ensembles. *Environmental Research Letters*, 16(8), 084022.

Wood, R. R. & Ludwig, R. (2020): Analyzing Internal Variability and Forced Response of Subdaily and Daily Extreme Precipitation Over Europe. *Geophysical Research Letters*, 47(17), e2020GL089300.

Wu, C., Lin, Z. & Liu, X. (2020): The global dust cycle and uncertainty in cmip5 (coupled model intercomparison project phase 5) models. *Atmospheric Chemistry and Physics*, 20(17), 10401–10425.

Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., Jie, W., Zhang, J., Liu, Y., Zhang, L., Zhang, F., Zhang, Y., Wu, F., Li, J., Chu, M., Wang, Z., Shi, X., Liu, X., Wei, M., Huang, A., Zhang, Y. & Liu, X. (2019): The Beijing Climate Center Climate System Model (BCC-CSM): the main progress from CMIP5 to CMIP6. *Geoscientific Model Development*, 12(4), 1573–1600.

Yue, S. & Rasmussen, P. (2002): Bivariate frequency analysis: discussion of some useful concepts in hydrological application. *Hydrological Processes*, 16(14), 2881–2898.

Zink, M., Samaniego, L., Kumar, R., Thober, S., Mai, J., Schäfer, D. & Marx, A. (2016): The German drought monitor. *Environmental Research Letters*, 11(7), 074002.

Zscheischler, J. & Fischer, E. M. (2020): The record-breaking compound hot and dry 2018 growing season in Germany. *Weather and Climate Extremes*, 29, 100270.

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., van den Hurk, B., AghaKouchak, A., Jézéquel, A., Mahecha, M. D., Maraun, D., Ramos, A. M., Ridder, N. N., Thiery, W. & Vignotto, E. (2020): A typology of compound weather and climate events. *Nature Reviews Earth & Environment*, 1, 333–347.

Zscheischler, J., Orth, R. & Seneviratne, S. I. (2015): A submonthly database for detecting changes in vegetation-atmosphere coupling. *Geophysical Research Letters*, 42(22), 9816–9824.

ZSCHEISCHLER, J. & SENEVIRATNE, S. I. (2017): Dependence of drivers affects risks associated with compound events. *Science Advances*, 3(6), e1700263.

ZSCHEISCHLER, J., WESTRA, S., VAN DEN HURK, B. J. J. M., SENEVIRATNE, S. I., WARD, P. J., PITMAN, A., AGHAKOUCHAK, A., BRESCH, D. N., LEONARD, M., WAHL, T. & ZHANG, X. (2018): Future climate risk from compound events. *Nature Climate Change*, 8, 469–477.