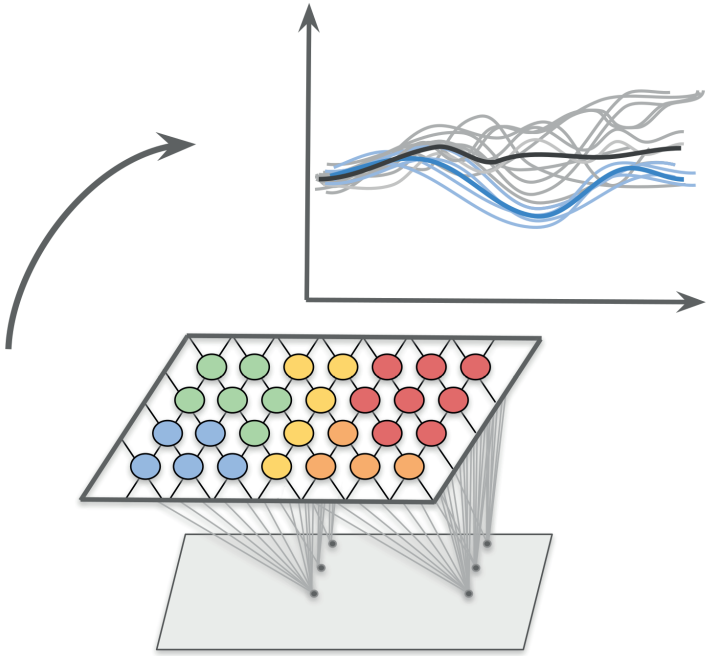
MAX-PLANCK-INSTITUT
FÜR METEOROLOGIE

Towards improved seasonal climate predictions with artificial intelligence:

An application on summer teleconnections



Julianna Carvalho Oliveira

Hamburg 2023

Berichte zur Erdsystemforschung
Reports on Earth System Science

261
2023

Hinweis

Die Berichte zur Erdsystemforschung werden vom Max-Planck-Institut für Meteorologie in Hamburg in unregelmäßiger Abfolge herausgegeben.

Sie enthalten wissenschaftliche und technische Beiträge, inklusive Dissertationen.

Die Beiträge geben nicht notwendigerweise die Auffassung des Instituts wieder.

Die "Berichte zur Erdsystemforschung" führen die vorherigen Reihen "Reports" und "Examensarbeiten" weiter.

Anschrift / Address

Max-Planck-Institut für Meteorologie
Bundesstrasse 53
20146 Hamburg
Deutschland

Tel./Phone: +49 (0)40 4 11 73 - 0
Fax: +49 (0)40 4 11 73 - 298

name.surname@mpimet.mpg.de
www.mpimet.mpg.de

Notice

The Reports on Earth System Science are published by the Max Planck Institute for Meteorology in Hamburg. They appear in irregular intervals.

They contain scientific and technical contributions, including PhD theses.

The Reports do not necessarily reflect the opinion of the Institute.

The "Reports on Earth System Science" continue the former "Reports" and "Examensarbeiten" of the Max Planck Institute.

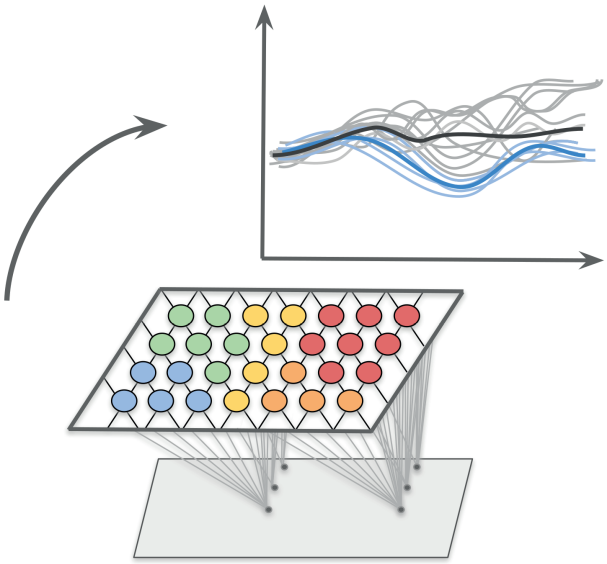
Layout

Bettina Diallo and Norbert P. Noreiks
Communication

Copyright

Towards improved seasonal climate predictions with artificial intelligence:

An application on summer teleconnections



Julianna Carvalho Oliveira

Hamburg 2023

Julianna Carvalho Oliveira
aus Itabuna, Brasilien

Max-Planck-Institut für Meteorologie
The International Max Planck Research School on Earth System Modelling
(IMPRS-ESM)
Bundesstrasse 53
20146 Hamburg

Universität Hamburg
Erdsystemwissenschaften
Bundesstr. 55
20146 Hamburg

Tag der Disputation: 15. Dezember 2022

Folgende Gutachter empfehlen die Annahme der Dissertation:
Prof. Dr. Johanna Baehr
Dr. Eduardo Zorita

Vorsitzender des Promotionsausschusses:
Prof. Dr. Hermann Held

Dekan der MIN-Fakultät:
Prof. Dr.-Ing. Norbert Ritter

Titelgrafik: Julianna Carvalho Oliveira

Berichte zur Erdsystemforschung / Max-Planck-Institut für Meteorologie 261
Reports on Earth System Science / Max Planck Institute for Meteorology 2023

ISSN 1614-1199 - doi: 10.17617/2.3489354

"Damit das Mögliche entsteht, muss immer wieder das Unmögliche versucht werden."

— Hermann Hesse

Dedicated to my parents Silvana and Zeca.

ABSTRACT

Recurrent large-scale atmospheric circulation patterns, or teleconnections, exert a prominent effect on the Euro-Atlantic surface climate. In summer, teleconnections are amongst the main drivers for high-impact climatic processes such as heatwaves, and hence several relevant socio-economic sectors could benefit from their credible seasonal prediction. However, dynamical climate models show limited capability to reproduce summer teleconnections. This problem is further compounded by the complex physical mechanisms influencing their predictability, which are still not well understood. While conventional statistical tools offer only a limited assessment of these physical mechanisms, artificial intelligence (AI) outperforms these tools, learning complex relationships from data and thereby advancing physical understanding. Here, I promote the combination of observations and dynamical climate modelling with AI to overcome some of these limitations and to achieve improved predictions of European summer climate a season ahead.

I implement this novel AI-dynamical approach in two complementary steps: I first refine the assessment of summer teleconnections in observations and a model, and then I apply this knowledge to improve Euro-Atlantic summer seasonal climate predictions. I use the AI classifier Self-Organising Maps (SOM) to characterise the observed and modelled variability of the two dominant Euro-Atlantic summer teleconnections in the 20th century: the summer North Atlantic Oscillation (NAO) and summer East Atlantic Pattern (EA). I find that while the ensemble dynamical prediction system can reproduce summer NAO and EA spatial features, it shows limited model performance in reproducing their frequency of occurrence.

I use SOM to illustrate that the seasonal predictability of summer teleconnections is associated with North Atlantic sea surface temperatures (SST), however this influence varies in intensity with time and is more relevant for summer EA than for summer NAO. I go beyond standard forecast practices by applying these SST predictors to constrain the credibility of summer climate predictions in a dynamical ensemble prediction system. I show that, particularly for years during which EA dominates, summer climate predictions up to 4 months ahead can be significantly improved in parts of Europe using these SST predictors.

With the use of an AI causal inference tool I find that although extratropical NA SST in spring show a causal link with EA in the second half of the 20th century in observations, the evaluated dynamical ensemble prediction shows limited performance to reproduce this causal link. However, I do find that those ensemble simulations that reproduce this causal link show improved surface climate prediction credibility over those that do not. Overall, my findings promote the use of combined AI-dynamical approach to improve seasonal predictions and could even be applied operationally, benefiting actual seasonal predictions of European summer climate.

ZUSAMMENFASSUNG

Wiederkehrende atmosphärische Strömungsmuster, auch „Telekonnektionen“ genannt, haben einen großen Einfluss auf das Wetter und Klima im Euro-Atlantikraum. Im Sommer gehören Telekonnektionen zu den wichtigsten Treibern für potenziell folgenreiche klimatische Prozesse wie z. B. Hitzewellen. Verschiedene wichtige sozioökonomische Bereiche könnten daher von einer zuverlässigen saisonalen Vorhersage dieser Telekonnektionen profitieren. Dynamische Klimamodelle sind jedoch nur begrenzt in der Lage, sommerliche Telekonnektionen zu reproduzieren. Dieses Problem wird noch dadurch verschärft, dass die komplexen physikalischen Mechanismen, die Telekonnektionsvorhersagbarkeit beeinflussen, noch nicht hinreichend verstanden sind. Herkömmliche statistische Instrumente haben klare Grenzen, wenn es um die Analyse dieser physikalischen Mechanismen geht. Künstliche Intelligenz (KI) hingegen übertrifft die klassische Statistik, indem sie komplexe Beziehungen aus den Daten erlernt und dadurch Verständnis nicht-linearer physikalischer Zusammenhänge ermöglicht. In dieser Dissertation kombiniere ich Klimabeobachtungen und dynamische Klimamodellierung mit KI, um Nichtlinearitäten im Telekonnektionen besser zu erfassen und so bessere Vorhersagen des europäischen Sommerklimas für eine Saison im Voraus zu ermöglichen.

Ich nutze diesen neuartigen KI-dynamischen Ansatz in zwei komplementären Schritten: Zunächst verbessere ich die Analyse der sommerlichen Telekonnektionen in den Beobachtungen und einem Modell, und dann wende ich dieses Wissen an, um saisonale Klimavorhersagen für den euro-atlantischen Sommer zu verbessern. Ich verwende den KI-Klassifikator Self-Organising Maps (SOM), um die beobachtete und modellierte Variabilität der beiden wichtigsten euro-atlantischen Telekonnektionen im 20. Jahrhundert zu charakterisieren: die sommerliche Nordatlantische Oszillation (NAO) und das Sommer-Ost-Atlantik-Muster („East Atlantic pattern“, EA). Ich stelle fest, dass das dynamische Klimavorhersagesystem zwar die räumlichen Merkmale der sommerlichen NAO und des EA reproduzieren kann, aber bei der Reproduktion der Häufigkeit ihres Auftretens klare Defizite aufweist.

Die SOM-Analyse zeigt auch, dass die saisonale Vorhersagbarkeit der sommerlichen Telekonnektionen mit den Meeresoberflächentemperaturen (SST) des Nordatlantiks zusammenhängt, wobei dieser Einfluss in seiner Intensität mit der Zeit variiert und für die sommerliche EA wichtiger ist als für die sommerliche NAO. Ich gehe über übliche Vorhersagepraktiken hinaus, indem ich die so identifizierten SST-Prädiktoren anwende, um die Glaubwürdigkeit der Sommer-Klimavorhersagen in einem dynamischen Ensemble-Vorhersagesystem besser abzuschätzen. Ich zeige, dass insbesondere für die Jahre, in denen die EA dominiert, die Vorhersage des Sommerklimas in Teilen Europas mit Hilfe dieser SST-Prädiktoren bis zu 4 Monate im Voraus signifikant verbessert werden kann. Mit Hilfe einer KI-basierten kausalen Detektionsmethode stelle ich fest, dass die außertropische NA-SST im Frühjahr zwar einen beobachtbaren, kausalen Zusammenhang mit EA in der zweiten Hälfte des 20. Jahrhundert hat, die untersuchten Ensemblesimulationen diesen kausalen Zusammenhang jedoch nur begrenzt reproduzieren können. Dennoch kann in den Fällen,

in denen die Ensemblesimulationen den kausalen Zusammenhang erfolgreich reproduzieren, eine höhere Zuverlässigkeit der Vorhersage des Oberflächenklimas erreicht werden, als in den übrigen Fällen. Meine Ergebnisse zeigen so auf, dass ein Einsatz eines kombinierten KI-dynamischen Ansatzes zur Verbesserung der saisonalen Vorhersagen führen kann. Dieser Ansatz könnte sogar operationell angewendet werden, um tatsächliche saisonale Vorhersagen des europäischen Sommerklimas zu verbessern.

PUBLICATIONS RELATED TO THIS DISSERTATION

Appendix A:
Carvalho-Oliveira, J., Borchert, L.F., Zorita, E., Baehr, J. (2022). "Self-Organizing Maps Identify Windows of Opportunity for Seasonal European Summer Predictions". *Frontiers in Climate*, 4, https://doi.org/10.3389/fclim.2022.844634.

Appendix B:
Carvalho-Oliveira, J., Di Capua, G., Borchert, L.F., Rousi, E., Donner, R., Zorita, E., Baehr, J. "Causal associations and predictability of the East Atlantic Pattern." - *to be submitted*.

ACKNOWLEDGEMENTS

Arriving at this finish line would not have been possible without an immense support from incredible people. First of all I would like to deeply thank my supervisors Johanna Baehr, Leonard Borchert and Eduardo Zorita for their guidance and motivation through all the challenges I encountered in this PhD journey. I would like to especially thank Johanna for being such a wonderful mentor through all these years. From the most brilliant scientific ideas and thorough feedback to many thoughtful advices on life, thank you for being such an inspiring person. I am also extremely grateful to Leo, whom I have had the pleasure to work closely with over the past 6-7 years, and in so many ways encouraged me and supported me to arrive where I am today. I would also like to thank Edu for his support, helpful advices and insights, and for always being available when I needed. I would like to further thank Thomas Ludwig for having co-supervised me in the beginning of my PhD, and my panel chair Jochem Marotzke for his constructive feedback and support over these years.

I also would like to thank Efi Rousi, Giorgia di Capua and Reik Donner, whom I had the pleasure to work with in Potsdam, and then through so many hours over Zoom. Our discussions motivated me immensely and I am very grateful for the opportunity of collaborating with you. A special thank you goes to Mauro Cirano, from whom I learned so much since my undergrad times and provided a huge support when I most needed. I would like to further thank the IMPRS-ESM school for providing such an amazing and valuable structure to something so chaotic as a PhD. Thanks to Michaela Born, Cornelia Kampmann, and particularly to Antje Weitz for her immense support through the darkest days of pandemic.

The Climate Modelling group has been a big component of my journey as a scientist, from masters to PhD. I would like to thank all of you for contributing to such an inspiring and friendly environment. I am grateful to your constructive feedback and interesting discussions, and to all the fun moments in person and online – it is a pleasure working with you! A special thank you goes to Lara for carefully revising this dissertation. I would like to further thank the KSI group for the friendly environment, and to Marlene for the company in the beginning of my PhD.

I cannot imagine surviving these PhD years without support of the amazing friends I made in Hamburg. I could not be more thankful to Deb, Lari, Ingrid, Lara and Fede for all sorts of support you gave me, from printing posters and cooking for me in the busiest times to amazing trips and countless climbing sessions together. And to all my dearest old friends Roos, Wilton, Gus, Je, Nati, Mari, Anita, Rafa, Sote and Beca: thank you for being present no matter how far! A massive thank you goes to Roos, for the endless hours skyping, yearly autumn pepernoten sessions, incredible climbing trips and for carefully revising this dissertation.

Finally, I owe my sincerest gratitude to my wonderful family, especially to my parents, sister, amazing dogs and dearest husband. Oliverinhas, obrigada por sempre terem acreditado em mim, pelo amor incondicional e por me apoiarem em toda essa jornada. Lipe, holding your hand since the very first days of this PhD has been the most precious thing to me. Thank you for making this possible!

CONTENTS

UNIFYING TEXT

1

TOWARDS IMPROVED SEASONAL CLIMATE PREDICTIONS WITH ARTIFICIAL INTELLIGENCE: AN APPLICATION ON SUMMER TELECONNECTIONS

1.1 INTRODUCTION

Predicting changes in the atmosphere is not trivial. We are used to regularly monitor numerical weather predictions issued by our local weather services, despite knowing that chances are high that those convey only little information to help us make decisions. An even more difficult endeavour is to predict trends in the evolution of the atmosphere a few months ahead, or *seasonal climate prediction*[1] (e.g. Doblas-Reyes et al., 2013). Such predictions rely on physics-driven dynamical climate models which resolve physical equations, thereby describing the dynamics of a simplified modelled version of the atmosphere. As opposed to the physics-based approach of dynamical climate models, artificial intelligence (AI) methods can learn complex nonlinear relationships from atmospheric data without relying on physical laws (e.g. Scher, 2018). This dissertation promotes the combination of AI and physics-based climate predictions to improve our ability to issue reliable climate predictions a season ahead. I specifically apply this AI-dynamical approach to investigate one of the hardest climate prediction problems: seasonal climate prediction of the European summer.

To illustrate this choice, let us assume that we want to predict whether next summer in Europe is going to be warmer than usual, perhaps to advise a vineyard owner on the best timing for their harvest. As a first approximation, we could consider in our model only external forcing from increasing concentration of anthropogenic greenhouse gases (IPCC, 2013) and issue a prediction of *likely warmer* temperatures. But besides external forcing, much of the seasonal climate variability in Europe is determined by the chaotic nature of the atmosphere. Coupling processes with slowly varying components of the Earth system further contribute to the complexity of the problem (e.g. Shuila and Kinter III, 2006). A reliable seasonal prediction thus requires accurate domain knowledge and thorough process understanding, which makes seasonal forecasting for Europe particularly challenging. State-of-the-art climate models provide European seasonal predictions with marginal credibility, in particular for the summer season (e.g. Mishra et al., 2019).

Over the past decade, increasing effort has been made towards improving the reliability of seasonal summer climate predictions to provide useful information to climate-sensitive sectors (e.g. Vitart and Robertson, 2018). Recurrent summer atmospheric circulation patterns, the so-called *teleconnections*, are amongst the drivers for high-impact climatic processes, including heatwaves and floods (e.g. Teng et al., 2013). Hence, a reliable seasonal prediction of summer teleconnections would benefit a range of users in sectors such as agriculture, energy and policy-making. Such

1 In seasonal climate *predictions* – or *forecasts* – we are interested in deviations from the "usual" climate state, which is often determined by a climatological mean calculated over a time period of 30 years. Please note that the terms "forecast" and "prediction" are used as synonyms in this thesis.

a reliable prediction becomes even more relevant with climate change, given the frequency of extreme weather and climate events projected to increase (Rahmstorf and Coumou, 2011).

Yet, the mechanisms underlying summer teleconnections are not well understood nor well represented in current climate models (e.g. Carvalho-Oliveira et al., 2022; Delgado-Torres et al., 2022). I show in this dissertation how advances in AI combined with climate modelling offer a powerful framework to improve our understanding of these mechanisms and achieve improved seasonal summer predictions. In the remainder of this chapter, I provide a foundation to understand how I reach this goal. I discuss important aspects concerning summer teleconnections, thereby highlighting relevant scientific caveats that lead to the framing of my research questions.

1.1.1 *What are teleconnections?*

We start off by having a close look at the organised chaos of the atmosphere. At timescales longer than 10 days, atmospheric variability in the extratropics tends to be organised in recurrent space-time circulation patterns, the so-called atmospheric *modes of variability* (Trenberth, 2022). These modes can link climate variability across remote distances via atmospheric pathways, an important property known as teleconnection (Trenberth, 2022).[2] This ability to link remote regions is closely related to the presence of planetary-scale atmospheric waves known as Rossby waves, which act to restore atmospheric energy budget imbalances (Liu and Alexander, 2007). The El Niño-Southern Oscillation (ENSO) is the most prominent example of a teleconnection. It results from ocean-atmosphere coupling and is responsible for connecting fluctuations in the sea surface temperature (SST) taking place in the tropical Pacific with changes in weather occurring across large parts of the globe (e.g. Trenberth et al., 1998).

Teleconnections can be divided into three classes according to their geographical extent and causes (Fig.1.1; e.g. Coumou et al., 2018). In summer, the main teleconnection class comprises the zonally symmetric modes, which in the northern hemisphere is known as the Northern Annular Mode (NAM, Thompson and Wallace, 2000), also referred to as the Arctic Oscillation. NAM corresponds to hemispheric-wide variations in position and strength of the westerly winds, featuring pressure anomalies of opposing signs between the Arctic and midlatitudes (Fig.1.1a). Its most prominent expression is over the North Atlantic, best known as the North Atlantic Oscillation (NAO, Walker, 1924).

During summer in the Euro-Atlantic region, the NAO and East Atlantic Pattern (EA, Wallace and Gutzler, 1981) are the main examples of a second class of teleconnections, the regional long-wave variations occurring in a sector of the globe (Fig.1.1b,c, respectively). Summer NAO and EA stand out as teleconnections because they are also linked to a third class of teleconnections, related to the amplification of Rossby waves. This process can promote trapping and focusing effects of tropospheric winds, known as the circumglobal wave-train (Branstator, 2002; Wu and Lin, 2012)[3]. It pro-

2 For this reason, atmospheric modes of variability are often simply referred to as "teleconnections". Both terms are used interchangeably in this thesis.

3 Two other teleconnections become relevant when describing the large-scale atmospheric circulation in the Euro-Atlantic sector during winter and spring: the Scandinavian Pattern and the East Atlantic

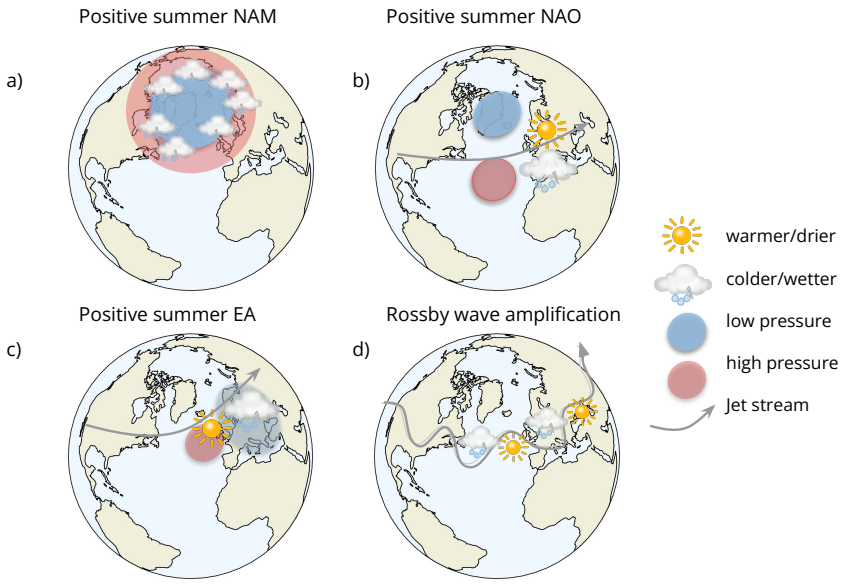


Figure 1.1: Schematics representing the major types of summer atmospheric teleconnections recurring in the Euro-Atlantic: a) Northern Annular Mode in positive phase. b) North Atlantic Oscillation in positive phase. c) East Atlantic Pattern in positive phase. d) Rossby wave amplification forming a circumglobal wave-train leading to synchronised weather events along its path. Adapted from Coumou et al., 2018.

duces zonally oriented chains of perturbations around the globe, associated with persistent alternating hot-dry and cold-wet conditions which are prone to development of severe weather extremes in summer (Fig.1.1d, e.g. Coumou et al., 2014). The summer NAO and EA teleconnections are the focus of this dissertation.

1.1.1.1 *Summer North Atlantic Oscillation*

"Thus it seems well established that there is a sort of oscillation in the pressure of the air between a centre of action at high pressure and another neighbouring one at low pressure. (...) For the moment, we cannot say how these great changes in air pressure take place, which extend over areas so large that they often cover an entire hemisphere." — Hildebrandsson [1897]

Hildebrandsson, 1897 is considered the forerunner for all future studies that led to the characterisation of the NAO. Walker, 1924 was the first study to define the NAO as a dipole of alternating sea level pressure (SLP) changes between the Azores and Iceland, the so-called Azores High and Icelandic Low. This spatial SLP oscillation results from the redistribution of atmospheric mass between the Arctic and the subtropical North Atlantic, with significant influence on weather across much of the Northern Hemisphere. The two most common definitions of the summer NAO are station-based and empirical orthogonal function (EOF)-based (Hurrell et al., 2003).

Western Russia. In summer, however, these patterns exert a relatively weak effect in comparison to the NAO and EA, thus being usually neglected (Bueh and Nakamura, 2007; Hall and Hanna, 2018; Josey et al., 2011; Zha et al., 2022).

While the station-based definition calculates the normalised SLP difference between the Azores and Iceland, the EOF-based is the time series associated with the first EOF of June-July-August SLP in the Euro-Atlantic region (70°–40°E, 25°–80°N), explaining between 22-39% of SLP variance depending on the investigated time period and dataset used (e.g. Comas-Bru and Hernández, 2018; Folland et al., 2009).

During negative phases of the NAO, the meridional pressure gradient is weak and shows enhanced pressure over the Arctic and lower pressures in the subtropics. These conditions lead to a decrease in the strength of the jet stream, which shifts equatorwards, thereby moving the path along which midlatitude storms preferred form – the storm tracks. In summer, this is associated with increased precipitation and cooler conditions over northwestern Europe, and warm and dry conditions in southern Europe (Folland et al., 2009). Summer 2012 is an outstanding example of such impact, when a sustained negative phase of the NAO was linked to widespread flooding events in the United Kingdom and strong drought and wildfires in Spain (Dong et al., 2013b). Positive NAO phases in summer show reversed conditions (see schematics in Fig.1.1b).

1.1.1.2 *Summer East Atlantic Pattern*

About sixty years after the seminal work by Sir Gilbert Walker defining the NAO (Walker, 1924), Wallace and Gutzler, 1981 were the first to identify the EA teleconnection. As opposed to the well-established characterisation of the NAO, the exact definitions and spatial features of the EA are still a matter of debate[4]. While some studies describe the EA as an SLP seesaw similar to the NAO, albeit more zonally distributed and shifted southwards (e.g. Barnston and Livezey, 1987; Neddermann et al., 2018), other studies refer to a single anomalous SLP centre located west of the British Isles (e.g. Comas-Bru and McDermott, 2014; Moore et al., 2013; Ossó et al., 2018). In order to focus on the physical processes underlying the EA, throughout this dissertation I refer to "East Atlantic Pattern", the EA, as the second dominant teleconnection in the Euro-Atlantic sector after the NAO. I take this general terminology to provide a framing of the physical aspects that are already known to influence (or are influenced by) this teleconnection, thus disregarding at this point which methods were used (see Sec. 1.1.4 for a discussion on the methods). Despite these discrepancies, perhaps the most important EA feature lies in its spatial overlap with the NAO, implying that the EA modulates NAO strength (Woollings et al., 2010). This coupling between EA and NAO teleconnections has been the focus of several recent studies, which showed that this NAO-EA interplay controls climate across different spatio-temporal scales, with important ecological and societal impacts (Bastos et al., 2016; Hall and Hanna, 2018; Zubiate et al., 2017).

Although the characteristics of EA in winter are usually well described in literature, only a few studies have focused on its component in other seasons (e.g. Ossó et al., 2018; Wulff et al., 2017). The most common definition of the summer EA corresponds to the second EOF of June-July-August SLP in the Euro-Atlantic region, which explains between 10-17% of SLP variance (Comas-Bru and Hernández, 2018).

4 To this day, several terms and methodologies have been proposed to analyse the EA, turning comparison among studies a challenge by itself. EA has also been referred to as the "Summer East Atlantic" (Wulff et al., 2017), the "Zonal Pressure Gradient" (Neddermann et al., 2018), the "Sea Level Pressure Index" (Ossó et al., 2018) or "Atlantic Low" and "Atlantic Ridge" (Cassou et al., 2005).

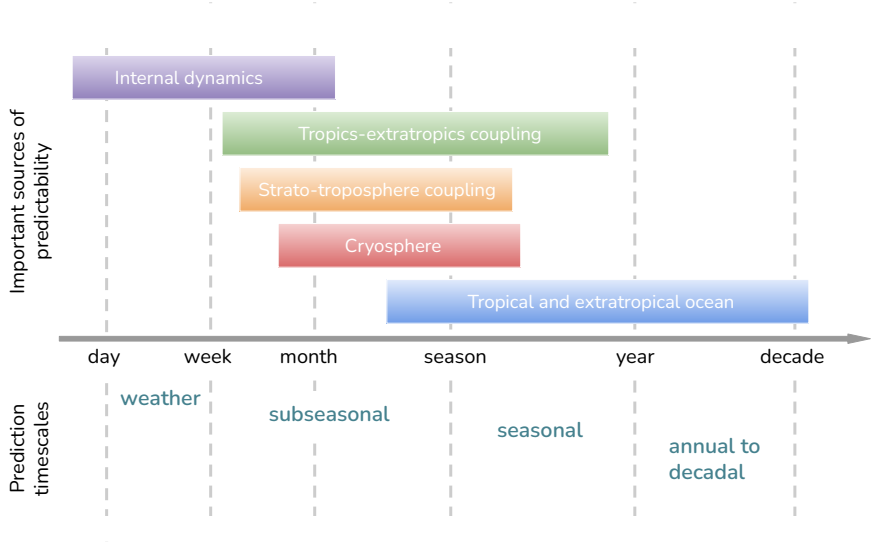


Figure 1.2: Schematics representing the main sources of summer predictability for Euro-Atlantic teleconnections across timescales. Adapted from Merryfield et al., 2020.

A positive phase of winter EA is characterised by anticyclonic conditions west off the British Isles[5]. Such conditions have been associated with decreased precipitation in the United Kingdom (Comas-Bru and McDermott, 2014), drier conditions over Western Europe, and below-average surface temperatures in southern Europe (Moore et al., 2011).

In summary, in this section I present an overview of the main characteristics and most common definitions of the Euro-Atlantic summer teleconnections, introducing the focus of my study on the summer NAO and EA. In Sec. 1.1.4 I reintroduce this topic by discussing the challenges associated with the analysis of these teleconnections. In preparation for this discussion, we first need to understand which physical processes influence the summer NAO and EA, and in particular how these processes impact their *predictability*.

1.1.2 *Sources of predictability*

Most teleconnections in the extratropics, including the summer NAO and EA, are intrinsic features of the atmosphere. They are primarily driven by internal dynamical processes, such as interactions between eddies and the mean flow (e.g. Barnes and Hartmann, 2010), which have short predictability timescales (Weiland et al., 2021; see Fig.1.2). Teleconnections may also arise through interactions with external drivers, or *forcings*, such as the stratosphere (Domeisen et al., 2020), the cryosphere (Hall et al., 2017) and in particular the ocean (Gastineau and Frankignoul, 2015).

External forcings are important for influencing the temporal variability of teleconnections beyond weather timescales, often increasing their persistence and thereby the theoretical capability to predict them at longer timescales (e.g. Ossó et al., 2020;

5 I use the polarity originally defined in Wallace and Gutzler, 1981, and followed by a number of studies (e.g. Comas-Bru and McDermott, 2014). However, some studies use a reversed polarity, with a positive phase of EA described as cyclonic conditions west off the British Isles (e.g. Barnston and Livezey, 1987; Wulff et al., 2017).

Wang and Ting, 2022). That is, external forcings influence the *predictability* of teleconnections across timescales and constrain the ability of climate models to predict their evolution in time.

Despite most predictability studies for the Euro-Atlantic region focusing on winter (e.g. Dunstone et al., 2016; Scaife et al., 2014), recent studies showed evidence of external forcings active in the summer season (Fig.1.2). One of the most prominent examples is the coupling between tropics and extratropics, which is typically associated with ENSO at seasonal timescales. ENSO-related SSTs can be accurately predicted by most climate models several months in advance, thus constituting a primary source of seasonal predictability (Doblas-Reyes et al., 2013). But while an impact of ENSO on both winter NAO and summer EA has been suggested by several studies (e.g. Mezzina et al., 2020; Wulff et al., 2017, respectively), ENSO exerts only a weak influence on summer NAO (e.g. Folland et al., 2009; Hall et al., 2017).

Stratospheric and cryospheric processes have also been linked to increased predictability of teleconnections. Several studies conclude that both stratospheric (e.g. Hansen et al., 2017; Scaife et al., 2005) and cryospheric (e.g. Cohen and Jones, 2011) forcings exert an important influence on winter NAO, by primarily modulating the jet stream strength. However, only a few studies have suggested an influence on summer NAO (e.g. Hall et al., 2017; Wang and Ting, 2022), and no such link has yet been found for EA neither in winter (Maidens et al., 2021) nor summer.

Yet, there is increasing evidence that both tropical and extratropical oceanic forcings play a key role for the predictability of NAO and EA across a range of timescales (e.g. Athanasiadis et al., 2020; Czaja and Frankignoul, 2002). Particularly for the summer case, the North Atlantic has been shown to influence seasonal predictability of these teleconnections via persisting SST patterns, a topic that I further discuss in Sec. 1.3.

So far I have shown how summer teleconnections impact the surface climate in the Euro-Atlantic region, and discussed the main external processes influencing their seasonal predictability. An important question thus arises: are climate models able to predict the variability of summer climate and to reproduce summer teleconnections at seasonal timescales?

1.1.3 *Summer seasonal climate predictions*

Despite the discussed prospects for summer predictability on seasonal timescales, current state-of-the-art climate models generate predictions with rather limited predictive *skill* (e.g. Doblas-Reyes et al., 2013; Mishra et al., 2019). Generally speaking, skill refers to the ability of a climate model (or prediction system) to produce a useful prediction – what I refer to as *credibility* in Sec. 1.1. A common approach to verify the skill of a climate model is to compare its predictions against some reference system, such as observations from past events. In this way, verification of a climate model consists of performing predictions for the past in so-called *hindcast* mode (Troccoli, 2010).

A hindcast, also known as re-forecast, is a prediction initialised from past conditions that can be verified against known past events. Such predictions are often performed as a set of realisations known as *ensemble* prediction, which is designed to account for a sample of the possible physical pathways or climatic states given the

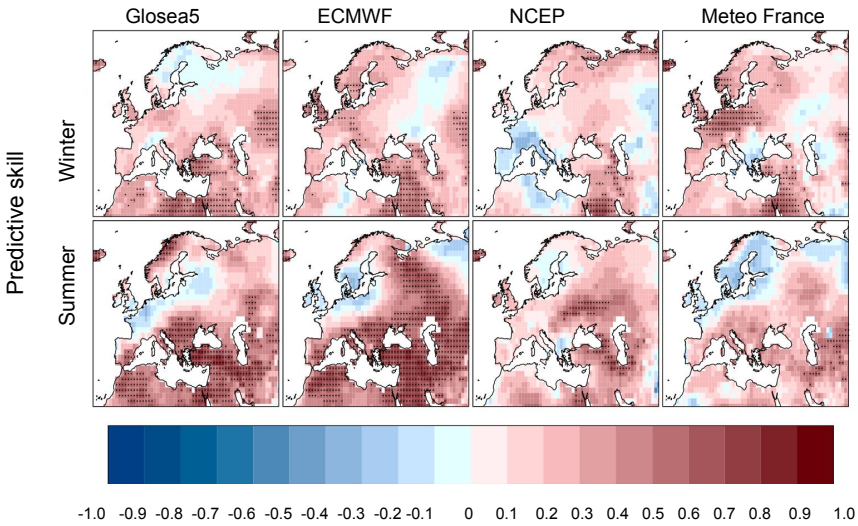


Figure 1.3: Comparison of predictive skill amongst four prediction systems, for summer and winter. Predictive skill is assessed with anomaly correlation coefficient (ACC) between the predicted ensemble mean of each individual climate model of EUROSIP and the observed seasonal summer (June-July-August mean) temperature obtained from Era-Interim reanalysis for the period 1992–2012. Forecasts are initialised in May. Areas covered in red correspond to positive correlations, and indicate an agreement between model and reanalysis. This agreement is significant at the 95% confidence interval for areas covered by dots. Adapted from Mishra et al., 2019.

distribution of initial conditions (Troccoli, 2010). The skill of an ensemble of predictions is usually assessed by averaging all realisations (denoted *ensemble mean*), thus reducing the unpredictable component of the model (Eade et al., 2014).

To illustrate these concepts, I show in Fig.1.3 a comparison of skill assessment amongst four seasonal prediction systems, referring to the prediction of summer and winter air temperature over Europe during the period 1992–2012. The skill of each climate model is calculated by comparing its ensemble mean against Era-Interim reanalysis, calculated with anomaly correlation coefficient[6] (ACC, Collins, 2002).

Firstly, we can see that predictions in winter show higher skill than those in summer. Moreover, all climate models are particularly limited in skilfully predicting air temperature in northwestern and central Europe in both seasons. Likewise, climate models tend to show low skill in reproducing the observed variability of summer teleconnections (e.g. Beverley et al., 2019; Delgado-Torres et al., 2022; Lledo et al., 2020). Several factors have been proposed to explain these limitations, including inappropriate model resolution (Cattiaux et al., 2013), biases in the jet stream position (Pithan et al., 2016), and ensemble overdispersion due to initialisation issues (Ho et al., 2013).

This limited skill of seasonal climate predictions has motivated several studies to perform a conditional skill assessment (Fig.1.4), using a concept also known as *windows of opportunity* (Mariotti et al., 2020). Windows of opportunity concern identifying conditions associated with enhanced predictability for a given problem, such as

6 The anomaly correlation coefficient (ACC) varies between values -1 and 1, with ACC = 1 indicating a perfect agreement between model and reference.

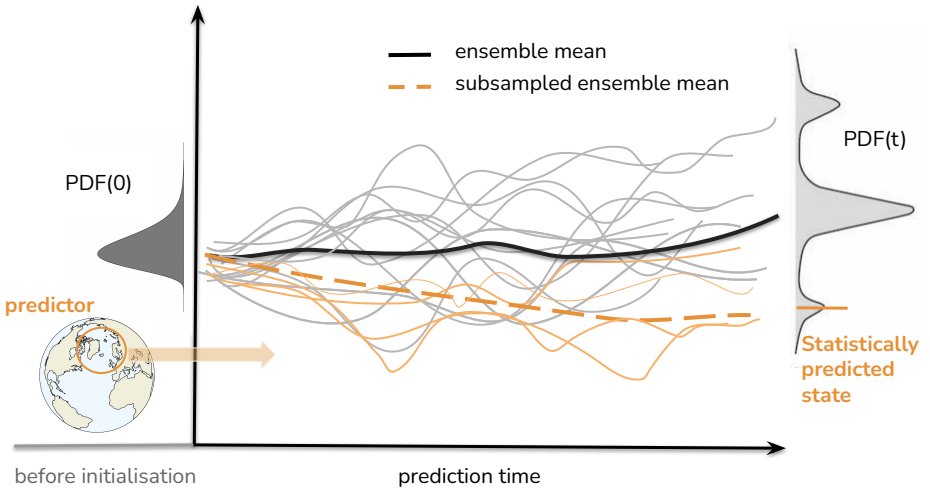


Figure 1.4: Illustration of a conditional skill assessment with ensemble subsampling. An ensemble of predictions (thin lines) for a given target variable is generated by integrating the model from initial conditions represented by the probability density function at time o (PDF(o)). The ensemble at prediction time t (lead time) consists of a sample of PDF(t), which represents the distribution of all possible climatic states at time t. To represent this ensemble, a mean over all members gives the ensemble mean (thick black line). In ensemble subsampling, physical predictors known to influence the target variable are used to select a sample of the possible states generated by the ensemble (orange thin lines). A mean over these selected ensemble members gives the subsampled ensemble mean (orange dashed line).

via a physical process known to exert an important influence (e.g. Carvalho-Oliveira et al., 2021). Several studies demonstrated improved seasonal climate skill by combining this concept with *ensemble subsampling*, which selects a sample of ensemble members informed by a chosen predictor to calculate the ensemble mean (Carvalho-Oliveira et al., 2022; Dobrynin et al., 2018; Neddermann, 2019).

In this section, I discuss the relevance of targeting teleconnections to detect conditions that favour more predictable states in summer seasonal predictions. This leads us to a final aspect that needs to be considered before implementing such a task: how do we properly define, identify and quantify teleconnections?

1.1.4 *On the complexities of analysing teleconnections*

Although several methods have been proposed to monitor teleconnections, extracting physically relevant information from the chaotic atmosphere is a challenging task (e.g. Shepherd, 2014). In this section, I discuss the linear and nonlinear approaches, commonly used in teleconnection analysis. I highlight their limitations and motivate how AI methods belonging to the field of *machine learning* can overcome those to a large extent. This discussion, therefore, leads to scoping out the general goal of my dissertation and the framing of my overarching research question.

1.1.4.1 *Linear approach*

The first approach to quantify a teleconnection was a correlation map depicting the NAO, illustrating the linear relationship between polar SLP and indices derived from fixed stations across the Northern Hemisphere (Exner, 1913). As introduced in Sec. 1.1.1, station-based monitoring of NAO is amongst the most common methods to analyse its variability. While this method provides a continuous record of the teleconnection, it has the disadvantage of not capturing spatial structures nor changes in the position of the Azores High and Iceland Low over time. This approach is less used in the monitoring of the EA (Comas-Bru and Hernández, 2018), which often relies on EOF-based analysis.

In EOF analysis, large atmospheric datasets are reduced to a set of orthogonal spatial patterns, generated such that each pattern explains as much variance as possible. Each given spatial pattern is associated with a time series that represents its amplitude over time (principal components, PC). That is, EOF analysis is not based on physical principles, but rather on mathematical ones: a field of SLP, for example, is split into independent orthogonal patterns ordered by relative variance, which may be interpreted as teleconnections. Consequently, spatial patterns concerning the same physical process (the same teleconnection) with similar variances may be wrongly split into different EOF patterns (North et al., 1982). The orthogonality constraint may also lead to the blending of patterns deriving from unrelated physical processes (e.g. Reusch et al., 2007), thus making a physical interpretation very difficult.

Another limitation of an EOF-based analysis for the summer NAO and EA concerns the assumptions of stationarity and spatial symmetry between opposite phases, which seem to be only partly adequate (e.g. Cassou et al., 2004; Rieke et al., 2021; Weisheimer et al., 2019). Changes in the variability of a teleconnection over time (nonstationarity) can generate nonphysical EOF patterns, artificial trends and amplitude changes in the associated PCs (Tremblay, 2001). Moreover, the symmetry assumption relies on the hypothesis that positive and negative phases of a teleconnection show identical spatial structure of the anomalies but opposing polarity, which does not agree with observations of the summer NAO or EA (e.g. Cornes et al., 2013). To better understand the implications arising from these limitations, we need to take another perspective, the one introduced by a nonlinear approach.

1.1.4.2 *Nonlinear approach*

The nonlinear paradigm of *climate regimes* to analyse teleconnections in the Euro-Atlantic was first introduced in Cassou et al., 2004 for the winter, and later in Cassou et al., 2005 for the summer case. This perspective is similar to the concept of "weather regimes"[7] (Vautard, 1990) used in synoptic meteorology, and can be seen as the temporal integration of the latter. These studies applied the so-called *k-means* clustering algorithm on atmospheric data (e.g. Michelangeli et al., 1995) to determine four teleconnection patterns, which in summer are reminiscent of the positive and negative phases of the NAO and EA in the linear approach. The four summer teleconnections

7 *Weather regimes* describe recurrent and persistent states of the atmospheric circulation over a given specific region, often computed using k-means clustering (e.g. Michelangeli et al., 1995). Clustering is a multivariate statistical method that groups data into a pre-defined number of classes (clusters) according to a measure of similarity.

are referred to as NAO in positive (also known as Blocking) and negative phases, Atlantic Ridge (similar to the positive phase of EA) and Atlantic Low (similar to the negative phase of EA). This approach offers a compromise to overcome the spatial symmetry assumption imposed by EOF, thus accommodating the observed spatial asymmetry between positive and negative phases. This asymmetry is strongest for the NAO during winter, but it is also present in the summer case for both NAO and EA (see. Fig.3 in Cattiaux et al., 2013 for a visual example).

Yet, the most interesting aspect of the nonlinear paradigm concerns analysing the frequency of occurrence of a given teleconnection. From this perspective, the atmosphere is viewed as a complex dynamical system (Lorenz, 1963), where low-frequency climate variability[8] is reflected in the preferred *transition* between teleconnections and changes in their *occurrence*. Relating these changes with the variability of external forcings could therefore provide important information on the predictability of teleconnections and thereby identify windows of opportunity for skilful summer climate predictions. Which is exactly what I intend to do in this dissertation. However, some drawbacks become evident when using clustering methods for this purpose. In order to best implement this task, I need to introduce two optimised AI algorithms belonging to the group of machine learning (ML) methods.

1.1.4.3 *Machine learning approach*

The first ML algorithm I use in this thesis is the so-called *Self-Organising Maps* (SOM, Kohonen, 1984). SOM belongs to the field of unsupervised neural networks, which is comprised of algorithms designed to learn and classify patterns from data without prior knowledge of any labelling. As a neural network-based classifier, SOM consists of two fully-connected layers that can be visualised as a two-dimensional grid, denoted SOM map. As for all neural network methods, one drawback of SOM is that it contains hyperparameters[9] that must be chosen a priori by the user. Carrying out a sensitivity test of these parameters is therefore required, upon evaluation of the research question and calculation of performance metrics (Forest et al., 2020). The biggest advantage of this method over clustering on the analysis of teleconnections is that it treats the data as a *continuum* and preserves the *topology*[10] of the data (e.g. Leloup et al., 2007). To understand these concepts, let us assume we apply SOM on a three-dimensional SLP dataset corresponding to the Euro-Atlantic region.

During training, SOM spans the entire SLP data space to identify a set of *modes* that best represent the nearby SLP fields, while still describing the distribution of the dataset when considered all together. The learning process is iterative and controlled by a neighbourhood function, responsible for preserving a topological ordering of the data. This property assures that neighbouring modes correspond to similar SLP patterns and widely separated ones correspond to dissimilar patterns (Kohonen, 2013). Once trained, our SOM map therefore offers an efficient way to visualise and

8 Low-frequency variability stands for timescales between 10 and 50 days, i.e. longer timescales than baroclinic systems (e.g. storms), but shorter than a season (e.g. Hannachi et al., 2017).

9 In machine learning, a *hyperparameter* refers to a parameter whose value controls the learning process.

10 A topology is a system of subsets in the data, which can be interpreted as neighbouring, related data points. A topology preserving classifier implies that neighbouring patterns presented in the output space are also neighbouring patterns in the input space. This property preserves quantitative relationships in the dataset and thus facilitates the interpretation of SOM-derived patterns as physical patterns.

analyse transitions amongst modes of SLP that comprise a continuum of teleconnection patterns in the physical space.

While SOM has been applied to analyse both winter and summer teleconnections in the Euro-Atlantic region (e.g. Johnson et al., 2008; Polo et al., 2011; Rousi et al., 2017, Gu and Gervais, 2022), these studies concerned observations or historical model simulations at synoptic or decadal timescales. In the first part of this dissertation, I use the SOM framework to analyse the interannual variability of summer teleconnections in the Euro-Atlantic during the 20th century, both in observations and simulations from an initialised seasonal prediction system (Sec. 1.2). I further explore SOM to analyse dependencies between teleconnections and other climatic variables. I give a particular focus to identify potential North Atlantic precursors, which I then test in the model (Sec. 1.3).

Once these potential precursors are identified, I focus on understanding their underlying mechanisms and causal-effect relationships in the second part of this dissertation (Sec. 1.4). Causality in climate sciences has been conventionally assessed in two ways: via correlations or climate modelling. On one hand, correlations cannot be interpreted as a measure of causality (Pearl et al., 2000). Interactions with other variables, such as indirect links or common drivers, as well as autocorrelation effects (Runge et al., 2019) can lead to spurious links. On the other hand, dedicated climate model experiments can be designed to isolate effects (e.g. by applying different forcings) and presumably enable a causal analysis (e.g. Osborne et al., 2020). But besides high computational costs, model biases in the representation of relevant physical processes (e.g. Beverley et al., 2019) can make the interpretation of causal-effect relationships using this approach challenging.

Alternatively, I make use of a second ML method, the so-called *Causal Effect Networks* (CEN, Runge et al., 2015) to investigate causal relationships between the North Atlantic and summer teleconnections. CEN is based on the *Conditional Independence Causal Discovery* framework (Pearl et al., 2000) and is suitable for testing hypotheses of causal-effect relationships among physical variables (e.g. Di Capua et al., 2020b; Kretschmer et al., 2016). CEN is able to detect confounding effects, overcoming several disadvantages imposed by correlation analysis, and thereby facilitating process understanding. As a last step, I use CEN as a foundation to investigate the variability of predictive skill in an initialised seasonal prediction system. Thus, in my dissertation I aim at answering the following overarching question:

How can machine learning help addressing the complexities of predicting summer Euro-Atlantic teleconnections at seasonal timescales?

I give an overview of the two publications related to this dissertation in the next sections, providing answers to my specific research questions (Sec. 1.2-1.4). I then provide a summary of my main conclusions and final remarks in Sec. 1.5. We start off with Sec. 1.2, where I present the first neural network-based analysis of initialised simulations generated by the Max Planck Institute Earth System Model (MPI-ESM) seasonal prediction system.

The central goal of this dissertation is established: investigate with a ML framework to what extent summer teleconnections in the Euro-Atlantic can be predicted one season ahead. To achieve this, I first use the neural network-based classifier SOM to characterise these teleconnections, both in observations and in the model (Carvalho-Oliveira et al., 2022, see Appendix A). SOM outperforms both linear (e.g. Liu et al., 2006) and clustering (e.g. Astel et al., 2007; Budayan et al., 2009) methods in partitioning data, and allows the evaluation of intermediate modes associated with a given teleconnection (e.g. Hunt et al., 2013). These aspects motivate my first research question:

1. How does SOM characterise the interannual variability of summer Euro-Atlantic teleconnections occurring in the 20th century?

I assess the interannual variability of summer Euro-Atlantic teleconnections in terms of spatial features and frequency of occurrence. I focus on the high summer (Cassou et al., 2005), and evaluate monthly July and August SLP data spanning the period 1902–2008. I use ERA-20C reanalysis (Poli et al., 2016) for observations, and model simulations from a 30-member hindcast ensemble with MPI-ESM in mixed resolution (hereafter: MR-30). The ensemble is initialised every May between 1902 and 2008. I use ERA-20C data to train SOM and thereby define a continuum of observed summer teleconnections. Next, I perform this SOM training for each individual ensemble member in the model, and compare the results to observations.

I find an optimum SOM map size with 3x4 dimensions, i.e. generating 12 SOM modes. My sensitivity analysis finds that larger SOM map sizes (e.g., 5x5 as in Polo et al., 2011, for daily data) show qualitatively similar modes although split in duplicate patterns. Since SOM is an unsupervised ML method, the next step consists of labelling these SOM modes guided by expert knowledge (e.g. Kashinath et al., 2021). In this case, labelling means to associate a SOM mode with a summer teleconnection, i.e. summer NAO or EA in positive or negative phases. I choose the methodology used in Cassou et al., 2005 as a reference to label the modes using a similarity analysis[11].

Observed variability

The SOM approach reveals a strong spatial asymmetry between modes belonging to reversed phases of NAO and EA. This is a robust feature that can be visualised in both summer SLP and 500 hPa geopotential height fields (Fig.A.2). I find an eastward shift of the pressure dipole (in particular the southern lobe) for a positive phase, as opposed to a negative phase, as reported for the winter NAO case (Cassou et al., 2004; Luo et al., 2018). SOM modes related with summer NAO occur more frequently (56%) than those with summer EA (44%). Nonlinearities of the method are expressed in the difference of occurrence between positive and negative phases, with summer NAO (EA) occurring 31% (20%) of the times in negative and 25% (24%) in positive

11 In the paper (Appendix A), I adopt the nonlinear terminology of Cassou et al., 2005 and refer to NAO and EA teleconnections as positive and negative NAO, Atlantic Ridge and Atlantic Low (see Sec. 1.1.4.2). In the text, I keep the NAO and EA terminology used so far for consistency.

phases. These findings are in line with the observed trends in the summer NAO index (e.g. Hanna et al., 2015).

As opposed to EOF or clustering, SOM provides a detailed characterisation of the summer teleconnections spatial features. Each teleconnection can be represented by a set of SOM modes (e.g. 3 for positive summer NAO), which show variations in the intensity and position of the cyclonic and anticyclonic pressure centres. For example, pronounced pressure centre shifts are evident for summer EA in negative phase (SOM modes 11 and 12 in Fig.A.2), a teleconnection that has been previously associated with heatwave occurrence in Europe (e.g. Duchez et al., 2016). These large scale shifts in the circulation act as a dynamic driver for variability of the surface climate modulated by the jet stream (e.g. Belmecheri et al., 2017), and can be visualised by anomalies of air temperature associated with each SOM mode (Fig.A.10). Interestingly, I find that the most frequent SOM mode associated with a negative summer EA (SOM mode 12) is associated with negative temperature anomalies in Europe. And only when anticyclonic conditions shift from Greenland towards Europe (leading to the occurrence of SOM mode 11), positive temperature anomalies occur (Neddermann et al., 2018; Wulff et al., 2017). Predicting the occurrence of such mode is therefore more informative than predicting a single index for summer EA, as usually done in conventional forecast practices.

How does the model perform?

I find that MR-30's performance in reproducing spatial features varies amongst different teleconnections, being highest for a positive, and lowest for a negative phase of EA (0.85 and 0.7 mean correlations, respectively; Fig.A.3). Moreover, the model performance is higher for a negative than for a positive phase of NAO (0.8 and 0.73 mean correlations, respectively). While results for the spatial features suggest a moderate performance for MR-30, comparing modelled and observed frequencies of occurrence reveals a stronger limitation in the model.

This discrepancy between simulated spatial and temporal features agrees with results from performance assessments of other models (e.g. Cattiaux et al., 2013; Cortesi et al., 2017; Delgado-Torres et al., 2022; Fabiano et al., 2020). Despite MR-30 showing a high intra-ensemble variability, it tends to overestimate the observed frequencies of summer NAO. I find that MR-30 shows particularly high underestimation in predicting the occurrence of a negative phase of EA, thus limiting its potential to predict early warning of warmer than average summers. In contrast, MR-30 is mostly able to predict the observed frequency of a positive phase of EA, and could represent with its occurrence windows of opportunity for skilful summer predictions (Mariotti et al., 2020).

To illustrate this concept and motivate my next steps, I propose in Fig.1.5 a thought experiment: what if we could predict which of these teleconnections is most likely to occur in a summer – how would seasonal summer predictive skill improve? In this ideal scenario, the dominant summer teleconnection is a priori known and can be used as a predictor for a conditional skill assessment with ensemble subsampling (see Fig.1.4). That is, instead of calculating an ensemble mean with the full ensemble to compare with observations, I calculate a "refined" ensemble mean using my predictors. I perform this test with MR-30 and find a significant skill improvement in

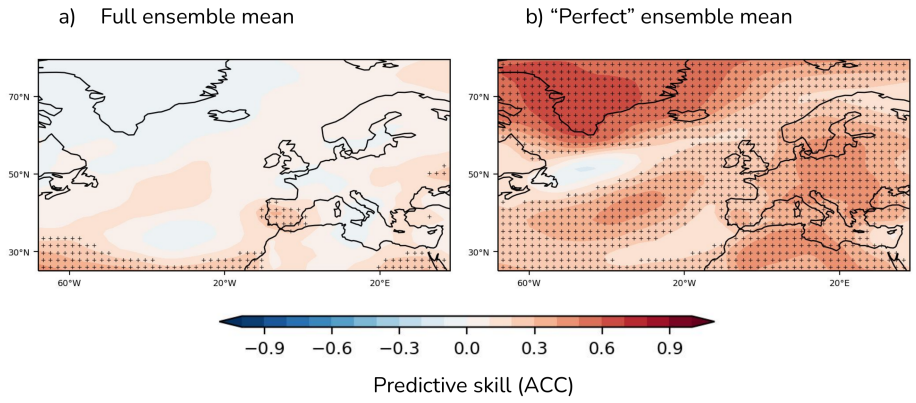


Figure 1.5: How would summer predictive skill improve if summer teleconnections could be "perfectly" predicted by MR-30? a) July-August SLP predictive skill for a mean over the full 30-member ensemble against ERA-20C in 1902-2008 (measured as anomaly correlation coefficient, ACC). b) Same as a), but for a "perfect" ensemble mean, where only ensemble members that predict the observed summer teleconnection are selected. Stippling represents a correlation significant at the 95% level.

the prediction of summer SLP 3-4 months ahead of time in the Euro-Atlantic for the "refined" ensemble mean, compared to the "full" (compare Fig.1.5a,b).

This improvement has two main implications. Firstly, it represents an upper limit of skill that could be reached with MR-30 if the model were to always simulate the correct summer teleconnection. Note that despite overall significant skill increase, ACC values are constrained by the model ability in reproducing the teleconnections (Fig.A.3). Secondly, it reflects a reduction of the ensemble spread, which grows too large because of weaknesses in the ensemble generation in the data assimilation system (e.g. Ho et al., 2013). Subsampling acts towards enhancing the signal of a given teleconnection by excluding ensemble members that drifted away from the correct physical path (e.g. Dobrynin et al., 2018), thereby increasing skill. In the next section, I take a step further and show how such a conditional skill assessment can be achieved in practice, using information from spring predictors generated by SOM.

1.3 LINKING SUMMER PREDICTIVE SKILL AND THE NORTH ATLANTIC

Several recent studies have suggested an influence of the North Atlantic on the seasonal predictability of summer atmospheric circulation (e.g. Dunstone et al., 2019). Both extratropical (Ossó et al., 2018) and tropical (Neddermann et al., 2018) SSTs in spring have been shown to persist into summer and influence summer EA, whereas an SST tripole pattern (Czaja and Frankignoul, 1999) has been suggested to influence summer NAO (Gastineau and Frankignoul, 2015; Hall et al., 2017). In this section, I analyse North Atlantic SSTs with SOM to derive a set of spring SST predictors that can be tested in MR-30 at prediction start (i.e. May initialisation). Identifying these SST predictors thus enables a conditional skill assessment with ensemble subsampling, as illustrated in Fig.1.5, but applicable in forecast mode. My guiding research question is:

2. How does predictive skill of summer hindcasts in MR-30 depend on North Atlantic SST at initialisation?

Here, my main hypothesis is that the seasonal predictability of summer teleconnections may be forced by the North Atlantic (e.g. Osborne et al., 2020), and that each spring SST predictor may excite a preferable summer teleconnection. This oceanic influence could arise from interacting diabatic heating and eddy vorticity forcings, altering the lower-tropospheric baroclinity over regions of strong meridional SST gradients (Nie et al., 2016; Peng et al., 2003). Persisting spring SSTs have been suggested to force the summer jet stream position, which shifts as a response to changes in the baroclinicity. Such mechanism has been linked to both summer NAO and EA teleconnections (Dunstone et al., 2019; Gastineau and Frankignoul, 2015; Osborne et al., 2020).

Identifying SST precursors

To identify spring SST predictors, I calculate monthly April SST composites in ERA-20C preceding each summer SOM mode obtained in Sec. 1.2 (Fig.A.4). The identified predictors are reminiscent of SST tripole and horseshoe patterns, which have been previously suggested to influence summer NAO (Hall et al., 2017) and summer EA (Duchez et al., 2016; Gastineau and Frankignoul, 2015; Ossó et al., 2018). I evaluate the linear relationship between these predictors and summer surface climate in the Euro-Atlantic, to identify target regions where the influence of spring SST is significant (e.g. Fig.A.5 for summer SLP). By calculating similarity with the SST predictors, I classify each April North Atlantic SST field in the assimilation experiment used to initialise MR-30. That is, upon initialisation of MR-30, a spring SST predictor carries subsampling information of which summer teleconnection is likely to dominate in a given year. This information thereby allows me to perform a conditional skill assessment of MR-30.

Conditional skill assessment

I perform this skill assessment for summer surface climate predictions by first calculating a mean over the full ensemble, and analysing summer SLP, 2-metre air temperature and 500 hPa geopotential heights predictions 3-4 months ahead. I find that summer surface climate skill in MR-30 seems to depend on the characteristics of North Atlantic SSTs present at the initialisation (Figs.A.6, A.11). This dependence could presumably reflect the variability in strength and location of meridional SST gradients (e.g. Kushnir et al., 2002), as well as their persistence (e.g. Deser et al., 2003). Skill spatial features differ amongst SST predictors, and are overall limited to the mid-North Atlantic and seldom significant over land. I find the highest skill values over the ocean for positive EA SST predictors, particularly offshore of the British Isles for summer SLP.

When performing subsampling, I find that spring SST predictors are not sufficient to indicate the dominant summer teleconnection, illustrated by a marginal skill improvement in summer SLP skill. There are two main reasons why these results are not particularly surprising. Firstly, besides oceanic forcing, other predictors have been

suggested to influence summer NAO (e.g. snow cover, Screen, 2013, Matsumura et al., 2014; sea ice concentration, Hall et al., 2017; stratosphere-troposphere coupling, Wang and Ting, 2022, to name a few) and summer EA (e.g. tropical precipitation, Wulff et al., 2017). Secondly, a correlation analysis suggests only a regional influence of spring SST on summer surface climate (Fig.A.5), dependent on the characteristics of each SST regime.

However, I find that spring SST predictors are partly able to indicate the phase of the dominant summer teleconnection via subsampling. This phase-based grouping can be interpreted with SOM's topology preserving feature, which assures that the position of modes in the SOM map reflects their similarity in the physical space (Fig.A.9). Besides, each phase is typically associated with a preferred jet stream position, which is more northerly (southerly) when summer NAO and EA are in positive (negative) phase (e.g. Dorado-Liñán et al., 2022; Trouet et al., 2018; Woollings et al., 2010).

While results of skill improvement for 2-metre air temperature and 500 hPa geopotential heights are marginal, I achieve regional enhanced SLP skill when summer EA occurs. This skill improvement is statistically significant over Scandinavia for a positive phase. For a negative phase, improvement takes place south of 55° N but is mostly statistically insignificant. These regions of skill improvement usually coincide with areas of significant correlation with the SST predictor (Fig.A.5). I find only marginal skill improvement for summer NAO, which could be due to MR-30's limited performance in representing it (Sec. 1.2, e.g. Osborne et al., 2020) or may indicate that other external forcings play a more important role for its seasonal predictability.

Next, I verify my results by performing a conditional skill assessment on an independent ensemble generated by the MPI-ESM-MR-based seasonal prediction system, covering the recent period 1980-2016 (Dobrynin et al., 2018). In comparison to MR-30's centennial scale simulations (1902-2008), I find a stronger dependence of summer EA to spring SST in the recent years covered by the independent ensemble. This is shown by a higher regional improvement across different variables, which significantly increases for 2-metre air temperature in Scandinavia (Fig.A.8, right column). The areas of improvement, however, do not always coincide with those in MR-30. Although these differences could be due to initialisation of each system, they raise the question whether the relationship between North Atlantic SSTs and EA is stationary, and most importantly whether it is causal. I address these questions in the next section, delving into the mechanisms behind this relationship by implementing a systematic evaluation of observations and MPI-ESM-MR within a causal-effect framework.

1.4 Causal pathways towards skilful summer seasonal predictions

I demonstrate in Sec. 1.3 that EA seasonal predictability is influenced by North Atlantic SST, whereby improved skill of summer climate predictions can be achieved. While there is no consensus on the physical processes underlying this link between North Atlantic SST and summer EA, two mechanisms have been proposed in observations: Rossby wave activity driven by ENSO-related *tropical* SST forcing (Wulff et al., 2017), and shifts in the jet stream position driven by *extratropical* North Atlantic SST forcing (Ossó et al., 2018). A link between EA and tropical SST forcing is, how-

ever, likely to be nonstationary (Rieke et al., 2021). This therefore limits its use to constrain summer predictive skill to only certain periods of time.

Furthermore, a link between extratropical North Atlantic SST and summer EA (hereafter: SST-EA link) has been suggested for a recent period starting from 1979 (Ossó et al., 2018). The characteristics of this link at longer timescales, however, are not yet known. Here, I address this gap by disentangling the causal-effect pathways between extratropical North Atlantic SST and summer EA over a long observational record. I go beyond previous studies by applying an AI-based causal discovery tool to investigate the variability of predictive skill in the initialised ensemble MR-30. I therefore aim at answering the following two research questions:

3. To what extent are spring extratropical North Atlantic SSTs causal drivers for the summer EA?

4. How does skill of MPI-ESM-MR at predicting European surface climate depend on its representation of Euro-Atlantic causal links?

I answer these questions in two complementary steps. First, I analyse observations covering the 20th century with the ERA-20C reanalysis (Poli et al., 2016). Then, I analyse three independent sets of MPI-ESM-MR simulations: a pre-industrial control run (piControl), a historical run, and the initialised ensemble MR-30 used in Sec. 1.3.

Causal drivers of summer EA

I start addressing the first research question by carrying out a variability analysis for the period 1908-2008. In Fig.1.6a I use two indices to represent the variability of spring extratropical North Atlantic SST (in green) and summer EA (in orange) (see Sec. B.3.2 for details). It is possible to visualise that both SST and EA undergo a strong variability through the analysed period. I find significant correlations of 0.54 during the last 30 years, in line with Ossó et al., 2018. Correlations are insignificant when considering the full period, dropping to 0.2. In Fig.1.6b I compute averaged correlations over a 20-year period to visualise the low-frequency variability of this link. I find that the SST-EA link is nonstationary, showing a significant change in strength between first (blue) and second (pink) halves of the 20th century (hereafter: *early* and *late* periods, respectively).

I speculate that this nonstationarity could result from a confounding effect due to an external process that modulates the strength of this link. Testing for confounding effects in climatic data is not possible with conventional statistical methods such as correlations. I therefore perform a causality analysis with the AI-based CEN (Sec. 1.1.4.3) for early and late periods, separately. CEN performs an interactive conditional independence test (Pearl et al., 2000), assessing whether the SST-EA link is significant when conditioned on another potential predictor, such as ENSO-related tropical SSTs (Wulff et al., 2017). Although I find a spurious, noncausal link in the early period, I show that the SST-EA link in the late period is causal, confirming suggestions of previous studies (Ossó et al., 2018, 2020). This causal link has an estimated mean strength expressed by a path-coefficient (Runge et al., 2015) of about 0.2, i.e. a 1 standard deviation change in SST causes a 0.2 standard deviation change in

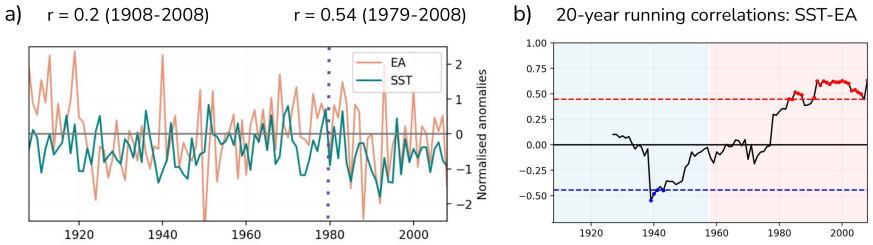


Figure 1.6: Observed variability of the SST-EA link in the 20th century. a) Time series of SST (green) and EA (orange) indices smoothed by a 3-year running mean (see text for definitions). Purple dashed line illustrates the beginning of the period investigated in Ossó et al., 2018. b) Running-correlation between SST and EA indices for a 20-year window. Coloured markers indicate significant correlations at the 95% confidence interval, illustrated by dashed lines.

the EA 3-4 months later (Fig.B.3). These findings shed light on the mechanisms proposed to influence summer EA, suggesting a more active role of extratropical North Atlantic SSTs than previously thought.

Is the model able to reproduce the causal link?

To answer the second research question, I first evaluate whether MPI-ESM-MR can reproduce the observed SST-EA causal link. I find that the model is unable to reproduce the SST-EA causal link when analysing both a 1000-year piControl simulation (Giorgetta et al., 2011; Mauritsen et al., 2012), and a historical simulation (period 1850-2005) under natural and anthropogenic forcing following CMIP5 (Dobrynin et al., 2018). Yet, I find that the model is able to reproduce this link in the initialised MR-30 ensemble at certain times, suggesting a possible role of initialisation. Skill is nevertheless limited, with MR-30 tending to underestimate the SST-EA link strength (Fig.B.6), which might explain its low skill in predicting summer seasonal European climate (e.g. Carvalho-Oliveira et al., 2022; Neddermann et al., 2018). However, I do find that those ensemble simulations that reproduce the SST-EA causal link tend to show improved skill in certain regions, over those that do not (Fig.B.6). This finding suggests that a causal analysis could be particularly useful to identify target regions under strong influence of a predictor (i.e. SST in our case), where predictions could be potentially skilful.

I illustrate these ideas in Fig.1.7, showing a causal map that represents the causal influence of spring SST on summer SLP and air temperatures in observations. I identify a target region west of the British Isles under strong causal influence of SST. I test the hypothesis that ensemble simulations able to reproduce a causal link with SST in this region are more likely to show local skilful predictions, in comparison to simulations that do not. I assess this by randomly choosing one ensemble member amongst the 30-member ensemble every year during the late period, a process which I repeat 2000 times. I then assess the skill of all sampled 45-year long time-series against observations in predicting summer surface climate 3-4 months ahead.

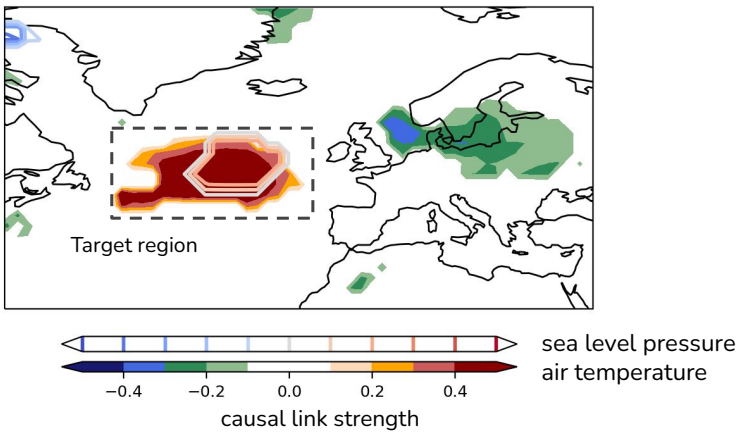


Figure 1.7: Could regions under causal influence from a SST predictor be a target for skilful predictions? Plot shows the causal influence of an SST predictor calculated from spring North Atlantic SSTs onto summer SLP (contours) and 2-metre air temperature (shading). This causal influence is expressed in terms of a causal link strength coefficient calculated with the AI-tool CEN, which denotes by how much summer SLP or air temperature change (in standard deviation) when the SST predictor changes by 1 standard deviation.

I find higher skill for predictions of SLP, 500 hPa geopotential heights and 2-metre air temperature in simulations which reproduce the causal link with SST, over those that do not. These findings suggest that AI-based causal discovery could complement conventional predictive skill assessments to identify opportunities for skilful seasonal predictions.

1.5 SUMMARY AND CONCLUSIONS

This dissertation proposes an AI-dynamical approach to advance the understanding of the physical mechanisms influencing the seasonal predictability of Euro-Atlantic summer teleconnections, thereby achieving improved seasonal summer climate predictions a season ahead. In this section I present a summary of my conclusions, revising the research questions raised throughout this dissertation, and providing answers to each of them. I finish this section with an outlook, drawing final thoughts concerning the findings presented in this dissertation.

1.5.1 *Answers to the research questions*

In Sec. 1.2, I implement an AI-dynamical approach with the neural network-based classifier SOM. I specifically assess the variability of summer Euro-Atlantic teleconnections during 1902-2008, first in observations, and then in the initialised ensemble MR-30. SOM outperforms the conventional statistical methods used in teleconnection analysis, overcoming their assumptions of stationarity, spatial symmetry and orthogonality (e.g. Liu et al., 2006). Moreover, as opposed to clustering, SOM en-

ables the characterisation of intermediate atmospheric modes associated with each teleconnection. These aspects led to my first research question:

1. How does SOM characterise the interannual variability of summer Euro-Atlantic teleconnections occurring in the 20th century?

- Summer teleconnections are characterised by a set of 12 SOM modes, which for reversed phases of summer NAO or EA show a pronounced spatial asymmetry, in agreement with observations.
- SOM modes related to summer NAO occur more frequently (56%) than those to summer EA (44%). In line with observations, I find differences in the occurrence of each teleconnection among positive and negative phases, with summer NAO (EA) occurring 31% (20%) of the times in negative and 25% (24%) in positive phases.
- For a given teleconnection, I find zonal shifts in the pressure centres amongst SOM modes, which in turn are reflected as anomalies in the summer 2-metre air temperature fields. This spatial characterisation becomes particularly relevant when analysing the occurrence of negative summer EA, which can be a driver for European heatwaves.
- I find that while MR-30 can moderately reproduce the spatial features of summer NAO and EA, it has a very limited performance in reproducing their timing.
- MR-30 shows particularly high underestimation in predicting the occurrence of a negative summer EA, thus limiting its potential to predict early warning of warmer than average summers.
- By contrast, MR-30 shows a high agreement with observations when reproducing a positive summer EA, both in terms of spatial features and timing. Hence, the occurrence of positive summer EA could represent windows of opportunity for skilful summer predictions in this model.

Motivated by these findings, in Sec. 1.3 I investigate how the occurrence of summer teleconnections relates to the predictive skill of Euro-Atlantic summer surface climate in MR-30. I verify my results analysing an independent seasonal prediction system covering the recent period 1980-2016 (Dobrynin et al., 2018). Furthermore, I take an AI-dynamical approach to investigate how the predictability of summer teleconnections is linked to North Atlantic SSTs, which have been suggested to influence their occurrence (Gastineau and Frankignoul, 2015; Osborne et al., 2020). I derive a set of SST predictors, allowing me to carry out a conditional skill assessment of MR-30 with ensemble subsampling. Hence, my second research question is:

2. How does predictive skill of summer hindcasts in MR-30 depend on North Atlantic SST at initialisation?

- Summer predictive skill in MR-30 depends on spring North Atlantic SST anomalies present at the initialisation of hindcasts, particularly for SLP. I classify these

anomalies in a set of 12 SST predictors, reminiscent of SST tripole and horseshoe patterns, previously associated with the predictability of summer teleconnections.

- Skill spatial features differ amongst SST predictors, and are overall limited to the mid-North Atlantic and seldom significant over Europe. I find the highest skill values over the ocean for positive EA SST predictors, particularly offshore of the British Isles for summer SLP.

- I find that the SST predictors are partly able to indicate the phase of the dominant summer teleconnection via ensemble subsampling, whereby I achieve improved SLP skill 3-4 months ahead in parts of Europe. This improvement is significant over parts of northern Europe and Scandinavia for a positive summer EA, where correlation with the SST predictor is highest. In contrast to summer EA, I find a marginal improvement for summer NAO, suggesting that its seasonal predictability is weakly associated with North Atlantic SST.

- I find a higher dependence of the summer predictive skill to spring North Atlantic SST over the more recent period covered by the independent ensemble (1980-2016), in comparison to MR-30's centennial simulations (1902-2008). This is illustrated by a higher regional improvement across different variables for the independent ensemble, which particularly increases for 2-metre air temperature in Scandinavia.

In summary, these findings contribute to the growing body of evidence that the predictable component of summer teleconnections at seasonal timescales may be larger than previously thought, thereby suggesting an active role of the North Atlantic in particular for summer EA. In Sec. B.3.2, I further investigate the specific role of extratropical North Atlantic SST (hereafter: SST) in forcing the summer EA, using the AI-based causal discovery tool CEN. I go beyond previous studies (Ossó et al., 2018, 2020) and analyse a long observational record to assess the variability of the SST-EA link, and testing if SST is a causal driver for summer EA. I then evaluate three independent sets of MPI-ESM-MR simulations to compare whether the model is able to reproduce the characteristics this link, and how this may affect summer predictive skill. I address the following two research questions:

3. To what extent are spring extratropical North Atlantic SSTs causal drivers for the summer EA?

- I find that the observed SST-EA link is nonstationary during the 20th century, showing distinct variability between early (1902-1957) and late (1958-2008) periods.

- I perform a causal discovery analysis and find that, whereas the SST-EA link is spurious in the early period, SST is a causal driver for the summer EA in the late period. I calculate an estimated causal influence on summer EA of 0.2 standard deviation change, when spring SST changes by 1 standard deviation.

- I test for a confounding influence of ENSO-related tropical forcing, which has been suggested as another predictor for summer EA. I find that the SST-EA

causal link remains significant when conditioned on ENSO-related tropical forcing, suggesting a more active role of extratropical North Atlantic SSTs than previously thought.

4. How does skill of MPI-ESM-MR at predicting European surface climate depend on its representation of Euro-Atlantic causal links?

- I find that whereas piControl and historical simulations with MPI-ESM-MR are unable to reproduce the SST-EA causal link detected in observations, the initialised ensemble MR-30 is able to reproduce this link at certain times, suggesting a possible role of initialisation.
- MR-30's performance in reproducing the SST-EA link is, nevertheless, limited. I find that the model tends to underestimate the SST-EA link strength, which can explain its low skill in predicting summer seasonal European climate.
- However, I find that MR-30 simulations that can reproduce the SST-EA causal link tend to show regional improved skill, over those that do not. I show this impact on skill for a region of strong SST causal influence west of the British Isles. I find higher skill for predictions of SLP, 500 hPa geopotential heights and 2-metre air temperature in simulations which reproduce the causal link with SST, over those that do not.

In summary, I show that MPI-ESM-MR has limited performance in reproducing a causal link between spring SST and summer EA, offering an explanation why this model shows poor skill in predicting European summer climate at seasonal timescales. My findings suggest that AI-based causal discovery could complement conventional predictive skill assessments to identify opportunities for skilful seasonal predictions.

1.5.2 *"No such thing as a free lunch"*

Perhaps one of the most important assumptions in AI is reflected in the *No free lunch* theorem (Wolpert, Macready, et al., 1995): there is no universal learning algorithm. I show in this dissertation how AI can benefit our endeavour to predict summer climate a season in advance. Any combination of AI and climate modelling, however, requires a very careful consideration of the problem at hand. The tools used in this dissertation might not be best suited for other problems, and collaboration between climate and computer scientists can be crucial to tailor the best AI-dynamical approach. My dissertation is the effort of such an approach, whereby I overcome some of the challenges for seasonal prediction of the European summer climate.

APPENDICES

A

SELF-ORGANIZING MAPS IDENTIFY WINDOWS OF OPPORTUNITY FOR SEASONAL EUROPEAN SUMMER PREDICTIONS

The work in this appendix has been published as:

Carvalho-Oliveira, J., Borchert, L.F., Zorita, E., Baehr, J. (2022). "Self-Organizing Maps Identify Windows of Opportunity for Seasonal European Summer Predictions". Frontiers in Climate, 4, https://doi.org/10.3389/fclim.2022.844634.

SELF-ORGANIZING MAPS IDENTIFY WINDOWS OF OPPORTUNITY FOR SEASONAL EUROPEAN SUMMER PREDICTIONS

Julianna Carvalho-Oliveira [1,2,3], Leonard F. Borchert [4,5], Eduardo Zorita [2], Johanna Baehr[1]

[1] Institute for Oceanography, Center for Earth System Research and Sustainability, Universität Hamburg, Hamburg, Germany

[2] Helmholtz-Zentrum Hereon, Institute of Coastal Systems - Analysis and Modeling, Geesthacht, Germany

[3] International Max Planck Research School on Earth System Modelling, Max Planck Institute for Meteorology, Hamburg, Germany

[4] École Normale Supérieure, LMD, Institut Pierre Simon Laplace (IPSL), Paris, France

[5] Sorbonne Universités (SU/CNRS/IRD/MNHN), LOCEAN Laboratory, Institut Pierre Simon Laplace (IPSL), Paris, France

Received: 28 December 2021 / Accepted: 10 February 2022 / Published: 01 April 2022

A.1 ABSTRACT

We combine a machine learning method and ensemble climate predictions to investigate windows of opportunity for seasonal predictability of European summer climate associated with the North Atlantic jet stream. We particularly focus on the impact of North Atlantic spring sea surface temperatures (SST) on the four dominant atmospheric teleconnections associated with the jet stream: the summer North Atlantic Oscillation in positive and negative phases, the Atlantic Ridge, and Atlantic Low. We go beyond standard forecast practices by not only identifying these atmospheric teleconnections and their sea surface temperature precursors, but by making use of these identified precursors in the analysis of a dynamical forecast ensemble. Specifically, we train the neural network-based classifier Self-Organising Maps (SOM) with ERA-20C reanalysis and combine it with model simulations from the Max Planck Institute Earth System Model in mixed resolution (MPI-ESM-MR). We use two different sets of 30-member hindcast ensembles initialised every May, one for training and evaluation between 1902-2008, and one for verification between 1980-2016, respectively. Among the four summer atmospheric teleconnections analysed here, we find that Atlantic Ridge simulated by MPI-ESM-MR shows the best agreement with ERA-20C, thereby representing with its occurrence windows of opportunity for skillful summer predictions. Conversely, Atlantic Low shows the lowest agreement, which might limit the model skill for early warning of warmer than average summers. In summary, we find that spring SST patterns identified with a SOM analysis can be used to guess the dominant summer atmospheric teleconnections at

initialisation and guide a sub-selection of potential skillful ensemble members. This holds especially true for Atlantic Ridge and Atlantic Low, and is unclear for summer NAO. We show that predictive skill in the selected ensemble exceeds that of the full ensemble over regions in the Euro-Atlantic domain where spring SST significantly correlates with summer SLP. In particular, we find a significant improvement in predictive skill for SLP, geopotential height at 500 hPa, and 2 metre temperature at 3-4 months lead time over Scandinavia, which is robust among the two sets of hindcast ensembles.

A.2 INTRODUCTION

Seasonal predictability of European summer climate is closely linked to the leading modes of atmospheric teleconnections associated with the North Atlantic jet stream. In the Euro-Atlantic region, the jet stream controls the location of the storm track and modulates the occurrence of weather systems, thus acting as a dynamical control for large-scale temperature and precipitation regimes (e.g. Bladé et al., 2012; Dong et al., 2013a). Yet, current state-of-the-art seasonal prediction systems often show biased representations of the jet stream strength and position (Beverley et al., 2019), posing a constraint to the skillful prediction of large-scale features of the summer climate in the North Atlantic-European sector a season ahead (e.g. Dunstone et al., 2016). A further limitation is that dynamical seasonal prediction systems tend to produce overdispersive ensembles, for which the forecast uncertainty is higher than the forecast error (Ho et al., 2013). In contrast, relatively small forecast uncertainties presumably indicate more predictable climate states, which in turn reveals *windows of opportunity* for more skillful forecasts (Mariotti et al., 2020). Identifying conditions that favour more predictable states - the aim of the present study - is hence a crucial step to achieve improved seasonal forecasts.

Here, we go beyond standard forecast practices by combining an ensemble seasonal prediction system with the neural network-based classifier Self-Organising Maps (SOM) (Kohonen, 1984). This approach identifies a sub-ensemble in which simulated North Atlantic sea surface temperatures (SST) at the initialisation of the prediction system (i.e. April) are linked to the seasonal predictability of the two dominant modes of variability associated with the North Atlantic jet stream: the summer North Atlantic Oscillation (NAO) and East Atlantic Pattern (EA) e.g. Bastos et al., 2016; Folland et al., 2009.

Several studies suggested an influence of spring North Atlantic SST on the predictability of NAO and EA. Neddermann et al., 2018 showed that tropical North Atlantic SST in spring can be a predictor for a zonal pressure difference mode that resembles the EA, while Ossó et al., 2018 found that the source of predictability for the EA lies on the temperature gradient between subpolar and subtropical gyres. Gastineau and Frankignoul, 2015 and Hall et al., 2017 suggested that a similar temperature gradient may influence the predictability of NAO as well. Going beyond these studies, we use North Atlantic SST patterns in spring as predictors for both the NAO and the EA.

Traditionally, NAO and EA are defined as the first two empirical orthogonal functions (EOF) of summer SLP in the Euro-Atlantic region (Barnston and Livezey, 1987; Folland et al., 2009). Cassou et al., 2005 proposed an alternative approach using k-

means clustering (k=4) and defined four modes of summer variability: the NAO in positive and negative phases, and the Atlantic Ridge (At. Ridge) and Atlantic Low (At. Low). While the At. Low resembles the positive phase of the EA (Barnston and Livezey, 1987), the At. Ridge resembles the negative phase. NAO in positive phase and the At. Low are associated with warmer and drier conditions in northern and central Euro-Atlantic regions, and colder and wetter conditions over south. Generally, a negative NAO phase and At. Ridge show a reverse pattern (e.g. Cassou et al., 2005).

In this study we use SOM as an alternative tool to EOF to identify the main atmospheric teleconnections. We compare how well simulations with the Max Planck Institute Earth System Model in mixed resolution (MPI-ESM-MR) represent the spatial and temporal variability of sea level pressure and its co-variability with spring SST in the Euro-Atlantic sector. Specifically, we use two different sets of 30-member hindcast ensembles initialised every May, one for training and evaluation between 1902-2008, and one for verification between 1980-2016, respectively. We give particular focus to the influence of specific SST patterns in spring on the seasonal predictability of the main atmospheric teleconnections.

We adopt a SOM perspective over traditional EOF analysis for two main reasons. Firstly, a clear limitation for EOF is that all decomposed basis vectors must be orthogonal, which may lead to nonphysical or blended patterns (e.g. Reusch et al., 2005). Secondly, an EOF analysis requires stationarity, which cannot be assumed for century-long analysis of the North Atlantic jet stream, and likely neither to the SST-SLP relationship investigated here (e.g. Rieke et al., 2021; Weisheimer et al., 2019; Woollings et al., 2015).

Besides neglecting orthogonality and stationarity assumptions, SOM provides advantages to visualise and interpret spatial and temporal variability associated with the data. It assumes that the data exist on a continuum instead of in distinct categories, which are organised such that similar SOM modes are displayed close together in the SOM map. This fine classification allows for efficient application of SOM to explore large-scale, slow varying processes, with a number of successful applications for climate characterisation (e.g. Johnson, 2013; Polo et al., 2011). Furthermore, SOM has been recently applied as an assessment tool to pre-select both skillful models in an multi-model ensemble (Mignot et al., 2020), as well as skillful ensemble members in a single-model ensemble prediction system (Oliveira et al., 2020).

In this study we combine the concept of ensemble subsampling (e.g. Dobrynin et al., 2018) with SOM, and evaluate the potential of spring North Atlantic SST at initialisation in predicting skillful ensemble members in MPI-ESM-MR. We further assess to what extent this selection affects the predictive skill of European summer climate, in comparison to a traditional predictive skill analysis. The manuscript is structured as follows: Sect.A.3 describes our methodology, the SOM algorithm and datasets used for training and evaluation. We characterise the main observed and simulated summer atmospheric teleconnections in the Euro-Atlantic domain in Sect.A.4.1. In Sect.A.4.2 we perform a SOM training to identify regions in the domain with potential for skill improvement, and in Sect.A.4.3 we evaluate the hindcast skill. We present a discussion of the results in Sect.A.5, followed by conclusions in Sect.A.6.

A.3 METHODOLOGY

We organise our methodology in four main parts: data (Sect.A.3.1), training (Sect.A.3.2), evaluation (Sect.A.3.3) and verification (Sect.A.4.4). In Sect.A.3.1, we describe the reanalysis and test ensemble used for training and evaluation, an independent ensemble used for verification and the preprocessing methods adopted in our analysis. A brief introduction to the SOM algorithm, the steps taken for training SOM, and our predictor analysis is described in Sect.A.3.2. We characterise ensemble subsampling and hindcast skill analysis in Sect.A.3.3, and describe an independent verification in Sect.A.4.4. In Fig.A.1 we show a sketch representing our methods workflow.

A.3.1 *Data*

A.3.1.1 *Reanalysis*

We use monthly means of sea level pressure (SLP), geopotential height at 500 hPa (Z500), air temperature at 2m (T2m) and sea surface temperature (SST) from ERA-20C (Poli et al., 2016) and ERA-Interim (Dee et al., 2011) reanalysis products, spanning 1902-2008 and 1980-2016, respectively. For the three atmospheric variables we evaluate the North Atlantic-European sector covering 70°W-40°E, 25°-80°N, and for SST covering 90°W-40°E, 5°-80°N.

A.3.1.2 *Ensemble simulations*

We use two sets of 30-member hindcast simulations with the Max Planck Institute Earth System Model in its mixed-resolution setup (MPI-ESM-MR, Dobrynin et al., 2018). The atmospheric component ECHAM6 Stevens et al., 2013 has a resolution of T63L95, with an approximate horizontal resolution of 200 km (1.875°) and 95 vertical layers up to 0.01 hPa. The oceanic component MPI-OM Jungclaus et al., 2013 has a resolution of TP04L40, with an approximate horizontal resolution of 40 km (0.4°) and 40 vertical layers. External forcing is taken from CMIP5 Giorgetta et al., 2013. The first set of hindcasts is used for training and evaluation, and covers the period 1902-2008 (hereafter: test ensemble), and the second is used for an independent verification (Sect.A.3.4, see Fig.A.1) and covers 1980-2016 (hereafter: independent ensemble).

The test ensemble is initialised on 1st of May every year from 1902-2008, with initial conditions taken from an assimilation experiment. The assimilation experiment is performed using the MPI-ESM-MR with full-field nudging by Newtonian relaxation towards all atmospheric and ocean levels except in the boundary layer. The atmosphere conditions of vorticity, divergence, three-dimensional temperature and two-dimensional pressure are taken from ERA-20C. In the ocean, three-dimensional daily mean salinity and temperature anomalies are nudged at a relaxation time of approximately 10 days. The ocean state is derived in an ocean-only simulation performed with MPI-OM forced with the atmospheric variables from ERA-20C. The three-dimensional atmospheric and ocean fields of the assimilation experiment form the initial conditions, and ensemble members are generated by small perturbations of the atmospheric state by disturbing the diffusion coefficient in the uppermost layer.

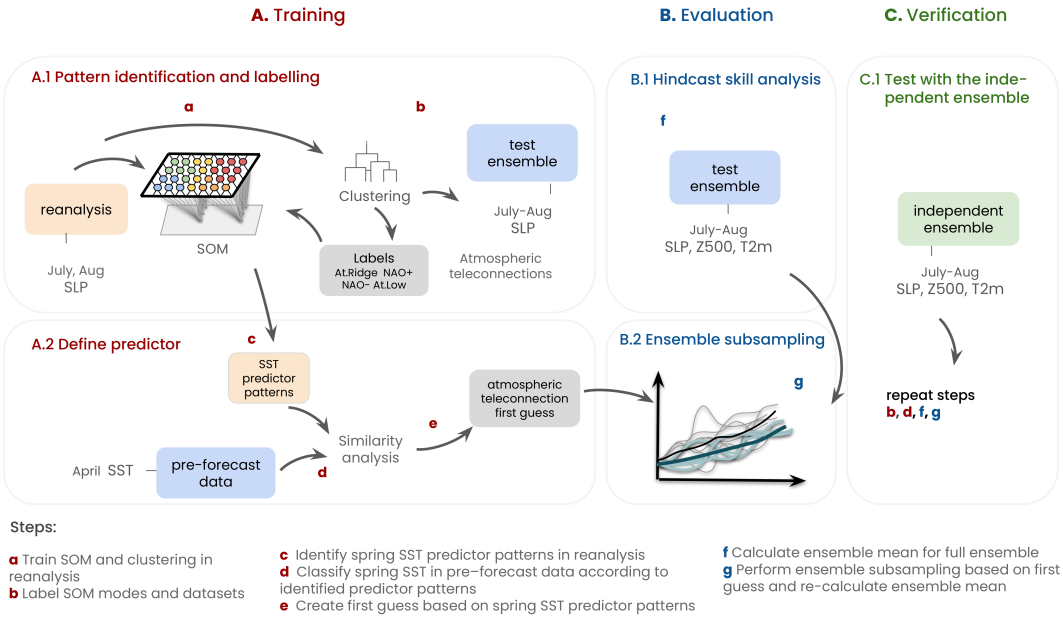


Figure A.1: Sketch of workflow. *A.1*: We use SOM to identify atmospheric teleconnections dominating the large-scale variability over the Euro-Atlantic region in 1902-2008. The final SOM output is a map consisting of twelve SOM modes. We train SOM with monthly July and August sea level pressure (SLP) in the reanalysis data. This same data is used to train a k-means clustering (k=4) to be used as a reference, whose centroids are labelled NAO+, NAO-, At. Ridge or At. Low as proposed in Cassou et al., 2005. *A.2*: We calculate pattern correlation between SOM modes and the k-means centroids to associate each summer SOM mode to a summer atmospheric teleconnection. We repeat this process with the k-means centroids to label each JA SLP field of the reanalysis and the hindcast. Lastly, we identify SST predictor patterns by calculating composite April SST patterns in the reanalysis w.r.t. SOM modes. Those SST predictor patterns are used to classify each North Atlantic April SST field in the pre-forecast data using pattern correlation. Given that each SST predictor pattern precedes a summer SOM mode associated with NAO+, NAO-, At.Ridge or At. Low, we are able to define a summer atmospheric teleconnection "first guess" before each initialisation. *B.1*: Using the test ensemble, we evaluate the hindcast skill for 3-4 month lead time considering the full ensemble to calculate the ensemble mean. *B.2*: We perform ensemble subsampling to re-calculate the ensemble mean by selecting only ensemble members whose labels agree with the "first guess", and compare with the full ensemble. *C.1*: We label, classify and subsample the independent ensemble as verification of our method.

Similarly, the independent ensemble is initialised in 1st of May every year from 1980-2016 from an assimilation experiment where ERA-Interim data (Dee et al., 2011) is assimilated in the atmospheric model component, and ORA-S4 data (Balmaseda et al., 2013) and National Snow and Ice Data Center observations (Comiso, 1995) are assimilated in the ocean/sea ice component.

We use July and August monthly means (3-4 months lead time) of SLP, Z500 and T2m from test and independent ensembles. At every gridpoint we compute anomalies by removing mean seasonal cycle and linear trend, in order to eliminate the centennial-scale climate change signal. We use July-August seasonal means (JA) to focus on the low-frequency dominant summer atmospheric teleconnections. In addition, we calculate spatial averages weighted by the cosine of latitude to take into account the dependence of the gridpoint density on latitude.

A.3.2 *Training*

A.3.2.1 *Pattern identification and labelling*

We use monthly July and August SLP fields from ERA-20C reanalysis to train *Minisom*, a Python implementation of Self-Organising Map (SOM) (Vettigli, 2019). SOM is a non-linear method based on unsupervised learning with two-layer neural networks (Kohonen, 1984). SOM's architecture allows for a reduced and ordered representation of high-dimensional datasets by a smaller set of variables. In a typical two-layer SOM, the input layer corresponds to feature vectors from the training dataset, while the output layer is the *SOM map*. The SOM map is a topological ordering of neurons usually in 2D grid (denoted SOM_{ij}, where i and j are the grid indices of the SOM map). This layer is fully connected to the input layer via weight vectors with the same dimension as the feature vectors. The lattice structure of the layers permit to calculate a measure of distance (here Euclidean distance) and identify the shortest path between neurons of both layers, thereby assigning as Best Matching Unit the respective closest neuron in the SOM map, iteratively. A fundamental property of SOM is the topological ordering: neighbouring neurons SOM_{ij} represent similar neurons in the input data space and therefore share similar properties. Hereafter we adopt *mode* as terminology to refer to neurons in the SOM map. For details on the SOM algorithm see Kohonen, 2013.

We train a 3x4 rectangular lattice of neurons (i.e. SOM_{34}) and choose training parameters as a compromise to keeping low quantisation and topological errors, while achieving a detailed view on the representative SOM modes associated with the summer atmospheric teleconnections defined in Cassou et al., 2005. We find that optimum parameters are *i)* initial spread of the neighbourhood function $\sigma(0) = 0.01$ (Gaussian) and *ii)* learning rate 0.8, for a maximum of 100000 iterations in batch training mode. Our tests with larger SOM sizes (e.g. 4x4, 5x5) showed qualitatively similar modes to the chosen 3x4 lattice, although showing duplicate patterns. Hence, a 3x4 SOM lattice balanced the need to represent the main summer atmospheric teleconnections with the least number of modes possible. The final output of the SOM training is a 12-mode SOM map (hereafter: SOM master). In addition to the SOM master, we perform a similar SOM training for each individual ensemble member of the test ensemble to allow comparison between model and reanalysis.

Next, we label each SOM mode as either At. Ridge, At. Low or NAO in positive or negative phase to allow comparison with other studies using EOF or k-means. We label by calculating pattern correlation between each SOM mode and the four centroids of a k-means clustering of ERA-20C JA SLP means, previously labelled according to Cassou et al., 2005; Cattiaux et al., 2013. We obtain similar labels if classifying the SOM modes with hierarchical agglomerative clustering using ward dissimilarity (Jain and Dubes, 1988). Note that during learning, each observation in the training dataset is associated with only one mode SOM_{ij} in the SOM map. In other words, this analysis assigns one dominant atmospheric teleconnection per summer each year.

A.3.2.2 *Define predictor*

We derive a set of twelve SST predictor patterns by calculating monthly April North Atlantic SST composites in the reanalysis w.r.t. the SOM master. That is, each SST composite pattern is a mean over the years associated with the input vector connected to a specific SOM mode in summer. We assume that each pattern is a potential predictor for one of the four atmospheric teleconnections (At. Low, At. Ridge, NAO+ or NAO-). To test this in the model, we classify each April North Atlantic SST field in the pre-forecast data according to the SST predictors via maximum pattern correlation. Thus, the most similar SST predictor defines a summer atmospheric teleconnection "first guess" before each initialisation, which we use as subsampling criteria in Sect.A.3.3.

A.3.3 *Evaluation*

A.3.3.1 *Hindcast skill analysis & ensemble subsampling*

The hindcast skill of MPI-ESM-MR against reanalysis data is assessed using pointwise detrended anomaly correlation coefficient (ACC) (Collins, 2002), which describes the model ability to reproduce the reference anomalies.

For single-model initialised ensemble prediction systems, ensemble subsampling consists of a post-processing technique that pre-selects potentially skillful ensemble members prior to a predictive skill assessment, given a statistical link to sources of high predictability (e.g. Dobrynin et al., 2018). We perform an ensemble subsampling with the aim of leveraging the MPI-ESM-MR ensemble prediction system before calculating the ensemble mean and assessing the hindcast skill. This procedure retains only ensemble members anticipated by the SST predictor to re-calculate the ensemble mean, i.e. selects a subsample of potential skillful ensemble members per year. That is, it allows us to investigate the conditional predictability of the SST predictor on the summer atmospheric circulation. This selection of ensemble members is used to re-calculate the ensemble mean of SLP, Z500 and T2m.

A.3.4 *Verification*

We use the independent ensemble (Sect.A.3.1.2) to evaluate the robustness of our spring SST predictors in selecting potential skillful ensemble members at the ini-

tialisation of summer predictions for the period 1980-2016. Even though "test" and "independent" ensembles derive from completely independent seasonal prediction systems, we acknowledge that an overlapping period (1980-2008) is present in the verification, implying that the conclusions drawn in this section must be taken with care. We stress, however, that excluding this overlapping period would lead to a very limited statistical analysis of the model predictive skill covering only the 8-year period 2009-2016, and thus preventing us from reproducing the analysis performed in Sect.A.3.3. Once more observed and predicted years become available, it would be relevant to perform this analysis without overlap with the training period.

To perform the verification, we firstly label the independent ensemble as described in Sect.A.3.2.1. Next, using the assimilation experiment as our pre-forecast data, we classify SST as described in Sect.A.3.2.2. Lastly, we perform a hindcast skill analysis and ensemble subsampling (Sect.A.3.3), comparing to ERA-Interim (Dee et al., 2011) reanalysis.

A.4 RESULTS

A.4.1 *Dominant summer atmospheric teleconnections*

We use a 3x4 SOM to represent the summer (July-August) sea level pressure variability spanning 1902-2008 in the Euro-Atlantic region using the reanalysis (Fig.A.2 shows the labelled SOM master, where the position of some modes differ from the original SOM map, see Fig.A.9). We group together SOM modes associated with the same main atmospheric teleconnections (i.e. NAO+, NAO-, At. Ridge or At. Low) and identify two main groups based on the meridional position of cyclonic and anticyclonic pressure centres. The first group comprises SOM modes 1-6, with NAO+ and At. Ridge modes, the second group comprises SOM modes 7-12, with NAO- and At. Low modes. In terms of the preferred jet stream position, the first group of atmospheric teleconnections is usually associated with northerly or central positions, while the second with southerly or central (e.g. (Trouet et al., 2018; Woollings et al., 2010). For simplicity, we refer to the former as northern jet group, and the latter as southern jet group.

NAO modes in positive (SOM modes 1-3) and negative (SOM modes 7-10) phases are located in opposite corners of the original SOM master, showing the highest topological distance and therefore the least level of similarity (c.f. Fig.A.9). Clear spatial asymmetries revealed by the non-linear method can be observed between the two phases. SOM modes 4-6 and 11-12 cover a range of wavy patterns related to At. Ridge (SOM modes 4-6), and At. Low (SOM modes 11-12). While NAO modes show significant moderate correlation to the correspondent first EOF of SLP (e.g. SOM mode 1: 0.36, $p < 0.05$ and SOM mode 10: -0.54, $p < 0.05$), no significant pattern correlation can be found with the second EOF of SLP. Yet, some modes bear high similarity with At. Ridge (SOM modes 4-6) and At. Low (SOM modes 11-12), reported in Cassou et al., 2005.

Next, we assess the agreement between atmospheric teleconnections in the reanalysis and those simulated by MPI-ESM-MR in the test ensemble (Fig.A.3 a, b). As opposed to a traditional evaluation using the ensemble mean, we instead analyse each ensemble member separately to assess the intra-ensemble variability. Spatially, we

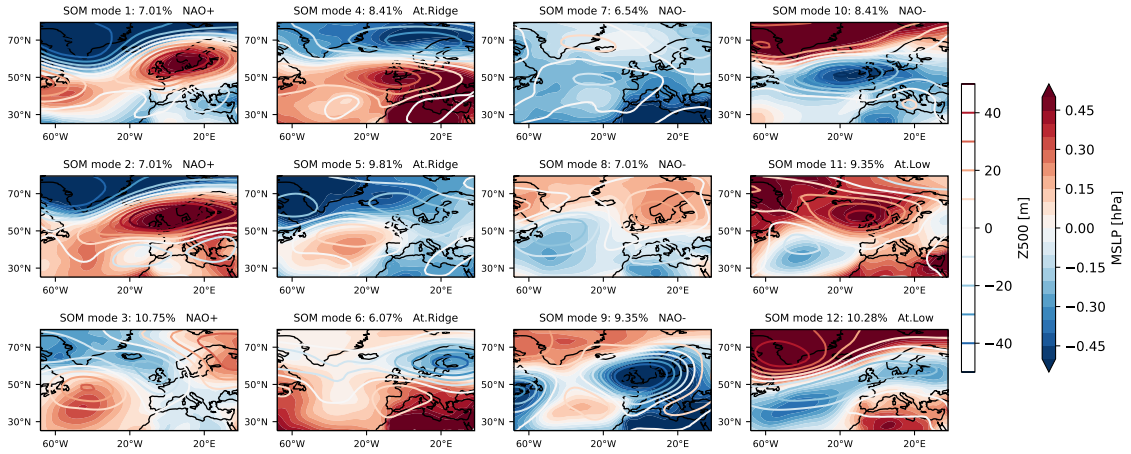


Figure A.2: SOM master representing the dominant summer atmospheric teleconnections in the North Atlantic European sector during 1902 - 2008, trained with ERA-20C July and August sea level pressure (SLP). Note that some modes were re-positioned with respect to the original SOM map to facilitate the interpretation of results.

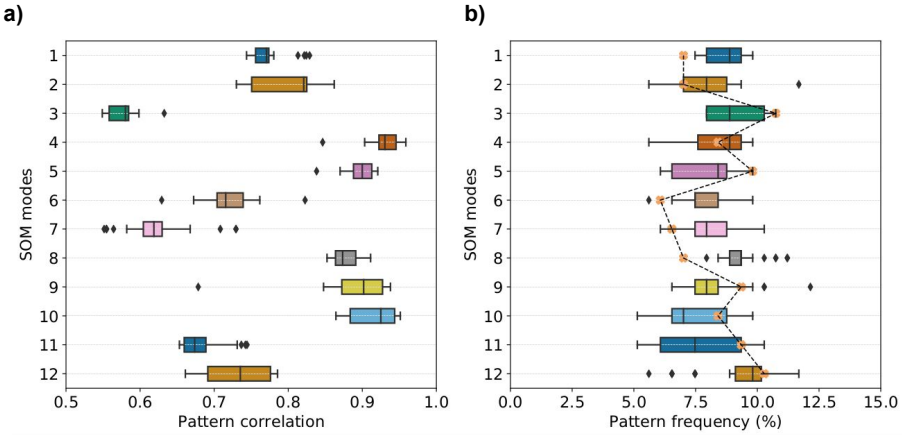


Figure A.3: Model agreement with reanalysis: a) Pattern correlation between observed and simulated SOM modes (per ensemble member). b) Associated frequency of observed and simulated SOM modes (per ensemble member). X markers and dashed line represent observed frequencies.

find overall moderate pattern correlation values, with At. Low modes (SOM modes 11-12) and one NAO+ mode (SOM mode 3) differing the most between model and reanalysis. In contrast, At. Ridge modes SOM 4-5 show the best agreement. While pattern frequency results (Fig.A.3b) show high intra-ensemble variability, the model fails at times to encompass the observed frequency (e.g. SOM modes 1, 6, 8). The model tends to underestimate the frequency of At. Low (SOM modes 11-12), and

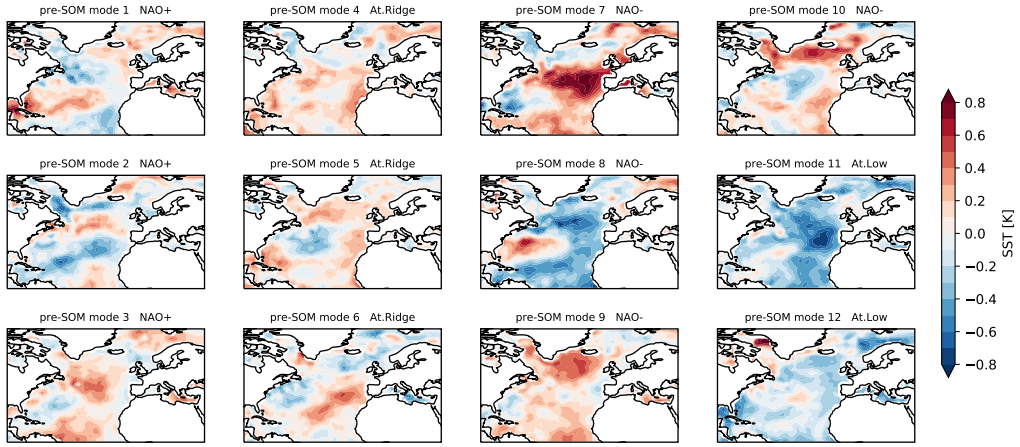


Figure A.4: Spring sea surface temperature (April SST) predictor patterns in the reanalysis w.r.t. the SOM master (Fig.A.2), for the period 1902-2008. Each pattern of SST precedes a summer SOM mode, as labelled.

overestimate NAO+ modes (SOM modes 1-2). We hypothesise that these limitations in the model representation of the observed summer variability pose a constraint on the credibility of summer predictions in the Euro-Atlantic region one season ahead based on this prediction system.

A.4.2 *Target regions for skill improvement and link to SST*

In this section, we evaluate the relationship between spring SST and summer SLP to target regions for potential skill improvement in the model. The set of spring SST predictors (Fig.A.4) show reasonable agreement with previous studies (e.g. Gastineau and Frankignoul, 2015). Preceding summers dominated by At. Ridge (hereafter: pre-At. Ridge), we find SST tripole patterns with warm SST anomalies in the tropical North Atlantic and western subpolar gyre off Newfoundland, and mostly cold SST in the subtropical western North Atlantic. At. Low summers follow major North Atlantic basinwide cooling, with a warming pattern near Greenland and off Newfoundland (hereafter: pre-At. Low). In contrast, negative NAO summers mostly follow horseshoe-like SST patterns with warm anomalies except over the subtropical gyre (hereafter: pre-NAO-). One case stands out, however, showing instead a SST predictor similar to the At. Low ones (c.f. Fig.A.4, pre-SOM mode 8). Lastly, we find no consistent mean SST pattern preceding positive NAO summers, but a set of three different SST tripole patterns with a common cooling over the North Sea and off the coast of northwest Africa (hereafter: pre-NAO+).

We evaluate the linear relationship between April North Atlantic SST predictors and JA SLP using the reanalysis (Fig.A.5). Areas of significant correlation indicate where the specific SST predictor might influence the summer circulation, and thus serve as reference to interpret any skill improvement in the model. We find that SST predictors have significant influence on JA SLP in At. Ridge (pre-SOM modes 4-6) and At. Low cases (pre-SOM modes 11-12), and in less extent to NAO+ (pre-SOM modes 1-2) and NAO- (pre-SOM modes 8-10). We find reasonable agreement with correlation results for the SST predictors and JA T2m or JA Z500 (not shown). From

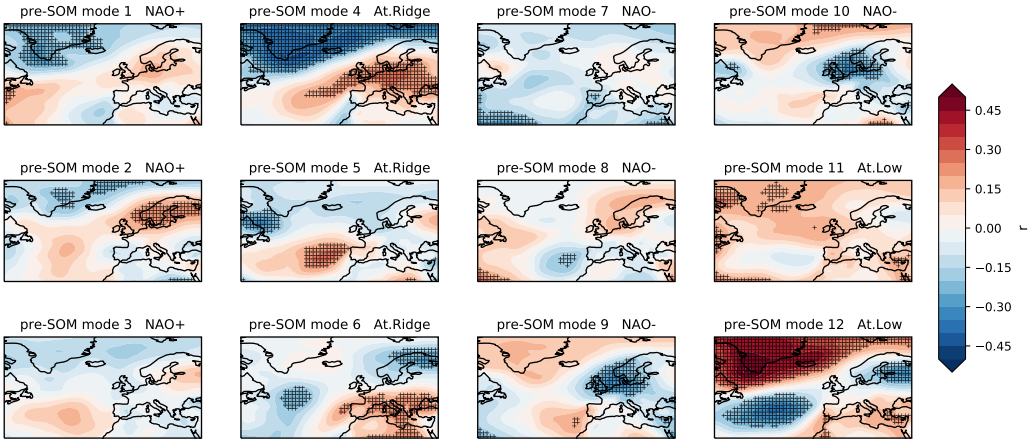


Figure A.5: Summer sea level pressure (JA SLP) point-wise correlated to timeseries associated with April sea surface temperature (SST) predictor patterns for the reanalysis over the period 1902-2008. Stippling represents correlation significant at the 95% level.

Fig.A.5 we might expect to find skill improvement only for the summer circulation over Greenland, Scandinavia, central Europe and a region over the ocean west of Iberia Peninsula using our SST predictors.

A.4.3 *Windows of opportunity for MPI-ESM-MR based on SST*

Firstly, we use the test ensemble to evaluate whether the SST-SLP relationship found in the reanalysis holds in the model. Next, we use the independent ensemble to test the robustness of this relationship and to assess the potential of skill improvement for other variables.

We evaluate the predictive skill at 3-4 months lead time for JA SLP in the test ensemble, for SST predictors of each of the four main summer atmospheric teleconnections separately to distinguish the contribution of SST (Fig.A.6). As a first step we analyse ACCs before subsampling in the ensemble space, considering the full ensemble (Fig.A.6a, d, g, j). We find distinct predictive skill for each group of SST predictor, with SST predictors for At. Ridge (i.e. pre-At. Ridge) showing highest skill off the Iberian coast, but no skill over land. The remaining groups show very limited skill overall.

We perform ensemble subsampling by selecting ensemble members according to the April SST classification in the pre-forecast data in predicting northern (At. Ridge or NAO+) or southern jet (At. Low or NAO-) groups. For example, Fig.A.6b shows the ACC calculated for the ensemble mean over members which predict either NAO+ or At. Ridge, for those years where April SST in the pre-forecast data classifies as pre-NAO+. A more strict selection allowing only one dominant atmospheric teleconnection per summer leads to a weak improvement. We include difference plots to illustrate the effect of performing ensemble subsampling on the predictive skill, in comparison to the traditional analysis using the full ensemble mean. Physically, a positive difference in correlation (selected minus full ensemble) thus represents the model predictive skill that could be achieved by this prediction system, if corrected

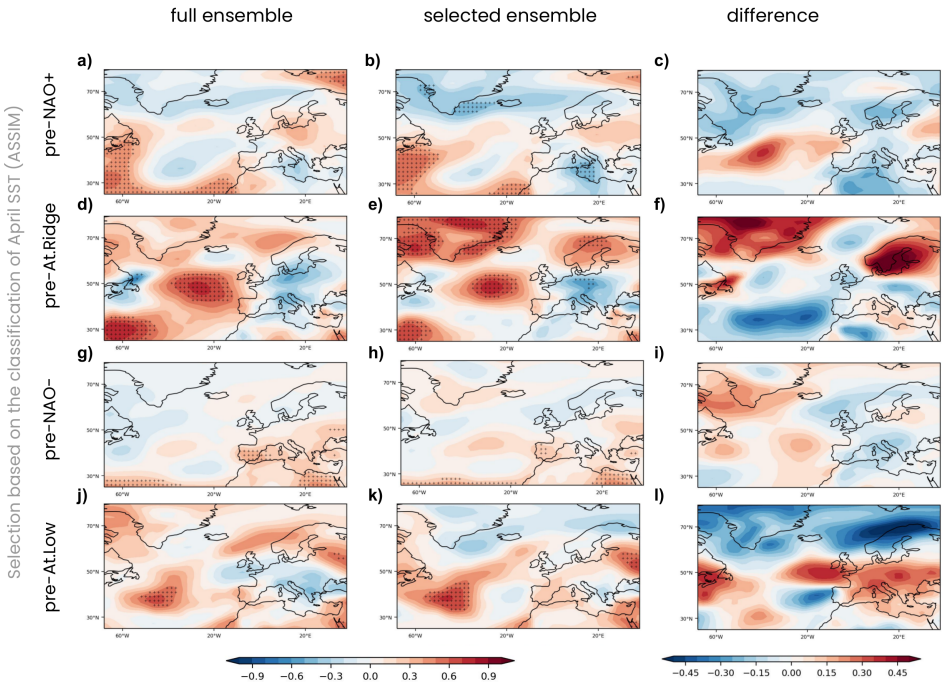


Figure A.6: Evaluation of SST as predictor in the test ensemble. Anomaly correlation coefficients (ACC) for sea level pressure (SLP), comparing the test ensemble to the reanalysis in 1902-2008. On the first row (uppermost) ACCs are calculated for years where April SST months in the pre-forecast data are classified as preceding NAO+, i.e. pre-NAO+. Similarly, the remaining three rows show ACCs for the case of pre-Atl. Ridge, pre-NAO- and pre-At. Low, respectively. Column-wise, ACCs are presented as follows: a), d), g), j) the ensemble mean is taken over the full ensemble; b), e), h), k) the ensemble mean is taken over the selected ensemble; c), f), i), j) differences in ACC between full and selected ensemble mean. The criteria for the selected ensemble is explained in Stippling represents correlation significant at the 95% level.

using our method. In contrast, a negative difference in correlation suggests that the SST predictor is not sufficient to perform a first guess, or that the model is unable to accurately simulate the atmospheric teleconnection under the circumstances given.

Comparing selected and full ensemble in Fig.A.6, we find significant regional improvement in the predictive skill of JA SLP for SST predictors of At. Ridge and Low. Improvement is highest for At. Ridge over Greenland and Scandinavia, reaching ACCs above 0.6. These regions of improvement in skill agree with the expectation based on Fig.A.5, SOM modes 4-6. This finding further agrees with Fig.A.2b, c, which shows that At. Ridge is the atmospheric teleconnection best represented by MPI-ESM-MR, thus having the potential to benefit the most from a physics informed subsampling. For NAO+ and NAO-, improvement is limited to the region over the ocean off the Iberian Peninsula, insignificant at the 95% level. We speculate that this is partly due to the less accurate representation of some NAO SOM modes by MPI-ESM-MR (see Fig.A.2b, c), in addition to a weaker SST-SLP relationship than for At. Ridge and Low modes (see Fig.A.5).

To contribute to the interpretation of predictive skill improvement, we test the benefit of refining the model ensemble with a "perfect" selection using the test ensemble. The "perfect" ensemble selection assumes that the dominant summer atmospheric teleconnection in the Euro-Atlantic domain is known each year in advance, thereby estimating the potential skill if such atmospheric teleconnection would be perfectly predicted by the model. We stress that, as opposed to the "selected" ensemble, such an analysis is only possible in hindcast mode, thus not being reproducible in real forecast mode. We analyse northern and southern jet groups separately to distinguish the effect of subsampling for SLP. In Fig.A.7, we select only ensemble members which agree with the atmospheric teleconnection label predicted by the reanalysis in a given year to calculate the ensemble mean. We find major skill improvement in the Euro-Atlantic sector for SLP predictions at 3-4 months lead time, with about half of the ensemble members selected to re-calculate the ensemble mean (Fig.A.7g). The main area of improvement fairly agrees with the position of the jet: over Scandinavia for the northern group and over south-western Europe for the southern group. This suggests that the area of skill improvement depends on on the skill of the model in simulating the relationship between predictor and target.

A.4.4 *Test with the independent ensemble*

Next, we use the independent ensemble to test whether these findings hold for an independent hindcast dataset covering the period of 1980-2016 (Fig.A.8). In addition to SLP, we test how the selection of ensemble members based on the SST-SLP relationship impacts the predictive skill of T2m and Z500. We analyse two groups and calculate ACCs for SST predictors for northern jet (pre-At. Ridge or NAO+) and southern jet (pre-At. Low or NAO-), similarly to Fig.A.7. Analysing the full ensemble (first column), we find for both SLP and Z500 that the northern jet group (Fig.A.8a, g) shows higher skill for northern Europe in comparison to the southern jet group (Fig.A.8j, p). Still, improvement for SLP and Z500 skill (Fig.A.8c, i, l, r) takes place in similar areas as for the independent ensemble (Fig.A.6 and Fig.A.11), albeit less pronounced. For T2m in the selected ensemble (Fig.A.8d-f), we find skill improvement for north-western Europe, with ACCs over Scandinavia reaching significant

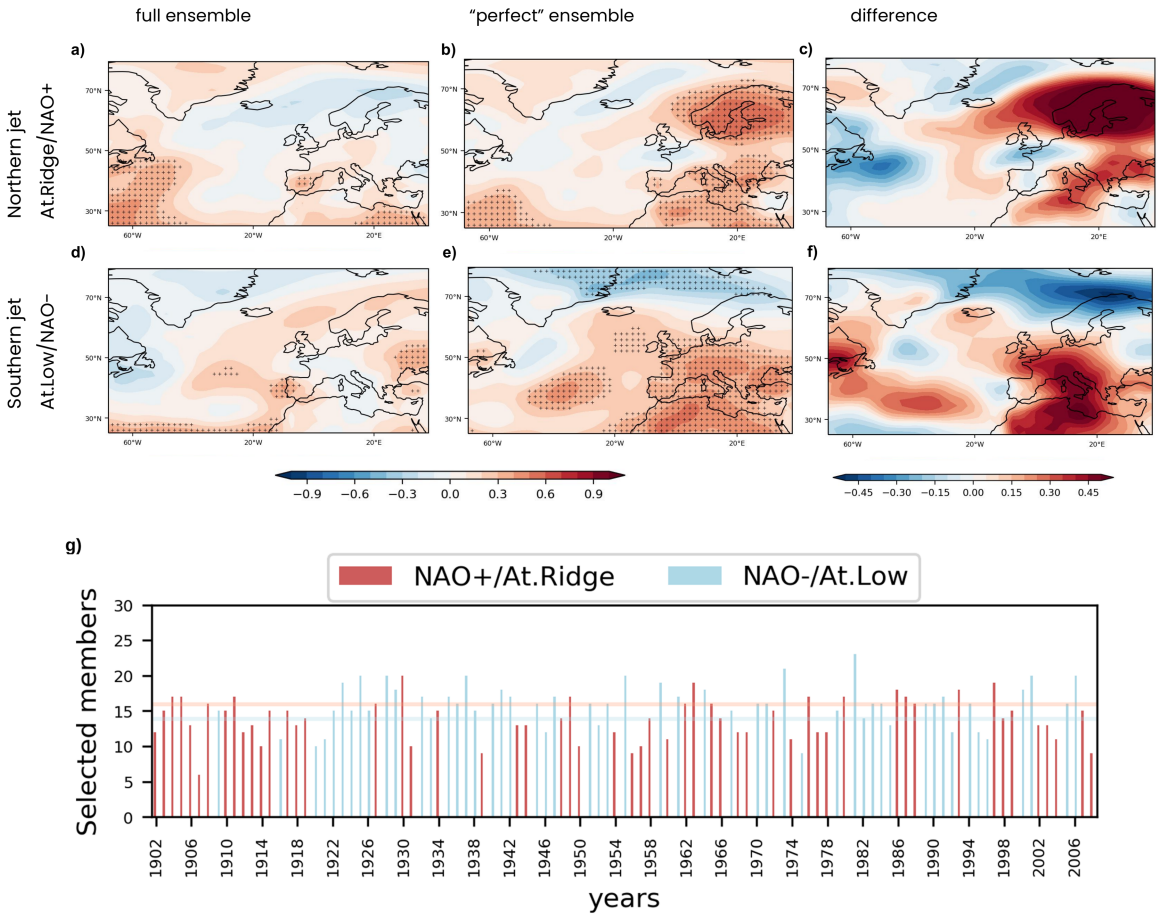


Figure A.7: "Perfect" model approach. Anomaly correlation coefficients (ACC) for sea level pressure (SLP), comparing the test ensemble to the reanalysis in 1902-2008. On the upper row ACCs are calculated for years where the observed dominant summer atmospheric teleconnection is classified as either At.Ridge or NAO+. Conversely, in the lower row ACCs are calculated for At. Low or NAO- cases. ACCs are presented as follows: a) and d) the ensemble mean is taken over the full ensemble; b) and e) only members whose classification matches the observed are selected to calculate the ensemble mean; c) and f) differences in ACC between full and selected ensemble mean. Stippling represents correlation significant at the 95% level. g) Bar plots show the fraction of ensemble members selected per year: red bars for observed At.Ridge or NAO+, blue bars for At. Low or NAO- cases. Horizontal lines show the mean fraction for each case.

values above 0.5 (at 95% level) for spring SST indicating northern jet group (pre-At. Ridge/NAO+). We find no significant T2m skill improvement for the southern jet group (Fig.A.8m-o). Despite not reaching statistical significance, a correspondence in the spatial pattern improvement for SLP, Z500 and T2m (Fig.A.8c, f, i) alludes that the ensemble subsampling based on the spring SST - summer SLP relationship influences the predictive skill of Z500 and T2m. This suggests the effect of the large-scale atmospheric circulation as a dynamical driver of temperature variability at the seasonal timescale, and illustrates the importance of improving the simulation of summer atmospheric teleconnections within the ensemble as a step to achieve skillful predictions of European climate.

A.5 DISCUSSION

Dynamical seasonal forecasts of European summer climate have until recently mostly shown negligible skill (e.g. Mishra et al., 2019), presumably a consequence of model errors in representing important drivers of summer climate variability (e.g. Beverley et al., 2019; Ossó et al., 2020). While recent work has shown evidence for skillful European summer rainfall predictions (Dunstone et al., 2018), model skill seemed to stem primarily from thermodynamical drivers. Identifying processes and predictors that provide forecast opportunities for skillful long-lead predictions is hence of paramount importance.

In this paper we use SOM to enable the identification of dynamical predictors for a pre-defined set of SOM modes. These SOM modes represent twelve stages of the four main atmospheric teleconnections driving the large-scale summer variability in the Euro-Atlantic region (Cassou et al., 2005). While defining these discrete samples is a clear limitation of our approach, using SOM allows us to distinguish spring SST patterns that can be used to indicate conditions of potentially high summer predictive skill at prediction start. Conversely, we were unable to find skillful SST predictors for composites based on a k-means clustering (k=4), possibly due to the blending of patterns and loss of information (not shown).

We show in this study that April North Atlantic SST patterns can be used to select potentially skillful ensemble members for MPI-ESM-MR summer hindcasts over specific target regions. We find that these regions depend on the dominant atmospheric teleconnection in a given summer, and on the strength of its relationship with spring SST. Based on these two criteria, we find that spring SST precursors for At. Ridge summers correspond to conditions under which the MPI-ESM-MR seasonal predictions may be expected to have particularly high skill over target regions, thus representing forecasts of opportunity (Mariotti et al., 2020).

We identify spring SST precursor patterns for At. Ridge summers which show reasonable agreement with results reported by Ossó et al., 2018 using Maximum Covariance Analysis. The authors described a pattern consisting of a spring SST dipole between subpolar and subtropical North Atlantic that persists into summer, forced by anomalous winter atmospheric circulation. Ossó et al., 2020 showed that this SST pattern drives a poleward displacement of the jet stream through changes in the background baroclinicity. Ossó et al., 2018 suggested that the summer atmospheric response to this oceanic forcing is imprinted at the surface as an anticyclonic anomaly that resembles the At. Ridge.

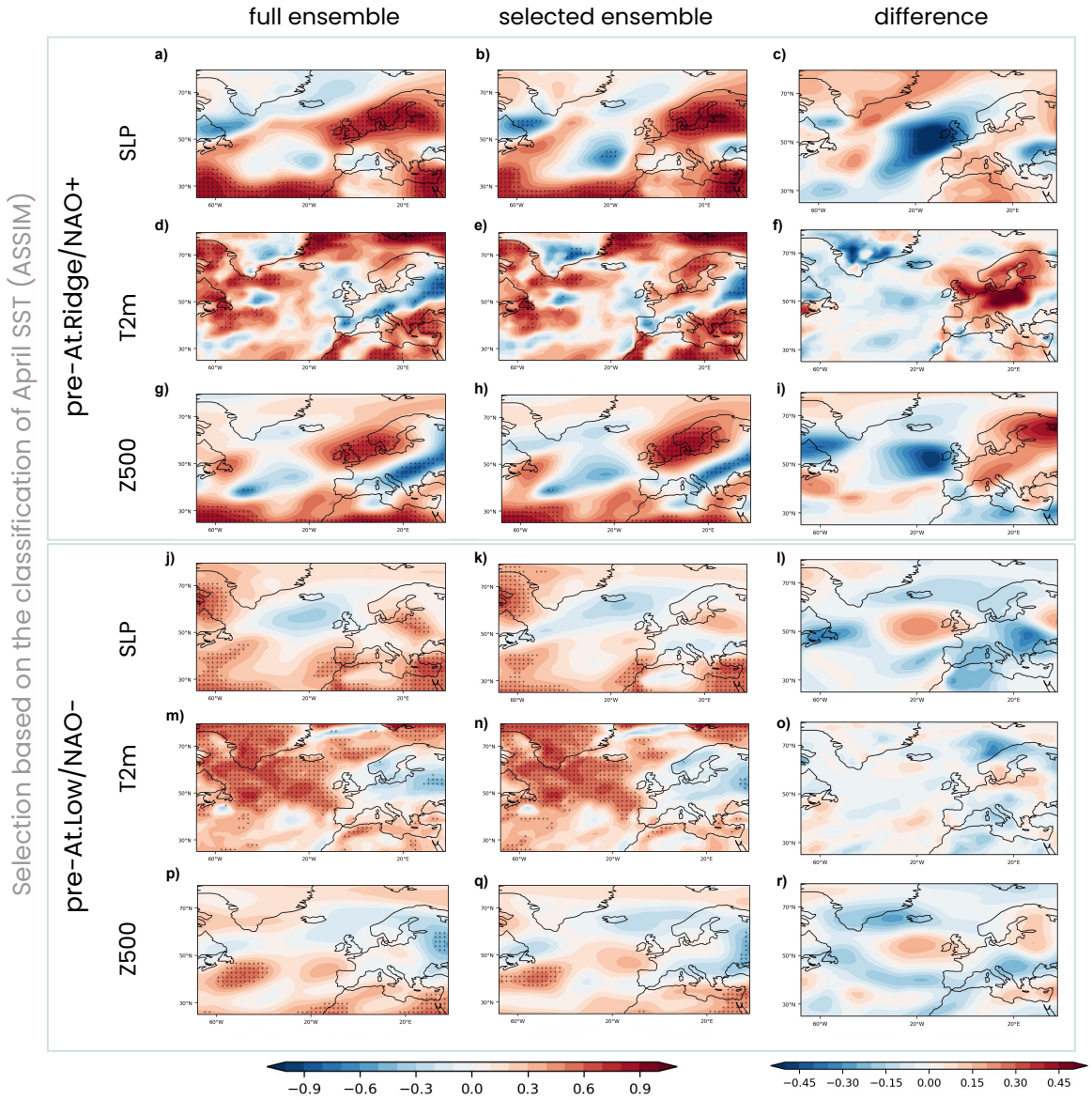


Figure A.8: Evaluation of SST as predictor in the independent ensemble. Anomaly correlation coefficients (ACC) for sea level pressure (SLP), temperature at 2 m height (T2m) and geopotential height at 500 hPa (Z500), as labelled, comparing the independent ensemble to Era-Interim in 1980-2016. On the upper three rows ACCs are calculated for years where April SST months in the pre-forecast data are classified as preceding either At.Ridge or NAO+, i.e. pre-At.Ridge/NAO+. Conversely, the lower three rows show ACCs for the case of pre-Atl.Low/NAO-. Column-wise, ACCs are presented as follows: a), d), g), j), m), p) the ensemble mean is taken over the full ensemble; b), e), h), k), n), q) the ensemble mean is taken over the selected ensemble; c), f), i), l), o), r) differences in ACC between full and selected ensemble mean. Stippling represents correlation significant at the 95% level.

Identifying SST initial conditions that are linked to summers dominated by At. Ridge 3-4 months ahead is an important step towards skillful European seasonal summer predictions. At. Ridge has been associated with northerly wind anomalies over western Europe (Craig and Allan, 2021), leading to a widespread below-average surface temperature distribution (Cassou et al., 2005; Comas-Bru and Hernández, 2018). We find a similar imprint on the surface climate for SOM modes associated with At. Ridge (Fig.A.10), illustrated by composites of summer air temperature anomalies. In combination with cyclonic conditions over the Mediterranean region, At. Ridge has been additionally associated with easterly wind anomalies, thereby influencing the occurrence of dry spells and drought conditions over western Europe (Haarsma et al., 2009; Rousi et al., 2021).

In addition to At. Ridge, our analysis of the relationship between spring SST and summer SLP (Fig.A.5) suggests that spring North Atlantic SST precursors for At. Low could predict potentially skillful ensemble members for MPI-ESM-MR summer hindcasts (e.g. Fig.A.5, pre-SOM mode 12). While we find improvement in skill over large areas of central Europe and off the coast of England, the improvement does not reach significance (95%) over most areas, except over southeastern Europe. We speculate that this is due to the more limiting representation of this atmospheric teleconnection in MPI-ESM-MR (Fig.A.3a, b), which might imply a model limitation for early warning of warmer than average summers. This finding agrees with Neddermann, 2019, which show that a teleconnection pattern similar to At. Low in MPI-ESM-MR shows a different structure of the centres of action in comparison to the reanalysis.

In contrast to At. Ridge and Low, only a few studies have suggested an active role of spring North Atlantic SST as a driver for summer NAO at seasonal to interannual timescales (e.g. Gastineau and Frankignoul, 2015; O'Reilly et al., 2017). Osborne et al., 2020 investigated the effect of North Atlantic SST on atmospheric circulation responses over the Euro-Atlantic region, speculating that summer NAO-related SST anomalies might feedback onto the At. Ridge and Low teleconnections rather than the summer NAO. Our analysis using North Atlantic SST as precursor for NAO show only marginal predictive skill improvement, for both NAO in positive and negative phases. This suggests that April North Atlantic SST has a limited impact on the seasonal predictability of summer NAO, potentially explained by this SST-NAO inconsistency.

Though we only investigate the role of spring North Atlantic SST, other predictors have been suggested to influence predictability of the summer atmospheric teleconnections analysed here. In particular, Hall et al., 2017 suggested that summer NAO is dependent on the positioning of the polar front jet, and thus on Arctic sea ice. Another potential predictor for summer NAO is stratospheric temperature, which in winter establishes a downward connection from the stratosphere to the surface, leading to enhanced surface predictability (e.g. Ayarzagüena and Serrano, 2009). For winter NAO, a combination of these precursors in autumn with ensemble subsampling allowed unprecedented skillful prediction of the NAO index in MPI-ESM-MR (Dobrynin et al., 2018). Including such predictors to inform the ensemble subsampling would be an interesting focus for future work.

A.6 CONCLUSIONS

We combine SOM and seasonal climate predictions with MPI-ESM-MR to investigate the seasonal predictability of European summer climate associated with the North Atlantic jet stream. We show that April North Atlantic SST patterns can be used to select potentially skillful ensemble members for MPI-ESM-MR summer hindcasts over specific target regions. Our main findings are:

- Our SOM analysis shows that among the four main summer atmospheric teleconnections MPI-ESM-MR best represents Atlantic Ridge, showing the highest agreement with the reanalysis both spatially and in frequency of occurrence. Conversely, spatial representation of Atlantic Low in MPI-ESM-MR agrees the least with the reanalysis, and the model underestimates frequency of occurrence.

- The use of SOM composites of North Atlantic SST patterns as spring SST predictors is relevant for Atlantic Ridge and Atlantic Low teleconnections, and limited for NAO. Greenland and Scandinavia are the areas over land with most potential spring SST influence on northern jet positions.

- Using the test ensemble (1902-2008), we find significant skill improvement for summer SLP at 3-4 month lead time over Greenland and Scandinavia for predictions initialised with SST predictors for Atlantic Ridge, for a selected ensemble predicting northern jet atmospheric teleconnections (Atlantic Ridge or NAO+).

- Using the independent ensemble (1980-2016), we find significant skill for summer SLP at 3-4 month lead time over northern Europe for predictions initialised with SST predictors for northern jet atmospheric teleconnections (Atlantic Ridge or NAO+). For a selected ensemble predicting northern jet teleconnections we find significant skill improvement over Scandinavia for both Z500 and T2m.

- A spatial correspondence in the improvement of Z500 and T2m for the northern jet group suggests the effect of large-scale atmospheric circulation as a dynamical driver of European temperature variability at the seasonal timescale. This finding highlights the importance of accurately representing summer atmospheric teleconnections in dynamical seasonal prediction systems as a necessary condition towards skillful predictions of the European climate.

Our findings offer an interesting avenue for the use of SOM in further research on windows of opportunity for seasonal climate predictions. Since our analysis only relies on SST information prior to the initialisation of the prediction system, our methodology can be extended and further applied to operational ensemble prediction systems.

DATA AVAILABILITY STATEMENT The data analysed in this study is subject to the following licenses/restrictions: All data are stored at the DKRZ in archive and can be made accessible upon request (https://www.dkrz.de/up). Requests to access these datasets should be directed to German Climate Computing Center (DKRZ) https://www.dkrz.de/up.

CONFLICT OF INTEREST STATEMENT The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS JC-O wrote the article and performed the analysis. JC-O, LFB, JB and EZ conceived the study and contributed to the writing and interpretation of the results. All authors contributed to the article and approved the submitted version.

FUNDING This research is a contribution to the Excellence Cluster CliCCS - Climate, Climatic Change, and Society at the University of Hamburg, funded by the DFG through Germany's Excellence Strategy EXC 2037 Project 390683824 (JB). This research has been additionally supported by the Helmholtz-Inkubator projector Reduced Complexity Models Redmod (JC-O, EZ); the European Commission, Horizon 2020 (grant no. EUCP (776613)); and the ANR Tremplin–ERC project HARMONY,grant no. ANR-20-ERC9-0001 (LB).

ACKNOWLEDGMENTS The authors would like to thank the Climate Modelling group at Universität Hamburg for helpful discussions. LB was supported by the European Commission, Horizon 2020 (grant no. EUCP (776613)); the ANR Tremplin–ERC project HARMONY (grant no. ANR-20-ERC9-0001). JC-O and EZ were supported by the Helmholtz Association. JB was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2037 'CLICCS - Climate, Climatic Change, and Society' – Project Number: 390683824, contribution to the Center for Earth System Research and Sustainability (CEN) of Universität Hamburg. The model simulations were performed using the high-performance computer at the German Climate Computing Center (DKRZ).

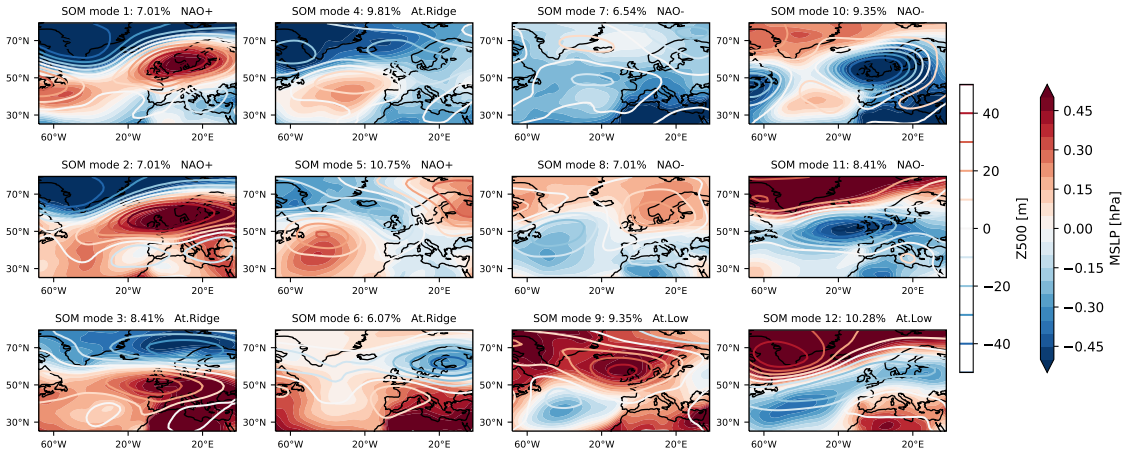


Figure A.9: Original SOM map representing the dominant summer atmospheric teleconnections in the North Atlantic European sector during 1902 - 2008, trained with ERA-20C July and August sea level pressure (SLP).

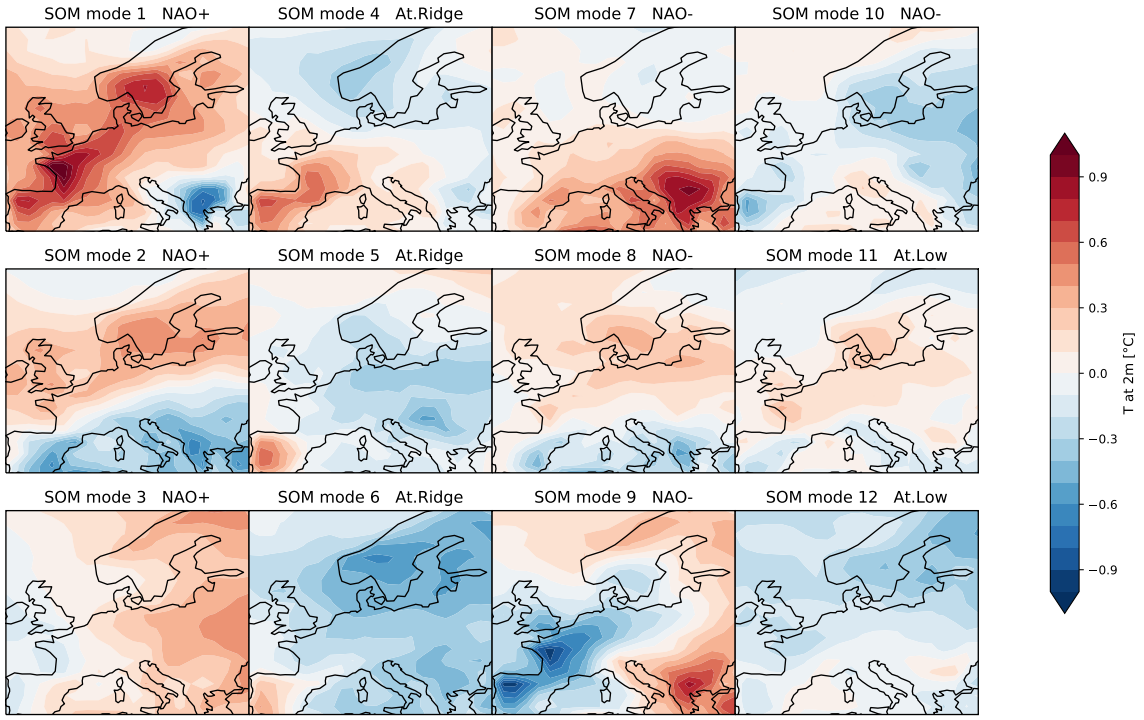


Figure A.10: Composite of air temperature at 2 metre in the reanalysis w.r.t. the SOM master (Fig.2), for the period 1902-2008.

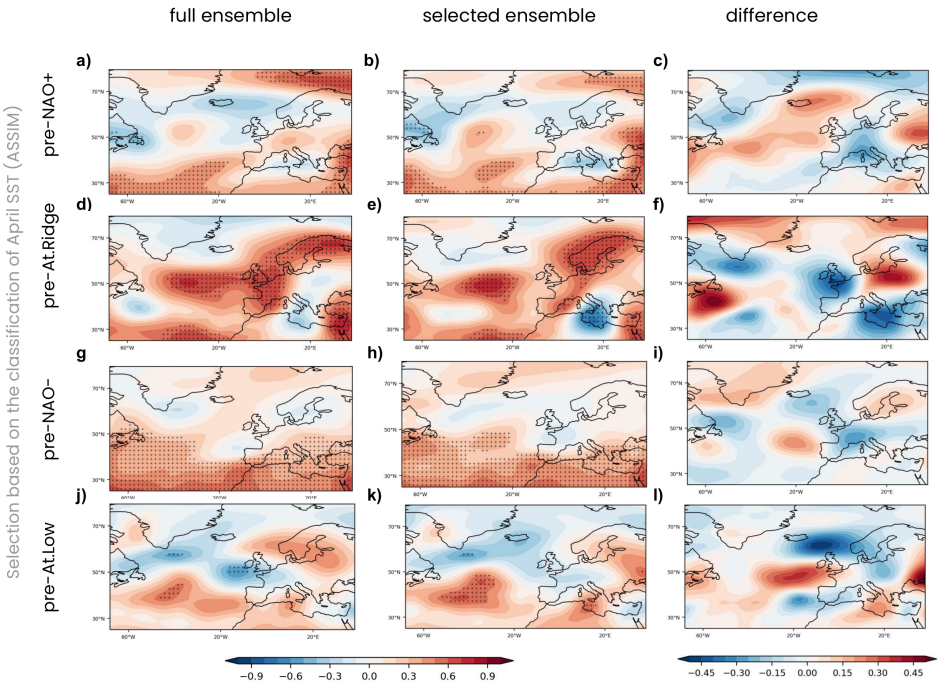


Figure A.11: Evaluation of SST as predictor in the test ensemble. Anomaly correlation coefficients (ACC) for geopotential height at 500 hPa (Z500), comparing the test ensemble to the reanalysis in 1902-2008. On the first row (uppermost) ACCs are calculated for years where April SST months in the pre-forecast data are classified as preceding NAO+, i.e. pre-NAO+. Similarly, the remaining three rows show ACCs for the case of pre-Atl. Ridge, pre-NAO- and pre-At. Low, respectively. Column-wise, ACCs are presented as follows: a), d), g), j) the ensemble mean is taken over the full ensemble; b), e), h), k) the ensemble mean is taken over the selected ensemble; c), f), i), j) differences in ACC between full and selected ensemble mean. The criteria for the selected ensemble is explained in Stippling represents correlation significant at the 95% level.

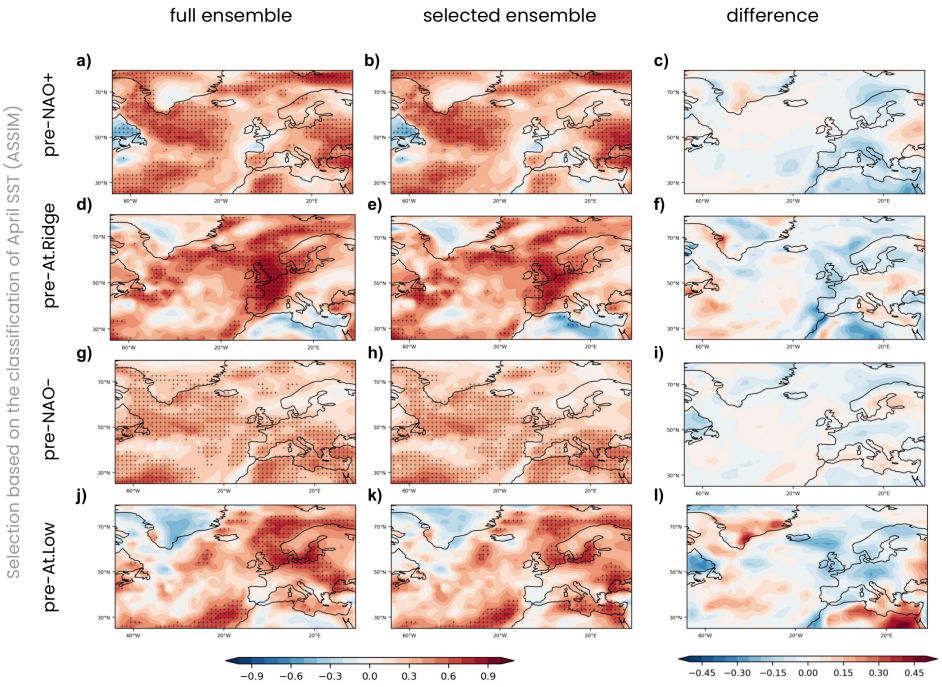


Figure A.12: As Fig.S3, for air temperature at 2 metre.

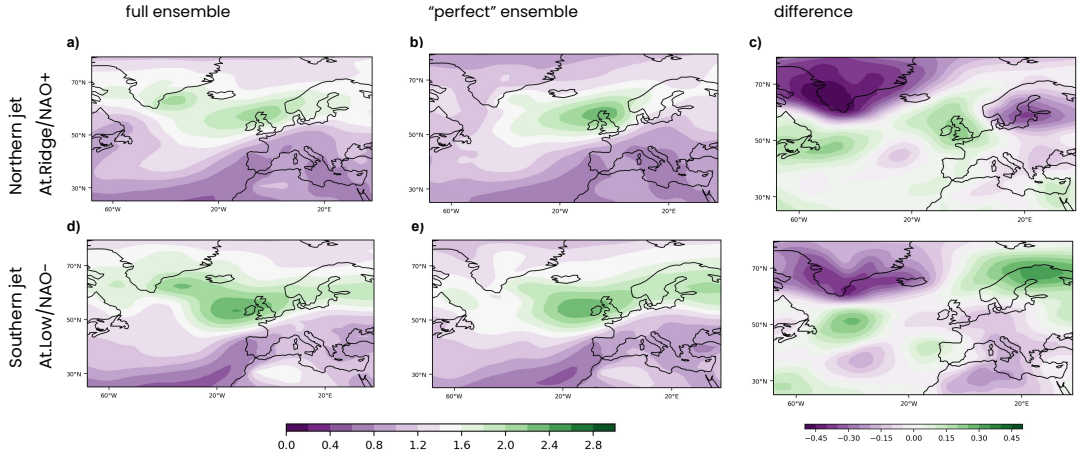


Figure A.13: "Perfect" model approach. Root mean square error (RMSE) for sea level pressure (SLP), comparing the test ensemble to the reanalysis in 1902-2008. On the upper row RMSE is calculated for years where the observed dominant summer atmospheric teleconnection is classified as either At.Ridge or NAO+. Conversely, in the lower row RMSE is calculated for At. Low or NAO- cases. RMSEs are presented as follows: a) and d) the ensemble mean is taken over the full ensemble; b) and e) only members whose classification matches the observed are selected to calculate the ensemble mean; c) and f) differences in RMSE between full and selected ensemble mean.

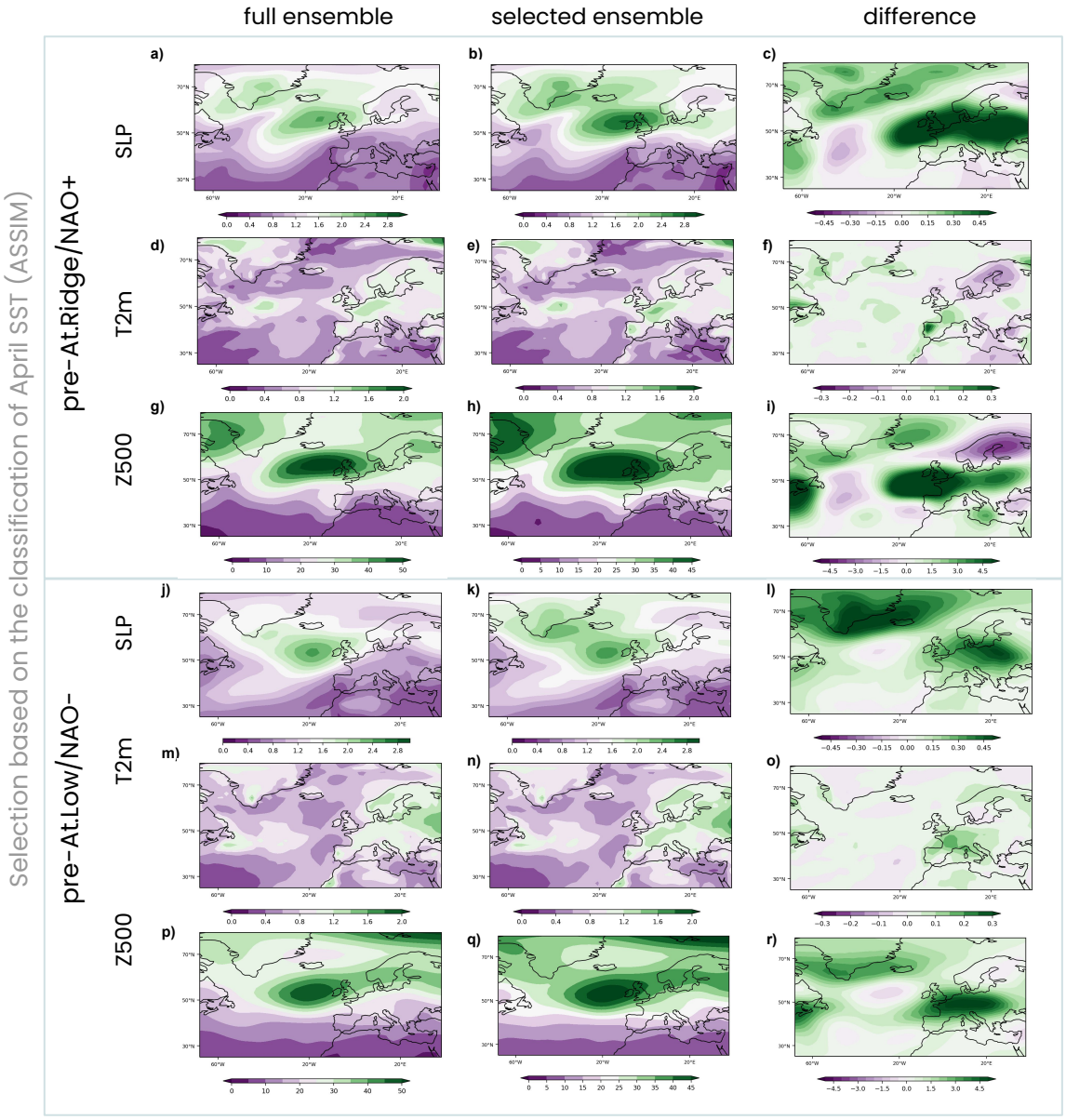


Figure A.14: Evaluation of SST as predictor in the independent ensemble. Root mean square error (RMSE) for sea level pressure (SLP), temperature at 2 m height (T2m) and geopotential height at 500 hPa (Z500), as labelled, comparing the independent ensemble to Era-Interim in 1980-2016. On the upper three rows RMSE is calculated for years where April SST months in the pre-forecast data are classified as preceding either At.Ridge or NAO+, i.e. pre-At.Ridge/NAO+. Conversely, the lower three rows show RMSE for the case of pre-Atl.Low/NAO-. Column-wise, RMSEs are presented as follows: a), d), g), j), m), p) the ensemble mean is taken over the full ensemble; b), e), h), k), n), q) the ensemble mean is taken over the selected ensemble; c), f), i), l), o), r) differences in RMSE between full and selected ensemble mean.

B

CAUSAL ASSOCIATIONS AND PREDICTABILITY OF THE EAST ATLANTIC PATTERN

The work in this appendix will be submitted as:
Carvalho-Oliveira, J., Di Capua, G., Borchert, L.F., Rousi, E., Donner, R., Zorita, E., Baehr, J. (*to be submitted*). "Causal associations and predictability of the East Atlantic Pattern."

JCO designed the research and performed the analysis. All authors contributed through discussions on the interpretation of the results. JCO prepared the manuscript with contributions from all co-authors. This work might be subject of further changes before it is submitted.

CAUSAL ASSOCIATIONS AND PREDICTABILITY OF THE EAST ATLANTIC PATTERN

Julianna Carvalho-Oliveira [1,2,3], Giorgia di Capua[4], Leonard F. Borchert [5], Efi Rousi[4], Reik Donner[4,6], Eduardo Zorita [2], Johanna Baehr[1]

[1] Institute for Oceanography, Centre for Earth System Research and Sustainability, Universität Hamburg, Hamburg, Germany

[2] Helmholtz-Zentrum Hereon, Institute of Coastal Systems - Analysis and Modeling, Geesthacht, Germany

[3] International Max Planck Research School on Earth System Modelling, Max Planck Institute for Meteorology, Hamburg, Germany

[4] Potsdam Institute for Climate Impact Research, Potsdam, Germany

[5] Climate Statistics and Climate Extremes, Centre for Earth System Research and Sustainability, Universität Hamburg, Hamburg, Germany

[6] Department of Water, Environment, Construction and Safety, Magdeburg-Stendal University of Applied Sciences, Magdeburg, Germany

B.1 ABSTRACT

We apply the causal inference-based tool Causal Effect Networks to evaluate the influence of spring North Atlantic extratropical surface temperatures (SST) on the summer East Atlantic Pattern (EA) seasonal predictability during the 20th century. First in the ERA-20C reanalysis, we find that a meridional SST gradient in spring (SST index) causally influences the summer EA, with an estimated causal effect expressed by a β-coefficient of about 0.2 (a 1 standard deviation change in spring SST index causes a 0.2 standard deviation change in the EA 3-4 months later). We only find this link to be causal, however, in a late period consisting of 51 years (1958 - 2008). When performing the analysis on 45-year-long timeseries randomly sampled in this late period, we find the strength of the causal link to depend on interannual variability, suggesting a possible modulation by an external physical mechanism. In addition to the summer EA, we find that spring SST has an estimated causal effect of about -0.2 on summer 2-metre air temperatures over northwestern Europe, possibly mediated by summer EA. Second, we test whether MPI-ESM-MR can reproduce the observed causal links. We use a pre-industrial and a historical simulation, and a 30-member initialised seasonal prediction ensemble with MPI-ESM-MR to test the model performance in reproducing the detected causal links in ERA-20C and to evaluate whether this performance might leave an imprint in the model predictive skill of European summer climate. While we find that MPI-ESM-MR is mostly unable to reproduce the causal link between spring SST and the summer EA among the different datasets, the 30-member initialised ensemble can moderately reproduce a causal link between spring SST and summer 2-metre air temperatures over a region west of the British Isles. We perform a predictive skill assessment conditioned on the spring

SST causal links for July-August sea level pressure, 500 hPa geopotential height and 2-metre air temperatures for predictions initialised in May. Our results suggest that MPI-ESM-MR's performance in reproducing the spring SST causal links constrains the seasonal prediction skill of European summer climate.

B.2 INTRODUCTION

The summer East Atlantic Pattern (EA) is amongst the important atmospheric teleconnections influencing weather and climate in the Euro-Atlantic region (e.g. Bastos et al., 2016; Comas-Bru and McDermott, 2014). Along with the summer North Atlantic Oscillation (NAO), these teleconnections are often used to describe the combined changes in latitude and speed of the North Atlantic jet stream (Woollings et al., 2010) – one of the major modulators of mid-latitude weather extremes (e.g. Rousi et al., 2022). Understanding the predictability associated with these teleconnections is therefore of paramount importance. Although several recent studies have focused on predictability of the NAO (Athanasiadis et al., 2020; Domeisen et al., 2018; Klavans et al., 2021; O'Reilly et al., 2019), the EA has received less attention. Here, we apply a causal inference-based tool to evaluate the influence of North Atlantic extratropical surface temperatures (SST) on the predictability of EA at seasonal timescales.

The most common description of the EA pattern features a well-defined sea level pressure (SLP) centre of action south of Iceland and west of the British Isles, usually defined as the second leading empirical orthogonal function (EOF) of SLP in the Euro-Atlantic region (e.g. Moore et al., 2013). Wallace and Gutzler, 1981 define a positive phase of the EA as characterised by a centre of action exhibiting anticyclonic conditions, featuring the northward extension of the Azores High. A positive EA has been associated with below-average surface temperatures (Cassou et al., 2005; Comas-Bru and Hernández, 2018) and dry spells in parts of Europe (Rousi et al., 2021). Conversely, anomalous cyclonic conditions offshore of Ireland have been suggested to influence heatwaves in Europe for a negative EA phase (e.g. Duchez et al., 2016). Comparing to a nonlinear approach (e.g. Carvalho-Oliveira et al., 2022; Cassou et al., 2004), a positive EA phase is reminiscent of Atlantic Ridge, whereas a negative EA phase resembles the Atlantic Low. A common feature amongst the different EA definitions is that its centre of action is positioned along the NAO nodal line, thus ultimately modulating the location and strength of the NAO dipole and the North Atlantic storm track (Woollings et al., 2010). That is, summer climate predictability in the Euro-Atlantic region is closely linked to EA variability.

While there is no consensus on the physical processes driving the EA, spring North Atlantic sea surface temperatures (SSTs) have been proposed to influence EA variability and predictability. Gastineau and Frankignoul, 2015 suggested that summer 500-geopotential height anomalies in the Euro-Atlantic significantly co-vary with a spring North Atlantic tripole pattern in observations over the 20th century. Moreover, Carvalho-Oliveira et al., 2022 suggested that spring North Atlantic SSTs can influence predictive skill of summers dominated by EA in initialised simulations. Based on linear regression analyses of the period 1979–2017, Ossó et al., 2018 and Ossó et al., 2020 proposed a physical mechanism whereby anomalous extratropical North Atlantic SSTs in spring may persist into summer and influence shifts in the eddy-driven jet stream, imprinting at the surface a SLP pattern that resembles the

EA. In particular, these studies suggest that this mechanism is forced by changes in baroclinicity of the lower troposphere associated with a strong meridional SST gradient in spring located between subpolar and subtropical North Atlantic. The authors hypothesised that the delayed atmospheric response in summer, and not in spring, could be explained by the seasonal evolution of both SST gradient and jet stream position, modulated by a positive coupled ocean–atmosphere feedback that operates primarily in summer.

Nevertheless, while the linear regression-based analysis provided in Ossó et al., 2018 suggest a contribution of spring SST on the summer SLP variability, this approach does not imply causation. Disentangling the complex causal-effect pathways underlying the mechanism proposed in Ossó et al., 2020 over a long observational record is a crucial step to evaluate EA predictability in dynamical climate models. Hence, in this paper we use Causal Effect Network (CEN, Kretschmer et al., 2016; Runge et al., 2015) to test the hypothesis of spring SST causally driving a response in the summer SLP and temperature fields in the Euro-Atlantic sector during the 20th century. CEN overcomes spurious correlations due to autocorrelation, indirect effects, or common drivers (Runge et al., 2019), and has been successfully used to complement hypothesis testing for other teleconnections (e.g. Di Capua et al., 2020a).

Although dynamical seasonal forecasts of European summer climate usually show very little skill (e.g. Mishra et al., 2019), recent studies suggest that improving the representation of teleconnections can increase forecast skill (Carvalho-Oliveira et al., 2022; Oliveira et al., 2020; Schuhen et al., 2022). The physical mechanism connecting SST variability and jet stream dynamics proposed in Ossó et al., 2020 thus offers a framework to more generally assess the influence of SST on seasonal predictability of the EA – the aim of the present study. Therefore, we use CEN to firstly investigate under which circumstances spring extratropical North Atlantic SSTs causally influence the summer EA and its associated impact on surface climate. Secondly, we analyse pre-industrial, historical and initialised simulations with the MPI-ESM-MR to test the model performance in reproducing the observed SST - EA link, and to identify how this performance might constrain the seasonal prediction skill of European summer climate.

B.3 METHODOLOGY

B.3.1 *Reanalysis and model data*

We investigate the SST - EA link first using ERA-20C reanalysis (Poli et al., 2016), and then using model simulations with the Max Planck Institute Earth System Model in its mixed-resolution setup (MPI-ESM-MR, Dobrynin et al., 2018). We use monthly means of sea level pressure, sea surface temperature (SST), and air temperature at 2 metre height (T2m). We focus our analysis on the 101-year long period spanning 1908-2008 in the model and observations.

In MPI-ESM-MR, the atmospheric component ECHAM6 (Stevens et al., 2013) has a resolution of T63L95, with a nominal horizontal resolution of 200 km (1.875°) and 95 vertical layers up to 0.01 hPa. The oceanic component MPI-OM (Jungclaus et al., 2013) is coupled to ECHAM6 and has a resolution of TP04L40, with an approximate

horizontal resolution of 40 km (0.4°) and 40 vertical layers. External forcing is taken from CMIP5 (Giorgetta et al., 2013).

We investigate how MPI-ESM-MR performs in reproducing the SST - EA link among three independent sets of MPI-ESM-MR simulations. The datasets comprise a pre-industrial control run (piControl), a historical run, and a 30-member seasonal initialised hindcast ensemble (MR-30). Comparing the performance of each set against reanalysis enables us to distinguish the role of forcing (from piControl to historical), and of assimilation (historical to initialised ensemble) on the model skill.

The pre-industrial coupled atmosphere/ocean control run piControl has a total length of 1000 years (period 1850-2849) (Giorgetta et al., 2011), with forcing constant in time: orbital parameters and greenhouse gases concentration are fixed at 1850 values; spectral solar irradiance remains constant as the solar cycle average over 1844-1856, and monthly ozone concentrations are fixed at the 11-year average over 1850-1860 (Mauritsen et al., 2012). The historical simulation runs from 1850 to 2005 under natural and anthropogenic forcing following CMIP5 (Dobrynin et al., 2018).

Lastly, the hindcast ensemble MR-30 is initialised on 1st of May every year from 1902-2008, with initial conditions taken from an assimilation experiment (e.g. Oliveira et al., 2020). In the assimilation experiment, Newtonian relaxation (*nudging*) is used in full-field mode towards all atmospheric and ocean levels except in the boundary layer. The atmosphere conditions of vorticity, divergence, three-dimensional temperature and two-dimensional pressure are assimilated with ERA-20C data. In the ocean, three-dimensional daily mean salinity and temperature anomalies are nudged at a relaxation time of approximately 10 days. The ocean state is derived in an ocean-only simulation performed with MPI-OM forced with the atmospheric variables from ERA-20C. The three-dimensional atmospheric and ocean fields of the assimilation experiment form the initial conditions, from which 30 ensemble members are generated by perturbing the atmospheric state with slightly disturbed diffusion coefficient in the uppermost layer.

B.3.2 *Pre-processing and climate indices*

We compute anomalies at every gridpoint by removing mean seasonal cycle and linear trend, satisfying data input requirements for the CEN algorithm. We analyse bimonthly means in March-April (MA) and April-May (AM) for spring SST and July-August (JA) SLP and T2m. In MR-30, we use the assimilation experiment to obtain spring SST fields, and the hindcast ensemble at lead times 3-4 months to obtain summer SLP, T2m and 500 hPa geopotential height (Z500). We apply area-weighting by multiplying each value with the cosine of its latitudinal location to take into account the dependence of the gridpoint density on latitude.

We test the influence of spring extratropical North Atlantic SSTs in the summer EA using the SST index proposed in Ossó et al., 2018. We calculate the SST index by subtracting the average SST over the eastern box (35°W-20°W, 35°-42°N) from the average SST over the western box (52°W-40°W, 42°-52°N), represented in green colours in Fig.B.2f. We analyse the SST index for both March-April and April-May means.

We calculate the EA index to analyse the summer EA teleconnection. As a first step, we define a reference EA index as the second principal component (PC) of the

leading empirical orthogonal function (EOF) of JA anomalies of sea level pressure over the Euro-Atlantic sector 70°W-40°E, 25°-80°N calculated from the ERA-20C reanalysis data (e.g. Comas-Bru and McDermott, 2014). Next, each EA index in the model simulations from MPI-ESM-MR are calculated by projecting each ensemble member onto the EA reference EOF. We consider a positive phase of the EA index when characterised by a centre of positive SLP anomalies that lies south of Iceland and west of the British Isles (e.g. Comas-Bru and McDermott, 2014; Wallace and Gutzler, 1981, Fig.B.1a).

To fully analyse the impact of spring SST on the summer SLP variability, we include the SLP index proposed in Ossó et al., 2018, in addition to the EA index. The SLP index is calculated as JA SLP anomalies averaged over the region 45°N-55°N; 25°W-5°W indicated by a blue box in FigB.2c. We check whether the strength and timing of the SST-SLP relationship is consistent among the two indices.

Finally, we evaluate whether spring SST influences temperatures over Europe with a $T2m_{CE}$ index, defined as the average summer air temperatures over the central European region 46°N-55°N; 11°E-34°E, represented by the red box in Fig.B.2i. All climate indices are standardised to have mean of zero and SD of 1 to allow for comparison.

Using the aforementioned climate indices, we perform linear regressions and correlations to analyse the linear relationship between the predictor spring SST and the target variables summer EA, SLP, and $T2m_{CE}$ indices. We use a two-tailed Student's t-test to calculate the statistical significance.

B.3.3 *Causal effect network*

We use Causal Effect Network (CEN, Kretschmer et al., 2016; Runge et al., 2015) to test whether spring SST causally influences the variability of summer SLP and temperature fields in the Euro-Atlantic sector during the 20th century. CEN is machine learning tool which implements the so-called Peter and Clark momentary conditional independence algorithm (PC-MCI, Runge et al., 2019) for causal discovery. CEN iteratively calculates partial correlations amongst a set of time-series to test if a link between a potential precursor and a target variable at a certain time lag is: *i)* considered spurious, and can be explained by the combination of other time-series at different lags (i.e. conditional independence); or *ii)* considered causal, and cannot be explained by the combined influence of other investigated variables (i.e. conditional dependence). We stress, however, that the term *causal* should be interpreted as causal relative to the set of investigated variables, under the assumptions considered in the CEN algorithm (e.g. stationarity of time-series). We refer the reader to Kretschmer et al., 2016 for a thorough description of the CEN algorithm.

We visualise the output of CEN as a process graph, where circles represent the investigated variables, and arrows indicate the strength and the direction of the causal links. The strength is expressed by the standardised regression coefficient, denoted β-coefficient, and defined as the expected change of Y_t in units of its standard deviation (SD) induced by raising $X_{t-\tau}$ by 1 SD, while keeping all other potential precursors constant. Moreover, CEN outputs the autocorrelation path coefficient, which represents the causal influence of a variable on itself, as opposed to the Pearson autocorrelation.

B.4.1 *Characteristics of the observed link: temporal and spatial variability*

The spatial pattern of the summer EA in its positive phase is characterised by large-scale cyclonic conditions, except at the anticyclonic centre of action located south of Iceland and west of the British Isles (Fig.B.1a). A typical surface climate imprint of the summer EA in positive phase shows below-average temperatures in continental Europe (Fig.B.1b) and below-average precipitation in the British Isles and northwestern Europe (Fig.B.1c). As a first approach to evaluate the influence of extratropical SSTs in the summer EA, we use the SST index defined in Ossó et al., 2018. A linear correlation analysis suggests that strength of the relationship between the SST index in spring and the EA index in summer is time dependent (Fig.B.1f). Considering a period of 101 years (1908-2008), this relationship is weak (r = 0.2, Fig.B.1e). However, considering indeed only the latest 51 years(1958-2008, Fig.B.1d), correlation reaches significant values (r = 0.5). The temporal variability of this relationship is well illustrated for correlations calculated using a 20-year running window, which shows a reverse in the sign of correlations starting from 1945, and highlights an increase in the strength beyond 1958 (Fig.B.1f). This analysis suggests that the spring SST - summer EA relationship is nonstationary. Hence, we distinguish the following three periods to scope the remaining analysis: i) *early period*: 1908 - 1957; ii) *late period*: 1958-2008, and iii) *full period*: 1908-2008.

We assess the spatial features of the SST index influence on the summer atmospheric circulation in the different periods to further explore the variability of the spring SST - summer EA relationship. We analyse average lag of 3 months, i.e. April-May (AM) SST and July-August (JA) SLP means (3-months lag). Correlation maps show distinct patterns in early and late periods. We find that significant correlations between the precursor SST index and the summer sea level pressure (SLP) takes place over a region in the North Atlantic which reasonably coincides with the location of the EA teleconnection centre of action during the late period (Fig.B.2b, c.f. Fig.B.1a). The location of this region seems to oscillate about 45° N, remaining south of this latitude in the early period (Fig.B.2a), while located northwards in the late one (Fig.B.2b). Surrounding this high correlation region, the sign of correlations is opposite between early and late periods. We find similar results using March-April (MA) SST means, only in weaker strength.

Regression maps further suggest that spring SST anomalies persist into summer and then influence atmospheric circulation (Fig.B.2d-f). Positive values of the AM SST index in spring are associated with summer warm anomalies east of Newfoundland and cool anomalies west of Iberia, leading to concomitant anticyclonic conditions in the ocean located south of Greenland. In the late period, these anticyclonic conditions coincide with the position of the EA centre of action.

Moreover, we test whether the SST index influences JA T2m via the EA. We find significant correlations between the AM SST index and JA T2m, showing a similar pattern as in Fig.B.1b corresponding to JA EA - T2m. We find that correlations between AM SST index and JA T2m show distinct patterns between early and late periods (Fig.B.2g,h). A positive phase of the SST index in spring precedes a positive phase of the summer EA (e.g. Fig.B.2e), which in turn can be associated with below-

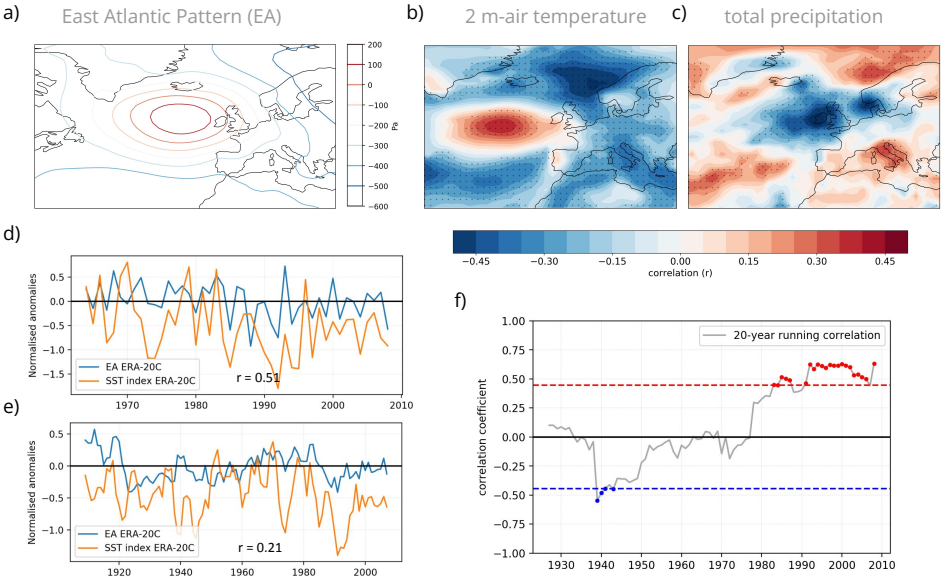


Figure B.1: Variability and linear relationships of EA in ERA-20C. a) Positive phase of the EA teleconnection, defined as the second EOF of July-August SLP. b) Pointwise correlation of EA index with concurrent July-August anomalies of 2-metre air temperatures in the full period. c) Same as b), for July-August anomalies of total precipitation. d) Time series of SST (orange) and EA (blue) indices for the late and e) the full periods. The full period is smoothed by a 3-year running mean. f) Running-correlation between SST and EA indices for a 20-year window. Coloured markers indicate significant correlations at the 95% confidence interval, illustrated by dashed lines.

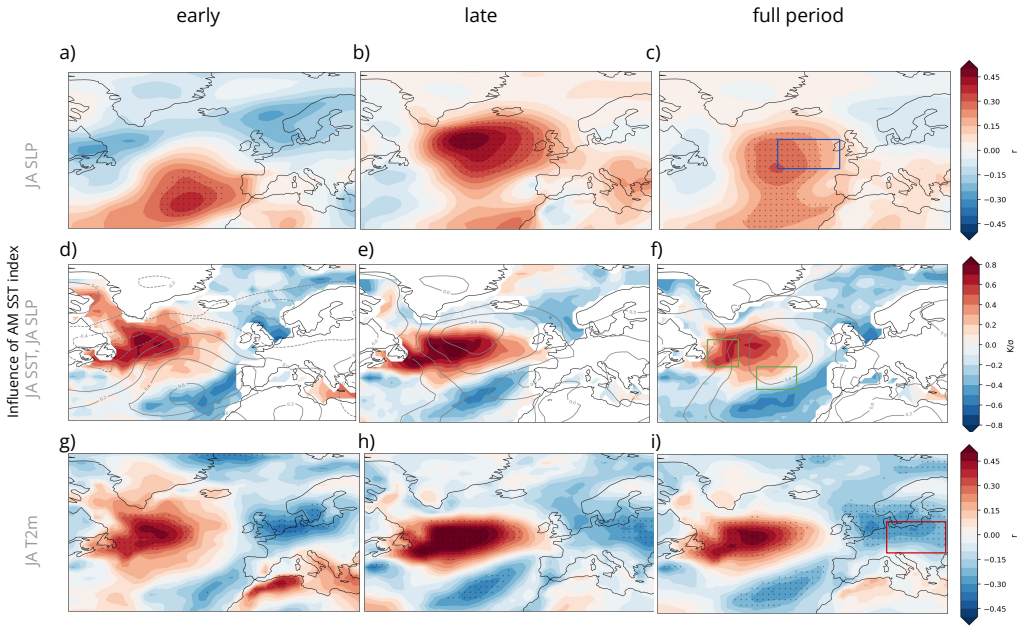


Figure B.2: Distinct spatial characteristics of the spring SST influence on the summer circulation over the 20th century (for ERA-20C). Top row (a-c) shows point-wise correlation coefficients for the April-May SST index and July-August SLP means considering *early* (1908-1957), *late* (1958-2008) and *full* periods (1908-2008), respectively. Middle row (d-f) shows linear regression maps of July-August SST anomalies (shading) and SLP (contours) against the precursor SST index (normalised). Contour interval is 0.2 $hPa\sigma^{-1}$. Bottom row (g-i) shows point-wise correlation coefficients for the April-May SST index and July-August air temperature at 2 metre height. Stippling indicates correlations significant at the 95% confidence level, calculated with a Student's t-test. Boxes illustrate the regions used to calculate the SLP and SST indices (c, f, respectively) as in Ossó et al., 2018, and the $T2m_{CE}$ index, as described in the text.

average temperatures, primarily over central Europe. In summary, this analysis reveals that spring extratropical oceanic forcing of the summer atmospheric circulation has a marked temporal and spatial variability over the 20th century, only projecting onto the EA pattern over the late period. This variability might pose a constraint on the predictive skill of European summer climate based on spring extratropical SST over certain periods of time.

B.4.2 *Investigating causality*

To further test the robustness of the SST-EA relationship in ERA-20C, we evaluate whether spring extratropical SSTs and summer EA are conditionally dependent. Specifically, we test the hypothesis that spring SST is a causal driver for the summer EA, thus excluding autocorrelation effects or common drivers which could lead to spurious links. We perform a causality analysis using the CEN algorithm (Runge et al., 2015).

First, we build one CEN for each of the three investigated periods in ERA-20C. Besides the EA and SST indices, we include two additional indices in the CEN. The first is the SLP index, defined in Ossó et al., 2018 and illustrated by the blue box in Fig.B.2c. We thus test whether differences between early and late periods (c.f. Sec.B.4.1) are reflected in distinct timing or strength among the EA and SLP indices with SST. The second index concerns summer air temperatures averaged over the region represented by the red box in Fig.B.2i ($T2m_{CE}$), which shows significant correlations with SST. We test whether spring SST causally drives changes in air temperature over central Europe and under which circumstances this holds true. Our CEN analysis focuses on 3 and 4 months lag only.

Over the late period, we confirm that the spring SST index is a causal driver for both the summer EA, and the summer SLP index, however at distinct time lags (Fig.B.3a). The strength of the causal link is expressed by the standardised regression coefficient, denoted β-coefficient in CEN. At 4-month lag, we find $\beta_{SST \to EA} \approx 0.22$, which means that a change of 1 standard deviation (SD) in the March-April SST index leads to a change of 0.22 SD in July-August EA, when conditioned on $T2m_{CE}$ and SLP indices. We find a causal link of similar strength at 3-month lag $\beta_{SST \to SLP} \approx 0.21$ between April-May SST and July-August EA, as well as $\beta_{SST \to T2m_{CE}} \approx -0.2$ between April-May SST and July-August $T2m_{CE}$. Although we speculate that the link $SST \to T2m_{CE}$ is mediated via the summer EA, we are unable to confirm this mediation with a CEN analysis focusing on 3-4 months lag. We find no significant causal links when analysing the full or early period.

Next, we test the sensitivity of the detected causal links between spring SST and summer SLP to slight differences in the analysed years. By removing 6 randomly selected years (12% of tested years in the late period) in each new CEN over 500 iterations , we test whether the causal links are particularly subjected to interannual variability (Fig.B.3c-f). We find high variability in the strength of the links $\beta_{SST \to EA}$ and $\beta_{SST \to SLP}$ (Fig.B.3b, e), ranging from zero (i.e. no causal link) to 0.5, with median values corresponding to β-coefficients calculated in Fig.B.3a. This sensitivity in the causal link strength due to sampling suggests that the spring SST - summer SLP relationship might be modulated by an external physical mechanism, i.e. an additional actor excluded from this CEN.

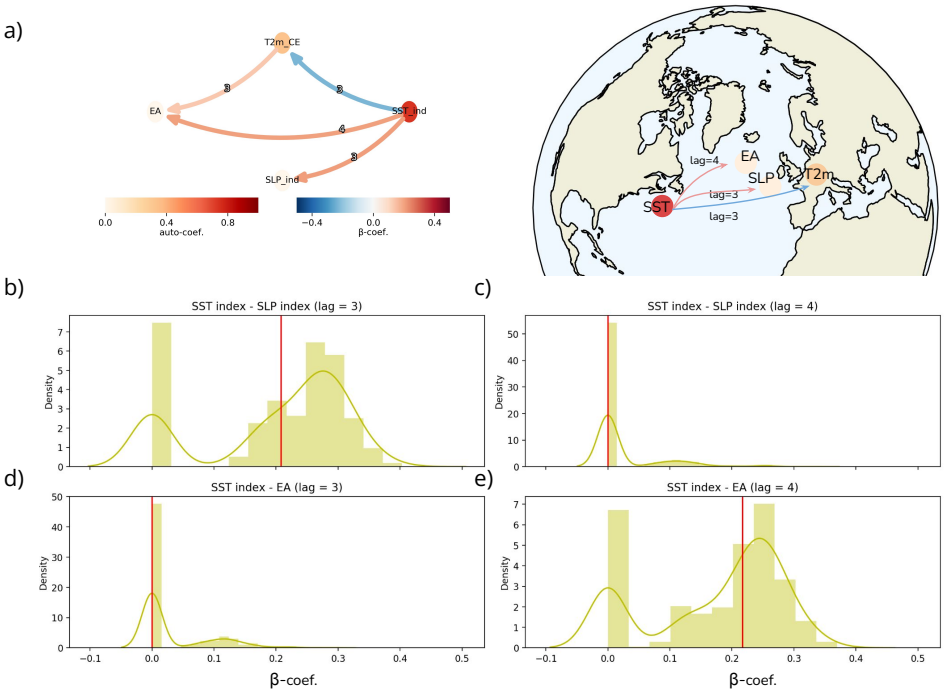


Figure B.3: Causal effect network analysis for the late period (1958-2008) in ERA-20C. a) causal graph between SST index, EA teleconnection, SLP index and $T2m_{CE}$ (*left*), and schematic illustrating the causal pathways in the Euro-Atlantic region. The strength and direction of the causal links is given by the β-coefficient and is represented by the arrows, whereas the auto-correlation path coefficient is represented for each variable by the respective circle colour. The numbers over each arrow represent the time tag (in months) when the strongest causal link between each variable pair is detected. b) Region used to calculate the SLP index (black box). c-f) Sensitivity of the causal links shown as the PDF of β-coefficients calculated for a random sample selection of 45 years, iterated 500 times, between the variables: SST and SLP indices at lag 3 (a) and lag 4 (d), and SST index and EA at lag 3 (e) and lag 4 (f). Only causal links with p_value < 0.1 are shown. Red lines show the correspondent β-coefficients represented in (a).

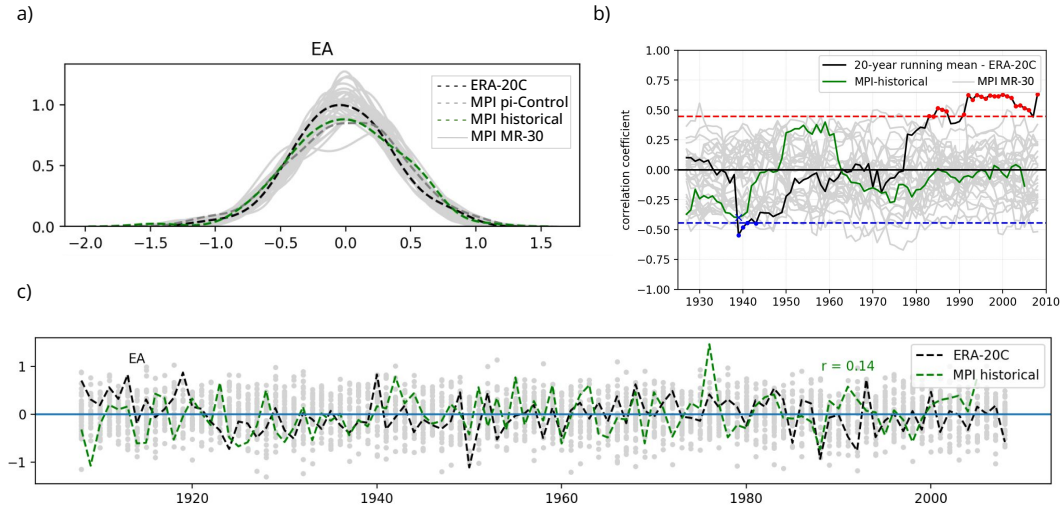


Figure B.4: Model skill in reproducing summer EA and its link with spring SST. a) probability density functions (PDF) of the summer EA and c) time series of the summer EA: light grey colours in a) and c) represent individual ensemble members, and dashed grey line shows the piControl. b) Running-correlation between SST and EA indices for a 20-year window, for ERA-20C (black line), the ensemble mean (grey line) and the historical simulation (green line). Coloured markers indicate significant correlations at the 95% confidence interval, illustrated by the horizontal dashed lines.

B.4.3 *Does MPI-ESM reproduce the observed link?*

We now test whether the causal links detected in ERA-20C during the late period can be reproduced by MR-30. As a first step, we compare the model ability to reproduce the temporal variability of the observed summer EA. We find that MPI-ESM is overall able to reproduce the range of variability (Fig.B.4a) but shows different levels of skill in reproducing the summer EA amongst the different simulation sets. Historical simulations show low agreement with ERA-20C (r = 0.14), whereas MR-30 initialised simulations tend to mostly encompass the observed variability (Fig.B.4c).

Next, we evaluate the model skill in reproducing the spring SST - summer EA relationship. We find that the model shows limited skill, with MR-30 capturing the temporal variability of the relationship in the early, but not in the late period (Figs.B.4b, B.5). A comparison between correlation maps computed for the evaluated periods shows that while historical simulations do not show agreement in the spatial pattern of the spring SST - summer EA relationship against observations, the MR-30 ensemble mean shows an improvement in reproducing the mechanism (Fig.B.5d-f). These results motivate us to assess whether the model is able to reproduce any of the observed causal links, or whether it shows different causal paths than those observed.

We build three different CEN sets to evaluate, respectively, piControl, historical and initialised simulations with MR-30. The variables analysed in the CEN sets are first SST, EA and SLP indices and the time lag of interest is spring - summer (3 and

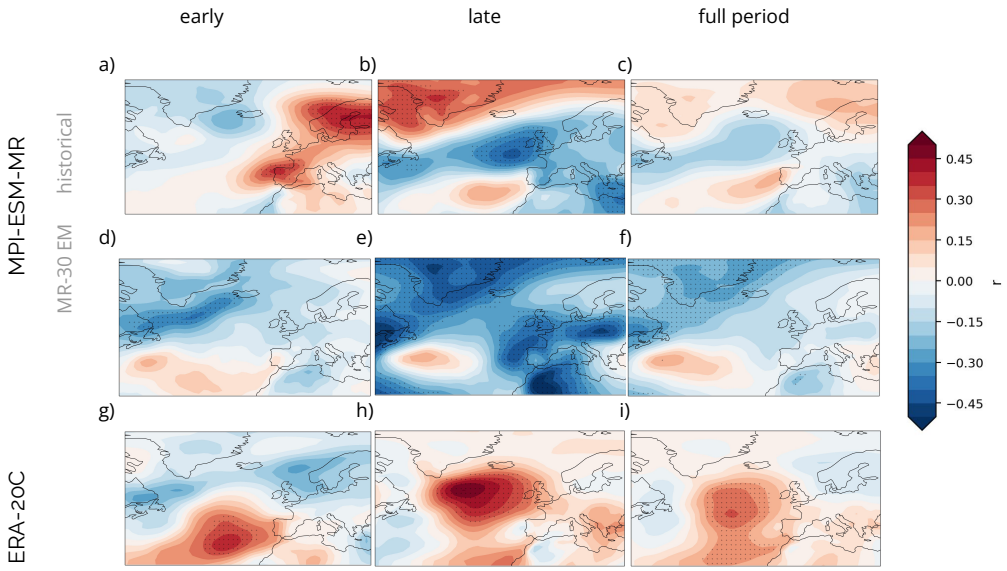


Figure B.5: Comparison of the spatial characteristics of the SST-SLP relationship over the 20th century (model against ERA-20C). Correlation maps show point-wise correlation coefficients for the April-May SST index and July-August SLP means considering *early* (1908-1957; a,d,g), *late* (1958-2008; b,e,h) and *full* periods (1908-2008; c,f,i), respectively. Top row shows results for the MPI-ESM-MR historical simulation, middle row for MPI-ESM-MR 30-member ensemble, and bottom row for ERA-20C.

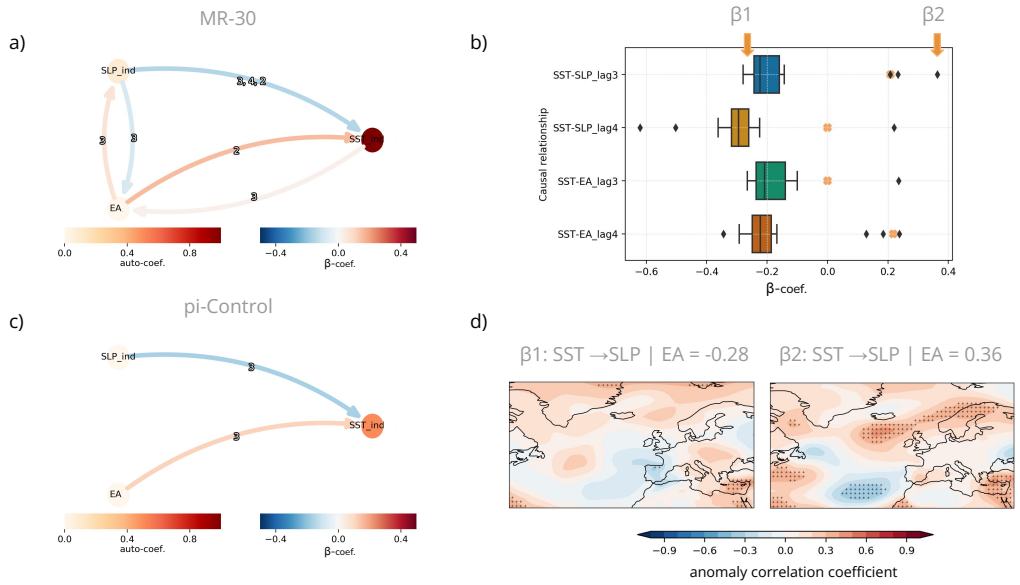


Figure B.6: Causal effect networks for MPI-ESM. a) Causal graph between SST index, EA teleconnection and SLP index for the MPI-ESM-MR 30-member ensemble (MR-30) considering the full period. b) Sensitivity of the causal links between SST, SLP and EA indices at 3 and 4-month lags in the late period. Boxplots show β-coefficients calculated for a random selection of 45 years, sampling one random ensemble member amongst the 30-member set per year. This process is repeated 2000 times and only β-coefficients different from zero are shown (here denoted *MR-30 causal ensemble*). Orange "x" markers represent the β-coefficient calculated from ERA-20C (red lines in Fig.B.3). c) Same as (a) for a 1000-year long piControl simulation with MPI-ESM-MR. Only causal links with p_value < 0.1 are shown. d) Comparison of the impact on SLP predictive skill for 3-4 month lead time in MR-30 against ERA-20C for timeseries showing opposite β-coefficient strengths: a MR-30 causal timeseries with (left) β_1 = -0.28, and (right) β_2 = 0.36. Predictive skill is quantified with anomaly correlation coefficients for the late period. β_1 and β_2 are highlighted in (b) by orange arrows.

4 months lag). While no causal links are found in the historical simulations, we find opposite causal links than those in ERA-20C for the piControl simulation, suggesting an atmospheric forcing into the ocean (e.g. $\beta_{EA \to SST} \approx 0.22$), but no detected causal influence from the ocean on the atmosphere.

Analysing the initialised simulations, we first exploit the full 30-member ensemble MR-30 to build a CEN for the full period (1908-2008), where each constructed timeseries thus comprises 3030 years. We find that MR-30 is able to reproduce a weakly positive SST - EA link (i.e. $\beta_{SST \to EA|SLP} = 0.03$) at 3-month lag, but not at 4-month lag as detected in ERA-20C during the late period, and in much weaker strength (i.e. $\beta_{SST \to EA|SLP} = 0.22$, ERA-20C). No causal links from SST to EA or SLP indices are found when analysing only the late period (1958-2008). Next, we therefore investigate the causal link sensitivity to the sample size and focus on 45-year long timeseries covering the late period, allowing a direct comparison with the sensitivity analysis performed in ERA-20C (Fig.B.6b-e).

B.4.4 *Sensitivity analysis and impact on predictive skill*

We perform a two-step sampling method in our sensitivity analysis with MR-30. First, 45-years are randomly selected in the late period (1958-2008). Second, one ensemble member amongst the full 30-member ensemble is randomly selected in every year. We iterate this process 2000 times, thus generating 2000 45-year-long timeseries for each SST, EA and SLP variables. In each iteration, we build one CEN to analyse whether any causal associations are detected for the sampled SST, SLP and EA timeseries. Our sensitivity results suggest that the model is mostly unable to reproduce the observed links between SST and EA or SLP indices (Fig.B.6b), showing only in very rare cases β-coefficients in the positive range as in ERA-20C (Fig.B.3).

We hypothesise that this MR-30 limitation in reproducing the causal links detected in ERA-20C might constrain the skilful prediction of European summers a season ahead. As a first test, we focus on two particular β-coefficients, namely β_1 = -0.28 and β_2 = 0.36, corresponding to the link SST → SLP at 3-month lag illustrated by orange arrows in Fig.B.6b. In other words, we analyse two cases with strong causal link strength but in opposite signs, with β_2 lying in the observed ERA-20C range.

We perform a predictive skill assessment for each MR-30 causal timeseries respective to β_1 and β_2 against ERA-20C, checking whether the strength of the causal link has a fingerprint in the predictive skill of JA SLP. We find a better agreement between model and reanalysis for β_2 than for β_1, with significant anomaly correlation coefficients (ACC) particularly over the region where spring SST is significantly correlated to summer SLP in ERA-20C (e.g. FigB.2b). However, since positive causal links are only seldom present in MR-30, we are unable to identify a robust fingerprint in the predictive skill related to any of the links between SST and EA or SLP indices.

Nevertheless, identifying a robust fingerprint of spring SST on summer predictive skill could be an important step towards targeting forecasts of opportunity (Mariotti et al., 2020). The correlation analysis in Fig.B.2 suggests that spring SST could influence summer T2m variability over the Euro-Atlantic region in ERA-20C during the late period (Fig.B.7a). Therefore, we perform an additional causality analysis in ERA-20C to highlight where in the T2m field a causal influence of spring SST is expected, and whether this causal relationship could be used to investigate an effect on MR-30's predictive skill.

We compute a causal map (Di Capua et al., 2020b) that represents the beta coefficients calculated for the link between AM SST index and each grid point of the JA T2m field, i.e. $\beta_{SST \to T2m}$ (Fig.B.7b). Results are similar if we calculate the causal link conditioned to either EA or SLP index with a 3-month lag (not shown). In other words, the linear influence of spring SST index on summer T2m cannot be exclusively explained by the linear influence of neither EA nor SLP index in spring. We find two causal regions of opposite signs. The first region shows negative causal links and is located in northwestern Europe, partly encompassing the area used to calculate the $T2m_{CE}$ index expressed in the causal graph in Fig.B.3a. This can be interpreted as an increase of 1 SD in the spring SST index (e.g. warming over subpolar, and cooling over subtropical North Atlantic) causally driving a decrease of about 0.3 SD in the summer T2m field in northwestern Europe. The second region shows positive causal links with $\beta_{SST \to T2m} > 0.5$ and partly overlaps the region of highest correlation between AM SST and JA SLP (e.g. FigB.7b). We therefore compute a causal map for the

link $\beta_{SST \to SLP}$ at 3-month lag, to also highlight in the JA SLP field where AM SST is a causal influence (FigB.7d). We find that a positive causal region associated with AM SST is present in both JA SLP and T2m fields, illustrated by a grey box in FigB.2b,d and denoted *Ridge*. Targeting this causal region, we test the hypothesis that predictive skill of the summer surface climate in MR-30 might be higher for timeseries able to reproduce the causal link strength in ERA-20C ($\beta_{SST \to T2m} > 0.5$), than for those unable to reproduce the link ($\beta_{SST \to T2m} = 0$). We focus on the positive causal region identified in Fig.B.7b,d to calculate a T2m index corresponding to the average over 40°N-55°N; 15°W-34°W (denoted $T2m_{Ridge}$). We perform a CEN sensitivity analysis to investigate this causal link, whereby each constructed CEN consists of SST, $T2m_{Ridge}$ and SLP indices calculated with MR-30. We perform a two-step sampling method to generate 500 timeseries consisting of 45-years randomly selected in the ensemble space during the late period – similarly to the analysis performed for SST, EA and SLP. At 3-month lag, we find that MR-30 is able to reproduce a range of beta coefficients for SST $\to$ $T2m_{Ridge}$, encompassing the observed link 16% of the times (Fig.B.7e). That is, 16% of random combinations in the MR-30 ensemble space result in a MR-30 causal timeseries which represents a causal influence of the SST index in spring (April-May) onto the $T2m_{Ridge}$ in summer (JA).

To test our hypothesis, we evaluate whether the strength of this causal link (i.e. $\beta_{SST \to T2m_{Ridge}}$ at 3-month lag, Fig.B.7c) is imprinted on MR-30's skill in predicting summer SLP, T2m and Z500 for the *Ridge* region a season ahead. We quantify the predictive skill with ACC using ERA-20C as a reference, for two opposite cases in MR-30: i) timeseries showing strong β-coefficients lying in the range $0.6 < \beta < 0.8$ and ii) timeseries showing beta-coefficients = 0, i.e. non-causal. We find 25 samples in i), and we therefore randomly select 25 samples in ii) to enable a direct comparison. We calculate the ACC for each of the total 50 samples, averaging over the Ridge region (Fig.B.7f). We find that a random selection in the ensemble space tends to show higher median and maximum values for the predictive skill of SLP, T2m and Z500 when MR-30 reproduces the causal link $\beta_{SST \to T2m_{Ridge}}$, than when the causal link is absent.

B.5 DISCUSSION

The framework of *forecasts of opportunity* (Mariotti et al., 2020) in seasonal prediction has been increasingly explored to identify physical processes which lead to enhanced predictability and forecast skill. Such a strategy has been particularly useful for summer (Carvalho-Oliveira et al., 2022) and winter (Dobrynin et al., 2018) seasonal predictions in the European region, where predictive skill is limited. Here, we target the summer EA to understand how its seasonal predictability is influenced by spring North Atlantic SSTs using the causal inference-based tool CEN.

Using ERA-20C, our CEN analysis confirms that the spring SST index proposed in Ossó et al., 2018 causally influences the variability of summer SLP in the Euro-Atlantic region with a 3-4 months delay during the late period. Specifically, we find that a 1 SD change in the spring SST index first drives a 0.2 SD change in the summer SLP index at 3-month lag (e.g. March-April SST $\to$ June-July SLP index), and then drives a 0.2 SD change a month later in the summer EA (e.g. March-April SST $\to$ July-August EA, Fig.B.3a). While EA and SLP indices are highly correlated (r = 0.82), the

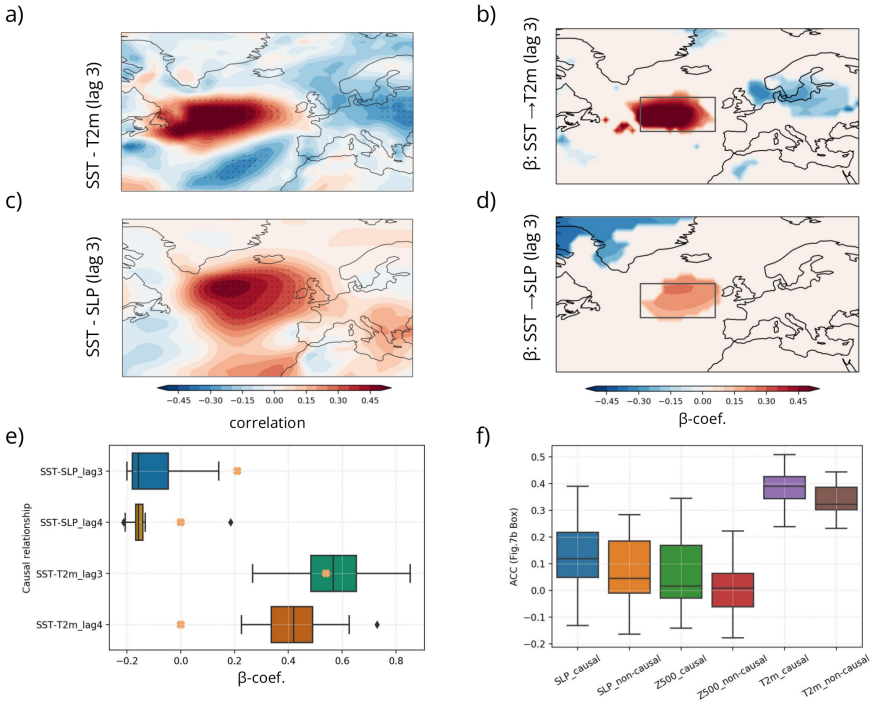


Figure B.7: Spring extratropical SST causal associations and impact on MR-30 predictive skill. Point-wise correlation coefficients for (a) April-May SST index and July-August T2m; (c) April-May SST index and July-August SLP. Correlations are calculated for ERA-20C during the late period, as in Fig.B.2i,b, respectively. Correspondent causal maps, showing the causal associations between (b) April-May SST index and July-August T2m $\beta_{SST \rightarrow T2m}$; (d) April-May SST index and July-August SLP $\beta_{SST \rightarrow SLP}$. Grey boxes highlight the region of strongest causal influence and represents the area used to calculate the T2m index used in (e), and is denoted $T2m_{Ridge}$ in the text. e) Sensitivity of CEN built with the SST, SLP and T2m indices for MR-30. Boxplots show β-coefficients calculated for a random selection of 45 years, sampling one random ensemble member amongst the 30-member set per year. This process is repeated 500 times and only β-coefficients different from zero are shown. Orange "x" markers represent the β-coefficient calculated for a CEN built with the SST, SLP and T2m indices from ERA-20C for the late period. Only causal links with p_value < 0.1 are shown. f) Comparison of the impact on summer surface climate predictive skill in MR-30 against ERA-20C for causal and non-causal MR-30 timeseries. Mean ACCs are shown for July-August sea level pressure (SLP), 2-metre air temperature (T2m) and 500 hPa geopotential height (Z500), averaged over the region highlighted by the grey box. See the text for further description.

position of the area used to calculate the SLP index (Fig.B.2c) only partly overlaps the EA centre of action (Fig.B.1a), which extends further northwest. We speculate that the northward migration of the North Atlantic Jet stream during summer (e.g. Hallam et al., 2022) could explain the delay of a month between the causal link of SST index and EA/SLP indices.

Besides extratropical SSTs, ENSO-related tropical forcing has been suggested to influence the summer EA over more recent decades (1979 - 2016, e.g. O'Reilly et al., 2018; Wulff et al., 2017). As opposed to the mechanism proposed in Ossó et al., 2018, Wulff et al., 2017 suggested that the summer EA is forced by diabatic heating anomalies in the tropical Pacific and Caribbean, and it is characterised by an extratropical Rossby wave train with a centre of action west of the British Isles. The CEN analysis proposed in this paper could therefore be extended to include tropical SST predictors, thus testing how the causal links discussed here could be affected by the influence of additional drivers.

Our findings suggest that the causal links detected in ERA-20C are nonstationary during the 20th century, being present only in the late period. Nonstationarity in teleconnections has been reported by several studies (e.g. Weisheimer et al., 2019; Woollings et al., 2015). In particular, Rieke et al., 2021 used a 700-year pre-industrial control run with MPI-ESM-LR to investigate the tropical link of the summer EA (Wulff et al., 2017) with a statistical model, and showed that the link had a nonstationary behaviour, being present in some multidecadal epochs but not in others. Detecting nonstationarity in the causal links discussed here has an important consequence for the application on predictive skill in seasonal forecasting, implying a limited use of such causal links to target forecasts of opportunity.

Yet, our causality analysis with CEN offers an alternative assessment of MPI-ESM-MR's performance, enabling a direct comparison of the causal links reproduced by the model with those detected in reanalysis. We find that the causal links between spring SST index and summer EA and T2m are absent in piControl and historical simulations, but appear in some 45-year-long timeseries sampled in the initialised ensemble MR-30, thus suggesting a role of initialisation (Fig.B.6). We use a random ensemble subsampling to perform a predictive skill assessment conditioned to MR-30's performance in reproducing causal links between spring SST and both summer EA and $T2m_{Ridge}$. As a result, one ensemble member is randomly chosen among the 30-member per year. Alternatively, performing an ensemble subsampling to calculate an ensemble mean over a subset of ensemble members could provide a better analysis of MR-30's potential. Nevertheless, our results suggest that MR-30's limited performance in reproducing these causal links, in particular between spring SST and the summer EA, might explain its low skill in predicting summer seasonal European climate (e.g. Carvalho-Oliveira et al., 2022; Neddermann et al., 2018).

B.6 CONCLUSIONS

We apply the causal inference-based tool CEN to evaluate the influence of spring North Atlantic extratropical SSTs on the predictability of summer EA and its associated impact on surface climate at seasonal timescales. Our main findings are:

- Analysing ERA-20C, we find that the observed relationship between spring SST index and summer EA is nonstationary during the 20th century, showing

distinct spatial patterns between early (1902-1957) and late (1958-2008) periods. The estimated causal influence of spring SST on summer EA is of $\beta \approx 0.2$.

- We find that this relationship is only considered causal over the late period. A sensitivity analysis of its strength during the late period shows high variability, suggesting that the presence or absence of specific years plays an important role in the quantification of the causal link. This implies that an external physical mechanism not included in our analysis might modulate the spring SST - summer EA causal link.

- In addition to summer EA, we find that the spring SST index causally influences summer T2m ($\beta \approx$ -0.2) over a region in northwestern Europe, and the Ridge region located west of the British Isles ($\beta \approx 0.5$). This causal influence is possibly mediated by the EA.

- We find that piControl and historical simulations are unable to reproduce the causal links detected in ERA-20C. In contrast, our CEN analysis with the full initialised ensemble MR-30 reveals a weakly positive causal link between spring SST and summer EA ($\beta \approx 0.03$).

- However, for 45-year-long timeseries randomly sampled in MR-30, we find that the initialised ensemble is mostly unable to reproduce the spring SST - summer EA link.

- In contrast, MR-30 shows a moderate performance in reproducing the spring SST - summer $T2m_{Ridge}$ causal link. We find that MR-30 tends to show improved predictive skill for summer surface climate predictions over the Ridge region when the spring SST - summer $T2m_{Ridge}$ causal link is correctly reproduced by the model.

In this paper, we show that MPI-ESM-MR has limited performance in reproducing a causal link between spring SST and summer EA amongst uninitialised and initialised model datasets. Our causality analysis therefore sheds light on the limitations of this model in providing skilful seasonal predictions of summer climate, particularly over areas which undergo a significant EA influence. Finally, our results for the initialised ensemble MR-30 show that ensemble members able to reproduce a causal link to spring SST have a potential for regional skill improvement, thereby illustrating how this causality framework could be used to target forecasts of opportunity.

ACKNOWLEDGMENTS The authors would like to thank the Climate Modelling group at Universität Hamburg and the Climate Extremes group at VU Amsterdam for helpful discussions. Model simulations were performed using the high-performance computer at the German Climate Computing Center (DKRZ).

BIBLIOGRAPHY

Astel, Aleksander, Stefan Tsakovski, Pierluigi Barbieri, and Vasil Simeonov (2007). "Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets." In: *Water research* 41.19, pp. 4566–4578.

Athanasiadis, Panos J, Stephen Yeager, Young-Oh Kwon, Alessio Bellucci, David W Smith, and Stefano Tibaldi (2020). "Decadal predictability of North Atlantic blocking and the NAO." In: *NPJ Climate and Atmospheric Science* 3.1, pp. 1–10.

Ayarzagüena, Blanca and Encarna Serrano (2009). "Monthly characterization of the tropospheric circulation over the Euro-Atlantic area in relation with the timing of stratospheric final warmings." In: *Journal of Climate* 22.23, pp. 6313–6324.

Balmaseda, Magdalena Alonso, Kristian Mogensen, and Anthony T Weaver (2013). "Evaluation of the ECMWF ocean reanalysis system ORAS4." In: *Quarterly journal of the royal meteorological society* 139.674, pp. 1132–1161.

Barnes, Elizabeth A and Dennis L Hartmann (2010). "Testing a theory for the effect of latitude on the persistence of eddy-driven jets using CMIP3 simulations." In: *Geophysical Research Letters* 37.15.

Barnston, Anthony G and Robert E Livezey (1987). "Classification, seasonality and persistence of low-frequency atmospheric circulation patterns." In: *Monthly weather review* 115.6, pp. 1083–1126.

Bastos, Ana, Ivan A Janssens, Célia M Gouveia, Ricardo M Trigo, Philippe Ciais, Frédéric Chevallier, Josep Penuelas, Christian Rödenbeck, Shilong Piao, Pierre Friedlingstein, et al. (2016). "European land CO 2 sink influenced by NAO and East-Atlantic Pattern coupling." In: *Nature communications* 7.1, pp. 1–9.

Belmecheri, Soumaya, Flurin Babst, Amy R Hudson, Julio Betancourt, and Valerie Trouet (2017). "Northern Hemisphere jet stream position indices as diagnostic tools for climate and ecosystem dynamics." In: *Earth Interactions* 21.8, pp. 1–23.

Beverley, Jonathan D, Steven J Woolnough, Laura H Baker, Stephanie J Johnson, and Antje Weisheimer (2019). "The northern hemisphere circumglobal teleconnection in a seasonal forecast model and its relationship to European summer forecast skill." In: *Climate dynamics* 52.5-6, pp. 3759–3771.

Bladé, Ileana, Brant Liebmann, Didac Fortuny, and Geert Jan van Oldenborgh (2012). "Observed and simulated impacts of the summer NAO in Europe: implications for projected drying in the Mediterranean region." In: *Climate dynamics* 39.3-4, pp. 709–727.

Branstator, Grant (2002). "Circumglobal teleconnections, the jet stream waveguide, and the North Atlantic Oscillation." In: *Journal of Climate* 15.14, pp. 1893–1910.

Budayan, Cenk, Irem Dikmen, and M Talat Birgonul (2009). "Comparing the performance of traditional cluster analysis, self-organizing maps and fuzzy C-means method for strategic grouping." In: *Expert Systems with Applications* 36.9, pp. 11772–11781.

Bueh, Cholaw and Hisashi Nakamura (2007). "Scandinavian pattern and its climatic impact." In: *Quarterly Journal of the Royal Meteorological Society: A journal of the*

atmospheric sciences, applied meteorology and physical oceanography 133.629, pp. 2117–2131.

Carvalho-Oliveira, Julianna, Leonard F Borchert, Eduardo Zorita, and Johanna Baehr (2022). "Self-organizing maps identify windows of opportunity for seasonal European summer predictions." In: *Frontiers in Climate* 4.

Carvalho-Oliveira, Julianna, Leonard Friedrich Borchert, Aurélie Duchez, Mikhail Dobrynin, and Johanna Baehr (2021). "Subtle influence of the Atlantic Meridional Overturning Circulation (AMOC) on seasonal sea surface temperature (SST) hindcast skill in the North Atlantic." In: *Weather and Climate Dynamics* 2.3, pp. 739–757.

Cassou, Christophe, Laurent Terray, James W Hurrell, and Clara Deser (2004). "North Atlantic winter climate regimes: Spatial asymmetry, stationarity with time, and oceanic forcing." In: *Journal of Climate* 17.5, pp. 1055–1068.

Cassou, Christophe, Laurent Terray, and Adam S Phillips (2005). "Tropical Atlantic influence on European heat waves." In: *Journal of climate* 18.15, pp. 2805–2811.

Cattiaux, Julien, Benjamin Quesada, Ara Arakélian, Francis Codron, Robert Vautard, and Pascal Yiou (2013). "North-Atlantic dynamics and European temperature extremes in the IPSL model: sensitivity to atmospheric resolution." In: *Climate dynamics* 40.9-10, pp. 2293–2310.

Cohen, Judah and Justin Jones (2011). "A new index for more accurate winter predictions." In: *Geophysical Research Letters* 38.21.

Collins, Matthew (2002). "Climate predictability on interannual to decadal time scales: The initial value problem." In: *Climate Dynamics* 19.8, pp. 671–692.

Comas-Bru, Laia and Armand Hernández (2018). "Reconciling North Atlantic climate modes: revised monthly indices for the East Atlantic and the Scandinavian patterns beyond the 20th century." In: *Earth System Science Data* 10.4, pp. 2329–2344.

Comas-Bru, Laia and Frank McDermott (2014). "Impacts of the EA and SCA patterns on the European twentieth century NAO–winter climate relationship." In: *Quarterly Journal of the Royal Meteorological Society* 140.679, pp. 354–363.

Comiso, Josefino C (1995). *SSM/I sea ice concentrations using the bootstrap algorithm*. Vol. 1380. National Aeronautics and Space Administration, Goddard Space Flight Center.

Cornes, Richard C, Philip D Jones, Keith R Briffa, and Timothy J Osborn (2013). "Estimates of the North Atlantic Oscillation back to 1692 using a Paris–London westerly index." In: *International Journal of Climatology* 33.1, pp. 228–248.

Cortesi, Nicola, Nube Gonzalez-Reviriego, Albert Soret, and Francisco J Doblas-Reyes (2017). *Weather regimes: ECMWF seasonal forecasts verification. Technical report.* Tech. rep. Barcelona Supercomputing Center.

Coumou, Dim, Giorgia Di Capua, Steve Vavrus, Lei Wang, and Simon Wang (2018). "The influence of Arctic amplification on mid-latitude summer circulation." In: *Nature Communications* 9.1, pp. 1–12.

Coumou, Dim, Vladimir Petoukhov, Stefan Rahmstorf, Stefan Petri, and Hans Joachim Schellnhuber (2014). "Quasi-resonant circulation regimes and hemispheric synchronization of extreme weather in boreal summer." In: *Proceedings of the National Academy of Sciences* 111.34, pp. 12331–12336.

Craig, Philip M and Richard P Allan (2021). "The role of teleconnection patterns in the variability and trends of growing season indices across Europe." In: *International Journal of Climatology*.

Czaja, Arnaud and Claude Frankignoul (1999). "Influence of the North Atlantic SST on the atmospheric circulation." In: *Geophysical Research Letters* 26.19, pp. 2969–2972.

— (2002). "Observed impact of Atlantic SST anomalies on the North Atlantic Oscillation." In: *Journal of Climate* 15.6, pp. 606–623.

Dee, Dick P, S M Uppala, AJ Simmons, Paul Berrisford, Paul Poli, Shinya Kobayashi, U Andrae, MA Balmaseda, G Balsamo, d P Bauer, et al. (2011). "The ERA-Interim reanalysis: Configuration and performance of the data assimilation system." In: *Quarterly Journal of the royal meteorological society* 137.656, pp. 553–597.

Delgado-Torres, Carlos, Deborah Verfaillie, Elsa Mohino, and MG Donat (2022). "Representation and annual to decadal predictability of Euro-Atlantic weather regimes in the CMIP6 version of the EC-Earth coupled climate model." In: *Journal of Geophysical Research: Atmospheres*, e2022JD036673.

Deser, Clara, Michael A Alexander, and Michael S Timlin (2003). "Understanding the persistence of sea surface temperature anomalies in midlatitudes." In: *Journal of Climate* 16.1, pp. 57–72.

Di Capua, Giorgia, Marlene Kretschmer, Reik V Donner, Bart van den Hurk, Ramesh Vellore, Raghavan Krishnan, and Dim Coumou (2020a). "Tropical and mid-latitude teleconnections interacting with the Indian summer monsoon rainfall: a theory-guided causal effect network approach." In: *Earth System Dynamics* 11.1, pp. 17–34.

Di Capua, Giorgia, Jakob Runge, Reik V Donner, Bart van den Hurk, Andrew G Turner, Ramesh Vellore, Raghavan Krishnan, and Dim Coumou (2020b). "Dominant patterns of interaction between the tropics and mid-latitudes in boreal summer: causal relationships and the role of timescales." In: *Weather and Climate Dynamics* 1.2, pp. 519–539.

Doblas-Reyes, Francisco J, Javier García-Serrano, Fabian Lienert, Aida Pintó Biescas, and Luis RL Rodrigues (2013). "Seasonal climate predictability and forecasting: status and prospects." In: *Wiley Interdisciplinary Reviews: Climate Change* 4.4, pp. 245–268.

Dobrynin, Mikhail, Daniela IV Domeisen, Wolfgang A Müller, Louisa Bell, Sebastian Brune, Felix Bunzel, André Düsterhus, Kristina Fröhlich, Holger Pohlmann, and Johanna Baehr (2018). "Improved teleconnection-based dynamical seasonal predictions of boreal winter." In: *Geophysical Research Letters* 45.8, pp. 3605–3614.

Domeisen, Daniela IV, Gualtiero Badin, and Inga M Koszalka (2018). "How predictable are the Arctic and North Atlantic Oscillations? Exploring the variability and predictability of the Northern Hemisphere." In: *Journal of Climate* 31.3, pp. 997–1014.

Domeisen, Daniela IV, Christian M Grams, and Lukas Papritz (2020). "The role of North Atlantic–European weather regimes in the surface impact of sudden stratospheric warming events." In: *Weather and Climate Dynamics* 1.2, pp. 373–388.

Dong, Buwen, Rowan T Sutton, Tim Woollings, and Kevin Hodges (2013a). "Variability of the North Atlantic summer storm track: mechanisms and impacts on European climate." In: *Environmental Research Letters* 8.3, p. 034037.

Dong, Buwen, Rowan Sutton, and Tim Woollings (2013b). "The extreme European summer 2012." In: *Bulletin of the American Meteorological Society* 94.9, s28–s32.

Dorado-Liñán, Isabel, Blanca Ayarzagüena, Flurin Babst, Guobao Xu, Luis Gil, Giovanna Battipaglia, Allan Buras, Vojtěch Čada, J Julio Camarero, Liam Cavin, et al. (2022). "Jet stream position explains regional anomalies in European beech forest productivity and tree growth." In: *Nature communications* 13.1, pp. 1–10.

Duchez, Aurélie, Eleanor Frajka-Williams, Simon A Josey, Dafydd G Evans, Jeremy P Grist, Robert Marsh, Gerard D McCarthy, Bablu Sinha, David I Berry, and Joël JM Hirschi (2016). "Drivers of exceptionally cold North Atlantic Ocean temperatures and their link to the 2015 European heat wave." In: *Environmental Research Letters* 11.7, p. 074004.

Dunstone, Nick, Doug Smith, Steven Hardiman, Rosie Eade, Margaret Gordon, Leon Hermanson, Gillian Kay, and Adam Scaife (2019). "Skilful real-time seasonal forecasts of the dry Northern European summer 2018." In: *Geophysical Research Letters* 46.21, pp. 12368–12376.

Dunstone, Nick, Doug Smith, Adam Scaife, Leon Hermanson, Rosie Eade, Niall Robinson, Martin Andrews, and Jeff Knight (2016). "Skilful predictions of the winter North Atlantic Oscillation one year ahead." In: *Nature Geoscience* 9.11, pp. 809–814.

Dunstone, Nick, Doug Smith, Adam Scaife, Leon Hermanson, David Fereday, Chris O'Reilly, Alison Stirling, Rosie Eade, Margaret Gordon, Craig MacLachlan, et al. (2018). "Skilful seasonal predictions of summer European rainfall." In: *Geophysical Research Letters* 45.7, pp. 3246–3254.

Eade, Rosie, Doug Smith, Adam Scaife, Emily Wallace, Nick Dunstone, Leon Hermanson, and Niall Robinson (2014). "Do seasonal-to-decadal climate predictions underestimate the predictability of the real world?" In: *Geophysical research letters* 41.15, pp. 5620–5628.

Exner, Felix M (1913). "Uber monatliche Witterungs-anomalien auf der nordlichen Erdhalfte im Winter." In: *Sitzb. Mathem.-Naturw, Kl., Acad. der Wissenshaften*, pp. 1165–1240.

Fabiano, F, HM Christensen, K Strommen, P Athanasiadis, A Baker, R Schiemann, and S Corti (2020). "Euro-Atlantic weather Regimes in the PRIMAVERA coupled climate simulations: impact of resolution and mean state biases on model performance." In: *Climate Dynamics* 54.11, pp. 5031–5048.

Folland, Chris K, Jeff Knight, Hans W Linderholm, David Fereday, Sarah Ineson, and James W Hurrell (2009). "The summer North Atlantic Oscillation: past, present, and future." In: *Journal of Climate* 22.5, pp. 1082–1103.

Forest, Florent, Mustapha Lebbah, Hanane Azzag, and Jérôme Lacaille (2020). "A survey and implementation of performance metrics for self-organized maps." In: *arXiv preprint arXiv:2011.05847*.

Gastineau, Guillaume and Claude Frankignoul (2015). "Influence of the North Atlantic SST variability on the atmospheric circulation during the twentieth century." In: *Journal of Climate* 28.4, pp. 1396–1416.

Giorgetta, Marco A, Johann Jungclaus, Christian H Reick, Stephanie Legutke, Jürgen Bader, Michael Böttinger, Victor Brovkin, Traute Crueger, Monika Esch, Kerstin Fieg, et al. (2013). "Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5." In: *Journal of Advances in Modeling Earth Systems* 5.3, pp. 572–597.

Giorgetta, Marco, Johann Jungclaus, Christian Reick, Stephanie Legutke, Victor Brovkin, Traute Crueger, Monika Esch, Kerstin Fieg, Ksenia Glushak, Veronika Gayler, et al. (2011). "CMIP5 simulations of the Max Planck Institute for Meteorology (MPI-M) based on the MPI-ESM-LR model: The piControl experiment, served by ESGF." In.

Gu, Qinxue and Melissa Gervais (2022). "Diagnosing Two-Way Coupling in Decadal North Atlantic SST Variability Using Time-Evolving Self-Organizing Maps." In: *Geophysical Research Letters* 49.8, e2021GL096560.

Haarsma, Reindert J, Frank Selten, Bart Vd Hurk, Wilco Hazeleger, and Xueli Wang (2009). "Drier Mediterranean soils due to greenhouse warming bring easterly winds over summertime central Europe." In: *Geophysical research letters* 36.4.

Hall, Richard J and Edward Hanna (2018). "North Atlantic circulation indices: links with summer and winter UK temperature and precipitation and implications for seasonal forecasting." In: *International Journal of Climatology* 38, e660–e677.

Hall, Richard J, Julie M Jones, Edward Hanna, Adam A Scaife, and Róbert Erdélyi (2017). "Drivers and potential predictability of summer time North Atlantic polar front jet variability." In: *Climate Dynamics* 48.11-12, pp. 3869–3887.

Hallam, Samantha, Simon A Josey, Gerard D McCarthy, and Joël J-M Hirschi (2022). "A regional (land–ocean) comparison of the seasonal to decadal variability of the Northern Hemisphere jet stream 1871–2011." In: *Climate Dynamics*, pp. 1–22.

Hanna, Edward, Thomas E Cropper, Philip D Jones, Adam A Scaife, and Rob Allan (2015). "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." In: *International Journal of Climatology* 35.9, pp. 2540–2554.

Hannachi, Abdel, David M Straus, Christian LE Franzke, Susanna Corti, and Tim Woollings (2017). "Low-frequency nonlinearity and regime behavior in the Northern Hemisphere extratropical atmosphere." In: *Reviews of Geophysics* 55.1, pp. 199–234.

Hansen, Felicitas, Richard J Greatbatch, Gereon Gollan, Thomas Jung, and Antje Weisheimer (2017). "Remote control of North Atlantic Oscillation predictability via the stratosphere." In: *Quarterly Journal of the Royal Meteorological Society* 143.703, pp. 706–719.

Hildebrandsson, Hugo Hildebrand (1897). *Quelques recherches sur les centres d'action de l'atmosphère.* Norstedt & Söner.

Ho, Chun Kit, Ed Hawkins, Len Shaffrey, Jochen Bröcker, Leon Hermanson, James M Murphy, Doug M Smith, and Rosie Eade (2013). "Examining reliability of seasonal to decadal sea surface temperature forecasts: The role of ensemble dispersion." In: *Geophysical Research Letters* 40.21, pp. 5770–5775.

Hunt, Freja K, Joël J-M Hirschi, Bablu Sinha, Kevin Oliver, and Neil Wells (2013). "Combining point correlation maps with self-organising maps to compare ob-

served and simulated atmospheric teleconnection patterns." In: *Tellus A: Dynamic Meteorology and Oceanography* 65.1, p. 20822.

Hurrell, James W, Yochanan Kushnir, Geir Ottersen, and Martin Visbeck (2003). "An overview of the North Atlantic oscillation." In: *Geophysical Monograph-American Geophysical Union* 134, pp. 1–36.

IPCC (2013). *Summary for Policymakers Climate Change 2013: The Physical Science Basis.* Tech. rep. Cambridge, UK: Cambridge., 571—657.

Jain, Anil K and Richard C Dubes (1988). *Algorithms for clustering data*. Prentice-Hall, Inc.

Johnson, Nathaniel C (2013). "How many ENSO flavors can we distinguish?" In: *Journal of Climate* 26.13, pp. 4816–4827.

Johnson, Nathaniel C, Steven B Feldstein, and Bruno Tremblay (2008). "The continuum of Northern Hemisphere teleconnection patterns and a description of the NAO shift with the use of self-organizing maps." In: *Journal of Climate* 21.23, pp. 6354–6371.

Josey, Simon A, Samuel Somot, and Mikis Tsimplis (2011). "Impacts of atmospheric modes of variability on Mediterranean Sea surface heat exchange." In: *Journal of Geophysical Research: Oceans* 116.C2.

Jungclaus, JH, Nils Fischer, Helmuth Haak, K Lohmann, J Marotzke, D Matei, U Mikolajewicz, D Notz, and JS Von Storch (2013). "Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model." In: *Journal of Advances in Modeling Earth Systems* 5.2, pp. 422–446.

Kashinath, Karthik, Mayur Mudigonda, Sol Kim, Lukas Kapp-Schwoerer, Andre Graubner, Ege Karaismailoglu, Leo Von Kleist, Thorsten Kurth, Annette Greiner, Ankur Mahesh, et al. (2021). "ClimateNet: an expert-labeled open dataset and deep learning architecture for enabling high-precision analyses of extreme weather." In: *Geoscientific Model Development* 14.1, pp. 107–124.

Klavans, Jeremy M, Mark A Cane, Amy C Clement, and Lisa N Murphy (2021). "NAO predictability from external forcing in the late 20th century." In: *NPJ climate and atmospheric science* 4.1, pp. 1–8.

Kohonen, Teuvo (1984). *Self-Organization and Associative Memory*. Springer.

— (2013). "Essentials of the self-organizing map." In: *Neural networks* 37, pp. 52–65.

Kretschmer, Marlene, Dim Coumou, Jonathan F Donges, and Jakob Runge (2016). "Using causal effect networks to analyze different Arctic drivers of midlatitude winter circulation." In: *Journal of climate* 29.11, pp. 4069–4081.

Kushnir, Y, WA Robinson, I Bladé, NMJ Hall, Sutton Peng, and R Sutton (2002). "Atmospheric GCM response to extratropical SST anomalies: Synthesis and evaluation." In: *Journal of Climate* 15.16, pp. 2233–2256.

Leloup, Julie A, Zouhair Lachkar, Jean-Philippe Boulanger, and Sylvie Thiria (2007). "Detecting decadal changes in ENSO using neural networks." In: *Climate dynamics* 28.2, pp. 147–162.

Liu, Yonggang, Robert H Weisberg, and Christopher NK Mooers (2006). "Performance evaluation of the self-organizing map for feature extraction." In: *Journal of Geophysical Research: Oceans* 111.C5.

Liu, Zhengyu and Mike Alexander (2007). "Atmospheric bridge, oceanic tunnel, and global climatic teleconnections." In: *Reviews of Geophysics* 45.2.

Lledo, Llorenç, Irene Cionni, Veronica Torralba, Pierre-Antoine Bretonniere, and Margarida Samso (2020). "Seasonal prediction of Euro-Atlantic teleconnections from multiple systems." In: *Environmental Research Letters* 15.7, p. 074009.

Lorenz, Edward N (1963). "Deterministic nonperiodic flow." In: *Journal of atmospheric sciences* 20.2, pp. 130–141.

Luo, Dehai, Xiaodan Chen, and Steven B Feldstein (2018). "Linear and nonlinear dynamics of North Atlantic Oscillations: A new thinking of symmetry breaking." In: *Journal of the Atmospheric Sciences* 75.6, pp. 1955–1977.

Maidens, Anna, Jeff R Knight, and Adam A Scaife (2021). "Tropical and stratospheric influences on winter atmospheric circulation patterns in the North Atlantic sector." In: *Environmental Research Letters* 16.2, p. 024035.

Mariotti, Annarita, Cory Baggett, Elizabeth A Barnes, Emily Becker, Amy Butler, Dan C Collins, Paul A Dirmeyer, Laura Ferranti, Nathaniel C Johnson, Jeanine Jones, et al. (2020). "Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond." In: *Bulletin of the American Meteorological Society* 101.5, E608–E625.

Matsumura, Shinji, Xiangdong Zhang, and Koji Yamazaki (2014). "Summer Arctic atmospheric circulation response to spring Eurasian snow cover and its possible linkage to accelerated sea ice decrease." In: *Journal of Climate* 27.17, pp. 6551–6558.

Mauritsen, Thorsten, Bjorn Stevens, Erich Roeckner, Traute Crueger, Monika Esch, Marco Giorgetta, Helmuth Haak, Johann Jungclaus, Daniel Klocke, Daniela Matei, et al. (2012). "Tuning the climate of a global model." In: *Journal of advances in modeling Earth systems* 4.3.

Merryfield, William J, Johanna Baehr, Lauriane Batté, Emily J Becker, Amy H Butler, Caio AS Coelho, Gokhan Danabasoglu, Paul A Dirmeyer, Francisco J Doblas-Reyes, Daniela IV Domeisen, et al. (2020). "Current and emerging developments in subseasonal to decadal prediction." In: *Bulletin of the American Meteorological Society* 101.6, E869–E896.

Mezzina, Bianca, Javier García-Serrano, Ileana Bladé, and Fred Kucharski (2020). "Dynamics of the ENSO teleconnection and NAO variability in the North Atlantic–European late winter." In: *Journal of Climate* 33.3, pp. 907–923.

Michelangeli, Paul-Antoine, Robert Vautard, and Bernard Legras (1995). "Weather regimes: Recurrence and quasi stationarity." In: *Journal of the atmospheric sciences* 52.8, pp. 1237–1256.

Mignot, Juliette, Carlos Mejia, Charles Sorror, Adama Sylla, Michel Crépon, and Sylvie Thiria (2020). "Towards an objective assessment of climate multi-model ensembles–a case study: the Senegalo-Mauritanian upwelling region." In: *Geoscientific Model Development* 13.6, pp. 2723–2742.

Mishra, Niti, Chloé Prodhomme, and Virginie Guemas (2019). "Multi-model skill assessment of seasonal temperature and precipitation forecasts over Europe." In: *Climate Dynamics* 52.7, pp. 4207–4225.

Moore, GWK, RS Pickart, and IA Renfrew (2011). "Complexities in the climate of the subpolar North Atlantic: A case study from the winter of 2007." In: *Quarterly Journal of the Royal Meteorological Society* 137.656, pp. 757–767.

Moore, GWK, IA Renfrew, and Robert S Pickart (2013). "Multidecadal mobility of the North Atlantic oscillation." In: *Journal of Climate* 26.8, pp. 2453–2466.

Neddermann, Nele-Charlotte, Wolfgang A Müller, Mikhail Dobrynin, André Düsterhus, and Johanna Baehr (2018). "Seasonal predictability of European summer climate re-assessed." In: *Climate Dynamics*, pp. 1–18.

Neddermann, Nele (2019). "Seasonal prediction of European summer climate: a process-based approach." PhD thesis. Universität Hamburg Hamburg.

Nie, Yu, Yang Zhang, Gang Chen, and Xiu-Qun Yang (2016). "Delineating the barotropic and baroclinic mechanisms in the midlatitude eddy-driven jet response to lower-tropospheric thermal forcing." In: *Journal of the Atmospheric Sciences* 73.1, pp. 429–448.

North, Gerald R, Thomas L Bell, Robert F Cahalan, and Fanthune J Moeng (1982). "Sampling errors in the estimation of empirical orthogonal functions." In: *Monthly weather review* 110.7, pp. 699–706.

O'Reilly, Christopher H, Antje Weisheimer, Tim Woollings, Lesley J Gray, and Dave MacLeod (2019). "The importance of stratospheric initial conditions for winter North Atlantic Oscillation predictability and implications for the signal-to-noise paradox." In: *Quarterly Journal of the Royal Meteorological Society* 145.718, pp. 131–146.

Oliveira, Julianna C, Eduardo Zorita, Vimal Koul, Thomas Ludwig, and Johanna Baehr (2020). "Forecast opportunities for European summer climate ensemble predictions using Self-Organising Maps." In: *Proceedings of the 10th International Conference on Climate Informatics*, pp. 67–71.

Osborne, Joe M, Mat Collins, James A Screen, Stephen I Thomson, and Nick Dunstone (2020). "The North Atlantic as a driver of summer atmospheric circulation." In: *Journal of Climate* 33.17, pp. 7335–7351.

Ossó, Albert, Rowan Sutton, Len Shaffrey, and Buwen Dong (2018). "Observational evidence of European summer weather patterns predictable from spring." In: *Proceedings of the National Academy of Sciences* 115.1, pp. 59–63.

— (2020). "Development, Amplification, and Decay of Atlantic/European Summer Weather Patterns Linked to Spring North Atlantic Sea Surface Temperatures." In: *Journal of Climate* 33.14, pp. 5939–5951.

O'Reilly, Christopher H, Tim Woollings, and Laure Zanna (2017). "The dynamical influence of the Atlantic Multidecadal Oscillation on continental climate." In: *Journal of Climate* 30.18, pp. 7213–7230.

O'Reilly, Christopher H, Tim Woollings, Laure Zanna, and Antje Weisheimer (2018). "The impact of tropical precipitation on summertime Euro-Atlantic circulation via a circumglobal wave train." In: *Journal of Climate* 31.16, pp. 6481–6504.

Pearl, Judea et al. (2000). "Models, reasoning and inference." In: *Cambridge, UK: CambridgeUniversityPress* 19.2.

Peng, Shiling, Walter A Robinson, and Shuanglin Li (2003). "Mechanisms for the NAO responses to the North Atlantic SST tripole." In: *Journal of Climate* 16.12, pp. 1987–2004.

Pithan, Felix, Theodore G Shepherd, Giuseppe Zappa, and Irina Sandu (2016). "Climate model biases in jet streams, blocking and storm tracks resulting from missing orographic drag." In: *Geophysical Research Letters* 43.13, pp. 7231–7240.

Poli, Paul, Hans Hersbach, Dick P Dee, Paul Berrisford, Adrian J Simmons, Frédéric Vitart, Patrick Laloyaux, David GH Tan, Carole Peubey, Jean-Noël Thépaut, et

al. (2016). "ERA-20C: An atmospheric reanalysis of the twentieth century." In: *Journal of Climate* 29.11, pp. 4083–4097.

Polo, Irene, Albin Ullmann, Pascal Roucou, and Bernard Fontaine (2011). "Weather regimes in the Euro-Atlantic and Mediterranean sector, and relationship with West African rainfall over the 1989–2008 period from a self-organizing maps approach." In: *Journal of Climate* 24.13, pp. 3423–3432.

Rahmstorf, Stefan and Dim Coumou (2011). "Increase of extreme events in a warming world." In: *Proceedings of the National Academy of Sciences* 108.44, pp. 17905–17909.

Reusch, David B, Richard B Alley, and Bruce C Hewitson (2005). "Relative performance of self-organizing maps and principal component analysis in pattern extraction from synthetic climatological data." In: *Polar Geography* 29.3, pp. 188–212.

— (2007). "North Atlantic climate variability from a self-organizing map perspective." In: *Journal of Geophysical Research: Atmospheres* 112.D2.

Rieke, Ole, Richard J Greatbatch, and Gereon Gollan (2021). "Nonstationarity of the link between the Tropics and the summer East Atlantic pattern." In: *Atmospheric Science Letters*, e1026.

Rousi, E, U Ulbrich, HW Rust, and C Anagnostolpoulou (2017). "An NAO climatology in reanalysis data with the use of self-organizing maps." In: *Perspectives on Atmospheric Sciences*. Springer, pp. 719–724.

Rousi, Efi, Kai Kornhuber, Goratz Beobide-Arsuaga, Fei Luo, and Dim Coumou (2022). "Accelerated western European heatwave trends linked to more-persistent double jets over Eurasia." In: *Nature communications* 13.1, pp. 1–11.

Rousi, Eftychia, Frank Selten, Stefan Rahmstorf, and Dim Coumou (2021). "Changes in North Atlantic atmospheric circulation in a warmer climate favor winter flooding and summer drought over Europe." In: *Journal of Climate* 34.6, pp. 2277–2295.

Runge, Jakob, Peer Nowack, Marlene Kretschmer, Seth Flaxman, and Dino Sejdinovic (2019). "Detecting and quantifying causal associations in large nonlinear time series datasets." In: *Science Advances* 5.11, eaau4996.

Runge, Jakob, Vladimir Petoukhov, Jonathan F Donges, Jaroslav Hlinka, Nikola Jajcay, Martin Vejmelka, David Hartman, Norbert Marwan, Milan Paluš, and Jürgen Kurths (2015). "Identifying causal gateways and mediators in complex spatio-temporal systems." In: *Nature communications* 6.1, pp. 1–10.

Scaife, AA, A Arribas, E Blockley, A Brookshaw, RT Clark, N Dunstone, R Eade, D Fereday, CK Folland, M Gordon, et al. (2014). "Skillful long-range prediction of European and North American winters." In: *Geophysical Research Letters* 41.7, pp. 2514–2519.

Scaife, Adam A, Jeff R Knight, Geoff K Vallis, and Chris K Folland (2005). "A stratospheric influence on the winter NAO and North Atlantic surface climate." In: *Geophysical Research Letters* 32.18.

Scher, Sebastian (2018). "Toward Data-Driven Weather and Climate Forecasting: Approximating a Simple General Circulation Model With Deep Learning." In: *Geophysical Research Letters* 45.22, pp. 12–616.

Schuhen, Nina, Nathalie Schaller, Hannah C Bloomfield, David J Brayshaw, Llorenç Lledó, Irene Cionni, and Jana Sillmann (2022). "Predictive Skill of Teleconnection

Patterns in Twentieth Century Seasonal Hindcasts and Their Relationship to Extreme Winter Temperatures in Europe." In: *Geophysical Research Letters*, e2020GL092360.
Screen, James A (2013). "Influence of Arctic sea ice on European summer precipitation." In: *Environmental Research Letters* 8.4, p. 044015.
Shepherd, Theodore G (2014). "Atmospheric circulation as a source of uncertainty in climate change projections." In: *Nature Geoscience* 7.10, pp. 703–708.
Shuila, J and J Kinter III (2006). "Predictability of seasonal climate variations: A pedagogical review." In.
Stevens, Bjorn, Marco Giorgetta, Monika Esch, Thorsten Mauritsen, Traute Crueger, Sebastian Rast, Marc Salzmann, Hauke Schmidt, Jürgen Bader, Karoline Block, et al. (2013). "Atmospheric component of the MPI-M Earth system model: ECHAM6." In: *Journal of Advances in Modeling Earth Systems* 5.2, pp. 146–172.
Teng, Haiyan, Grant Branstator, Hailan Wang, Gerald A Meehl, and Warren M Washington (2013). "Probability of US heat waves affected by a subseasonal planetary wave pattern." In: *Nature Geoscience* 6.12, pp. 1056–1061.
Thompson, David WJ and John M Wallace (2000). "Annular modes in the extratropical circulation. Part I: Month-to-month variability." In: *Journal of climate* 13.5, pp. 1000–1016.
Tremblay, L-Bruno (2001). "Can we consider the Arctic Oscillation independently from the Barents Oscillation?" In: *Geophysical research letters* 28.22, pp. 4227–4230.
Trenberth, Kevin E. (2022). "Teleconnections and Patterns of Variability." In: *The Changing Flow of Energy Through the Climate System*. Cambridge University Press, 162–179. DOI: 10.1017/9781108979030.014.
Trenberth, Kevin E, Grant W Branstator, David Karoly, Arun Kumar, Ngar-Cheung Lau, and Chester Ropelewski (1998). "Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures." In: *Journal of Geophysical Research: Oceans* 103.C7, pp. 14291–14324.
Troccoli, Alberto (2010). "Seasonal climate forecasting." In: *Meteorological Applications* 17.3, pp. 251–268.
Trouet, V, F Babst, and M Meko (2018). "Recent enhanced high-summer North Atlantic Jet variability emerges from three-century context." In: *Nature communications* 9.1, pp. 1–9.
Vautard, Robert (1990). "Multiple weather regimes over the North Atlantic: Analysis of precursors and successors." In: *Monthly weather review* 118.10, pp. 2056–2081.
Vettigli, Giuseppe (2019). *MiniSom: minimalistic and NumPybased implementation of the Self Organizing Map. Release 2.1.5*. http://github.com/JustGlowing/minisom.
Vitart, Frédéric and Andrew W Robertson (2018). "The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events." In: *npj Climate and Atmospheric Science* 1.1, pp. 1–7.
Walker, Gilbert T (1924). "Correlations in seasonal variations of weather. I. A further study of world weather." In: *Mem. Indian Meteorol. Dep.* 24, pp. 275–332.
Wallace, John M and David S Gutzler (1981). "Teleconnections in the geopotential height field during the Northern Hemisphere winter." In: *Monthly weather review* 109.4, pp. 784–812.
Wang, Lei and Mingfang Ting (2022). "Stratosphere-Troposphere Coupling Leading to Extended Seasonal Predictability of Summer North Atlantic Oscillation and Boreal Climate." In: *Geophysical Research Letters* 49.2, e2021GL096362.

Weiland, Ruud Sperna, Karin van der Wiel, Frank Selten, and Dim Coumou (2021). "Intransitive Atmosphere Dynamics Leading to Persistent Hot–Dry or Cold–Wet European Summers." In: *Journal of Climate* 34.15, pp. 6303–6317.

Weisheimer, Antje, Damien Decremer, David MacLeod, Christopher O'Reilly, Tim N Stockdale, Stephanie Johnson, and Tim N Palmer (2019). "How confident are predictability estimates of the winter North Atlantic Oscillation?" In: *Quarterly Journal of the Royal Meteorological Society* 145, pp. 140–159.

Wolpert, David H, William G Macready, et al. (1995). *No free lunch theorems for search*. Tech. rep. Technical Report SFI-TR-95-02-010, Santa Fe Institute.

Woollings, Tim, Christian Franzke, DLR Hodson, B Dong, Elizabeth A Barnes, CC Raible, and JG Pinto (2015). "Contrasting interannual and multidecadal NAO variability." In: *Climate Dynamics* 45.1-2, pp. 539–556.

Woollings, Tim, Abdel Hannachi, and Brian Hoskins (2010). "Variability of the North Atlantic eddy-driven jet stream." In: *Quarterly Journal of the Royal Meteorological Society* 136.649, pp. 856–868.

Wu, Zhiwei and Hai Lin (2012). "Interdecadal variability of the ENSO–North Atlantic Oscillation connection in boreal summer." In: *Quarterly Journal of the Royal Meteorological Society* 138.667, pp. 1668–1675.

Wulff, C Ole, Richard J Greatbatch, Daniela IV Domeisen, Gereon Gollan, and Felicitas Hansen (2017). "Tropical forcing of the summer East Atlantic pattern." In: *Geophysical Research Letters* 44.21, pp. 11–166.

Zha, Jinlin, Cheng Shen, Deming Zhao, Jinming Feng, Zhongfeng Xu, Jian Wu, Wenxuan Fan, Meng Luo, and Liya Zhang (2022). "Contributions of External Forcing and Internal Climate Variability to Changes in the Summer Surface Air Temperature Over East Asia." In: *Journal of Climate*, pp. 1–64.

Zubiate, Laura, Frank McDermott, Conor Sweeney, and Mark O'Malley (2017). "Spatial variability in winter NAO–wind speed relationships in western Europe linked to concomitant states of the East Atlantic and Scandinavian patterns." In: *Quarterly Journal of the Royal Meteorological Society* 143.702, pp. 552–562.

EIDESSTATTLICHE VERSICHERUNG - DECLARATION ON OATH

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.
I hereby declare upon oath that I have written the present dissertation independently and have not used further resources and aids than those stated.

Hamburg, October 2022

Julianna Carvalho Oliveira

Hinweis / Reference

Die gesamten Veröffentlichungen in der Publikationsreihe des MPI-M
„Berichte zur Erdsystemforschung / Reports on Earth System Science“,
ISSN 1614-1199

sind über die Internetseiten des Max-Planck-Instituts für Meteorologie erhältlich:
http://www.mpimet.mpg.de/wissenschaft/publikationen.html

All the publications in the series of the MPI -M
„Berichte zur Erdsystemforschung / Reports on Earth System Science“,
ISSN 1614-1199

are available on the website of the Max Planck Institute for Meteorology:
http://www.mpimet.mpg.de/wissenschaft/publikationen.html